

**MODELO DE DESARROLLO PARA EL ALMACENAMIENTO,
ANÁLISIS E INTERPRETACIÓN DE INFORMACIÓN APLICADO EN
UN SISTEMA DE INFORMACIÓN PARA VIGILANCIA
EPIDEMIOLÓGICA
"MAISIVE"**



**Ricardo Emilio Lombana Quiñones
Richard Andrés Rojas Rosero**

Director: Ing. Diego Mauricio López Gutiérrez

Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Departamento de Telemática
Línea de Investigación en Ingeniería de Sistemas Telemáticos
Popayán, Febrero de 2004

*A mi padre por su paciencia y por creer en mi,
A mi madre por su voz de aliento en todo momento,
A mis hermanos por inspirarme a continuar,
A toda mi familia por mantenerme en su corazón,
A mis amigos por los momentos vividos y por los que vendrán,
A Carolina por soportarme y complementarme la vida,
A todos, muchas gracias por se así...*

Ricardo Lombana

A Dios.

A Yolanda, mi madre.

A Carlos Alirio, mi padre

A Nilssa, Jairo, Ivan y Doris, mis hermanos

A Libia, mi otra mamá

A toda mi familia

A Mary Luz

A mis amigos.

Richard Rojas

AGRADECIMIENTOS

Los autores manifiestan su agradecimiento:

A Dios,

A la FIET por el conocimiento inculcado,

A Diego Mauricio López por apoyarnos en la propuesta de este trabajo de investigación,

A Juan Carlos Corrales por su interés y aporte en la culminación de este proyecto,

Al Dr. Milton Lombana por su asesoría y colaboración.

TABLA DE CONTENIDO

INTRODUCCIÓN.....	1
1. ESTADO DEL ARTE DE LOS SISTEMAS DE VIGILANGIA EPIDEMIOLÓGICA	5
1.1. CONCEPTOS Y GENERALIDADES.....	5
1.1.1. Definiciones	5
1.1.2. Sistemas de Información Epidemiológica	7
1.1.3. Panorama Mundial	9
1.1.4. Panorama Nacional	11
1.1.5. Panorama Local.....	21
1.2. HERRAMIENTAS PARA ANÁLISIS DE INFORMACIÓN EPIDEMIOLÓGICA	22
1.2.1. SIG EPI.....	22
1.2.2. EPI INFO	23
1.2.3. EPIDAT	25
1.2.4. STATA.....	26
1.2.5. SIGSA	28
2. CONCEPTOS TECNOLÓGICOS	30
2.1. DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS (KDD)	30
2.2. ALMACENES DE DATOS (DATA WAREHOUSING)	32
2.2.1. Concepto de Data Warehousing.....	32
2.2.2. Diferencias entre un OLTP y un DW.....	34
2.3. MINERÍA DE DATOS (DATAMINING).....	35
2.3.1. Fundamentos de Datamining	35
2.3.2. Alcance del Datamining.....	36
2.3.3. Beneficios claves del uso de Datamining.....	37
3. MODELO DE DESARROLLO PARA EL ALMACENAMIENTO, ANÁLISIS E INTERPRETACIÓN DE INFORMACIÓN EPIDEMIOLÓGICA.....	39
3.1. DESCRIPCIÓN DEL MODELO	39
3.1.1. Fundamentos del Modelo	39
3.1.2. Estructura para Descripción del Sistema	40
3.1.3. Modelo del Proceso de Desarrollo	40
3.1.3.1. Características principales del proceso de desarrollo	40
3.1.3.2. Fases de Referencia	42
3.1.4. MODELO DE ORGANIZACION DEL TALENTO	43
3.2. MODELO DE DESARROLLO	44
3.2.1. FASE I: Estudio Preliminar	44
3.2.1.1. Descripción General de la Fase	44
3.2.1.2. Etapas de Referencia	45
3.2.2. FASE II: Diseño del almacén de datos	64
3.2.2.1. Descripción General de la Fase	64

3.2.2.2.	Etapas de Referencia	65
3.2.3.	FASE III: Implementación del almacén de datos	89
3.2.3.1.	Descripción General de la Fase	89
3.2.3.2.	Etapas de Referencia	90
3.2.4.	FASE IV: Analisis de información	101
3.2.4.1.	Descripción General de la Fase	101
3.2.4.2.	Etapas de Referencia	102
3.2.5.	FASE V: Despliegue e interpretación.....	123
3.2.5.1.	Descripción General de la Fase	123
3.2.5.2.	Etapas de Referencia	124
3.2.6.	FASE VI: Evaluación	129
3.2.6.1.	Descripción General de la Fase	129
3.2.6.2.	Etapas de Referencia	130
4.	DESCRIPCIÓN DE LA IMPLEMENTACIÓN DEL MAISIVE	135
5.	CONCLUSIONES Y RECOMENDACIONES	140
BIBLIOGRAFIA.....	144

INDICE DE TABLAS

Tabla 1.	Concepto de Data warehouse.....	33
Tabla 2.	Diferencias entre un sistema operacional y un Data warehouse.....	35
Tabla 3.	Estructura para Descripción del Sistema con soporte para la toma de decisiones.....	40
Tabla 4.	Descripción de las Fases de Referencia.....	42
Tabla 5.	Estructura del Grupo de gestión	43
Tabla 6.	Estructura del Grupo de desarrollo	43
Tabla 7.	Notación UML para actores y casos de uso de negocio.....	51
Tabla 8.	Notación de modelado UML para Trabajadores y Entidades de Negocio.....	51
Tabla 9.	Plantilla para la descripción de riesgos.....	57
Tabla 10.	Categorización de recursos computacionales	60
Tabla 11.	Descripción de Escenarios a nivel del Estudio Preliminar	70
Tabla 12.	Plantilla para la descripción de requisitos	73
Tabla 13.	Cuadro comparativo de SGBD.....	78
Tabla 14.	Matriz de Decisión.....	79

INTRODUCCIÓN

Actualmente se plantea la necesidad de redefinir la salud pública como “nueva”, con lo cual se quiere referir a las situaciones caracterizadas por avances en el conocimiento de los problemas de salud, en la tecnología y la legislación sobre temas afines con salud, con cambios en la organización y gestión de las instituciones encargadas de la salud colectiva.

La Organización Mundial de la Salud, desde esta nueva perspectiva ha impulsado la promoción de la salud y la prevención de las enfermedades, dando un énfasis primordial al autocuidado de la salud, a la auto e inter ayuda entre la población, a la creación de entornos favorables y a la organización de servicio comunitarios de salud. [LVE 00].

La vigilancia en salud pública se entiende como un proceso continuo y sistemático de obtención de información, análisis, interpretación y divulgación, con la clara intencionalidad de generar acciones de promoción de la salud, prevención o control, bien sea en el momento mismo, en el mediano plazo o en el largo plazo.

Al considerar la salud pública, como el análisis y transformación de las condiciones de salud, mediante respuestas sociales poblacionales; es necesario hacer referencia a los tres componentes esenciales que la constituyen: la vigilancia demográfica, la vigilancia epidemiológica y la vigilancia de los sistemas de salud. [JSC 94].

La vigilancia epidemiológica según el centro para el control de enfermedades de Atlanta, “es la recolección sistemática, el análisis e interpretación de información de salud esencial para la planeación, implantación y evaluación de la práctica de salud pública, así como la diseminación oportuna de estos datos entre los que necesitan conocerlos”. Estos datos analizados e interpretados deben llevar al planteamiento de acciones para la prevención y para el control.

Existen algunos elementos previos al diseño del sistema de vigilancia epidemiológico que son indispensables para contribuir al logro de los objetivos, entre ellos se tienen un sistema de información en funcionamiento, el apoyo y respaldo de los funcionarios de salud encargados de la

toma de decisiones y basados en la motivación, en el conocimiento de las utilidades y beneficios de la vigilancia, una adecuada capacitación en el área de vigilancia del personal de salud que participa en el mismo y la existencia de una estructura administrativa que permita y facilite el desarrollo de las intervenciones planteadas en el sistema de vigilancia epidemiológica.

La vigilancia debe ser específica para un país, un departamento, un municipio, una región o una localidad, ya que responde a una situación y condiciones de salud también específicas. Generalmente, es difícil plantear en forma adecuada las acciones de vigilancia para una población definida en un tiempo dado, por lo tanto, hay que tener en cuenta la selectividad y la flexibilidad al plantearlas con el fin de optimizar los recursos asignados para realizarla.

Los datos de los eventos epidemiológicos deben ser analizados para constituirse en información o insumo útil para la toma de decisiones, ejecución y evaluación de intervenciones orientadas a la promoción, prevención, curación y recuperación del estado de salud bien sea a nivel individual o colectivo.

Sin embargo, aunque nos encontramos en la denominada “sociedad de la información”, debido a que se destina gran cantidad de recursos en la recolección, almacenamiento, procesado y análisis de información, el conocimiento más valioso no lo podemos encontrar a simple vista, sino que suele aparecer oculto entre los datos recogidos en forma de patrones o reglas que relacionan entre sí otras partes más superficiales de la información.

Este conocimiento se ha venido obteniendo, tradicionalmente, mediante análisis manual, aplicando la inferencia inductiva sobre el conjunto de datos de partida, sin embargo, la recolección y almacenamiento de los datos se realiza a un ritmo cada vez mayor.

La información que manejan los actuales sistemas de vigilancia en salud pública son susceptibles de generar un volumen de datos imposible de analizar de forma manual y es por eso que se cuenta cada vez con computadoras potentes para automatizar el proceso inductivo, para analizar de forma inteligente los datos existentes y extraer de ellos ese conocimiento oculto y de gran importancia en el campo de la Salud pública.

La toma de decisiones con relación a las actividades del sistema de vigilancia debe ser analizada y sustentada teniendo en cuenta los objetivos planteados; como es un proceso continuo y dinámico se requiere que periódicamente se definan o redefinan las prioridades y se evalúen las metas propuestas.

Es muy importante para las organizaciones prestadoras del servicio de salud pública que pretendan implementar un sistema de soporte a la toma de decisiones tener un modelo de negocio consistente y que, además, sus operaciones estén soportadas sobre aplicaciones eficientes que permitan el óptimo manejo de la información para su correcto análisis e interpretación.

Cuando se construyen aplicaciones software es de invaluable ayuda el contar con un modelo adecuado de desarrollo para dicho fin puesto que permite a los involucrados tener claros los puntos a seguir para la consecución de software con alto nivel de calidad y que se ajuste eficientemente a las necesidades de la organización. El propósito de un modelo de desarrollo para el almacenamiento, análisis e interpretación de información epidemiológica (MAISIVE) es garantizar el éxito de la aplicación de soporte a la toma de decisiones, sustentada en la definición de los procesos del sistema y en los criterios para la escogencia de las tecnologías a emplear para su desarrollo.

El MAISIVE tiene como prioridad la entrega de un modelo de desarrollo que aporte tanto a la implementación de servicios de Información, como a la base de conocimientos universal en el área de servicios de información para Salud pública del Departamento del Cauca, teniendo en cuenta aspectos importantes de planificación, requerimientos, análisis y diseño.

Paralelamente el MAISIVE contribuye significativamente a los objetivos establecidos para el Programa No. 3, Generación de Referencias Metodológicas para el Desarrollo de Soluciones, del Plan de Trabajo 1998-2003 del GIT, constituyendo también un aporte significativo para el nodo de Infraestructura Metodológica del Departamento de Telemática de la Universidad del Cauca, además de su evidente contribución al sistema de información para vigilancia epidemiológica (SIGSA) que tiene implementado la Dirección Departamental de Salud del Cauca.

El documento del trabajo de grado se encuentra estructurado de la siguiente manera:

En primera instancia, se encuentra la introducción que contiene la presentación y el enfoque que se adoptó en la realización del trabajo de grado.

En el primer capítulo se describen los conceptos y generalidades sobre la vigilancia epidemiológica en salud pública. Se presentan los panoramas relacionados con los sistemas de vigilancia epidemiológica a nivel mundial, nacional y local, además de una revisión de herramientas existentes para el análisis de información.

El segundo capítulo presenta la descripción general y detallada del modelo de desarrollo propuesto. Este comprende los pilares conceptuales del modelo, los componentes del proceso de desarrollo, y el modelo de referencia organizacional. Se expone el proceso de desarrollo, constituido por fases de Referencia, la infraestructura de actividades, y los productos de trabajo.

El tercer capítulo contiene una descripción general de los procesos de almacenamiento, análisis e interpretación de información aplicados al sistema de vigilancia epidemiológica con el fin de validar y ejercitar el MAISIVE.

Finalmente se exponen las conclusiones y recomendaciones del trabajo de grado realizado.

Como anexos se tienen los siguientes documentos:

Anexo A – Marco Conceptual. Contiene una descripción detallada de los aspectos conceptuales que sirvieron como guía y soporte para la construcción del MAISIVE.

Anexo B – Descripción de los procesos de almacenamiento, análisis e interpretación de información implementados en el sistema de vigilancia epidemiológica. Contiene los productos obtenidos en cada fase del desarrollo del sistema que se construyó para validar y ejercitar el MAISIVE.

1. ESTADO DEL ARTE DE LOS SISTEMAS DE VIGILANCIA EPIDEMIOLÓGICA

1.1. CONCEPTOS Y GENERALIDADES

1.1.1. DEFINICIONES

La palabra vigilancia se refiere a la actividad de ver (en una persona o un área) presumiblemente con el propósito de dirigir, supervisar o controlar. En salud, la vigilancia se ha utilizado para identificar aquellos aspectos importantes que se relacionan con la generación de enfermedades o alteraciones de los procesos normales.

El perfil epidemiológico ha ido cambiando en los países en vías de desarrollo, de ahí la importancia de reflexionar sobre la filosofía y el papel de la salud pública, el cual en la actualidad esta orientado hacia la promoción de la salud y a la prevención y control de los nuevos retos que afronta.

Actualmente se plantea la necesidad de redefinir la salud pública como “nueva”, con lo cual se quiere referir a las situaciones caracterizadas por avances en el conocimiento de los problemas de salud, en la tecnología y la legislación sobre temas afines con salud, con cambios en la organización y gestión de las instituciones encargadas de la salud colectiva.

El informe de Marc Lalonde(1974) es considerado como la primera declaración contemporánea relacionada con la nueva salud pública. Lalonde, en este informe señala la importancia de dirigir las acciones de salud hacia la prevención “.....las principales causas de enfermedad y muerte...tienen sus raíces en tres elementos de este concepto: biología humana, ambiente y estilo de vida, Es evidente, por lo tanto, que las vastas sumas gastadas en tratar las enfermedades podrían ser asignadas en primer lugar a la prevención.” [LVE 00].

La Organización Mundial de la Salud, desde esta nueva perspectiva ha impulsado la promoción de la salud y la prevención de las enfermedades, dando un énfasis primordial al autocuidado de la

salud, a la auto e inter ayuda entre la población, a la creación de entornos favorables y a la organización de servicio comunitarios de salud. [LVE 00].

La gestión administrativa reconoce que una manera de elevar su eficiencia está en hacer el mejor uso de los recursos de información que ya existen dentro de la organización. Sin embargo, a pesar de que esto se viene intentando desde hace muchos años, no se tiene todavía un uso efectivo de los mismos. La razón principal es la manera en que han evolucionado las computadoras, basadas en las tecnologías y sistemas de información. La mayoría de las organizaciones hacen lo posible por conseguir buena información, pero el logro de ese objetivo depende fundamentalmente de su arquitectura actual, tanto del hardware como del software.

La vigilancia epidemiológica según el centro para el control de enfermedades de Atlanta, “es la recolección sistemática, el análisis e interpretación de información de salud esencial para la planeación, implantación y evaluación de la práctica de salud pública, así como la diseminación oportuna de estos datos entre los que necesitan conocerlos”. Estos datos analizados e interpretados deben llevar al planteamiento de acciones para la prevención y para el control [MVP 04].

A medida que los países se desarrollan, se generan situaciones que afectan el estado de salud poblacional, es así como en la actualidad se tiene que dar una respuesta desde salud a la coexistencia de enfermedades infecciosas y enfermedades crónicas y a otros problemas como la aparición de SIDA, la contaminación ambiental, la aparición de nuevos virus y enfermedades, la resistencia a los antibióticos, entre otros, los cuales deben ser identificados rápidamente y monitoreados por el sistema de vigilancia.

Teniendo en cuenta que para efectuar vigilancia epidemiológica el insumo indispensable son los datos y la información, los elementos de la misma son aquellos que conducen y permiten el planteamiento de las intervenciones requeridas. Algunos de ellos son:

- La producción, la recolección y la notificación de datos e información
- La tabulación, consolidación, procesamiento y almacenamiento de datos e información
- El análisis de datos e información
- La divulgación, publicación y distribución de los datos e información

- La evaluación tanto de los datos e información como de las intervenciones planteadas
- Información demográfica, ambiental y de condiciones de vida
- Resultados de investigaciones epidemiológicas o en salud
- Registros específicos utilizados en salud

Existen algunos elementos previos al diseño del sistema de vigilancia epidemiológico que son indispensables para contribuir al logro de los objetivos, entre ellos se tienen un sistema de información en funcionamiento, el apoyo y respaldo de los funcionarios de salud encargados de la toma de decisiones y basados en la motivación, en el conocimiento de las utilidades y beneficios de la vigilancia, una adecuada capacitación en el área de vigilancia del personal de salud que participa en el mismo y la existencia de una estructura administrativa que permita y facilite el desarrollo de las intervenciones planteadas en el sistema de vigilancia epidemiológica.

Los datos de los eventos epidemiológicos deben ser analizados para constituirse en información o insumo útil para la toma de decisiones, ejecución y evaluación de intervenciones orientadas a la promoción, prevención, curación y recuperación del estado de salud bien sea a nivel individual o colectivo.

Para alcanzar el logro de los objetivos del sistema de vigilancia epidemiológica se requiere que este funcione como un sistema de información.

1.1.2. SISTEMAS DE INFORMACIÓN EPIDEMIOLÓGICA

Para el efectivo funcionamiento del sistema de vigilancia es fundamental asegurar la calidad, la confiabilidad, la oportunidad, la cantidad, la relevancia y la actualidad de la información epidemiológica en salud; esta información debe reflejar las necesidades reales de la comunidad y del ambiente para que sea utilizada efectivamente en la toma de las decisiones.

Los insumos requeridos para el adecuado funcionamiento del sistema de información son lógicamente los datos y la información.

El trabajo consiste en recolectar, organizar, analizar y contextualizar los datos para transformarlos en información, la cual mediante un proceso de análisis, evaluación y obtención de conclusiones útiles se convierte en conocimiento. De esta manera el conocimiento se transforma en el nuevo factor de poder económico y en el activo más importante de la estructura social.

La información es considerada como una representación de la realidad, como la expresión de conocimientos, la base para el funcionamiento de las organizaciones, el soporte para los procesos gerenciales, un recurso crítico de las organizaciones y un eslabón que une todos los componentes de la organización

Al estructurar los datos e interpretarlos se obtiene una información que tiene sentido general, sin embargo esta información solo podrá ser utilizada por aquellas personas que tienen unas determinadas expectativas y criterios para su interpretación y convertirla en conocimiento.

En salud el objetivo es crear condiciones para generar la información y distribuirla en forma adecuada a los actores en salud, que al juzgar su importancia la valoren para emplearla en el desarrollo de la salud del individuo, la familia y la comunidad y de los espacios en que se desenvuelven.

Un sistema de información desarrolla cuatro actividades básicas: entrada, almacenamiento, procesamiento de datos y salida de información [MVP 04].

Entrada de información

Proceso en el cual se toman los datos requeridos para procesar la información, las entradas se pueden hacer manual o automáticamente. En la primera el usuario aporta la información directamente y en la segunda, los datos provienen de otros sistemas. En salud, constituye datos de entrada la caracterización del usuario, como identificación, localización, tipo de afiliación, estado de salud, etc.

Almacenamiento de la información

Es un proceso en el cual se guarda la información en archivos que pueden ser recuperados en cualquier momento. La historia clínica del usuario es el almacenamiento de la información referente al mismo.

Procesamiento de la información

Permite la transformación de los datos fuentes en resultados por la aplicación de mecanismos o indicadores que soporten la toma de decisiones. En consulta externa el cálculo de indicadores sobre el desarrollo de proyectos desarrollados como porcentajes de citología tomadas en un período determinado, son producto del procesamiento de datos.

Salida de información

Es la capacidad de un sistema para sacar la información procesada hacia otro sistema o usuario. Un informe estadístico sobre un indicador es una salida de información que facilita la planeación de actividades en un área específica

1.1.3. PANORAMA MUNDIAL

La Organización Mundial de la Salud definió la vigilancia epidemiológica como: “El escrutinio permanente y la observación activa de la distribución y propagación de las infecciones y de factores relacionados con suficiente exactitud en cantidad y calidad para un control eficaz”. En esta definición se enfatiza en la necesidad de contar con un insumo generado en condiciones de calidad y cantidad para poder de esta manera realizar el control eficaz acorde con la situación identificada.

El Centro de Control de Enfermedades de Atlanta planteó la vigilancia epidemiológica como “información para la acción” definición a la cual el doctor Llopis, asesor de la Organización Panamericana de la Salud, complementó como “información epidemiológica para la acción”.

A través de estos conceptos se identifican actividades básicas de vigilancia como son:

- Recolección sistemática de datos de enfermedades.
- Procesamiento y evaluación de datos
- Análisis epidemiológico de datos
- Formulación de soluciones y medidas de control para los diferentes niveles responsables de las mismas

Así mismo, se plantea que la vigilancia va más allá de la simple recolección de datos sobre enfermedades porque sin un análisis y publicación la información no podrá ser efectiva y utilizada en forma apropiada. Los análisis deben ser rápidos, precisos y completos y la publicación puede ser simple o compleja, incluyendo el análisis respectivo.

En el documento “Programa Análisis de la Situación en Salud y sus Tendencias” la Organización Panamericana de la Salud, plantea que la vigilancia epidemiológica debe tener como función principal el control de enfermedades y riesgos y en particular la capacidad de respuesta de los servicios. Insiste en que la vigilancia no debe limitarse a la recolección, procesamiento y acumulación de datos e informaciones que no producen respuestas satisfactorias. Las técnicas de la vigilancia deben ser acordes con las características epidemiológicas del problema y a las condiciones, necesidades y capacidad de respuesta de cada país y sistema.

La vigilancia epidemiológica parte de una activa consecución de datos e información sobre los eventos, los cuales deben ser analizados para constituirse en información o insumo útil para la toma de decisiones, ejecución y evaluación de intervenciones orientadas a la promoción, prevención, curación y recuperación del estado de salud bien sea a nivel individual o colectiva.

Al plantear un sistema de vigilancia epidemiológica se debe recordar que:

- Tiene como finalidad la prevención y control de los problemas y situaciones que modifican el estado de salud de la población.
- La aplicación exige de un trabajo activo y permanente, multidisciplinario, intersectorial ya que los problemas y sus soluciones requieren de la participación de diferentes sectores.

- Se puede aplicar a cualquier evento en salud como una enfermedad, un factor de riesgo o una discapacidad.
- El proceso de la vigilancia epidemiológica es más eficaz, en la medida en que se parte de la percepción poblacional sobre sus problemas y situación de salud y se cuenta con su participación en el desarrollo, evaluación y control del mismo.

La estructuración de un sistema de vigilancia epidemiológica, facilita y contribuye a que se efectúe la prevención y un control eficaz de los eventos que modifican o alteran el proceso de salud de una comunidad, el cual a su vez depende de la capacidad de respuesta efectiva y eficaz y para ello es necesario contar con un sistema de vigilancia epidemiológica funcional.

Para alcanzar el logro de los objetivos del sistema de vigilancia epidemiológica se requiere que este funcione como un sistema integrado por tres componentes principales: uno de entrada que esta conformado por los datos epidemiológicos y estadísticos; uno de proceso en el cual debe efectuarse el análisis de la información recolectada y la selección de alternativas de intervención y finalmente uno de salida o de producto donde se plantean los resultados obtenidos a partir de la ejecución de las intervenciones y la evaluación de los cambios de la situación de salud.

1.1.4. PANORAMA NACIONAL

La vigilancia epidemiológica es considerada como la expresión más conocida de la actividad epidemiológica y en el contexto de salud pública, es el proceso que provee la información proveniente de las unidades notificadoras de todos los niveles, para orientar la toma de decisiones y planificar las intervenciones de prevención y control [MVP 04].

En Colombia con la ley 100 de 1993, se estableció el Sistema General de Seguridad Social en Salud y la oferta de servicios de salud se organiza en empresas como: instituciones prestadoras del servicio de salud (IPS), empresas promotoras de salud (EPS), empresas sociales del estado (ESES), aseguradoras de riesgos profesionales (ARP), con paquetes como el plan obligatorio de salud (POS), y el plan de atención básica (PAB), entre otros.

En el marco de la ley 100/93, la vigilancia epidemiológica es considerada como una herramienta de utilidad de salud pública y de los planes de beneficio, ya que permite: conocer el comportamiento de los eventos que se relacionan con el proceso salud-enfermedad, caracterizar la población beneficiaria de los planes de seguridad social en salud, para desarrollar intervenciones orientadas a la promoción, prevención y control de la salud y la enfermedad.

Para dar cumplimiento a lo anterior, cada uno de los actores del sistema tienen responsabilidades específicas, es así como:

El ministerio de salud

Debe dirigir y controlar el sistema nacional de vigilancia en salud pública a través de la normatización, asesoría técnica y financiera, planteamiento de lineamientos y evaluación.

Las direcciones seccionales de salud

Son las encargadas de la gerencia del sistema de vigilancia epidemiológica a este nivel, adoptar y adaptar las normas de carácter nacional, apoyar las direcciones locales en el control de brotes y emergencias, prestar asesoría a las direcciones locales y a las instituciones prestadoras del servicio de salud en cuanto a las intervenciones de vigilancia y evaluación se refiere.

Las direcciones locales de salud

Son las responsables de la gerencia, asesoría, vigilancia y control del sistema de vigilancia epidemiológica; así como de garantizar la implementación de las acciones del plan de atención básica municipal.

Las empresas promotoras de salud

Son las responsables por las acciones individuales de diagnóstico, tratamiento, notificación y rehabilitación de eventos sujetos a vigilancia epidemiológica de los afiliados y beneficiarios.

Las instituciones prestadoras de servicios de salud

Son las encargadas de operativizar las responsabilidades de las empresas promotoras de salud, la notificación, aplicación de protocolos de vigilancia epidemiológica, asesoría técnica a nivel institucional y análisis de la información para orientar y definir las intervenciones.

Las aseguradoras de riesgos profesionales

Son las responsables de los planes y programas de salud ocupacional y de las intervenciones de vigilancia en esta área.

Los individuos

Son los responsables por su autocuidado y su participación activa en las intervenciones de vigilancia epidemiológica.

El sistema de vigilancia epidemiológica, requiere que se le suministre la información necesaria y a su vez debe estar en capacidad de suministrarla oportunamente y con las características necesarias para facilitar la toma de decisiones en el proceso de planeación de las intervenciones según las necesidades de los diferentes actores que participan en el sistema.

En Colombia el decreto 1562 del 22 de junio de 1984, reglamenta parcialmente los títulos VII, XI de la ley 9 de 1979, en cuanto a la vigilancia y control epidemiológico y medidas de seguridad. En su capítulo IV artículo 27, normatiza la obligatoriedad de la información epidemiológica así: “La información epidemiológica es obligatoria para todas las personas naturales o jurídicas, residentes o establecidas en el territorio nacional, dentro de los términos de responsabilidad, clasificación, periodicidad, destino y claridad, señalados en el presente decreto y las disposiciones especiales que con fundamento en el mismo o en la Ley dicte el Ministerio de Salud”. [MVP 04].

En el artículo 34 determina la obligatoriedad de informar en los siguientes términos: “Toda persona natural o jurídica deberá informar en el menor tiempo posible, a la autoridad de salud más cercana, cualquier fenómeno que se presente en las personas, animales y medio ambiente, que en una u otra forma comprometa o pueda comprometer la salud de la población”.

Por lo anterior, las instituciones prestadoras del servicio de salud, las empresas promotoras de salud, direcciones de salud deben aportar la información a través de la notificación epidemiológica que constituye el insumo básico para la formulación y desarrollo de las medidas de prevención y control epidemiológico.

La reforma en seguridad social en Colombia, introdujo cambios estructurales y de funcionamiento en el sector salud, las cuales han llevado a la reformulación de procesos y actividades de los diferentes actores del sistema. Como consecuencia de lo anterior, se han generado nuevos requerimientos de información y es indispensable que se adecue el subsistema de información, para lograr una gestión coherente con el desarrollo de la reforma de seguridad social en salud.

De los diferentes diagnósticos del subsistema de información, se han evidenciado coincidencias en algunos aspectos tales como:

- Ausencia de un plan de desarrollo, producto de la concertación con los diferentes actores.
- Los procesos funcionales como los modelos de análisis de información del sistema general de seguridad social en salud, se encuentran actualmente en permanente adecuación.
- Los requerimientos informacionales se caracterizan por ser: diversos, duplicados y no racionales, en los niveles de dirección, administración y control.
- Existen algunos instrumentos estandarizados de recolección y flujo de datos.
- Hay voluntad política y técnica para su desarrollo.
- Se realiza la asignación de recursos financieros y adquisición de recursos tecnológicos para mantener su desarrollo.

Analizando los aspectos anteriores, es urgente la necesidad de efectuar adecuación al subsistema de información en lo siguiente:

- La generación de información debe apoyar en primera instancia la gestión local.
- El registro de la información debe orientarse hacia el usuario y hacia las operaciones del sistema.
- La información debe tener un nivel de desagregación, de tal manera que facilite la utilización por todos y cada uno de los actores del sistema.
- Hay necesidad de modernizar los procesos de información, apoyándose en la tecnología de las comunicaciones y computacional, previo rediseño de los procesos funcionales teniendo en cuenta el sistema general de seguridad social en salud.

- A través de la concertación, unificar los requerimientos informacionales necesarios como insumo para los diferentes niveles de dirección,
- Propiciar el desarrollo de la cultura informacional y la aplicación de la tecnología.

Como consecuencia de lo anterior el Ministerio de Salud, a través de la Resolución No. 2542 de julio de 1998, diseño un sistema único de Información, Sistema Integral de Información en Salud (S.I.I.S), en el cual se define el sistema de información como el "conjunto de usuarios, normas, procedimientos, recursos físicos, lógicos y financieros, que interrelacionados registran, procesan, almacenan, recuperan y distribuyen información para soportar la operación, la toma de decisiones y el control del sistema general de seguridad social en salud".

Cuando todos los actores del sistema de salud actúen con base en la cultura de la información, se estará contribuyendo a garantizar una información que permita el seguimiento y control y se constituya en apoyo los procesos gerenciales e institucionales, facilitando la adecuada y eficiente toma de decisiones, con el fin de responder a los fundamentos del servicio público en salud y de esta forma el Ministerio con la función otorgada a través de la Ley 100 en su artículo 173, en el cual expresa que "debe reglamentar la recolección, transferencia y difusión de la información en el subsistema al que concurren obligatoriamente todos los integrantes del sistema general de seguridad social en salud".

El Sistema de información de una organización es considerado como uno de los elementos fundamentales para la evaluación efectiva de los programas y proyectos en las Instituciones de servicios. Tanto la información de planeación como la desempeño debe provenir de un sistema de información bien diseñado, estructurado y en funcionamiento, que se convierta en un sistema de apoyo, abierto, moderno y facilite su aplicación en el sistema de Vigilancia Epidemiológica.

El Sistema de Vigilancia en Salud Pública, **SIVIGILA**, hace referencia a la relación organizada de usuarios, normas, procedimientos y recursos (financieros, técnicos y humanos) para la recopilación de datos, análisis, interpretación y divulgación de información sobre eventos de salud, de forma sistemática y continua, para su uso esencial en la orientación de la toma de decisiones en salud pública [MDS 99].

El SIVIGILA es definido como:

"Artículo 2. – Definiciones. Adóptense las siguientes definiciones técnicas para el Sistema de Vigilancia en Salud Pública:

a) Salud Pública: Hace referencia al conjunto de prácticas sociales y funciones del estado existentes en una sociedad, tendientes a desarrollar, preservar o mejorar la salud colectiva, a través de las condiciones, determinantes y significados de la salud para la población. En este contexto, la salud se constituye en un bien público, de interés y responsabilidad de todos, que debe producirse desde la sociedad y el estado.

b) Vigilancia en Salud Pública: Hace referencia al proceso sistemático y constante de la recolección, el análisis y la interpretación de datos específicos, así como la divulgación de información para utilizarla en la orientación, planificación, ejecución y evaluación de la práctica en salud pública. Como tal esta definición responde a una evolución conceptual de la vigilancia epidemiológica.

c) Eventos en Salud: Es el conjunto de sucesos o circunstancias que pueden modificar o incidir en la situación de salud de una comunidad. Se entenderán como eventos en salud para efectos del presente decreto enfermedades, lesiones de causa externa, causas de muerte, factores protectores, factores de riesgo, acciones de protección específica y de detección temprana de enfermedades, muestras biológicas o químicas y demás determinantes asociados.

d) Autoridad en Salud: Se consideran autoridades en salud, a todos aquellos que tienen la función de tomar decisiones de orden público en salud dentro del sector. Son autoridades en salud a Nivel Nacional, el Ministro de Salud, el Director General de promoción y Prevención, el Director del INS, el Director del INVIMA, y el Superintendente Nacional de Salud; a nivel de las entidades territoriales los Gobernadores, Alcaldes, Secretarios de Salud o quien haga sus veces, así como los funcionarios a quienes se les hayan asignado funciones de vigilancia y control del orden público en salud y del servicio público de la seguridad social en salud.

e) Usuarios del Sistema: Es el conjunto de personas y organizaciones dentro y fuera del sector salud y el Sistema Integral de Seguridad Social, que participan en el funcionamiento del Sistema en

uno o varios componentes y que requieren de la información allí construida para tomar decisiones que afectan la salud de la población.

f) Factores de Riesgo / Protectores del Ambiente: Se definen como tales a aquellas condiciones físicas, biológicas, químicas, y ergonómicas del medio ambiente de carácter natural o antropogénico, que determinan, en la población expuesta a ellos, una mayor / menor probabilidad de aparición de enfermedad o lesiones de causa externa.

g) Factores de Riesgo / Protectores del Consumo: Son aquellos elementos que forman parte de alimentos, medicamentos, cosméticos, bebidas alcohólicas y no alcohólicas, y demás productos de uso y/o consumo humano, de origen natural, animal, vegetal o industrial, que determinan en la población que los consume una mayor / menor probabilidad de aparición de la enfermedad o de lesiones de causa externa.

h) Factores de Riesgo / Protectores del Comportamiento humano: Son aquellas circunstancias donde está comprometido el ser humano con sus rasgos de personalidad, pensamientos, actitudes, motivaciones, prácticas, reglas y normas, que determinan en la población una mayor / menor probabilidad de aparición de enfermedad o de lesiones de causa externa.

i) Protección específica: Hace referencia a la aplicación de acciones y/o tecnologías que permitan y logren evitar la aparición inicial de la enfermedad mediante la protección frente al riesgo.

j) Detección temprana: Hacer referencia a los procedimientos que identifican en forma oportuna y efectiva la enfermedad, facilitan su diagnóstico en estados tempranos, el tratamiento oportuno y la reducción de su duración y el daño que causan evitando secuelas, incapacidad y muerte.

k) Orden público en salud: Hace referencia al conjunto de medidas y condiciones establecidas por la ley, para garantizar la salud de la colectividad y prevenir o controlar la ocurrencia de enfermedades, lesiones o muerte en la población.

l) Prestadores de servicios de salud: Hace referencia al conjunto de consultorios, centros de salud, puestos de salud e IPS públicas o privadas de I, II o III nivel, dedicadas a la prestación de todas aquellas atenciones de promoción, prevención, diagnóstico, tratamiento y rehabilitación

dirigidas a las personas como parte del plan de beneficios del SGSSS o por fuera del plan de beneficios a personas no afiliadas al sistema.

(...)

Artículo 9.- Tipo de eventos a vigilar. Los eventos en salud sometidos a vigilancia dentro del Sistema de Vigilancia en Salud Pública, serán aquellos cuya importancia para la salud colectiva sea evidente, teniendo en cuenta criterios de frecuencia, gravedad, costo efectividad, posibilidades de prevención, transmisibilidad e interés público; que además requieren ser enfrentados con medidas de salud pertinentes de acuerdo con la variabilidad de su comportamiento en la población.

Artículo 10.- Componentes del Sistema de Vigilancia. Los eventos en salud a ser vigilados se organizan para efectos del Sistema en los siguientes componentes:

Componente de vigilancia de eventos transmisibles y de fuente común:

Se consideran aquellas enfermedades, que por sus condiciones de transmisibilidad, pueden ser de alto poder epidémico y/o en todo caso su ocurrencia se reconoce como un riesgo para la salud colectiva, dentro de estas se incluyen las enfermedades sujetas a vigilancia estricta en el ámbito internacional. Además se incluyen aquellas enfermedades cuya presencia se relaciona con una fuente común de exposición. En este componente se incluye como un módulo especial y con un modelo operativo específico, la vigilancia de enfermedades infecciosas nosocomiales.

Componente de vigilancia de la mortalidad evitable

Se vigila la ocurrencia de muertes que pudiesen haber sido evitadas si la población hubiese tenido acceso oportuno y adecuado a acciones de prevención primaria, detección temprana y/o tratamiento.

Componente de vigilancia sobre acciones de protección específica, detección temprana y atención de enfermedades de interés en salud pública

Se vigila la realización de acciones de protección específica y detección temprana en la población y por tanto la gestión que al respecto realizan las instituciones responsables de su realización y garantía en el Sistema General de Seguridad Social en Salud.

Componente de vigilancia de enfermedades no transmisibles

Se vigilan las enfermedades no transmisibles y/o las lesiones de causa externa, que por su impacto en la salud colectiva y por los altos costos de su atención, se consideran de importancia en salud pública para el país.

Componente de vigilancia de factores de riesgo/protectores del ambiente, consumo y del comportamiento

Se refiere a la vigilancia de aquellos eventos o situaciones relacionadas con una exposición individual o colectiva de la población a factores de riesgo/protectores del ambiente, del consumo y del comportamiento.

Componente de vigilancia por laboratorio

Da cuenta del proceso de vigilancia fundamentado en la realización de procedimientos de laboratorio, sobre especímenes de interés en salud pública.

Componente de vigilancia de eventos especiales en salud

Se refiere a la vigilancia de eventos en salud que por su naturaleza y por el objetivo mismo de implementación de la vigilancia hacen necesario el diseño de un modelo específico de operación que no es compatible con ninguno de los componentes ya diseñados en el Sistema; las estrategias utilizadas para la observación del evento podrían diferir entre una entidad territorial y otra y con frecuencia mucha información necesaria para la vigilancia se provee fuera del sector. El objetivo es el monitoreo de una intervención.

El Ministerio de Salud en coordinación con otras organizaciones del nivel nacional y con las entidades territoriales, establecerá los eventos en salud que serán incluidos para la vigilancia en todas las entidades territoriales, en cada componente del Sistema de Vigilancia en Salud Pública [MDS 99].

Artículo 11.- Uso de los resultados derivados de la vigilancia. Los resultados derivados de la operación de cada uno de los componentes del Sistema de Vigilancia en Salud Pública, podrán ser usados para:

- Conocer el comportamiento de los eventos en salud sujetos a vigilancia.
- Detectar oportunamente brotes o epidemias
- Documentar la distribución y propagación de los eventos en salud en vigilancia.
- Facilitar la investigación epidemiológica y de laboratorio.
- Definir medidas de promoción, prevención y control.
- Evaluar medidas de promoción, prevención y control implementadas.
- Planificar las acciones en salud pública.
- Identificar factores de riesgo / protectores.
- Identificar grupos poblacionales expuestos a riesgo y o factores protectores

El SIVIGILA desde su propósito de orientar la salud colectiva, responde a los siguientes principios:

- El sistema se organiza con el objeto de anticiparse a las posibles condiciones de riesgo/protección para la población, y en los casos específicos donde no es posible, para detectar tempranamente los eventos de salud que conduzcan al deterioro de la salud del grupo.
- Es preciso que el sistema opere permanentemente de tal forma que sea posible la valoración continua del comportamiento de los eventos bajo vigilancia y en consecuencia la construcción continua de los escenarios epidemiológicos de cada uno.

Con respecto al flujo de información de los componentes del SIVIGILA, el sistema se encuentra siempre en alerta, con el fin de detectar los casos de los componentes descritos y tomar decisiones inmediatas. Desde este punto de vista su periodicidad es diaria, sin embargo semanalmente existe la tarea de reportar al nivel nacional la totalidad de casos presentada por evento con el fin de unificar y estandarizar la información.

1.1.5. PANORAMA LOCAL

En el departamento se adoptan los lineamientos que presenta el SIVIGILA. Como iniciativa local se están utilizando al máximo las tecnologías de la información en el almacenamiento, análisis e interpretación de información epidemiológica a través del Sistema de Información Geográfica en Salud Pública (SIGSA).

En el proyecto "Análisis Espacial y Generación de Capas de Información Temática - Ambiental para los Modelos SIG en Salud - SIGSA" se desarrolló el Sistema de Información Geográfico en Salud mediante Convenio Interadministrativo entre la Universidad del Cauca y la Dirección Departamental de Salud del Cauca, que presenta la interfaz de usuario de generación de mapas temáticos y de registro de eventos en salud y de información medio-ambiental [DSC 03].

SIGSA se establece como un sistema de información geográfico aplicado a un sistema de vigilancia epidemiológica en el Cauca. Este sistema nos puede proveer información confiable, continua y sistematizada para el desarrollo de las actividades de salud pública, la investigación epidemiológica clínica y comunitaria, además de un desarrollo curricular universitario basado en la evidencia.

Mediante este sistema se despliegan los patrones de morbi-mortalidad distribuidos de una forma específica en diferentes regiones o municipios correlacionándolos con otros fenómenos locales. Además se tiene un mejor conocimiento, de cara a la toma de decisiones, de hacia dónde se difunden los problemas de salud a través del tiempo y de cuál es el significado que tienen las variables ambientales para la salud de la población.

En SIGSA se desarrollaron mapas temáticos con un cubrimiento actual de 8 municipios del Cauca (Timbío, Santander de Quilichao, Puerto Tejada, Silvia, Guapi, López de Micay, Timbiquí- y Popayán).

1.2. HERRAMIENTAS PARA ANÁLISIS DE INFORMACIÓN EPIDEMIOLÓGICA

1.2.1. SIG EPI

SIGepi es un Sistema de Información Geográfica (SIG) diseñado para aplicaciones en Epidemiología y Salud Pública. Ofrece una compilación de técnicas, procedimientos y métodos para el análisis de datos epidemiológicos. Los mismos se presentan de manera simplificada, en un ambiente amigable y en múltiples idiomas.

SIGepi es un producto desarrollado por el Área de Análisis de Salud y Sistemas de Información (AIS) de la Organización Panamericana de la Salud (OPS) como parte del Proyecto de Cooperación Técnica "Aplicación de los Sistemas de Información Geográfica en Epidemiología y Salud Pública".

- **Características**

- Creación de mapas temáticos utilizando medidas e indicadores de salud.
- Respuesta a preguntas sobre el impacto de variables espaciales y geográficas en problemas de salud
- Cálculo, estandarización y suavizamiento espacial de tasas
- Estadísticas descriptivas, correlación y regresión lineal.
- Identificación de Áreas y Poblaciones Críticas ó Prioritarias.
- Construcción de un índice compuesto como por ejemplo, índice de necesidades básicas insatisfechas en salud, índice de pobreza, etc
- Identificación y detección de conglomerados (cluster) espaciales y en tiempo-espacio.
- Medición de la asociación de factores de exposición ambiental/ecológicos y eventos de salud para estudios epidemiológicos de caso-control y cohortes.
- Evaluación de acceso a los servicios de salud.

- **Características del Sistema**

- Win 9x/NT/Me/2000/XP
- 128 Mb RAM

- CD-ROM
- Acrobat Reader

1.2.2. EPI INFO

Epi Info es un programa de dominio público diseñado por el Centro para el Control de Enfermedades de Atlanta (CDC) de especial utilidad para la Salud Pública. Tiene un sistema fácil para construir bases de datos, analizarlos con las estadísticas de uso básico en epidemiología y representarlos con gráficos y mapas.

La primera versión para MSDOS se realizó en 1982 y la última, la versión 6, en 1996. La versión para Windows salió en Junio de 2000 y posteriormente salió una nueva versión Epi Info 2002, la cual está en proceso de traducción a varios idiomas incluyendo el español que viene disponible con la versión de distribución.

Epi Info es actualmente una herramienta para el apoyo de la vigilancia epidemiológica, cálculo de tamaños muestrales, diseño de bases de datos con su respectivo análisis estadístico. Una ventaja de esta base de datos es su uso de dominio público bajo el auspicio de la Organización Panamericana de la Salud y la Organización Mundial de la Salud (OPS-OMS), el cual se puede acceder en diversos idiomas y su distribución es gratuita.

Epi Info está compuesto por un conjunto de programas que conforman este paquete estadístico.

- Subprograma EPED: es un sistema para programar textos o diseñar cuestionarios con apoyo de un tutorial o con el uso del manual de instrucciones incorporado al paquete de Epi Info.
- Subprograma ENTER: tiene la capacidad de crear archivos provenientes del EPED, cuyo propósito es servir de ficheros.
- Subprograma CHECK: permite proteger la introducción de los datos a fin de evitar errores de digitación.

- Subprograma ANALYSIS: permite bajo lectura de la base de datos o ficheros, realizar órdenes estadísticas o gráficas (listado, frecuencias, tablas, gráficos y estadísticos epidemiológicos).

En resumen, la ventaja de emplear estas herramientas en la investigación y en el trabajo del epidemiólogo, es que ofrecen velocidad, precisión y flexibilidad en el manejo de los datos.

- **Características de Epi Info 2002**

- Compatibilidad máxima con los estándares de la industria, incluyendo:
 - Microsoft Access y otras bases de datos mediante SQL y ODBC
 - Visual Basic, Versión 6
 - Browsers del World Wide Web y HTML
- Escalabilidad, de manera que organizaciones externas a la CDC puedan crear módulos adicionales
- Epi Map, el cual es un SIG (Sistema de Información Geográfico) compatible.
- NutStat, un programa de antropometría de nutrición que calcula porcentajes y z-uentas usando la CDC 2000 o la referencia del crecimiento de 1978 CDC/WHO
- Regresión Logística de Kaplan-Meier-Meier
- Análisis de Supervivencia, programa para el análisis exploratorio univariado de datos. El programa calcula tablas de vida usando las estimaciones Kaplan Meier, para la función de supervivencia y proporciona estadísticas resumen del tiempo de supervivencia media.

- **Características del Sistema**

- Win 9x/NT/Me/2000/XP
- 32 Mb RAM
- Se recomienda el procesador de 200 Mhz.
- Por lo menos 50 megabytes de espacio libre en disco duro

1.2.3. EPIDAT

Programa desarrollado por el Servicio de Información sobre Salud Pública de la Cancillería de Sanidad de Servicios Sociales de la Junta de Galicia (España) y el Programa de Análisis de la Situación de Salud y sus Tendencias (SHA) de la Organización Panamericana de la Salud.

Epidat 2.1 es un paquete informático dirigido a epidemiólogos y a otros profesionales de la salud que habitualmente manejan datos tabulados. Sus objetivos son complementar a otros paquetes estadísticos que utilizan bases de datos y ofrecer un instrumento que facilite y apoye la enseñanza de la epidemiología. Esta nueva versión multilingüe permite seleccionar entre Catalán, Español, Gallego, Inglés y Portugués el idioma de trabajo.

Esta versión está desarrollada en entorno Windows y consta de los siguientes módulos:

- Ajuste de tasas—Permite introducir los datos manualmente o importarlos en formato *.DBF. Pueden ajustarse simultáneamente tantas tasas como desee el usuario, tanto por el método directo como por el indirecto.
- Muestreo— Calcula el tamaño muestral mínimo para realizar estudios de estimación de prevalencias y medias, estudios caso-control y de cohortes, verificación de la calidad de lotes. También genera números pseudoaleatorios.
- Concordancia— Analiza la concordancia entre dos o múltiples observadores, con dos o múltiples categorías de clasificación. También contrasta hipótesis de igualdad de capas.
- Tablas— Analiza tablas 2x2 que provengan de estudios transversales, caso-control o cohortes, en tablas simples o estratificadas. En los estudios caso-control pareados admite hasta 4 controles por caso. Incorpora el método Bootstrap para el cálculo de intervalos de confianza. En tablas mxn realiza cálculos específicos si las categorías son ordenadas. Incorpora el método de captura-recaptura de Chandra-Sekar-Deming y el de detección de clusters de Knox.
- Pruebas diagnósticas— Calcula la sensibilidad, especificidad y valores predictivos de pruebas diagnósticas simples, en serie y en paralelo. En las pruebas combinadas realiza los cálculos tanto si dispone de los valores de las celdas como de valores de sensibilidad y especificidad de cada prueba. Permite dibujar curvas ROC calculando el área bajo la curva y el punto óptimo de corte si se dispone de la prevalencia y de la relación de costes.

- Inferencia— Permite contrastar hipótesis de igualdad de medias y proporciones, con una o dos muestras, dependientes o independientes y calcular intervalos de confianza.
 - Priorización— Propone un método sencillo (CENDES/OPS) de determinación de prioridades en salud a partir de indicadores y áreas definidos por el usuario.
 - Epidat 2.1 es un programa de libre distribución desarrollado por instituciones públicas. Por ello, no sólo se permite, sino que se agradece su difusión y cualquier tipo de crítica o comentario que ayude a mejorar futuras versiones.
- **Requerimientos mínimos:**
 - Computador IBM-PC o compatible
 - Disquetera de 3,5" de 1.44 Mb
 - Procesador 386 o superior
 - Windows 3.xx o superior
 - 7 Mb de espacio libre disponible en disco.
 - 4 Mb de memoria RAM

1.2.4. STATA

Stata, es un paquete estadístico integrado para Windows, Macintosh y Unix, está diseñado para investigadores profesionales y es distribuido en 129 países.

Stata es utilizado por investigadores médicos, bioestadísticos, epidemiólogos, economistas, sociólogos, científicos, políticos, especialistas en geografía, psicólogos, científicos sociales y otros investigadores profesionales que necesitan analizar datos o información estadística.

Stata es un ambiente para manipular y analizar datos usando métodos estadísticos y gráficos, es un paquete integrado y no una colección de módulos por separado. Tiene la capacidad de entremezclar administración de datos estadísticos y comandos gráficos.

- **Capacidades**

- Administración de datos. Transformación de datos, match-merge, procesamiento por grupos, anexar archivos, ordenar, ODBC, ingresos externos, transposición filas-columnas, marcación, funciones de texto.
- Estadísticas básicas. Sumarios, tabulación cruzada, correlaciones, t tests, equivalencia de test de varianza, test de proporciones, intervalos de confianza.
- Modelos lineales. Regresión, estimados de varianza Huber/White/ sandwich, variables instrumentales, mínimos cuadrados de tres etapas, regresiones inconexas aparentes, regresiones forzadas, regresión quantile, GLS.
- Modelos lineales generalizados. Diez funciones de enlace, enlaces definidos por el usuario, siete distribuciones, estimaciones ML y IRLS, nueve estimadores de varianza, siete residuales.
- Binarios, contadores y variables limitadas dependientes. Logística, probit, regresión Poisson, tobit, regresión truncada, logística condicional, logit multinomial, binomios negativos, modelos zero-inflated, modelos de selección Heckman, efectos marginales.
- Series de Tiempo Panel data/cross sectional. Regresiones aleatorias y de efectos fijos, GEE, efectos aleatorios Poisson y binomiales negativos, Arellano Bond, regresiones de variables instrumentales, regresiones con distorsiones AR.
- Métodos no paramétricos. Wilcoxon - Mann - Whitney, rangos firmados Wilcoxon, Kruskal-Wallis, correlaciones Spearman y Kendall, Kolmogorov-Smirnov, binomial CIs exactos,
- ANOVA/MANOVA, Diseños balanceados y no balanceados.
- Métodos multivariantes,
- Análisis de conglomerados
- Métodos de remuestreo y simulación.
- Prueba de modelos y soporte post-estimación.
- Gráficos
- Métodos de encuestas
- Análisis de encuestas
- Análisis de supervivencia
- Herramientas para epidemiologistas
- Series de tiempo

- Máxima probabilidad
- Transformar y normalizar tests.
- Otros métodos estadísticos
- Comandos de matrix
- Lenguaje de programación
- Capacidades de Internet

1.2.5. SIGSA

El Sistema de Información Geográfico en Salud (SIGSA) presenta la interfaz de usuario de generación de mapas temáticos y de registro de eventos en salud y de información medio-ambiental como parte de las áreas de la organización de la Dirección Departamental de Salud del Cauca (DDSC) encargadas de la vigilancia en salud pública, participación social, sistemas de información e infraestructura en salud.

En SIGSA se desarrollaron mapas temáticos con un cubrimiento actual de 8 municipios del Cauca. Estos mapas temáticos son:

- *Sistema Nacional de Vigilancia en Salud Pública - SIVIGILA*
 - Mapas de Incidencia y Casos (Morbilidad y Mortalidad)
 - Tuberculosis
 - Muerte Materna
 - Muerte Perinatal
 - Mapa de atención de eventos en municipios no endémicos
 - Mapa de eventos municipios endémicos y mesoendémicos Malaria falciparum
 - Mapa de eventos municipios endémicos y mesoendémicos Malaria vivax
 - Información de casos
 - Leishmaniasis (Cutánea, Mucosa, Visceral)
 - Dengue (Clásico, Hemorrágico)

- *Proyecto ETV - Enfermedades transmitidas por vectores*
 - Estadísticas de muestras tomadas de gota gruesa por localidad.
 - Información para las enfermedades de Leshmaniasis (cutánea, mucosa y visceral) y Dengue (clásico y hemorrágico)
 - Puntos de diagnóstico para análisis de Gota Gruesa.
 - Puntos de información para toma de muestras de Gota Gruesa.

- *Proyecto Medio Ambiente*
 - Estado sanitario de los municipios
 - Calidad del agua según el informe mensual del Laboratorio de la DDSC
 - Plantas de tratamiento en los municipios indicados
 - Mataderos
 - Basureros

- *Proyecto de Tuberculosis (Análisis de Cohorte)*

Resultados del tratamiento en la cohorte de casos pulmonares nuevos con bk positivo, en el que se muestran los índices anuales de efectividad y de captación del programa.

La importancia principal del uso de SIGSA radica en explicar el significado de la variable geográfica en el complejo salud-enfermedad, de manera que sea posible contribuir a su comprensión. Mediante este sistema se despliegan los patrones de morbi-mortalidad distribuidos de una forma específica en diferentes regiones o municipios correlacionándolos con otros fenómenos locales.

Además se tiene un mejor conocimiento, de cara a la toma de decisiones, de hacia dónde se difunden los problemas de salud a través del tiempo y de cuál es el significado que tienen las variables ambientales para la salud de la población.

2. CONCEPTOS TECNOLÓGICOS

2.1. DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS (KDD)

Históricamente al proceso de encontrar patrones útiles en los datos, se ha conocido con diferentes nombres: descubrimiento de conocimiento en bases de datos (KDD), minería de datos (data mining), extracción de conocimiento (knowledge extraction), descubrimiento de información (information discovery), recolección de información (información harvesting), arqueología de datos (data archaeology) y procesamiento de patrones (pattern processing).

El término KDD se acuñó en el año 1989 en el primer workshop de KDD para referirse a todo el proceso de encontrar conocimiento útil en los datos.

KDD se ha desarrollado y continúa desarrollándose con base en las investigaciones realizadas en los campos del aprendizaje de máquina, reconocimiento de patrones, bases de datos, estadística, inteligencia artificial, sistemas expertos, visualización de datos y computación de alto rendimiento [WGC 91]. La meta común es la extracción de conocimiento de los datos en el contexto de grandes bases de datos.

Según [WGC 91] el descubrimiento de conocimiento es la extracción no trivial de información implícita, previamente desconocida y potencialmente útil, a partir de un conjunto de datos.

En esta definición, los datos corresponden a un conjunto de hechos (registros en una base de datos) y un patrón es una expresión en algún lenguaje que describe un subconjunto de datos o un modelo aplicable a ese subconjunto. El término proceso implica que KDD se compone de varios pasos entre los cuales están:

1. Desarrollo y entendimiento del dominio de la aplicación, el conocimiento relevante y los objetivos del usuario final. Este paso requiere cierta dependencia usuario/analista, pues intervienen factores como: conocer los cuellos de botella del dominio, saber qué

partes son susceptibles de un procesamiento automático y cuáles no, cuáles son los objetivos, los criterios de rendimiento exigibles, para qué se usarán los resultados que se obtengan, compromisos entre simplicidad y precisión del conocimiento extraído, etc.

2. Creación del conjunto de datos objetivo, seleccionando el subconjunto de variables o ejemplos sobre los que se realizará el descubrimiento. Esto implica consideraciones sobre la homogeneidad de los datos, su variación a lo largo del tiempo, estrategia de muestreo, grados de libertad, etc.
3. Preprocesado de los datos: eliminación de ruido, estrategias para manejar valores ausentes, normalización de los datos, etc.
4. Transformación y reducción de los datos. Incluye la búsqueda de características útiles de los datos según sea el objetivo final, la reducción del número de variables y la proyección de los datos sobre espacios de búsqueda en los que sea más fácil encontrar una solución. Este es un paso crítico dentro del proceso global, que requiere un buen conocimiento del problema y una buena intuición, y que, con frecuencia, marca la diferencia entre el éxito o fracaso de la minería de datos (paso 7).
5. Elección del tipo de sistema para minería de datos. Esto depende de si el objetivo del proceso de KDD es la clasificación, regresión, agrupamiento de conceptos (clustering), detección de desviaciones, etc. (en [UGR 96] pueden verse en detalle los diferentes métodos de minería de datos).
6. Elección del algoritmo de minería de datos.
7. Minería de datos. En este paso se realiza la búsqueda de conocimiento con una determinada representación del mismo. El éxito de la minería de datos depende en gran parte de la correcta realización de los pasos previos, por parte del usuario.
8. Interpretación del conocimiento extraído, con posibilidad de iterar de nuevo desde el primer paso. La obtención de resultados aceptables dependerá de factores como: definición de medidas del interés del conocimiento (de tipo estadístico, en función de su sencillez, etc.) que permitan filtrarlo de forma automática, existencia de técnicas de

visualización para facilitar la valoración de los resultados o búsqueda manual de conocimiento útil entre los resultados obtenidos.

9. Consolidación del conocimiento descubierto, incorporándolo al sistema, o simplemente documentándolo y enviándolo a la parte interesada. Este paso incluye la revisión y resolución de posibles inconsistencias con otro conocimiento extraído previamente.

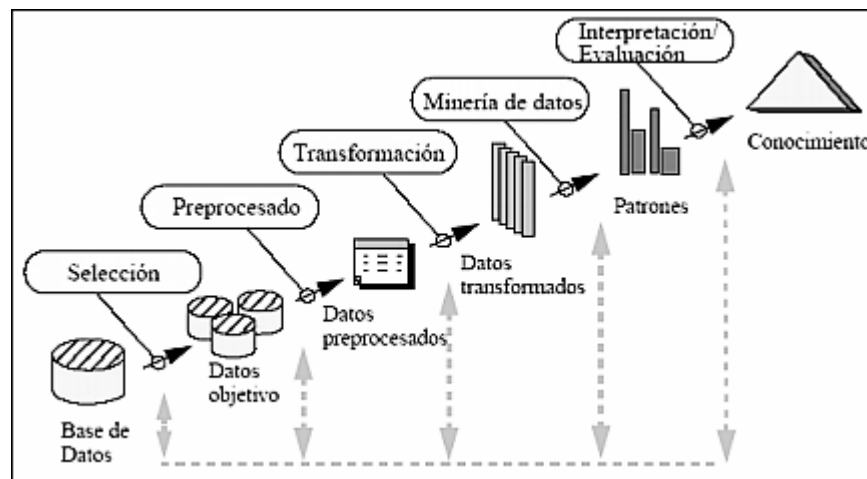


Figura 1. Proceso KDD

2.2. ALMACENES DE DATOS (DATA WAREHOUSING)

2.2.1. CONCEPTO DE DATA WAREHOUSING

Data warehousing (DW) es el centro de la arquitectura para los sistemas de información en la década de los '90. Soporta el procesamiento informático al proveer una plataforma sólida, a partir de los datos históricos para hacer el análisis de información. Facilita la integración de sistemas de aplicación no integrados. Organiza y almacena los datos que se necesitan para el procesamiento analítico, informático sobre una amplia perspectiva de tiempo.

La innovación de la Tecnología de Información dentro de un ambiente data warehousing, puede permitir a cualquier organización hacer un uso más óptimo de los datos, como un ingrediente clave para un proceso de toma de decisiones más efectivo. Las organizaciones tienen que aprovechar sus recursos de información para crear la información de la operación del negocio, pero deben considerarse las estrategias tecnológicas necesarias para la implementación de una arquitectura completa de data warehouse.

Se puede caracterizar un data warehouse haciendo un contraste de cómo los datos de un negocio almacenados en un data warehouse, difieren de los datos operacionales usados por las aplicaciones de producción.

Base de Datos Operacional	Data Warehouse
Datos Operacionales	Datos del negocio para Información
Orientado a la aplicación	Orientado al sujeto
Actual	Actual + histórico
Detallada	Detallada + más resumida
Cambia continuamente	Estable

Tabla 1. Concepto de Data warehouse

El ingreso de datos en el data warehouse viene desde el ambiente operacional en casi todos los casos. El data warehouse es siempre un almacén de datos transformados y separados físicamente de la aplicación donde se encontraron los datos en el ambiente operacional.

Un data warehouse se crea al extraer datos desde una o más bases de datos de aplicaciones operacionales. La data extraída es transformada para eliminar inconsistencias y resumir si es necesario y luego, cargadas en el data warehouse. El proceso de transformar, crear el detalle de tiempo variante, resumir y combinar los extractos de datos, ayudan a crear el ambiente para el acceso a la información Institucional. Este nuevo enfoque ayuda a las personas individuales, en todos los niveles de la empresa, a efectuar su toma de decisiones con más responsabilidad.

2.2.2. DIFERENCIAS ENTRE UN OLTP Y UN DW

Los sistemas tradicionales de transacciones y las aplicaciones de Data Warehousing (DW) son polos opuestos en cuanto a sus requerimientos de diseño y sus características de operación. Es de suma importancia comprender perfectamente estas diferencias para evitar caer en el diseño de un Data Warehouse como si fuera una aplicación de transacciones en línea (OLTP).

Las aplicaciones de OLTP están organizadas para ejecutar las transacciones para los cuales fueron hechos, como por ejemplo: mover dinero entre cuentas, un cargo o abono, una devolución de inventario, etc. Por otro lado, un Data Warehouse está organizado en base a conceptos, como por ejemplo: clientes, facturas, productos, etc.

Otra diferencia radica en el número de usuarios. Normalmente, el número de usuarios de un Data Warehouse es menor al de un OLTP. Es común encontrar que los sistemas transaccionales son accesados por cientos de usuarios simultáneamente, mientras que los Data Warehouse sólo por decenas. Los sistemas de OLTP realizan cientos de transacciones por segundo mientras que una sola consulta de un Data Warehouse puede tomar minutos. Otro factor es que frecuentemente los sistemas transaccionales son menores en tamaño a los Data Warehouses, esto es debido a que un Data Warehouse puede estar formado por información de varios OLTP's.

Existen también diferencia en el diseño, mientras que el de un OLPT es extremadamente normalizado, el de un Data Warehouse tiende a ser desnormalizado. El OLTP normalmente está formado por un número mayor de tablas, cada una con pocas columnas, mientras que en un Data Warehouse el número de tablas es menor, pero cada una de éstas tiende a ser mayor en número de columnas.

Los OLTP son continuamente actualizados por los sistemas operacionales del día con día, mientras que los Data Warehouse son actualizados en batch de manera periódica.

En la Tabla 2 se enmarcan las principales diferencias entre los sistemas operacionales y las aplicaciones de Data warehouse.

Sistema Operacional	Almacén de Datos
Almacena datos actuales	Almacena datos históricos
Almacena datos de detalle	Almacena datos de detalle y datos agregados a distintos niveles
Bases de datos medianas (100Mb – 1Gb)	Bases de datos grandes (100Gb – 1Tb)
Los datos son dinámicos (actualizables)	Los datos son estáticos
Los procesos (transacciones) son repetitivos	Los procesos no son previsibles
El número de transacciones es elevado	El número de transacciones es medio o bajo
Tiempo de respuesta pequeño (segundos)	Tiempo de respuesta variable (segundos – horas)
Dedicado al procesamiento de transacciones	Dedicado al análisis de datos
Orientado a los procesos de la organización	Orientado a la información relevante
Soporta decisiones diarias	Soporta decisiones estratégicas
Sirve a muchos usuarios (administrativos)	Sirve a técnicos de dirección

Tabla 2. Diferencias entre un sistema operacional y un Data warehouse

2.3. MINERÍA DE DATOS (DATAMINING)

2.3.1. FUNDAMENTOS DE DATAMINING

Las técnicas de Datamining son el resultado de un largo proceso de investigación y desarrollo de productos. Esta evolución comenzó cuando los datos de negocios fueron almacenados por primera vez en computadoras, y continuó con mejoras en el acceso a los datos, y más recientemente con tecnologías generadas para permitir a los usuarios navegar a través de los datos en tiempo real.

La Minería de Datos toma este proceso de evolución más allá del acceso y navegación retrospectiva de los datos, hacia la entrega de información prospectiva y preactiva. Minería de Datos está listo para su aplicación en la comunidad de negocios porque está soportado por tres tecnologías que están suficiente maduras:

- Recolección masiva de datos
- Potentes computadoras con multiprocesadores
- Técnicas de Minería de Datos

Las bases de datos comerciales están creciendo a un ritmo sin precedentes. Un reciente estudio sobre los proyectos de Data Warehouse encontró que el 19% de los que contestaron están por encima del nivel de los 50 Gigabytes, mientras que el 59% espera alcanzarlo en un periodo de tres meses. Esta investigación fue realizada en el año 1997, esto nos lleva a pensar en la magnitud de las bases de datos de hoy. En algunas industrias, tales como las de ventas al por menor, estos números pueden ser aún mayores. MCI Telecommunications Corp. cuenta con una base de datos de 3 terabytes + 1 terabyte de índices y cabecera. La necesidad paralela de motores computacionales mejorados puede ahora alcanzarse de forma menos costosa y más efectiva con tecnología de computadores con multiprocesamiento paralelo. Las técnicas de Minería de Datos han existido por lo menos desde hace 10 años, pero solo han sido implementadas recientemente como herramientas maduras, confiables, entendibles que consistentemente son más eficientes que los métodos estadísticos clásicos.

En la transformación de los datos de negocios a información de negocios, cada nuevo paso se basa en el anterior. Por ejemplo, el acceso a datos dinámicos es crítico para las aplicaciones de navegación de datos (drill through applications), y la habilidad para almacenar grandes bases de datos es crítica para la Minería de Datos.

Los componentes esenciales de la tecnología de Minería de Datos han estado bajo desarrollo por décadas, en áreas de investigación como estadísticas, inteligencia artificial y aprendizaje de máquinas. Hoy, la madurez de estas técnicas, junto con los motores de bases de datos relacionales de alta performance, hicieron que estas tecnologías fueran prácticas para los entornos de Data Warehouse actuales.

2.3.2. ALCANCE DEL DATAMINING

El nombre de Minería de Datos deriva de las similitudes entre buscar valiosa información en grandes bases de datos y minar una montaña para encontrar una veta de metales valiosos. Ambos procesos requieren examinar una inmensa cantidad de material, o investigar inteligentemente hasta encontrar exactamente donde residen los valores. Dadas bases de datos

de suficiente tamaño y calidad, la tecnología de Minería de Datos puede generar nuevas oportunidades de negocios al proveer estas capacidades:

- *Predicción automatizada de tendencias y comportamientos.* La Minería de Datos automatiza el proceso de encontrar información predecible en grandes bases de datos. Preguntas que tradicionalmente requerían un intenso análisis manual, ahora pueden ser contestadas directa y rápidamente desde los datos. Un típico ejemplo de problema predecible es el marketing apuntado a objetivos (trgeted marketing). Por ejemplo, problemas que incluyen pronósticos de problemas financieros e identificar segmentos de población que probablemente respondan similarmente a eventos dados.
- *Descubrimiento automatizado de modelos previamente desconocidos.* Las herramientas de Minería de Datos barren las bases de datos e identifican modelos en un solo paso. Otros problemas incluyen detectar transacciones fraudulentas por ejemplo de tarjetas de crédito e identificar datos anormales que pueden representar errores de tipeado en la carga de datos.

2.3.3. BENEFICIOS CLAVES DEL USO DE DATAMINING

1. Contribuye a la toma de decisiones tácticas y estratégicas proporcionando un sentido automatizado para identificar información clave desde volúmenes de datos generados por proceso tradicionales de e-business.
2. Permite a los usuarios dar prioridad a decisiones y acciones mostrando factores que tienen una mayor relevancia en un objetivo.
3. Proporciona poderes de decisión a los usuarios del negocio que mejor entienden el problema y el entorno y es capaz de medir las acciones y los resultados de la mejor forma.

4. Modelos descriptivos: En un contexto de objetivos definidos en los negocios permite a la Minería de Datos a organizaciones, sin tener en cuenta la industria o el tamaño, obtener soluciones que impactan en los resultados finales de la cuenta de resultados.

5. Modelos predictivos: Permiten que relaciones no descubiertas e identificadas a través del proceso de Minería de Datos sean expresadas como reglas de negocio o modelos predictivos.

3. MODELO DE DESARROLLO PARA EL ALMACENAMIENTO, ANÁLISIS E INTERPRETACIÓN DE INFORMACIÓN EPIDEMIOLÓGICA

3.1. DESCRIPCIÓN DEL MODELO

El modelo de desarrollo para el almacenamiento, análisis e interpretación de información aplicado a un sistema de vigilancia epidemiológica, MAISIVE, se fundamenta en el Modelo para la Construcción de Soluciones (MCS) [CES 01]. Los componentes constitutivos de MAISIVE son, por lo tanto, los macro-componentes constitutivos que se exponen en MCS.

MAISIVE es el resultado de un estudio exhaustivo producto de diferentes áreas que han centrado sus esfuerzos en el análisis de información como lo es el proceso KDD (Knowledge Discovery in Databases), en la adecuada planificación y construcción de los Datawarehouse para el almacenamiento de datos y en la correcta utilización de técnicas de Minería de Datos para el análisis de información y obtención de conocimiento novedoso y potencialmente útil.

3.1.1. FUNDAMENTOS DEL MODELO

Ingeniería del Sistema

Una aproximación de desarrollo adecuada debe tener su énfasis en la ingeniería del sistema, lo cual implica que la calidad del sistema debe ser vista como la habilidad del mismo para satisfacer las necesidades y expectativas de su ambiente, además debe asegurar primero la comprensión del todo antes de abordar las partes. Este es un concepto esencial para MAISIVE.

Orientación a Objeto

Parte importante del paradigma de Orientación a Objeto es tener presente en todo momento que se diseña con reuso y se diseña para reuso. Las técnicas de orientación a objeto son imprescindibles para la automatización de procesos que propone MAISIVE.

Descripción formal mediante UML

Se utiliza esta técnica de descripción formal ya que provee un lenguaje de modelamiento que favorece la comunicación de las abstracciones del sistema entre los integrantes del equipo humano del proyecto y los clientes/usuarios.

3.1.2. ESTRUCTURA PARA DESCRIPCIÓN DEL SISTEMA

En esta sección se presentan los modelos en los que se basa MAISIVE como referentes para describir el sistema de vigilancia epidemiológica con soporte para toma de decisiones. Cada modelo planteado debe usarse como una herramienta de comunicación valiosa en el proceso de definir y describir la solución propuesta.

MODELOS DE ALTO NIVEL	COMPONENTES PRINCIPALES
MODELO PARA EL ESTABLECIMIENTO DE RESPONSABILIDADES	Modelo de Negocio Modelo de Casos de Uso Arquitectura de Referencia del sistema Lista de Riesgos y Estrategias Plan del Proceso Caso del Negocio
MODELO PARA DESCRIPCIÓN DEL SISTEMA	Modelo de diseño Modelo de Despliegue Plan de pruebas
MODELO DE IMPLEMENTACIÓN DEL SISTEMA	Prototipo Operacional Manuales de Usuario

Tabla 3. Estructura para Descripción del Sistema con soporte para la toma de decisiones.

3.1.3. MODELO DEL PROCESO DE DESARROLLO

3.1.3.1. Características principales del proceso de desarrollo

El proceso de desarrollo propuesto por MAISIVE se describe detalladamente más adelante. A continuación se presentan las características principales del proceso:

- *Dirigido por Objetivos:* Se busca que los implicados en el desarrollo del trabajo se concentren principalmente en los objetivos más que en las actividades.
- *Iterativo e Incremental:* Es útil para manejar la complejidad de un proyecto se propone utilizar aproximaciones iterativas, las cuales deben ser planeadas y seguidas como si se tratara de un proyecto a menor escala.
- *Permite Gestionar Riesgos:* Un riesgo es una variable que puede incidir en el éxito de un proyecto, y puede causar efectos adversos tales como contratiempos, sobrecostos, hasta la cancelación del mismo. Este modelo propone técnicas que permiten asegurar una gestión adecuada de los riesgos del proyecto.
- *Guiado por Casos de Uso:* La razón de ser de un sistema es servir a usuarios ya sean humanos u otros sistemas; un caso de uso es una facilidad que un sistema debe proveer a sus usuarios. Los casos de uso reemplazan la antigua especificación funcional tradicional y constituyen la guía fundamental establecida para las actividades a realizar durante todo el proceso.
- *Centrado en la Arquitectura:* La arquitectura de un sistema de comercio electrónico involucra sus elementos más significativos, y está influenciada por consideraciones de desarrollo como sistemas legados y requisitos no funcionales. Es importante aclarar que los casos de uso dirigen la arquitectura del sistema.
- *Basado en Componentes:* En sistemas de soporte para la toma de decisiones de mediana y alta complejidad, el uso de componentes juega un papel importante, ya que permite dividir el sistema en componentes con interfaces bien definidas, que posteriormente se ensamblarán para generar el sistema objetivo. Esta característica en un proceso de desarrollo permite que el sistema objetivo se vaya creando a medida que se obtienen o que se desarrollan y maduran sus componentes

- *Permite rastreo sistemático del progreso:* Se efectúa un seguimiento del proyecto basado más en la evaluación absoluta del esfuerzo restante que en porcentaje de terminación de actividades.

3.1.3.2. Fases de Referencia

Este modelo aplica la técnica de desarrollo incremental por medio de iteraciones, buscando de esta forma proveer al cliente de una capacidad operacional inicial para validar su funcionalidad y mejorarla mediante la realimentación de las iteraciones

FASES	OBJETIVOS GENERALES	MODELOS
ESTUDIO PRELIMINAR	Definir el alcance del proyecto y sus Objetivos principales.	Modelo para el Establecimiento de Responsabilidades.
DISEÑO DEL ALMACEN DE DATOS	Definir una arquitectura de referencia para la construcción del sistema de vigilancia epidemiológica con soporte para la toma de decisiones, y eliminar y mitigar los riesgos críticos.	Modelo para el Establecimiento de Responsabilidades para la iteración
IMPLEMENTACION DEL ALMACEN DE DATOS	Implementar la arquitectura de referencias que soportará los procesos de almacenamiento para el análisis de información.	Modelo para Descripción del Sistema y Modelo de Implementación del Sistema.
ANALISIS DE INFORMACION	Formulación de los lineamientos para la implementación del proceso de análisis de información y su aporte en el adecuado procesamiento de la información y los recursos.	Modelo de Implementación del Sistema.
DESPLIEGUE E INTERPRETACION	Presentar nuevo conocimiento a las partes interesadas que den soporte a la toma de decisiones en la solvencia de conflictos o inquietudes dentro del sector de salud pública.	Modelo de Implementación del Sistema.
EVALUACION	Evaluar los resultados obtenidos por parte del cliente, del equipo de desarrollo y gestionar características nuevas que se puedan añadir dependiendo de si aportan nuevo conocimiento.	

Tabla 4. Descripción de las Fases de Referencia

3.1.4. MODELO DE ORGANIZACION DEL TALENTO

La adecuada organización del talento humano es de vital importancia para asegurar que el equipo humano de desarrollo del proyecto esté debidamente organizado y motivado, es por esto que se debe invertir todo el tiempo que sea requerido.

GRUPO DE GESTION		
CARGO	ÁREAS DE DESEMPEÑO	TAREA ESENCIAL
Director	1. Liderazgo del Proyecto 2. Integración del Producto 3. Calidad del Producto	1. Construir un ambiente de trabajo adecuado 2. Asegurar la coherencia del Producto 3. Garantizar la satisfacción del cliente
Asesor	1. Gestión de Reusabilidad 2. Calidad del Proceso	1. Enriquecer la base de conocimiento/experiencia 2. Mejorar sistemáticamente el proceso
Consultor	1. Producto 2. Proceso	1. Evaluar aspectos claves 2. Validar métodos y decisiones

Tabla 5. Estructura del Grupo de gestión

GRUPO DE DESARROLLO	
CARGO/POSICIÓN	RESPONSABILIDADES BÁSICAS
Analista del sistema	1. Modelar la Organización 2. Elaborar el Modelo de casos de uso para los procesos del sistema
Arquitecto del sistema	3. Diseñar la arquitectura para soportar los procesos de almacenamiento y análisis de información
Especialista en Minería de Datos	1. Analizar y buscar patrones 2. Procesar datos 3. Proponer el modelo de análisis

Tabla 6. Estructura del Grupo de desarrollo

3.2. MODELO DE DESARROLLO

3.2.1. FASE I: ESTUDIO PRELIMINAR

La presente sección describe la primera fase del modelo de desarrollo como el Estudio Preliminar que da inicio a los lineamientos para la implementación del almacenamiento, análisis e interpretación de información aplicado a un sistema de vigilancia epidemiológica. Entendiendo que la implementación de un sistema de vigilancia epidemiológica con soporte a la toma de decisiones no es un proyecto sino un proceso a mediano y largo plazo.

3.2.1.1. Descripción General de la Fase

- **Objetivos Generales**

Definición del alcance y viabilidad del proceso a ejecutar y sus objetivos principales.

- **Objetivos Específicos**

- Identificar las funcionalidades que soportará el sistema de vigilancia epidemiológica para el análisis e interpretación de información.
- Definir una arquitectura inicial que soporte la funcionalidad básica del sistema de vigilancia epidemiológica para el análisis e interpretación de información.
- Elaborar la planeación inicial para el proceso.
- Definir el caso de negocio.

- **Justificación**

El desarrollo de esta fase al inicio de un proceso que soportará el análisis e interpretación de información en un sistema de vigilancia epidemiológica tiene gran relevancia debido a que el correcto entendimiento de los procesos internos de la organización son el punto de partida para determinar la viabilidad del proceso a implementar y enfocar los esfuerzos en la consecución de un producto final acorde con las necesidades de la organización y que mejore los procesos internos de la misma.

- **Etapas de Referencia**

1. Planeación de la Fase.
2. Modelado de la Organización.
3. Gestión Inicial de Riesgos para el Proceso.
4. Planeación Inicial de Proceso.
5. Determinación la Viabilidad del Proceso.

- **Productos de trabajo**

1. Declaración inicial del negocio.
2. Lista de comprobaciones para la fase.
3. Modelo de negocio.
4. Arquitectura de referencia.
5. Lista de riesgos para el proceso.
6. Plan de trabajo para el proceso.
7. Lista de comprobaciones de la fase.
8. Caso del negocio.

3.2.1.2. Etapas de Referencia

1. Etapa I : Planeación de la Fase

Propósito

Organizar y definir actividades, tiempos, objetivos y recursos para el desarrollo del Estudio Preliminar para el proceso a implementar.

Actividades

1. Declarar el negocio según la perspectiva del cliente.
2. Establecer los criterios de evaluación.

Actividad 1. Declarar el negocio según la perspectiva del cliente

- **Descripción**

La finalidad de esta actividad es lograr un correcto entendimiento sobre las ideas del cliente en torno a las implicaciones que tiene la implementación de procesos de análisis e interpretación de información en el sistema de vigilancia epidemiológica existente.

- **Actores**

Director del proceso

Analista del sistema

Arquitecto del sistema

Asesor/es de la organización para una descripción adecuada de los procesos a implementar.

- **Técnicas**

Entrevistas

En la mayoría de los casos las personas entrevistadas son gerentes o usuarios finales que proporcionan datos claves sobre el sistema de vigilancia epidemiológica existente. Para mayor información en [JAS 98].

Para obtener el mayor provecho de las entrevistas es aconsejable aplicar los siguientes lineamientos:

- Recolección completa de la estructura administrativa de la organización, por ejemplo, visión, misión, organización del personal, departamentos, entre otros.
- Previo a las entrevistas, se debe realizar una asesoría sobre los tipos de sistemas de información, las diferencias entre los sistemas estratégicos y los sistemas técnico-operacionales y explicar que la implementación de arquitecturas de soporte para procesos de análisis e interpretación de información no es un proyecto sino un proceso de implementación a mediano y largo plazo. Para mayor información en el **Anexo A** de este documento.
- Documentar cada sesión de las entrevistas previo acuerdo de las dos partes, así como también la información adicional reunida por el equipo de trabajo que realizará el proceso con el fin de crear y mantener un banco de información de referencia.

- Si la arquitectura propuesta por el cliente requiere interacción con sistemas ajenos a la organización se debe recopilar la información necesaria sobre el funcionamiento de estos sistemas y realizar un estudio preliminar de la interacción con ellos.

- **Productos de Trabajo**

Declaración inicial del negocio, documento que contiene:

- Propósito de la implementación de procesos de análisis e interpretación de información en el sistema de vigilancia epidemiológica.
- Tipo de usuarios finales a los cuales van dirigidos los procesos a implementar en el sistema de la organización. Con la descripción de las funciones que tendrán estos usuarios con los procesos a implementar.
- Beneficios que se buscan para la organización con la inversión en procesos para la toma de decisiones.
- Arquitectura inicial propuesta por el cliente.
- Lista inicial de requisitos.

Actividad 2. Establecer los criterios de evaluación

- **Descripción**

Esta actividad define los criterios a evaluar satisfactoriamente para dar por terminado el Estudio Preliminar del Proceso.

- **Actores**

Director del proceso

- **Técnicas**

La lista de comprobaciones de la declaración inicial del negocio sirve como base para implementar una lista propia para el equipo de desarrollo del proceso de acuerdo a la experiencia en anteriores proyectos y las características particulares de cada uno de ellos.

Definición del alcance del proyecto

- ¿Se obtuvo una descripción suficientemente clara del sistema de vigilancia epidemiológica con soporte a la toma de decisiones y una adecuada declaración de su propósito?
- ¿Se realizó una clara identificación de actores?
- ¿Se tiene un modelo del negocio que representa realmente los procesos de negocio necesarios?

Viabilidad de la arquitectura inicial

- ¿Satisface la arquitectura propuesta las necesidades del Cliente y de los Usuarios?
- ¿Es factible realizar una implementación del sistema de vigilancia epidemiológica con soporte a la toma de decisiones según la arquitectura propuesta?
- ¿Se han considerado alternativas a ésta?
- ¿La arquitectura propuesta utiliza apropiadamente la tecnología existente?
- ¿Se han evaluado sus criterios de eficiencia, tolerancia a fallas, adaptabilidad y robustez?
- ¿La arquitectura planteada posibilita el crecimiento o evolución del sistema a implementar?
- ¿Se han realizado todas las evaluaciones pertinentes indicadas en esta guía?

Identificación de riesgos críticos

- ¿Se han identificado todos los riesgos críticos?
- ¿Se evaluaron los elementos del entorno de desarrollo y del entorno de ejecución, y se han propuesto medidas adecuadas al respecto?
- ¿Se ha elaborado un presupuesto consistente y acorde a las necesidades?
- ¿Se ha elaborado un cronograma o un plan de trabajo a seguir realista?

Conveniencia del caso del negocio

- ¿Es el Caso del Negocio lo suficientemente satisfactorio para justificar la continuación del proyecto?

• **Productos de Trabajo**

Lista de comprobaciones para la fase: Contiene una lista de comprobaciones para el estudio preliminar del proceso.

2. Etapa II : Modelado de la Organización

Propósito

Modelar los procesos de la organización y definir los requisitos iniciales para los procesos de almacenamiento, análisis e interpretación de información en el sistema de vigilancia epidemiológica. Además, se pretende recopilar la mayor información sobre los procesos internos para el análisis e interpretación de información que la organización realiza actualmente.

Actividades

1. Identificar los procesos del negocio.
2. Definir el alcance del proceso y tipos de usuarios.
3. Describir la Arquitectura Inicial para soportar el proceso.

Actividad 1. Identificar los procesos del negocio

- **Descripción**

Con esta actividad se pretende hacer una revisión de los procesos de negocio que la organización posee con el objetivo de estudiar el impacto que los procesos de almacenamiento, análisis e interpretación de información del sistema de vigilancia epidemiológica puedan tener sobre los procesos de la organización.

- **Actores**

Analista del sistema

Entrevistados

- **Técnicas**

Técnicas para encontrar hechos

El analista del sistema tiene como primera actividad realizar el análisis de los procesos de negocio de la organización con el fin de capturar los requisitos bases para la implementación de los

procesos de almacenamiento, análisis e interpretación de información en el sistema de vigilancia epidemiológica. Algunas técnicas para la adecuada recolección de esta información son:

- *Establecer una asociación de usuarios, gestión y grupos:* Es esencial involucrar tanto a los usuarios como a la gestión para asegurar que el almacén de datos contenga información que satisfaga los requerimientos de la organización para los procesos de análisis e interpretación de información que se desean implementar en el sistema de vigilancia epidemiológica.
- *Entrevistas:* Se utiliza para reunir información de personas o grupos. Estas entrevistas se deben realizar al número de usuarios seleccionados anteriormente basados en su valor dentro de la organización y realizar un análisis de sus puntos, preguntas y necesidades de acceso a los datos. De acuerdo a estas necesidades se construirán prototipos de implementación para el análisis de información y se prueban para que los usuarios finales puedan experimentar y modificar sus requerimientos mediante un proceso iterativo. Este es un paso inicial para obtener un consenso general sobre las necesidades.
- *Cuestionarios:* Esto permite que el Analista del sistema reúna información relacionada con diversos aspectos de un sistema de un grupo grande de personas. Para este propósito generalmente se emplean formatos estandarizados. Sin embargo este método no permite observar las expresiones o reacciones de los encuestados.
- *Revisión de registros:* La revisión de registros y reportes proporciona al Analista del sistema información de interés respecto a la organización y a sus operaciones. Aunque esta información no puede indicar la forma en que se desarrollan las actividades dentro de la organización, si pueden servir de gran ayuda para familiarizarse con las operaciones que necesitan apoyo y las relaciones formales dentro de la organización.

Se puede encontrar más información acerca de estas técnicas en [JAS 98] y [WIB 00].

Modelado del negocio

El modelado del negocio se constituye como la estrategia para modelar los procesos de negocio que posee una organización. Los beneficios que proporciona el modelado del negocio son los de definir mejores requisitos para los procesos de almacenamiento, análisis e interpretación de información que se implementarán en el sistema de vigilancia epidemiológica y que serán usados por los usuarios finales.

- *Modelo de casos de uso de negocio:* Este observa el negocio desde una perspectiva externa. El modelo de casos de uso de negocio incluye una descripción de los actores de negocio y de los casos de uso de negocio y las interacciones entre los actores de negocio y los casos de uso de negocio (Diagramas de casos de uso de negocio).
- *Diagramas de Casos de Uso del Negocio:* Define la interacción entre las entidades fuera del negocio y los procesos de negocio. Un diagrama de Casos de Uso de negocio representa visualmente la interacción entre los servicios primarios (casos de uso de negocio) que el negocio provee y para quien son proveídos dichos servicios (actores de negocio).



Estereotipo UML	Nombre	Definición UML
	Actor de negocio	Alguien o algo fuera del negocio que interactúa con él.
	Caso de uso de negocio	Una secuencia de acciones que ejecuta el negocio y que produce un resultado observable para un actor de negocio particular.

Tabla 7. Notación UML para actores y casos de uso de negocio

- *Modelo de objetos de negocio:* Este modelo detalla cómo los procesos de negocio son implementados internamente. El modelo de objetos de negocio incluye una descripción de los empleados del negocio (trabajadores de negocio), las cosas (entidades de negocio) que el negocio manipula y como los trabajadores de negocio manipulan las entidades de negocio para ejecutar los procesos de negocio (Diagramas de clases).



Estereotipo UML	Nombre	Definición UML
	Trabajador del negocio	Rol o conjunto de roles dentro del negocio. Un trabajador de negocio interactúa con otro y manipula entidades de negocio.
	Entidad del negocio	Son cosas manejadas o usadas por los trabajadores de negocio.

Tabla 8. Notación de modelado UML para Trabajadores y Entidades de Negocio

- **Productos de Trabajo**

Modelo de negocio, compuesto por:

- Modelo de casos de uso de negocio
- Modelo de objetos de negocio
- Glosario de términos

Se puede encontrar más al respecto de esta técnica en [RSC 01].

Actividad 2. Definir el alcance del proceso y tipos de usuarios

- **Descripción**

Hay muchas maneras para desarrollar procesos de almacenamiento y análisis de información como tantas organizaciones existen. Sin embargo, hay un número de dimensiones diferentes que necesitan ser consideradas, como lo son: Alcance del almacenamiento y análisis de información, redundancia de datos y tipos de usuarios finales. En esta actividad se propone definir el alcance de los procesos de almacenamiento y análisis de información así como los tipos de usuario final.

- **Actores**

Director del Proceso

Analista del sistema

Arquitecto del sistema

- **Técnicas**

Alcance del Proceso

El alcance de la implementación de procesos de almacenamiento y análisis de información en un sistema de información epidemiológica puede ser tan amplio como toda la información estratégica de la organización desde su inicio, o puede ser tan limitado como una base de datos personal para un solo gerente durante un año. En la práctica, en la amplitud del alcance, el mayor valor para la organización es el almacén de datos y el análisis de información que potencialmente permite realizar y lo más caro y consumidor de tiempo es crearlo y mantenerlo. Como consecuencia de

ello, la mayoría de las organizaciones comienzan con almacenes de datos funcionales, departamentales o divisionales y luego los expanden como usuarios que proveen retroalimentación.

Redundancia de Datos

Hay tres niveles esenciales de redundancia de datos que las organizaciones deberían considerar en sus opciones para la implementación de procesos de almacenamiento con fines de análisis de información:

- *Almacén Virtual o "Point to Point"*: Significa que los usuarios finales pueden acceder a bases de datos operacionales directamente, usando cualquier herramienta que posibilite "la red de acceso de datos". Este enfoque provee flexibilidad así como también la cantidad mínima de datos redundantes que deben cargarse y mantenerse. Los depósitos virtuales de datos proveen un punto de partida para que las organizaciones determinen qué usuarios finales están buscando realmente.
- *Almacenes Centrales*: Es una única base de datos física, que contiene todos los datos para un área funcional específica, departamento, división o empresa. Los almacenes centrales se seleccionan por lo general donde hay una necesidad común de los datos informáticos y un número grande de usuarios finales ya conectados a una red o computadora central. Pueden contener datos para cualquier período específico de tiempo. Comúnmente, contienen datos de sistemas operacionales múltiples. Los almacenes centrales son accesibles desde un lugar y deben cargarse y mantenerse sobre una base regular.
- *Almacenes Distribuidos*: Son aquellos en los cuales ciertos componentes del depósito se distribuyen a través de un número de bases de datos físicas diferentes. Cada vez más, las organizaciones grandes están tomando decisiones a niveles más inferiores de la organización y a la vez, llevando los datos que se necesitan para la toma de decisiones a la red de área local (Local Area Network - LAN) o computadora local que sirve al que toma decisiones.

No se puede pensar en un único enfoque. Cada opción adapta un conjunto específico de requerimientos y una buena estrategia de almacenamiento de datos, lo constituye la inclusión de las tres opciones.

Tipo de Usuario Final

Es importante notar que hay una gama cada vez más amplia de usuarios finales, cada una con su propio conjunto de requerimientos para los datos, acceso, flexibilidad y facilidad de uso. En general, se puede considerar tres grandes categorías:

- Ejecutivos y gerentes
- "Power users" o "Buzo de Información" (analistas financieros y de negocios, ingenieros, etc.)
- Usuarios de soporte (de oficina, administrativos, etc.).

Actividad 3. Describir la Arquitectura Inicial para soportar el proceso

- **Descripción**

Para esta actividad se pretende describir los recursos técnicos mínimos con los que debe contar el cliente para soportar los procesos de almacenamiento, análisis e interpretación de información que se desean implementar en el sistema de vigilancia epidemiológica de la organización. La realización de esta actividad es de gran importancia puesto que de acuerdo al cliente, se puede contar o no con la infraestructura tanto técnica como logística para afrontar correctamente la implementación de procesos para análisis de información.

- **Actores**

Arquitecto del sistema

- **Técnicas**

Para el entendimiento de la arquitectura básica necesaria que soportará los procesos para análisis de información del sistema de vigilancia epidemiológica, se deben comprender las operaciones que se ejecutan dentro de la arquitectura del almacén de datos y los componentes que las realizan. Este entendimiento nos permite encontrar los elementos constituyentes de la arquitectura de soporte para el análisis de información. Elementos como:

- Base de datos operacional / Nivel de bases de datos externos.

- Nivel de acceso a la información.
- Nivel de acceso a los datos.
- Nivel de directorio de datos.
- Nivel de gestión de proceso.
- Nivel de mensaje de la aplicación.
- Nivel del almacén de datos.
- Nivel de organización de datos.

Para mayor información en el **Anexo A** de este documento.

- **Productos de Trabajo**

Arquitectura de Referencia, que identifica los elementos constituyentes y sus características.

3. Etapa III : Gestión Inicial de Riesgos para el proceso

Propósito

La identificación de los riesgos iniciales que presenta la implementación de procesos para el análisis de información es una tarea clave para analizar su impacto y definir estrategias para afrontarlos. Esta etapa es de gran importancia ya que puede influir en la viabilidad del proceso y en la definición temporal de las actividades de planeación.

Actividades

1. Identificar riesgos.

Actividad 1. Identificar riesgos

- **Descripción**

La definición de los factores que incidirán en el éxito de la implementación del proceso y la identificación de las posibles soluciones para estos es el principal objetivo de esta actividad. Los diferentes tipos de riesgos se deben analizar y priorizar teniendo en cuenta el posible impacto

sobre el proceso y definir los posibles planes de contingencia para los riesgos verdaderamente importantes.

- **Actores**

Director del proyecto, quien definirá y categorizará los riesgos más importantes para el proceso apoyado en las valoraciones que hagan el analista y el arquitecto del sistema.

Analista del sistema.

Arquitecto del sistema.

- **Técnicas**

Categorizar los riesgos

Para identificar claramente el agente que los produce, como por ejemplo definir riesgos:

- *Relacionados con el tamaño del sistema:* Tamaño estimado para el proceso (LDC, PF, otras), confianza en la estimación, tamaño relativo con relación a otros procesos, cantidad de cambios a los requisitos proyectados antes y después de la entrega, etc.
- *Relacionados con el impacto en la organización:* Cómo impacta los procesos de almacenamiento y análisis de información a los diferentes procesos de la organización como son los usuarios finales, el análisis de datos actual, los beneficios, etc.
- *Relacionados con el cliente:* Existen experiencias anteriores con el cliente, el cliente está dispuesto a destinar tiempo a la especificación de requisitos, etc.
- *Relacionados con el entorno de desarrollo:* Existen herramientas para gestión de este tipo de procesos, hay herramientas apropiadas para soportar el análisis y diseño del sistema de soporte a la toma de decisiones, existen expertos a los cuales se le pedirá soporte sobre las herramientas, etc.
- *Relacionados con la tecnología:* Existen nuevas tecnologías en la implementación de estos procesos que la organización no haya manejado, existen requisitos de rendimiento muy estrictos, etc.
- *Relacionados con la experiencia del equipo de desarrollo:* Existe suficiente personal disponible, esta comprometido el equipo para toda la duración del proceso, el personal esta adecuadamente formado, etc.

- *Relacionados con la confiabilidad de los datos:* Existen en la organización herramientas de limpieza de datos especializadas y la forma de programar de los clientes proporcionan redes de seguridad, los usuarios ayudaron a definir los requerimientos de limpieza de datos, ya que son ellos los que mejor conocen los datos y pueden informar sobre qué tipo de datos sucios deben salir y cómo limpiarlos, etc.

Definir el impacto: Para cada riesgo identificado con el fin de hacer una distinción de los verdaderamente importantes.

Para mayor información en [DMI 99].

- **Productos de Trabajo**

Lista de riesgos para el proceso, documentados con la siguiente plantilla.

No. De Orden: Nombre del Riesgo	
Descripción	Una breve descripción del riesgo.
Categoría	Por ejemplo: técnico, del cliente, etc.
Impacto	Componente del riesgo que se afecta: Rendimiento, Costos, Mantenibilidad, Planificación.
Efecto	Nivel en el cual afecta al proceso: Despreciable, marginal, crítica, catastrófica.
Estrategia	Definición de las estrategias iniciales para afrontar el riesgo.

Tabla 9. Plantilla para la descripción de riesgos

4. Etapa IV : Planeación Inicial del Proceso

Propósito

El propósito de la planeación inicial del proceso es establecer planes razonables para administrar el proceso. Esto involucra el desarrollo de estimaciones para el trabajo a ser ejecutado, estableciendo los compromisos necesarios y definiendo el plan para ejecutar el trabajo.

Actividades

1. Elaborar el plan de trabajo.

Actividad 1. Elaborar el plan de trabajo

- **Descripción**

Esta actividad tiene como fin definir una estimación inicial costos para el desarrollo de las actividades futuras del proceso, tiempo, y además estimar el esfuerzo y los recursos para la realización de las actividades por parte de los responsables.

- **Actores**

Director del proyecto.

Analista del sistema.

Arquitecto del sistema.

- **Técnicas**

Estimaciones de tiempo

Aproximaciones de las horas, días o meses de esfuerzo necesario para el desarrollo de los procesos de almacenamiento, análisis e interpretación de información a implementar. Algunos métodos comunes para estimar el tiempo de desarrollo son los siguientes:

- *Método Histórico:* Se basa en registros con respecto a procesos de desarrollo anteriores.
- *Método Intuitivo:* Basado en la intuición para realizar cálculos de tiempo. Se debe tener en cuenta que es difícil estimar con un alto grado de precisión, sin embargo la popularidad de este método se debe a que es rápido para obtener una estimación.
- *Método Estándar:* En este se identifica y cuantifican (con pesos individuales) los factores que afectan más drásticamente al proceso, tales como las características del personal, los detalles del sistema y la complejidad del proyecto.

Para mayor información en [JAS 98].

Estimaciones de costo y esfuerzo

La estimación del costo y del esfuerzo para un proceso que involucre desarrollo o implementación de software en conjunto con hardware nunca será una ciencia exacta debido a que las variables humanas, técnicas, de entorno, políticas que pueden afectar al costo final del proceso y al esfuerzo aplicado para desarrollarlo son complejas. Algunas de las técnicas y modelos utilizados para el cálculo de costo y el esfuerzo son las siguientes:

- *Método Histórico:* Aquí se tienen en cuenta los cálculos obtenidos en proyectos similares anteriores.
- *Técnicas de Descomposición:* Estas técnicas proponen que la estimación de costos puede llevarse en forma jerárquica hacia abajo o en forma jerárquica hacia arriba (Bottom Up). La primera se enfoca primero en los costos a nivel del sistema, así como los costos del manejo de configuración, de control de calidad, la integración del sistema, de entrenamiento y de documentación. La segunda estima el costo de desarrollo de cada módulo o subsistema, tales costos se agrupan para obtener el costo total.

Para mayor información en [ITG 03].

Requisitos de tiempo calendario

Reuniones con la gerencia, revisiones de los procesos a implementar, asesorías y capacitación, etc. consumen tiempo adicional para la implementación de los procesos de almacenamiento y análisis de información por lo tanto se necesitan técnicas para administrar dichos tiempos. Técnicas como:

- *Diagramas de Barras:* Esta es una de las técnicas de planificación más simple en la cual se usan diagramas de barras que muestran cada actividad en un proyecto y la cantidad de tiempo que se tomará esa actividad. Este método fue desarrollado por Henry L. Gantt, con lo que comúnmente a sus diagramas se les llama diagramas de Gantt.
- *Diagramas de Eventos Críticos:* Estos diagramas muestran los eventos significativos en la conclusión de un proyecto y la secuencia en la que deben llevarse a cabo. Difieren de los diagramas de barras en que representan puntos de conclusión, no tareas individuales de realizar.

- *Diagramas de PERT:* Debido a que los proyectos constan de eventos y actividades, el diagrama de PERT usa nodos y rutas para representar la interrelación de las actividades del proyecto. Los nodos indican eventos y las rutas las actividades necesarias para moverse de un nodo a otro.

Para mayor información en [JAS 98].

Estimación de recursos computacionales críticos

Para establecer un listado de recursos computacionales críticos se plantea hacer una definición de la categoría de los recursos requeridos para la implementación de los procesos de almacenamiento y análisis de información y que pueden incidir en forma definitiva en la construcción del sistema deseado. La siguiente es una posible categorización de recursos computacionales respecto a su tipo:

Categoría	Tipo	Ejemplos
Hardware	Plataformas	- Estaciones de Trabajo o PC y el número requerido. - Servidores, especificación de memoria, capacidad de procesamiento y canales de comunicación.
	Conexiones de Internet	MODEM, ADSL, etc.
Software	Sistemas Operativos	Windows family, Linux, Unix, Mac.
	Herramientas de desarrollo	Compiladores, Entornos de desarrollo, Generadores de código, herramientas de diseño gráfico, herramientas para modelado de software, motores de bases de datos, etc.
	Herramientas de Gestión	Herramientas de gestión de requisitos, herramientas para automatización de pruebas, software para monitoreo de tareas, software para gestión de versiones, hojas de cálculo, Procesadores de texto, etc.
	Software para seguridad	Sistemas para detección de intrusiones, software para encriptación, Antivirus, etc.
	Otros	-Software cliente o de servidor para transacciones (SET, Monederos electrónicos). - Certificados

Tabla 10. Categorización de recursos computacionales

- **Productos de Trabajo**

Plan de trabajo para el proceso, documento que contiene:

- Relación de actividades a realizar, duración estimada y esfuerzo

- Lista de recursos computacionales críticos para el proyecto
- Descripción del equipo humano inicial y asignación de responsabilidades
- Descripción de los recursos requeridos y las respectivas ventanas de disponibilidad
- Estimación de los costos

5. Etapa V : Determinación de la Viabilidad del Proceso

Propósito

Evaluar todo el trabajo realizado y elaborar el Caso del Negocio concluyendo con una definición de la viabilidad de implementación de los procesos de almacenamiento, análisis e interpretación de información en el sistema de vigilancia epidemiológica.

Actividades

1. Evaluar los productos de trabajo de la fase
2. Elaborar el caso del negocio

Actividad 1. Evaluar los productos de trabajo de la fase

- **Descripción**

Esta actividad permite comprobar a través de los criterios de evaluación establecidos en la Etapa I si al finalizar la fase se realizaron correctamente todas las actividades propuestas y si los resultados obtenidos con la elaboración de los productos satisfacen los objetivos de la implementación de los procesos de análisis de información.

- **Productos de Trabajo**

Lista de comprobaciones para la fase, con indicación de cumplimiento en cada uno de los criterios de evaluación establecidos.

Actividad 2. Elaborar el caso del negocio

- **Descripción**

El Director del proceso realiza una primera aproximación de la apuesta económica y los beneficios que se van a conseguir con la implementación de procesos para análisis de información. Es aquí donde se debe justificar la continuación del proceso.

- **Actores**

Director del proceso

- **Técnicas**

La elaboración del caso del negocio es una tarea compleja y lleva consigo gran cantidad de trabajo investigativo que involucra aspectos fundamentales tales como la apuesta económica y la recuperación de la inversión. Para la formulación del proceso es necesario definir más claramente los costos y beneficios buscados.

Costos

Se identifican diversos costos en la implementación de procesos de almacenamiento y análisis de información, tales como:

- *Costos Preliminares:* Son los de planificación, diseño y modelamiento.
- *Costos Iniciales:* Como son los de la plataforma de Hardware, software de la base de datos y herramientas de transferencia y limpieza de datos.
- *Costos en Procesamiento:* Son los de mantenimiento de datos, desarrollo de aplicaciones y capacitación y soporte.

Beneficios

Los beneficios que se deben formular en la elaboración de un proyecto de esta envergadura son:

- *Tácticos:* En la impresión y emisión de reporte reducidos, demanda reducida para consulta de clientes, entrega más rápida de información a los usuarios, etc.

- *Estratégicos* (Potencialidad): Como las aplicaciones y herramientas de acceso para los usuarios finales, decisiones con mayor información, toma de decisiones más rápida, capacidad de soporte a la información organizacional, etc.
- *A obtener*: Para la organización, para los usuarios, para la organización en tecnologías de información.

- **Productos de Trabajo**

Caso del negocio, que establece claramente:

- Beneficios que se quieren lograr.
- Costo estimado para la implementación de los procesos de almacenamiento y análisis de información en el sistema de vigilancia epidemiológica.
- Duración estimada para el desarrollo del proyecto.
- Apreciación sobre la conveniencia del proyecto.

3.2.2. FASE II: DISEÑO DEL ALMACÉN DE DATOS

En esta sección se presentan los lineamientos a seguir propuestos en la segunda fase del modelo de desarrollo para el diseño del almacén de datos como uno de los procesos para la toma de decisiones en un sistema de vigilancia epidemiológica.

3.2.2.1. Descripción General de la Fase

- **Objetivos Generales**

- Configuración de la arquitectura de referencia y diseño del almacén de datos que soportará la información del sistema de vigilancia epidemiológica para su análisis e interpretación.
- Establecer un proceso iterativo para la consecución de un sistema soporte para la toma de decisiones.

- **Objetivos Específicos**

- Estructurar los requisitos esenciales del proceso para esta iteración.
- Definir la arquitectura para el almacén de datos.
- Diseñar el almacén de datos para esta iteración.
- Documentar el proceso de diseño.
- Gestionar los riesgos del proceso.
- Completar el plan para el desarrollo del proceso.

- **Justificación**

Esta fase es sumamente importante para la correcta ejecución del proceso puesto que es aquí donde se refinan los requerimientos y se establece el proceso iterativo que terminara con la implementación del almacén de datos que soportará los procesos para el análisis e interpretación de la información del sistema de vigilancia epidemiológica existente.

- **Etapas de Referencia**

1. Planeación de la Fase.

2. Refinamiento de requisitos
3. Selección de Arquitecturas y del Sistema Gestor de Bases de Datos.
4. Gestión de Riesgos para la iteración.
5. Diseño del Almacén de Datos.
6. Planeación del Proceso iterativo.

- **Productos de trabajo**

1. Declaración del proceso iterativo.
2. Lista de comprobaciones para la fase.
3. Modelo de casos de uso para la iteración.
4. Lista de requisitos no funcionales.
5. Modelo de despliegue.
6. Lista de administración de riesgos para la iteración.
7. Modelo de diseño del almacén de datos.
8. Plan de trabajo para el almacén de datos de la iteración.

3.2.2.2. Etapas de Referencia

1. Etapa I : Planeación de la Fase

Propósito

Declarar los procesos del área de la organización que se implementarán como parte de un proceso iterativo que dará soporte al análisis e interpretación de información en el sistema de vigilancia epidemiológica de la organización.

Actividades

1. Declarar el proceso iterativo.
2. Establecer los criterios de evaluación.

Actividad 1. Declarar el proceso iterativo

- **Descripción**

La finalidad de esta actividad es declarar una aplicación piloto de alcance limitado con un reembolso medible para los usuarios y la gestión. Estos mismos criterios (alcance limitado, reembolso medible y beneficios claros para la organización) se aplican a cada iteración de implementación del proceso. La implementación incremental reduce riesgos y asegura que la magnitud del proceso permanezca manejable en cada iteración.

- **Actores**

Director del proceso

Analista del sistema

Arquitecto del sistema

Grupo de usuarios entrevistados.

- **Técnicas**

Seleccionar el tamaño de datos de la iteración

Una pauta general para la carga de cada iteración es: “La iteración debe contener datos que sean lo bastante grandes como para ser significativos y lo bastante pequeños como para ser rápidamente gestionables”.

Hay, prácticamente hablando, un número infinito de maneras para definir los datos del proceso iterativo a implementar. Lo que hace que la iteración del proceso de almacenamiento y análisis de información a implementar sea significativa depende completamente de quien es el primer usuario de la iteración. En este punto hay que tener mucho cuidado. Por un lado, el arquitecto del sistema debe tener cuidado de no poner demasiados datos de análisis en la iteración a implementar. Por otro lado, debe tener cuidado de no incluir datos tan pequeños que la espontaneidad de descubrimiento por el analista de Sistemas de Información de apoyo a la toma de decisiones (DSS) este limitada.

Seleccionar un área para la iteración

Una decisión importante para la implementación de los procesos de almacenamiento y análisis de información en la iteración es la selección del proceso o área de la organización a implementar. El modelo de datos de alto nivel es útil en la selección del área sujeto a implementar en los procesos de análisis de información en esta iteración. Las diferentes opciones que se aplican en la implementación de procesos de almacenamiento para análisis de información en la iteración son:

- implementar una sola área sujeto,
- implementar un subconjunto de una sola área sujeto, o
- implementar un subconjunto de múltiples áreas sujeto.

Para mayor información en [WIB 00], [WIC 00] y [WIU 00].

- **Productos de Trabajo**

Declaración del proceso iterativo a implementar, documento que contiene:

- Descripción del área de la organización a modelar.
- Los procesos del área.
- El tamaño de los datos a gestionar.

Actividad 2. Establecer criterios de evaluación

- **Descripción**

En esta actividad se definen los aspectos fundamentales que deben obtener evaluación satisfactoria para poder dar por terminada la fase de diseño del almacén de datos. El diseño del almacén de datos debe implementar en su totalidad los casos de uso que presente la implementación de la iteración del proceso.

- **Actores**

Director del proceso

Analista del sistema

Arquitecto del sistema

- **Técnicas**

Se pueden distinguir los siguientes puntos a tener en cuenta como criterios de evaluación de esta fase:

Especificación de Requisitos

- ¿Se han identificado los requisitos, actores, y casos de uso necesarios para diseñar la arquitectura del sistema?
- ¿Se han detallado lo suficiente los requisitos como para satisfacer los objetivos de esta fase?
- ¿Satisface el diseño los requisitos funcionales y no funcionales especificados en los casos de uso?

Definición De La Arquitectura

- ¿Satisface la arquitectura los requisitos determinados para el sistema y sus usuarios?
- ¿Se han adaptado suficientemente los casos de uso a la arquitectura ya existente, en el caso de que el cliente la posea?
- ¿Se ha ajustado la arquitectura para encajar con sistemas heredados existentes en la organización tales como sistemas contables o sistemas de bases de datos?
- ¿Se ha investigado lo suficiente con el fin de garantizar que los sistemas heredados suministran funcionalidades que se puedan reutilizar?
- ¿Es la línea base para la arquitectura del sistema lo suficientemente robusta para soportar las demandas de crecimiento del sistema con el tiempo y el crecimiento del número de usuarios del mismo?
- ¿Es la línea base de la arquitectura coherente con los requisitos no funcionales tales como disponibilidad, seguridad, carga, tiempo de respuesta, etc.?

Gestión de riesgos

- ¿Se han identificado los riesgos críticos para la iteración?
- ¿Se han eliminado convenientemente los riesgos críticos de acuerdo a las estrategias planteadas o los planes propuestos para este fin?
- ¿Se ha creado un plan de contingencia para afrontar los riesgos que no se puedan eliminar?

Material de apoyo para usuarios finales

- ¿Se tienen preparadas ayudas, manuales o documentos de apoyo para los usuarios relacionados con esta iteración?

- **Productos de Trabajo**

Lista de Comprobaciones para la fase, contiene una lista de comprobaciones basada en la presentada en las técnicas de esta actividad con las modificaciones específicas que se puedan añadir según el criterio de los gestores del proceso y sus características.

2. Etapa II : Refinamiento de Requisitos

Propósito

Realizar la captura de los requisitos esenciales para la definición de los procesos de la iteración con el objetivo de que estos sirvan para definir sus responsabilidades principales.

Actividades

1. Definir las responsabilidades para la iteración.
2. Definir los requisitos no funcionales.

Actividad 1. Definir las responsabilidades para la iteración

- **Descripción**

Se propone definir la funcionalidad para los procesos de almacenamiento y análisis de información en el área de la iteración seleccionada con el fin de capturar los elementos de comportamiento más importantes. Para esta finalidad se utiliza el modelado de casos de uso con el objetivo de recoger los requisitos funcionales.

- **Actores**

Analista del sistema

- **Técnicas**

Reingeniería de los Procesos del Negocio

El diseño de un almacén de datos para soportar los procesos de análisis de datos, con frecuencia involucra lo que se piensa en términos más amplios y con conceptos del negocio más difíciles de definir que en el diseño de un sistema operacional. Al respecto, un almacén de datos está bastante cerca a Reingeniería de los Procesos del Negocio (Business Process Reengineering) [CFL 95].

Perfiles de Usuarios

Un perfil de usuario debería identificar quiénes son los usuarios del almacén de datos a implementar, dónde se ubican y cuántos necesita soportar. La información sobre cómo cada grupo espera usar los almacenes de datos, ayudará a analizar los diversos estilos de uso.

Casos de Uso

El modelo de casos de uso es un modelo del sistema que contiene actores, casos de uso y sus relaciones (Diagrama de casos de uso). En general este modelo describe lo que hace el sistema para cada tipo de usuario. Un caso de uso delimita la forma en que un actor usa el sistema. Representan trozos de funcionalidad que el sistema ofrece con el fin de aportar un resultado de valor para los actores. Los Casos de Uso pueden verse como acciones secuenciales y sus alternativas cuando un actor interactúa con el sistema. Debido a que primero se ha desarrollado un modelo de negocio, este se utiliza para encontrar los casos de uso y actores.

La descripción del escenario permite tener una visión textual del caso de uso con el fin de mejorar su comprensión. Para la descripción de los escenarios se propone la siguiente plantilla:

Nombre del Caso de Uso	Es un nombre descriptivo dado al caso de uso acompañado de un número de orden.
Iniciador	Se identifica el actor iniciador del Caso de Uso.
Propósito	Se describe el propósito para el cual fue creado el caso de uso.
Resumen	Se describe el propósito para el cual fue creado el caso de uso.

Tabla 11. Descripción de Escenarios a nivel del Estudio Preliminar

Más información al respecto de esta técnica en [IGJ 99].

- **Productos de Trabajo**

Modelo de casos de uso, el cual contiene:

- Diagrama de casos de uso de los procesos de la iteración del sistema de vigilancia epidemiológica con soporte para el análisis e interpretación de información.
- Descripción de escenarios.
- Descripción de actores.

Actividad 2. Definir los requisitos no funcionales

- **Descripción**

El objetivo de esta actividad es obtener los requisitos no funcionales para la implementación de los procesos soportados por la iteración. Los requisitos no funcionales son requisitos que no pueden asociarse a ningún caso de uso en concreto.

- **Actores**

Analista del sistema

Arquitecto del sistema

- **Técnicas**

Para obtener los requisitos no funcionales se realiza un proceso similar al de recolección de la información de la organización. Esta lista de requisitos se emplea en conjunto con el Modelo de Casos de Uso para realizar el análisis y diseño de los procesos de la iteración a implementar. Algunos ejemplos de requisitos no funcionales son:

Seguridad

Quizá la más común en cuanto a requisitos adicionales para procesos de almacenamiento y análisis de información de una organización se refiere, debido obviamente al valor de la información.

Disponibilidad

Se convierte en un requisito relevante en los periodos en que el sistema será accesado para soportar la toma de decisiones de la organización.

Excepciones de los datos

Mientras que los componentes del almacén de datos trabajan de acuerdo al modelo descrito para casi todos los datos, hay pocas excepciones útiles que necesitan ser discutidas:

- La **data resumida pública** se almacena y administra en el mismo almacén de datos, pero su cálculo fue realizado fuera de él.
- La **data externa**, que corresponde a la data almacenada en fuentes externas a la organización pero que es utilizada para soportar los procesos de toma de decisiones.
- Otro excepcional tipo de datos es el **detalle de los datos permanentes**, que resulta de la necesidad de una organización para almacenar la data a un nivel detallado permanentemente por razones éticas o legales. La organización simplemente no puede dejar los detalles porque en futuros años, en el caso de una demanda, una notificación, un edificio en disputa, etc., se incrementaría la exposición de la organización.

Consideraciones para el Diseño

El tipo de consideraciones para el diseño del almacén de datos que soportará los procesos de análisis e interpretación de información del sistema de vigilancia epidemiológica esta muy ligado con las características que debe tener un proceso de almacenamiento de esta índole. Características como: Orientación al tema, Integridad, de tiempo variante y no volátil. Para mayor información en el **Anexo A** de este documento.

Restricciones de Implementación

Muchas veces se limita la codificación e implementación de procesos de almacenamiento y análisis de información debido a que la organización ofrece para dicha implementación ciertos lenguajes de programación y manejadores de bases de datos con los que ya cuenta. Aunque se debe recalcar que la implementación de este tipo de procesos requiere de una inversión que abarca la adquisición de nuevas tecnologías en caso de ser necesario.

Planes de Expansión

Como cualquier proceso de almacenamiento y análisis de información debe evolucionar y los datos que gestiona llegan a ser más accesibles, empleados externos al sistema podrían descubrir también el valor de estos datos. Como estos procesos continúan creciendo en sofisticación y uso por lo que al momento de diseñar los procesos se deben tener en cuenta estos planes de expansión.

- **Productos de Trabajo**

Lista de requisitos no funcionales, Para la descripción de estos requisitos se plantea el siguiente formato:

No. De Orden: Nombre del Requisito no funcional	
Tipo	Identifica la clase de requisito (de diseño, de expansión, etc.), según lo descrito en las técnicas de la actividad 2 de la Etapa II de esta fase.
Descripción	Una breve descripción del requisito.
	Componente del riesgo que se afecta: Rendimiento, Costos, Mantenibilidad, Planificación.
Efecto	Nivel en el cual afecta al proceso: Despreciable, marginal, crítica, catastrófica.
Estrategia	Definición de las estrategias iniciales para afrontar el riesgo.

Tabla 12. Plantilla para la descripción de requisitos

3. Etapa III : Selección de Arquitecturas y del Sistema Gestor de Bases de Datos

Propósito

El éxito de los procesos de almacenamiento y análisis de información comienza cuando se seleccionan e integran satisfactoriamente tres elementos claves para la arquitectura de soporte: La arquitectura total del almacén de datos, la arquitectura del servidor y el sistema gestor de bases de datos. En esta etapa se pretende realizar la selección e integración de estos elementos claves. Si se escoge incorrectamente, los procesos implementados se convertirán en una gran empresa con problemas difíciles de trabajar en su entorno, costoso para arreglar y difícil de justificar.

Actividades

1. Seleccionar la arquitectura del almacén de datos.
2. Seleccionar la configuración del servidor.
3. Seleccionar el sistema gestor de bases de datos.

Actividad 1. Seleccionar la arquitectura del almacén de datos

- **Descripción**

La implementación de los procesos de almacenamiento y análisis de información comienzan con la estructura lógica y física de la base de datos del depósito más los servicios requeridos para operar y mantenerlo. En esta actividad se selecciona la arquitectura que soportara el almacén de datos para los procesos de análisis de la organización.

- **Actores**

Arquitecto del sistema

- **Técnicas**

La plataforma física puede centralizarse en una sola ubicación o distribuirse regional, nacional o internacionalmente. A continuación se dan las siguientes alternativas de arquitectura:

Arquitectura Centralizada

El almacén de datos centralizado refleja todos los aspectos del negocio. Un plan para almacenar los datos de la organización, que podría obtenerse desde fuentes múltiples internas y externas, es consolidar la base de datos en un almacén de datos integrado. El enfoque consolidado proporciona eficiencia tanto en la potencia de procesamiento como en los costos de soporte.

Arquitectura Global

La arquitectura global distribuye información por función, con datos financieros sobre un servidor en un sitio, los datos de comercialización en otro y los datos de fabricación en un tercer lugar. Los datos son consolidados lógicamente pero se almacena por separado sin las bases de datos físicas relacionadas, en los mismos sitios físicos o en diferentes.

Arquitectura por Niveles

Esta arquitectura almacena datos altamente resumidos sobre una estación de trabajo del usuario, con resúmenes más detallados en un segundo servidor y la información más detallada en un tercero.

Para mayor información en el **Anexo A** de este documento.

Actividad 2. Seleccionar la configuración del servidor

- **Descripción**

Al decidir sobre una estructura de depósito distribuida o centralizada, también se necesita considerar los servidores que retendrán y entregarán los datos. El tamaño de su implementación (y las necesidades de su empresa para escalabilidad, disponibilidad y gestión de sistemas) influirá en la elección de la arquitectura del servidor.

- **Actores**

Arquitecto del sistema

- **Técnicas**

Se debe combinar la configuración de plataformas de los servidores, mientras se decide cómo aprovechar los saltos casi constantes de la potencia del procesador. Algunas configuraciones de servidor son:

Servidores de un solo procesador

Los servidores de un sólo procesador son los más fáciles de administrar, pero ofrecen limitada potencia de procesamiento y escalabilidad. Además, un servidor sólo presenta un único punto de falla, limitando la disponibilidad garantizada del almacén de datos.

Multiprocesador Simétrico

Las máquinas de multiprocesamiento simétrico (Symmetric MultiProcessing - SMP) aumentan mediante la adición de procesadores que comparten la memoria interna de los servidores y los dispositivos de almacenamiento de disco. Se puede adquirir la mayoría de SMP en configuraciones mínimas (es decir, con dos procesadores) y levantar cuando es necesario, justificando el crecimiento con las necesidades de procesamiento.

Procesamiento en Paralelo Masivo

Una máquina de procesamiento en paralelo masivo (Massively Parallel Processing - MPP), conecta un conjunto de procesadores por medio de un enlace de banda ancha y de alta velocidad. Cada nodo es un servidor, completo con su propio procesador (posiblemente SMP) y memoria interna. Para optimizar una arquitectura MPP, las aplicaciones deben ser "paralelizadas" es decir, diseñadas para operar por separado, en partes paralelas.

Acceso de memoria no uniforme

La dificultad de mover aplicaciones y los sistemas de bases de datos a agrupaciones o ambientes realmente paralelos ha conducido a nuevas y recientes arquitecturas, tales como el acceso de memoria no uniforme (Non Uniform Memory Access - NUMA). NUMA crea una sola gran máquina SMP al conectar múltiples nodos SMP en un solo (aunque físicamente distribuida) banco de memoria y un ejemplo único de Sistema Operativo. NUMA facilita el enfoque SMP para obtener los beneficios de performance de las grandes máquinas MPP (con 32 o más procesadores), mientras se mantiene las ventajas de gestión y simplicidad de un ambiente SMP estándar.

Para mayor información en el **Anexo A** de este documento.

Actividad 3. Seleccionar el Sistema Gestor de Bases de Datos

- **Descripción**

Del lado del software, la complejidad y el alto costo de los sistemas gestores de bases de datos (SGBD) fuerzan a tomar decisiones drásticas y balances comparativos inevitables, con respecto a la integración, requerimientos de soporte, desempeño, eficiencia y confiabilidad. En esta actividad

se debe realizar un comparativo de SGBD dependiendo de los procesos a ser soportados en la organización para seleccionar con fundamentos el software de la base de datos que implementará la gestión de la información para la toma de decisiones en el sistema de vigilancia epidemiológica.

- **Actores**

Analista del sistema

Arquitecto del sistema

- **Técnicas**

Tipos de SGBD

De acuerdo a su costo y las funcionalidades soportadas, existen diversos tipos de SGBD. Algunos tipos comunes de SGBD son:

- *SGBD Relacionales* (Relational Data Base Management Systems - RDBMS): Los RDBMS son muy flexibles cuando se usan con una estructura de datos normalizada. En una base de datos normalizada, las estructuras de datos son no redundantes y representan las entidades básicas y las relaciones descritas por los datos (por ejemplo productos, comercio y transacción de ventas). La performance de los RDBMS tradicionales es mejor para consultas basadas en claves que para consultas basadas en el contenido como es el caso de procesos para análisis de información.
- *SGBD Multidimensionales* (*MultiDimensional Databases - MDDBS*): Para el soporte de depósitos a gran escala y para mejorar el interés hacia las aplicaciones de procesamiento analítico en línea (OLAP), los proveedores han añadido nuevas características al RDBMS tradicional. Estas, también llamadas características super relacionales, incluyen el soporte para hardware de base de datos especializada, tales como la máquina de base de datos Teradata.
- *SGBD Relacional-Objeto*: Una limitación de un RDBMS y un MDDB, es la carencia de soporte para tipos de datos no tradicionales como imágenes, documentos y clips de video/ audio. Muchos RDBMS almacenan los datos complejos como objetos grandes binarios (Binary Large Objects - BLOBs). En este formato, los objetos no pueden ser indexados, clasificados, o buscados por el servidor. Los DBMS relacional-objeto, de otro lado, almacenan los datos

complejos como objetos nativos y pueden soportar las grandes estructuras de datos encontradas en un ambiente orientado a objetos.

La selección del DBMS está también sujeta al servidor de hardware que se usa

Algunos RDBMS, como el DB2 Paralelo, Informix XPS y el Oracle Paralelo, ofrecen versiones que soportan operaciones paralelas. El software paralelo divide consultas, uniones a través de procesadores múltiples y corre estas operaciones simultáneamente para mejorar la performance. Se requiere el paralelismo para el mejor desempeño en los servidores MPP grandes y SMP agrupados. No es aún una opción con MDDBS o DBMS relacional-objeto.

Realizar comparativas entre los SGDB

En esta comparativa se resumen los pros y los contras de los diferentes tipos de SGDB para operaciones del almacén de datos a implementar. Un ejemplo de esta comparativa podría ser:

Características	SGBD			
	Relacional	Super-Relacional	Multidimensional	Objeto-Relacional
Estructuras Normalizadas				
Tipos de datos abstractos				
Paralelismo				
Estructuras Multidimensionales				
Drill-Down				
Rotación				
Operaciones dependientes de datos				
Entre otras...				

Tabla 13. Cuadro comparativo de SGDB

Matriz de Decisión del Almacén de Datos

Contiene algunos ejemplos de cómo afectan estos criterios de decisión en la elección de una arquitectura de servidor/ almacén de datos.

PARA ESTOS AMBIENTES ...			ELIJA ...		
Requerimientos comerciales	Usuarios	Soporte de Sistemas	Arquitectura	Servidor	DBMS
Alcance: departamental	Pequeña: ubicación única	Local mínimo	Consolidado paquete	Procesador único o SMP	MDDB
Usos: análisis de datos		central promedio			
Alcance: departamental	Grande: analistas en una sola ubicación y los usuarios informáticos dispersos	Local mínimo	Seccionado: Detalle en central. Resumen en local	Grupos de SMP para central;	RDBMS para central
Usos: análisis más informática		central promedio			
Alcance: empresa	Grande: geográficamente disperso	Central fuerte	Centralizado	Grupos de SMP	Objeto-relacional con soporte Web
Usos: análisis más informática					
Alcance: departamental	Pequeña: pocas ubicaciones	Central fuerte	Centralizado	MPP	RDBMS con soporte paralelo
Usos: investigación					

Tabla 14. Matriz de Decisión

- **Productos de Trabajo**

Modelo de Despliegue, documento que relaciona:

- La arquitectura del almacén de datos.
- La configuración del servidor.
- El sistema gestor de bases de datos.

4. Etapa IV : Gestión de Riesgos para la Iteración

Propósito

El propósito de esta etapa la identificación de los riesgos importantes para la iteración con el fin de determinar estrategias para su administración y eliminación.

Actividades

1. Identificar los riesgos para la iteración.
2. Administrar los riesgos de la iteración.

Actividad 1. Identificar los riesgos para la iteración

- **Descripción**

En esta actividad se pretende identificar los riesgos para la iteración no identificados en el Estudio Preliminar.

- **Actores**

Director del proceso

Analista del sistema

Arquitecto del sistema

- **Técnicas**

Para identificar los riesgos de esta iteración se pueden utilizar las estrategias definidas para la actividad 1 de la etapa III del Estudio Preliminar.

Cabe aclarar que para esta actividad de la fase de Diseño del Almacén de Datos, se debe prestar especial cuidado al manejo de los datos y los riesgos o comportamientos adicionales que esta actividad demande, como son:

- La administración debe reconocer que el mantenimiento de la estructura del almacén de datos es tan crítico como el mantenimiento de cualquier otra aplicación de misión crítica.
- La gestión debe comprender también que con la implementación de procesos de almacenamiento de esta envergadura, se crearán nuevas demandas sobre sus sistemas operacionales, como: Demandas para mejorar datos, demandas para una data consistente, demandas para diferentes tipos de datos, etc.

Actividad 2. Administrar los riesgos de la iteración

- **Descripción**

Una vez que se hayan identificado los riesgos que afectan al proceso iterativo, el Director del proyecto los analiza y los refina con el fin de definir la estrategia para administrarlos según la importancia o impacto definido para cada uno de ellos.

- **Actores**

Director del proceso

- **Técnicas**

Árboles de Decisión

Con esta técnica se pretende tener una mejor comprensión entre las causas del riesgo y las acciones que permitirán manejar el riesgo en el futuro“. Los árboles de decisión son diagramas que representan en forma secuencial condiciones y acciones; muestra que condiciones se consideran en primer lugar, cuales en segundo lugar y así sucesivamente. Este también permite ver la relación que existe entre cada condición y el grupo de acciones permisibles asociadas con ella” [JAS 98].

Definición de Estrategias

Las estrategias para administrar el riesgo dependen de la posición o políticas de la Dirección del proceso. Estas son algunas posturas sobre los riesgos que se puede encontrar para afronta la planeación de estos:

- *Convivir con el riesgo:* En esta estrategia se convive con el riesgo en el transcurso del proceso. Esta estrategia es definida usualmente para riesgos con bajo impacto.
- *Eliminación:* Esta estrategia busca la solución del riesgo mediante la eliminación del mismo. Esta estrategia se define para riesgos que tienen gran impacto en el proceso y cuyas alternativas para afrontarlo no son muy eficaces para su fin.
- *Mitigación:* Esta estrategia busca reducir el impacto de la ocurrencia del riesgo, mediante el estudio de sus causas o estudiando estrategias para anticiparse a su ocurrencia.
- *Contingencias:* Esta estrategia busca definir salidas alternativas ante la ocurrencia del riesgo cuando las estrategias de mitigación no tienen efecto.

Se pueden encontrar algunas estrategias adicionales y más sobre estas técnicas en [EHA 98].

- **Productos de Trabajo**

Listado de administración de riesgos para la iteración, documento que contiene la lista de riesgos y las estrategias para administrarlos.

5. Etapa V : Diseño del Almacén de Datos

Propósito

El propósito de esta etapa es la realización del diseño que dará soporte conceptual al almacén de datos a implementar. Se deben definir el diseño lógico y físico del almacén para una adecuada documentación del proceso teniendo en cuenta las orientaciones de diseño para este tipo de desarrollos.

Actividades

1. Especificar el modelado de datos.
2. Definir la infraestructura técnica para el diseño físico.

Actividad 1. Especificar el modelado de datos

- **Descripción**

Para esta actividad, con el análisis de los requisitos de usuario hecho hasta el momento y con el discernimiento de las fuentes necesarias del sistema de información de la organización y externas, se puede especificar el modelo de datos que se implementará en el almacén de datos.

- **Actores**

Analista del sistema

- **Técnicas**

Diseño Conceptual

Partiendo de la recogida y análisis de requisitos se puede proponer el diseño conceptual del almacén de datos que soportará los procesos de la iteración. El modelo de datos conceptual permite la especificación detallada de una base de datos multidimensional o relacional. Esta especificación se construye mediante un lenguaje gráfico que permite describir las estructuras de datos y algunas restricciones de integridad. Un ejemplo del diseño conceptual es el diagrama de Entidad-Relación. Para mayor información en [FCA 00].

Diseño Lógico

La realización de este diseño puede partir formalmente del diseño conceptual propuesto, aunque no es relevante esta declaración. En el diseño lógico se define el tipo de modelado de datos a implementar. El modelado de datos de mayor uso en la actualidad es el modelado multidimensional ya que se puede implementar en un SGBD relacional.

- *Modelado Multidimensional:* En una base de datos multidimensional, la información se representa como matrices multidimensionales, cuadros de múltiples entradas o funciones de varias variables sobre conjuntos finitos. Cada una de estas matrices se denomina Cubo Multidimensional. En un esquema multidimensional se representa una actividad que es objeto de análisis (hecho) y las dimensiones que caracterizan la actividad (dimensiones). La información relevante sobre el hecho (actividad) se representa por un conjunto de indicadores (medidas o atributos del hecho). La información descriptiva de cada dimensión se representa por un conjunto de atributos (atributos de dimensión). La representación gráfica del esquema multidimensional dependerá del modelo de datos utilizado (relacional, ER, UML, OO, etc.)

Para mayor información en [FCA 00].

Actividad 2. Definir la infraestructura técnica para el diseño físico

- **Descripción**

Realizar los procesos de análisis de información requiere de una infraestructura técnica adoptada en el SGBD. En esta actividad se define la infraestructura para soportar el procesamiento analítico en línea de la información.

- **Actores**

Analista del sistema

- **Técnicas**

Tendencias

El trabajo analítico de la información se describe a menudo con las preguntas de tipo "por qué" en lugar de preguntar "cuánto". Este tipo de procesamiento de datos o información se llama a menudo "Proceso Analítico En Línea", OLAP abreviado. Al implementar un proceso de almacenamiento con soporte para OLAP, un análisis acertado de los requisitos es necesario. Basado en estos requisitos, la infraestructura técnica puede escogerse. Para esto, varias opciones están disponibles: la principal diferencia esta en la manera en que se guardan los datos. Existen tres tendencias principales:

- *OLAP Relacional (ROLAP):* Al usar las bases de datos relacionales para OLAP, los conceptos básicos de OLAP como las dimensiones, cubos y jerarquías tienen que ser trazadas a las estructuras de los datos relacionales. Las bases de datos relacionales sólo tienen tablas, columnas y relaciones entre las tablas, se debe diseñar un modelo de datos que permita implementar tablas para las dimensiones, tablas para los cubos y tablas o columnas para las jerarquías. El método más ampliamente usado para hacer esto es el método de modelado de *esquema en estrella*: se tiene una tabla para cada dimensión con una llave primaria y varias columnas adicionales (atributos) y una tabla para el cubo con una columna para cada medida y las columnas importantes con las relaciones a todas las tablas de la dimensión requeridas.

Para mayor información de la técnica de esquema en estrella en [RKI 97] y [RLM 98].

- *OLAP Multidimensional (MOLAP):* Al usar SGBD multidimensional (MDBMS) para almacenar datos para OLAP, no hay necesidad para transformar o trazar el modelo lógico multidimensional al modelo físico. Los MDBMS tienen incorporados estos objetos: los objetos básicos en un MDBMS son las dimensiones y los cubos. Esto posibilita una implementación más fácil y directa de los requisitos del usuario final que se han modelado lógicamente en el modelo multidimensional. El objetivo de los sistemas MOLAP es almacenar físicamente los datos en estructuras multidimensionales de forma que la representación externa y la representación interna coincidan.
- *OLAP Híbrido (HOLAP):* Al combinar las fortalezas de ROLAP (escalabilidad, estándares abiertos) y las ventajas de MOLAP (el poder analítico, el uso intuitivo y directo para implementar OLAP), una tercera opción para obtener OLAP puede crearse: OLAP híbrido (HOLAP). El carácter híbrido viene de la manera combinada de guardar los datos: los datos detallados (qué es volumen más alto de datos en un sistema de OLAP: muchos archivos) se guarda en una base de datos relacional, mientras los datos agregados (agregaciones definidas por las jerarquías) se almacenarán en estructuras multidimensionales en un MDBMS. El sistema HOLAP se accederá con herramientas de consulta y análisis multidimensional (MOLAP); el sistema de HOLAP decide, dependiendo del tipo de dato a ser leído, si: el dato está disponible en la base de datos multidimensional (agregó los niveles) y por consiguiente puede accederse directamente, o si el dato tiene que ser leído o calculado desde los datos almacenados en la base de datos relacional; en este caso, el dato se lee y despliega a través de una herramienta multidimensional.

Para mayor información en [HFR 00].

Pasos para el Diseño

La realización de estos pasos para el diseño del almacén de datos será una actividad fácil y rápida de realizar teniendo en cuenta todo el proceso de análisis que se tiene hasta el momento de la iteración en proceso de implementación:

- *Paso 1: Elegir un proceso de la organización para modelar:* Actividad de la organización soportada por el sistema de vigilancia epidemiológica del cual se puede extraer información

con el propósito de construir el almacén de datos para soportar los procesos de análisis de información.

- *Paso 2: Decidir el gránulo de presentación del proceso:* El gránulo es el nivel de detalle al que se desea almacenar información sobre la actividad a modelar. En un almacén de datos que soporta procesos de análisis de información se almacena información a un nivel de detalle fino, no porque se vaya a interrogar al almacén a ese nivel sino porque ello permite clasificar y estudiar (analizar) la información desde muchos puntos de vista.
- *Paso 3: Identificar las dimensiones que caracterizan el proceso:* Se identifican las dimensiones que caracterizan la actividad al nivel de detalle que se ha elegido. Las dimensiones se pueden obtener cuestionando: ¿Cuándo se produce la actividad?, ¿Cuál es el objeto de la actividad?, ¿Dónde se produce la actividad?, ¿Quién es el destinatario de la actividad?, etc. De cada dimensión se debe decidir los atributos (propiedades) relevantes para el análisis de la actividad. Entre los atributos de una dimensión existen jerarquías naturales que deben ser identificadas (p.e.: día, mes, año).
- *Paso 4: Decidir la información a almacenar sobre el proceso:* La tabla de hechos corresponde a la información (sobre la actividad) que se desea almacenar en cada tupla de la tabla de hechos y que será el objeto del análisis. Por ejemplo, precio, unidades, importes, etc.

Para mayor información en [ORL 02] y [MCG 03].

Orientaciones de Diseño

Algunas consideraciones para el diseño físico del almacén de datos:

- *Usar claves primarias sin significado:* Las claves de las dimensiones deben ser generadas artificialmente. Claves de tipo entero (4 bytes) son suficientes para dimensiones de cualquier tamaño (2^{32} valores distintos). Esta clave se debe aplicar también a la dimensión Tiempo.
- *Evitar normalizar:* El ahorro de espacio no es significativo además de multiplicarse las sentencias SQL con JOIN en su definición.
- *Incluir la dimensión Tiempo:* En un almacén de datos muchas consultas son restringidas y parametrizadas por criterios relativos a periodos de tiempo.

- *Dimensiones que cambian:* Se considera relevante el caso en que, en el mundo real, para un valor de una dimensión, cambia el valor de un atributo que es significativo para el análisis sin cambiar el valor de su clave. Existen 3 estrategias para el tratamiento de los cambios en las dimensiones: Realizar la modificación, Crear un nuevo registro o Crear un nuevo atributo.
- *Definición de agregados:* El almacenamiento de datos agregados por distintos criterios de agregación en la tabla de hechos mejora la eficiencia del almacén de datos.

Para mayor información en [ORL 02] y [MCG 03].

- **Productos de Trabajo**

Modelo de Diseño del Almacén de Datos, documento que contiene:

- Diseño lógico
- Diseño físico
- Documentos de soporte para la definición del diseño físico.

6. Etapa VI : Planeación del proceso iterativo

Propósito

El propósito es realizar una actualización del plan inicial del proceso realizado en el Estudio Preliminar a través de la planeación de la iteración, con el fin de documentar la implementación de los procesos del negocio durante todo el proceso, teniendo en cuenta las estimaciones de tiempo, esfuerzo o de recursos necesarios de acuerdo a los avances en el proceso. Además se tiene ahora una idea más clara del proceso y las estimaciones antes muy tentativas pueden ahora ser más precisas y de mayor confiabilidad.

Actividades

1. Elaborar el plan de trabajo para la iteración.

Actividad 1. Elaborar el plan de trabajo para la iteración

- **Descripción**

En esta instancia del proceso se actualiza la organización de actividades necesarias para llevarse a cabo la implementación del proceso de almacenamiento para la iteración declarada en la fase. Es de tener en cuenta que la distribución temporal de algunas actividades para la iteración pueden no corresponder a las estimadas en el plan de trabajo inicial del proceso puesto que pueden existir riesgos adicionales para cada iteración.

- **Actores**

Director del proceso

- **Técnicas**

Para hacer las estimaciones de la iteración se puede hacer uso de las técnicas mencionadas en la actividad 1 de la etapa IV del Estudio Preliminar. Se debe prestar especial atención al método histórico debido a que ya se realizaron estimaciones previas para la implementación de los procesos de almacenamiento y análisis de información. El Director del proyecto debe guardar una relación de las estrategias que usó para generar las estimaciones actuales así como las previas con el fin de que sirvan de guía para futuras y efectivas estimaciones con el uso del método histórico. Esto es importante para la organización ya que permite conocerse más así misma y por lo tanto ayuda a realimentar estrategias para fomentar el crecimiento organizacional.

- **Productos de Trabajo**

Plan de trabajo para el almacén de datos de la iteración, que contiene puntos definidos en la actividad 2 de la etapa IV del Estudio Preliminar y que servirá como complemento para el plan de trabajo del proceso.

3.2.3. FASE III: IMPLEMENTACIÓN DEL ALMACÉN DE DATOS

En esta sección se presentan los lineamientos propuestos en la tercera fase del modelo de desarrollo para la implementación del almacén de datos que dará soporte a los procesos de análisis e interpretación de la información en el sistema de vigilancia epidemiológica.

3.2.3.1. Descripción General de la Fase

- **Objetivos Generales**

Implementar el diseño físico del almacén de datos para definir los procesos de extracción, transformación y carga de datos.

- **Objetivos Específicos**

- Implementar el diseño físico del almacén de datos mediante la definición de los procesos de extracción, transformación y carga de datos.
- Gestionar los riesgos del proceso.
- Actualizar la documentación para el almacén de datos.

- **Justificación**

En esta fase se definen los procesos de administración de los datos que se depositaran en el almacén de datos. Es muy importante la realización de esta fase puesto que de la calidad de datos en el almacén depende el resultado en los procesos de análisis e interpretación de la información del sistema de vigilancia epidemiológica existente.

- **Etapas de Referencia**

1. Planeación de la Fase.
2. Definición de los procesos de carga y mantenimiento del almacén.
3. Formalización del proceso implementado.

- **Productos de trabajo**

1. Lista de procesos para la carga y mantenimiento de los datos.
2. Lista de comprobaciones para la fase.
3. Lista de administración de riesgos para el proceso iterativo.
4. Documentación actualizada del almacén de datos.

3.2.3.2. Etapas de Referencia

1. Etapa I: Planeación de la fase

Propósito

Organizar y definir las actividades para la implementación del almacén de datos del proceso iterativo seleccionado una vez definido su diseño físico.

Actividades

1. Establecer las tareas para los procesos de carga y mantenimiento de los datos
2. Establecer los criterios de evaluación

Actividad 1. Establecer las tareas para los procesos de carga y mantenimiento de los datos

- **Descripción**

En esta actividad se definen las tareas a realizar teniendo como objeto de estudio los procesos de extracción, transformación y carga de los datos. Estas tareas se deben listar partiendo de los requisitos analizados y modelados en el diseño físico del almacén de datos de la iteración.

- **Actores**

Director del proceso

Analista del sistema

- **Técnicas**

La definición de las tareas involucradas en los procesos de carga y mantenimiento del almacén de datos hacen referencia a dos componentes de la arquitectura del almacén de datos:

Sistema ETL (Extraction, Transformation, Load)

Realiza las funciones de extracción de las fuentes de datos (OLTP o externas), transformación (limpieza, consolidación, etc.) y la carga del almacén de datos, a través de tareas de: Extracción de datos, Filtrado de datos, Carga inicial del almacén y Refresco del almacén de datos.

- *Extracción:* Este proceso debe identificar los datos que han cambiado en las fuentes de datos, debe realizar la extracción (lectura) de datos, obtener datos agregados y realizar el mantenimiento del metadata.
- *Transformación:* Este proceso debe realizar la limpieza y transformación de datos, la integración de datos (cálculo de datos derivados), creación de claves, obtener datos agregados y realizar el mantenimiento del metadata.
- *Carga:* Este proceso debe realizar la carga de los datos al almacén de datos incorporando la indexación, obtención de datos agregados, realización de pruebas de calidad de la carga y realizar el mantenimiento del metadata.

La construcción del sistema ETL es responsabilidad del equipo de desarrollo del almacén de datos y requiere más del 50% del esfuerzo para la implementación del proceso de almacenamiento para análisis de información. Cabe resaltar que en la construcción del sistema se pueden utilizar herramientas de mercado o programas diseñados específicamente para estas operaciones.

Sistemas de Integridad y Seguridad

Se encargan de un mantenimiento global, copias de seguridad, planes de contingencia, etc.

Para mayor información en [JHO 03].

- **Productos de Trabajo**

Lista de tareas para los procesos de carga y mantenimiento del almacén de datos, documento que contiene la descripción de cada tarea del proceso de carga y mantenimiento.

Actividad 2. Establecer los criterios de evaluación

- **Descripción**

En esta actividad se definen los aspectos fundamentales que deben obtener evaluación satisfactoria para poder dar por terminada la fase de implementación del almacén de datos para la iteración en proceso.

- **Actores**

Director del proceso

Analista del sistema

- **Técnicas**

Se pueden distinguir los siguientes aspectos a tener en cuenta como criterios de evaluación de esta fase:

Especificación de Fuentes de Datos

- ¿Se han identificado todas las fuentes de datos que alimentarán al almacén de datos?
- ¿Se han identificado la inexistencia de fuente de datos para algunos requisitos?
- ¿Se han detallado lo suficiente la estructura de los datos como para satisfacer los objetivos de esta fase?
- ¿Satisface el diseño físico los requisitos de estructuración de datos para la implementación de los procesos de esta fase?

Definición de la Arquitectura

- ¿Se ha ajustado la arquitectura para identificar y encajar con los sistemas de carga y mantenimiento del almacén de datos?
- ¿Es la línea base para la arquitectura del sistema capaz de realizar futuros ajustes en sus componentes para satisfacer los objetivos de la fase?

Gestión de riesgos

- ¿Se han identificado los riesgos críticos para la iteración?

- ¿Se han implementado las estrategias para afrontar los riesgos críticos en la realización de los procesos de carga y mantenimiento del almacén de datos, ya sea con desarrollo de software interno o mediante la adquisición de herramientas comerciales?

Material de apoyo para usuarios finales

- ¿Se tienen preparadas ayudas, manuales o documentos de apoyo para los usuarios relacionados con esta iteración?
- ¿Son suficientes las ayudas para soportar la validación de estas funcionalidades del sistema por parte de dichos usuarios?

- **Productos de Trabajo**

Lista de Comprobaciones para la fase, contiene una lista de comprobaciones basada en la presentada en las técnicas de esta actividad con las modificaciones específicas que se puedan añadir según el criterio de los gestores del proceso y sus características.

2. Etapa II : Definición de los procesos de carga y mantenimiento del almacén

Propósito

El propósito de esta etapa es definir los procedimientos a seguir después de haber seleccionado y listado las tareas requeridas para los procesos de carga y mantenimiento del almacén de datos de la iteración. Es entendible que después de la implementación de la iteración, el sistema ETL aún se encuentre incompleto, teniendo en cuenta que es un proceso reiterativo en el cual se pueden y deben hacer este tipo de ajustes.

Actividades

1. Establecer los procesos ETL del almacén de datos

Actividad 1. Establecer los procesos ETL del almacén de datos

- **Descripción**

En esta actividad se establecen los procedimientos a realizar para la implementación de los procesos de extracción, transformación y carga de los datos. Los procedimientos corresponden a las actividades de las tareas listadas en la etapa I de esta fase.

- **Actores**

Analista del sistema

- **Técnicas**

La identificación de las actividades a realizar para las tareas del sistema ETL es la actividad más compleja en la implementación del proceso iterativo en curso dentro del proceso de almacenamiento. Hay muchas circunstancias que se deben tener en cuenta para la especificación del sistema:

- Múltiples fuentes de datos
- Inexistencia de fuentes de datos en algunos casos
- Reformateo de datos
- Resumen de datos
- Conversión de datos
- Alteración de las claves de la estructura
- Reestructuración de los atributos de los datos, entre otras.

La arquitectura de soporte para el proceso de almacenamiento seleccionada debe identificar el sistema ETL tanto como sea posible, pero es entendible que después de la implementación de la iteración, el sistema ETL aún se encuentre incompleto, teniendo en cuenta que es un proceso reiterativo en el cual se pueden y deben hacer este tipo de ajustes. La calidad de los datos es la clave del éxito de un almacén de datos.

Estrategias de Calidad

- Actuación sobre los sistemas operacionales: modificar las reglas de integridad, los disparadores y las aplicaciones de los sistemas operacionales para la obtención de una data de calidad desde su entrada a la organización.
- Documentación de las fuentes de datos.
- Definición del proceso de transformación de datos.
- Adopción de programas y herramientas comerciales para asegurar la calidad de los datos en caso de que el proceso en curso lo requiera.
- Nombramiento de un responsable de calidad del sistema (Data Quality Manager)

Proceso de Extracción

Este proceso debe realizar la extracción (lectura) de datos desde las fuentes de datos de la organización. En este punto se diferencian dos procesos para la extracción: durante la carga inicial del almacén de datos y para el mantenimiento del almacén.

- *Durante la carga inicial:* Existen dos consideraciones, analizadas en actividades anteriores:
 - Si los datos operacionales están mantenidos en un SGBD Relacional, la extracción de datos se puede reducir a consultas en SQL o rutinas programadas.
 - Si los datos operacionales están en un sistema propietario en el cual no se conoce el formato de los datos o en fuentes externas textuales, hipertextuales u hojas de cálculo, la extracción puede ser muy difícil y puede tener que realizarse a partir de informes o volcados de datos proporcionados por los propietarios que deberán ser procesados posteriormente.
- *Para el mantenimiento/refresco del almacén de datos:* Antes de realizar la extracción es preciso *identificar los cambios*. Es decir, identificar los datos operacionales (relevantes) que han sufrido una modificación desde la fecha del último mantenimiento. Algunos métodos para este punto son:
 - Carga total: Cada vez se empieza la extracción desde cero.
 - Comparación de instancias de la base de datos operacional.
 - Uso de marcas de tiempo (time stamping) en los registros del sistema operacional.

- Uso de disparadores en el sistema operacional.
- Uso del fichero de log de transacciones del sistema operacional.
- Uso de técnicas mixtas.

Proceso de Transformación

No importa cómo esté diseñado un programa o cuán hábilmente se use. Si se alimenta mala información, se obtendrá resultados incorrectos o falsos. Desafortunadamente, los datos que se usan satisfactoriamente en los sistemas operacionales pueden ser basura en lo que concierne a la aplicación del almacén de datos para análisis de información. Afortunadamente, las herramientas de limpieza de datos pueden ser de gran ayuda. En algunos casos, puede crearse un programa de limpieza efectivo. En el caso de bases de datos grandes imprecisas e inconsistentes, el uso de las herramientas comerciales puede ser casi obligatorio. Cada vez que se carga un nuevo conjunto de datos, la limpieza de datos comúnmente constituye cerca del 25 por ciento de lo que puede ser un proceso de cuatro semanas.

Decidir qué herramienta usar es importante y no solamente para la integridad de los datos. Si se equivoca, se podría malgastar semanas en recursos de programación o cientos de miles de dólares en costos de herramientas. Las herramientas comerciales pueden ayudar en cada uno de estos pasos. Sin embargo, es posible escribir sus propios programas para hacer el mismo trabajo. Los programas de limpieza de datos no proporcionan mucho razonamiento, por lo que las compañías necesitan tomar sus decisiones en forma manual, basados en información importante y reportes de auditoría de datos.

La limpieza de una data "sucia" es un proceso multifacético y complejo. Los pasos a seguir son los siguientes:

1. Analizar los datos a utilizar en el proceso para descubrir inexactitudes, anomalías y otros problemas. Con la ayuda de los usuarios que definieron los requisitos para la iteración, se debe realizar este proceso ya que son ellos los que mejor conocen los datos y pueden informar sobre qué tipo de datos sucios deben salir y cómo limpiarlos.
2. Transformar los datos para asegurar que sean precisos y coherentes.

3. Asegurar la integridad referencial, que es la capacidad del almacén de datos, para identificar correctamente al instante cada objeto del negocio.
4. Validar los datos que usa la aplicación para el análisis de datos para realizar las consultas de prueba.
5. Producir la metadata, una descripción del tipo de datos, formato y el significado relacionado al negocio de cada campo.
6. Finalmente, viene el paso crucial de la documentación del proceso completo para que se pueda ampliar, modificar y arreglar los datos en el futuro con más facilidad.

Para los pasos 2 y 3 se recomiendan los siguientes procedimientos:

- Implementar claves con estructura. Descomponer en valores atómicos.
- Unificar codificaciones. Por ejemplo las formas de declarar el género de una persona. (H-M, M-F, 1-0, Varón-Hembra, etc).
- Unificar estándares. Por ejemplo, para las unidades de medida, de tiempo, etc.
- Los valores duplicados deben ser eliminados. Mediante SQL o a través de restricciones en el SGBD del sistema operacional.
- La integridad referencial debe reconstruirse

En la práctica, se tendría que realizar múltiples pasos como parte de una operación única o cuando use una sola herramienta. En particular, limpiar la data y asegurar la integridad referencial son procesos interdependientes.

Para mayor información en [JHO 03] y en las orientaciones de diseño de [MCG 03].

Proceso de Carga

El proceso de carga consiste en mover los datos desde las fuentes operacionales o desde el almacenamiento intermedio que posea la arquitectura de referencia hasta el almacén de datos y cargar los datos en las correspondientes estructuras de datos. Se debe tener en cuenta que:

- La carga de datos puede consumir mucho tiempo.
- En la carga inicial del almacén de datos se mueven grandes volúmenes de datos.

- En los mantenimientos periódicos del almacén de datos se mueven pequeños volúmenes de datos.
- La frecuencia del mantenimiento periódico está determinada por el gránulo del almacén de datos.
- Se deben determinar las ventanas de carga más convenientes para no saturar la base de datos del sistema operacional.

Se deben tener en cuenta dos procedimientos posteriores a la carga: la indización y la obtención de agregados.

- *Indización:* Durante la carga se deja el índice del SBGD habilitado lo que implica un proceso de indización tupla a tupla, convirtiéndose en un proceso lento. Si se deja el índice del SGBD deshabilitado durante la carga y después se crea el índice (total o parcial), convirtiéndose en un proceso rápido.
- *Obtención de agregados:* Durante el proceso de extracción y en el transporte hacia la carga.

- **Productos de Trabajo**

Definición de los procedimientos de ETL del almacén de datos, documento que contiene los procedimientos y consideraciones realizadas en la implementación de estos procesos.

3. Etapa III : Formalización del proceso implementado

Propósito

En esta etapa el Director del proceso actualiza la documentación que se tiene del proceso. Actualización que comprende la lista de los riesgos y la documentación actualizada del almacén de datos para la fase de implementación.

Actividades

1. Gestionar los riesgos de la implementación.
2. Actualizar la documentación de apoyo para la iteración.

Actividad 1. Gestionar los riesgos de la implementación

- **Descripción**

En esta actividad se identifican y se formulan estrategias para gestionar los riesgos que presenta la implementación de los procesos de carga y mantenimiento en el almacén de datos de la iteración.

- **Actores**

Analista del sistema

- **Técnicas**

Para identificar y administrar los riesgos de implementación para la iteración se pueden utilizar las estrategias definidas para la actividad 1 y 2 de la etapa III de la Fase de Diseño del almacén de datos.

Para este punto del proceso se deben considerar factores como el tiempo de programación interna para la implementación de los procesos de carga y mantenimiento del almacén de datos o simplemente pensar en el costo de herramientas comerciales que realicen esta implementación. De todas formas, el director del proceso necesita evaluar el problema con realismo, los recursos internos disponibles para distribuirlos y seleccionar la solución que se adapte a la plantilla y presupuesto del proceso, o modificar la plantilla y el presupuesto para solucionar el problema.

- **Productos de Trabajo**

Lista de administración de riesgos para la iteración, documento que contiene la lista de riesgos y las estrategias para administrarlos.

Actividad 2. Actualizar la documentación de apoyo para la iteración

- **Descripción**

El objetivo de esta actividad es elaborar la documentación que servirá de referencia al equipo de trabajo para futuros desarrollos y para el personal de la organización que administrará los procesos una vez implementados. El equipo de desarrollo deberá elegir junto con el Director del proyecto la

forma cómo se presente esta documentación para que sea lo suficientemente clara y de utilidad para el lector.

- **Actores**

Director del proceso

Equipo de trabajo

- **Técnicas**

Manuales de soporte para el diseño del almacén de datos.

Diccionario de Datos del almacén de datos: Este documento contendrá la siguiente información acerca de la base de datos del sistema:

- Nombre y descripción de las tablas que conforman la BD.
- Campos que conforman cada tabla, indicándose también:
- Tipo de dato (Carácter, numérico, fecha, monto, etc.)
- Longitud (en bytes/caracteres).
- Número de cifras decimales en caso de que sea de tipo numérico.
- Claves primarias y foráneas de cada tabla.

Una sugerencia para no tener que repetir la información de campos que son utilizados por varias tablas (claves foráneas), es hacer un listado alfabético que describa cada uno de ellos. Esto no tiene ambigüedades siempre y cuando no se utilicen campos con el mismo nombre para propósitos diferentes.

- **Productos de Trabajo**

Documentación actualizada del almacén de datos

3.2.4. FASE IV: ANALISIS DE INFORMACIÓN

En este apartado se presentan los lineamientos a seguir que este modelo de desarrollo propone en su cuarta fase referente al análisis de información basado en el desarrollo de un sistema de vigilancia epidemiológica.

3.2.4.1. Descripción General de la Fase

- **Objetivos Generales**

Formulación de los lineamientos para la implementación del proceso de análisis de información y su aporte en el adecuado procesamiento de la información y los recursos

- **Objetivos Específicos**

- Capturar y estructurar los requisitos iniciales para llevar a cabo el proceso de Análisis de información.
- Realizar descubrimiento automatizado de modelos previamente desconocidos.
- Extraer información oculta y predecible de bases de datos.

- **Justificación**

La fase de Análisis de información es de gran importancia ya que es a partir de ella donde se da soporte a la toma de decisiones a través del descubrimiento de manera automática de nuevos y originales conocimientos ocultos a simple vista.

- **Etapas de Referencia**

1. Creación del Conjunto de Datos Objetivo.
2. Preparación de los Datos
3. Minería de Datos

- **Productos de Trabajo**

1. Reporte sobre la Recolección Inicial de Datos
2. Reporte sobre la Descripción de los Datos.
3. Reporte sobre la Exploración de los Datos.
4. Reporte sobre la Verificación de la Calidad de los Datos
5. Reporte sobre la Selección de los Datos.
6. Reporte sobre la Limpieza de los Datos.
7. Atributos Derivados
8. Registros Generados
9. Datos Unidos
10. Datos con nuevo formato
11. Datasets
12. Reporte sobre los Datasets
13. Documento sobre la técnica de Modelado.
14. Plan de Prueba
15. Descripción del Modelo Resultante.
16. Informe sobre la Evaluación del Modelo

3.2.4.2. Etapas de Referencia

1. Etapa I: Creación del Conjunto de Datos Objetivo

Propósito

Definir los requisitos iniciales para la creación del Conjunto de Datos Objetivo que serán necesitados por el proceso de análisis e interpretación en el sistema de vigilancia epidemiológica.

Actividades

1. Recolección inicial de Datos.
2. Descripción de los Datos.
3. Exploración de los Datos
4. Verificación de la Calidad de los Datos.

Actividad 1. Recolección inicial de Datos.

- **Descripción**

Esta actividad tiene como objetivo adquirir a través del proyecto los datos o acceso a los datos para iniciar apropiadamente el proceso de Análisis. Esta actividad incluye contar con datos cargados en caso de ser necesitados para lograr su entendimiento, por ejemplo, dado el caso en que se piense utilizar una herramienta en particular para el análisis, se debe preprocesar o cargar los datos en esta herramienta

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

Planeación de los Requisitos de los Datos

- Planee que tipo de información relacionada con los datos es necesitada (por ejemplo: es necesario utilizar datos que presentan determinados atributos o se requiere información adicional).
- Compruebe si toda la información exigida (para lograr el proceso de Análisis) está realmente disponible.
- Considere que la recolección de datos desde diferentes fuentes puede ocasionar problemas de calidad cuando se asocian (por ejemplo: localización de registros basados en los criterios del cliente puede sacar a flote inconsistencias de formatos, invalidez de datos, etc.).

Criterios de Selección

- Especifique los criterios de selección (por ejemplo: ¿Qué atributos son necesarios para propósitos específicos dentro del proceso de Minería de Datos? ¿Qué atributos han sido

identificado o cuales son irrelevantes? ¿Cuántos atributos se pueden manejar con las técnicas escogidas?).

- Seleccione las tablas o registros de interés
- Seleccione los datos de interés dentro de una tabla o registro.
- Calcule cuánto tiempo, en términos de información almacenada, se deba utilizar aún cuando no todo sea aprovechable, por ejemplo, tenemos disponible información relacionada con los últimos 18 meses, quizás solamente sean necesarios los últimos 6 meses para el ejercicio.

Para mayor información en [CDM 03] y [HPR 96].

- **Productos de Trabajo**

- Reporte sobre la Recolección inicial de datos

En este reporte se debe enumerar todo lo que sea posible, los diversos datos utilizados dentro del proyecto, junto con cualquier selección de requerimientos en la utilización de datos que requieren mayor detalle. El reporte de la recolección de datos también debe definir si existen atributos que son relativamente más importantes que otros.

Recuerde que cualquier valoración de la calidad de los datos no debe hacerse únicamente de fuentes de datos individuales sino también de datos que vienen de fuentes unidas o fusionadas. Los datos unidos pueden presentar problemas que no existen en las fuentes de datos individuales a causa de inconsistencias entre las fuentes.

Actividad 2. Descripción de los Datos.

- **Descripción**

Esta actividad tiene como finalidad describir las propiedades de los datos adquiridos antes de ser procesados e informar sobre los resultados obtenidos.

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

Análisis Volumétrico de los Datos

- Identifique los datos y métodos de captura
- Acceda a las fuentes de los datos.
- Utilice análisis estadístico si es apropiado.
- Reporte las tablas y sus relaciones.
- Compruebe el volumen de los datos, número de múltiplos, complejidad.
- ¿Los datos contienen entradas disponibles, en caso que necesiten ser normalizados o modificados para cargar la herramienta de Análisis?

Tipos de Atributos y Valores

- Compruebe la accesibilidad y disponibilidad de los atributos.
- Cheque los tipo de atributos (numéricos, simbólicos, taxonómicos etc.).
- Compruebe el rango de valores de los atributos.
- Analice las correlaciones de los atributos.
- Entienda el significado y los valores de cada atributo en términos de la organización.
- Para cada atributo calcule estadística básica (por ejemplo: distribución, promedio, máximo, mínimo, desviación estándar, varianza, moda etc.).
- Analice estas estadísticas y relaciónelas con los resultados y significados en términos de la organización.
- ¿Los atributos son aplicables para lograr los objetivos específicos del proceso de Análisis?
- ¿El significado del atributo es usado consistentemente?
- ¿Es necesario equilibrar los datos (dependiendo de las técnicas usadas para el modelado)?

- **Productos de Trabajo**

Reporte sobre la descripción de los datos

Este reporte describe los datos que han sido adquiridos incluyendo: el formato, la calidad, el número de registros y campos dentro de cada tabla, identificadores de las tablas y cualquier otra característica externa que se han descubierto en los datos.

Para mayor información en [JYN 92] y [CDM 03].

Actividad 3. Exploración de los Datos.

- **Descripción**

Esta actividad conlleva a inquietudes dentro del proceso de Análisis que pueden ser resueltas a través de queries, visualizaciones y reportes. Estos análisis pueden dirigir los objetivos del proceso de Minería de Datos y contribuir o refinar la descripción de los datos y la calidad de los reportes.

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

Exploración de datos

- Analice las propiedades de los atributos más interesantes con mayor detalle (por ejemplo: sacar estadísticas en sub-poblaciones de interés).
- Identifique las características de las sub-poblaciones.

Formulación de Supuestos para futuros análisis

- Considere y evalúe la información y los descubrimientos emitidos a través de los reportes descriptivos de los datos.
- Forme hipótesis e identifique acciones.
- Transforme las hipótesis, si es posible, hasta obtener las metas del proceso de Análisis.
- Clarifique los objetivos del proceso de Análisis para hacerlos más precisos. Búsquedas sin respuestas no son necesariamente inútiles pero es preferible búsquedas dirigidas directamente a encontrar los objetivos de la organización.
- Realice un análisis básico para verificar las hipótesis.

Para mayor información en [CDM 03].

- **Productos de Trabajo**

- Reporte sobre la exploración de los datos

Este reporte describe los resultados de esta actividad incluyendo los primeros descubrimientos o hipótesis iniciales y su posible impacto en el resto del proyecto. El reporte puede cubrir gráficas y esquemas que indican las características de los datos o conducirnos a interesantes subconjuntos de datos para mayor inspección

Actividad 4. Verificación de la Calidad de los Datos.

- **Descripción**

Esta actividad tiene la tarea de examinar y verificar la calidad de los datos, para ello se puede dirigir preguntas tales como: ¿Los datos están completos (cubren todos los casos requeridos)? ¿Son correctos o contienen errores y si hay errores que tan comunes son? ¿Hay valores perdidos en los datos? En ese caso, cómo son representados, dónde ocurren y que tan comunes son?

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

- Revisión de Información clave

- Identifique valores especiales y catalogue su significado
- Compruebe el campo de aplicación (por ejemplo: están todos los posibles valores representados).
- Realice una comprobación de la información clave.
- ¿El significado de los atributos y el contenido de los valores encajan adecuadamente?
- Identifique atributos perdidos y campos en blanco.
- Denote los datos perdidos.
- Compruebe los atributos con diferentes valores que tienen significados similares (por ejemplo: bajo en grasa, dieta).

- Compruebe que los valores estén escrito correctamente (por ejemplo: generalmente se presenta errores de mayúsculas y minúsculas).
- Compruebe las desviaciones, decida llegado el caso si una desviación es ruido o puede indicar un fenómeno interesante.
- Compruebe la credibilidad de los valores, por ejemplo: todos los campos tienen los mismos o casi los mismos valores.
- Haga una revisión de cualquier atributo que puede dar respuestas que chocan con el sentido común (por ejemplo: adolescentes con el alto ingreso).

Inconsistencias entre las Fuentes

- Compruebe consistencias y redundancias entre las diferentes fuentes.
- Plantee cómo tratar con el ruido.
- Detecte el tipo de ruido y qué atributos son afectados.
- En algunos casos puede ser necesario excluir algunos datos desde que ellos no afecten ni positiva ni negativamente en el proceso de análisis.
- Revise todas las supuestas en tanto sean válidas o no determinen información sobre los datos y el conocimiento

Para mayor información en [JMF 00].

• **Productos de Trabajo**

Reporte sobre la Verificación de la calidad de los Datos

Enumere los resultados de la verificación de la calidad de los datos; si hay problemas de calidad, enumere las posibles soluciones.

2. Etapas II: Preparación de los Datos

Propósito

Organizar y definir actividades, objetivos y recursos enmarcados hacia la preparación de los datos dentro del proceso de análisis de información en el sistema de vigilancia epidemiológica, además de definir los mecanismos para asegurar el seguimiento de la fase.

Actividades

1. Selección de Datos.
2. Limpieza de los Datos.
3. Estructuración de los Datos.
4. Integración de los Datos.
5. Formato de los Datos.
6. Creación de Datasets.

Actividad 1. Selección de Datos.

- **Descripción**

Esta actividad tiene como finalidad seleccionar los datos, es decir, decidir acerca de los datos que serán utilizados para llevar a cabo el proceso de análisis.

El criterio de decisión debe incluir relevancia sobre la calidad, restricciones técnicas y límites en los volúmenes y tipos de datos

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

Ejecución de Pruebas

- Recolectar apropiadamente datos adicionales (de diferentes fuentes, tanto internas como externas).
- Ejecutar pruebas significativas y correlacionales para decidir si los campos deben ser incluidos.
- Considerar la utilización de probar otras técnicas (por ejemplo una solución rápida puede involucrar la reducción del tamaño de las pruebas en el Dataset o la herramienta no es capaz de trabajar con el Dataset completamente lleno, en estos casos es necesario repetir pruebas). En ocasiones puede ser útil realizar pruebas de diferente importancia, es decir, datos o información que contienen diferentes atributos o de diferentes valores dentro de un mismo atributo.
- Chequear las técnicas disponibles para hacer un muestreo de los datos.

Revisión de Criterios de selección de datos

- Revisar los criterios de selección de datos (Fase IV, Etapa I, Actividad 1), en vista de los resultados obtenidos a partir de la Actividad de Selección de Datos es necesario incluir o excluir un conjunto de datos.
- Seleccionar diferentes subconjuntos de datos (por ejemplo: diferentes atributos, datos que reúnen solo ciertas condiciones).
- Basado en el criterio de selección de datos, decida si un atributo es más importante que otro y el peso que tiene el atributo acordado o seleccionado. Decida como manejar este peso basado en el contexto (es decir, aplicación, herramientas, etc.).

Para mayor información en [JYN 92], [CDM 03] y [FNG 97].

- **Productos de Trabajo**

Reporte sobre la Selección de los Datos

Este Reporte describe raciocinios sobre la inclusión o exclusión de datos dentro del proceso de análisis. Al final obtendremos un listado de los datos que serán usados o excluidos y las razones por las cuales se han tomado este tipo de decisiones.

Actividad 2. Limpieza de los Datos.

- **Descripción**

Esta actividad tiene como propósito elevar la calidad de los datos al nivel requerido por las técnicas de análisis seleccionadas. Esto puede involucrar selección de subconjuntos de datos limpios, inserción de valores predeterminados convenientemente o técnicas más ambiciosas como la estimación de datos perdidos

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

Tratamiento de Ruido

- Reconsiderar acerca de cómo tratar con el tipo de ruido observado.
- Corregir, remover o ignorar el ruido.

Valores Específicos

Decidir como tratar con valores específicos y su significado. El tema de valores específicos puede dar origen a muchos resultados extraños y que deben ser examinados cuidadosamente. Ejemplos de valores específicos pueden originarse a través de resultados tomados de una encuesta o estudio donde algunos interrogantes no fueron preguntados ni contestados. Esto podría producir un valor de '99' para un dato desconocido. Por ejemplo, 99 para el estado civil o afiliación política. Los valores específicos también pueden originarse cuando los datos son truncados, por ejemplo '00' para personas con 100 años de edad o todos los automóviles con 100.000 Km. de recorrido.

Revisión de Criterios de selección de datos

Revisar los Criterios de selección de datos (Fase IV, Etapa I, Actividad 1), en vista de la experiencia durante la actividad de limpieza de datos es necesario incluir/excluir un conjunto de datos.

Para mayor información en [JYN 92], [JYN 93].

- **Productos de Trabajo**

Reporte sobre la Limpieza de los Datos

Este Reporte describe las acciones y decisiones que fueron tomadas para resolver los problemas de calidad de los datos reportados durante el proceso de verificación de su calidad.

El reporte debe localizar qué problemas de calidad se resolvieron o si aún se mantienen en los datos. En caso de que se utilicen estos datos en el proceso de Análisis es preciso describir los posibles efectos que pueden tener sobre los resultados

Actividad 3. Estructuración de los Datos.

- **Descripción**

La finalidad de esta actividad es estructurar los datos seleccionados de tal forma que se puedan utilizar en el proceso de análisis, incluye operaciones de preparación y construcción de datos, así como la producción de atributos derivados, completando de esta manera nuevos registros o transformando valores para atributos existentes.

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

Utilización de Herramientas

- Comprobar los mecanismos de construcción disponibles en la lista de herramientas sugeridas para el proyecto.
- Decidir si es mejor ejecutar la construcción desde o sin la herramienta (es decir qué es más eficiente, exacto, repetitivo).

Atributos Derivados

- Decida si algún atributo debe normalizarse (por ejemplo cuando usamos algoritmos de agrupamiento entre edad e ingreso en pesos, el ingreso será el que predomina).
- Considere adicionar nueva información sobre atributos de relativa importancia adicionando nuevos atributos.
- ¿De qué manera los atributos perdidos pueden ser contruidos o atribuidos? (Decida el tipo de construcción, por ejemplo: agregado, promedio, inducción).
- Agregue nuevos atributos a los datos accedidos.
- Antes de adicionar atributos derivados, intente determinar si facilitan y de qué manera lo hacen dentro del proceso o en el algoritmo del modelo. No se debe derivar atributos simplemente para reducir el número de atributos entrantes.
- Otro tipo de atributos derivados es la transformación de atributos simples, que normalmente satisfacen las necesidades de acople en las herramientas para el modelo.

Transformación de Atributos simples

- Especifique los pasos o medios necesarios para realizar la transformación de los Atributos, por ejemplo, cambio a binario de un atributo numérico.
- Realice los pasos de transformación.

Para mayor información en [CDM 03].

• **Productos de Trabajo**

Atributos Derivados

Los atributos derivados son nuevos atributos que se construyen a partir de uno o más atributos existentes dentro del mismo registro. Un ejemplo podría ser:

área = largo * ancho

La necesidad de construir Atributos Derivados surge del resultado de experiencias que indican que no únicamente los datos almacenados en bases de datos u otras fuentes sean el único tipo de datos que deben ser usados en la construcción del modelo. Los atributos derivados podrían construirse debido a que:

- En un segundo plano, el conocimiento nos convence de que algún hecho es importante y que debe ser representado aunque actualmente no tengamos atributos para representarlos.
- La utilización de algoritmos para el modelo se ocupa solo de cierto tipo de datos, por ejemplo estamos utilizando regresión lineal y sospechamos que hay ciertos no lineamientos que no han sido incluidos en el modelo.
- El resultado propuesto por las técnicas de modelado puede hacer pensar que ciertos hechos no están cubiertos.

Registros Generados

Los registros generados son indiscutiblemente nuevos registros, que agregan nuevo conocimiento o representan nuevos datos que no están representados de otra manera, por ejemplo, si los datos están segmentados, puede ser útil generar un registro para representar los miembros de un prototipo en cada segmento con el fin de mejorar la etapa de procesado.

Actividad 4. Integración de los Datos.

- **Descripción**

Esta actividad tiene como finalidad realizar una integración de los datos, aplicando métodos con los cuales la información se combina desde múltiples tablas u otras fuentes de información para crear nuevos registros o valores.

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

Unificación de Datos

El unificar tablas se refiere a la unión de dos o más tablas que tienen información diferente sobre los mismos objetos. Ejemplo: un Centro de Salud tiene una tabla con información sobre las características generales de cada Puesto de Salud (ej. Ubicación, el tipo de Puesto de Salud), otra tabla con datos resumidos sobre Gastos Operacionales (ej. Gastos operacionales, porcentaje de gastos del año anterior) y otra con información sobre la demografía del área. Cada una de estas tablas contiene un registro para cada puesto de Salud. Éstos pueden unirse en una nueva tabla combinando campos de las tablas fuente.

- **Productos de Trabajo**

Datos Unidos

Se refiere a la generación de nuevos registros, y agregaciones. La agregación se refiere a operaciones donde nuevos valores son computados con la información recopilada de múltiples registros y/o tablas.

Actividad 5. Formato de los Datos.

- **Descripción**

La principal tarea de esta etapa es dar un nuevo formato a los datos. Las transformaciones en el formato se refieren principalmente a modificaciones sintácticas hechas a los datos que no cambian su significado pero que podrían ser requeridas por la herramienta que utiliza el modelo.

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

Reestructuración de Atributos y registros Reordenados

Podría ser importante cambiar el orden de los registros en el dataset. Quizás la herramienta del modelo requiere que los registros sean ordenados conforme al valor del atributo resultante. Una situación común es que se piden archivos del dataset que inicialmente se encuentran organizados de alguna manera pero el algoritmo necesita que ellos se encuentren ubicados al azar. Por ejemplo, cuando se usa redes neuronales generalmente es mejor que los registros estén presentados en forma aleatoria, aunque algunas herramientas manejan esto automáticamente sin intervención del usuario.

Cambios en el valor del nuevo Formato

Se refiere a cambios sintácticos hechos para satisfacer los requerimientos de la herramienta utilizada para aplicar las técnicas de modelado

- **Productos de Trabajo**

Datos con nuevo formato

Algunas herramientas tienen requisitos en el orden de los atributos, tal como el primer campo es un identificador único para cada registro o el último campo es el campo del resultado total.

Actividad 6. Creación de Datasets

- **Descripción**

Esta actividad es de gran importancia ya que es requerida para iniciar adecuadamente el proceso de análisis de los datos y es sobre el Dataset sobre el cual se va a trabajar para poder obtener información útil que ayude a la toma de decisiones en el área de la Salud pública.

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

Las técnicas utilizadas para esta actividad se refieren a las empleadas durante las actividades 1 a 5 de la Etapa II “Preparación de los Datos”.

- **Productos de Trabajo**

Datasets

Los Datasets son conjuntos de datos que se han procesado adecuadamente y que serán utilizados en la próxima Etapa.

Reporte sobre los Datasets

Este reporte describe detalladamente los atributos y características presentes en los Datasets generados durante la preparación de los datos.

3. Etapa III: Minería de Datos

Propósito

Organizar y definir actividades, objetivos y recursos para el desarrollo de la Etapa de Minería de Datos, además de definir los mecanismos para asegurar el seguimiento de la fase

Actividades

1. Selección de la Técnica de Modelado.
2. Generación del Diseño de la Prueba.
3. Construcción del Modelo.
4. Evaluación del Modelo.

Actividad 1. Selección de la Técnica de Modelado.

- **Descripción**

La finalidad de esta Actividad es seleccionar adecuadamente la técnica de modelado más apropiada que realice el análisis de los datos.

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

Conocimiento y Disponibilidad de Herramientas que soporten las técnicas seleccionadas.

Como primer paso se debe seleccionar una técnica de modelado. En el caso en que se apliquen técnicas múltiples es necesario realizar esta Actividad para cada técnica por separado.

Se debe tener en cuenta que no todas las herramientas y técnicas de modelado se pueden aplicar para resolver todo tipo de problema. Existen técnicas y herramientas que solo manejan solo cierto tipo de información, datos o atributos. Entre estas herramientas y técnicas también existen requisitos y restricciones de ley, el conocimiento de éstas nos permiten tener un panorama más amplio que nos ayuden a seleccionar la más adecuada.

Es posible que dentro de las herramientas disponibles para realizar el Análisis no contemos con la más adecuada, por eso es necesario realizar un estudio exhaustivo de cada una de ellas y de sus potencialidades al inicio del proyecto.

Para mayor información en [JHO 03].

- **Productos de Trabajo**

- Documento sobre la Técnica de Modelado

Este documento realiza una descripción de la Técnica o Técnicas de Modelado utilizadas. Muchas técnicas de modelado hacen supuestos específicos sobre los datos, calidad de los datos o el formato de los datos, estos factores también serán descritos en este documento, para esto podríamos tener en cuenta:

- Definir cualquier supuesto hecho por la técnica sobre los datos (por ejemplo calidad, formato, distribución).
- Comparar estos supuestos con aquéllos que se encuentran en el Informe de Descripción de Datos.
- Asegurar que estos supuestos se mantengan y regresen a la Fase de Preparación de Datos si es necesario.

Actividad 2. Generación del Diseño de la Prueba.

- **Descripción**

Antes de construir un Modelo, un procedimiento necesita ser definido para probar la calidad y validez del mismo, es por esto que esta actividad es de gran importancia ya que tiene como finalidad generar el diseño de la prueba y de su éxito dependerá conseguir los objetivos propuestos al inicio del proyecto.

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

- Diseño de la Prueba

Surge de la necesidad de generar un mecanismo o procedimiento para probar la calidad del modelo y su validez. Por ejemplo, en la utilización de técnicas de modelado como clasificación es

muy común la generación de datos erróneos como medida de calidad para los datos. Es por esto que en ocasiones es necesario separar el Dataset del conjunto de datos disponibles para preparación y pruebas, el modelo es construido sobre el conjunto de preparación y su calidad es estimada en el conjunto de prueba. También se debe tener en cuenta los siguientes apartes:

- Comprobar por separado los diseños de prueba existentes para cada objetivo del proceso de Análisis.
- Decidir con respecto a los pasos necesarios (número de iteraciones, número de despliegues etc.).
- Preparar los datos requeridos para la prueba

Para mayor información en [JHO 03].

- **Productos de Trabajo**

- Plan de Prueba

- Se describe el plan deseado para los modelos de preparación, prueba y evaluación. Un componente primario del diseño es decidir como dividir el Dataset disponible entre los datos de preparación, datos de prueba y el conjunto de pruebas de validación.

- **Actividad 3. Construcción del Modelo.**

- **Descripción**

- Esta actividad tiene como finalidad ejecutar la herramienta de modelado sobre el Dataset acondicionado para construir uno o más modelos.

- **Actores**

- Analista del Sistema

- Especialista en Minería de Datos

- **Técnicas**

Configuración de Parámetros

Con cualquier herramienta de modelado, hay a menudo un gran número de parámetros que pueden ser ajustados. Enumere los parámetros y sus valores escogidos, junto con la razón por la cual se escogieron.

Modelado

- Ejecute la técnica seleccionada sobre el Dataset de entrada para producir el modelo.
- Procese los resultados del análisis (por ejemplo corrección de reglas, árboles de despliegue).

- **Productos de Trabajo**

Descripción del Modelo Resultante

Describa el modelo resultante y evalúe su exactitud esperada, robustez y sus posibles limitaciones. Es necesario reportar sobre la interpretación de los modelos y cualquier dificultad encontrada. Para realizar esta descripción podría tener en cuenta:

- Describir cualquier característica del modelo actual que puede ser útil para el futuro.
- Registrar la configuración de los parámetros usados para producir el modelo.
- Dar una descripción detallada del modelo y cualquier característica especial.
- Para los modelos basados en reglas, enumere las reglas producidas más cualquier valoración de exactitud y alcance del modelo global o a través de reglas.
- Para los modelos opacos o no transparentes, enumere cualquier información técnica sobre el modelo (como la topología de la red neuronal) y cualquier descripción conductual producido por el proceso de modelado (como exactitud o sensibilidad).
- Describa la conducta e interpretación del modelo.
- Haga conclusiones con respecto a los patrones encontrados en los datos (cualquiera); a veces el modelo revela hechos importantes sobre los datos sin un proceso de valoración separado (por ejemplo que el resultado o la conclusión se duplica en una de las entradas)

Para mayor información en [JHO 03] y [CDM 03].

Actividad 4. Evaluación del Modelo.

- **Descripción**

Esta Actividad tiene como objetivo evaluar el modelo y asegurar que el modelo seleccionado encuentre el criterio de éxito en el proceso de Análisis. Esto es solamente una evaluación técnica basada en el resultado de las tareas de modelado

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

Construcción de Tablas o Campos Adicionales

Pueden construirse tablas o campos adicionales para determinar la manera cómo el modelo está prediciendo.

Revisión de los parámetros de configuración

De acuerdo a la evaluación del modelo, revise la configuración de los parámetros y reajústelos. Itere la evaluación y construcción del modelo hasta que encuentre el mejor modelo.

- **Productos de Trabajo**

Informe sobre la Evaluación del Modelo

- Resuma los resultados de esta etapa, enumere la calidad de los modelos generados (por ejemplo en términos de exactitud) y compare su calidad con respecto a otros.
- Los resultados de la prueba concuerdan con los procedimientos o estrategias planteados para la prueba (por ejemplo: preparación y prueba, validación, etc.).
- Compare los resultados de la evaluación y déles una interpretación.
- Cree clasificaciones de los resultados con respecto al éxito y el criterio de la evaluación

- Seleccione los mejores modelos.
- Interprete los resultados en términos de la organización (hasta donde posible en esta fase).
- Consiga comentarios sobre modelos por personas expertas en el campo del Análisis de información.
- Compruebe la viabilidad del modelo.
- Compruebe el impacto para los objetivos planteados durante el proceso de análisis.
- Compruebe el modelo contra la base de conocimiento dado para ver si la información descubierta es nueva y útil .
- Compruebe la fiabilidad del resultado.
- Analice la potencialidad para el despliegue de cada resultado.
- Si hay una descripción verbal del modelo generado (por ejemplo, a través de reglas), evalúe estas reglas; estas son lógicas, son factibles, hay demasiados o pocos, están en contra del sentido común?
- Evalúe los resultados.

3.2.5. FASE V: DESPLIEGUE E INTERPRETACIÓN

En este apartado se presentan los lineamientos a seguir que el modelo de desarrollo propone en su quinta fase para el despliegue e interpretación de información dentro de un sistema de vigilancia de información epidemiológica.

3.2.5.1. Descripción General de la Fase

- **Objetivos Generales**

Presentar nuevo conocimiento a las partes interesadas que den soporte a la toma de decisiones en la solvencia de conflictos o inquietudes dentro del sector de salud pública.

- **Objetivos Específicos**

- Realizar un plan que despliegue los resultados del proceso de análisis.
- Realizar un plan de supervisión y mantenimiento.
- Generar un reporte final.

- **Justificación**

Esta fase es de suma importancia puesto que refleja el resultado de los procesos de almacenamiento y análisis de información y soporta la toma de decisiones dentro de la organización, el cual es uno de los objetivos de mayor importancia.

- **Etapas de Referencia**

1. Definición del Plan de Despliegue
2. Definición del Plan de Supervisión y Mantenimiento.
3. Definición del Informe final.

- **Productos de Trabajo**

1. Plan de Despliegue.
2. Plan de Supervisión y Mantenimiento.
3. Informe Final
4. Presentación Final.

3.2.5.2. Etapas de Referencia

1. Etapa I: Definición del Plan de Despliegue

Propósito

Realizar la captura de requerimientos para la definición y ejecución del plan de despliegue, donde se refleje los resultados del proceso de almacenamiento y análisis de información.

Actividades

1. Elaboración del Plan de Despliegue.

Actividad 1. Elaboración del Plan de Despliegue.

- **Descripción**

La finalidad de esta Actividad es tomar los resultados de la evaluación y concluir una estrategia para el despliegue e interpretación de los resultados del proceso de Análisis

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

Con base en experiencia de grupos de desarrollo en despliegue de información [CDM 03] se plantean:

- Resumen de los resultados del despliegue.
- Desarrollo y evaluación de planes alternativos para el despliegue
- Decidir durante la aparición de conocimiento diferente o información resultante.
- ¿Cómo el conocimiento o la información se propagará a los usuarios finales?
- Discutir sobre cada modelo de despliegue o resultado del software se va a utilizar.
- Identificar los posibles problemas cuando despliegue los resultados (dificultades del despliegue).

- **Productos de Trabajo**

Plan de Despliegue

Es un Resumen de la estrategia para el despliegue e Interpretación incluyendo los pasos necesarios y cómo hacer para ejecutarlos.

2. Etapa II: Definición del Plan de Supervisión y Mantenimiento

Propósito

Realizar la captura de requerimientos para la definición y ejecución del plan de supervisión y mantenimiento, donde se defina una estrategia adecuada que conlleven y den soporte a la toma de decisiones.

Actividades

1. Elaboración del Plan de Supervisión y Mantenimiento.

Actividad 1. Elaboración del Plan de Supervisión y Mantenimiento.

- **Descripción**

El objetivo de esta Actividad es generar un plan de supervisión y mantenimiento. La supervisión y el mantenimiento son temas importantes si los resultados del proceso se convierten en parte del día a día y del entorno de la organización. Una preparación cuidadosa de una estrategia de mantenimiento ayuda a evitar innecesarios periodos de uso incorrecto de los resultados. Para supervisar el despliegue de los resultados del proceso de Almacenamiento y Análisis, el proyecto necesita un plan detallado sobre el proceso de supervisión. Este plan toma en cuenta el tipo específico de despliegue.

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

- Compruebe si existen aspectos dinámicos o cambiantes (es decir, qué cosas podrían cambiar en el ambiente?).
- ¿Cómo será supervisada la exactitud?
- ¿Cuándo se debe dejar de utilizar el modelo o los resultados del proceso de Análisis? Identifique criterios (validez, umbral de exactitud, nuevos datos, cambios en el dominio de la aplicación, etc.) ¿Qué podría pasar si el modelo o el resultado dejaría de utilizarse? (Actualice el modelo, establezca un nuevo proyecto, etc.).
- ¿Los objetivos de la organización usados por el modelo cambiarán con el tiempo?

Para mayor información en [JHO 03] y [CDM 03]

- **Productos de Trabajo**

Plan de Supervisión y Mantenimiento

Este plan es el resumen de las estrategias de supervisión y mantenimiento incluyendo los pasos necesarios para ejecutarlos.

3. Etapa III: Definición del Informe Final

Propósito

Divulgar y emitir los resultados del proceso de almacenamiento y análisis de información a las partes interesadas.

Actividades

1. Elaboración del Informe Final.

Actividad 1. Elaboración del Informe Final.

- **Descripción**

El director del proceso y su equipo escriben un reporte final. Dependiendo del plan de despliegue, este informe puede ser sólo un resumen del proyecto y de sus experiencias o puede ser una presentación final donde se expongan de manera comprensiva los resultados del proceso de Almacenamiento y Análisis.

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

- Identifique qué reportes se necesitan (presentación de diapositivas, sumario de administración, resultados detallados, explicación de los modelos, etc.).
- Analice qué tan bien han sido encontrados los objetivos iniciales del proceso de Análisis.
- Identifique los grupos que serán designados para el reporte.
- Haga un diseño preliminar de la estructura y contenido de los reportes.
- Seleccione los descubrimientos que serán incluidos en el reporte.

- **Productos de Trabajo**

- Informe Final

Al final del proyecto, habrá por lo menos un reporte final dónde todos los hilos se reúnen y encajan. Así pues, identificando los resultados obtenidos, el informe también debe describir el proceso, mostrar en que gastos se han incurrido, definir cualquier desviación del plan original, describir los planes de implementación y hacer algunas recomendaciones para un trabajo futuro.

- Presentación Final

Así como un Informe Final, puede ser necesario hacer una Presentación Final para resumir el proyecto - quizás ante el patrocinador o la gerencia, por ejemplo. La Presentación normalmente contiene un subconjunto de la información contenida en el Informe Final, pero estructurada de una manera diferente.

3.2.6. FASE VI: EVALUACIÓN

En este apartado se presentan los lineamientos a seguir que el modelo de desarrollo propone en su quinta fase referente a la Evaluación de los proceso de almacenamiento, análisis e interpretación de información en un sistema de vigilancia epidemiológica.

3.2.6.1. Descripción General de la Fase

- **Objetivos Generales**

Evaluar los resultados obtenidos por parte del cliente, del equipo de desarrollo y gestionar características nuevas que se puedan añadir dependiendo de si aportan nuevo conocimiento.

- **Objetivos Específicos**

- Evaluar los resultados obtenidos
- Revisar los procesos de almacenamiento y análisis de información
- Determinar próximos pasos que complementen los procesos de almacenamiento y análisis.

- **Etapas de Referencia**

1. Evaluación de Resultados
2. Revisión del Proceso.
3. Determinación de próximos pasos

- **Productos de Trabajo**

1. Informe de Resultados del Proceso de Almacenamiento.
2. Informe de Resultados del Proceso de Análisis.
3. Informe de Revisión del Proceso de Análisis.
4. Informe de la decisión tomada.

3.2.6.2. Etapas de Referencia

4. Etapa I: Evaluación de Resultados

Propósito

Evaluar los resultados generados por los procesos de almacenamiento, análisis e interpretación de información dentro del sistema de Vigilancia epidemiológica..

Actividades

17. Evaluación de resultados del proceso de almacenamiento.
18. Evaluación de resultados del proceso de análisis.

Actividad 1. Evaluación de resultados del proceso de almacenamiento.

- **Descripción**

Basados en el caso del negocio, actividad 2 de la Fase I, se realizan las evaluaciones al proceso de almacenamiento implementado teniendo en cuenta los costos y beneficios alcanzados.

- **Actores**

Director del proceso

- **Productos de Trabajo**

Informe de Resultados del proceso de almacenamiento. documento que contiene la relación de los costos y los beneficios obtenidos en la implementación del proceso.

Actividad 2. Evaluación de Resultados del proceso de análisis.

- **Descripción**

La finalidad de esta etapa es evaluar los resultados y valorar el grado en que el modelo encuentra los objetivos de la organización y busca determinar razones por la cual el modelo no pueda llegar a

ser eficiente. Otra opción de evaluación es probar el modelo(s) con programas de prueba sobre la aplicación real siempre y cuando el tiempo y el presupuesto lo permitan.

Los resultados del proceso de Análisis de información cubren modelos que evidentemente se relacionan con los objetivos originales de la organización y otros resultados que no necesariamente se relacionan con ellos, pero que no dejan de ser importantes, ya que pueden mostrar desafíos adicionales, revelar información o leves indicadores que se podrían convertirse en el punto de partida de futuros procesos de análisis.

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

Comparación de resultados

- Interprete los resultados en términos de la aplicación.
- Compruebe el impacto de los objetivos del proceso de Análisis.
- Compruebe los resultados del proceso de Análisis con relación a la base de conocimiento suministrada para determinar si la información descubierta es novedosa y útil.
- Evalúe y calcule los resultados con respecto a los criterios de éxito de la organización ¿es decir, el proyecto ha logrado los Objetivos originales de la organización?
- Confronte los resultados de la evaluación e intérpretelos.
- Compruebe el impacto de los resultados sobre los objetivos iniciales de la aplicación.
- ¿Hay nuevos objetivos en la organización para ser dirigidos posteriormente en el proyecto o en un nuevo proyecto?
- Declare conclusiones para futuros proyectos de Análisis de información.

- **Productos de Trabajo**

Informe de Resultados del proceso de análisis

Una buena manera de definir los resultados totales del proceso de Análisis de información es mediante la ecuación:

RESULTADO = MODELO + DESCUBRIMIENTO

En esta ecuación estamos definiendo que el resultado total de un proyecto de Análisis de datos no es justamente el modelo, aunque es muy importante, pero también es de igual o mayor importancia el descubrimiento que nosotros logramos obtener.

Aunque el modelo esta directamente relacionado con las inquietudes de la organización, el descubrimiento no necesita ser relacionado con éstos.

1. Etapa II: Revisión del proceso

Propósito

Gestionar los errores y peticiones del cliente y usuarios con el fin de generar la versión final entregable.

Actividades

19. Revisión del proceso de análisis.

Actividad 1. Revisión del proceso de análisis.

- **Descripción**

El objetivo de esta actividad es realizar una revisión minuciosa y detallada del proceso de análisis de información con el fin de determinar si hay algún factor importante o proceso que de alguna manera se haya pasado por alto.

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

Los siguientes aspectos conllevan a obtener y utilizar técnicas para lograr mejor rendimiento en la revisión del proceso de análisis:

- Dé una apreciación global sobre el proceso de Análisis.
- Analice el proceso de análisis en cada etapa:
 - ¿En retrospectiva, fue necesario?
 - ¿Fue ejecutado óptimamente?
 - ¿De qué manera podría mejorarse?
- Identifique anomalías.
- Identifique posibles acciones alternativas, caminos inesperados durante el proceso.

Para mayor información en [JHO 03] y [CDM 03].

- **Productos de Trabajo**

Informe de Revisión del proceso de análisis

En este informe se debe resumir la revisión del proceso de análisis y se debe plantear sugerencias para actividades que han sido extrañas y/o deben repetirse.

2. Etapa III: Determinación de próximos pasos

Propósito

Definir, proponer o sugerir actividades y procesos que conlleven a un mejoramiento en el desarrollo de sistemas que soporten la toma de decisiones.

Actividades

20. Determinar próximos pasos.

Actividad 1. Determinar próximos pasos.

- **Descripción**

Con base en los resultados de las etapas anteriores relacionados con la valoración y revisión del proceso esta actividad es de suma importancia ya que es aquí donde se necesita decidir y definir si el proyecto concluyó para trasladarse a la siguiente fase de Despliegue e Interpretación, o si por lo contrario, se deben iniciar más iteraciones o establecer nuevos proyectos de Almacenamiento y Análisis.

- **Actores**

Analista del Sistema

Especialista en Minería de Datos

- **Técnicas**

Búsqueda de posibles acciones

- Discuta posibles acciones junto con las razones a favor y en contra de cada opción.
- Analice el potencial para el despliegue de cada resultado.
- Estime el potencial para mejorar de proceso actual.
- Compruebe los recursos restantes para determinar si ellos permiten adicionar procesos iterativos (o si recursos adicionales pueden estar disponibles para ello).
- Recomiende continuaciones alternativas.
- Refine el plan del proceso.

Toma de Decisiones

- Clasifique las posibles acciones.
- Seleccione una de las posibles acciones.

- **Productos de Trabajo**

Informe de la Decisión tomada

Documento que contiene la decisión acerca de cómo proceder junto con un análisis razonado.

4. DESCRIPCIÓN DE LA IMPLEMENTACIÓN DEL MAISIVE

Este capítulo contiene la descripción general de los procesos de almacenamiento, análisis e interpretación de información aplicados al Sistema de Vigilancia Epidemiológica, implementado por la Dirección Departamental de Salud del Cauca (DDSC), con el fin de validar y ejercitar el Modelo de Desarrollo para el Almacenamiento, Análisis e Interpretación de Información aplicado en un Sistema de Información para Vigilancia Epidemiológica (MAISIVE) propuesto en este trabajo de grado.

El Sistema de Información para Vigilancia Epidemiológica implementado por la DDSC se denomina SIGSA (Sistema de Información Geográfico para Salud). El SIGSA es continuo y sistemático, basado en la población que puede proveer un aviso oportuno de cambios en eventos o riesgos en salud y que permite un mejor desarrollo de programas. De rutina, tal sistema recolecta datos sobre tres clases de indicadores (eventos, factores de riesgo y estrategias de intervención) para establecer un sistema de aviso oportuno (identificar tendencias y asociaciones y hacer predicciones sobre eventos) y para establecer un sistema de desarrollo de programas (evaluar la necesidad para estrategias de intervención, plantear e implementar esas estrategias y evaluar su efectividad).

La implementación de los procesos de Almacenamiento, Análisis e Interpretación de Información aplicados al SIGSA permitirá a la sección de Vigilancia en Salud Pública de la Dirección Departamental de Salud del Cauca, la gestión de eventos epidemiológicos para dar soporte a la toma de decisiones basados en los Mapas Temáticos generados por la relación de información temática recolectada en cada unidad notificadora.

Para el entendimiento de la organización se establece su administración a través de dos entes: la Sección de Vigilancia en Salud Pública de la Dirección Departamental de Salud del Cauca y los Hospitales Municipales del Departamento del Cauca. En base a los procesos de información que se generen en estos dos entes se establecieron las necesidades y requerimientos del SIGSA, así como el establecimiento y análisis de los requerimientos para la implementación de los procesos

de almacenamiento, análisis e interpretación de esta información para soportar la toma de decisiones en el SIGSA.

Los actores de la organización que se han identificado son: los pacientes, la comunidad, el coordinador PAB y el Ministerio de la Protección Social. Como trabajadores de la organización se identificaron: los promotores de salud, auxiliares de enfermería, coordinadores de los programas de prevención y control, estadistas, médicos, director de Hospital, jefe de sección de Vigilancia en Salud Pública.

La siguiente es una descripción breve de los tipos de usuarios que interactúan con los procesos implementados con el MAISIVE en el SIGSA:

- *Estadista.* Persona encargada de obtener reportes consolidados de eventos epidemiológicos.
- *Coordinador sección de Vigilancia en Salud pública.* Persona encargada de enviar al ministerio de Protección social, reportes consolidados de eventos epidemiológicos.
- *Analista Epidemiológico.* Persona encargada de analizar los procesos de prevención, control y vigilancia oportuna de enfermedades de notificación obligatoria haciendo uso de los procesos para la toma de decisiones que contiene el sistema.
- *Administrador.* Persona encargada de administrar la aplicación del SIGSA y supervisar los procesos implementados para el soporte a la toma de decisiones.

En la Figura 1. se presenta el Diagrama de Casos de Uso del sistema SIGSA que ya se encuentra implementado como proyecto de un trabajo de investigación relacionado con este trabajo de grado. En la Figura 1. se indican los principales casos que implementa el sistema. Para una mayor profundización en la descripción de los casos de uso del sistema se puede consultar el Anexo B - Descripción de los procesos de almacenamiento, análisis e interpretación de información implementados en el sistema de vigilancia epidemiológica, donde se presentan los productos generados para su implementación.

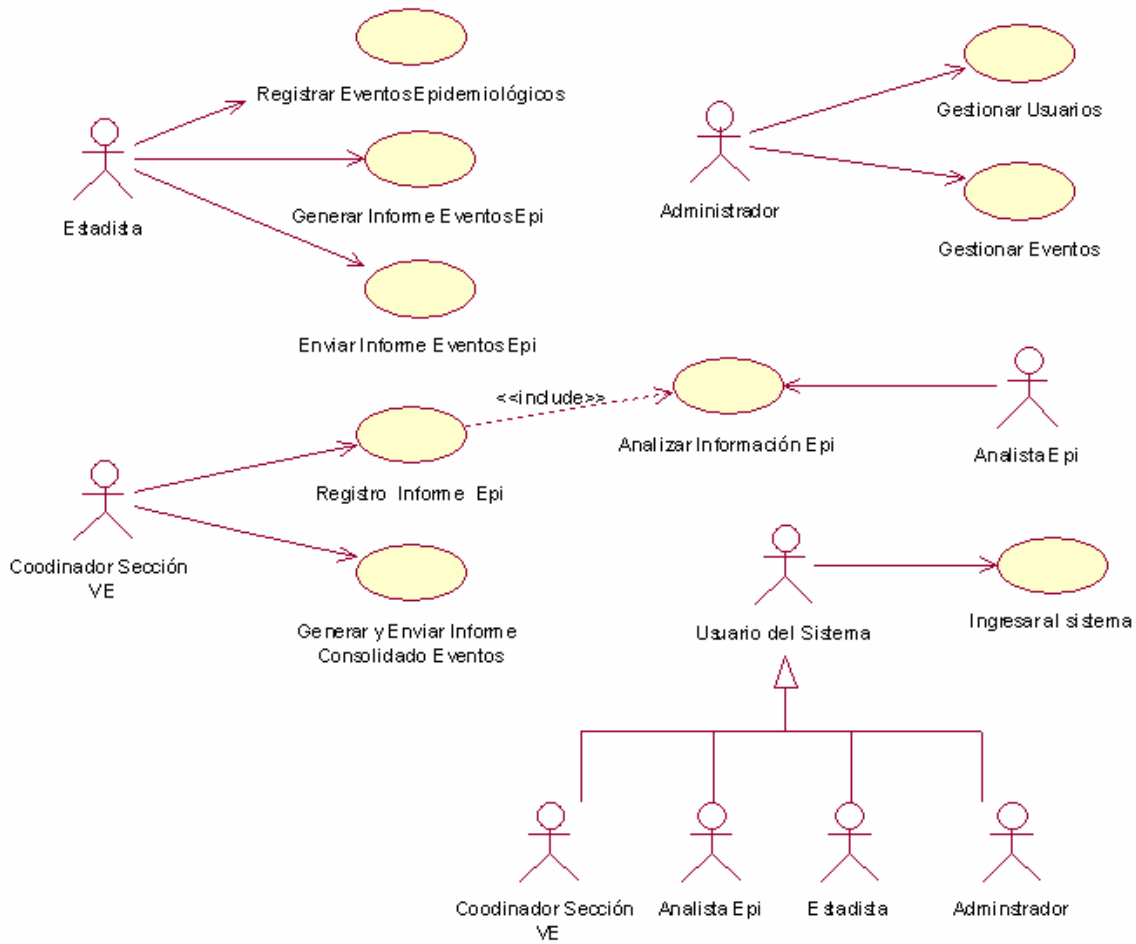


Figura 2. Diagrama de Casos de Uso del Sistema SIGSA

Los elementos constituyentes de la arquitectura de soporte para los procesos de almacenamiento representan la estructura total de datos, comunicación, procesamiento y presentación de información que existe para los usuarios finales que disponen de un equipo terminal dentro de la organización.

En la Figura 2. se presenta el ambiente de almacenamiento para el análisis de información en el SIGSA propuesto por el MAISIVE.



Figura 3. Ambiente de almacenamiento del Sistema SIGSA para análisis de información

Un análisis de información exitoso se soporta en la robustez de los procesos de almacenamiento. A su vez, esta robustez debe estar implementada sobre una adecuada arquitectura hardware y software. La correcta realización de todas las operaciones dentro de este ambiente concluirá en un sistema de soporte para la toma de decisiones exitoso.

Para el entendimiento de la arquitectura básica necesaria que soportará los procesos para análisis de información del sistema de vigilancia epidemiológica, se deben comprender las operaciones que se ejecutan dentro de la arquitectura del almacén de datos y los componentes que las realizan. Este entendimiento nos permite encontrar los elementos constituyentes de la arquitectura de soporte para el análisis de información. Elementos como:

- Base de datos operacional / Nivel de bases de datos externos.
- Nivel de acceso a la información.
- Nivel de acceso a los datos.
- Nivel de directorio de datos.

- Nivel de gestión de proceso.
- Nivel de mensaje de la aplicación.
- Nivel del almacén de datos.
- Nivel de organización de datos.

Para una mayor profundización en la arquitectura del proceso de almacenamiento del sistema de vigilancia epidemiológica se puede consultar el Anexo A, que contiene una descripción detallada de los aspectos conceptuales que sirvieron como guía y soporte para la construcción del MAISIVE; y el Anexo B, donde se presentan los productos generados para su implementación.

Vale la pena aclarar que por el enfoque que se le ha dado al documento, en el cual el principal objetivo es presentar y describir con un alto nivel de detalle el Modelo de Desarrollo propuesto, para este capítulo de descripción de los procesos de almacenamiento, análisis e interpretación de información implementados en el SIGSA tan solo se ha hecho una breve exploración de la base conceptual y un entendimiento del SIGSA que la organización tiene implementado. La aplicación del MAISIVE y los productos de trabajo generados en su implementación se encuentran descritos en mayor detalle en el Anexo B de este documento.

5. CONCLUSIONES Y RECOMENDACIONES

1. Un sistema de vigilancia con soporte para la toma de decisiones puede utilizarse para monitorear, evaluar y reajustar las intervenciones planteadas para la prevención y control de eventos que aumentan los riesgos de enfermar y morir de una comunidad e igualmente proporciona el insumo necesario para la prestación óptima de la atención en salud de la población.
2. Al plantear un sistema de vigilancia epidemiológica robusto para la toma de decisiones se debe recordar que tiene como finalidad la prevención y control de los problemas y situaciones que modifican el estado de salud de la población, lo que exige de un trabajo activo y permanente, multidisciplinario, intersectorial ya que los problemas y sus soluciones requieren de la participación de diferentes sectores. Partiendo de que el análisis de información epidemiológica se puede aplicar a cualquier evento en salud como una enfermedad, un factor de riesgo o una discapacidad.
3. Uno de los objetivos del sistema de vigilancia, es que la información generada como producto del mismo, al ser analizada se convierta en insumo para la toma de decisiones y se utilice en la orientación de las políticas sociales, asignación o resignación de recursos, inclusión de nuevos programas, acordes con las necesidades reales de la comunidad a fin de contribuir a mantener o mejorar las condiciones de vida de la población.
4. El sistema de vigilancia epidemiológica para la toma de decisiones requiere para su funcionamiento de un sistema de información operativo que brinde el conocimiento sobre la situación de salud de un área, de un municipio o departamento y que permita identificar las variaciones de esta y proporcione los elementos necesarios para la toma de decisiones. La información requerida debe ser completa, actualizada, objetiva, veraz, dinámica y referirse a las personas en las que suceden los eventos, al lugar y al tiempo en que se presentan, a las pruebas diagnósticas que confirman o descartan los casos y a las actividades que se desarrollan tanto en el individuo como en la comunidad.

5. La transferencia, almacenamiento, procesamiento, análisis y acceso a la información epidemiológica requiere de recurso humano capacitado para la realización de esta actividad y de una red automatizada de comunicaciones a la cual tenga acceso el personal responsable en cada una de las etapas, dependiendo del rol o papel que desempeña en el sistema, para lo cual se debe establecer la normatividad necesaria para controlar y garantizar el funcionamiento del sistema.

6. Cuando se implementan procesos de almacenamiento para análisis de información, el impacto de cambios puede originarse por un cambio en el ambiente de la organización y/o cambio en la tecnología. Un cambio en el ambiente de la organización puede cambiar las necesidades de información de los usuarios. Así, el contenido del almacén de datos se puede ver afectado y las aplicaciones para análisis de información y toma de decisiones pueden requerir cambios. Por otro lado, un cambio en la tecnología puede afectar la manera en que los datos operacionales son almacenados en el sistema de información, lo cual implicaría un ajuste en los procesos de Extracción, Transporte y Carga para adaptar las variaciones presentadas. Un cambio de cualquiera de los ya expuestos impacta los sistemas operacionales. Un cambio en el ambiente operacional puede cambiar el formato, estructura o significado de los datos operacionales usados como origen para el diseño e implementación del almacén de datos para análisis de información. De esta forma serían impactados los procesos de Extracción, Transformación y Carga de datos.

7. El valor de la implementación de procesos de almacenamiento para análisis de información se refleja en la comprensión de tres dimensiones:
 - *Mejorar la Entrega de Información:* Información completa, correcta, consistente, oportuna y accesible. Información que la gente necesita, en el tiempo que la necesita y en el formato que la necesita.
 - *Mejorar el Proceso de Toma de Decisiones:* Con un mayor soporte de información se obtienen decisiones más rápidas; así también, los usuarios involucrados adquieren mayor confianza en sus propias decisiones y las del resto, y logra un mayor entendimiento de los impactos de sus decisiones.
 - *Impacto Positivo sobre los Procesos Organizacionales:* Cuando a la gente se le da acceso a una mejor calidad de información, la organización puede lograr por sí sola:

- Eliminar los retardos de los procesos organizacionales que resultan de información incorrecta, inconsistente y/o no existente.
 - Integrar y optimizar procesos organizacionales a través del uso compartido e integrado de las fuentes de información.
 - Eliminar la producción y el procesamiento de datos que no son usados ni necesarios, producto de aplicaciones mal diseñados o ya no utilizados.
8. La implementación de procesos de almacenamiento, análisis e interpretación de información es una estrategia de largo plazo. Al querer implementar estos procesos, se debe evaluar el costo y el valor considerando un período de tiempo razonable para obtener beneficios. El retorno sobre la inversión, se comienza a percibir bastante más tarde del tiempo en el cual se realizó la inversión inicial. Si se calcula costo/valor desde una perspectiva de corto plazo, los costos serán significativamente más altos en proporción al valor.
9. El éxito de la implementación de procesos de almacenamiento, análisis e interpretación de información no está en su construcción, sino en usarlos para mejorar procesos organizacionales, operaciones y decisiones.
10. Las bases de datos integradas están ofreciendo un volumen desmesurado de información a las organizaciones. Para mejorar el servicio que estas prestan o su competitividad se requieren métodos de análisis rápidos y automatizados. Sin embargo, las posibilidades de análisis que dispone el usuario son limitadas. Tanto las tecnologías relacionales, como OLAP poseen tremendas capacidades para navegar por el almacén de datos de la organización pero no son suficientes, se necesitan de herramientas capaces de estructurar y priorizar la información para las necesidades concretas del usuario. Herramientas para análisis de información como las provistas por la Minería de Datos y Visualización ofrecen esta posibilidad.
11. Las herramientas para análisis de información deberían evaluarse primero por su capacidad de alcanzar unos niveles de precisión aceptable, y posteriormente diseñarse para ser integradas sin transformaciones con la tecnología informática de proceso de datos que se dispone dentro de la organización.

12. La evaluación eficaz de la implementación de procesos de almacenamiento para análisis de información en un sistema de vigilancia epidemiológica, debe partir del deseo de todos y cada uno de los participantes de conocer la evolución del proceso, su participación en el mismo, las facilidades, las limitantes, la satisfacción y el cumplimiento de los objetivos; así como del valor que representa la información producto de la evaluación en los ajustes a realizar al sistema. La evaluación, debe estar orientada a la reflexión que el grupo de desarrollo y que los usuarios finales asumen frente a los objetivos obtenidos en y con la ejecución de las intervenciones, realizado al interior del proceso permanentemente o como mínimo en períodos programados según necesidades.

BIBLIOGRAFIA

- [CDM 03] CRISPDM. "CRoss Industry Standard Process for Data Mining". <http://www.crisp-dm.org>. 2003.
- [CES 01] Carlos Enrique Serrano Castaño. "Modelo para construcción de soluciones v1.0." Popayán, 9 de Mayo de 2001
- [CFL 95] Colín-Flores. "Reingeniería de procesos de negocios". Management Today 1995
- [DMI 99] Departamento de Ciencias Matemáticas e Informática. "<http://dmi.uib.es/~bbuades/riesgos/>". Universidad de les Illes Balears. 1999.
- [DSC 03] Dirección Departamental de Salud del Cauca. "SIGSA - Sistema de Información Geográfica en Salud", Popayán, 2003.
- [EHA 98] E. Hall. "Managing Risk : Methods for Software Systems Development2. Adison Wesley. 1998.
- [FCA 00] Fernando Carpani. Tesis de Maestría "CMDM: Un Modelo Conceptual para la Especificación de Bases Multidimensionales". Agosto 2000.
- [FNG 97] Fernando Martín Sánchez; Nieves Ibarrola De Andrés; Guillermo López Campos. "Minería, Visualización y Descubrimiento de Conocimiento en Bases de Datos". Unidad de Bioinformática – BIOTIC. 1997.
- [HFR 00] Holger Frietsch, "ROLAP, MOLAP, HOLAP: How to determine which technology is appropriate". Promatis AG. 2000.
- [HPR 96] H. Gill, P. Rao. "Data warehousing la integración de información para la mejor toma de decisiones", Prentice-Hall, 1996.
- [IGJ 99] Ivar Jacobson, Grady Booch, James Rumbaugh. "El proceso unificado de desarrollo de software". Addison Wesley. 1999.
- [ITG 03] Instituto Tecnológico Ciudad de Guzmán. "<http://www.itcg.edu.mx/ingsoft/modelote.htm>". México. 2003.
- [JAS 98] James A. Senn. "Análisis y Diseño de Sistemas de Información". Segunda Edición. McGraw-Hill, 1998.
- [JHO 03] José Hernández Orallo. "Análisis y Extracción de Conocimiento en Sistemas de Información: Datawarehouse y Datamining". Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia. Julio 2003.
- [JMF 00] Jean Michel Franco. "El Data Warehouse y el Data Mining". Ed. Gestión 2000–1997.
- [JSC 94] J. Sepúlveda y Col. "Aspectos Básicos de la Vigilancia en Salud Pública para los Años Noventa". Artículo especial, Salud Pública de México, 1994, Vol.36, 70-82.

- [JYN 92] J. Han, Y. Cai, y N. Cercone. "Knowledge Discovery in Databases: An Attribute-Oriented Approach", VLDB Conference, Vancouver, Canada, 1992.
- [JYN 93] J. Han, Y. Cai y N.Cercone. "Data-Driven Discovery of Quantitative Rules in Relational Databases". En IEEE Transactions on knowledge and data engineering, vol. 5(1), Febrero 1993.
- [LVE 00] Leopoldo Vega. "La salud en el contexto de la nueva salud pública". Editorial el manual moderno, México, D.F. – Santafé de Bogotá, 2000
- [MCG 03] Matilde Celma Giménez. "Bases de Datos. <http://www.dsic.upv.es/~mcelma>". Dpto. Sistemas Informáticos y Computación. Universidad Politécnica de Valencia. 2003
- [MDS 99] Ministerio de Salud. "Decreto 6 de 1999 por el cual se reglamenta el Sistema de Vigilancia en Salud Publica en Colombia", Ministerio de Salud, 1999.
- [MVP 04] Mercedes Vizcaíno Pinilla. "Vigilancia Epidemiológica", Universidad Nacional de Colombia, 2004.
- [ORL 02] Olivia R. y Liu Sheng. "Data Warehouse Design: Auxiliary Examples and Concepts". MIS Department. University of Arizona. 2002.
- [RKI 97] Ralph Kimball. "The Data Warehouse Toolkit". Wiley Computer Publishing, John Wiley & Sons Inc., New York, Chichester, Brisbane, Toronto, Singapore, 1997.
- [RLM 98] Ralph Kimball, Laura Reeves, Margy Ross, Warren Thornthwaite: "The Data Warehouse Lifecycle Toolkit". Wiley Computer Publishing, John Wiley & Sons Inc., New York, Chichester, Brisbane, Toronto, Singapore, Weinheim, 1998
- [RSC 01] Rational Software Corporation. "Business Modeling with the UML and RAS". Rational Software White Paper. Pp. 1-9, 2001.
- [UGP 96] U. M. Fayyad, G. Piatetsky-Shapiro y P. Smyth. "From Data Mining to Knowledge Discovery". Eds., AAAI Press, Menlo Park, California, 1996.
- [UGR 96] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth y R. Uthurusamy. "Advances in Knowledge Discovery and Data Mining". The AAAI Press, Menlo Park, California, 1996.
- [UMF 96] U. M. Fayyad. "Data Mining and Knowledge Discovery: Making Sense Out of Data". En IEEE Expert, vol. 11(5), Oct. 1996.
- [WIB 00] William H. Inmon. "Building the Data Warehouse". 2000.
- [WGC 91] W. J. Frawley, G. Piatetski-Shapiro y C. J. Matheus. "Knowledge Discovery in Databases: An Overview". Knowledge Discovery in Databases, G. Piatetsky-hapiro y W. Frawley, AAAI-MIT Press, Menlo Park, California, 1991.
- [WIC 00] William H. Inmon. "Creating the Data Warehouse Data Model from the Corporate Data Model". 2000.
- [WIU 00] William H. Inmon. "Using the Generic Data Model". 2000.