

Modelo Generalizado de Dato para mejora de calidad del análisis de ingeniería de mantenimiento

Gabriel M. Cuello

Petrobras Energía, UTE Puesto Hernandez, Ruta Prov. 6, Km 20 (8319) , Neuquén.

Resumen. De entre los sistemas informáticos de mantenimiento, se destacan algunos con facilidades para el control de costos, otros con mayores agregados para registro de datos del mantenimiento preventivo y predictivo. Los análisis de información se limitan a los reportes que dichos sistemas brinden y es poco frecuente encontrar reportes que diagnostiquen la calidad de la información almacenada o de seguimiento de gestión de algún sector; tampoco son lo suficientemente flexibles como para soportar cambios en los modelos de política de mantenimiento. Teniendo en cuenta el marco de nuestro sistema de gestión de mantenimiento, el ciclo de administración del trabajo (CAT) constituye el centro de la gestión de los trabajos diarios de operación y mantenimiento. El ciclo consta de seis etapas: Identificación, Planificación, Programación, Asignación, Ejecución y Análisis; esta última etapa asociada normalmente con los resultados del ciclo. Cada etapa del mismo tiene vinculada información relevante al trabajo a realizar y a la etapa en sí misma mediante la cual se controla el CAT. En el presente trabajo se exponen y fundamentan las bases para definir un modelo de dato conceptual de mantenimiento que permite no solo el control del CAT, sino también el análisis de sus resultados. Este modelo garantiza la independencia de los análisis respecto del software que se utilice para la gestión del mantenimiento gracias a que se construye en función de los análisis pretendidos e incluye características de un Data Warehouse para el análisis cruzado de costos y pérdidas de producción.

1. Introducción

1.1. El alcance de los datos para la inferencia del entorno

En una actividad determinada, la recolección de datos tiene como fin mediciones para la gestión y control de la actividad. No obstante, muchas de las veces la carga de datos no implica que las mediciones sean realizables; los reportes no son flexibles y, eventualmente logrado el reporte, la experiencia práctica no condice con los resultados que ofrece el reporte. Estos dos factores pueden explicarse a través de un modelo. Este es el punto de partida para fundamentar la re-ingeniería de la información que manejan los procesos de mantenimiento y para determinar hasta qué punto el sistema brinda información correlacionada con la realidad.

En la figura 1 se representa con un rectángulo el universo de sucesos relacionados con los quehaceres de la zona de trabajo de un supervisor; no sólo de los trabajos de mantenimiento sino todo lo que pasa en el área que afecte directa o indirectamente los quehaceres del supervisor. El óvalo representa la inferencia que un supervisor de mantenimiento experimentado puede lograr sobre cierta parte de las actividades del universo. Por ejemplo, personal de mantenimiento experimentado conoce que tal o cual maniobra de producción implica una modificación parcial de cañerías y que normalmente no se solicita a mantenimiento, sino que la supervisa personal de producción; es posible que el reporte diario del servicio de modificación no llegue claramente al

sistema de mantenimiento, pero se infiere la actividad. El círculo representa la información en el sistema de mantenimiento.

Las decisiones de gestión normalmente se toman basadas en algunos datos y principalmente en la inferencia sobre las actividades; no sobre los datos. Así, el desafío de ingeniería consiste en lograr colocar el círculo dentro del rectángulo, centrar el círculo dentro del óvalo expandiendo el área cubierta por el sistema hacia la que cubre el óvalo; tratar de que la información cargada en el sistema sirva cada vez más para mejorar la toma de decisiones.

Mientras que es posible crear medidas sobre distintos aspectos del trabajo, no siempre es práctico. Se debe balancear el esfuerzo y el costo involucrado para coleccionar los datos con su utilidad real o potencial para la toma de decisiones [5].

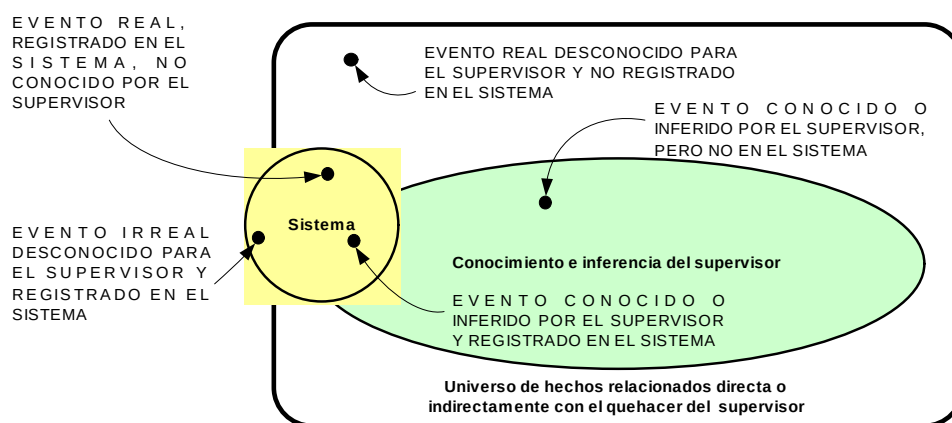


Figura 1: Posibles estados de los sucesos y su relación con el registro en el sistema y el conocimiento de un supervisor.

1.2. La gestión de mantenimiento. Toma de decisiones clásica.

En la figura 2 se muestra la visión inicial del área en cuanto a los aspectos de mantenimiento, que representa el deseo de mejora en cuanto a la maquinaria, y donde se representan necesidades básicas, comunes a diversos sistemas de gestión de mantenimiento.

Primordialmente, en las actividades de mantenimiento se distinguen dos grupos principales: actividades rutinarias y fallas. Los eventos rutinarios pueden ser debidos a políticas de mantenimiento preventivas y/o rutinas de operadores de producción que también registran acciones sobre los equipos; las fallas se registran para ser analizadas. En ambos casos, fue deseable registrar costos en mano de obra, materiales y pérdidas de producción; sin un adecuado sistema de gestión de mantenimiento y un ordenamiento de la información, esta tarea resultó ser difícil e imprecisa.

De los eventos rutinarios y de las fallas se analizan tiempos para optimizar rutinas, se realiza un ranking de fallas y se analizan las tareas de preventivo. Todo ello dedicado principalmente a aumentar el tiempo medio entre fallas, disminuir el tiempo medio para reparaciones y evaluar cuáles rutinas (de operación y mantenimiento) son realmente necesarias.

A pesar de lo básico del esquema de la figura 2, se concluyó que la importancia de la calidad de los datos de mantenimiento era un tema de base y que involucraba un cambio cultural importante.

Con el advenimiento de la implementación del proyecto de optimización de recursos operativos (ORO) en las áreas, se implantó el ciclo de administración de los trabajos (CAT) el cual provee un marco para la gestión del mantenimiento. El CAT constituye el centro de la gestión de los trabajos diarios de operación y mantenimiento; consta de seis etapas: Identificación, Planificación, Programación, Asignación, Ejecución y Análisis.

En principio, la etapa de análisis estaba asociada con los resultados del ciclo; el cuidado de la calidad de la información y el planeamiento de sus posibles usos futuros, condujo a agregar al análisis de resultados el flujo de información que cada etapa del CAT tiene vinculada consigo misma, mediante las cuales se controla la “salud” del ciclo.

La implementación del CAT ayudó a consolidar los procesos de mantenimiento e indirectamente contribuyó para poder precisar cuál información agrega valor al ciclo de mantenimiento.

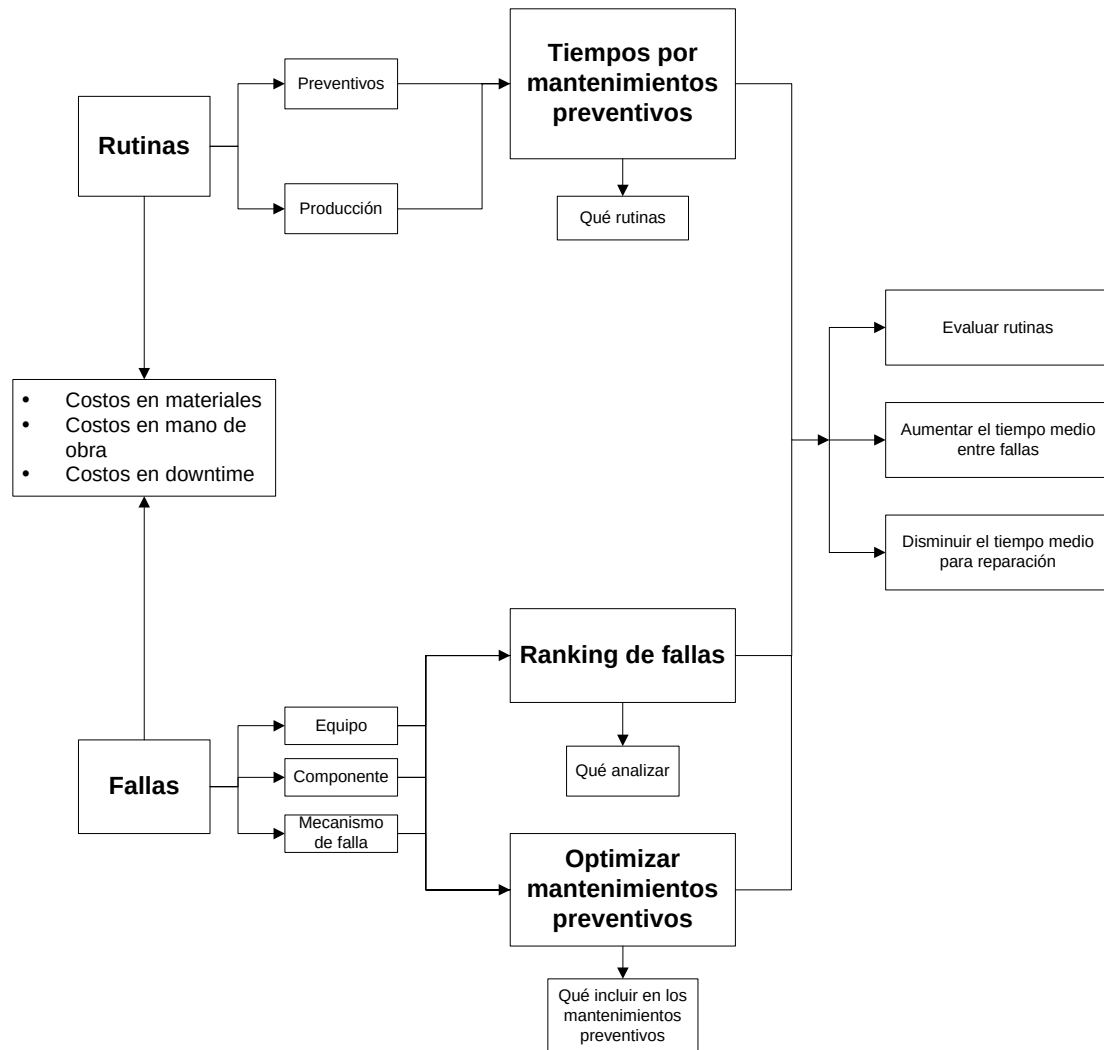


Figura2: visión inicial básica de ingeniería de mantenimiento

2. Toma de decisiones en mantenimiento basadas en modelos

La visión básica de mantenimiento pone en evidencia la importancia de la información. Antes de definir un concepto de dato genérico, se exploró qué otro tipo de actividades a nivel ingeniería de mantenimiento existían con el fin de otorgarle generalidad al dato para que aplique al mayor número de situaciones posible.

Para la toma de decisiones acerca del mantenimiento, reemplazo o confiabilidad de equipos [6], puede recurrirse al concepto de *optimización* mediante la construcción y solución de modelos matemáticos.

Una de las herramientas fundamentales para el enfoque científico sobre el manejo de tomas de decisión consiste en construir un modelo a evaluar, usualmente matemático, mediante el cual una variedad de decisiones alternativas pueden ser calculadas. Cualquier modelo es simplemente una

representación del sistema bajo estudio. En la aplicación de técnicas cuantitativas a problemas de gestión, el tipo de modelo utilizado es frecuentemente un modelo simbólico donde las relaciones del sistema se representan mediante símbolos y las propiedades del sistema mediante ecuaciones matemáticas.

En mantenimiento se deben tomar numerosos tipos de decisión. En el presente trabajo se consideraron los posibles modelos en lo que respecta a decisiones de reemplazo, inspección, overhaul y reparación de equipos, así como los de decisión de estructuras organizacionales, [6], listando las constantes y variables presentes en cada uno de los modelos con el propósito de definir cómo estructurar la información de mantenimiento para cubrir el mayor número de modelos posible. De esta manera, la gestión de la información del mantenimiento comienza desde lo global hacia el detalle: desde el modelo de toma de decisión hacia la forma de almacenar los datos en alguna herramienta para la gestión del mantenimiento. Al plantear tal análisis, se pretende crear mayor conciencia de uso de la información enfocada en el proceso que lo utiliza, evitando así los condicionamientos de las opciones particulares que posea un determinado software de gestión facilitando detectar sistemáticamente la distancia conceptual entre lo deseable (el proceso) y lo posible (medios de implementación del proceso).

Política	Modelo de eventos
Decisiones de reemplazo	
Tiempo óptimo para el reemplazo de equipos cuyo costo operativo aumenta con el uso	Aplica
Política de reemplazo óptima para dos máquinas, una de las cuales actúa de stand-by de la otra, cuando el costo operativo de las máquinas se incrementa con el uso	Aplica
Política de intervalo de reemplazo óptima que maximiza los beneficios netos devengados de operar un equipo un período de tiempo largo	Aplica parcialmente
Política de intervalo de reemplazo óptima que minimiza el costo total de mantenimiento y el reemplazo al operar un equipo un período de tiempo largo	Aplica
Política de reemplazo para equipamiento teniendo en cuenta las mejoras tecnológicas	Aplica parcialmente
Intervalo óptimo entre reemplazos preventivos en equipos sujetas a fallas	Aplica
Edad óptima para el reemplazo preventivo de equipos sujetos a falla	Aplica
Edad óptima para el reemplazo preventivo de equipos sujetos a falla, teniendo en cuenta el tiempo que toman los reemplazos en el caso de falla y en el caso de preventivo	Aplica
Intervalo óptimo entre reemplazos preventivos en equipos sujetas a fallas minimizando el downtime	Aplica
Reemplazos grupales: intervalo óptimo entre reemplazos de ítems sujetos a falla	Aplica
Decisiones de inspección	
Frecuencia de inspección óptima maximizando el beneficio económico	Aplica
Frecuencia de inspección óptima minimizando el downtime	Aplica
Frecuencia de inspección óptima para maximizar la disponibilidad de equipos utilizados en emergencias	Aplica
Plan óptimo de inspecciones cuya condición puede ser determinada solamente mediante la inspección, minimizando el costo total (inspección, reparación y no detección del equipo en falla)	Aplica parcialmente
Decisiones de overhaul y reparaciones	
Política de overhaul o reparación óptima de equipos sujetos a falla minimizando el costo total y producción perdida	No aplica
Decisiones de estructura organizacional	
Número óptimo de puestos de trabajo para atender carga de trabajo variable	Aplica
Esfuerzo de reparación óptimo de una cuadrilla de correctivos	Aplica parcialmente
Uso mixto óptimo de dos clases de equipos (o puestos de trabajo) similares para atender carga de trabajo variable	Aplica parcialmente
Número óptimo de máquinas por cuadrillas para repararlas	Aplica parcialmente
Tamaño óptimo de la fuerza laboral para atender carga de trabajo variable	Aplica

Tabla I: Políticas estudiadas

2.1. Coeficientes detectados

En la tabla I se resume el listado de las políticas estudiadas. De los modelos se desprende una lista de coeficientes relacionados con cada uno de ellos en particular, que listamos a continuación.

Las variables (coeficientes de los modelos) se determinaron a partir de los modelos. Se prestó especial interés a la fuente de información en la que se basaría cada una.

2.1.1. Decisiones de reemplazo

- Costo operativo
- Costo de mantenimiento (C_p)
- Costo de falla (C_f)
- Downtime por falla (DT_f)
- Downtime por efectuar el mantenimiento (DT_p)
- Costo de reemplazo
- Cantidad de reemplazos entre servicios
- Rédito económico (descontando los gastos de O&Mto)
- Densidad de probabilidad de los tiempos de falla del equipo
- Tiempo para ejecutar el preventivo (T_p)
- Tiempo para reparar la falla (T_f)

2.1.2. Decisiones de inspección

- Tiempo para realizar la inspección (T_i)
- Tiempo para reparar o reemplazar (T_r)
- Disponibilidad del equipo
- Costo de la inspección (C_i)
- Costo de reemplazo (C_r)

2.1.3. Decisiones acerca del overhaul y reparación

- Probabilidad de ir del estado I al J en un período si la acción a es llevada a cabo
- Costos asociados C IJ

2.1.4. Decisiones de estructura organizacional

- Tiempos medios entre arribos de pedidos de trabajo
- Tiempos medios de servicio por unidad de tiempo
- Costo del Downtime de un trabajo en espera (C_d)
- Costo de la cuadrilla o maquinaria

3. Modelo de eventos propuesto

En la figura 3 se esquematiza una línea de eventos para un equipo dado. El modelo generalizado de datos consiste en una sucesión de eventos cuya estructura se esquematiza en la tabla II. Salvo los coeficientes del punto 2.1.3, los demás encajan casi totalmente dentro de un esquema de eventos, quedando fuera de aquél coeficientes, entre otros, como la pérdida de producción del trabajo en espera; el evento está fuertemente orientado a que sea provisto al momento de generarse la necesidad particular.

Aquellos coeficientes en los modelos que pueden resultar de un grupo de eventos sí quedan incluidos como factibles; por ejemplo, la función densidad de probabilidad será obtenida a partir de varios eventos puntuales.

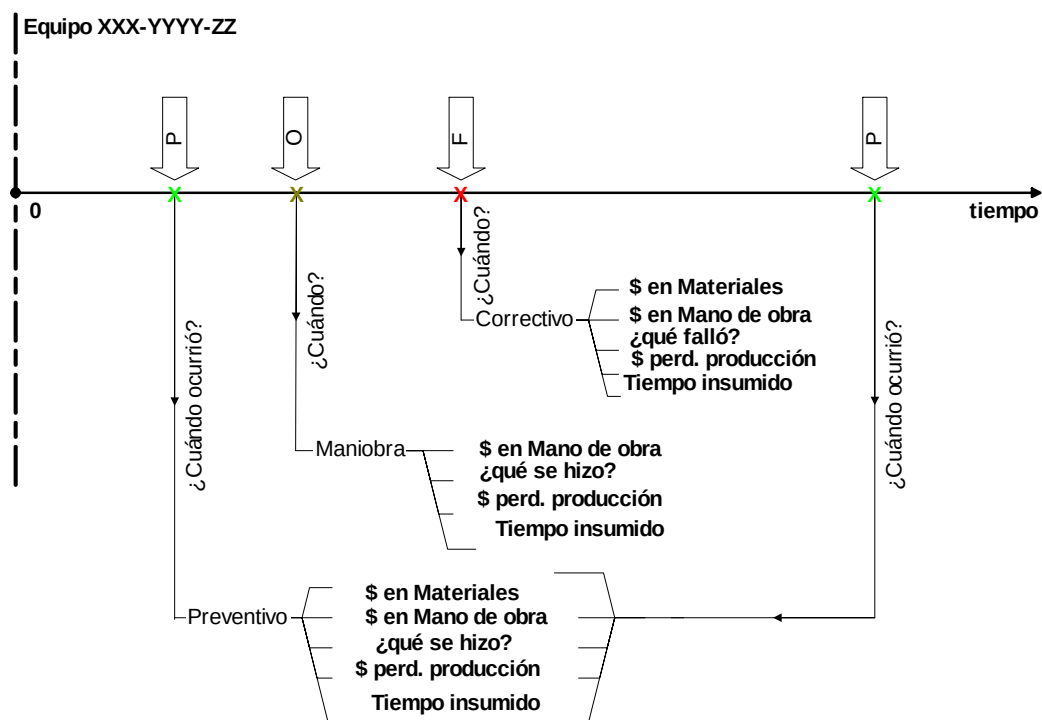


Figura 3: Modelo temporal de eventos por equipo.

Evento	Equipo	
	Fecha	
	Clase de actividad	Correctivo, preventivo, maniobras de producción, entre otras.
	Costos	Materiales
		Mano de obra
		Pérdida de producción
	Tiempo insumido por el evento	
	¿Quién ejecutó el evento?	
	¿Qué se hizo?	En el caso de correctivo, Componente-Mecanismo de falla
		En el preventivo, táctica de mantenimiento
		Si es una maniobra, cuál fue la actividad

Tabla II: Estructura conceptual de cada evento.

Como se puede apreciar en la tabla II, el evento reúne informaciones de diversas fuentes de origen; la información no está necesariamente en un solo sistema informático.

3.1. Aproximación parcial al uso del modelo de dato

En el área se lleva registro en diferentes sistemas del costo operativo, el downtime, las intervenciones a los pozos productores e inyectores, la mano de obra de personal de producción y mantenimiento. El modelo temporal de eventos integra toda esta información cubriendo casi todos los aspectos para el análisis. La figura 3 puede pensarse como una lista de eventos ordenada cronológicamente y con las columnas de costos, tiempos, ejecutantes, tipo del evento. Sobre ella es que se harán los diversos tipos de análisis.

En el área Puesto Hernández se está trabajando en pos de sostener instrumentar el modelo de evento. Como punto de partida, se implantó la recolección sistemática de fallas y actividades en el sistema de mantenimiento (MANTEC DB) mediante codificación ISO14224. En la figura 4 se muestra el detalle de la estructura de códigos [Cuello2004].

Para controlar la calidad de la información que se carga en el sistema MANTEC DB, se ha mantenido durante los últimos tres años hasta la fecha, una codificación del tipo de la figura 4, controlándose la calidad de la misma a través de sistemas auxiliares que detectan el mal uso de la codificación.

Los costos de mano de obra y materiales están integrados en el sistema MANTECDB, pero aún presenta dificultades la vinculación entre el registro del downtime en el sistema de producción y el evento que se registra en el sistema de mantenimiento MANTECDB.

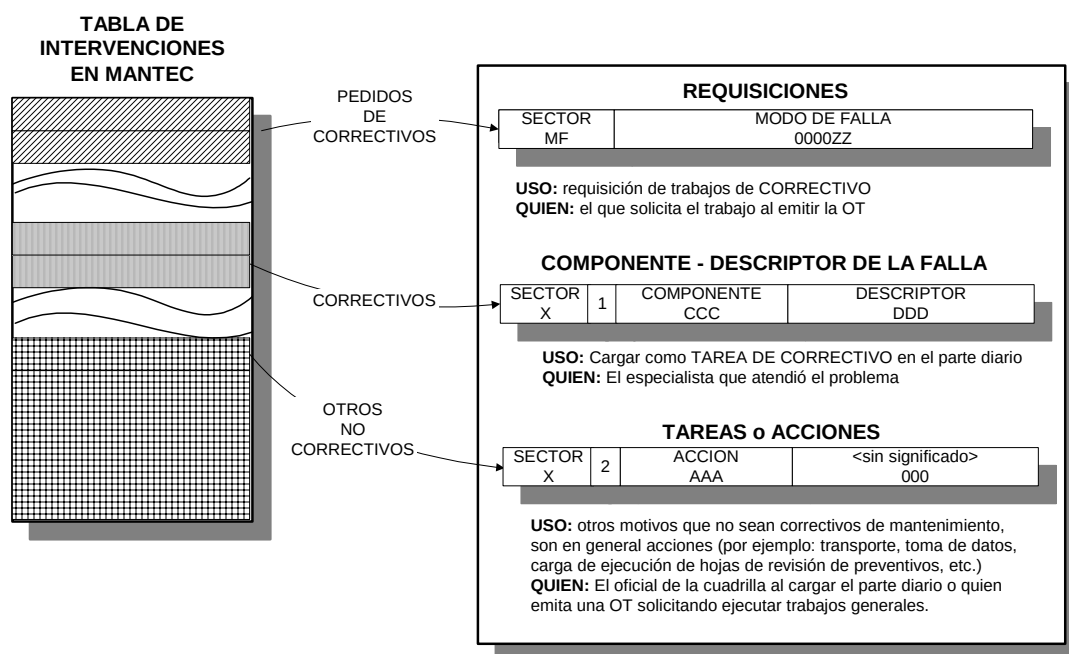


Figura 4: Codificación de las actividades de producción y mantenimiento.

3.2. Ejemplo de aplicación del modelo de dato a una política de mantenimiento.

A manera de ejemplo, se describe someramente el proceso de análisis para uno de los modelos cubiertos por el modelo de dato genérico.

Consideremos la política de intervalo óptimo entre mantenimientos preventivos de equipos sujetos a fallas. En la figura 5 se esquematiza el proceso para determinar el intervalo óptimo. Dado un cierto grupo de equipos, mediante el grupo de códigos componente-descriptor de falla se obtiene un conjunto de líneas de tiempo cronológicas para las fallas que se completa con la lista de equipos del mismo grupo que no ha presentado fallas. Luego se convierte el grupo cronológico en un gráfico de líneas de tiempos de supervivencia; este último conjunto sirve para determinar estadísticamente la

forma de la densidad de probabilidad $f(t)$ al ajustar una recta a la nube de puntos resultante en un gráfico *tasa de falla acumulada vs. tiempo de supervivencia*. Los costos de los eventos (C_f y C_p) y el downtime correspondiente completan las variables que requiere el modelo económico y son obtenidos a partir del registro de eventos que involucran el grupo de equipos bajo análisis.

En la figura,

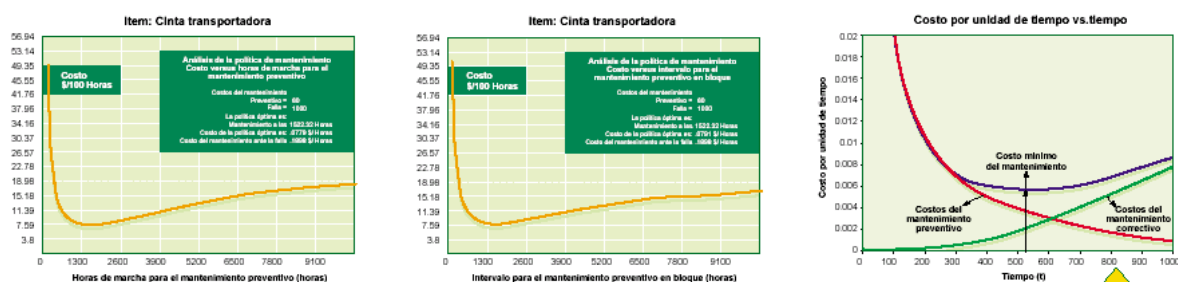
$$C(t_p) = \frac{C_p + C_f \cdot H(t_p)}{t_p}$$

es el modelo a optimizar y $H(t_p)$ representa el número esperado de fallas en el intervalo $(0, t_p)$, que está dado por

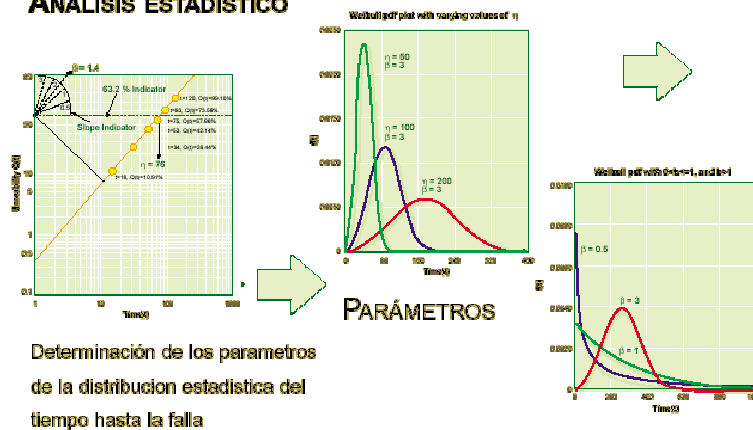
$$H(t_p) = \sum_{i=0}^{t_p} [1 + H(t_p - i - 1)] \cdot \int_i^{i+1} f(t) dt,$$

considerando a t_p subdividido en forma discreta. $H(t_p)$ es la parte estadística que utiliza la función densidad de probabilidad $f(t)$.

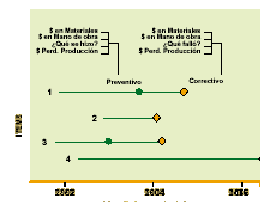
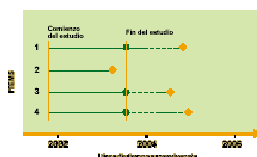
TOMA DE DECISIONES



ANÁLISIS ESTADÍSTICO



Determinación de los parámetros de la distribución estadística del tiempo hasta la falla



MODELO ECONÓMICO

$$C(t_p) = \frac{C_p + C_f \cdot H(t_p)}{t_p}$$

CONSOLIDACIÓN COSTOS

Figura 5: Ciclo de análisis del intervalo óptimo entre preventivos para equipos sujetos a fallas.

Al correr el modelo calculándolo para diferentes valores de t_p , se puede graficar la curva de costo por unidad de tiempo para cada intervalo de preventivo. Si la curva presenta un mínimo local, ésta será la zona donde se elegirá el t_p que presenta el costo mínimo (óptimo para la política); por el contrario, si la curva no presenta mínimos locales y es monótonamente decreciente, entonces conviene decidir trabajar a rotura el grupo de equipos.

3.3. Consideraciones para completar el contenido del modelo de dato mediante el uso de Data Warehouses

Para fundamentar la tendencia de organización de la información de mantenimiento, bajo este título se esboza someramente el negocio de la producción de petróleo y gas en un área petrolera tipo, sus sistemas informáticos (OLTP, *Online Transaction Processing*) asociados, se describen las actividades que el área realiza y, aclaradas condiciones de borde para su aplicación, se fundamenta la existencia de un Data Warehouse y algunos Data Marts para utilizar a modo de consolidación de todos los datos relacionados con cada evento.

Es claro que el negocio de la empresa también incluye sectores centralizados adicionales tales como finanzas, abastecimientos, legales, nuevos negocios, áreas de downstream, con actividades también altamente críticas y estratégicas, no obstante, se presentarán argumentos como para soportar la existencia de un Data Warehouse que sea extrapolable a todo nivel de la empresa teniendo en cuenta las condiciones de borde propuestas.

Un Data Warehouse es un conjunto de datos integrados orientados a una materia, que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una administración. Es un conjunto orientado porque los datos que se encuentran en Data Warehouse están relacionados a la operatoria de la empresa y son integrados porque los datos son tomados de distintos sistemas de la empresa, recopilados y almacenados en una forma coherente.

3.3.1. Descripción del negocio

Una área petrolera dada, tiene como objetivo de negocio ser rentable produciendo petróleo y gas. Por producir se entiende al acto de transportar fluido desde la perforación en la roca reservorio, tratarlo y entregar a la transportista el fluido tratado. Todo este proceso se realiza de acuerdo con las directrices corporativas que cada empresa adopte, que cubren aspectos de calidad, seguridad, cuidado del medio ambiente y la salud ocupacional.

La parte operativa principal de un área tipo dada está organizada por sectores que llevan a cabo las tareas relacionadas con una parte específica de cada pozo. Esta descripción no incluye los demás sectores existentes en la empresa, parte complementaria del negocio total (finanzas, exploración, abastecimientos, legales, etc.) que actúan de forma centralizada.

La unidad productora básica es el pozo productor. En la figura 6 se observa el esquema de un pozo donde se indican las áreas de influencia de cada sector: Reservorios, Ingeniería de producción, Producción. Se detallan esquemáticamente sus funciones:

- **Reservorios:** estudia el reservorio, hace seguimiento de la respuesta de producción de cada pozo y sus respectivas capas para ajustar los modelos simulados, estima sus reservas y hace el estudio de factibilidad de nuevos proyectos de explotación de reservas.
- **Ingeniería de Producción:** calcula y diseña la instalación de fondo, analiza y determina el régimen extractivo para cada pozo, realiza mantenimiento correctivo y preventivo de las instalaciones de fondo.
- **Producción:** opera y mantiene todos los equipos de superficie, realiza seguimiento diario de lo producido. Los equipos de superficie abarcan desde las líneas de conducción hasta las estaciones, las estaciones de bombeo y plantas de tratamiento y entrega.

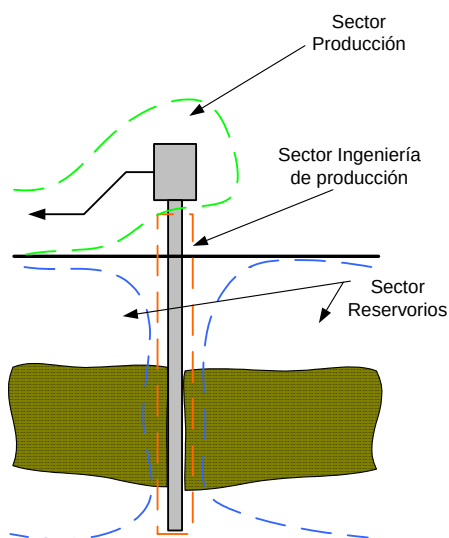


Figura 6: división del ámbito de trabajo para cada sector del área.

3.3.2. Sistemas OLTP y su relación con los sectores del área

Cada sector sigue, día a día, distintos aspectos relacionados con la evolución del comportamiento de los pozos y toma decisiones en función de ello. Existen diferentes bases de datos transaccionales para realizar registros relevantes. En la tabla II se describen a continuación las principales y qué sectores son los usuarios principales.

Sistema		Uso por parte de los sectores		
Sistema OLTP	Descripción	Reservorios	Ingeniería de producción	Producción
OFM	Sistema de seguimiento de la producción e inyección para explotación secundaria por formación y por pozo	Lo utiliza para monitorear la producción del reservorio y ajustar modelos en el simulador. A veces decide maniobras de mantenimiento de fondo que propone a ingeniería de producción.	Cálculo de instalaciones de fondo	Decisiones operativas para regular en superficie la producción de pozos y sistema de inyección secundaria.
TOW	Sistema de reporte de producción y mermas.	Alimenta el sistema OFM con datos de producción e inyección	Análisis con producción	Seguimiento diario de lo producido y lo perdido (mermas, downtime)
MANTEC	Sistema de gestión del mantenimiento de superficie	NO APLICA	NO APLICA	Transacciones diarias de cada trabajo que se realiza en superficie
DIMS	Sistema de gestión del mantenimiento de fondo	Análisis extractivo	Transacciones diarias de cada trabajo que se realiza en fondo	Consulta de sistema y reparaciones.
JDE	Sistema contable de la empresa	Imputaciones contables	Imputaciones contables	Imputaciones contables

Tabla II: Diversos OLTPs y su relación con los sectores del área.

3.3.3. Clasificación de las actividades de análisis que cada sector y jefatura realizan

Cada sector maneja parte de su tiempo en forma local, interna, y parte interactuando con los demás sectores. Las actividades típicas de un área petrolera pueden calificarse como sigue.

- análisis del régimen extractivo
- análisis de falla del equipamiento de fondo y de superficie
- informe mensual (con indicadores fijos, definidos corporativamente) por sector y su análisis
- análisis de los procesos de gestión

- análisis de distribución del agua de inyección para explotación secundaria
- análisis de la evolución del presupuesto
- generación del presupuesto anual
- reuniones intersectoriales
- análisis de los procesos de gestión

En la figura 7 se muestra una primera aproximación a cómo surgen las preguntas acerca de los datos en los sistemas. En general, las actividades individuales requieren sólo consultas a su respectivo OLTP. Sin embargo, a medida que las preguntas se realizan a un nivel cada vez más macro, surge la necesidad de consultas que involucren a varios OLTP.

El diagrama ilustra el flujo de datos para la consolidación de información de tres sistemas fuente (MANTEC, JDE y TOW) en un archivo de base de datos centralizado (ANALISIS.MDB), el cual a su vez se conecta con bases de datos de usuarios.

Sistemas Fuente:

- MANTEC:** Contiene una carpeta denominada "TABLAS" que genera "Consultas de acondicionamiento para consolidar tablas MANTEC".
- JDE:** Contiene un archivo denominado "LIBRO AA" que genera "Consultas de acondicionamiento para consolidar tablas JDE".
- TOW:** Contiene una carpeta denominada "TABLAS" que genera "Consultas de acondicionamiento para consolidar tablas TOW".

Proceso de Consolidación:

Las consultas generadas por cada sistema se dirigen a una columna centralizada en el archivo **ANALISIS.MDB**, etiquetada como "Consultas de acondicionamiento para consolidar tablas en forma cruzada entre sistemas". Esta columna recibe datos de las carpetas "TABLAS tbl..." de cada sistema fuente.

Conexión con Bases de Datos de Usuarios:

El archivo **ANALISIS.MDB** está conectado a dos bases de datos de usuarios:

- Base de datos Usuario 1:** Recibe datos de las consultas de MANTEC y JDE.
- Base de datos Usuario 2:** Recibe datos de las consultas de JDE y TOW.

 Las conexiones se establecen a través de "Vínculos ODBC".

Normalmente, cada consulta de gestión responde a la necesidad de tomar decisiones. Una empresa que dispone de información dispersa, no organizada, ubicada en distintas máquinas de usuarios responsables de algún sector, tarda mucho en encontrar y consolidar la información que busca para tomar una decisión; entonces, si una gerencia debe tomar una decisión en un 100% de tiempo dado y se pasa por ejemplo el 70% del tiempo buscando la información, sólo dispone del 30% restante

para analizarla y definir acciones a seguir. En cambio, si la información estuviera ordenada con facilitando la búsqueda de información, revertiría el proceso de decisión: 30% buscando datos, 70% meditando la decisión.

3.3.5. Condiciones de borde para implementar un dataware y sus Data Marts

Más que el presupuesto de la implementación, uno de los factores determinantes para decidir la implementación de Data Warehouses y Data Marts, es la necesidad o no de información ordenada que tenga la línea gerencial cuando el volumen de información es importante.

La necesidad de orden o no en los datos puede relacionarse con los siguientes aspectos:

Estilos de gestión. Se trata de un cambio cultural. Es humanamente complejo cambiar la forma de tomar decisiones; en algunos casos se confía más en la experiencia e inferencia que en la información ordenada y sistemática. En contraparte, decidir sólo a partir de los datos puede conducir a decisiones basadas en datos débiles.

Suponer que la percepción de que experiencia e inferencia alcanzan para tomar decisiones, tiene como desventajas:

- personalización del cargo en vez de sistematizar e impersonalizar su función
- desalentar la recolección sistemática de datos confiables
- en el mejor de los casos, generar OLTPs parciales difícilmente integrables al resto de los sistemas
- duplicación de los datos

El uso exclusivo de datos con poca atención a la experiencia, también induce a errores, pero por otras razones:

- no alentar la profundidad técnica inherente en los datos (conocen de sistemas pero no cuentan con la experiencia de campo suficiente como para determinar qué registrar)
- alto riesgo al tomar decisiones si es que no hacen hincapié para asegurar la calidad de los datos
- conclusiones desacertadas en función de datos no representativos de la realidad del área, sea porque hay datos incompletos, de poca credibilidad o de poca calidad.

Los dos estilos de gestión tienen en común el fin que persiguen. Normalmente una secuencia típica entre jefe y supervisor podría ser:

1. De tal tema: ¿cuáles son las causas que más contribuyen al problema?
2. ¿Qué se está haciendo para reducir las causas más importantes?
3. ¿En cuánto tiempo se piensa ver resultados al respecto?

En el primer caso, la jefatura basa sus decisiones en la experiencia empírica acumulada con las que realiza la inferencia, mientras que en el segundo caso, intenta fundamentarse con los hechos plasmados en los datos. El equilibrio es lo deseable; buena información, buena experiencia.

Contexto normativo de las empresas. Otro aspecto por el que la automatización de los datos y su consulta es importante se debe a un contexto de empresa. Actualmente, muchas empresas petroleras han comenzado a certificar normas de calidad y a someterse a evaluaciones de calidad oficiales (PNQ de Brasil, por ejemplo) buscando mejores calificaciones a nivel público, lo que implica mejores posibilidades de crédito financiero.

Dichas evaluaciones consultan numerosos aspectos de la empresa y, por ello, tener integrada la información que demandan es clave. Clave por dos razones: a) para responder de forma confiable y b) porque parte de la evaluación mide el grado de informatización de los procesos de la empresa.

Nuevos procesos. Cuando la empresa decide implementar nuevos procesos o sub-procesos, se los debería medir para seguir su grado de avance durante la implementación como para asegurar el

cambio cultural. En este caso la necesidad de gerencia consiste en medir los procesos y no los resultados del proceso en sí. El desconocimiento de los procesos y sub-procesos que componen cada sector, puede complicar el relevamiento de tal información. En cuanto a este tema, podría decirse que los datos deben responder tanto por la medición del proceso como por los resultados que éste arroja.

Asociación datos - software específico. Una duda de quienes tienen el poder de decidir implementar un data mart o data warehouses, es ¿para qué implementar esto si se está migrando de sistema? La idea de dato abstracto, generalizado y vinculado con los procesos de la empresa, no está suficientemente difundida. Creer que las preguntas que se puedan hacer al sistema OLTP dado estén condicionadas por su capacidad de brindar reportes es un error frecuente.

Sin limitar la visión a un solo OLTP por vez, se debería tener bien en claro qué datos se desean y cómo almacenarlos; de esta forma, si los sistemas OLTP migran o cambian, se sabrá de antemano qué parte de los procesos de empresa quedan sin medición o control y, recursivamente, se podrá decidir qué hacer.

Temor a lo indicado por los datos. El temor a un indicador resultante "malo" frena el avance de los Data Marts y Data Warehouses; "revisemos mejor los datos", "elimina este dato que causa ruido", entre otros comentarios, son expresiones que intentan minimizar la respuesta de un reporte extraído directamente de los datos del sistema. Se confía más en recopilar datos manual y artesanalmente que en lo que indique una consulta automática de un Data Warehouse. Básicamente, la calidad regular de los datos obliga a efectuar este tipo de controles. Sería deseable esforzarse en la calidad de los datos de forma que este aspecto de desconfianza se minimice.

3.3.6. Data Warehouse y Data Marts posibles para el caso de un área petrolera tipo

Si se tienen en cuenta las condiciones de borde antes mencionadas, una gerencia de área que, comprendiendo que sus decisiones estarán basadas en conocimiento que se apoya en información derivada de consolidar adecuadamente un gran volumen de datos, facilitará la implementación de un sitio de consulta común; la línea de conducción podrá idealizar el proceso de forma similar a la indicada en la figura 8. Las acciones derivadas de esta idealización conducen a –por lo menos— construir consultas para cada OLTP específico y cruzadas entre los diversos OLTPs.

Si consideramos nuevamente las tareas antes vistas de análisis que cada sector y jefaturas realizan, podemos clasificarlas en:

- Actividades vinculadas a un solo OLTP:
 - análisis del régimen extractivo
 - análisis de falla del equipamiento de fondo y de superficie
- Actividades que requieren cruzar varios OLTPs, que podríamos calificar como actividades que generen potenciales data marts:
 - Para las actividades individuales por sector:
 - informe mensual (con indicadores fijos, definidos corporativamente) por sector y su análisis
 - análisis de los procesos de gestión
 - Para las actividades compartidas tales como:
 - análisis de distribución del agua de inyección para explotación secundaria
 - análisis de la evolución del presupuesto

- o generación del presupuesto anual
- o reuniones intersectoriales análisis de los procesos de gestión

Si se piensa en la totalidad de las áreas petroleras de una compañía en un país, surgen preguntas en un nivel superior las que, por ejemplo, crucen presupuestos, inversiones, nivel de producción, fallas, entre otros, por área, por tecnología extractiva, por responsables de sector. En este contexto, se podría construir un Data Warehouse para responder a todas estas cuestiones.

4. Bases para el descubrimiento del conocimiento en los datos de mantenimiento

Independientemente de la tecnología empleada para poder obtener los datos transformados, para poder expandir el poder de inferencia de los datos, se plantea aprovechar las numerosas investigaciones acerca de minería de datos que actualmente se están realizando ([11] a [51]) y así, aplicar las diferentes metodologías para descubrir conocimiento en los datos volcados a los sistemas.

Estas investigaciones requieren contar con un gran volumen de datos. Si se tiene en cuenta el volumen de transacciones, que para el caso de Puesto Hernández consiste el grueso de la información está formado por

- Ordenes de trabajo: 01/01/2004 al 09/07/2005, 114.871,
- Partes Diarios: 01/01/2004 al 20/08/2005, 146.316 (cuadrillas de producción, servicios y mantenimiento),
- Libro Mayor: 01/01/2004 al 05/08/2005, 276.962 (sólo los movimientos relacionados con la operación),
- Down time: 01/01/2003 al 09/07/2005, 152.434,

se obtienen unos 611.000 [registros/año]; tal volumen de información justifica la automatización de muchas de las consolidaciones para diversos tipos de análisis.

Se educó acerca de cuatro conceptos al personal involucrado con los datos: Dato, volumen de los datos, Información y Conocimiento; figura 8. La forma de mostrar resultados fue a través de las reuniones periódicas donde se exploraban resultados partiendo desde el resultado (conocimiento) hacia el origen (el dato). En general, estas reuniones periódicas sirvieron para crear mediana conciencia en la carga de datos.

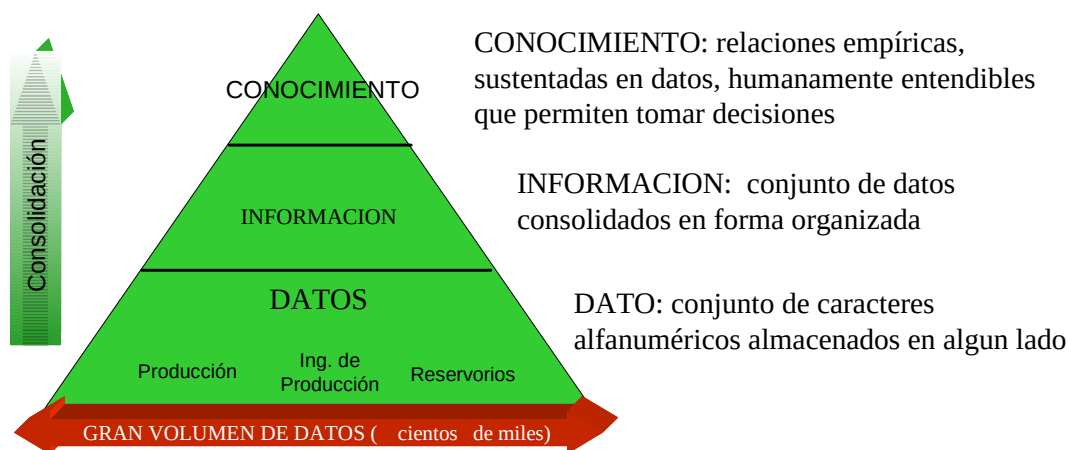


Figura 8: Datos involucrados en los análisis.

Para el control de la calidad de los datos se diseñaron dos programas:

- Control del mantenimiento preventivo (en cuanto a calidad del historial de mantenimiento)
- Generación y control de uso de códigos de tarea ISO14224 (aplicado a partes diarios)

4.1. Consolidación de datos

El esfuerzo del área acerca con respecto a la consistencia de la información se justifica si se desean aplicar las tecnologías de explotación de la información. El detalle de la pirámide esquematizada en la figura 8 se amplía en un mayor número de niveles en la figura 9; el proceso estándar para el descubrimiento de conocimientos en el gran volumen de datos [7].

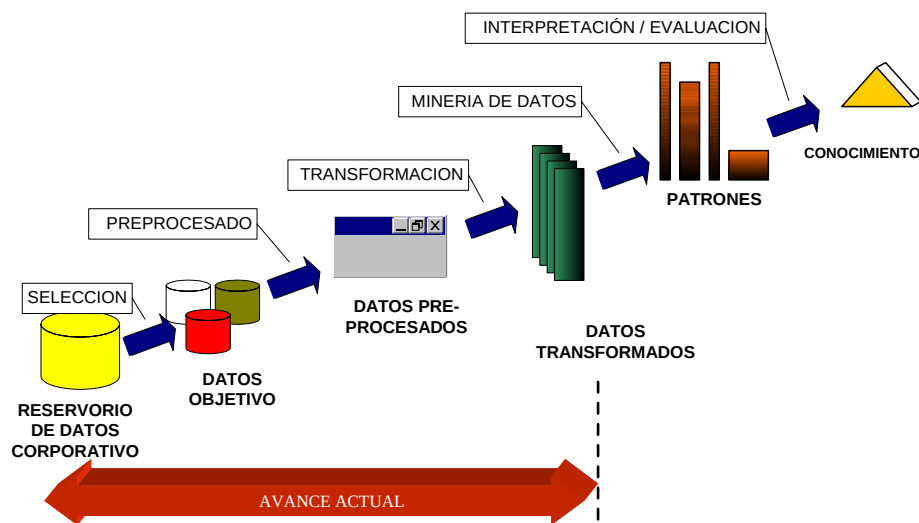


Figura 9: Proceso para la extracción de conocimiento.

El proceso de extracción del conocimiento consta de las siguientes etapas:

- Limpieza de datos (Data cleaning)
- Integración de datos (Data integration)
- Selección de datos (Data selection)
- Transformación de datos (Data transformation)
- Minería de Datos
- Evaluación de patrones (Pattern Evaluation)
- Presentación de los conocimientos (Knowledge presentation)

En la figura 9 se indica lo avanzado en el área Puesto Hernández a la fecha; se ha llegado hasta la etapa “Datos Transformados”.

Mediante las tareas de Minería de Datos se podrán realizar las siguientes actividades:

- Descripción de clases
- Asociación
- Clasificación

- Predicción
- Clustering (agrupamiento)
- Análisis de series a través del tiempo

No obstante lo deseable que estas actividades puedan ser, la base para resultados de análisis exitosos lo constituye la calidad de los datos volcados al sistema y todo el esfuerzo que implica lograr una calidad aceptable. Una vez estabilizado el control de la calidad de los datos, es razonable aplicar técnicas de minería de datos.

5. Conclusiones

La visión del CAT como un ciclo con un flujo de datos asociado, la fuerte dependencia de su calidad respecto de las personas que lo nutren, el planteo de un modelo generalizado de evento de mantenimiento y la integración de diversos OLTPs permiten

- Integración de los datos abstractos de acuerdo las necesidades del trabajo. (evitar que la herramienta o la versión de software nos domine)
- Ser más específicos al requerir modificaciones o herramientas de los especialistas en sistemas
- Definir los perfiles de acceso del personal de acuerdo con el impacto en la información
- Evitar la proliferación de consultas parciales a bases de datos que no favorezcan las vistas integradoras y que limiten la creatividad de análisis de los usuarios
- Tomar conciencia del volumen de información, de la calidad de los datos y de su impacto al tomar decisiones derivadas de éstos
- Tomar conciencia del grado de aplicación del CAT y la actual dependencia de las personas.
- Tomar conciencia de los indicadores y analizar su grado de utilidad y aplicación.

La implementación de sistemas de la magnitud de un Data Warehouse requiere además de una estrategia de implementación, de gerenciamiento. Es vital para el éxito de un programa así que la alta gerencia lo considere estratégico y marque esto como uno de los objetivos corporativos.

La mejor manera de generar el cambio cultural consiste en demostrar agilidad, precisión y confianza ante las diversas preguntas que una gerencia puede realizar. Para ello, se debe hacer un esfuerzo para determinar y responder las "preguntas más afines al negocio" y así poder construir el sistema; éstas no pueden hacerse adecuadamente si se desconocen los diferentes procesos y subprocesos dentro de la empresa.

Difundir el concepto de la figura 8 en la gerencia y mandos medios es un buen comienzo, una idea gráfica para ir sembrando la idea de un Data Warehouse o un Data Mart. Si además de este concepto, la gerencia lo asocia con la creciente demanda de prontitud de respuesta que hoy se le exige y, más aún, cuando se está consciente del volumen de datos implicado, tal combinación acelera la opción por esta nueva tecnología.

6. Trabajos futuros

Como trabajo futuro inmediato, se pretende afianzar el modelo de evento generalizado, intentando relacionar por sistemas el downtime a cada evento de falla.

En segundo lugar, se revisarán los indicadores para la gestión de mantenimiento y se validará la estructura de datos del evento genérico; si ésta sirve también para la generación del tablero de control de mantenimiento.

Se seguirá trabajando sobre la tipificación de los datos e investigando para que el modelo de evento generalizado permita: a) aplicar las distintas técnicas de minería de datos a conjuntos acotados de información, y b) alimentar las diferentes métricas para informes de gestión de mantenimiento.

7. Referencias

7.1. Libros

- [1] Barusso Lafraia J., (2001), *Manual de Confiabilidade, Manutenibilidade e Disponibilidade*, Qualitymark, ISBN: 85-7303-294-4.
- [2] Barlow R., Proschan F., (1996), *Mathematical Theory of Reliability*, Classics in applied mathematics, SIAM, ISBN:0-89871-369-2.
- [3] Date, C. J., (2001), *Introducción a los Sistemas de Bases de Datos*, Pearson Educación, ISBN 968-444-419-2, Cap. 21, 21.6 Procesamiento analítico en línea, 715-722.
- [4] Ebeling C. E., (1997) *An Introduction To Reliability And Maintainability Engineering*, McGraw Hill, ISBN 0-07-018852-1, Cap. 15, 15.3.1 Maximum likelihood estimator, pp. 375-376.
- [5] García Garrido S., (2003) *Organización y Gestión Integral de Mantenimiento*, Ed. Díaz de Santos, ISBN: 84-7978-548-9, Cap. 9.
- [6] Jardine A., (1994), *Maintenance, Replacement and Reliability*, Pitman Publishing, ISBN: 0-273-31654-0.
- [7] Kantardzic M., (2004), *Data Mining: Concepts, Models, Methods, and Algorithms*, John Wiley & Sons, ISBN: 0471228524, 2003.O'Connor, P., *Practical Reliability Engineering*, 4th Ed., John Wiley & Sons, Ltd., ISBN: 0-470-84463-9.
- [8] McClave J., Dietrich F. H., sincich T., *Statistics, 7th. ed*, Prentice Hall Inc., ISBN 0-13-471542-X, Cap. 11 y Cap. 12, pp.473-603.
- [9] Rausand M., Hoyland A., (2004), *System Reliability Theory, Models, Statistical Methods and Applications, 2nd Ed*, Wiley-Interscience, John Wiley & Sons Inc., ISBN: 0-471-47133-X.
- [10] Ruiz-Maya Perez, L. y Pliego J. M., (2001), *Estadística II: Inferencia, 2da ed.*, Ed. AC Madrid, ISBN 84-7288-196-2, Cap. 6, 6.1 Estimación Máximo-verisímil, pp. 178-208.
- [11] Sapag Chain, N., (1995), *Preparación y evaluación de proyectos*, 3ra Ed, McGraw-Hill, ISBN: 958-600-338-8.

7.2. Papers

- [11] Aboulnaga A., Chaudhuri S., *Self-tuning Histograms: Building Histograms Without Looking at Data*, In Proceedings of the ACM SIGMOD Intl. Conf. on Management of Data, 181-192, 1999.
- [12] Agrawal Rakesh, Ghosh Sakti, Imielinski Tomasz, Iyer Bala, Swami Arun, *An Interval Classifier for Database Mining*, Applications IBM Almaden Research Center.
- [13] Agarwal Ramesh C., Charu C. Aggarwal, V.V.V. Prasad, *A Tree Projection Algorithm For Generation of Frequent Itemsets*.
- [14] Agarwal Ramesh C., Aggarwal Charu C., V. V. V. Prasad, *Depth First Generation of Long Patterns*.

- [15] Agrawal R., Imielinsky T., Swami A., *Mining Association Rules between Sets of Items in Large Databases*, Proceedings of the 1993 ACM SIGMOD Conference.
- [16] Agrawal R., Srikant R., *Searching with Numbers*.
- [17] Agrawal R., Gosh S., Imielinski T., Iyer B., Swami A., *An Interval Classifier for Database Mining Applications*, Proceedings of the 18th VLDB Conference Vancouver, Canada, 1992.
- [18] Ali K., Manganaris S., Srikant R., *Partial Classification using Association Rules*.
- [19] Borgelt C., Kruse R., *Induction of Association Rules: Apriori Implementation*, 15th Conference on computational Statistics (Compstat 2002, Berlin, Germany), Physica Verlag, Heidelberg, Germany 2002.
- [20] Codd, E. F., *A Relational Model of Data for Large Shared Data Banks*, Communications ACM, Vol 13, Nr 6, June 1970.
- [21] Codd, E. F., *Extending the Database Relational Model to Capture More Meaning*, ACM Transactions on Database Systems, Vol 4, No 4, December 1979, pp. 397-434.
- [22] Donjerkovic D., Ioannidis Y. E., Ramakrishnan R., *Dynamic Histograms: Capturing Evolving Data Sets*, Proceedings of the 16th International Conference on Data Engineering, San Diego, California, USA, February 2000.
- [23] Graps A., *An Introduction to Wavelets*, IEEE Computational Science and Engineering, Summer 1995, Vol. 2, Num. 2, IEEE Computer Society.
- [24] Guha S., Koudas N., Shim K., *Data-Streams and Histograms*, ACM, 2001.
- [25] Guha S., Koudas N., Srivastava D., *Fast Algorithms For Hierarchical Range Histogram Construction*, ACM PODS 2002, June 3-6, Madison, Wisconsin, USA.
- [26] Han J., Fu Y., *Mining Multiple-Level Association Rules in Large Databases*, IEEE Transactions on Knowledge and Data Engineering, Vol. 11, No. 6, September/October 1999.
- [27] Heikki, D. H. and Smyth P., *Principles of data mining*, The MIT Press, ISBN:026208290X, 2001.
- [28] Hipp Jochen, Götter Ulrich, Nakhaeizadeh Gholamreza, *Algorithms for Association Rule Mining - A General Survey and Comparison*, SIGKDD Explorations, Volume 2, Issue 1 - page 58-64.
- [28] Huenerfauth M., *Decision Tree Problems*, CSE-391: Artificial Intelligence
- [29] Lecture Notes, University of Pennsylvania. Visitar: <http://www.cis.upenn.edu/~cse391/>.
- [30] Jain A. K., Murty M. N., Flynn P. J., *Data Clustering: A Review*, ACM Computing Surveys, Vol. 31, No. 3, September 1999.
- [31] Han, J., Kamber M., *Data Mining: Concepts and Techniques*, Morgan Kaufmann, ISBN 1558604898, 2000.
- [32] Jürgen Paetz and Rüdiger Brause, *A Frequent Patterns Tree Approach for Rule Generation with Categorical Septic Shock Patient*.
- [33] Littman M. L., Yihua W., *Chapter 3: Decision Tree Learning*, CS 536 lecture notes, Rutgers University, Fall 2003. Visitar: <http://www.cs.rutgers.edu/~mlittman/courses/ml03/>.
- [34] Mackenzie D., *Wavelets: Seeing the forest – and the trees*, Beyond Discovery, National Academy of Sciences, 2001.
- [35] *Statistics Toolbox for use with Matlab User's Guide Version 2*, Cap. 1, Cluster Analysis, pp. 50-62.

- [36] Marco Botta, Rosa Meo, Cinzia Malangone, *Association Rules Extraction with MINE RULE Operator*.
- [37] *Microsoft SQL Server OLAP Services 7.0: OLAP Manager Tutorial*.
- [38] Park S., Chen M., Yu P., *An Effective-Based Algorithm for Mining Association Rules*.
- [39] Pei J., *Data preprocessing*, CSE 626 – Data Mining Lecture Notes, University at Buffalo. Visitar: <http://www.cse.buffalo.edu/faculty/jianpei/teaching/datamining/>
- [40] Pei Han J., Yin Y., *Mining Frequent Patterns without Candidate Generation*.
- [41] Pei Jian, *Pattern-Growth Method for Frequent Pattern Mining*, Simon Fraser University, 2002.
- [42] Pei J., Behzad Mortazavi-Asl, Jianyong Wang, Helen Pinto, Qiming Chen, Umeshwar Dayal, and Mei-Chun Hsu, *Mining Sequential Patterns by Pattern-Growth: The PrefixSpan Approach*, IEEE transactions on knowledge and data engineering, vol. 18, No. 10, October, 2004.
- [43] Piatetsky-Shapiro G., Connell C., *Accurate Estimation of the Number of Tuples Satisfying a Condition*, ACM, 1984.
- [44] Possala V., Ioannidis Y., Haas P., Shekita, E., *Improved Histograms for Selectivity Estimation of Range Predicates*, In Proceedings of the ACM SIGMOD Conference, pages 294-305, 1996.
- [45] Shapiro, S. *Propositional, First-Order And High-Order Logics: Basic Definitions, Rules of Inference, and Examples*.
- [46] Shastri A., Govil R., *Optimal Discrete Entropy*, Applied Mathematics E-Notes, (2001), pp. 73-76. Visitar: <http://math2.math.nthu.edu.tw/~amen/>
- [47] Shannon C. E., *A Mathematical Theory of Communication*, Reprinted with corrections from The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948. (Enuncia el concepto de entropía de la información utilizada para determinar agrupadores)
- [48] Srikant R., Vu Q., Agrawal R., *Mining Association Rules with Item Constraints*.
- [49] Srikant R., Agrawal R., *Mining Generalized Association Rules*, Proceedings of the 21st VLDB Conference, Zurich, Switzerland, 1995.
- [50] Arjen van Ooyen, *Theoretical Aspects of Pattern Analysis A Simple Introduction to Principal Component and Cluster Analysis*, In: New Approaches for the Generation and Analysis of Microbial Typing Data (eds. L. Dijkshoorn, K. J. Towner & M. Struelens), Elsevier, Amsterdam, 2001, pp. 31-45. Visitar: <http://www.bio.vu.nl/enf/vanooyen/>.
- [51] Zhao Qiankun, Sourav Nanyang S. Bhowmick, *Association Rule Mining: A Survey*.