

BASE DE DATOS II

Trabajo Práctico 2: Data Warehouse

Ejercicio 1: Red de Clínicas

Enunciado:

Un grupo empresario dedicado a la prestación de servicios de salud administra una red de clínicas privadas distribuidas en distintos puntos de la provincia de La Rioja, Argentina. El grupo necesita desarrollar un almacén de datos decisional (Data Warehouse) para apoyar la toma de decisiones efectiva respecto de los servicios de atención médica brindados por las clínicas que integran su red de servicios.

Cada clínica perteneciente al grupo brinda atención a sus pacientes para distintos tipos de especialidades, tales como clínica médica, pediatría, cardiología, traumatología, ginecología, entre otras. Las consultas son atendidas, respondidas, resueltas o ejecutadas por diferentes profesionales médicos, dependiendo de la especialidad solicitada. A su vez, cada atención a un paciente por parte de un profesional puede tener una duración y un costo diferente.

Para facilitar la organización y el control de las actividades desarrolladas por las clínicas, el grupo ha dividido la provincia en los distintos departamentos que la conforman. De este modo, cada clínica administrada se encuentra ubicada en una localidad determinada, perteneciente a un departamento específico. En consecuencia, el grupo cuenta con varias clínicas en el territorio de la provincia, cada una ubicada en una localidad determinada, dentro de un departamento específico.

Se necesita, entonces, registrar la atención de consultas realizadas por los médicos a los pacientes en cualquiera de las clínicas pertenecientes al grupo, identificando, en cada oportunidad, la clínica, el profesional que realizó la atención, la especialidad, el paciente y la fecha en que se realizó. Además, para cada atención se desea conocer su duración (en minutos) y el importe abonado por el paciente, dado que el costo de cada consulta varía en cada caso según el profesional, el tipo de consulta, la especialidad, la complejidad de la patología, el tiempo demandado, la zona, la tecnología y la asistencia de personal auxiliar requerida.

El grupo empresario necesita contar con información que le permita afinar sus decisiones y analizar el comportamiento de las atenciones médicas realizadas en las distintas clínicas que administra, considerando diferentes dimensiones de análisis, tales como ubicación geográfica, especialidad, profesional, paciente y período de tiempo.

Consigna de resolución:

Proponer tres modelos alternativos para el diseño del Data Warehouse del grupo empresario, correspondientes, respectivamente, a los esquemas “Estrella”, “Copo de Nieve” y “Constelaciones”. Para cada uno de estos modelos, especificar:

- Tabla o Tablas de Hechos.
- Granularidad conveniente.
- Medidas de datos para el análisis.
- Tablas de Dimensiones.
- Jerarquías para las dimensiones que las acepten.
- Una representación gráfica simplificada del modelo propuesto.

El diseño deberá permitir responder, como mínimo, a los siguientes requerimientos de información:

¿Cuál es la cantidad total de atenciones realizadas por departamento, por clínica y por año?

¿Cuál es el importe total abonado y la duración promedio de las atenciones por especialidad, profesional y año?

¿Cuál es la cantidad de atenciones y el importe total abonado por localidad, clínica y mes?

¿Cuál es el importe total abonado por especialidad, departamento y año?

¿Cuál es la cantidad de pacientes atendidos por departamento, especialidad y año?

¿Podrá compararse la cantidad de atenciones y el importe total abonado entre las distintas clínicas que administra el grupo empresario, por departamento y año?

Ejercicio 2: Cadena de Supermercados

Enunciado:

Una empresa que posee una cadena de supermercados con establecimientos comerciales distribuidos en distintos sectores de una ciudad de La Rioja nos contrata para la realización de un trabajo técnico de Análisis de Datos. La empresa comercializa una amplia variedad de productos de consumo masivo y necesita desarrollar un Data Warehouse (DW) que permita analizar la actividad realizada en sus diferentes supermercados y facilite la toma de decisiones comerciales, financieras y de gestión.

A los efectos de organizar la distribución geográfica de sus establecimientos, la empresa divide la ciudad en diferentes zonas, y cada una de ellas comprende uno o más barrios. Por tanto, un supermercado se encuentra ubicado en un barrio determinado y pertenece a una zona específica de la ciudad. En este esquema, cada supermercado se encuentra identificado por un código y posee determinadas características, como nombre, dirección, superficie del local, cantidad de cajas y tipo de establecimiento.

Los supermercados de la cadena atienden a diferentes tipos de clientes. Entre ellos se encuentran clientes particulares, clientes registrados en la cadena (con programa de beneficios), clientes mayoristas y clientes corporativos. A su vez, los clientes registrados pueden pertenecer a diferentes categorías dentro del programa de fidelización, de acuerdo con su nivel de consumo o los beneficios obtenidos.

En otro orden, a través de sus supermercados, la empresa comercializa productos pertenecientes a diferentes rubros y categorías. Por ejemplo, almacén, bebidas, lácteos, carnes, frutas y verduras, limpieza, perfumería, electrónica y artículos para el hogar. Cada producto pertenece a una categoría y cada categoría a un rubro determinado. Los productos poseen información como código, descripción, marca, presentación y unidad de medida.

El precio de venta de un mismo producto de la cadena puede variar en función del período, del supermercado que se trate, de las promociones vigentes en cada uno de ellos y del tipo de cliente que lo adquiere. Asimismo, determinados productos pueden encontrarse alcanzados por descuentos, promociones especiales o beneficios para clientes registrados.

Por cada venta realizada en un supermercado de la cadena se registra los productos adquiridos en ella, así como la cantidad vendida, el precio unitario aplicado y el importe total correspondiente a cada uno de ellos. Además, se registra la fecha y hora de la venta, el supermercado donde se realizó la operación, el cliente -cuando se encuentra identificado- y el medio de pago utilizado.

En correspondencia con lo hasta aquí descripto, la empresa desea analizar especialmente el comportamiento de las ventas según diferentes franjas horarias, tales como madrugada, mañana, mediodía, tarde, tarde-noche y noche. También necesita comparar el comportamiento de las ventas durante distintos días de la semana, meses, trimestres y años.

A partir de la información que pueda almacenarse en el Data Warehouse, la gerencia de la empresa desea poder resolver situaciones críticas/difíciles/ confusas/encontradas, derivadas, entre otros aspectos, de los siguientes datos de análisis:

- Obtener el **importe total y la cantidad de unidades vendidas** por zona, barrio, supermercado y año.
- Obtener el **importe total de ventas** por zona, barrio, supermercado y mes.
- Analizar la **cantidad de productos vendidos e importe total** por rubro, categoría y producto.
- Obtener el **importe total y la cantidad de ventas** por tipo de cliente y año.
- Analizar las ventas según **franja horaria, supermercado y día de la semana**.
- Comparar el **precio promedio de venta** de los productos según supermercado, producto y período.
- Obtener el **importe total de las ventas realizadas durante promociones**, diferenciando las ventas con y sin descuento.

Consigna de resolución:

A partir del enunciado anterior, diseñe un Data Warehouse para la cadena de supermercados y proponga tres modelos alternativos, correspondientes a los siguientes esquemas: Estrella, Copo de Nieve y Constelaciones.

Para cada modelo deberá especificar:

- 1) Identificar: "Tabla/s de Hechos" involucradas, "Granularidad" de cada tabla de hechos identificada, "Medidas" que deberán almacenarse para permitir los análisis solicitados, "Tablas de Dimensiones" y "Jerarquías" existentes dentro de las dimensiones cuando correspondan.
- 2) Una representación gráfica simplificada del modelo con las relaciones de Hechos y dimensiones decididos.

Consideraciones para el diseño

- El diseño de los modelos propuestos debería permitir responder preguntas tales como:
 - ¿Cuánto se vendió por año, zona, barrio y supermercado?
 - ¿Qué rubros y productos generan mayores ingresos en cada supermercado?
 - ¿Qué tipos de clientes compran más y qué productos prefieren?
 - ¿En qué días y franjas horarias se concentra el mayor volumen de ventas?
 - ¿Cómo evolucionan las ventas y los precios de los productos a lo largo del tiempo?
 - ¿Qué supermercados, zonas y barrios presentan el mejor desempeño comercial?
 - ¿Cuál es el comportamiento de las ventas por tipo de cliente, rubro de producto y zona geográfica?
 - ¿Qué barrios y zonas generan mayores niveles de facturación?
 - ¿Cuál es la evolución de las ventas entre distintos años y meses?
- Al definir la granularidad de la tabla de hechos, considere que una venta puede incluir varios productos diferentes. Por lo tanto, deberá determinar si resulta conveniente registrar una fila por cada línea de venta o producto vendido, en lugar de una fila por cada ticket.
- Asimismo, considere que una misma venta puede involucrar:
 - un supermercado determinado;
 - un barrio y una zona determinados;
 - un cliente de un determinado tipo;
 - un producto perteneciente a un rubro y categoría;
 - una fecha y hora específica;
 - una cantidad de unidades;
 - un precio unitario;
 - un descuento o promoción;
 - un medio de pago.
- Para el esquema de Constelaciones, se deberá identificar al menos dos procesos de negocio diferentes que puedan ser analizados mediante tablas de hechos que compartan dimensiones. Por ejemplo, además del proceso de Ventas, podrá considerarse un segundo proceso relacionado con Promociones/Descuentos o Compras/Abastecimiento, justificando las tablas de hechos y las dimensiones compartidas.

Ejercicio 3. Hoteles de La Rioja

Enunciado:

El Gobierno Municipal de La Rioja desea mejorar la planificación y organización de las diferentes temporadas de los próximos años. Para ello, necesita disponer de información que le permita analizar el comportamiento de la actividad hotelera del Departamento Capital, identificar los períodos de mayor y menor demanda y evaluar la evolución de la actividad turística a lo largo del tiempo.

Con este objetivo, se propone desarrollar un almacén de datos decisional (Data Warehouse) que permita integrar y analizar información proveniente de los establecimientos hoteleros registrados en la ciudad.

La ciudad cuenta con diferentes establecimientos destinados al alojamiento de visitantes, tales como hoteles, apart-hoteles, hosterías y otros tipos de establecimientos. Cada uno posee determinadas características, como nombre, dirección, categoría, cantidad de habitaciones y capacidad de alojamiento.

Los establecimientos se encuentran organizados de acuerdo con su tipo de establecimiento y categoría, pudiendo existir varios hoteles pertenecientes a una misma categoría.

Los establecimientos reciben personas provenientes de diferentes lugares del país y del exterior, que visitan la ciudad por diferentes motivos: trabajo, vacaciones, estudio, etc. Para cada huésped se registra información sobre su país de procedencia y, cuando se trata de argentinos, la provincia de procedencia. Asimismo, los huéspedes pueden clasificarse según diferentes tipos de visitante, tales como turista individual, grupo turístico, turista corporativo, entre otros.

Cada establecimiento dispone de diferentes tipos de habitaciones, tales como individuales, dobles, triples, familiares y suites. Las habitaciones poseen una determinada capacidad y pueden ofrecerse bajo diferentes regímenes de alojamiento, como solo alojamiento, alojamiento con desayuno, media pensión o pensión completa.

Los visitantes pueden realizar reservas de alojamiento con anterioridad a su llegada. En cada reserva se registra el establecimiento, el turista o grupo de turistas, la fecha en que se realizó la reserva, la fecha prevista de ingreso, la fecha prevista de egreso, la cantidad de personas, la cantidad de habitaciones solicitadas, el tipo de habitación y el régimen de alojamiento elegido. También se registra el canal utilizado para realizar la reserva, que puede ser, entre otros, una reserva directa, una agencia de viajes, una plataforma de reservas en línea o el sitio web del establecimiento.

No todas las reservas se concretan. Algunas pueden ser canceladas o modificadas antes de la fecha prevista de ingreso. Por este motivo, se desea conservar información sobre el estado de cada reserva y poder diferenciar entre reservas realizadas, canceladas y efectivamente concretadas.

Cuando una reserva se concreta, se produce la correspondiente estadía del turista. Para cada estadía se registra la fecha efectiva de ingreso y egreso, la cantidad de huéspedes alojados, la cantidad de habitaciones ocupadas, el tipo de habitación utilizada y el régimen de alojamiento contratado.

El importe correspondiente a una estadía puede variar de acuerdo con el establecimiento, la categoría del alojamiento, el tipo de habitación, el régimen contratado, la cantidad de huéspedes, la duración de la estadía, la temporada turística, las promociones aplicadas y otros servicios contratados durante el alojamiento.

La demanda hotelera presenta variaciones a lo largo del año. Por tal motivo, el Gobierno desea identificar y analizar diferentes temporadas, considerando períodos de alta, media y baja demanda, así como períodos asociados a vacaciones, feriados, festividades, eventos culturales, deportivos o de otra naturaleza que puedan producir variaciones en la afluencia de visitantes y turistas.

Para cada período se desea analizar, entre otros aspectos, la cantidad de reservas, las reservas canceladas, la cantidad de huéspedes, las habitaciones ocupadas, las noches de alojamiento, el nivel de ocupación y los importes generados por la actividad hotelera.

Asimismo, se desea conocer el impacto que tienen las promociones y descuentos sobre la demanda. Una reserva o estadía puede estar asociada a una determinada promoción, por ejemplo, descuentos por temporada, promociones para determinados períodos, beneficios para grupos o promociones destinadas a turistas de determinadas procedencias.

El Gobierno necesita utilizar esta información para identificar patrones de comportamiento de la actividad hotelera y contar con elementos que permitan planificar las futuras temporadas turísticas, definir períodos de promoción, anticipar posibles incrementos de la demanda y orientar acciones destinadas a fortalecer la actividad turística de la ciudad.

Consigna de resolución:

Diseñar un modelo multidimensional para un Data Warehouse de la actividad hotelera de la ciudad capital de La Rioja, proponiendo tres modelos alternativos correspondientes, respectivamente, a los esquemas “Estrella”, “Copo de Nieve” y “Constelaciones”.

Para cada uno de los tres modelos se deberá especificar:

- Tabla o Tablas de Hechos involucradas.
- Granularidad conveniente de cada tabla de hechos.
- Medidas de datos necesarias para realizar los análisis solicitados.
- Tablas de Dimensiones.
- Jerarquías correspondientes a las dimensiones que las acepten.
- Relaciones entre las tablas de hechos y las dimensiones.
- Una representación gráfica simplificada del modelo propuesto.

El diseño deberá permitir responder, como mínimo, a los siguientes requerimientos de información:

- Obtener la cantidad de reservas realizadas, canceladas y concretadas por tipo de establecimiento, categoría y año.
- Obtener la cantidad de huéspedes y noches de alojamiento por hotel, temporada turística y año.
- Obtener el porcentaje de ocupación promedio y la cantidad de habitaciones ocupadas por hotel, temporada turística y año.
- Obtener el importe total y el importe promedio por estadía según categoría de establecimiento, temporada turística y año.
- Obtener la cantidad de turistas alojados según país de procedencia, hotel y año.
- Analizar la cantidad de noches de alojamiento según provincia de procedencia, categoría de hotel y temporada turística.
- Obtener la cantidad de reservas y el importe total generado según tipo de habitación, régimen de alojamiento y temporada turística.
- Analizar la cantidad de reservas según canal de reserva, tipo de establecimiento y temporada turística.
- Obtener la duración promedio de las estadías según hotel, temporada turística y procedencia de los turistas.
- Analizar el importe total de descuentos y promociones aplicados según hotel, temporada turística y tipo de promoción.
- Comparar la cantidad de reservas, noches de alojamiento e importe total generado entre los distintos hoteles y categorías durante las diferentes temporadas turísticas.
- Analizar la evolución de la actividad hotelera entre distintos años, identificando períodos de crecimiento o disminución de la demanda.
- Identificar las temporadas de mayor y menor demanda turística, considerando reservas, huéspedes, noches de alojamiento y nivel de ocupación.
- Identificar los hoteles y categorías de establecimientos que presentan mayores niveles de ocupación y facturación durante las diferentes temporadas.