

Taller #2

# Business Intelligence

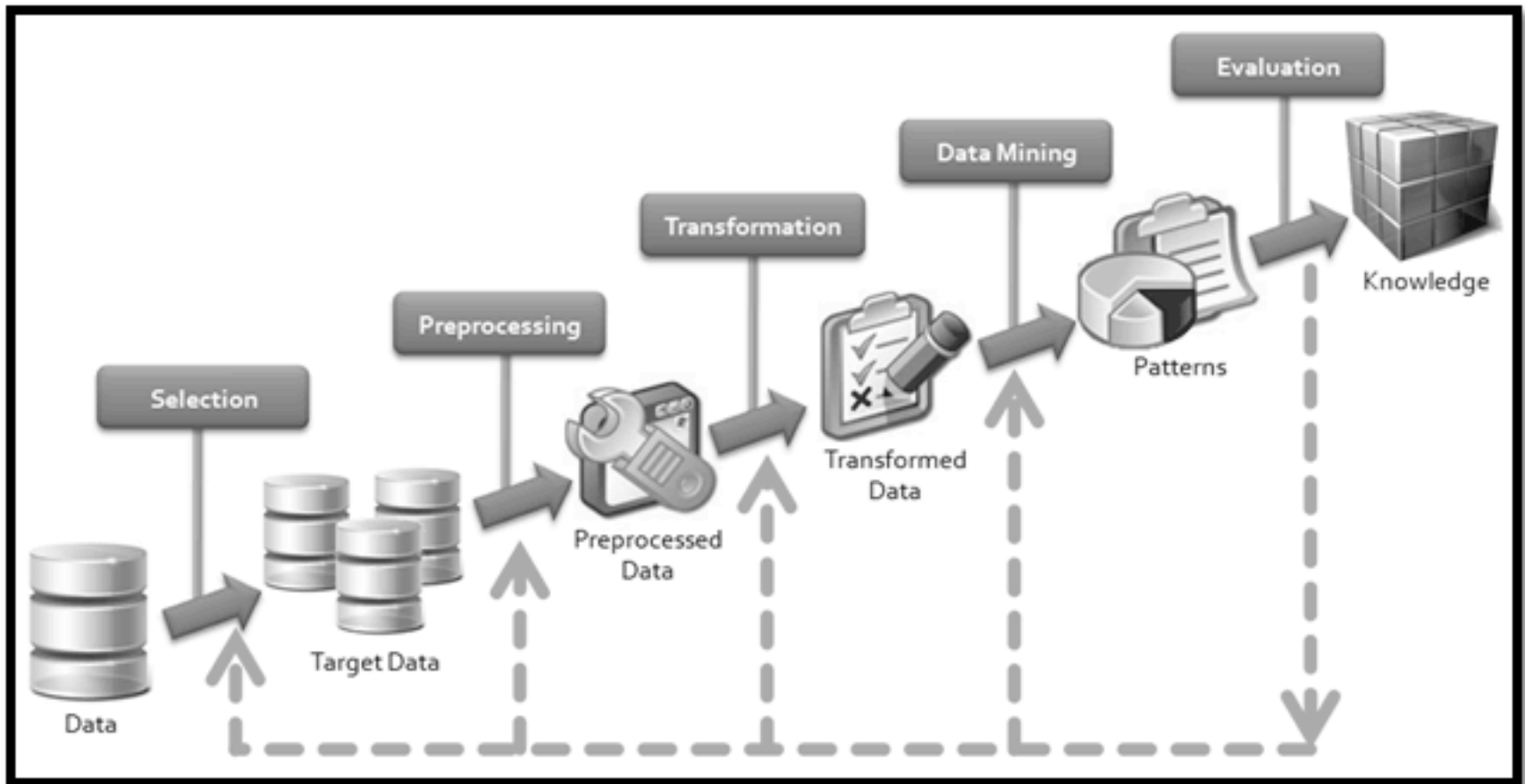
Carlos Reveco  
creveco@dcc.uchile.cl

Cinthya Vergara  
cvergarasilv@ing.uchile.cl

# Agenda

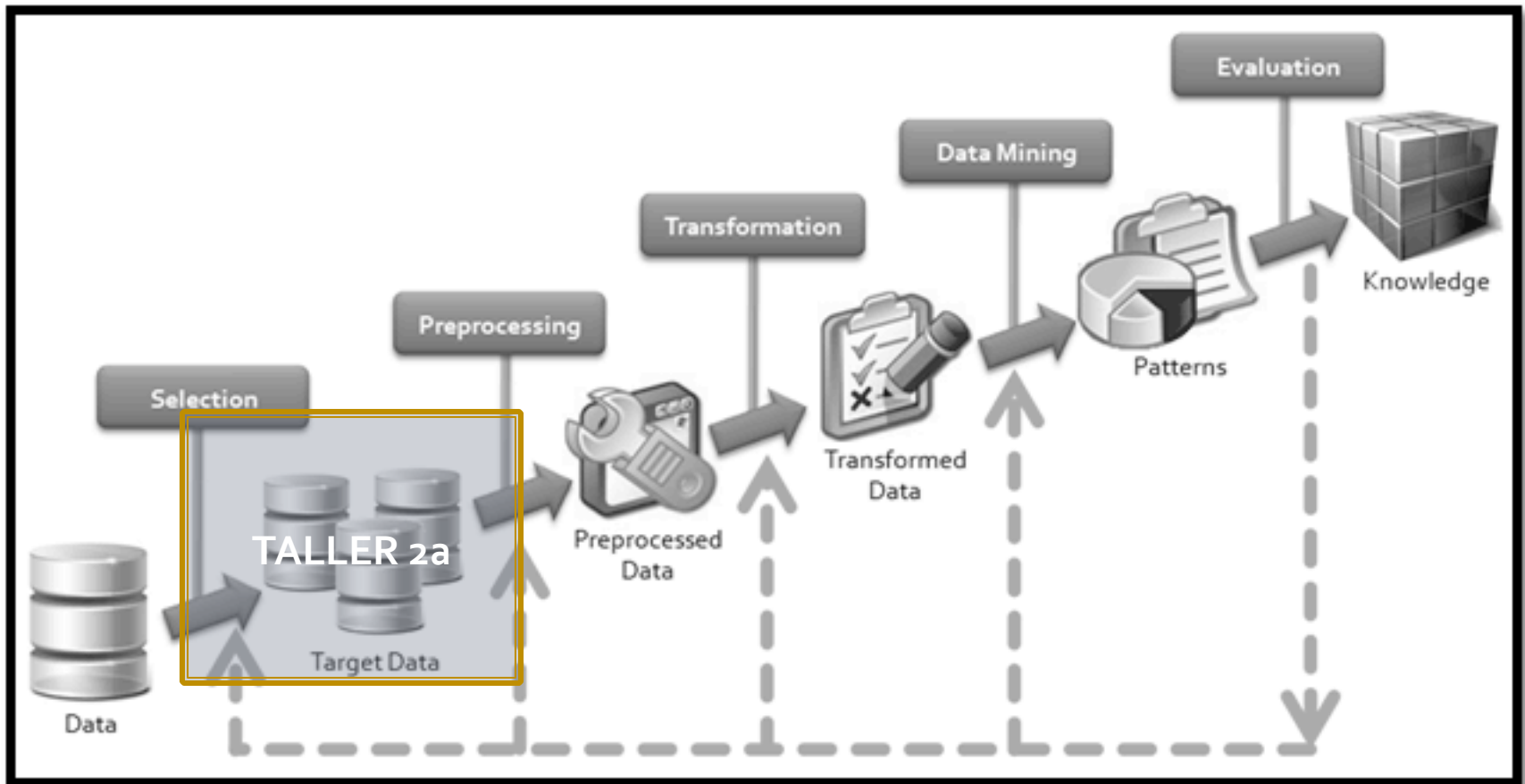
- **Taller#2**
  - **Modelamiento Multidimensional**
    - Antecedentes Generales
    - Ejercicio
  - **Laboratorio #2:**
    - **Desde Modelo Estrella a Consultas OLAP**
      - Herramienta JasperServer
      - Actividad práctica

# Proceso KDD



Knowledge Discovery in Databases → KDD

# Proceso KDD



Knowledge Discovery in Databases → KDD

# Taller #2

## Modelamiento Multidimensional

---

# Definición

- Un Data Warehouse es una **arquitectura**:  
“Colección de datos utilizada para la **toma de decisiones** la cual es **orientada al negocio**, **integrada**, que **varía en el tiempo y no volátil.**”
- **Orientada al negocio**: Definida para temas importantes del negocio
- **Integrada**: Usualmente construida integrando múltiples fuentes de datos heterogéneas
- **Varía en el tiempo**: Datos almacenados proveen información desde una perspectiva histórica
- **No volátil**: No ocurren muchos cambios a los datos disponibles en el data warehouse.

# Data Warehouse vs BBDD. Operacionales [1]

## B.D. operacionales

- Reflejan el “**ahora**”
- Constantemente actualizados
- **Mínima** redundancia
- Soportan **funciones operacionales** del día a día.

## Data Warehouse

- Reflejan la “**historia**”
- Rara vez **modificados o eliminados**
- **Alta** redundancia
- Soportan **requerimientos de información**

# Data Warehouse vs BBDD. Operacionales [2]

## B.D. operacionales

- **Transaccionales** (OLTP)
- Muchos usuarios concurrentes
- Principales prioridades:  
Alta disponibilidad y  
Rápido

## Data Warehouse

- **Analíticos** (OLAP)
- Pocos usuarios concurrentes
- Principales prioridades:  
Altamente flexible y  
orientado al usuario final.



# Implementación

## ■ Bases de datos Multidimensionales

- Creadas para almacenar información
- No difundidas
- Alto costo
- **EJEMPLO:** Cognos Power Cube

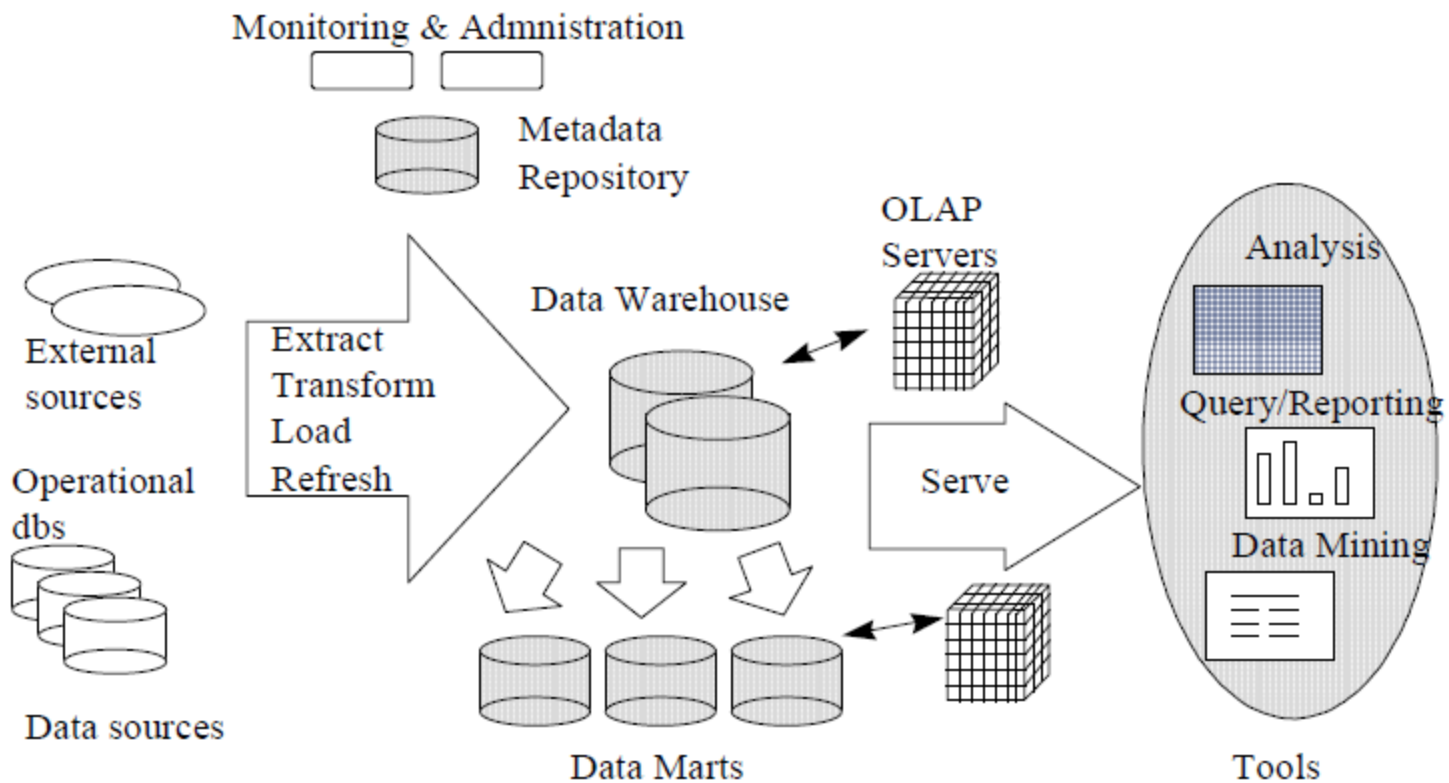
The logo for Cognos, featuring the word "COGNOS" in a bold, red, sans-serif font.

## ■ Bases de datos Relacionales

- Ya presentes en los sistemas operacionales
- Gracias al modelo estrella permiten almacenar información
  - Difundidas en la construcción de Data Warehouses
- **EJEMPLO:** Oracle, MySQL, PostgreSQL, SQLServer

The logo for Oracle, featuring the word "ORACLE" in a red, sans-serif font.The logo for MySQL, featuring the word "MySQL" in a blue, sans-serif font with a blue fish icon above the "Y".

# Arquitectura Data Warehousing



# Metodologías de Diseño

- **Top Down approach** (*Bill Inmon*)
  - Diseñar completamente el Data Warehouse y luego ir desplegando los Data Marts.
  - Esquema rígido y de largo desarrollo, pero fácil de mantener y muy integrado.
- **Bottom Up approach** (*Ralph Kimball*)
  - Desde la perspectiva del usuario: “me interesa sólo mi área de trabajo”
  - Desarrollo de proyectos departamentales (Data Marts) para una posterior construcción del Data Warehouse.
  - Alto ROI, rápido de construir, pero difícil de mantener.

# Data Warehouse / Data Mart

## Data Warehouse

- A nivel **corporativo**
- Modelo más **complejo** (modelo constelación)
- Se alimenta de **muchas fuentes** operacionales
- Desarrollo complejo al **largo plazo**
- Resultados en **el largo plazo**

## Data Mart

- A nivel **departamental**
- Puede ser suficiente un **simple** modelo estrella.
- Se alimenta de **pocas fuentes** operaciones.
- Desarrollo simple y al **corto plazo**
- Resultados en el **corto plazo**

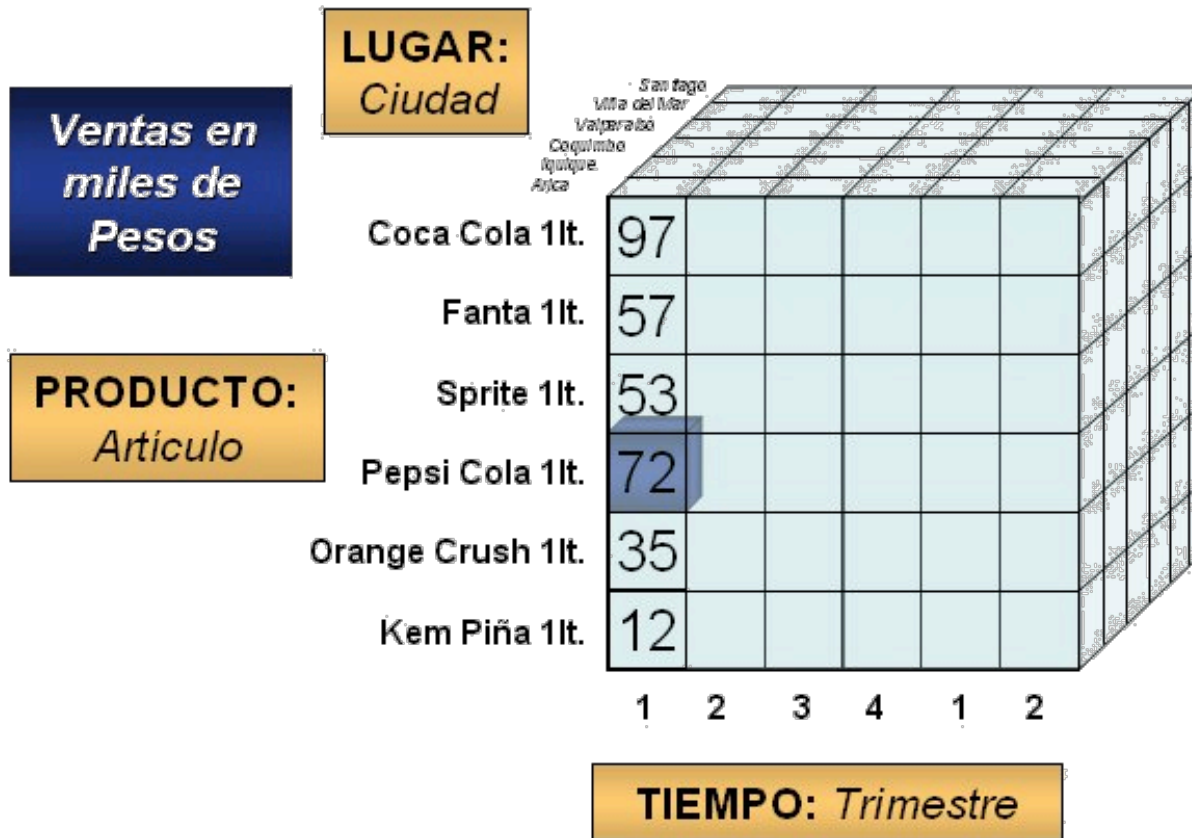
# Diseño del Data Warehouse [1]

## Conceptos fundamentales

- **Tablas de Dimensiones:** Tablas con los descriptores textuales del negocio sobre los cuales se quiere agregar o desagregar la información.
- **Tablas de Hechos (fact table):** Tabla central con las medidas e indicadores proyectados sobre las dimensiones.
- **Granularidad:** Nivel mínimo sobre el cual está diseñada una determinada dimensión.

# Diseño del Data Warehouse [2]

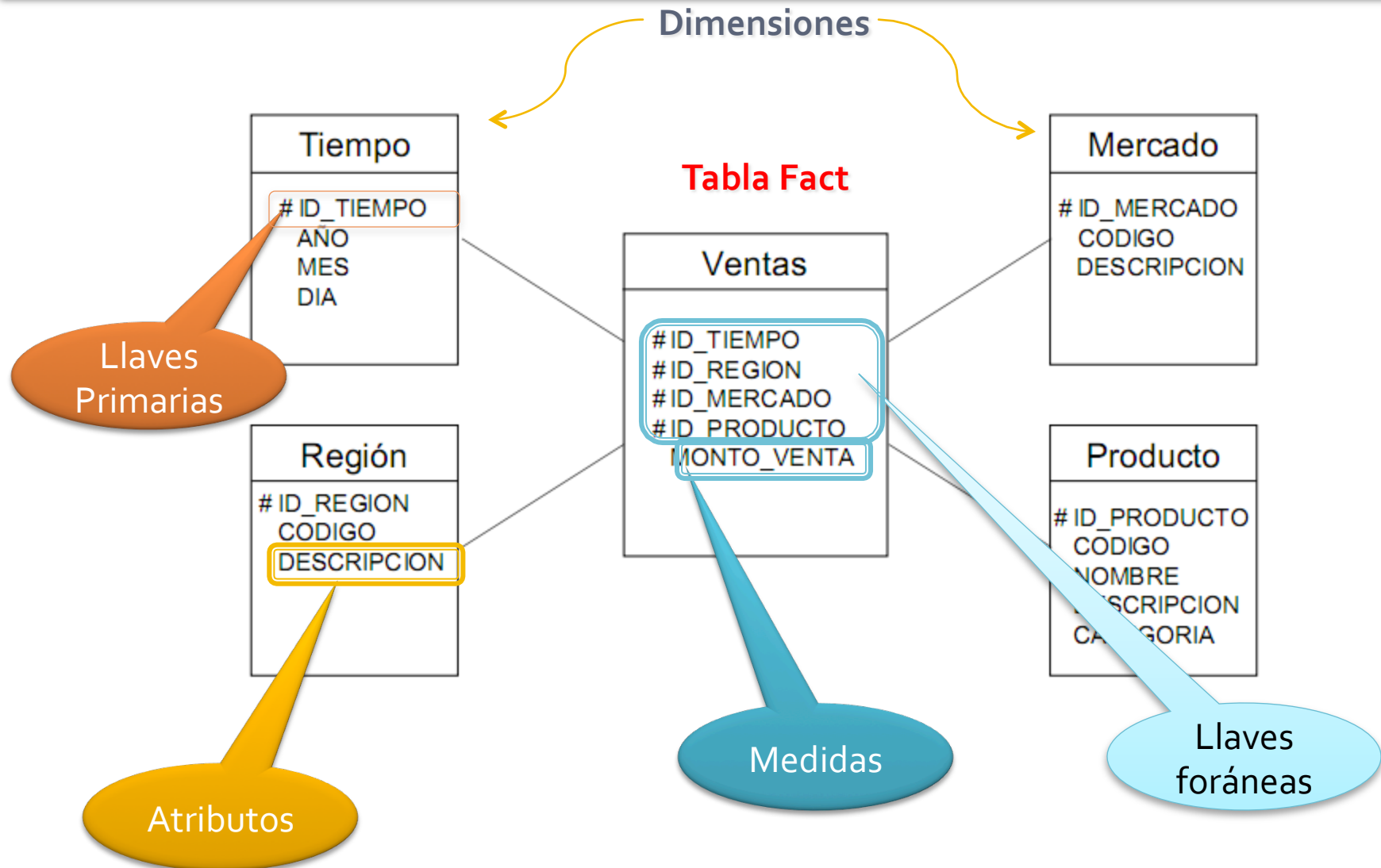
## Dimensiones y Granularidad



# Modelo Estrella

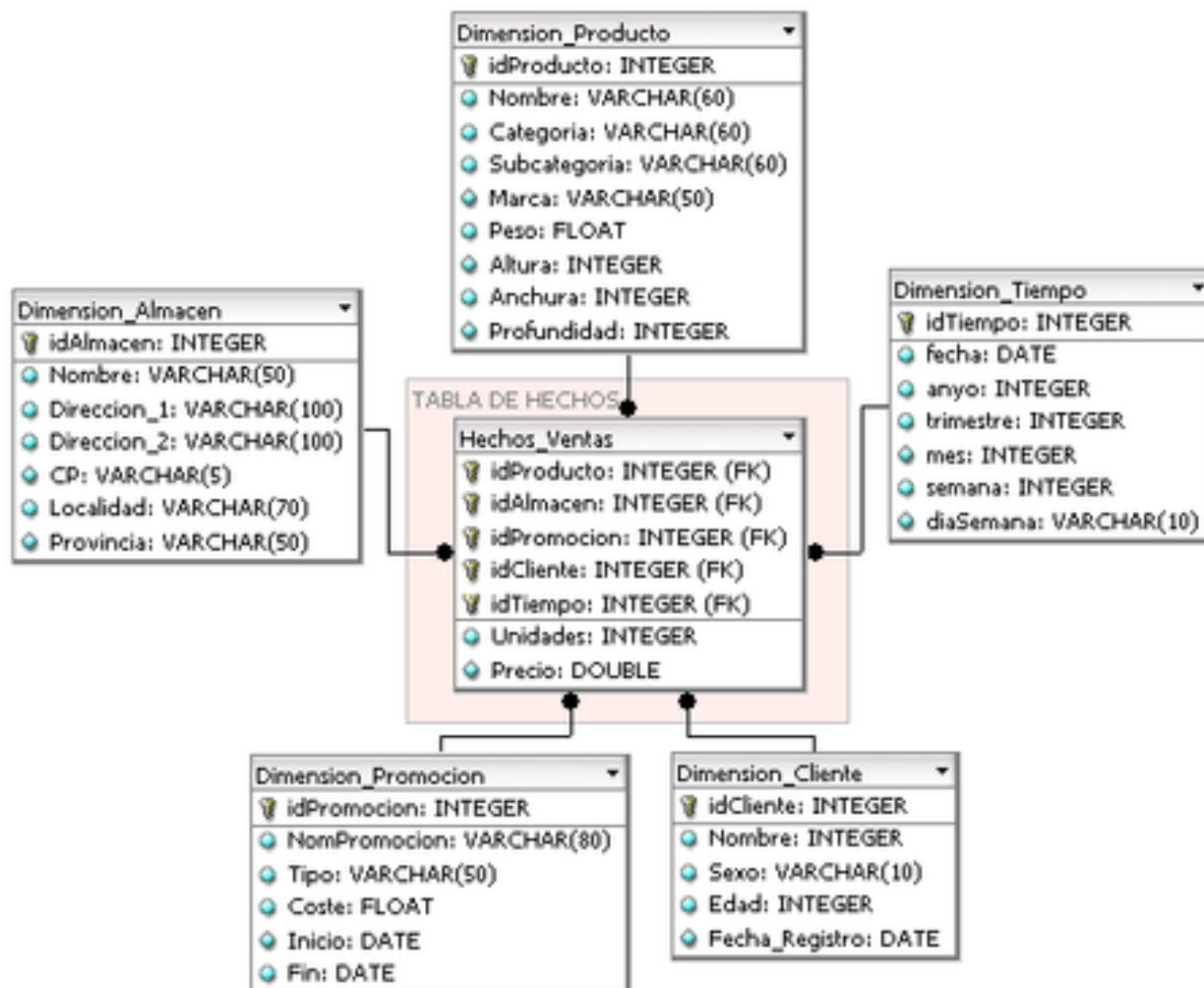
- **Tabla Fact** o de hechos constituida por:
  - Medidas: datos, indicadores etc. los cuales están caracterizados por:
- **Dimensiones** formadas por:
  - **Atributos** que caracterizan y contextualizan las medidas
- **Por lo tanto:**
  - **Estructura que lleva los datos (medidas) a Información**

# Modelo estrella



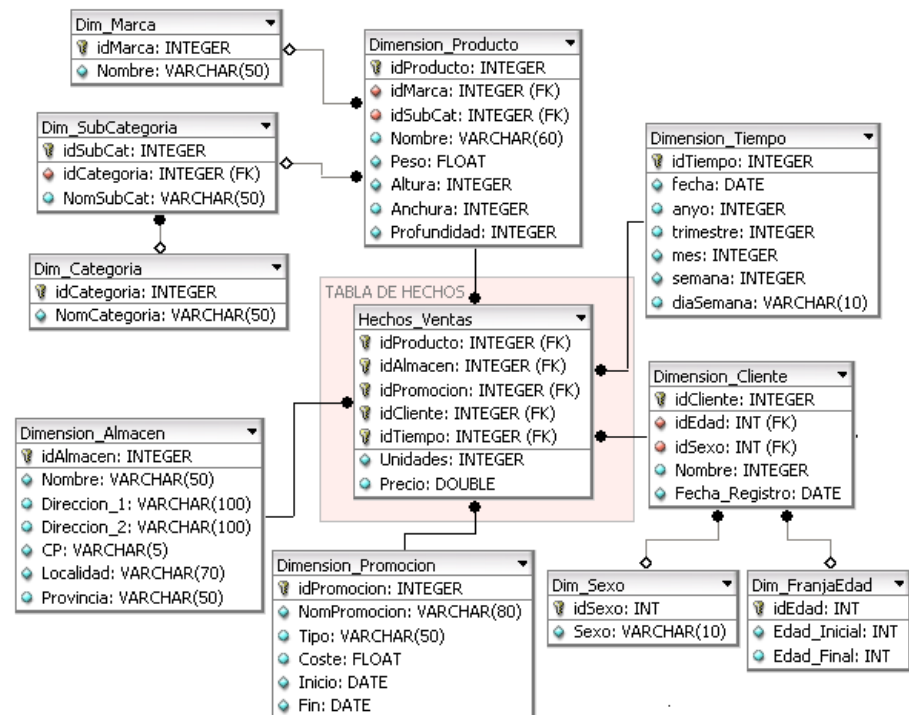
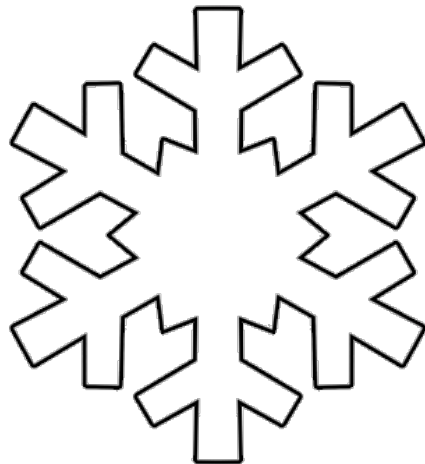


# Ejemplo Modelo Estrella



# Modelo Copo de Nieve

- Las dimensiones se implementa con más de una tabla para normalizar y así reducir el espacio de almacenamiento al eliminar la redundancia.

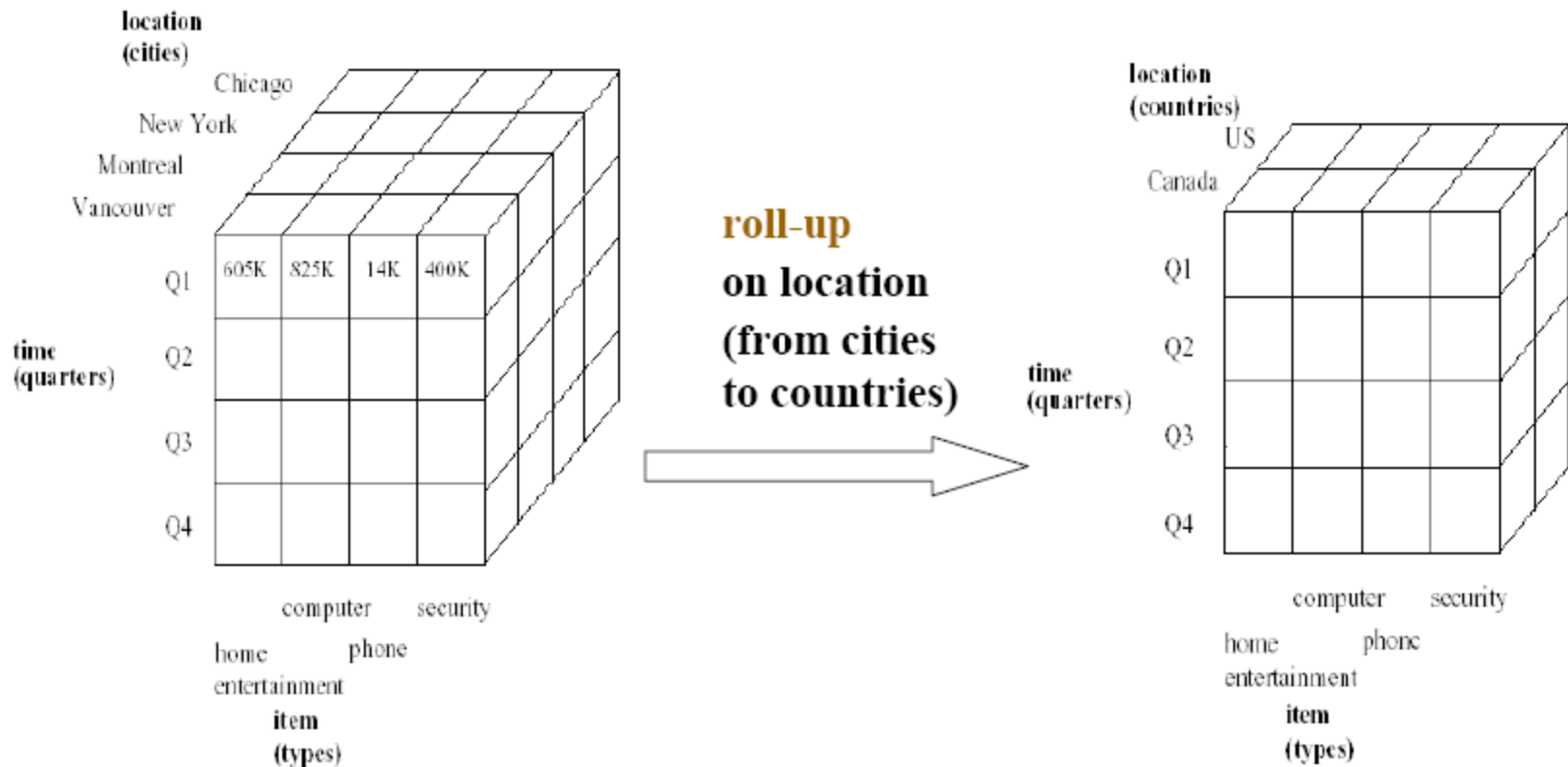


# Operaciones sobre un cubo OLAP

- Un usuario final debe poder realizar fácilmente las siguientes operaciones:
  - **Slice and Dice:** “Cortes” o filtros de la información en para poder examinar mejor.
  - **Roll up:** Agregar información sobre una determinada dimensión.
  - **Drill Down:** desagregar información sobre una determinada dimensión.
  - **Pivot:** Rotar el cubo sobre alguno de las dimensiones de interés.

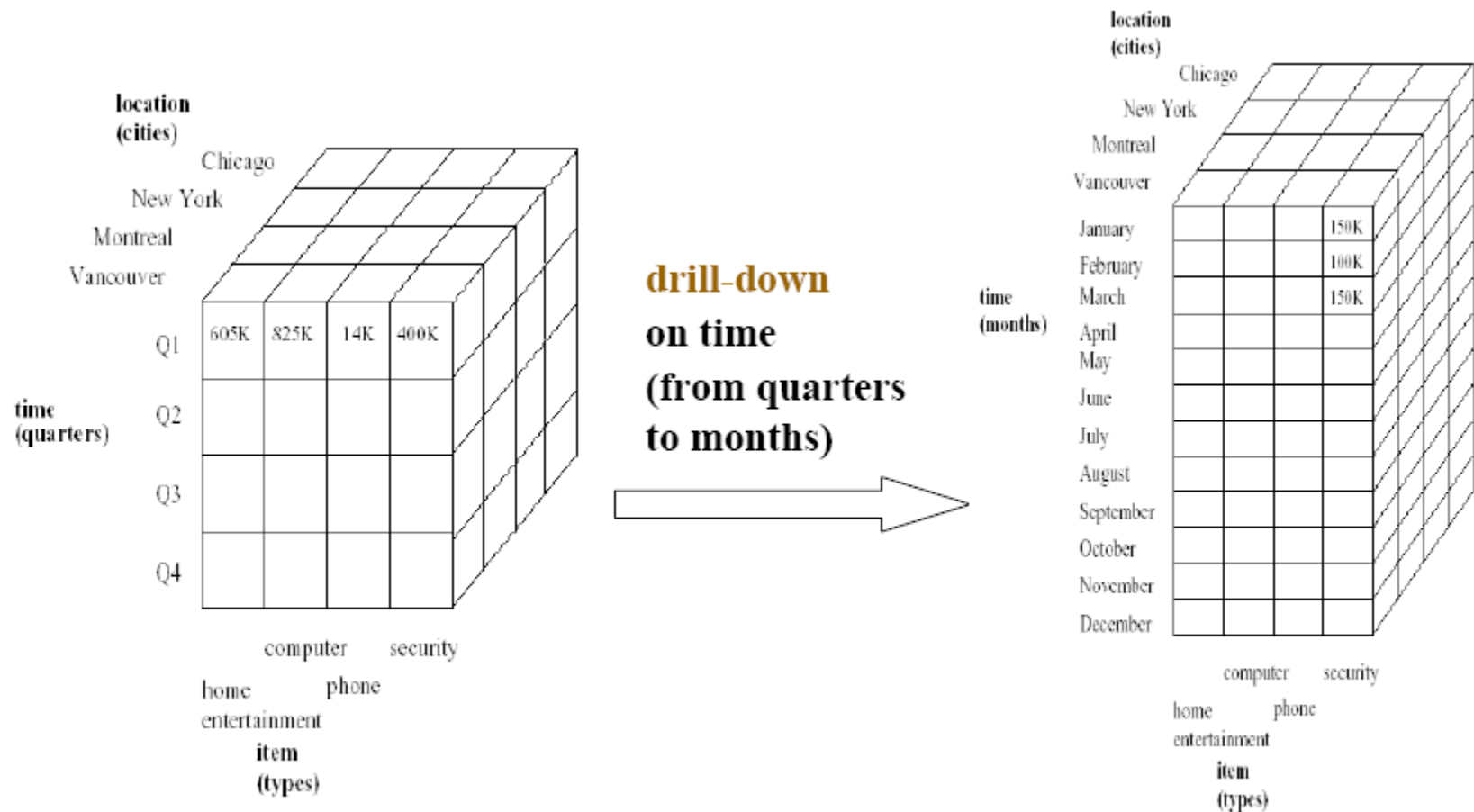
# Operaciones OLAP

## Roll-up



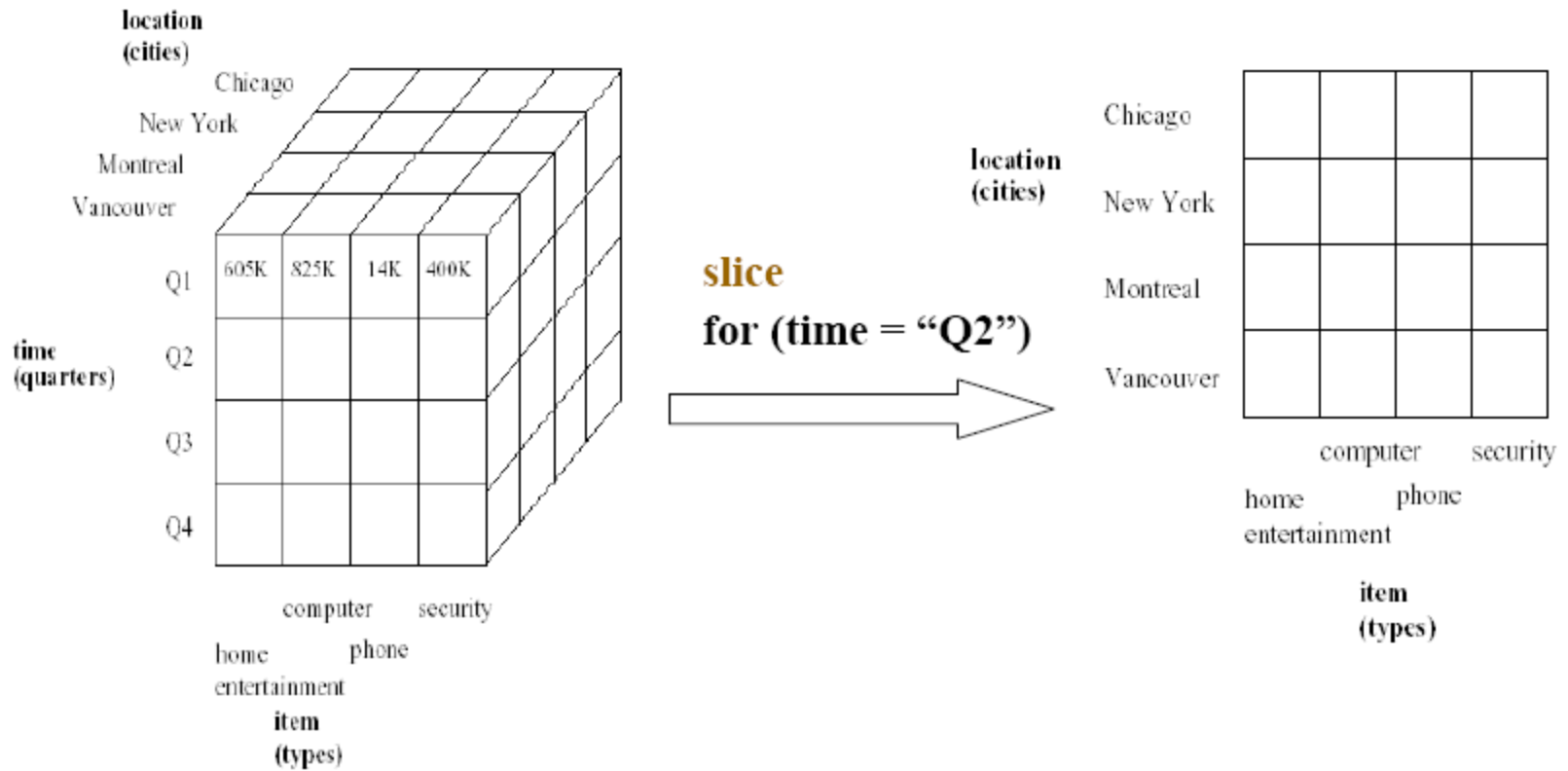
# Operaciones OLAP

## Drill Down



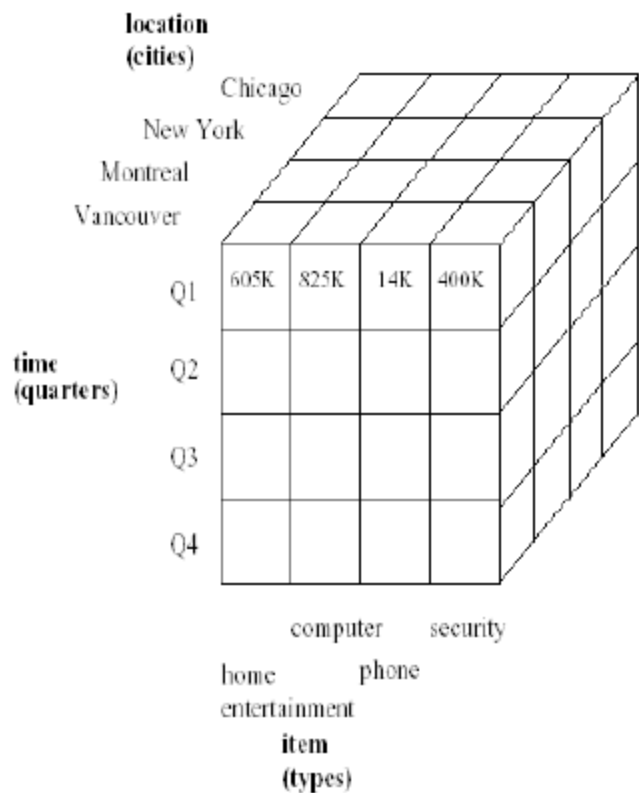
# Operaciones OLAP

## Slice



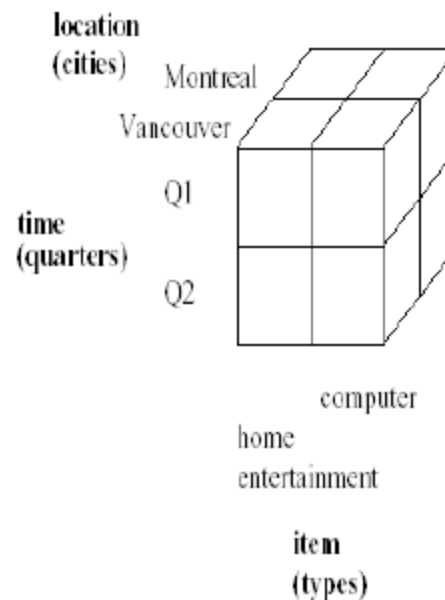
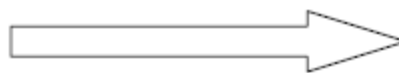
# Operaciones OLAP

## Dice



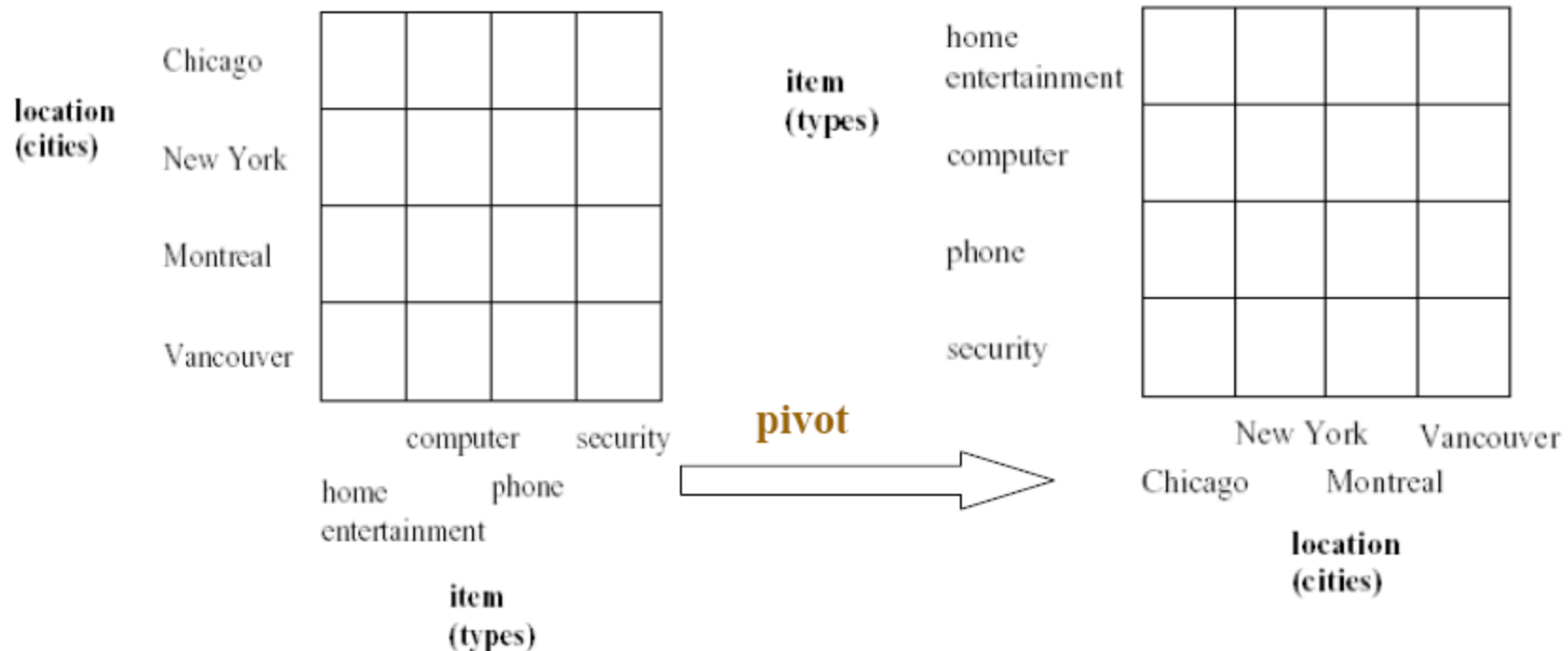
**dice for**

(location = "Montreal" or "Vancouver") and  
(time="Q1" or "Q2") and  
(item="home entertainment" or "computer")



# Operaciones OLAP

## Pivot





# “Kimball’s Top Ten Pitfalls”

## Errores comunes a evitar [1]

- **Pitfall 10:** Enamorarse de la tecnología y los datos, en vez de concentrarse en los requerimientos del negocio y objetivos.
- **Pitfall 9:** Fallar en adoptar una administración visionaria, con influencia, accesible y razonable cómo sponsor del DW.
- **Pitfall 8:** Atacar un proyecto galáctico de muchos años, en vez de perseguir un proyecto más manejable.
- **Pitfall 7:** Utilizar mucha energía en construir una estructura normalizada, quedándose sin presupuesto antes de construir un prototipo viable y operativo.

# “Kimball’s Top Ten Pitfalls”

## Errores comunes a evitar [2]

- **Pitfall 6:** Poner **más atención al rendimiento operacional y mejor desarrollo** que a una **interfaz rápida y fácil de utilizar**.
- **Pitfall 5: “Keep it simple”:** no desarrollar modelos muy complejos que usuarios finales no podrán utilizar, o que necesitarán una fuerte capacitación.
- **Pitfall 4:** Alimentar **modelos dimensionales con una visión stand-alone**, sin **considerar una arquitectura tecnológica adecuada**.
- **Pitfall 3:** **Cargar solamente datos agregados**.

# “Kimball’s Top Ten Pitfalls”

## Errores comunes a evitar [3]

- **Pitfall 2:** Presumir que el negocio, sus requerimientos, el análisis, los datos, la tecnología que lo soporta y la gente es **estática y no cambia en el tiempo.**
  - **Pitfall 1:** **Olvidar** que el éxito del Data Warehouse está directamente relacionado a la aceptación del usuario.
- Si un usuario no lo acepta cómo un sistema para la toma de decisiones, el proyecto fracasó.

# Ejercicio:

- Usted desea armar un modelo multidimensional para un Data Mart del área de ventas de la empresa.
- En el área de sistemas, le dicen que “todo lo datos” están disponibles en
  - host:**
  - user:**
  - pass:**
- Lamentablemente, le informan que no está disponible el modelo Entidad-Relación de la base de datos de ventas.

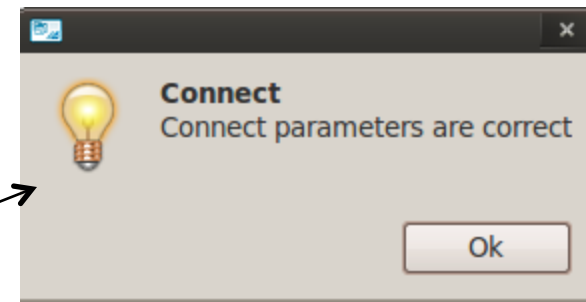
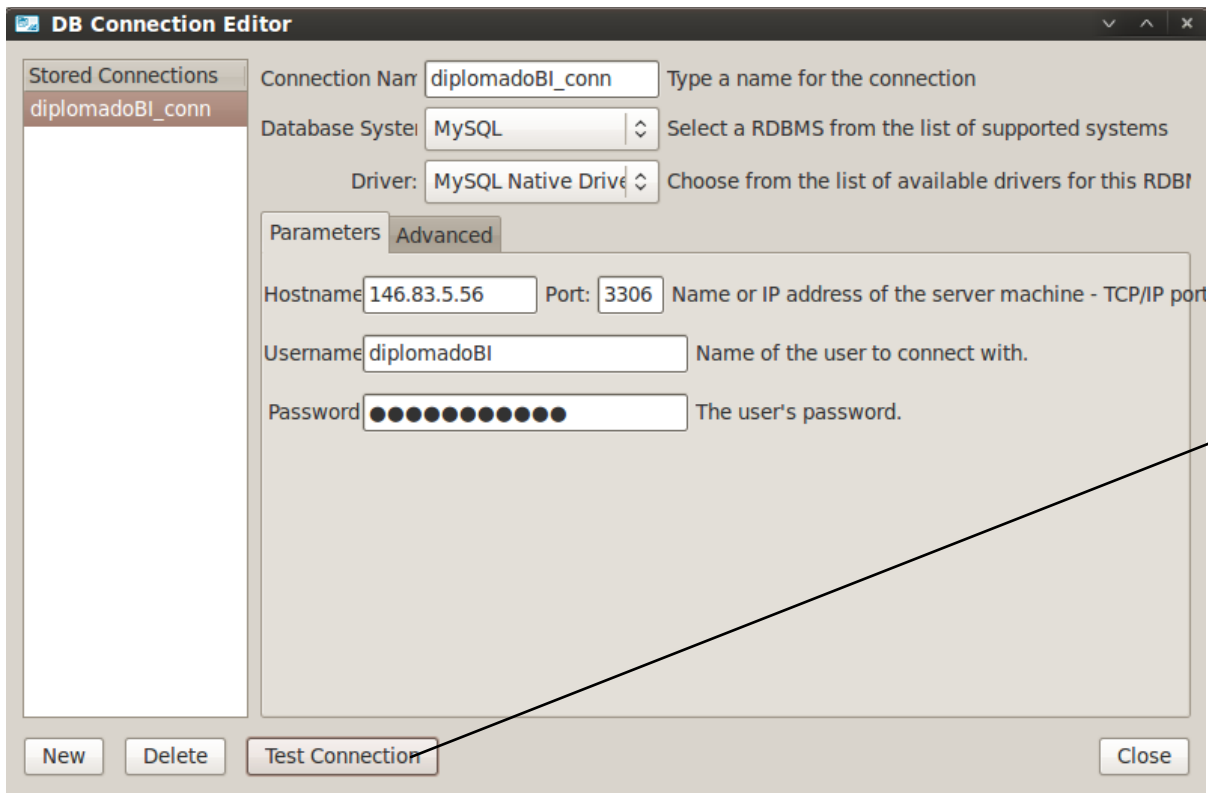
# Ejercicio:

- Usando **MySQL Workbench**:
  - Haga **ingeniería reversa** de la base de datos de ventas
  - Diseñe un **modelo estrella** orientado a la toma de decisiones respecto a las ventas.
  - **Identifique**
    - Medidas de la tabla de hechos y granularidad
    - Dimensiones, atributos

# Solución

## Ingeniería Reversa

- Creamos una conexión a repositorio remoto:



# Solución

## Ingeniería Reversa

**Reverse Engineer Database**

- Connection Options
- Connect to DBMS
- Select Schemata
- Fetch Object Info
- Select Objects
- Reverse Engineer
- Results

**Set parameters for connecting to a DBMS**

Stored Connection:  Select from saved connection settings

Database System:  Select a RDBMS from the list of supported systems

Driver:  Choose from the list of available drivers for this RDBMS

Parameters **Advanced**

Hostname:  Port:  Name or IP address of the server machine - TCP/IP port

Username:  Name of the user to connect with.

Password:  The user's password.

# Solución

## Ingeniería Reversa

**Reverse Engineer Database**

- Connection Options
- **Connect to DBMS**
- Select Schemata
- Fetch Object Info
- Select Objects
- Reverse Engineer
- Results

**Connect to DBMS and Fetch Information**

The following tasks will now be executed. Please monitor the execution.  
Press Show Logs to see the execution logs.

- Connect to DBMS
- Retrieve Schema List from Database

Execution Completed Successfully  
Fetch finished.

Show Logs      Cancel      Back      Next



# Solución


## Ingeniería Reversa

3

**Reverse Engineer Database**

- Connection Options
- Connect to DBMS
- **Select Schemata**
- Fetch Object Info
- Select Objects
- Reverse Engineer
- Results

**Select Schemata to Reverse Engineer**

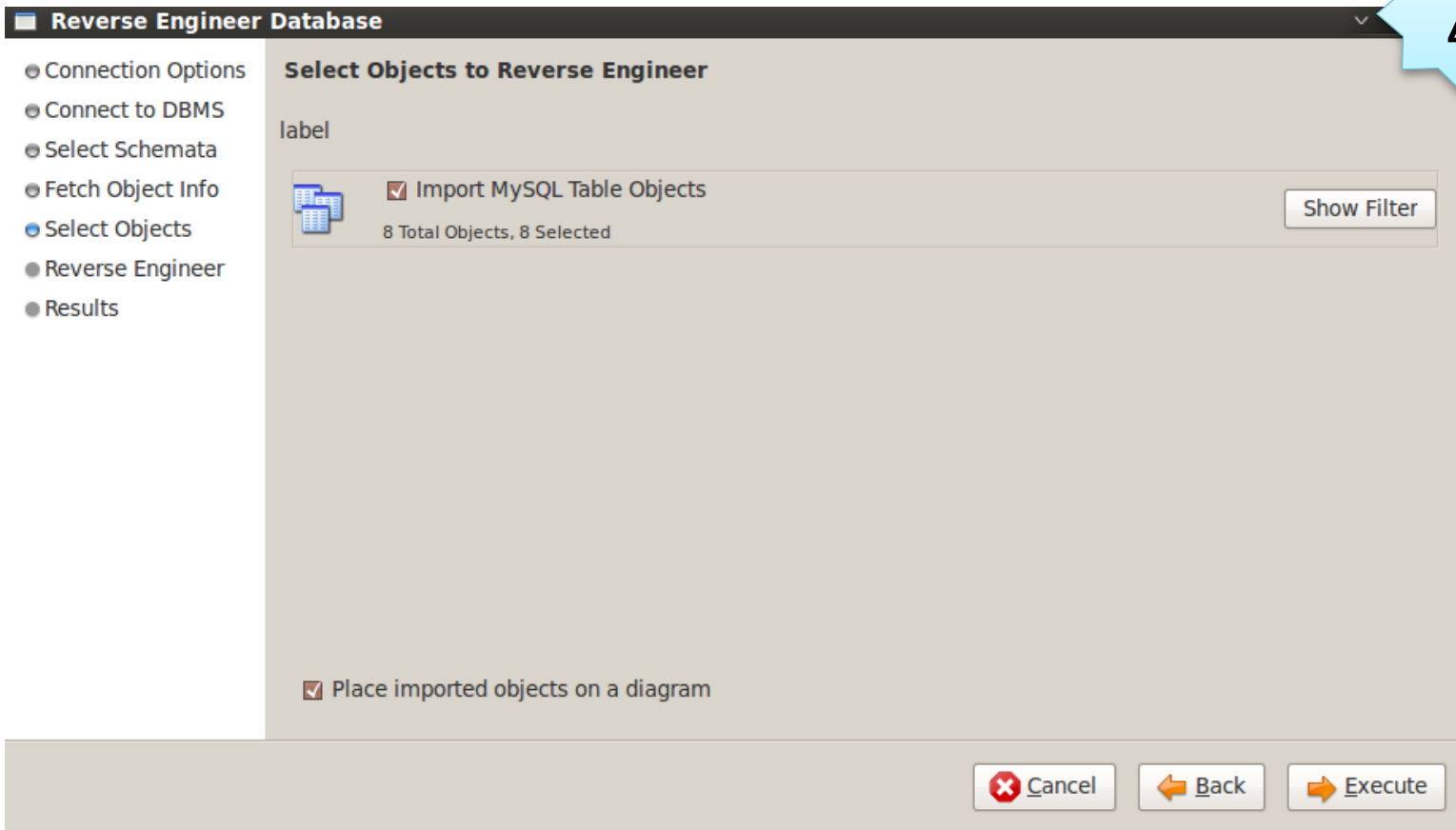
 **Select the Schemata to be Processed:**

- diplomado\_sales
- information\_schema
- taller1

Cancel Back Next

# Solución

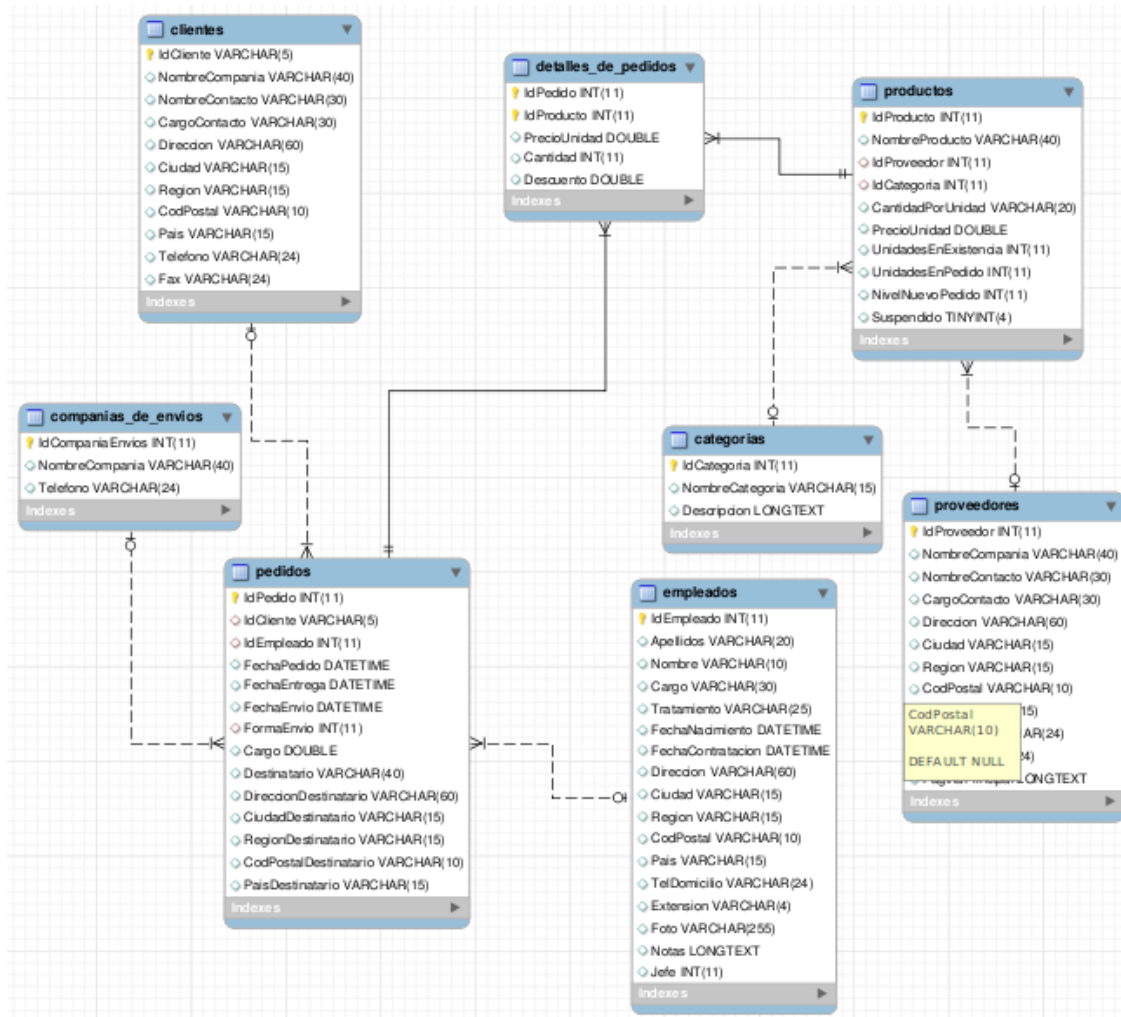
## Ingeniería Reversa



# Solución

## Ingeniería Reversa

5



# Solución

## Entendiendo el problema

- Venta de productos
  - Mediante pedidos
  - De diferentes empresas clientes y contactos
  - De diferentes proveedores
  - De diversas categorías
- Un mismo pedido puede tener
  - Más de un mismo productos
  - Dos o más productos diferentes
- Pueden haber descuentos

# Solución

## Creando la tabla de Hechos

- **Objetivo**

- Modelo orientado a las ventas

- **Granularidad**

- Detalle por **producto** vendido en cada pedido

- 1 pedido

- 2 unidades de Producto A

- 3 unidades de Producto B

- ➔ Para este pedido se deberían considerar dos nuevos registros en tabla de Hechos

# Solución

## Creando la tabla de Hechos

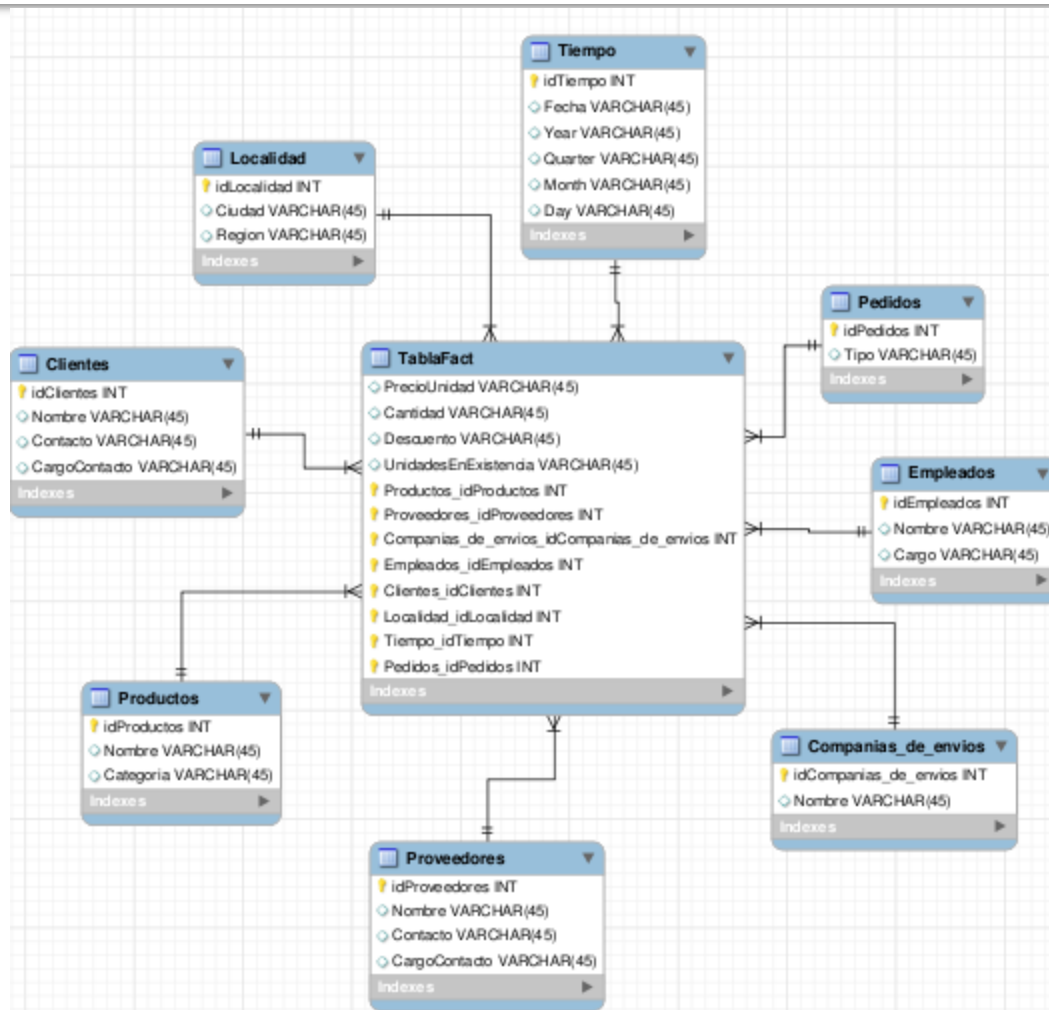
- **Objetivo**
  - Modelo orientado a las ventas
- **Medidas**
  - Precio por unidad
  - Cantidad
  - Descuento
  - Unidades en existencia

# Solución

## Las Dimensiones

- Cómo caracterizar las medidas:
  - **Tiempo**: Año, mes, día, etc.
  - **Clientes**: Nombre, contacto, cargo del contacto, etc.
  - **Proveedores**: Nombre, contacto
  - **Pedidos**: Tipo
  - **Compañías de Envíos**: Nombre
  - **Empleados**: Cargo, nombre
  - **Productos**: Categoría, nombre, etc.

# Solución Final





# Laboratorio #2

Desde modelo estrella a Consultas OLAP

# La herramienta: JasperServer

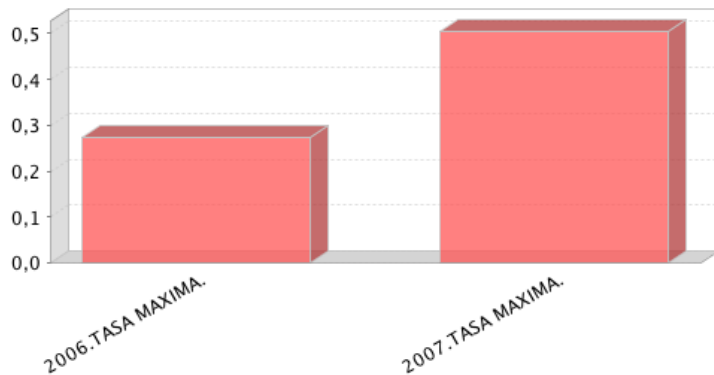
- Suite open source de Business Intelligence
  - <http://www.jaspersoft.com>
- Comprende
  - **Motor OLAP Mondrian**
  - Sistema de reporte JasperReport
  - iReport
  - JasperETL
- Alternativa a:
  - Pentaho
  - PALO

# JasperServer

## Porcentaje de centros con enfermedades de alto riesgo [tasa]



Chart



Slicer:

centros\_enfermos.

		Measures
Tiempo	Enfermedad	centros_enfermos
2006	+TASA MAXIMA	27,29%
2007	+TASA MAXIMA	50,38%

Slicer: This MDX query had no slicer (where-clause)

## Porcentaje de centros con enfermedades de alto riesgo [tasa]



Chart



Slicer:

- 2007.Caligidosis.
- 2007.Flavobacteriosis.
- 2007.Furunculosis atipica.
- 2007.Necrosis Pancreatica Infecciosa (IPN).
- 2007.Piscirickettsiosis (SRS).
- 2007.Renibacteriosis (BKD).
- 2007.Streptococosis.
- 2007.Vibriosis.
- 2007.Yersiniosis.

		Measures
Tiempo	Enfermedad	centros_enfermos
2007	Caligidosis	11,03%
	Flavobacteriosis	10,84%
	Furunculosis atipica	3,04%
	Necrosis Pancreatica Infecciosa (IPN)	41,83%
	Piscirickettsiosis (SRS)	50,38%
	Renibacteriosis (BKD)	15,59%
	Streptococosis	0,76%
	Vibriosis	11,60%
	Yersiniosis	0,57%

Slicer: This MDX query had no slicer (where-clause)

# Uso de la Herramienta

- Ingrese a:
- Espere instrucciones acerca del uso
- Familiarícese con el programa

# Actividad Práctica

1. Cree una consulta OLAP por cada dimensión del modelo
  - a) Indique ¿qué tipo de decisiones pueden ser tomadas con esta información?
2. Cree una consulta OLAP que involucre dos dimensiones del modelo
  - a) Indique ¿qué tipo de decisiones pueden ser tomadas con esta información?
3. Cree una consulta OLAP que involucre tres dimensiones del modelo
  - a) Indique ¿qué tipo de decisiones pueden ser tomadas con esta información?

# Actividad Practica

## ■ P1

- Estimar los mayores niveles de venta y margen a nivel regional, comunal y por cliente
- Determinar la relación entre los indicadores de venta y las Compañías de envíos
- Hacer una evaluación de desempeño basado en el nivel de ventas realizado por cada empleado
- Evaluar los tipos de pedido de acuerdo a sus niveles de venta
- Analisis de ventas por productos
- Analisis por proveedores
- Análisis de tendencia en el tiempo de los indicadores de venta

# Actividad Practica

- P<sub>2</sub>
  - Analisis de tendencia de ventas de Audio Menor para el periodo 2009, contrastado Con el precio promedio de venta por unidad
- P<sub>3</sub>
  - Análisis de tendencia de ventas de los productos de los proveedores de Audio Menor de la Región Metropolitana para los meses del 2009

# Solución P1

Clientes	Measures			
	Cantidad	Precio_Promedio_Unidad	Unidades_Existencia	Descuento_Promedio
-All Clientess	15.582	204.713	100.670	0,151
+I	1.287	196.673,544	8.675	0,146
+II	1.301	228.812,222	7.404	0,144
+III	1.011	231.756,129	5.668	0,163
+IV	998	201.938,905	7.171	0,147
+IX	948	178.118,571	7.085	0,15
+RM	752	254.326,207	5.765	0,158
+V	1.038	196.698,661	6.426	0,162
+VI	1.136	192.098,28	7.703	0,149
+VII	1.221	198.990	8.042	0,15
+VIII	1.520	182.223,831	10.105	0,154
+X	928	195.828,462	6.348	0,147
+XI	1.352	238.499,677	7.508	0,152
+XII	1.066	164.150	5.469	0,148
+XIV	1.024	222.745,556	7.301	0,151

Estimar los mayores niveles de venta y margen a nivel regional, comunal y por cliente



# Solución P1

	Measures			
Companias_envios	Cantidad	Precio_Promedio_Unidad	Unidades_Existencia	Descuento_Promedio
-All Companias_envios	15.582	204.713	100.670	0,151
Chilexpress	3.956	238.310,988	24.712	0,152
Take&Go	3.625	186.231,517	24.824	0,155
TravelFast	3.945	189.550,158	25.344	0,151
UPS	4.056	205.934,664	25.790	0,147

Determinar la relación entre los indicadores de venta y las Compañías de envíos

# Solución P1

Empleados			Measures				
(All)	Cargo	Nombre	▶ Cantidad	▶ Precio_Promedio_Unidad	▶ Unidades_Existencia	▶ Descuento_Promedio	
-All Empleadoss			15.582	204.713	100.670	0,151	
All Empleadoss	+Analista		2.260	204.725,593	15.332	0,154	
	+Consultor Experto		2.322	212.128,298	13.525	0,145	
	+Consultor Junior		951	220.675,039	6.215	0,159	
	+Ejecutivo Call center		1.053	189.638,276	7.837	0,158	
	+Gerente Comercial		694	279.222,323	4.796	0,146	
	+Gerente Finanzas		1.323	168.972,857	8.415	0,158	
	+Gerente General		1.005	173.030,323	6.401	0,16	
	+Gerente Operaciones		1.391	214.185,531	8.660	0,136	
	-Gerente Ventas		208	165.630	1.421	0,143	
	Gerente Ventas		BORIS GUERREZ	70	122.878,889	447	0,151
			CRISÓSTOMO PLAZA	91	120.353,636	636	0,135
			ZITA ALESON	47	342.190	338	0,144
	+Secretaria			2.852	230.086,591	18.135	0,152
	+Subgerente Comercial			1.523	160.147,36	9.933	0,152

Hacer una evaluación de desempeño basado en el nivel de ventas realizado por cada empleado

# Solución P1

Pedidos		Measures			
(All)	Tipo	• Cantidad	• Precio_Promedio_Unidad	• Unidades_Existencia	• Descuento_Promedio
-All Pedidoss		15.582	204.713	100.670	0,151
All Pedidoss	Mayoreo	9.795	207.687,546	63.868	0,154
	Mediana Escala	4.279	202.126,612	27.830	0,146
	Menudeo	1.508	192.282,553	8.972	0,148

Evaluar los tipos de pedido de acuerdo a sus niveles de venta

# Solución P1

Productos			Measures				
(All)	Categoria	Producto	Cantidad	Precio_Promedio_Unidad	Unidades_Existencia	Descuento_Promedio	
-All Productoss			15.582	204.713	100.670	0,151	
All Productoss	+Audio menor		5.660	214.876,396	36.452	0,153	
	+Computacion		25	68.990	85	0,17	
	+Electrodomesticos		4.822	196.771,99	31.496	0,15	
	-Linea blanca		5.075	201.724,663	32.637	0,151	
	Linea blanca	Biblioteca		237	138.050,606	1.591	0,152
		Centrifuga		283	300.684,444	1.668	0,137
		Cocina		344	188.056,667	1.899	0,169
		Computador		1.032	184.819,457	6.180	0,153
		Impresora		323	168.584,595	2.115	0,13
		Laptop		729	207.610	5.199	0,148
		Lavadora		181	200.313,529	1.984	0,143
		Mesa oficina		350	256.715	1.818	0,166
		Minicomponente		269	160.621,579	2.120	0,157
		Monitor		336	222.555,217	2.066	0,157
	Plasma		233	187.286,296	1.656	0,143	
Refrigerador		377	245.707,391	2.423	0,146		
Unidad de respaldo		381	172.233,902	1.918	0,147		

Analisis de ventas por productos

# Solución P1

Proveedores				Measures			
(All)	Region	Ciudad	Nombre	Cantidad	Precio_Promedio_Unidad	Unidades_Existencia	Descuento_Promedio
-All Proveedoreess				15.582	204.713	100.670	0,151
All Proveedoreess +IX				685	154.103,402	5.189	0,145
-RM				9.551	206.737,541	61.693	0,152
RM -Santiago				9.551	206.737,541	61.693	0,152
Santiago +Apple				902	232.710	4.735	0,151
+Dell Technology				1.454	193.613,762	10.071	0,153
+General Electric				785	191.941,456	5.344	0,147
+Lenovo				767	229.790	5.020	0,143
+LG				1.011	199.144,472	6.390	0,165
+Olidata				806	182.670,851	4.934	0,159
+Packard Bell				722	254.053,83	4.702	0,148
+Samsung				691	188.808,182	4.645	0,152
+Sony				1.578	187.990	10.137	0,147
+Toshiba				835	235.863,874	5.715	0,161
+V				2.758	194.816,625	15.733	0,151
+VII				1.770	220.868,661	12.176	0,144
+VIII				818	219.378,43	5.879	0,161

Análisis de proveedores

# Solución P1

Tiempo			Measures				
(All)	Ano	Mes	• Cantidad	• Precio_Promedio_Unidad	• Unidades_Existencia	• Descuento_Promedio	
-All Tiempos			15.582	204.713	100.670	0,151	
All Tiempos	+2000		1.519	220.441,282	10.221	0,146	
	+2001		1.483	157.141,351	9.493	0,156	
	+2002		1.595	216.463,171	10.615	0,151	
	+2003		1.505	165.115,628	9.660	0,154	
	+2004		1.457	199.399,756	10.551	0,146	
	+2005		1.767	235.062,464	9.440	0,15	
	+2006		1.714	190.140,235	11.106	0,159	
	+2007		1.950	226.376,266	12.010	0,157	
	+2008		1.558	224.344,546	10.531	0,145	
	-2009		1.034	202.895,797	7.043	0,148	
	2009	+Abril		128	177.147,895	972	0,139
		+Agosto		95	269.066,923	616	0,152
		+Enero		137	237.931,176	897	0,168
	+Febrero		57	206.632,857	640	0,111	
	+Julio		162	176.704,286	1.129	0,135	
	+Junio		185	276.132,857	1.017	0,149	
	+Marzo		159	148.779,474	1.014	0,152	
	+Mayo		111	132.990	758	0,181	

Análisis de tendencia de los indicadores de venta

# Solución P2

			Measures								Precio_Promedio_Unidad	
			Cantidad									
Productos			Tiempo								Tiempo	
			2009									
(All)	Categoria	Producto	2009	2009	2009	2009	2009	2009	2009	2009	2009	
			2009	2009	2009	2009	2009	2009	2009	2009	2009	
-All Productoss			1.034	128	95	137	57	162	185	159	111	202.895,797
All Productoss	-Audio menor		374	74	48	34	13	59	73	55	18	215.690
	Audio menor	Centrifuga	18			18						512.990
		Cocina	31		11			2	18			354.323,333
		Computador	87	23	21	3			14	19	7	262.740
		Impresora	10	1	1				3	5		204.240
		Laptop	57	8	8	4		4	4	12	17	37.212,222
		Lavadora	28					4	20	4		455.990
		Minicomponente	2							2		12.990
		Monitor	50	14					16		9	52.990
		Plasma	3								3	768.990
		Refrigerador	24	9	7				8			148.240
		Unidad de respaldo	64	19		9		5	9	22		239.990
	+Electrodomesticos		332	22	28	58	19	45	26	68	66	174.728,095
	+Linea blanca		328	32	19	45	25	58	86	36	27	214.707,391

Analisis de tendencia de ventas de Audio Menor para el periodo 2009, contrastado Con el precio promedio de venta por unidad

# Solución P3

						Measures								
						Cantidad								
						Tiempo								
Productos		Proveedores				2009	2009							
(All)	Categoria	(All)	Region	Ciudad	Nombre		2009	2009	2009	2009	2009	2009	2009	2009
							2009	2009	2009	2009	2009	2009	2009	2009
-All Productoss		+All Proveedoress				1.034	128	95	137	57	162	185	159	111
All Productoss	+Audio menor	-All Proveedoress				374	74	48	34	13	59	73	55	18
		All Proveedoress	+IX			20		1			16		3	
			-RM			252	55	29	22	8	27	71	26	14
			RM	-Santiago		252	55	29	22	8	27	71	26	14
				Santiago	+Dell Technology	62	12	8			16	15		11
					+General Electric	14	3				9		2	
					+Lenovo	8			4	4				
					+LG	19							19	
					+Olidata	2					2			
					+Packard Bell	30	25						5	
					+Samsung	69	1		18			50		
					+Sony	45	14	20		4		4		3
					+Toshiba	3		1				2		
			+V			36	2	18					16	
			+VII			59	17		12	5	16		5	4
			+VIII			7						2	5	
	+Electrodomesticos	+All Proveedoress				332	22	28	58	19	45	26	68	66
	+Linea blanca	+All Proveedoress				328	32	19	45	25	58	86	36	27

Análisis de tendencia de ventas de los productos de los proveedores de Audio Menor de la Región Metropolitana para los meses del 2009



Taller #2

# Business Intelligence

Carlos Reveco  
creveco@dcc.uchile.cl

Cinthya Vergara  
cvergarasilv@ing.uchile.cl