

# Inteligencia de Negocios

Seminario de Redes Inteligentes  
Oscar Duarte

Abril 15 del 2011

# Agenda

- Base de datos relacionales
- Bodegas de datos
- Minería de datos

# Bases de datos relacionales

# Bases de datos relacionales

- ¿Cómo almacenar un listado de libros? Tablas:

Título	Autor	Editorial
Cálculo	Thomas	Willey
Física	Sears	Oxford
Circuitos	Hayt	McGH

**Almacenamiento y recuperación de la información**

# Bases de datos relacionales

- ¿y si son varios autores?

Título	Autor	Editorial
Cálculo	Thomas - Freedman	Willey
Física	Sears - Jhonson - Ljung	Oxford
Circuitos	Hayt - Ljung	McGH

# Bases de datos relacionales

- ¿y si son varios autores?

Título	Autor	Editorial
Cálculo	Thomas - Freedman	Willey
Física	Sears - Jhonson - Ljung	Oxford
Circuitos	Hayt - Ljung	McGH

Título	Autor 1	Autor 2	Autor 3	Editorial
Cálculo	Thomas	Freedman		Willey
Física	Sears	Jhonson	Ljung	Oxford
Circuitos	Hayt	Ljung		McGH

# Bases de datos relacionales

- ¿y si son varios autores?

Título	Autor	Editorial
Cálculo	Thomas - Freedman	Willey
Física	Sears - Jhonson - Ljung	Oxford
Circuitos	Hayt - Ljung	McGraw

Título	Autor 1	Autor 2	Autor 3	Editorial
Cálculo	Thomas	Freedman		Willey
Física	Sears	Jhonson	Ljung	Oxford
Circuitos	Hayt	Ljung		McGraw

# Bases de datos *relacionales*

Libros

<b>Id</b>	<b>Título</b>
1	Cálculo
2	Física
3	Circuitos



# Bases de datos *relacionales*

Libros

Id	Título
1	Cálculo
2	Física
3	Circuitos

Autores

Id	Apellido
1	Thomas
2	Freedman
3	Sears
4	Jhonson
5	Ljung
6	Hayt

# Bases de datos *relacionales*

Libros

Id	Título
1	Cálculo
2	Física
3	Circuitos

Libros\_Autores

Id_libro	Id_autor
1	1
1	2
2	3
2	4
2	5
3	6
3	5

Autores

Id	Apellido
1	Thomas
2	Freedman
3	Sears
4	Jhonson
5	Ljung
6	Hayt

# Bases de datos *relacionales*

Id	Título
1	Cálculo
2	Física
3	Circuitos

Id_libro	Id_autor
1	1
1	2
2	3
2	4
2	5
3	6
3	5

Id	Apellido
1	Thomas
2	Freedman
3	Sears
4	Jhonson
5	Ljung
6	Hayt

La clave está en las claves

# Bases de datos relacionales

- Un sólido soporte **matemático**:
  - Álgebra relacional. Codd “A Relational Model of Data for Large Shared Data Banks.” (1970)
- Un **lenguaje** versátil de almacenamiento, modificación y consulta:
  - **SQL**: Structured Query Language

# Bases de datos relacionales

- Ventajas del modelo relacional:
  - Elimina la redundancia.
  - Reduce las inconsistencias.
  - Protege la integridad de los datos.
- Múltiples **motores**:
  - Oracle, MySql, Postgresql, Sql server

# SQL. Ejemplos

- Crear una base de datos:

```
CREATE DATABASE prueba;
```

- Crear una tabla:

```
CREATE TABLE Libros(id int  
auto_increment, título char(255));
```

- Insertar información:

```
INSERT INTO Libros(título) VALUES  
( 'Cálculo' );
```

# SQL. Ejemplos

- Recuperar información:

```
SELECT * FROM Libros WHERE título  
LIKE 'Cálculo';
```

- Recuperar información relacionada

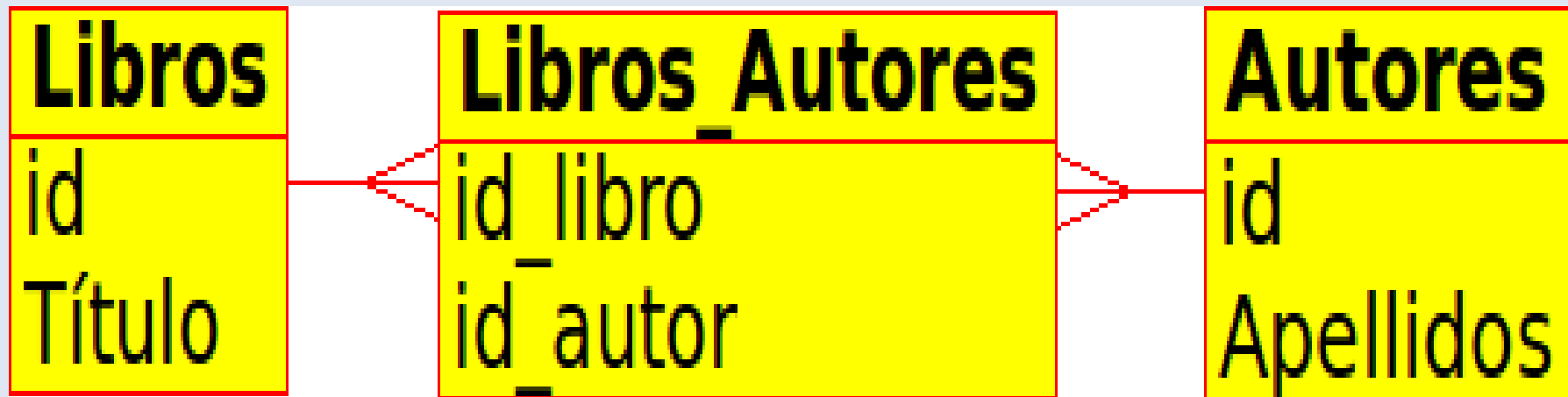
```
SELECT * FROM Libros
```

```
INNER JOIN Libros_Autores ON
```

```
Libros.id=Libros_Autores.id_libro ...
```

```
... WHERE Apellido LIKE 'Thomas';
```

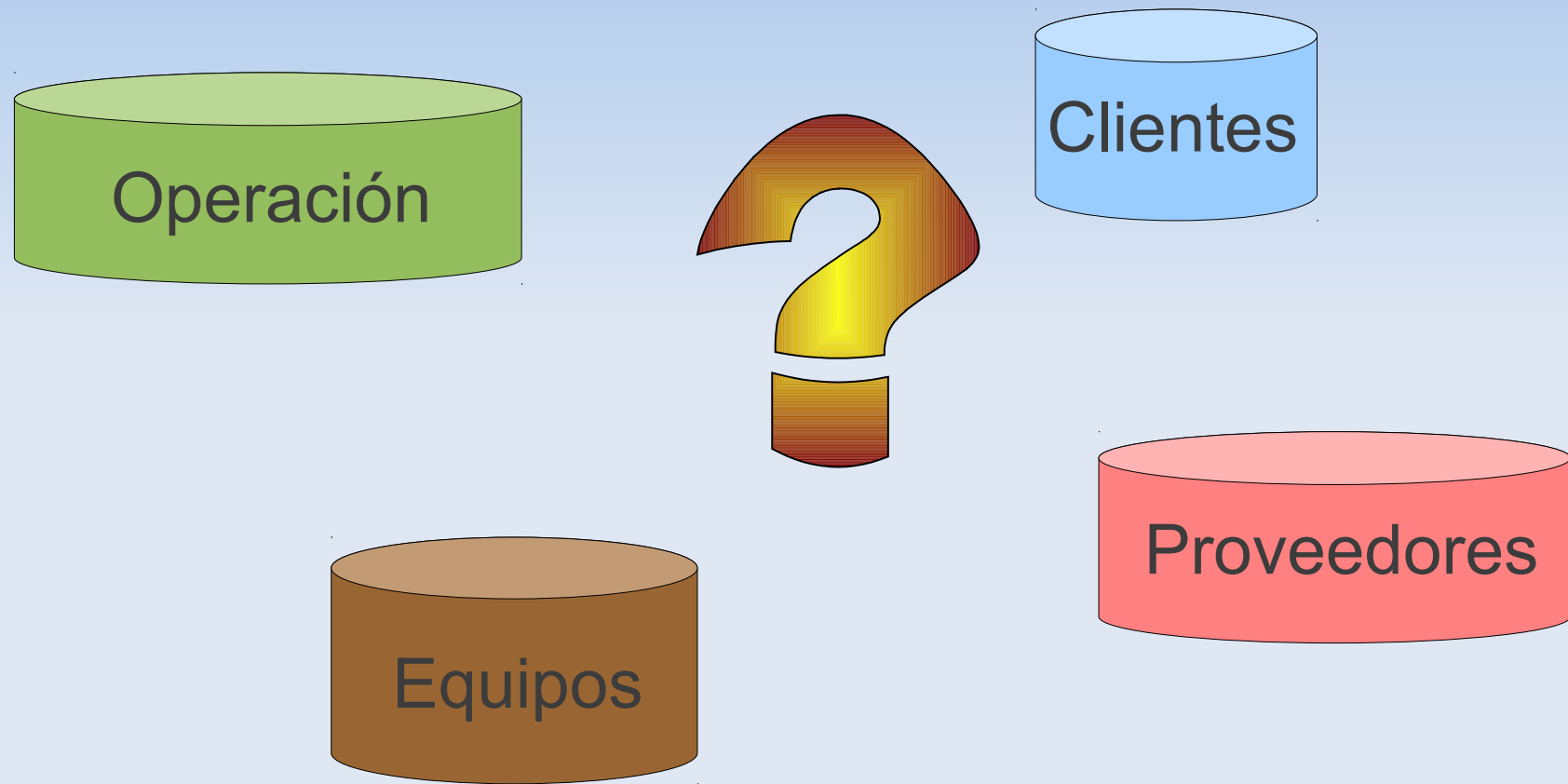
# El modelo Entidad-Relación. UML





# Bodegas de datos

# El problema: B.D. heterogéneas

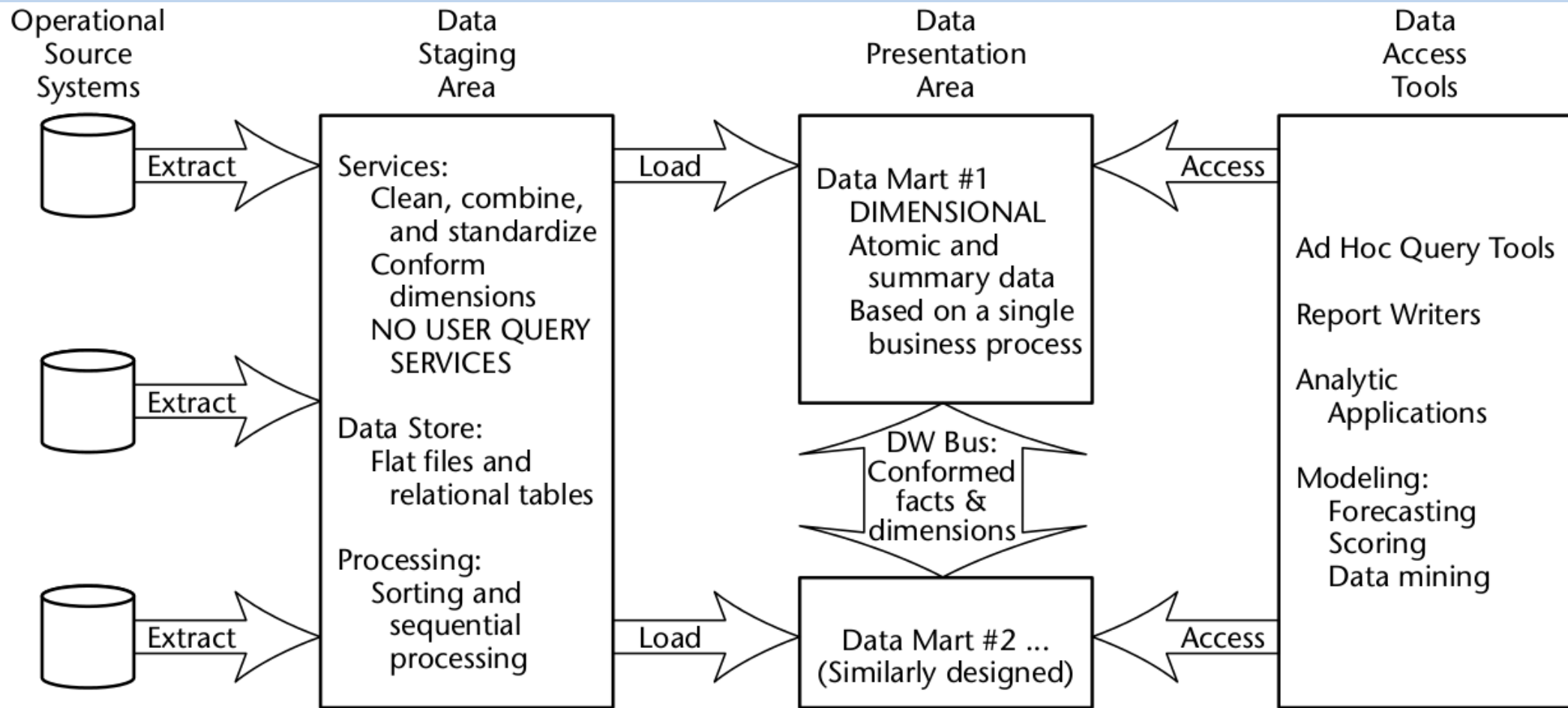


Una torre de babel

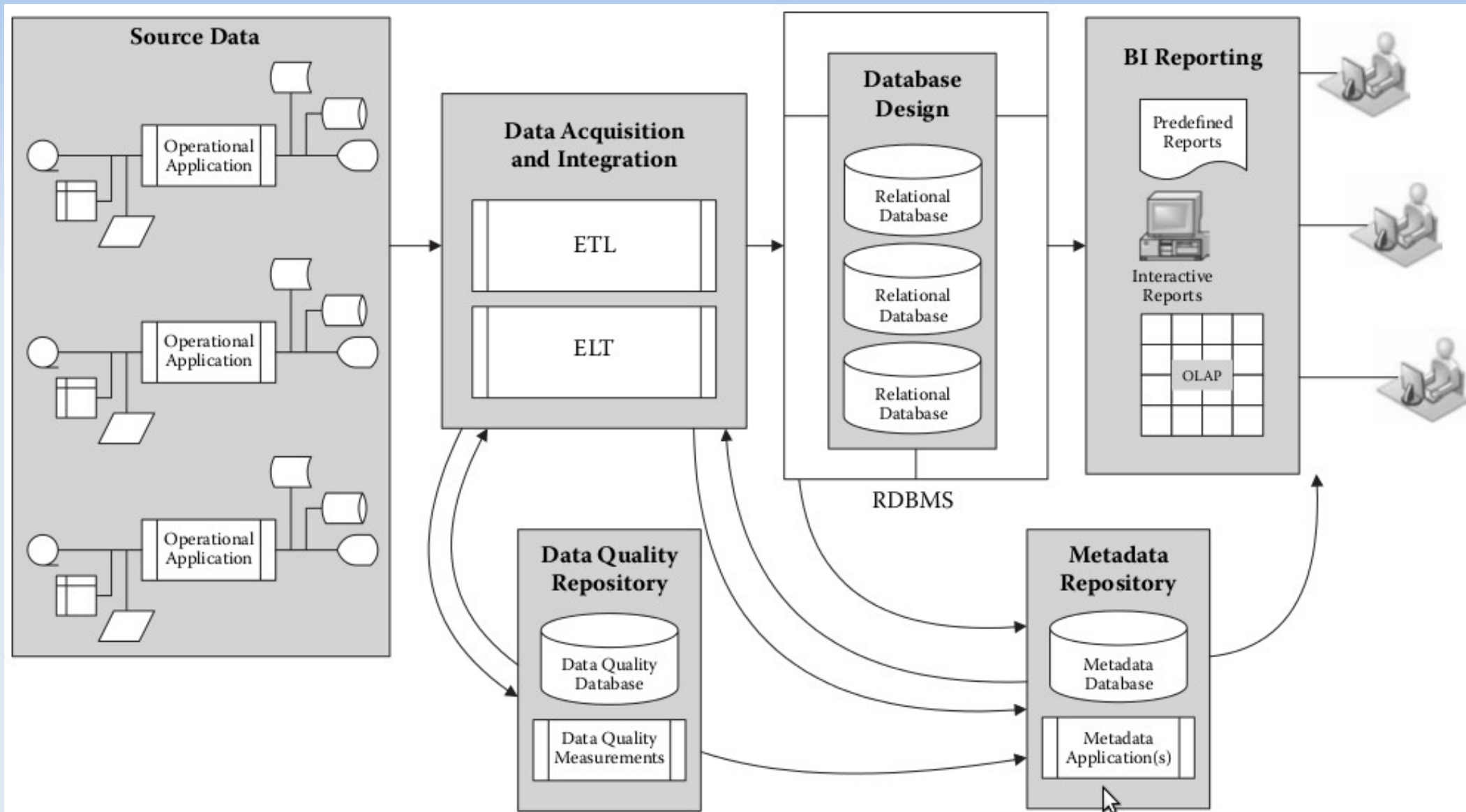
# Tras la solución

- Orientación temática.
- Integración de datos.
- No volatilidad.
- Contextos temporales.
- Una versión de la 'verdad'.
- Inversión a largo plazo.

# El esquema usual



# El esquema usual

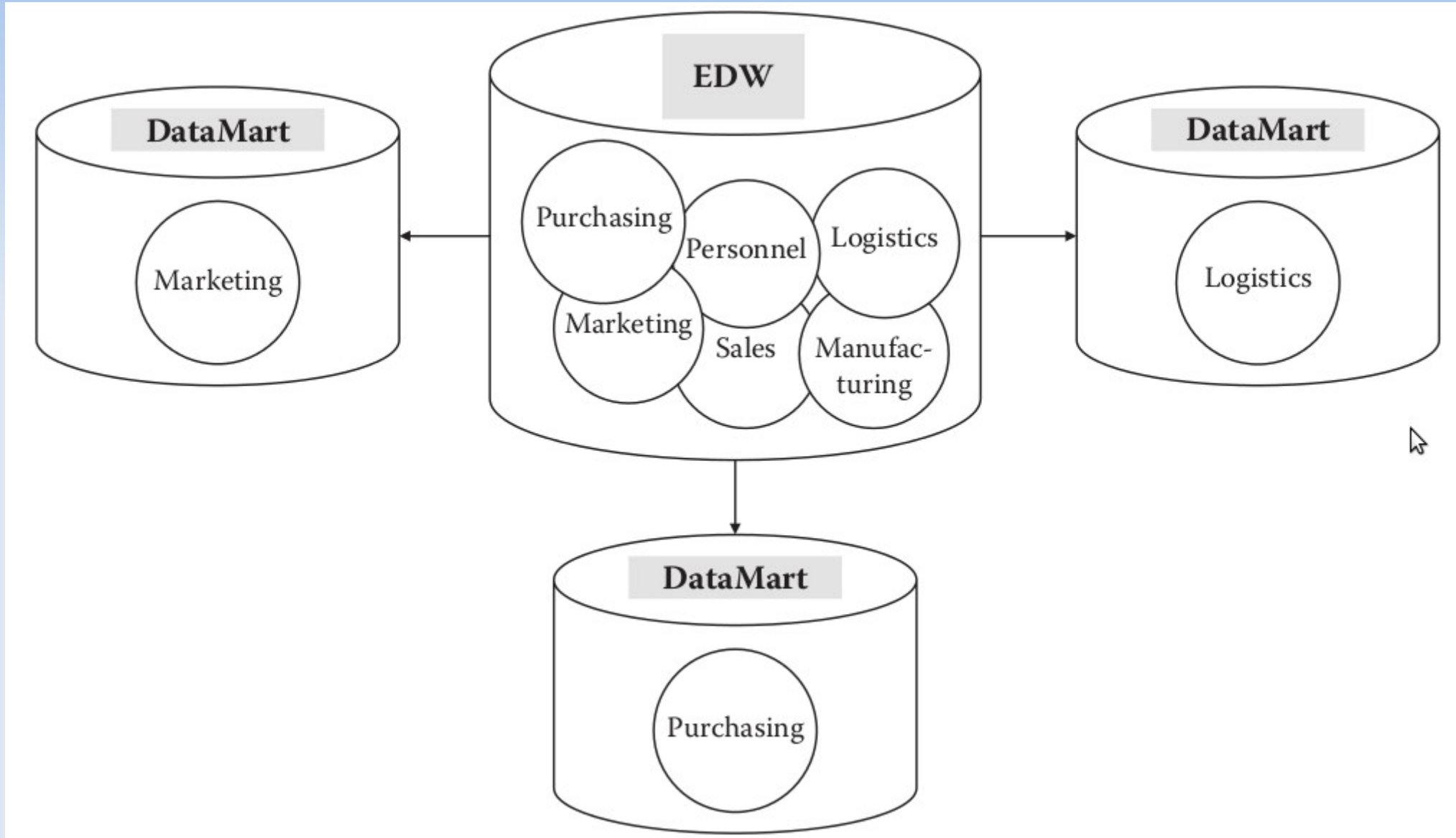


# De la operación a la bodega

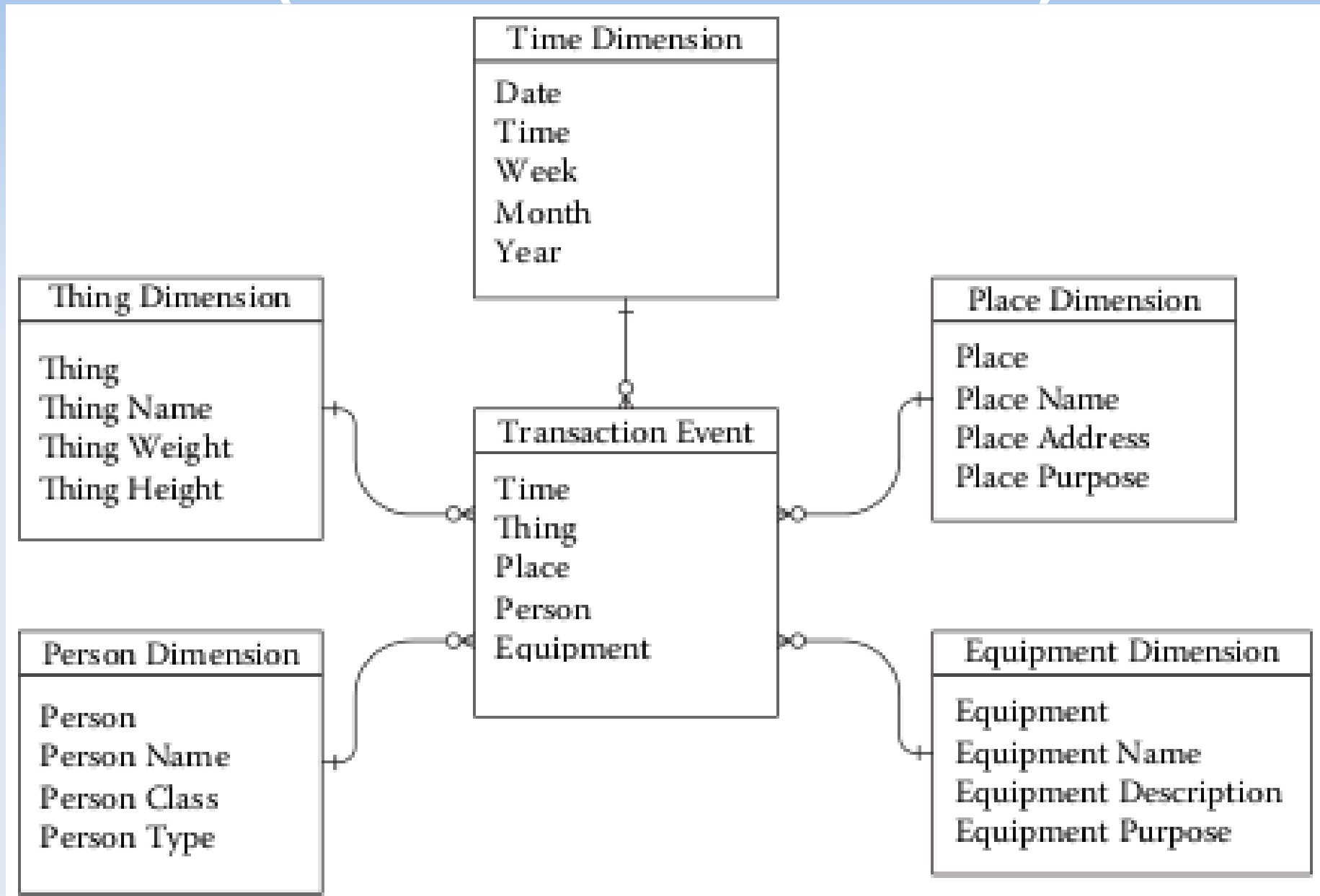
- ETL:
  - Extract, Transform, Load
- Limpieza
- Completez
- Coherencia
- Estandarización



# Data Warehouse / DataMarts



# Modelo dimensional (modelo en estrella)

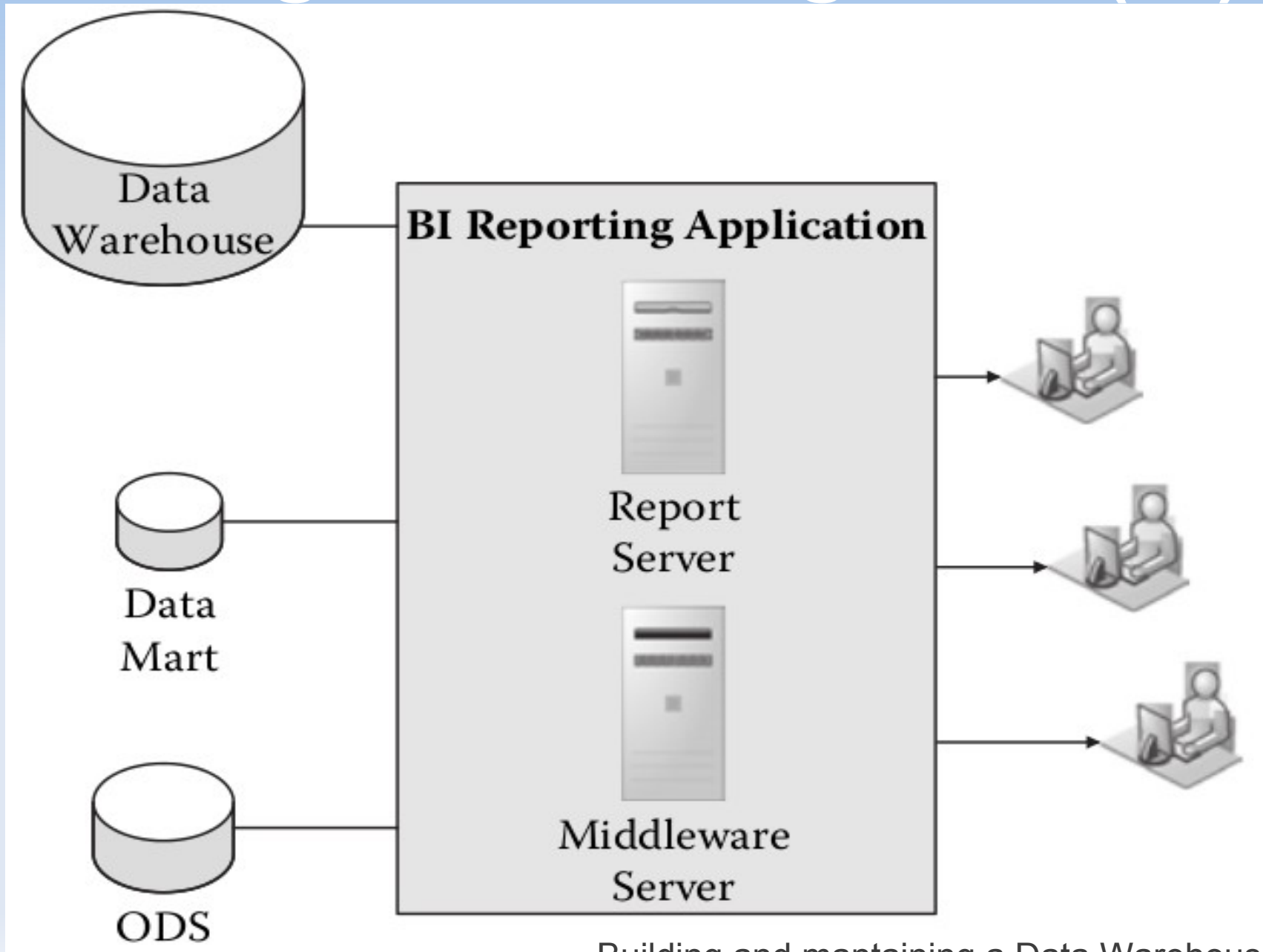




# Para destacar

- Los DataMarts no son sólo para resumir.
- Los modelos dimensionales no son soluciones de departamento, no de empresa.
- La definición del mínimo nivel de detalle de la tabla de eventos debe contemplar necesidades de negocio futuras.

# Reportes de Inteligencia de negocios (BI)



# BI - Factores de éxito

- Desempeño. Administración de recursos.
- Interfaz de usuario. Debe reflejar la empresa.
- Presentación de la arquitectura de datos. Origen de los datos.
- Respeto del modelo de datos.

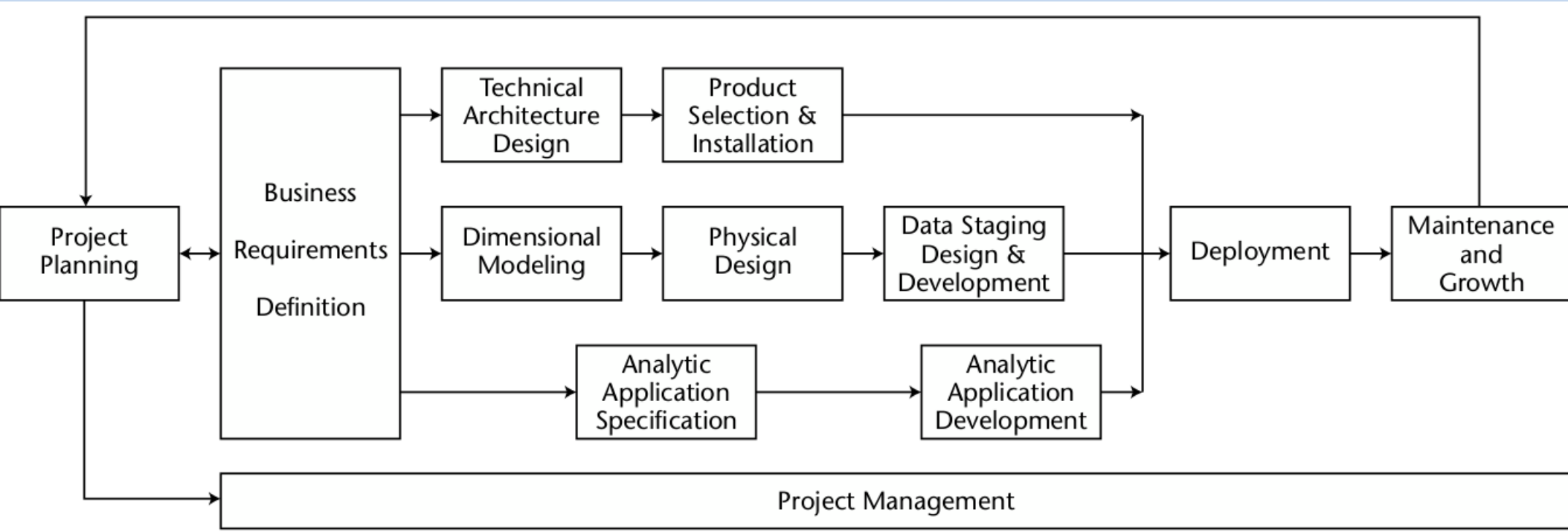
# (BI) - Factores de éxito

- Capacidad para responder preguntas.
- Movilidad.
- Flexibilidad.
- Disponibilidad.
- Enfoque al usuario: adaptabilidad, programación de tareas, proactividad y reactividad.

# Tipos de reporte

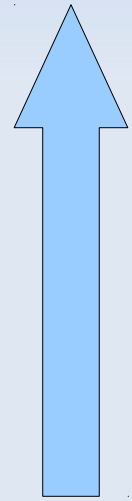
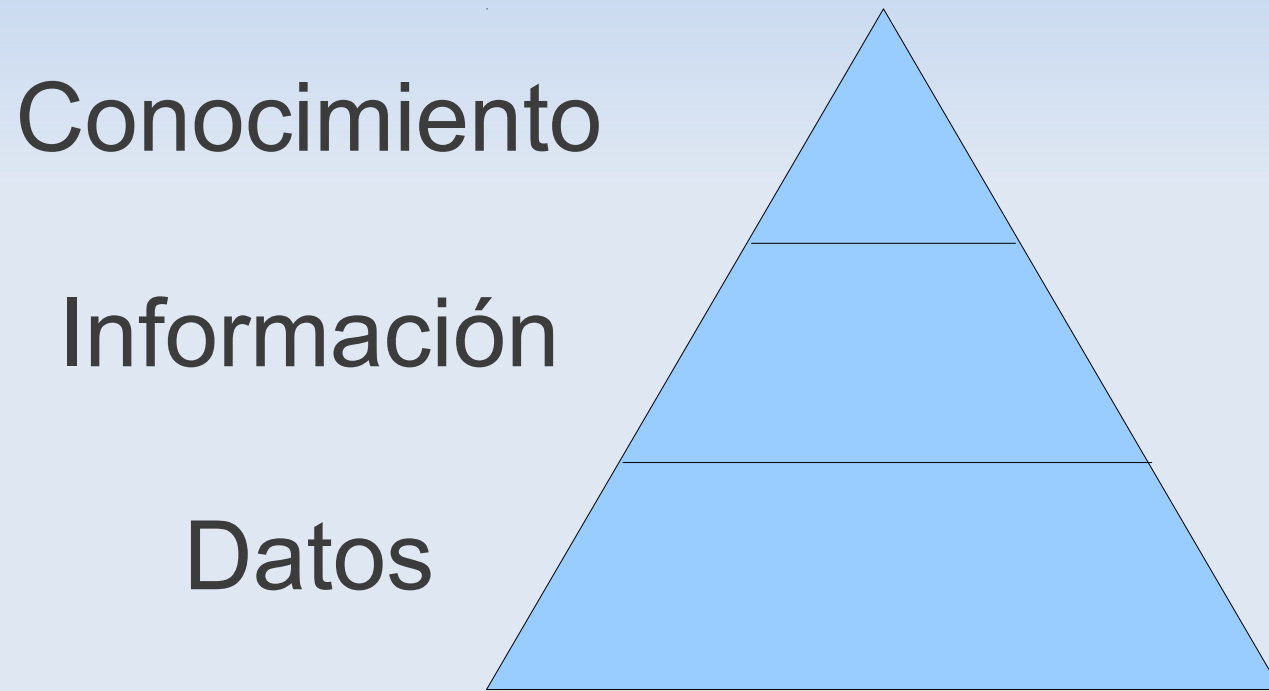
- Reportes predefinidos.
- Reportes interactivos.
- Reportes OLAP (*Online Analytical Process*)

# Ciclo de vida



# Minería de datos

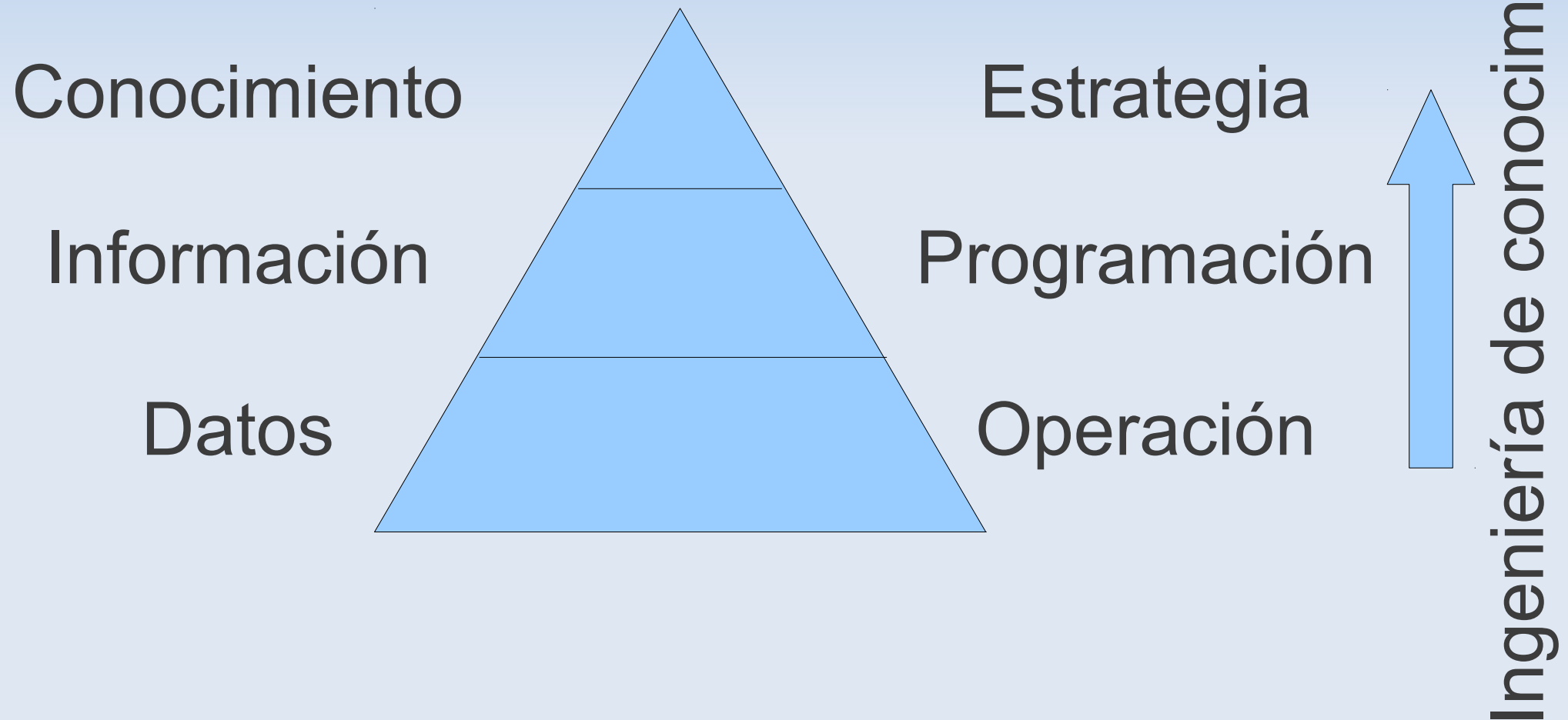
# Piramide de descubrimiento de conocimiento



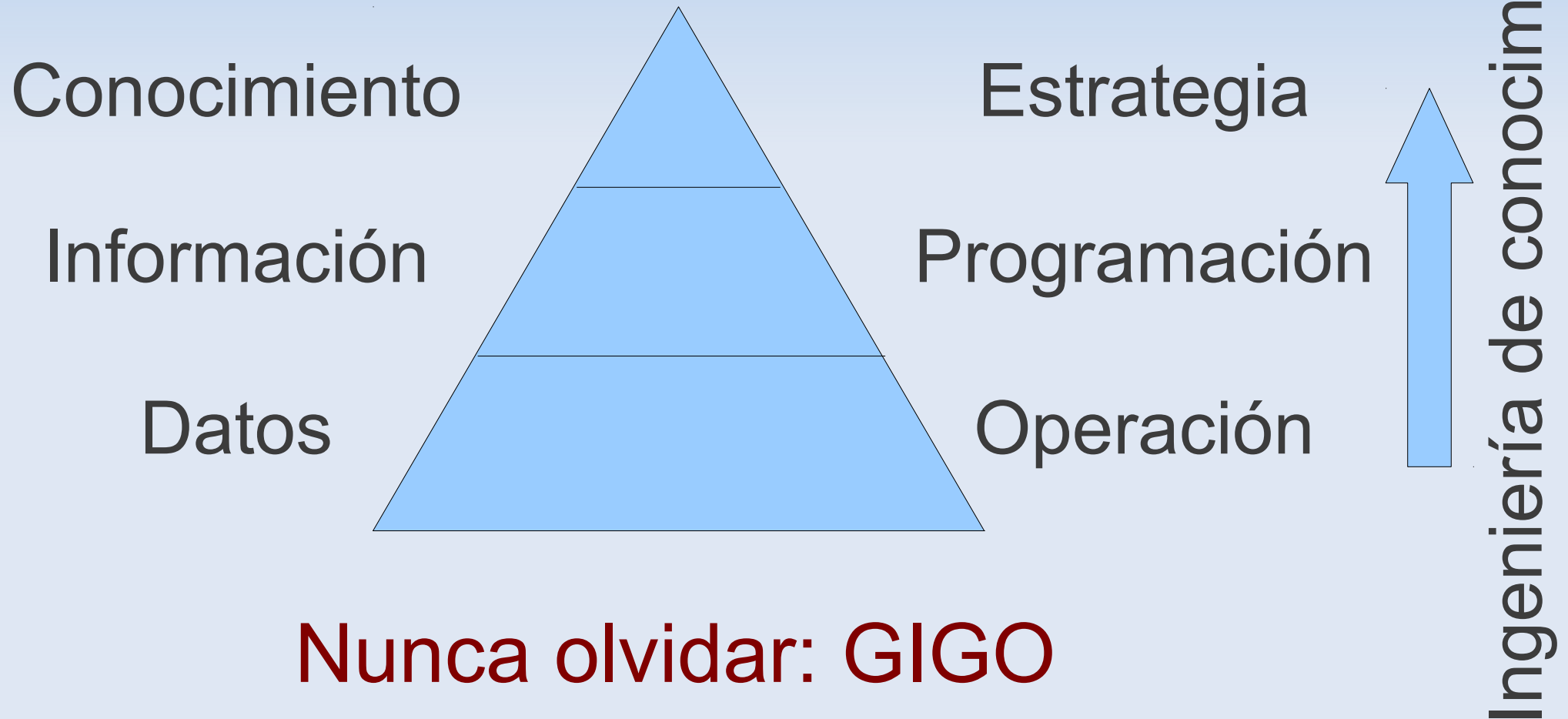
Ingeniería de conocimiento



# Piramide de descubrimiento de conocimiento



# Piramide de descubrimiento de conocimiento



# Minería de datos

- Tipos de análisis:
  - Exploratorios: Búsqueda de hipótesis y relaciones.
  - Confirmatorios: Pruebas de hipótesis.
- Es un ejercicio dinámico. Interacción usuario – herramienta.

# Minería de datos y estadística

- Se usan muchos conceptos y técnicas de la estadística.
- La minería de datos abarca el pre y pos-procesamiento de datos, por ejemplo la visualización.

# Minería de datos y estadística

- El volumen de información y la calidad de la misma hacen impráctico un análisis estadístico 'tradicional'.
- La estadística es más probatoria que exploratoria.

# Algunas técnicas usadas en Minería de datos

- Agrupamiento.
- Reglas de asociación.
- Redes Neuronales.
- Mapas autoorganizativos.
- Árboles de decisión.
- Algoritmos genéticos.
- Minería de texto.

# Algunas aclaraciones sobre la minería de datos

- **No** descubre automáticamente patrones escondidos en los datos.
- **No** está diseñada para analistas de negocios que 'no manejen números'.
- Los hallazgos **no** se traducen fácilmente en decisiones de negocio.
- **No** se circunscribe a los sistemas soporte de decisión.

# Calidad de los datos

- Calidad: 'Éxito o fracaso en cumplir las expectativas del cliente'. Deming.
- La calidad debe ser medible:
  - Completez
  - Latencia
  - Precisión
- Mejoramiento continuo de la calidad.



# Data Quality Service Level Agreement (SLA)

- Expectativas de calidad claras.
- Expectativas de calidad cuantificables y medibles .
- Herramientas de medición desde la Bodega de Datos.
- Publicación de resultados.
- Resultados usados en mejoramiento.