

# Bases de Datos I

Más allá de las Bases de Datos Relacionales

Lic.

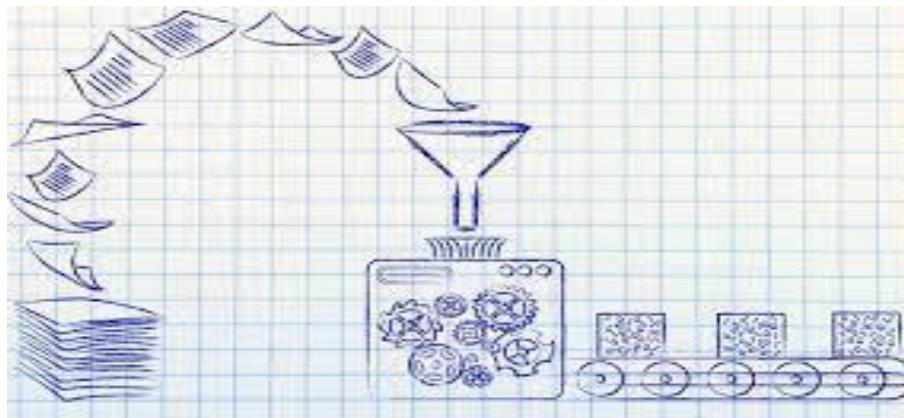
Dra. C. Lucina García Hernández  
Lic. Carlos León González

Lic.

Departamento de Computación  
Facultad de Matemática y Computación  
Universidad de La Habana

5 de diciembre de 2023

# Motivación

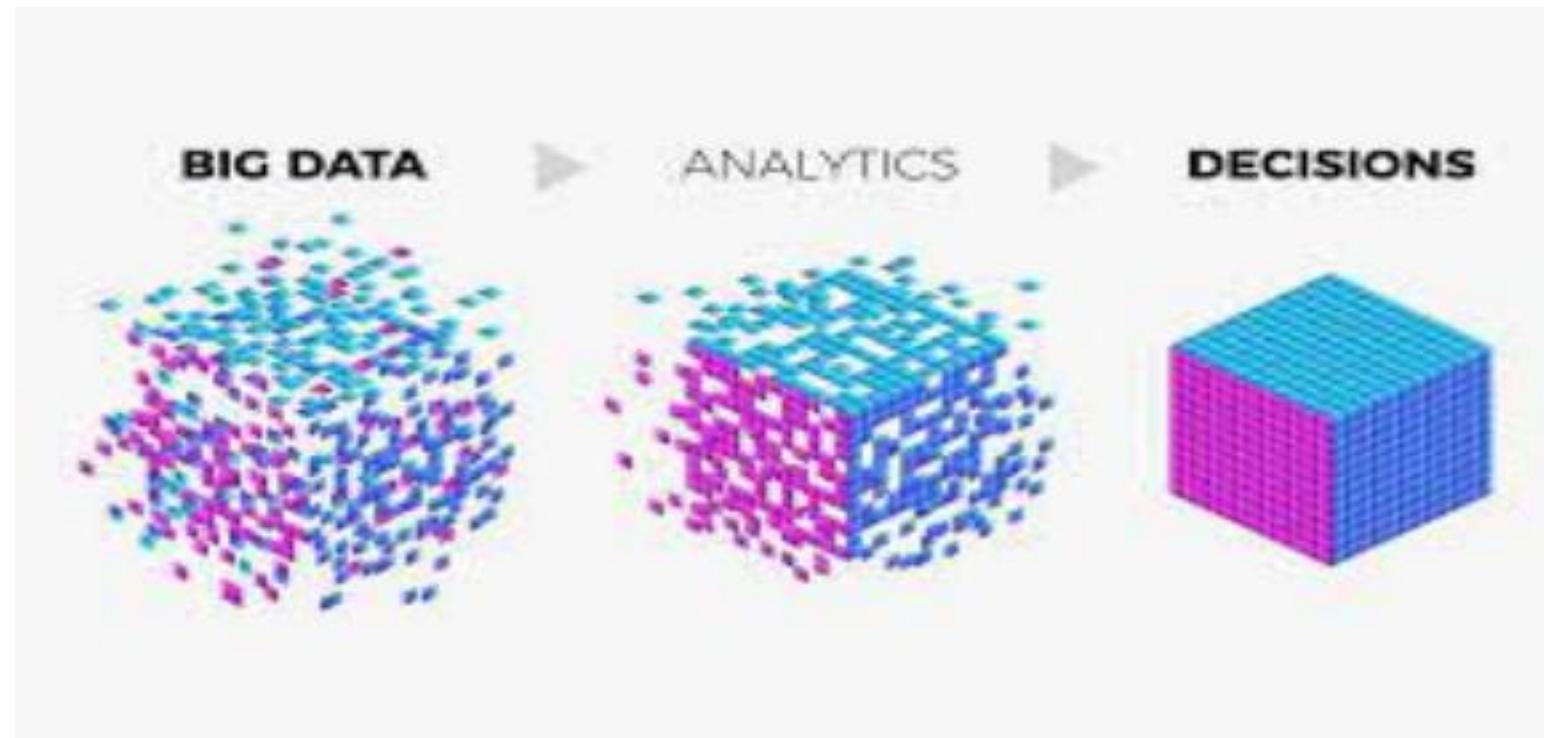


El término “base de datos” se escuchó por primera vez en 1963 en un simposio en California. Se la definió como “**un conjunto de “información” relacionada que se encuentra agrupada o estructurada**”.

# Motivación



El término “base de datos” se escuchó por primera vez en 1963 en un simposio en California. Se la definió como “**un conjunto de “información” relacionada que se encuentra agrupada o estructurada**”.



# Objetivos de la Conferencia 11

- Acercarse a los principales campos en la evolución y la vigencia de los sistemas de bases de datos desde la perspectiva del almacenamiento y el procesamiento de la información.
- Conocer la coexistencia de diferentes enfoques y plataformas en correspondencia con las funcionalidades requeridas y las exigencias de los procesos de toma de decisiones.

# La información como recurso estratégico

- La influencia impetuosa de los archivos digitales sobre la mayoría de las funciones académicas, científicas, empresariales, culturales y personales
- El teleprocesamiento de la información y la compartición de contenidos
- Las tecnologías de extracción y generación del conocimiento



# SISTEMAS DE INFORMACION

Recopilar

Procesar

Almacenar

Diseminar

Coordinar

Controlar

Analizar

Visualizar



# SISTEMAS DE INFORMACION

Recopilar

Procesar

Almacenar

Diseminar

Coordinar

Controlar

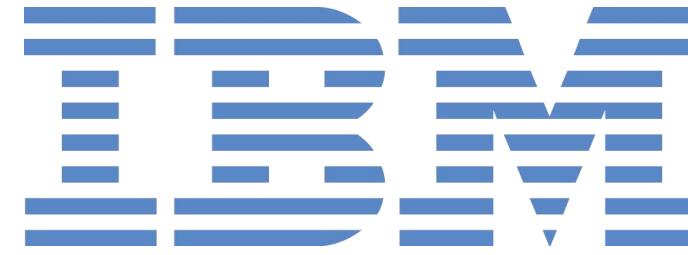
Analizar

Visualizar



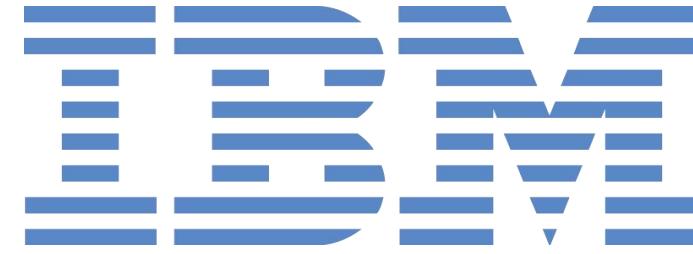
Toma de  
decisiones

# Sistemas de Información



1960's

# Sistemas de información



1960's

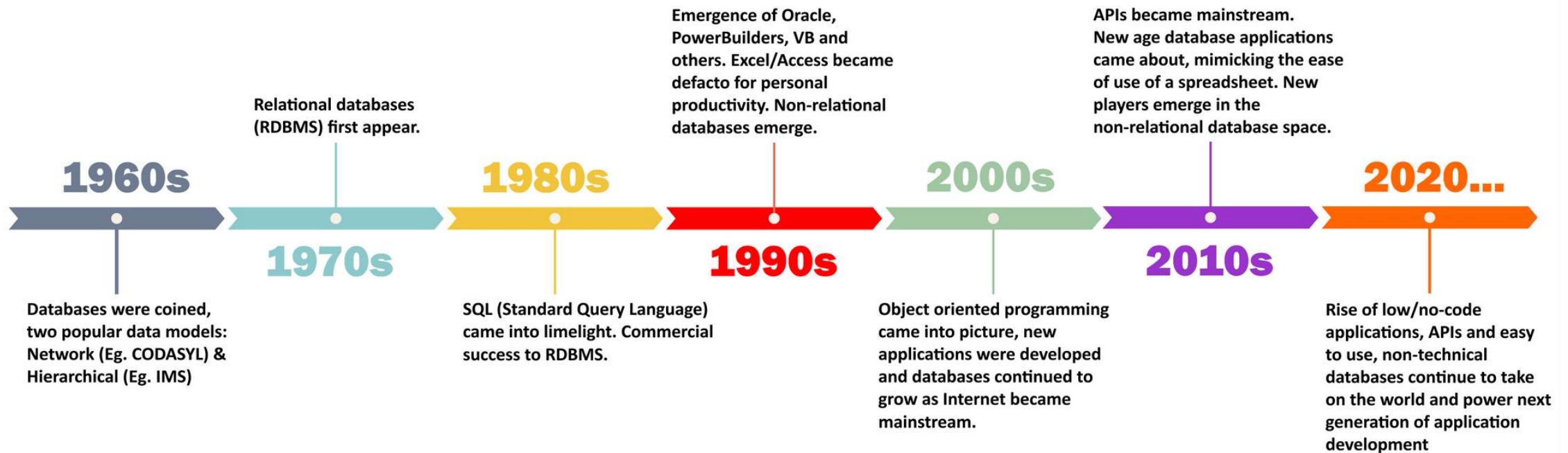
amazon.com®  
2000's

# Algo de historia comentada

Máquina tabuladora - [Herman Hollerith \(1860-1929\)](#)



## History of Databases (1960-2020)



¿Qué tipo de operaciones se realizaban  
comúnmente en las empresas?

# UNA TRANSACCIÓN

From account:	Savings	
To account:	Checking	
Amount:	1,000.00	
<b>Transfer</b>		

Steps

```
graph TD; Start((Start)) --> 1((1)); 1 --> 2((2)); 2 --> End((End));
```

Begin transaction

Withdraw 1,000 from savings

Deposit 1,000 into checking

End transaction

# ACID

# UNA TRANSACCIÓN

From account:	Savings	
To account:	Checking	
Amount:	1,000.00	
<b>Transfer</b>		

Steps

```
graph TD; Start([Start]) --> 1([1]); 1 --> 2([2]); 2 --> End([End]);
```

Begin transaction

Withdraw 1,000 from savings

Deposit 1,000 into checking

End transaction

**Atomicity**

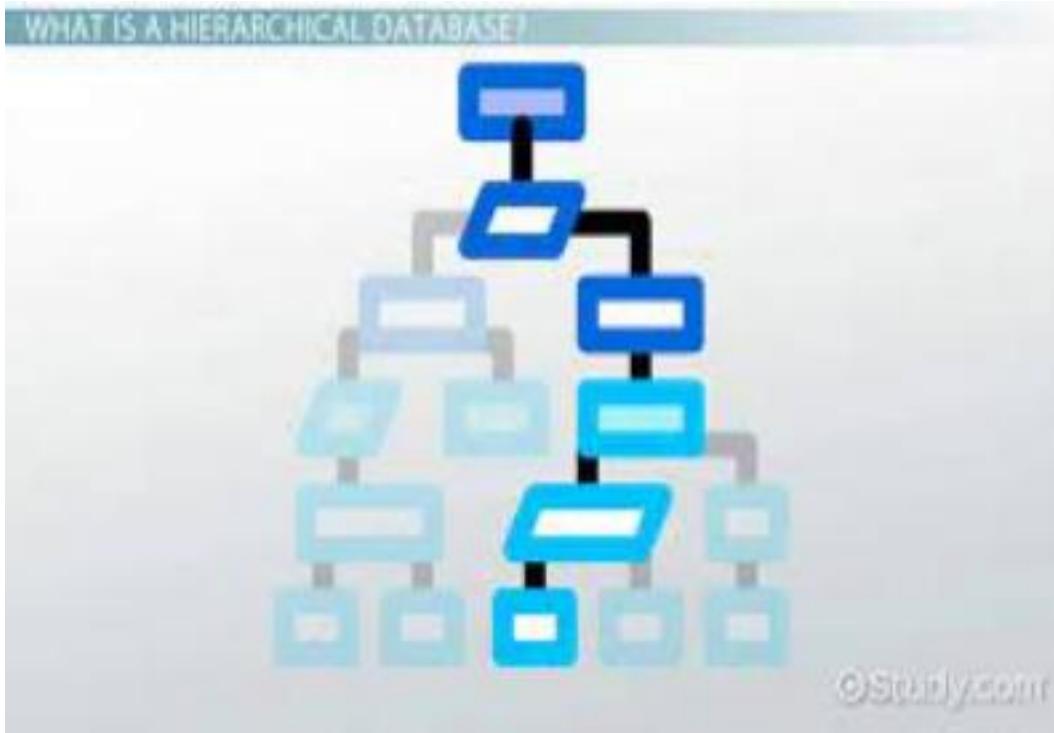
**Consistency**

**Isolation**

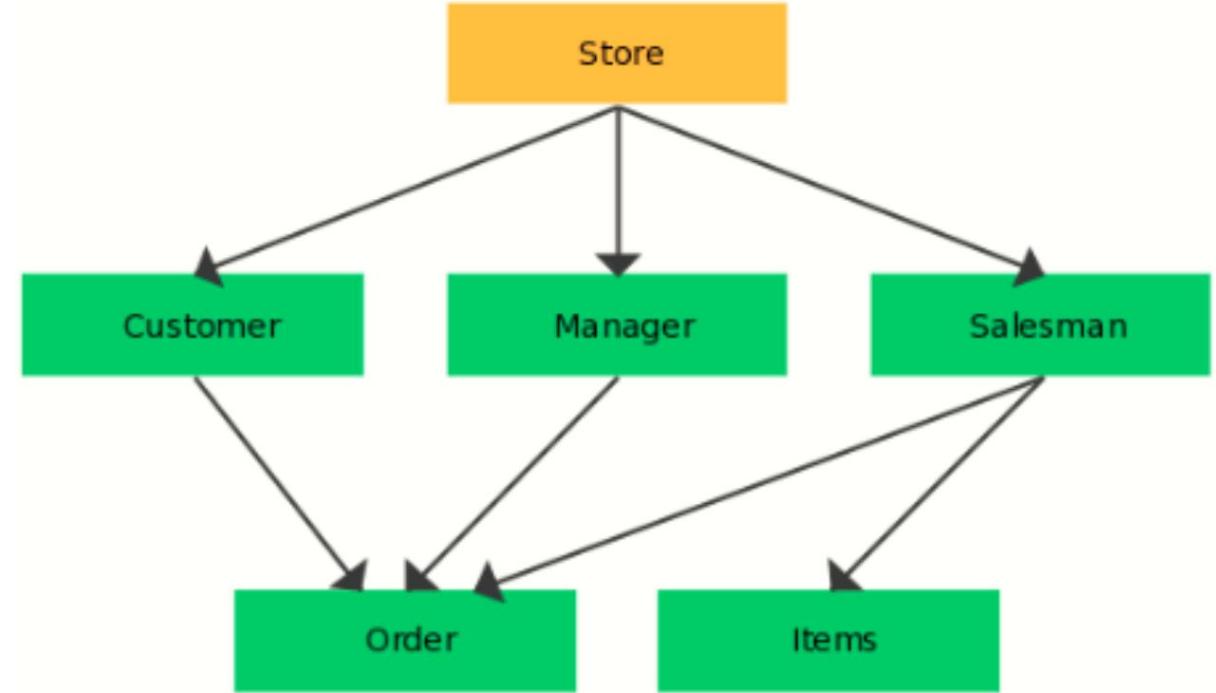
**Durability**

# SISTEMAS DE BASES DE DATOS TRANSACCIONALES

# Bases de datos pre-relacionales



Modelo jerárquico



Modelo reticular

# Bases de datos pre-relacionales

## VENTAJAS

Se garantiza la recuperación rápida de los datos en caso de interrelaciones predeterminadas

Numerosos fenómenos de la realidad pueden ser representados de manera natural

# Bases de datos pre-relacionales

## VENTAJAS

Se garantiza la recuperación rápida de los datos en caso de interrelaciones predeterminadas

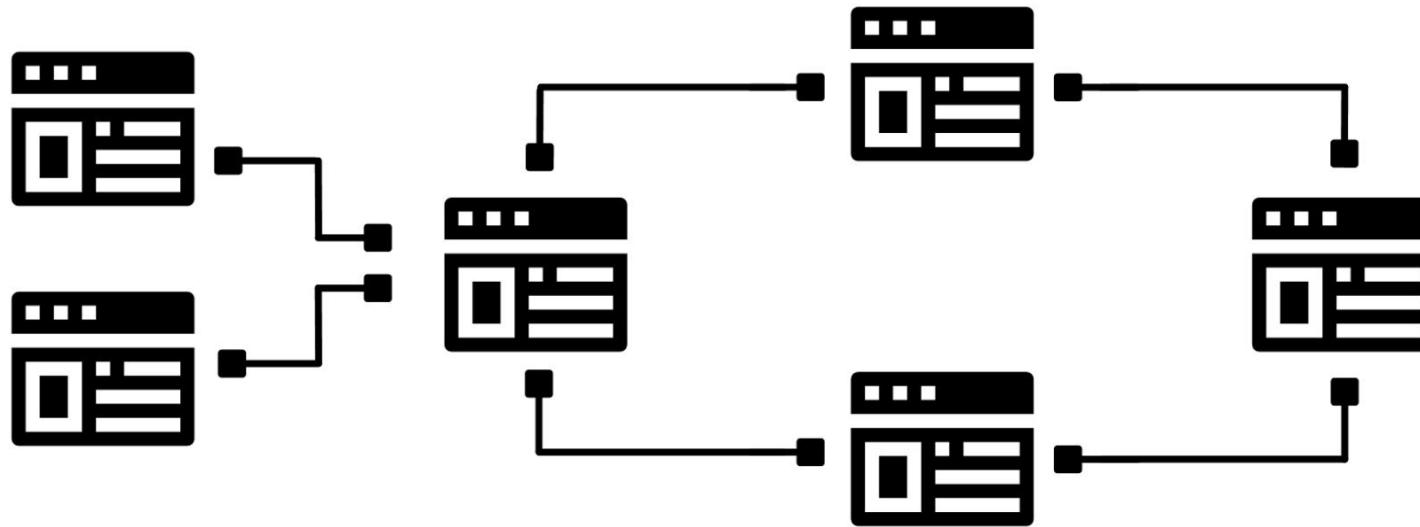
Numerosos fenómenos de la realidad pueden ser representados de manera natural

## DESVENTAJAS

Se requiere que las interrelaciones estén definidas antes de implementar y poner en explotación el sistema de bases de datos

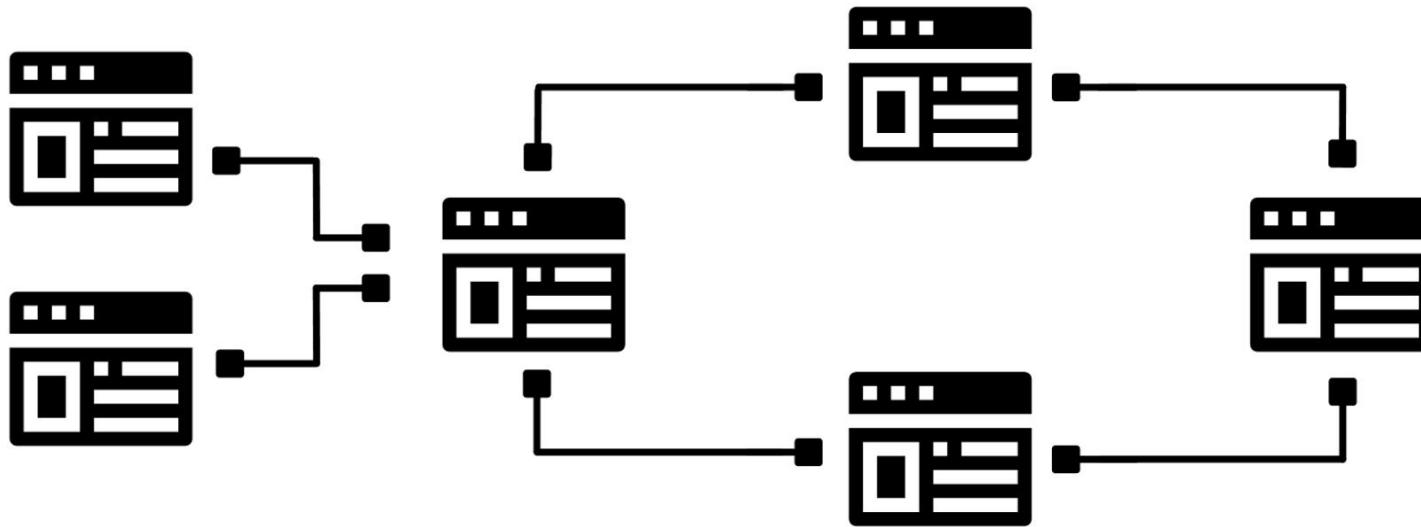
La recuperación de los datos basados en otros vínculos se hace muy difícil

# BASE DE DATOS RELACIONAL



Según C. J. Date (2000), el modelo relacional ha constituido el desarrollo más importante de la historia de las bases de datos

# BASE DE DATOS RELACIONAL



Según J. D. Ullman & J. Widow (2012), el modelo relacional es más que 35 años de historia, es el cimiento de los sistemas de gestión de bases de datos.



## VENTAJAS Y DESVENTAJAS DE LAS Bases de datos relacionales



# Bases de datos relacionales

## VENTAJAS

Tienen como fundamento un modelo de datos muy simple y fácil de comprender, que asegura la integridad de los datos

Admiten la creación de consultas con lenguajes de alto nivel simples, pero muy expresivos

Existen implementaciones extremadamente eficientes del modelo relacional y los lenguajes de consultas respectivos

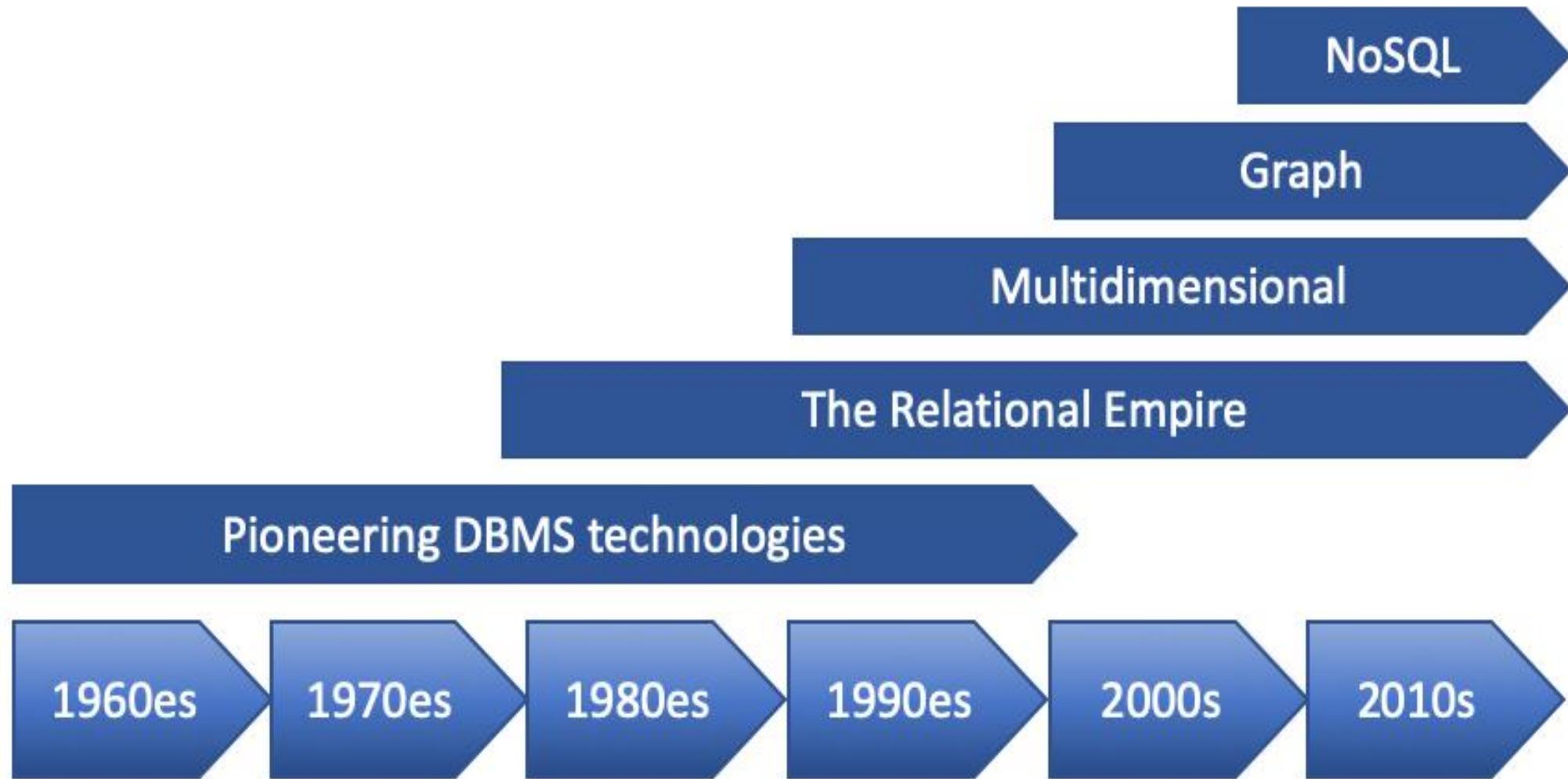
Se usan por los principales sistemas de gestión de bases de datos comerciales actuales

# Bases de datos relacionales

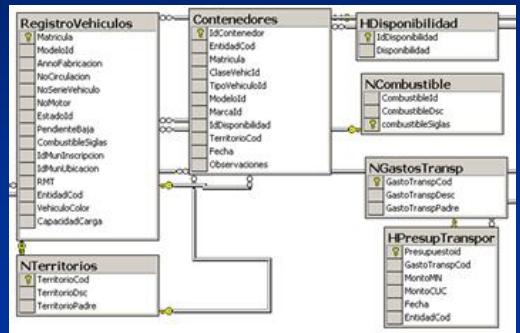
## DESVENTAJAS

- Manejan datos relativamente pequeños y uniformes con
- No admiten atributos semánticamente divisibles
- Se orientan a tuplas, lo que dificulta la segmentación
- Instrumentan consultas relativamente simples
- Soportan transacciones cortas y sin interacción humana
- Se basan en esquemas conceptuales estáticos
- Únicamente asegura las transacciones de tipo ACID

# Evolución de las bases de datos

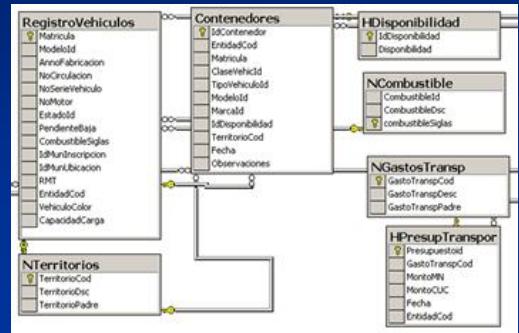


# Representar el contexto en función de los objetivos



Asegurar el  
funcionamiento  
cotidiano de los  
procesos operativos

# Representar el contexto en función de los objetivos



Asegurar el  
funcionamiento  
cotidiano de los  
procesos operativos

Examinar el  
comportamiento del  
negocio con vistas a la  
toma de decisiones



# Interrogantes en las empresas

## Preguntas usuales

- ¿Cuántas personas atendimos el último mes en la comunidad?
- ¿Qué marcas de habanos se venden a los países europeos?

# Interrogantes en las empresas

## Otras preguntas

- ¿En qué proporción han disminuido o no las enfermedades cardiovasculares en la comunidad, comparado con el año anterior?
- ¿Cuáles son las tendencias en las ventas de nuestros productos distintivos?
- Encontrar patrones de correlación entre las agencias de venta de habanos en Asia y Europa. Extraer similitudes y diferencias.

# ¿Son suficientes las bases de datos relacionales?

## PREMISAS

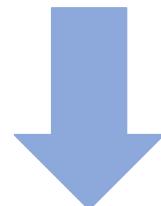
- 
- Los directivos necesitan una vista amplia de la organización que refleje el comportamiento del negocio.
  - Los especialistas en computación requieren mejores herramientas para manejar los datos de la organización en su conjunto desde diferentes perspectivas de análisis.

# ¿Son suficientes las bases de datos relacionales?

## PREMISAS

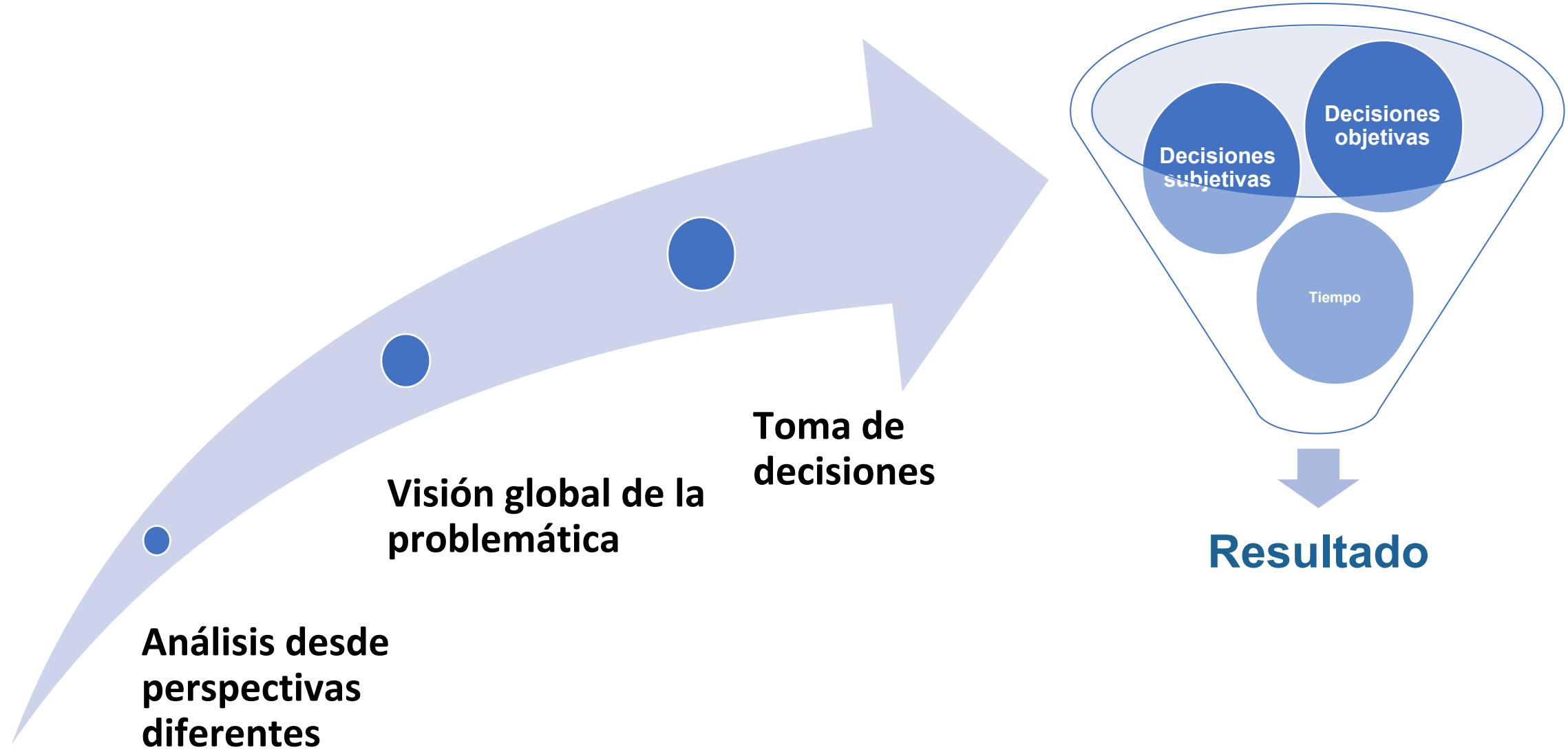
Los directivos necesitan una vista amplia de la organización que refleje el comportamiento del negocio.

Los especialistas en computación requieren mejores herramientas para manejar los datos de la organización en su conjunto desde diferentes perspectivas de análisis.

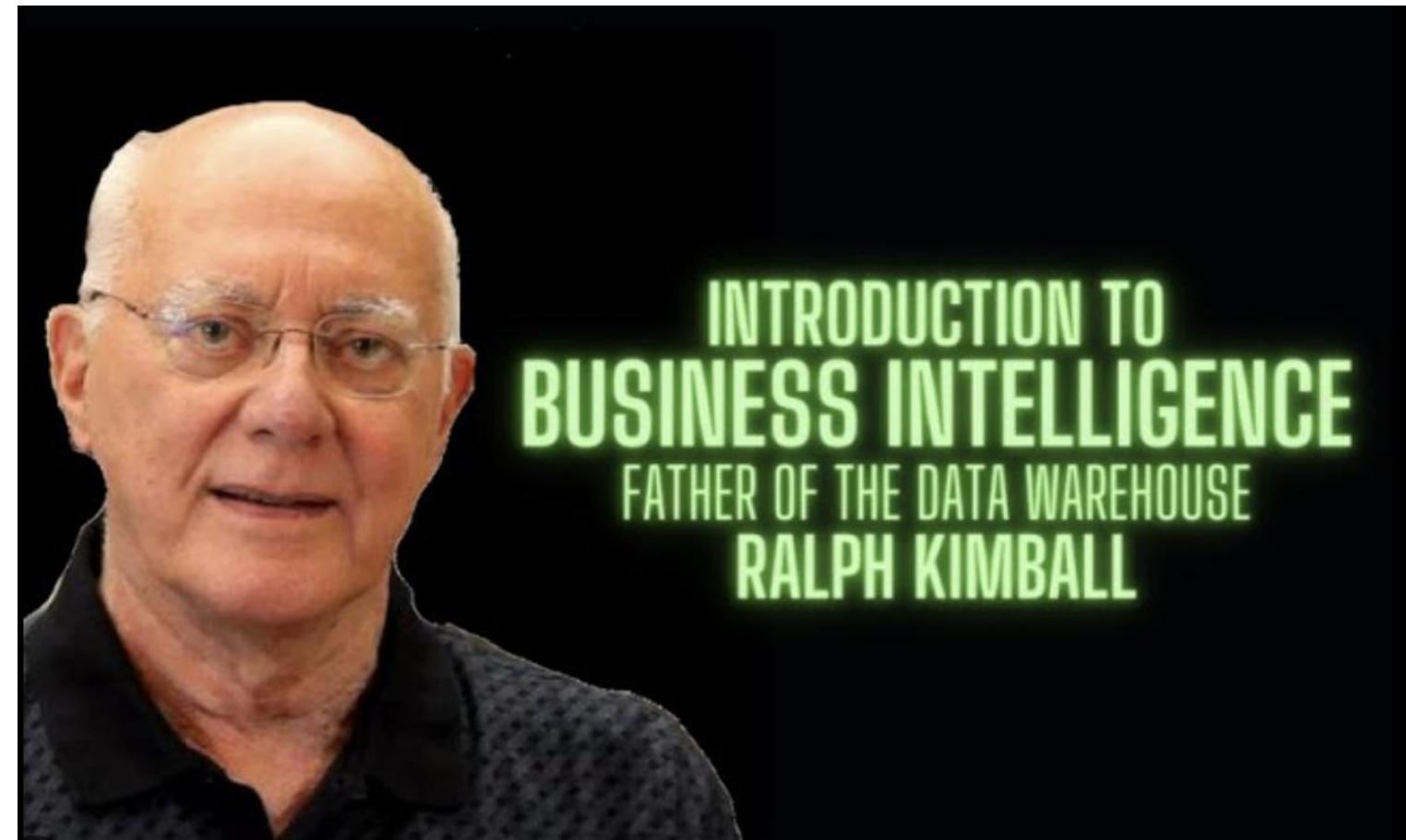


## Modelo Multidimensional de Datos

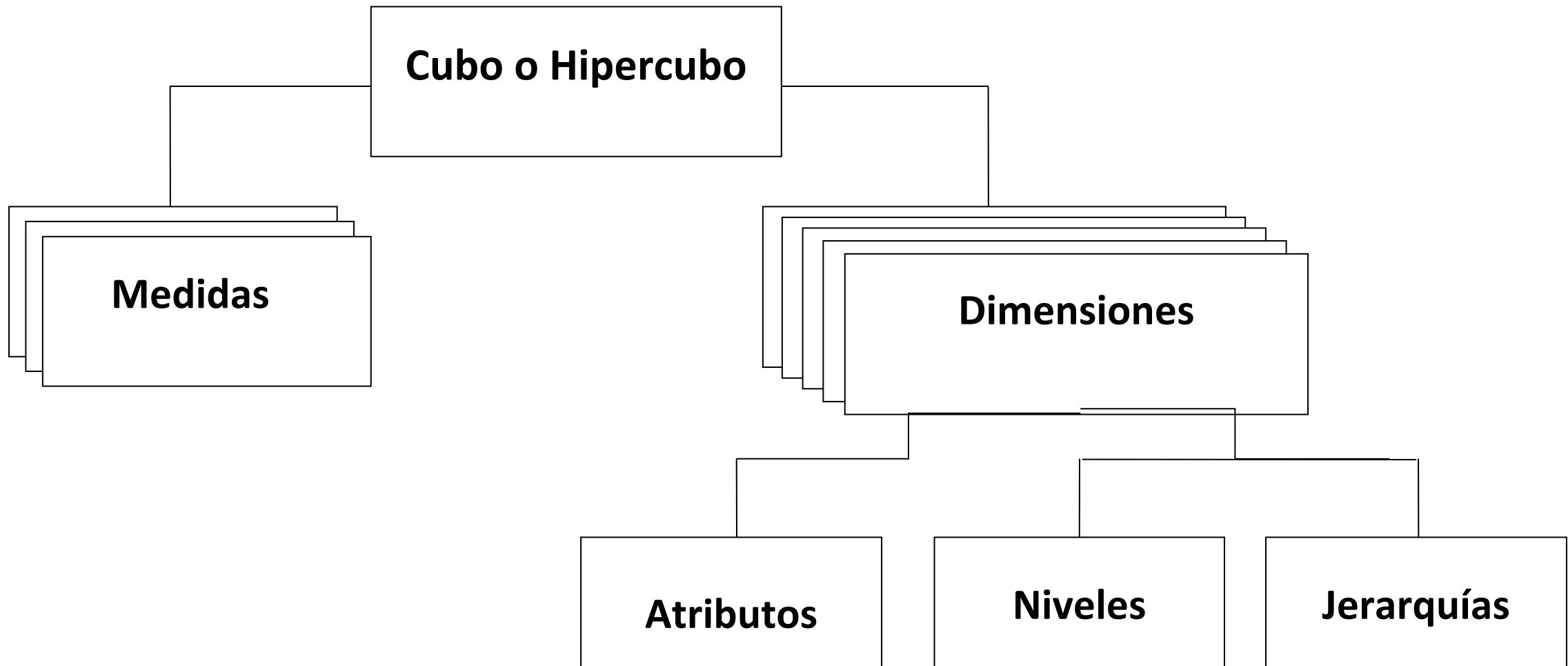
# Enfoque multidimensional



**“The ability to visualize something as abstract as a set of data in a concrete and tangible way is the secret of understandability.”**



# Modelo Multidimensional de Datos (MMD)



# EJEMPLO DE CUBO O HIPERCUBO

## DIMENSIONES

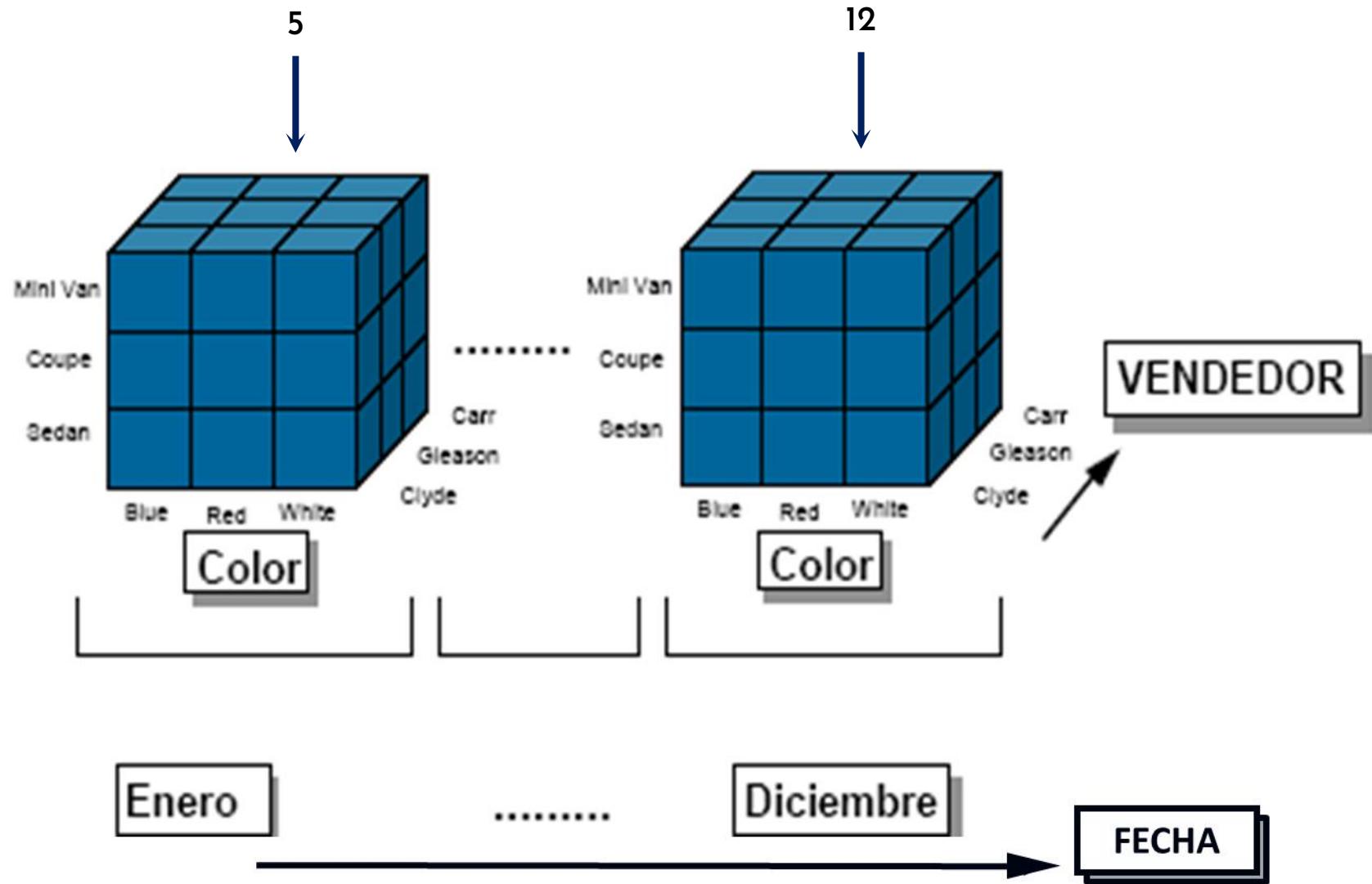
Modelo

Color

Vendedor

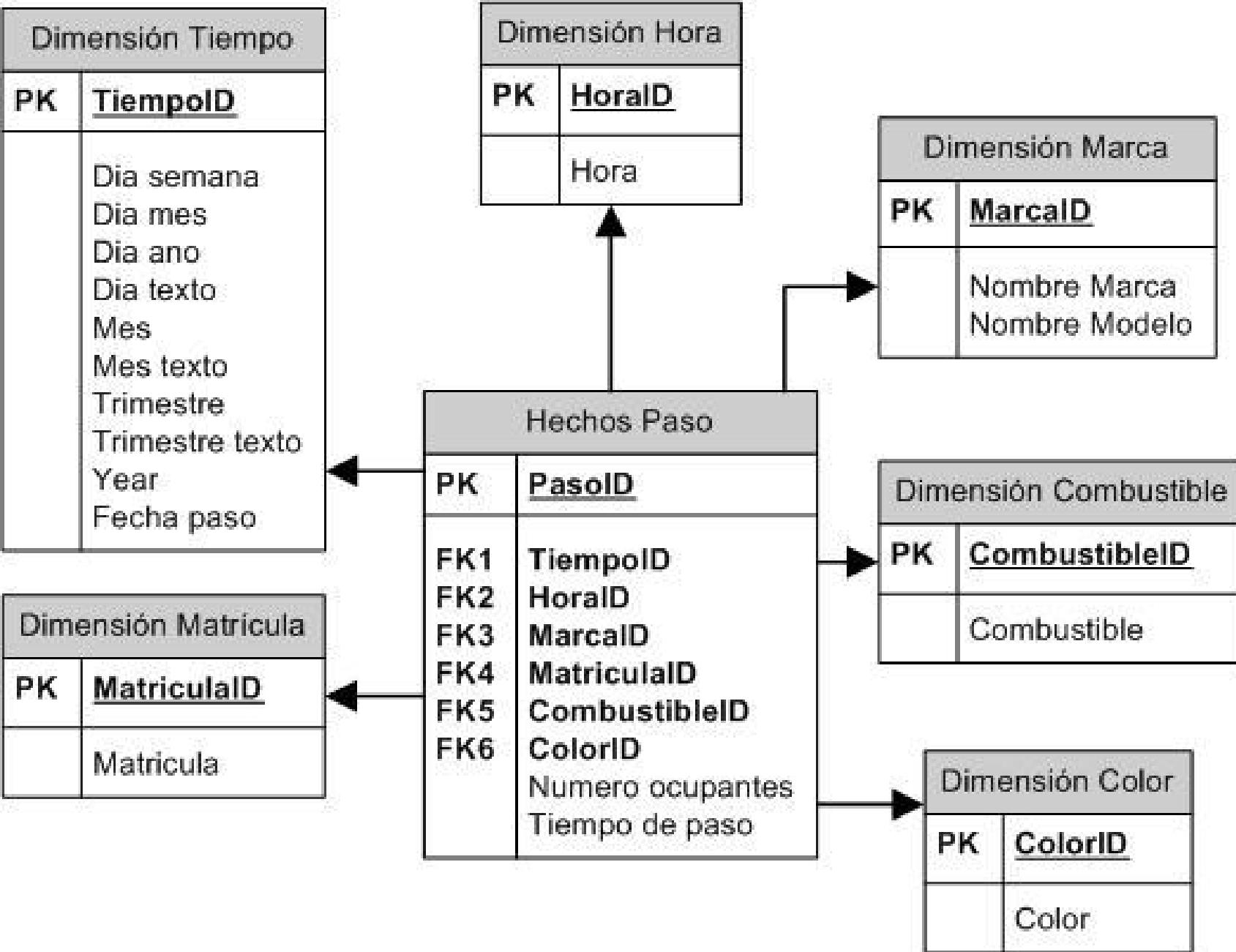
Fecha

M  
O  
D  
E  
L  
O





# ESQUEMA DE ESTRELLA



# Restricciones para el modelo multidimensional

## Granularidad

Expresa el alcance de la medición, o sea, el nivel de detalle de los hechos recogidos.

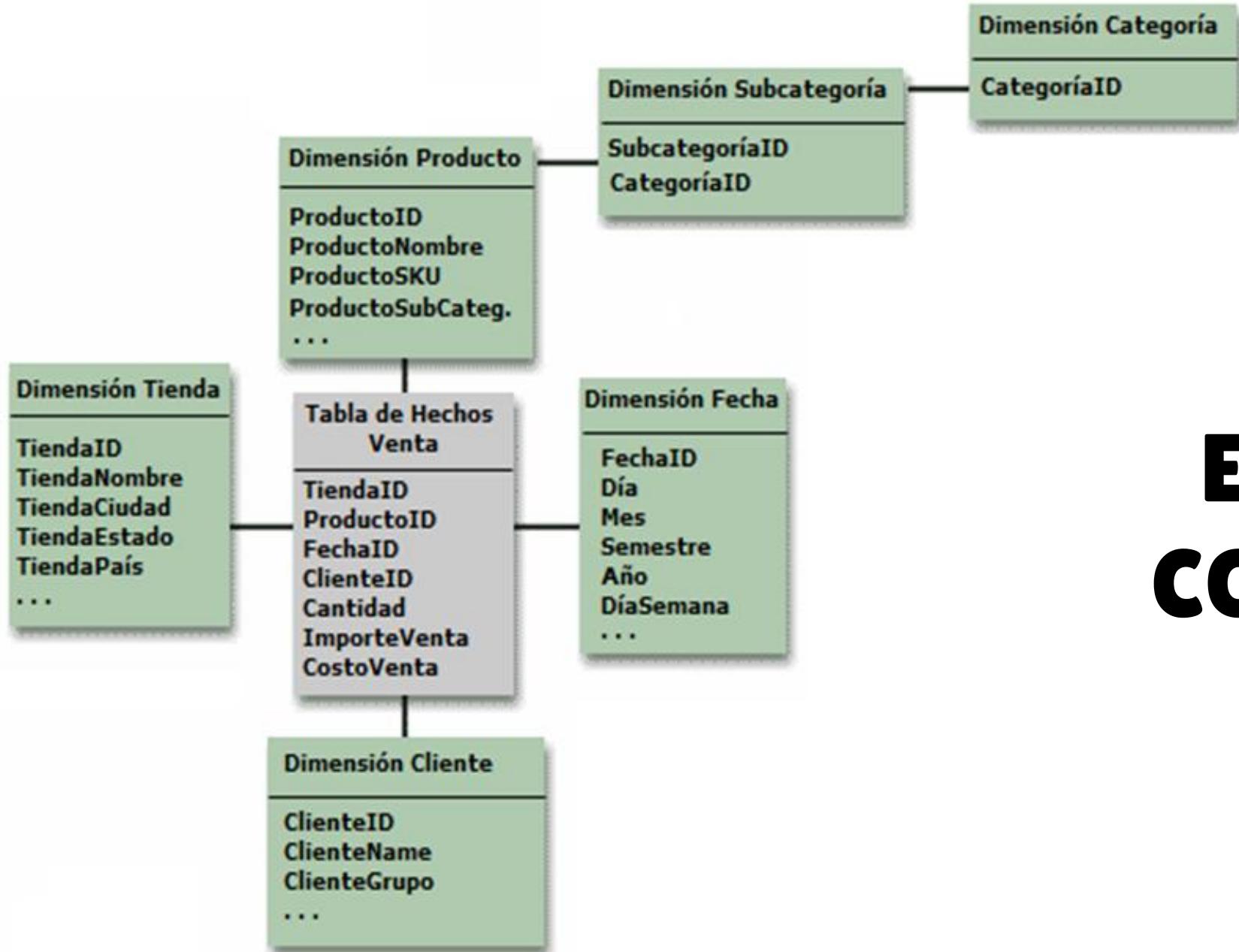
Declara qué es lo que exactamente representa una fila de la tabla de hechos.

## Aditividad

Asocia una función de agregación  $F_a$  a una medida a través de las dimensiones.

$F_a$ : SUM, AVG, MIN, ...

Posibilita la implementación del proceso de aumento (disminución) de detalle a lo largo de las jerarquías de las dimensiones.



# ESQUEMA DE COPO DE NIEVE

# Operaciones del modelo multidimensional (OLAP)

**For ALL Stores and Dates**

Product	Sales in USD
Soda	2,530
Milk	3,858
Juice	15,396
Total	21,784

Figure 4-8 Slice for product

**SLICE**

**DATE 1/1/2005 1/2/2005 1/3/2005 Total**

Metrics	Sales in	Sales in	Sales in	Sales in
	USD	USD	USD	USD
<b>STORE</b>				
CA	40	50	90	180
OR	3,115	3,340	1,267	7,722
LA	1,583	7,418	4,881	13,882
Total	4,738	10,808	6,238	21,784

Figure 4-9 Dice for store and date

**DICE**

**STORE CA OR LA Total**

Metrics	Sales in	Sales in	Sales in	Sales in
	USD	USD	USD	USD
Group-Class Beverage	180	7,722	13,882	21,784
Total	180	7,722	13,882	21,784

**Beverage -> Group**

STORE	CA	OR	LA	Total
Metrics	Sales in	Sales in	Sales in	Sales in
	USD	USD	USD	USD
Group	140	7,662	11,452	19,254
Pop	40	60	2,430	2,530
Water	60	1,452	9,106	15,396
Total	180	7,722	13,882	21,784

**Beverage -> Group -> Product**

STORE	CA	OR	LA	Total
Metrics	Sales in	Sales in	Sales in	Sales in
	USD	USD	USD	USD
PRODUCT	60	1,452	2,346	3,858
Milk	60	6,210	9,106	15,396
Juice	10	10	30	60
Total	40	50	90	180

Figure 4-12 Drill-down on product dimension

**DRILL-UP/DOWN**

**DATE 1/1/2005 1/2/2005 1/3/2005 Total**

Metrics	Sales in	Sales in	Sales in	Sales in
	USD	USD	USD	USD
<b>STORE</b>				
CA	40	50	90	180
OR	3,115	3,340	1,267	7,722
LA	1,583	7,418	4,881	13,882
Total	4,738	10,808	6,238	21,784

**CA -> Product**

Metrics	Sales in	Sales in	Sales in	Sales in
	USD	USD	USD	USD
PRODUCT	10	10	20	40
Milk	20	10	30	60
Juice	10	30	40	80
Total	40	50	90	180

Figure 4-14 Drill-across result

**DRILL-ACROSS**

**PRODUCT Milk Coke Juice Total**

Metrics	Sales in	Sales in	Sales in	Sales in
	USD	USD	USD	USD
<b>STORE</b>				
CA	40	60	80	180
OR	60	1,452	6,210	7,722
LA	2,430	2,346	9,106	13,882
Total	2,530	3,858	15,396	21,784

**Pivot**

Metrics	Sales in	Sales in	Sales in	Sales in
	USD	USD	USD	USD
PRODUCT	40	60	2,430	2,530
Milk	60	1,452	2,346	3,858
Juice	10	30	40	80
Total	180	7,722	13,882	21,784

Figure 4-11 Pivoting

**PIVOT**

**DATE 1/1/2005 1/2/2005 1/3/2005 Total**

Metrics	Sales in	Sales in	Sales in	Sales in
	USD	USD	USD	USD
<b>Group Class Group PRODUCT</b>				
Beverage	4,738	10,808	6,238	21,784
Total	4,738	10,808	6,238	21,784

**DATE 1/1/2005 1/2/2005 1/3/2005 Total**

Metrics	Sales in	Sales in	Sales in	Sales in
	USD	USD	USD	USD
<b>Group Class Group PRODUCT</b>				
Beverage	3,721	9,880	6,653	19,254
Pop	1,017	928	585	2,530
Water	4,738	10,808	6,238	21,784
Total	4,738	10,808	6,238	21,784

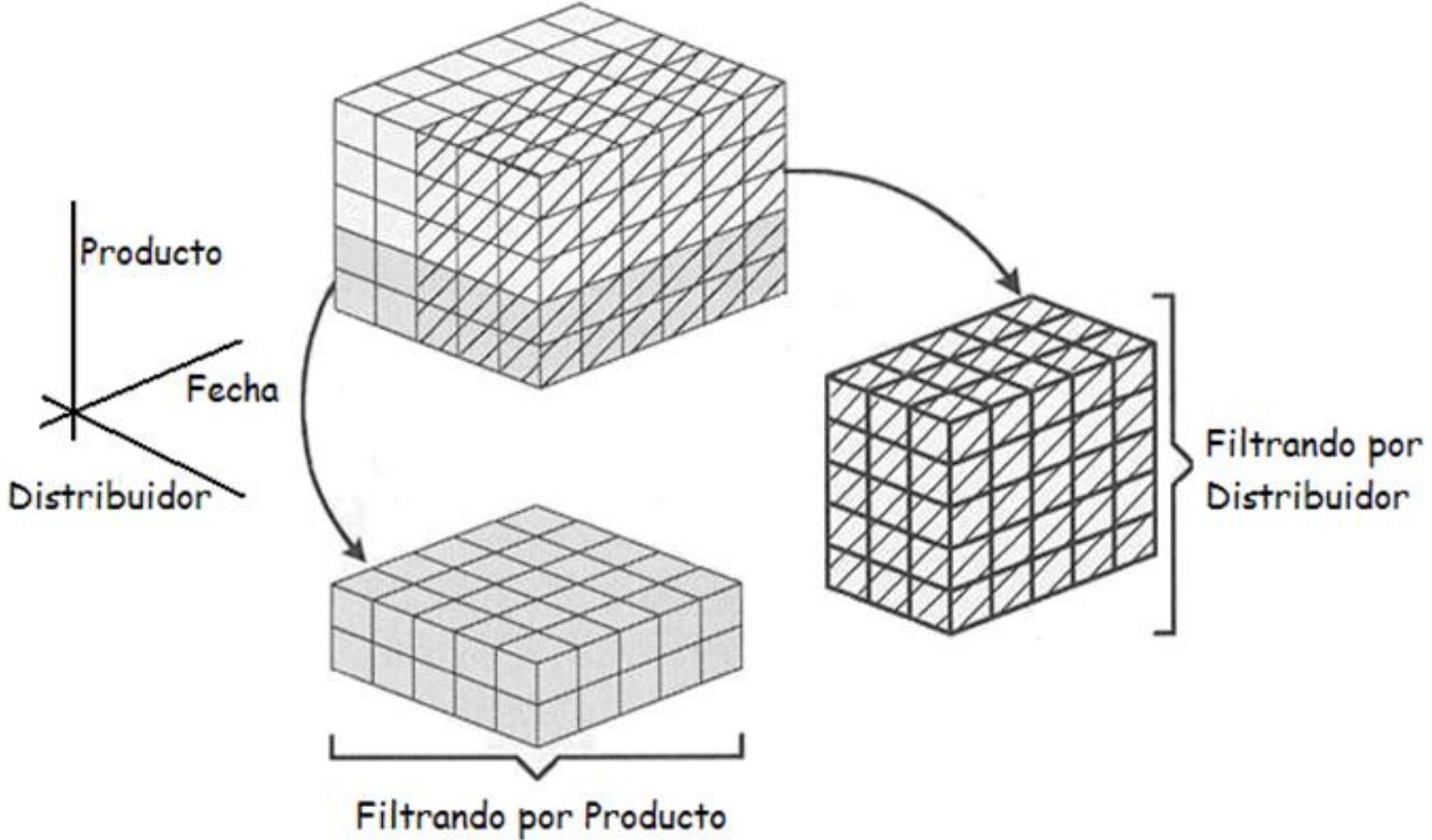
  

**DATE 1/1/2005 1/2/2005 1/3/2005 Total**

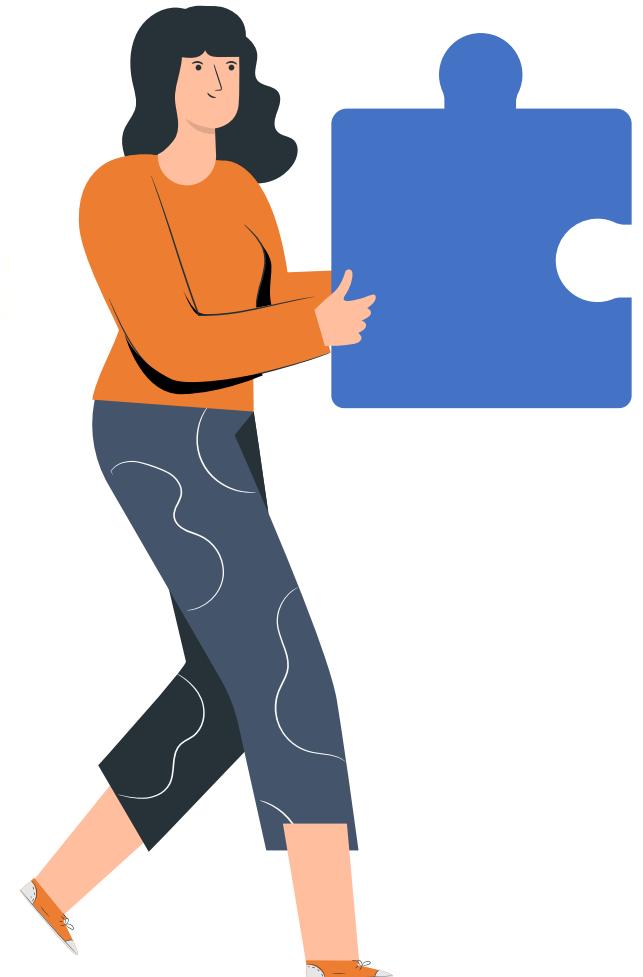
Metrics	Sales in	Sales in	Sales in	Sales in
	USD	USD	USD	USD
<b>Group Class Group PRODUCT</b>				
Beverage	1,141	1,431	1,286	3,858
Milk	2,580	8,449	4,367	15,396
Juice	3,721	9,880	5,653	19,254
Total	1,141	1,431	1,286	3,858

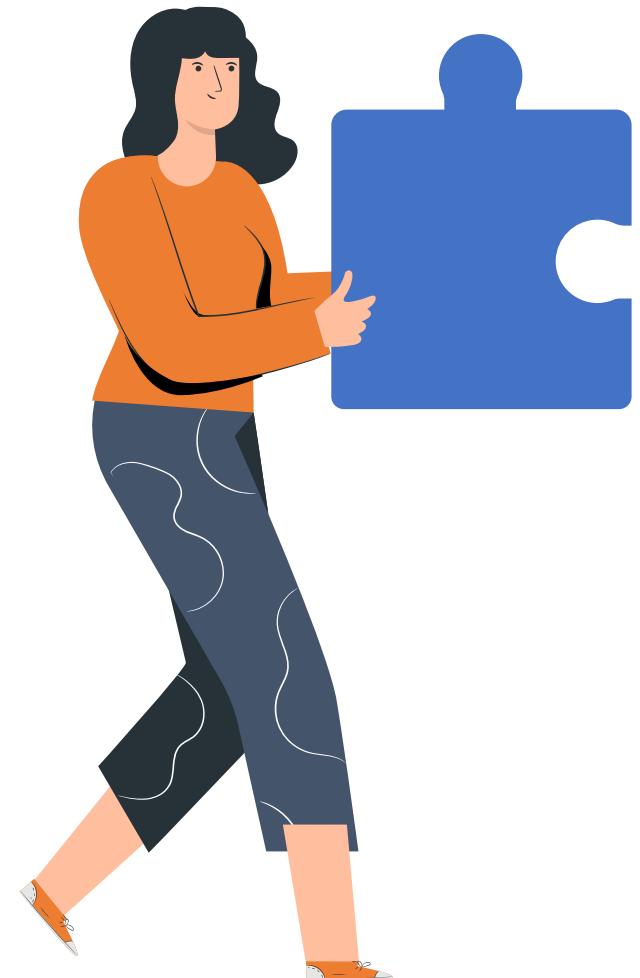
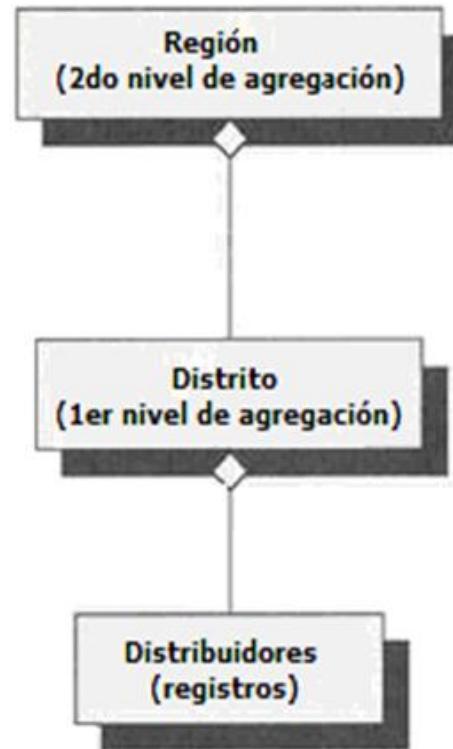
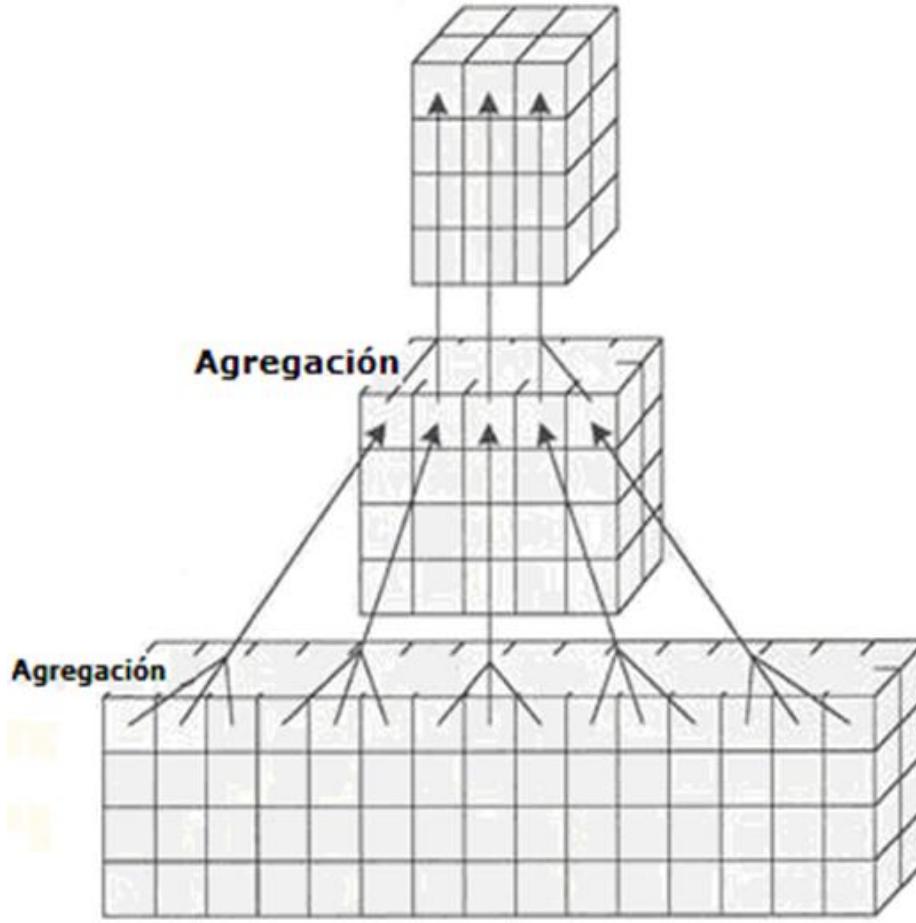
Figure 4-15 Roll-down, Roll-up

**ROLL-UP/DOWN**

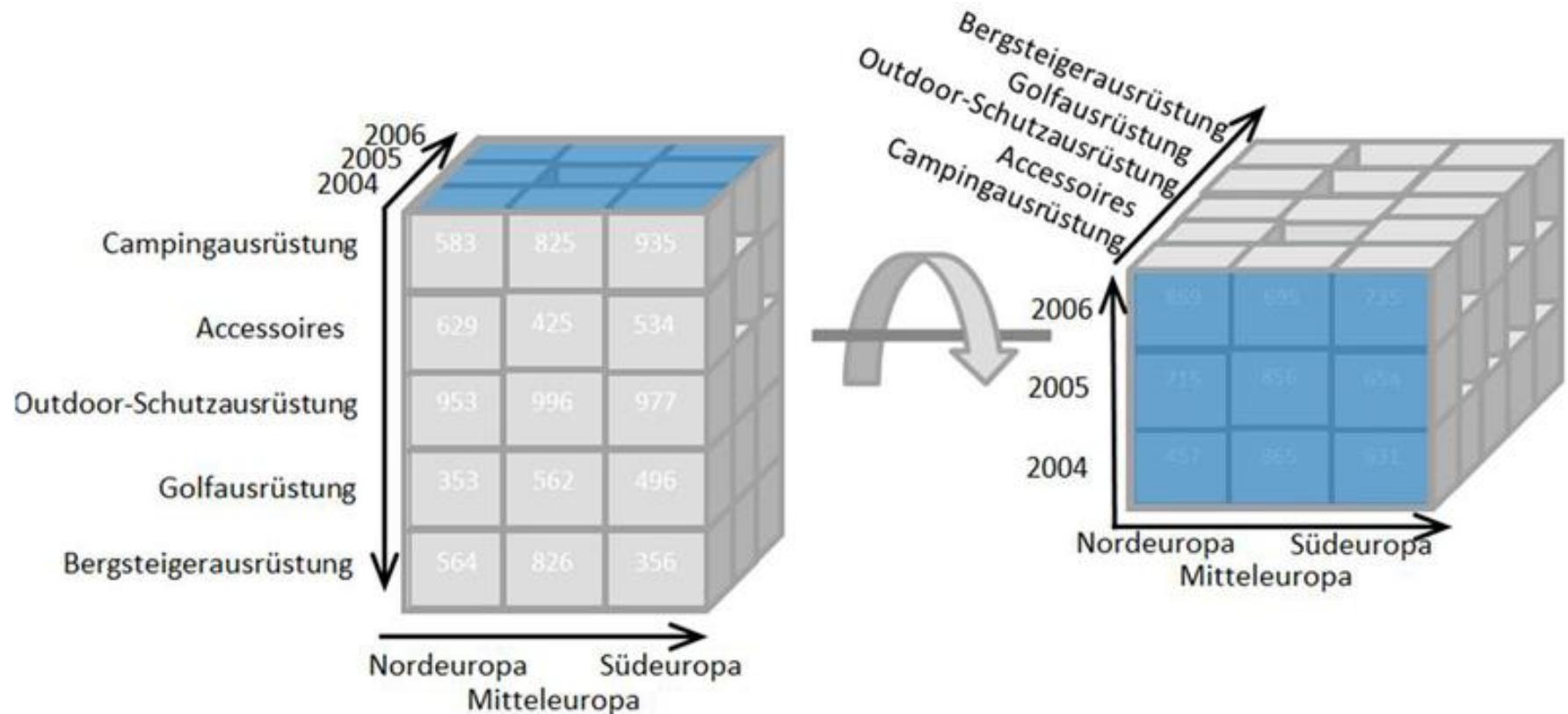


# *Slice & Dice*

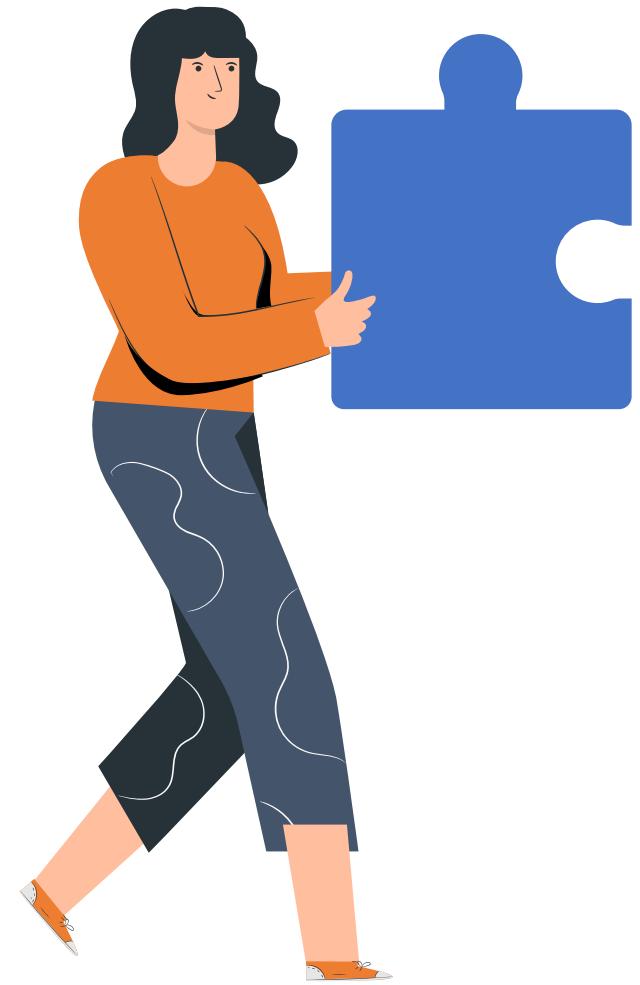




*Drill-Down & Drill-Up*



# Pivot



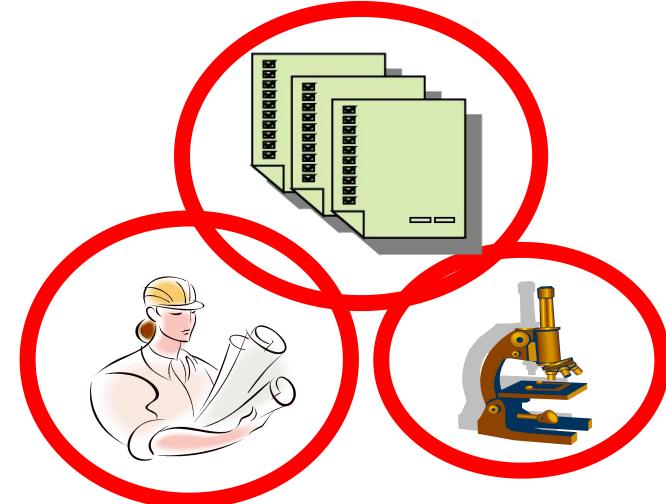
# SISTEMAS DE BASES DE DATOS INFORMATIVOS

**Datos dispersos  
y en diferentes formatos**

**Los especialistas de cada  
área procesan su  
información**

**Es muy difícil realizar análisis  
integrales y trazar estrategias**

**No existe una base  
informativa para los  
datos externos a la  
organización**



Integración de datos dispersos que permitan vistas disímiles

Procesamiento de la información en diferentes niveles de la organización

## Data Warehouse

Realización de análisis integrales y diseño de estrategias

Posibilidad de incorporar al análisis datos externos a la organización

# Definiciones de Data Warehouse (DW)

W. H. Inmon, 1992

"Un *Data Warehouse* es una colección de datos **integrada, orientada a sujetos, variante en el tiempo y no volátil**, utilizada como soporte para los procesos de toma de decisión."

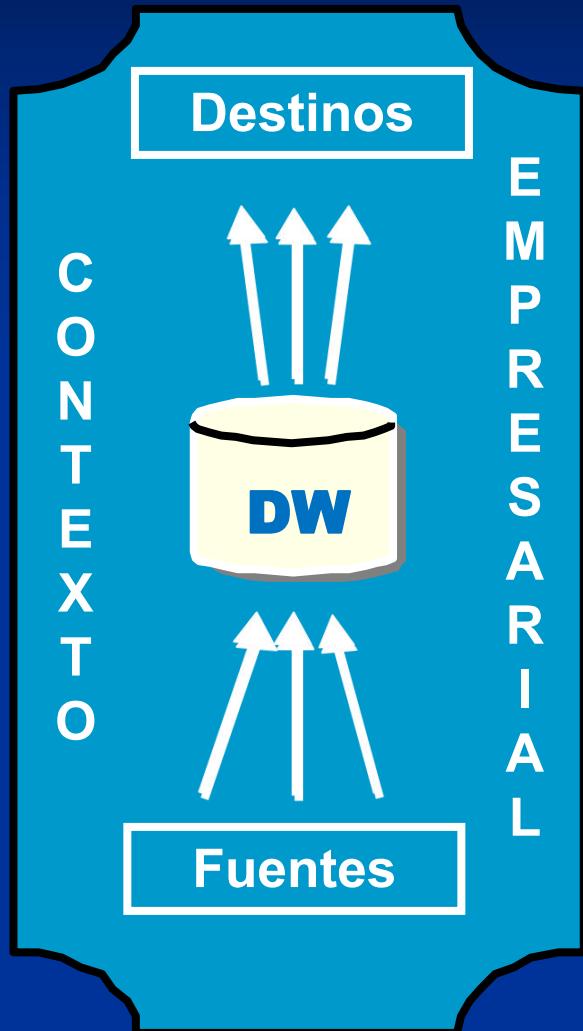
R. Kimball, 1996

"Un *Data Warehouse* es un sistema que **extrae, limpia, conforma y entrega los datos fuentes a un almacén de datos dimensional** que soporta la ejecución de consultas y el análisis con el propósito de la toma de decisiones."

# Claves para el diseño dimensional en un almacén de datos (DW, DM, ODS)

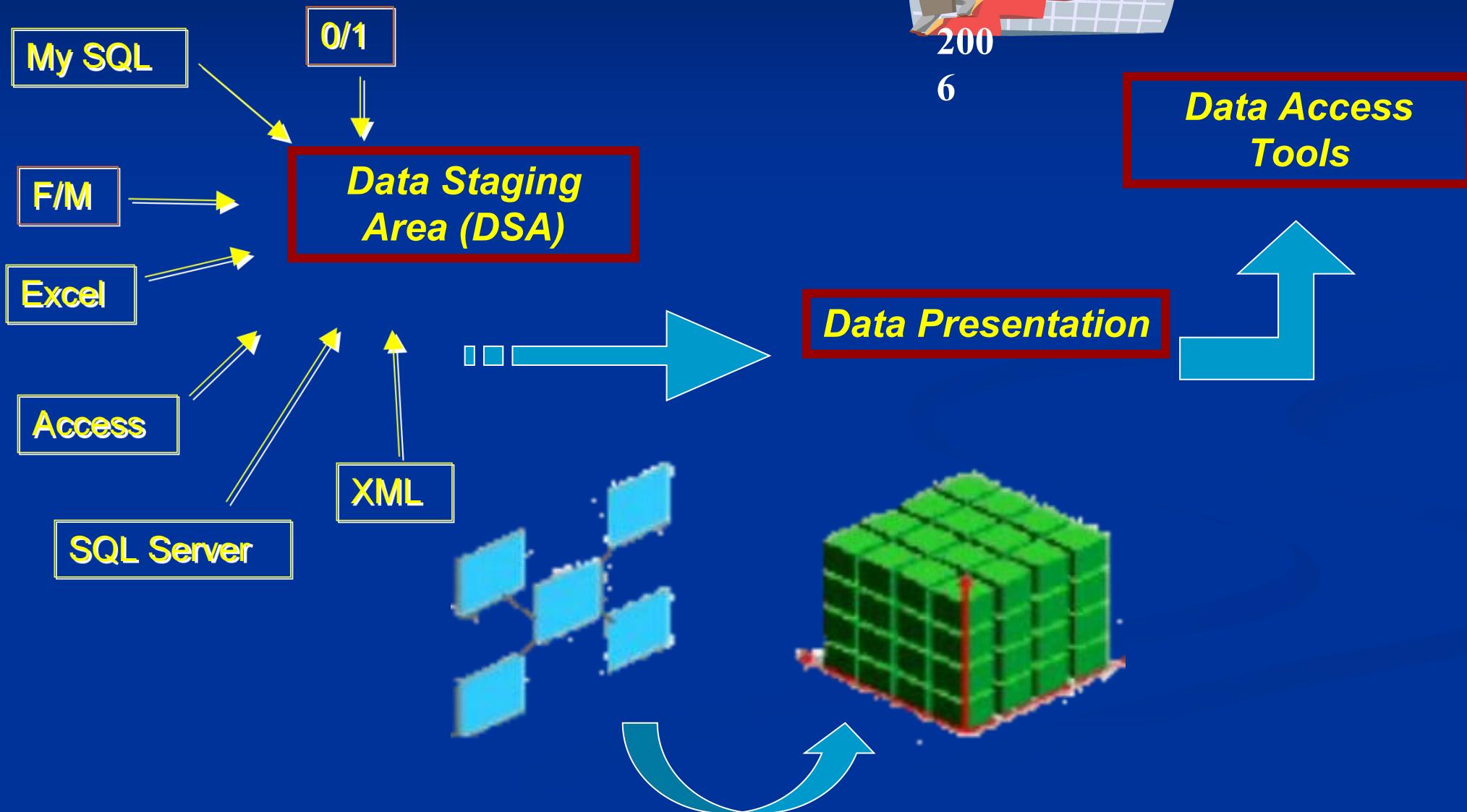


# Caracterización del *Data Warehousing*

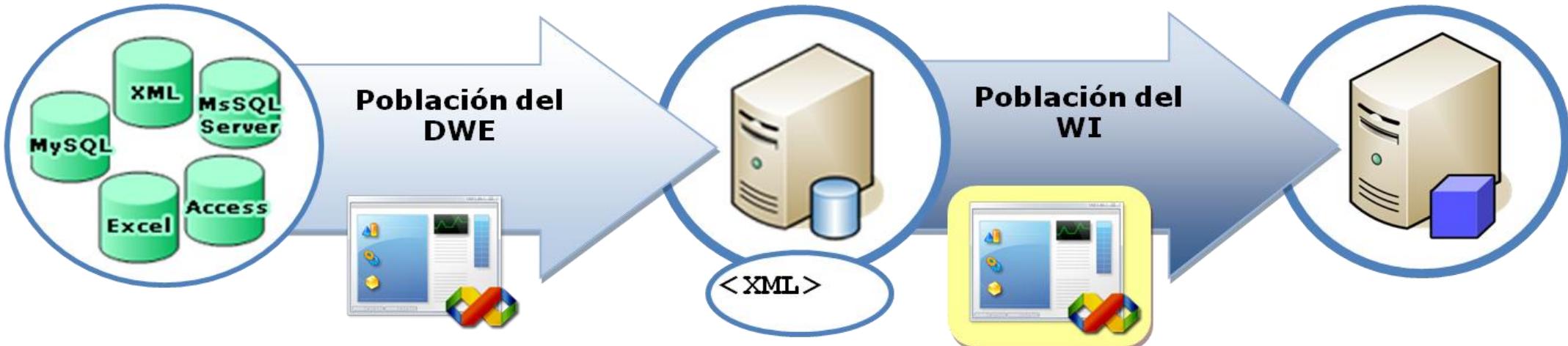


- La información se colecciona desde fuentes diferentes y dispersas y se sitúa en un lugar donde **se concilian las desigualdades**.
- Una vez situados los datos en un repositorio único, se permite que distintas aplicaciones obtengan **la información derivada** a partir de ellos.
- El sistema se utiliza en el contexto global de la organización.

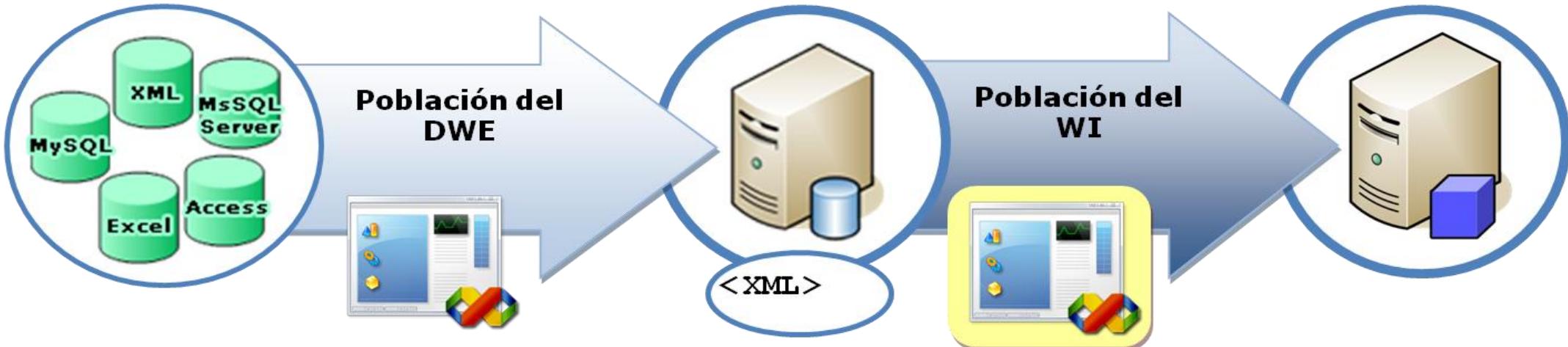
# Proceso de *data warehousing*



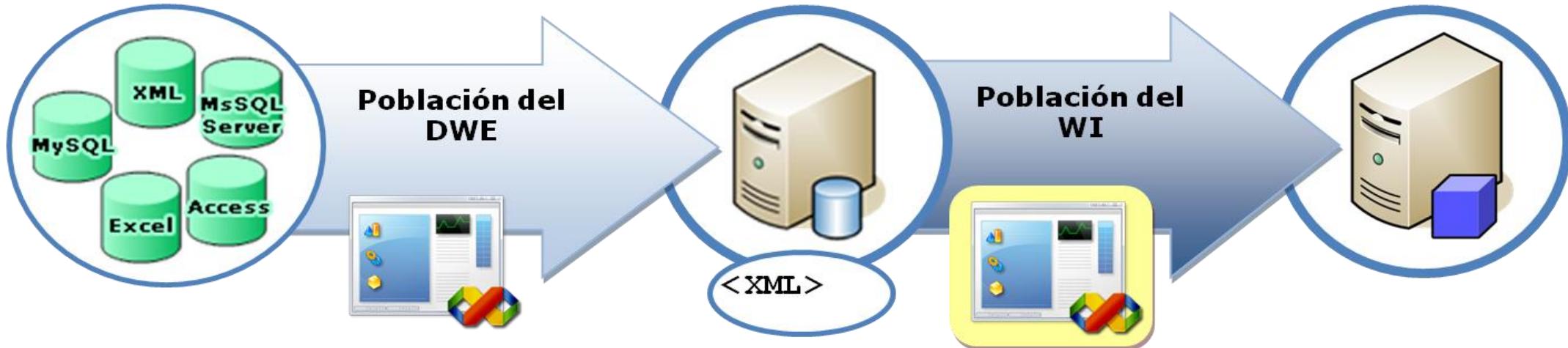
# Población del Data Warehouse: Procesos ETCL



# Población del Data Warehouse: Procesos ETCL



# Población del Data Warehouse: Procesos ETCL



**La parte más compleja en la construcción de un data warehouse consiste en la **carga** de los datos**

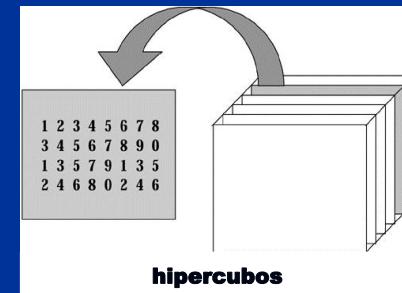
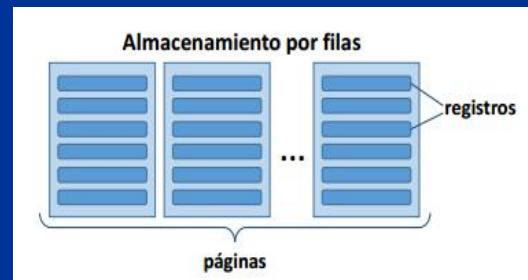
**Es imposible evaluar el diseño de un data warehouse hasta que los datos legados hayan sido cargados y visualizados**

# Alternativas de almacenamiento del *warehouse* informacional



## Modo Multidimensional

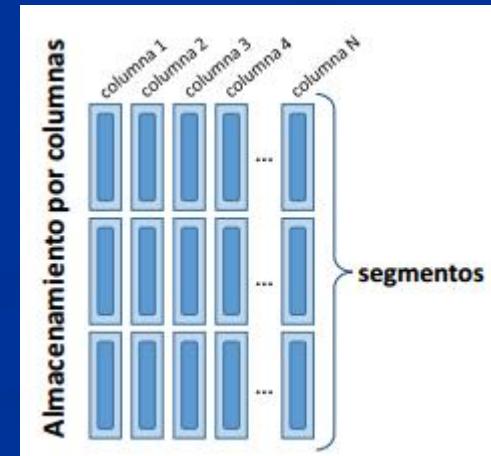
Almacenamiento por filas en hipercubos, creados en arrays multidimensionales de páginas



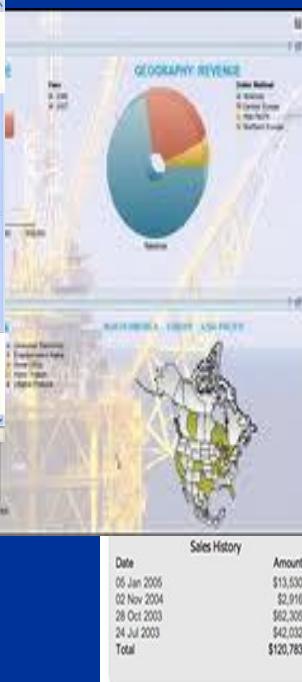
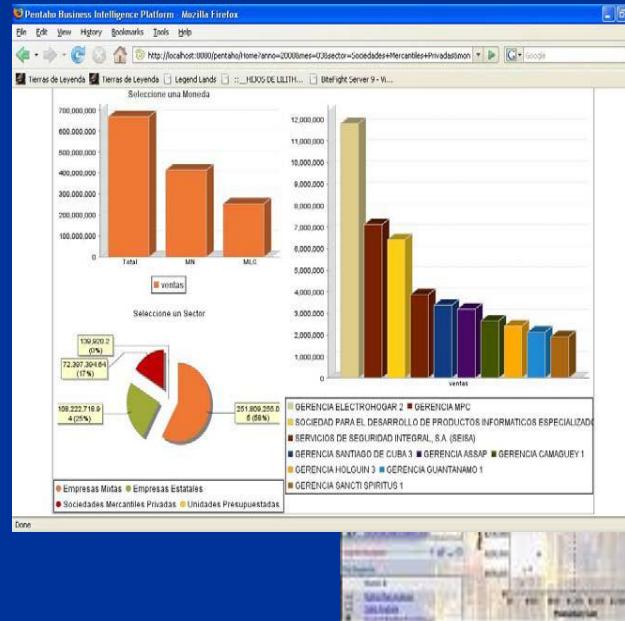
## Modo Tabular

Almacenamiento en memoria por columnas

Manejo de páginas de segmentos de una sola columna con valores repetidos



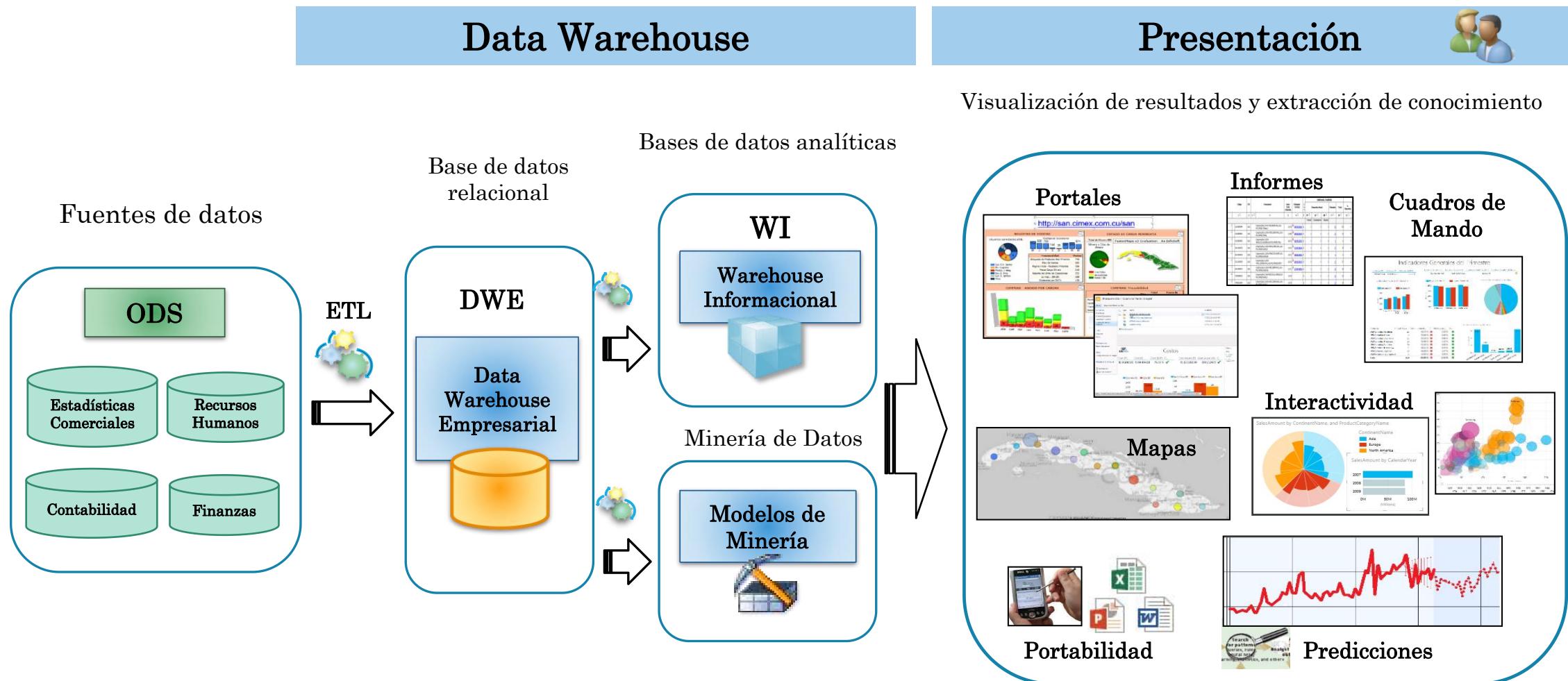
# Inteligencia de Negocios (BI – *Business Intelligence*)



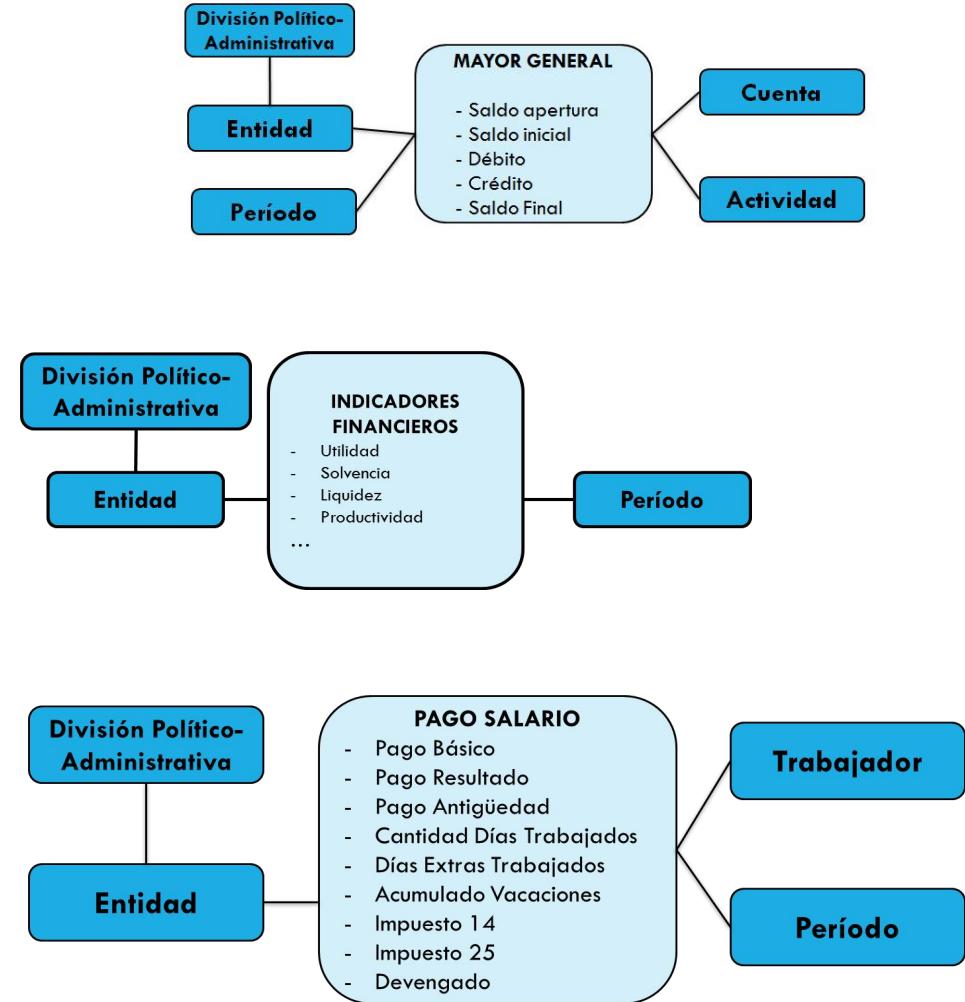
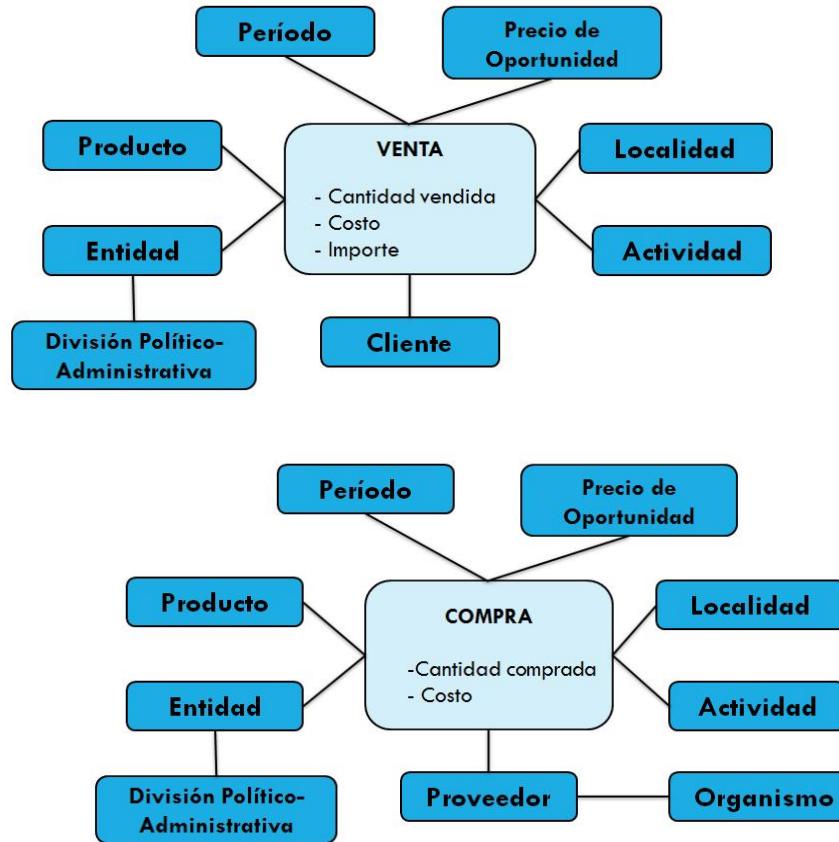


# Plataforma interactiva para el análisis del desempeño empresarial en CIMEX

# Concepción general de la solución



# Diseño de modelos analíticos



- Dimensiones compartidas
- Jerarquías en las dimensiones
- Granularidad mensual

## Resultados del experimento 2 | Consulta1

Obtener el cumplimiento mensual de los principales indicadores económicos para una entidad dada en los años 2014-2015

	A	B	C	D	E
1	Código	CCV12			
2	Año	(Multiple Items)			
3					
4	Row Labels	Ingresos % (R/P)	Status	Utilidades % (R/P)	Status
5	Enero 2014	100,27 %	🟡	194,37 %	🟢
6	Febrero 2014	90,94 %	🔴	108,90 %	🟡
7	Marzo 2014	76,00 %	🔴	121,76 %	🟢
8	Abril 2014	83,36 %	🔴	150,97 %	🟢
9	Mayo 2014	71,23 %	🔴	78,92 %	🔴
10	Junio 2014	101,91 %	🟡	205,70 %	🟢
11	Julio 2014	100,45 %	🟡	160,88 %	🟢
12	Agosto 2014	100,13 %	🟡	221,26 %	🟢
13	Septiembre 2014	100,41 %	🟡	205,96 %	🟢
14	Octubre 2014	84,60 %	🔴	77,44 %	🔴
15	Noviembre 2014	100,44 %	🟡	1,44 %	🔴
16	Diciembre 2014	112,82 %	🟢	74,14 %	🔴
17	Enero 2015	100,16 %	🟡	145,67 %	🟢
18	Febrero 2015	100,13 %	🟡	127,57 %	🟢
19	Marzo 2015	100,33 %	🟡	88,25 %	🔴
20	<b>Grand Total</b>	<b>94,73 %</b>	<b>🔴</b>	<b>130,21 %</b>	<b>🟢</b>
21					

Visualización de los KPI

PivotTable Fields

Show fields: (All)

Drag fields between areas below:

**FILTERS**

- Código
- Año

**ROWS**

- Mes

**COLUMNS**

- Σ Values

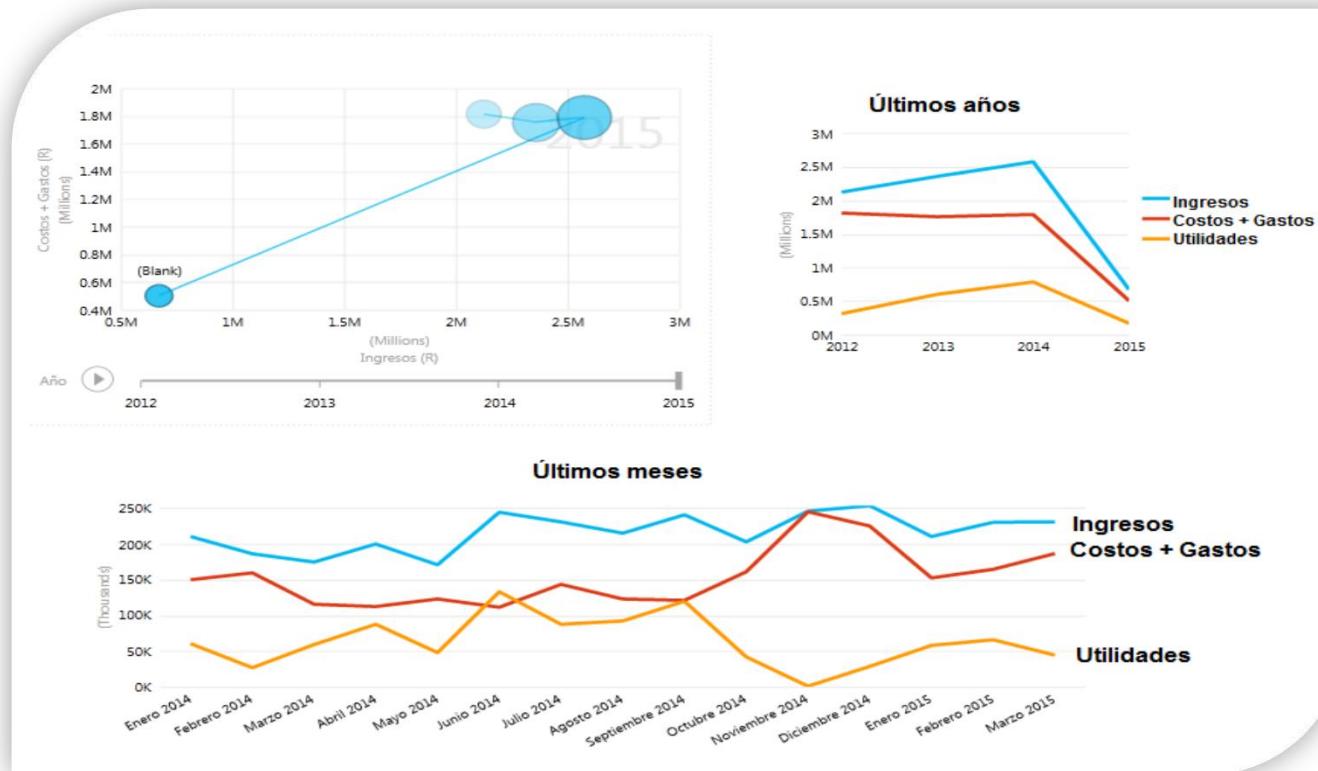
**VALUES**

- Ingresos Totales (P)
- Ingresos Totales (R)
- Ingresos Totales ...

Defer Layo...

## Resultados del experimento 2 | Consulta 2

Resumir el comportamiento histórico de los principales indicadores económicos para una entidad dada



Visualización del comportamiento histórico

## Resultados del experimento 2 | Consulta 3

Analizar la cantidad de altas laborales por provincias en el 2015 en la Corporación

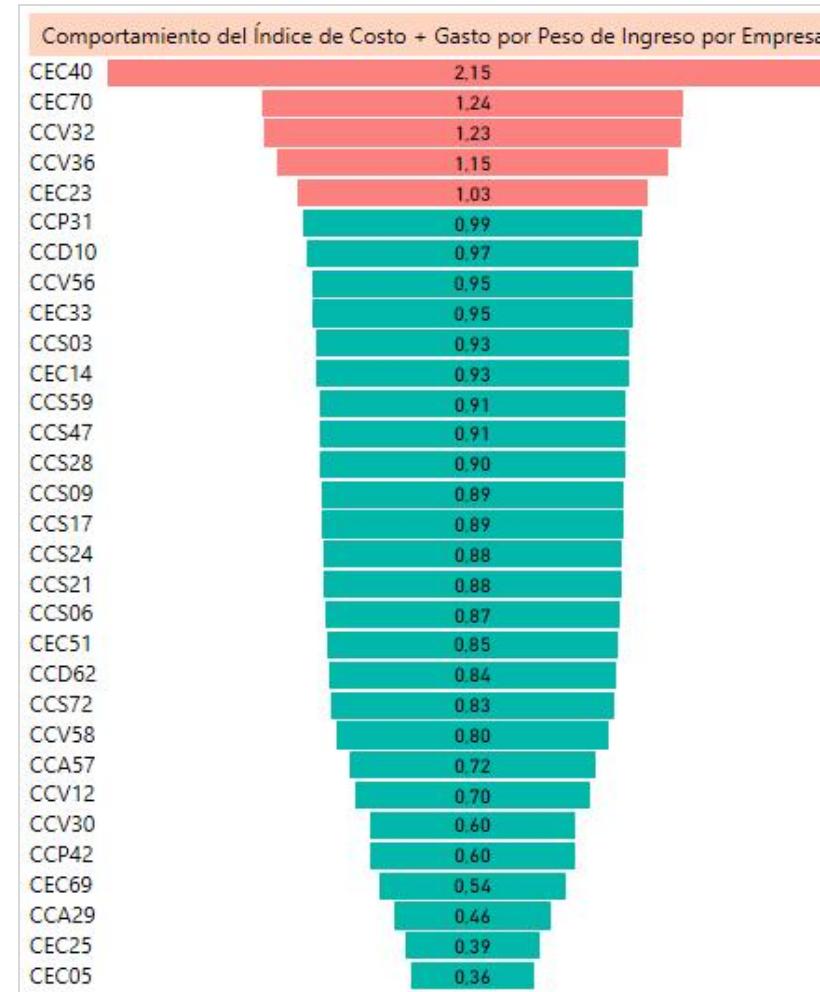


Visualización en un componente gráfico dinámico

## Resultados del experimento 2 | Consulta 4

Identificar las entidades más y menos eficientes del sistema empresarial desde la perspectiva financiera en el año 2014

Visualización intuitiva  
en términos de  
eficiencia económica



## Resultados del experimento 3



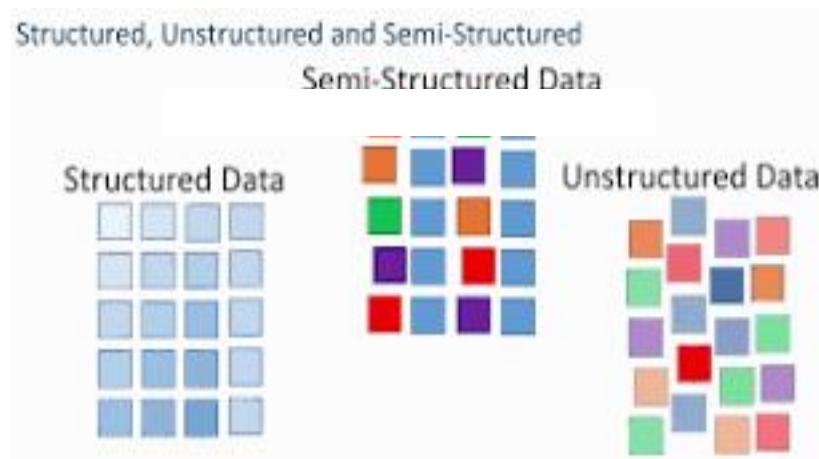
Período	Real	Estimado
01/01/2015	33206193,11	43814761,52
01/02/2015	40505405,07	45266393,80
01/03/2015	38819819,34	47481189,77
01/04/2015	42074889,08	47534231,04
01/05/2015	35177434,45	44381651,74
01/06/2015	41775217,59	40334050,11
01/07/2015	49162165,31	47088115,36
01/08/2015	47049971,99	44922337,74
01/09/2015	42220606,74	43382323,79
01/10/2015	42944970,15	43660189,98

Resultados de la predicción para el indicador *Ingresos*

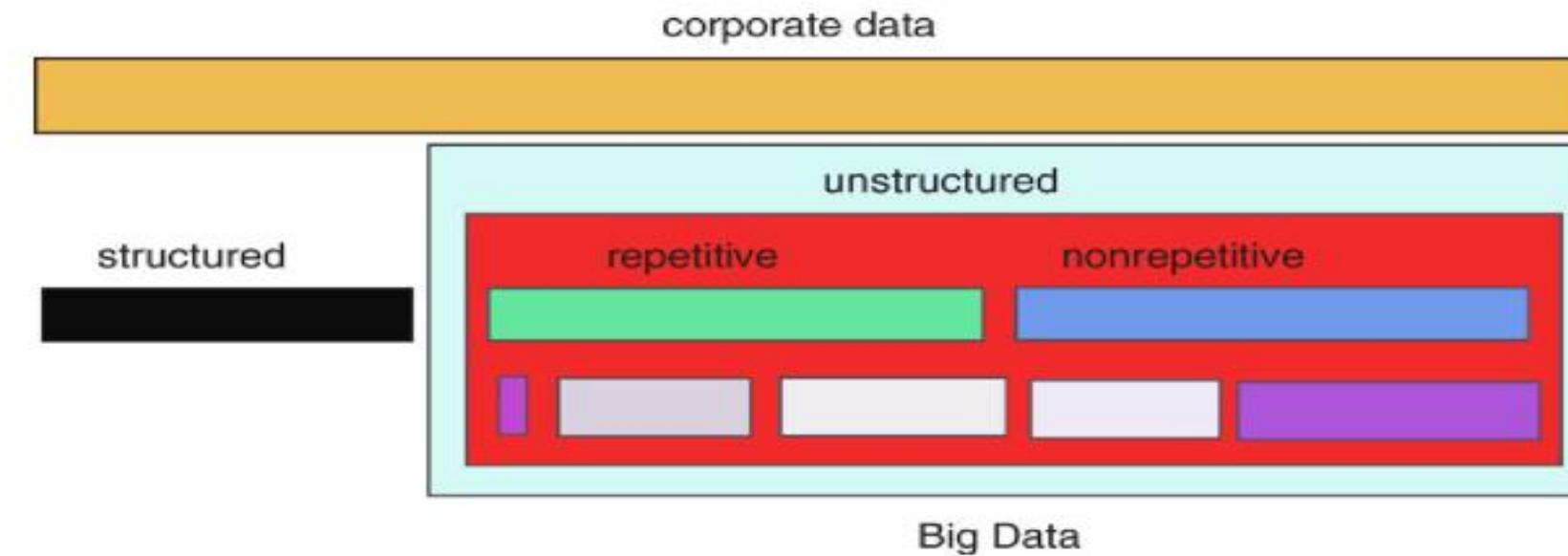
# ¿A qué deben responder las bases de datos ante el contexto actual?



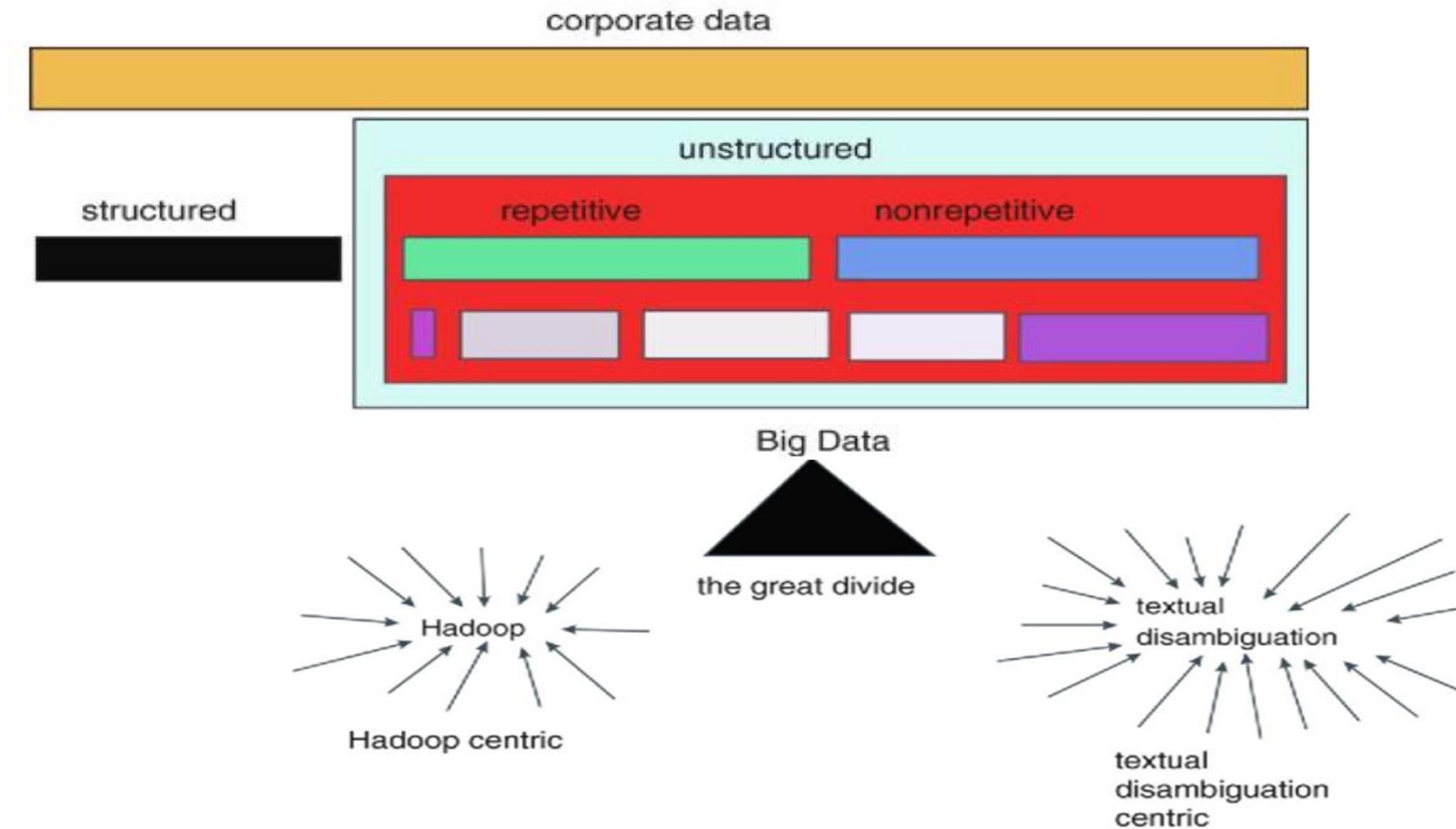
# Panorama vigente



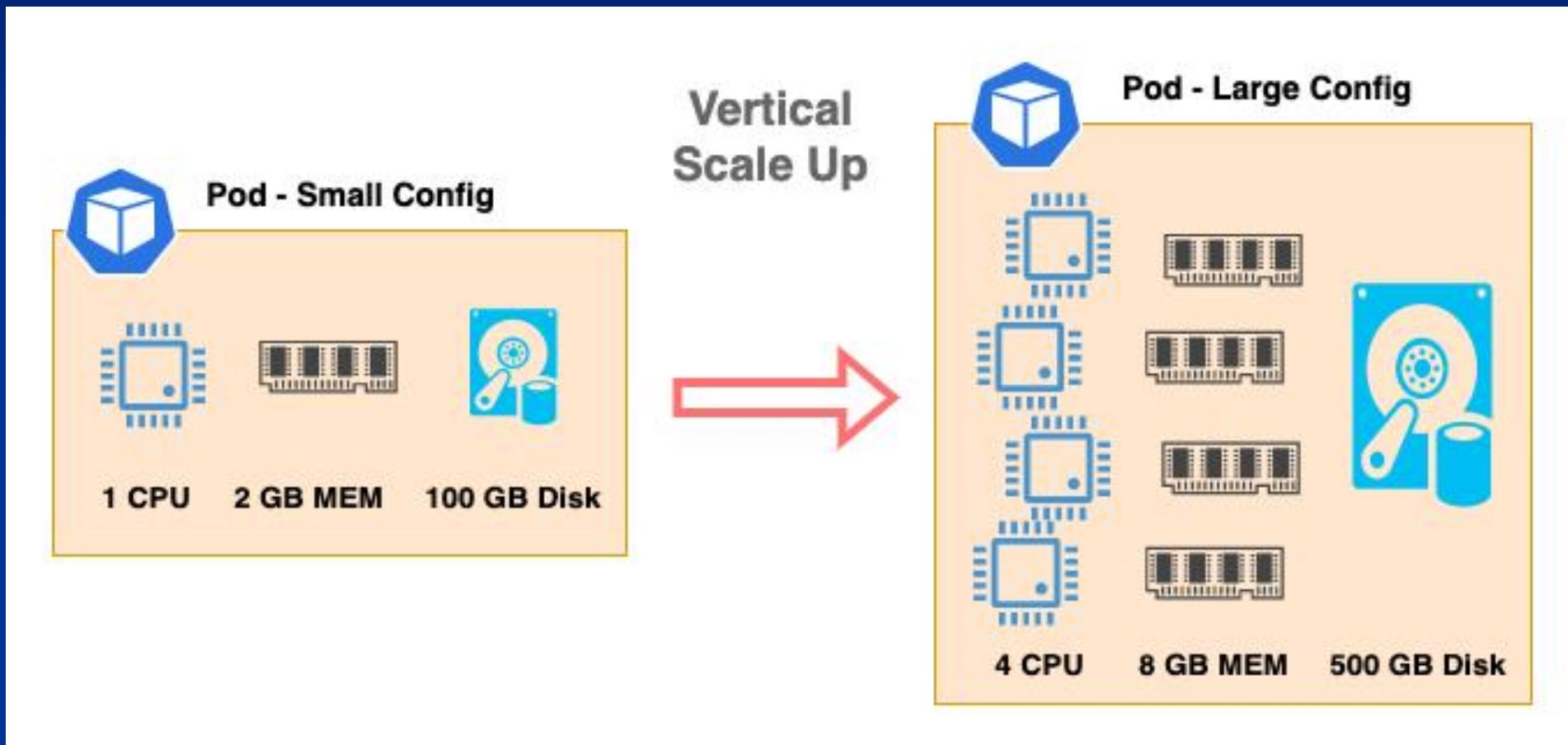
# Imagen de los datos en una organización



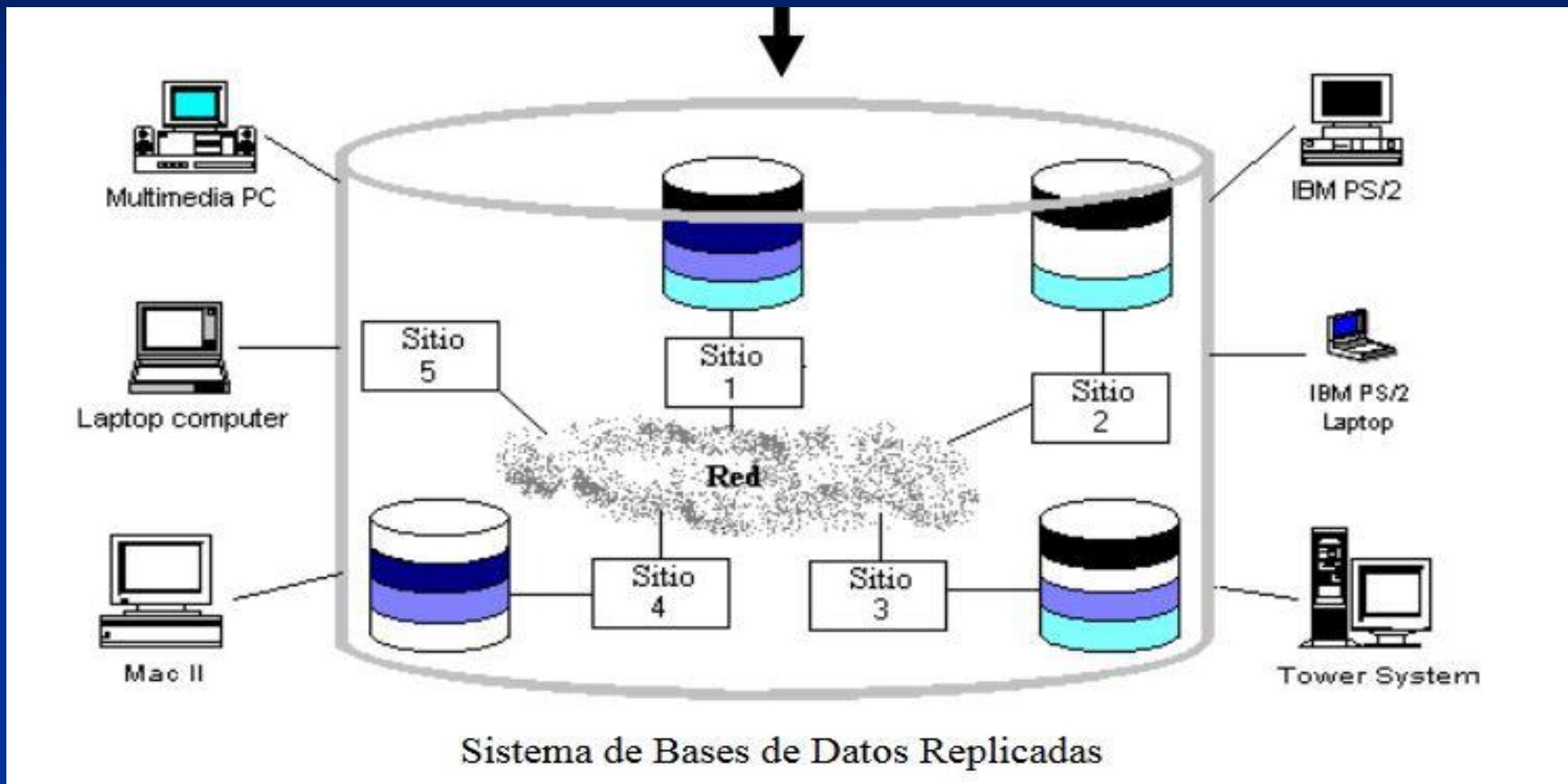
# Imagen de los datos en una organización



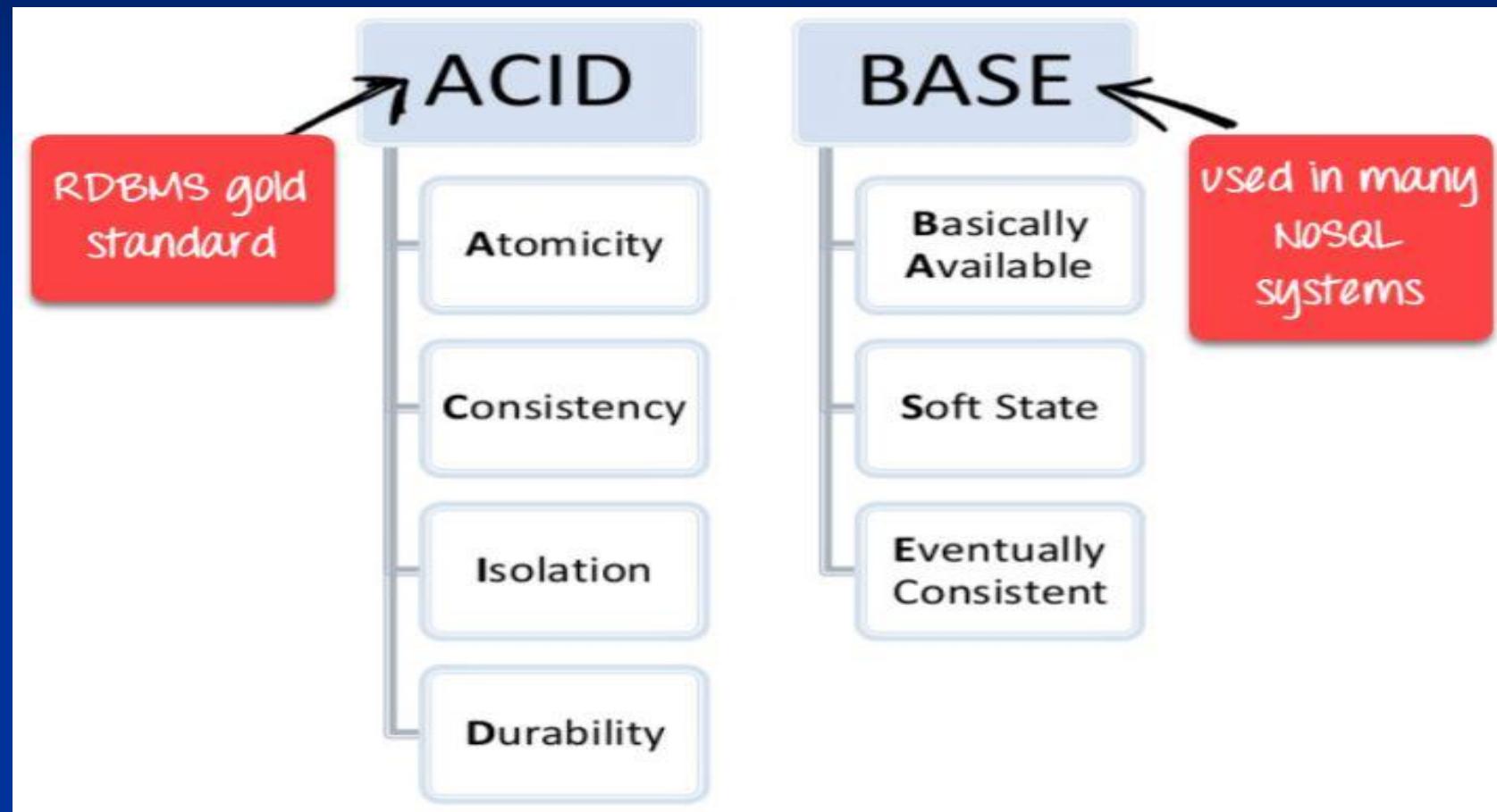
# Sustento de las Bases de Datos No Relacionales



# Sustento de las Bases de Datos No Relacionales



# Sustento de las Bases de Datos No Relacionales





- No requieren de la definición de un esquema de la base de datos



- No requieren de la definición de un esquema de la base de datos
- Almacenan los datos de forma no relacional (tablas hash)



- No requieren de la definición de un esquema de la base de datos
- Almacenan los datos de forma no relacional (tablas hash)
- Están diseñados para arquitecturas distribuidas



- No requieren de la definición de un esquema de la base de datos
- Almacenan los datos de forma no relacional (tablas hash)
- Están diseñados para arquitecturas distribuidas
- Pueden ser instalados en servidores de bajo costo (escalabilidad horizontal)



- No requieren de la definición de un esquema de la base de datos
- Almacenan los datos de forma no relacional (tablas hash)
- Están diseñados para arquitecturas distribuidas
- Pueden ser instalados en servidores de bajo costo (escalabilidad horizontal)
- No generan cuellos de botella

# Diversidad de bases de datos NoSQL



- Key/value data stores
- Document data stores
- Columnar data stores
- Graph data stores
- Object data stores
- External index data stores
- Time series data stores
- Typical requirements

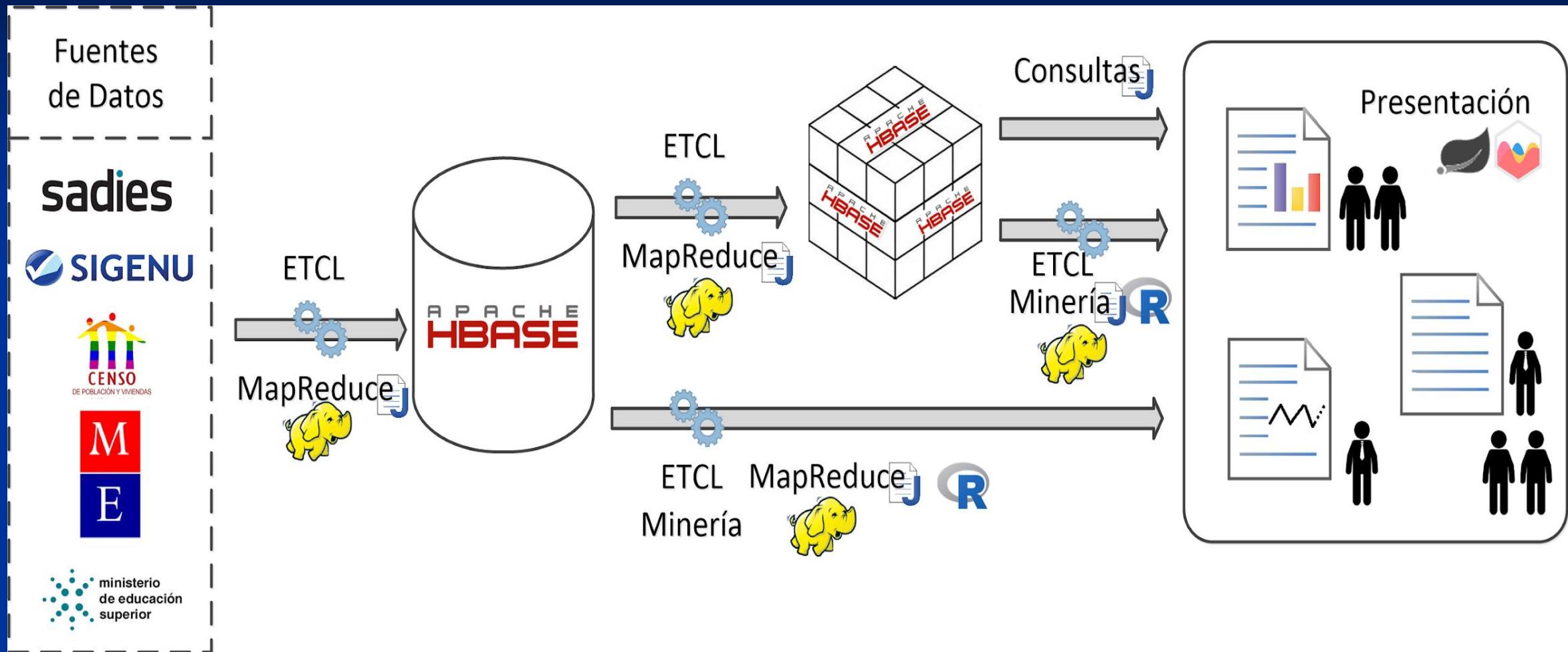
# Solución BI NoSQL para el análisis del ingreso a la educación superior (Orientación a columnas)

Llave	Datos								
99	Procesamiento 1	Color M	Sexo M	Municipio Manzanillo	Provincia Granma	Carrera Medicina	Modalidad CRD	Ocupación Madre Profesional/Dirigente	...
100	Procesamiento 1	Color B	Sexo F	Municipio Amancio Rodríguez	Provincia Las Tunas	Carrera Medicina	Modalidad CRD	Ocupación Madre Profesional/Dirigente	...
101	Color B	Sexo M	Municipio Melena del Sur	Provincia Mayabeque	Ocupación Madre Obrero	...			

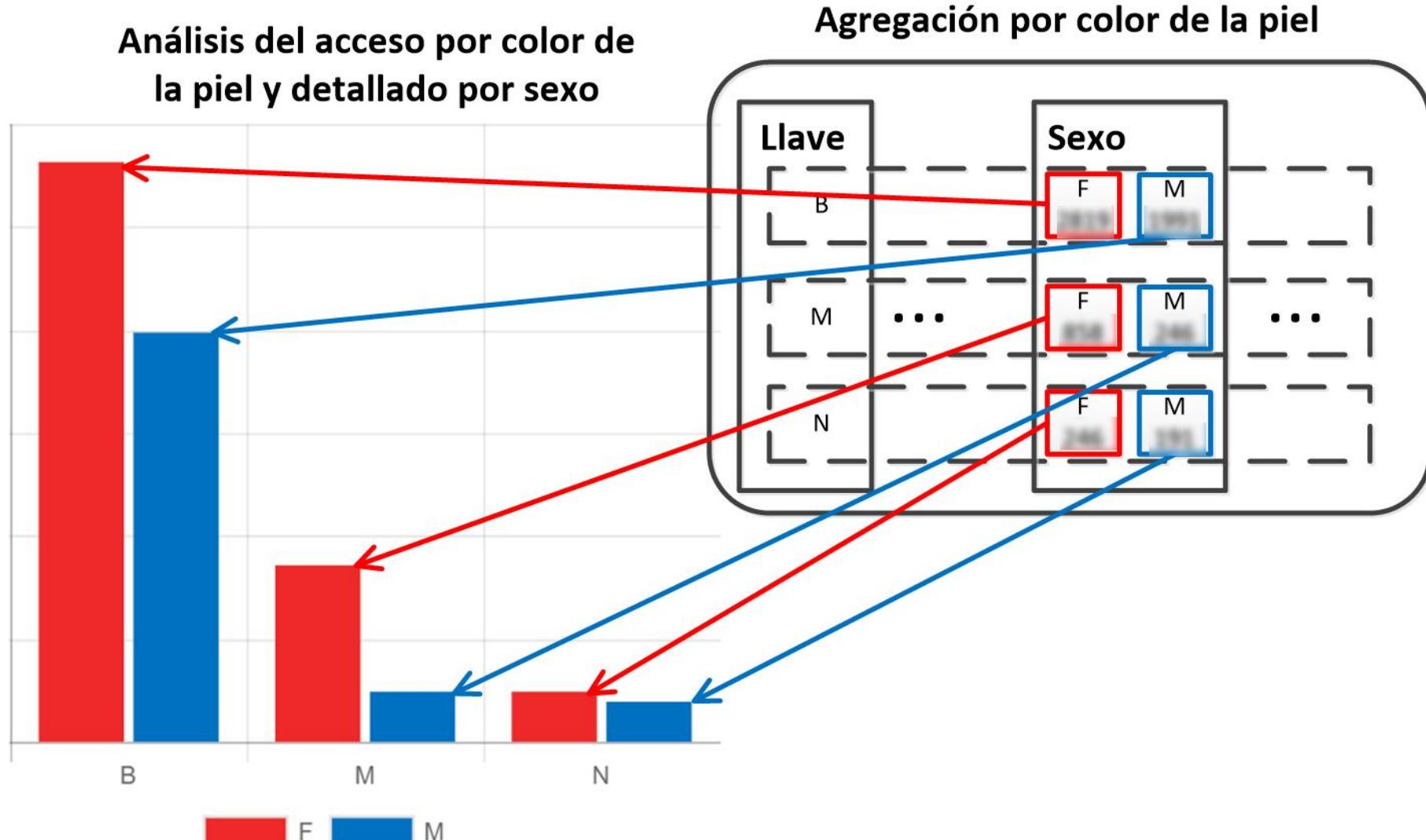
Datos derivados

Datos conciliados

Llave	Procesamiento	Color de la piel	Sexo	Ocupación Madre
Pinar del Río	1 2	B N M	F M	Profesional/Dirigente Obrero
Artemisa	1 2	B N M	F M	Profesional/Dirigente Obrero
La Habana	1 2	B N M	F M	Profesional/Dirigente Obrero
...	...	...	...	...

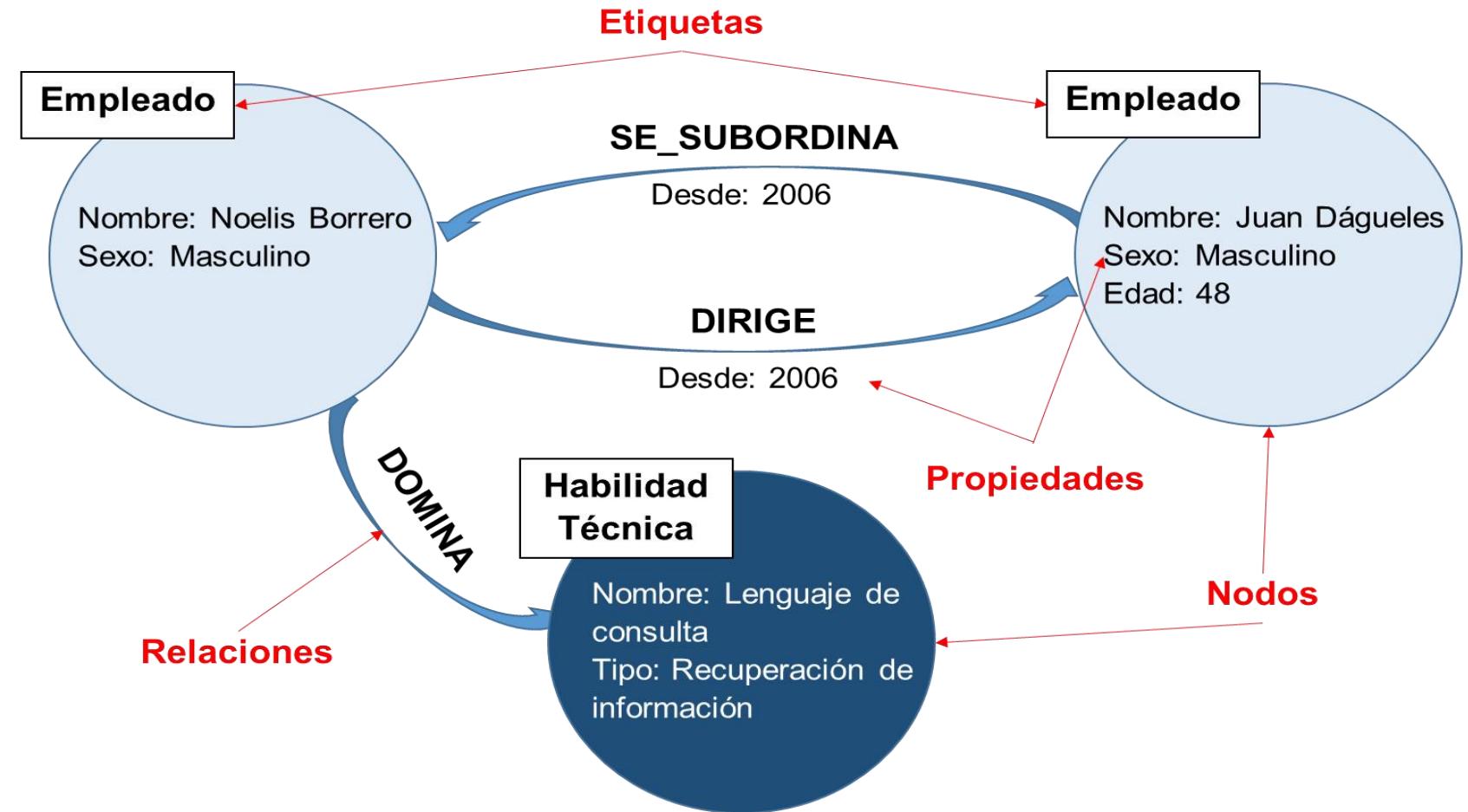


# Implementación de la Presentación de los Resultados

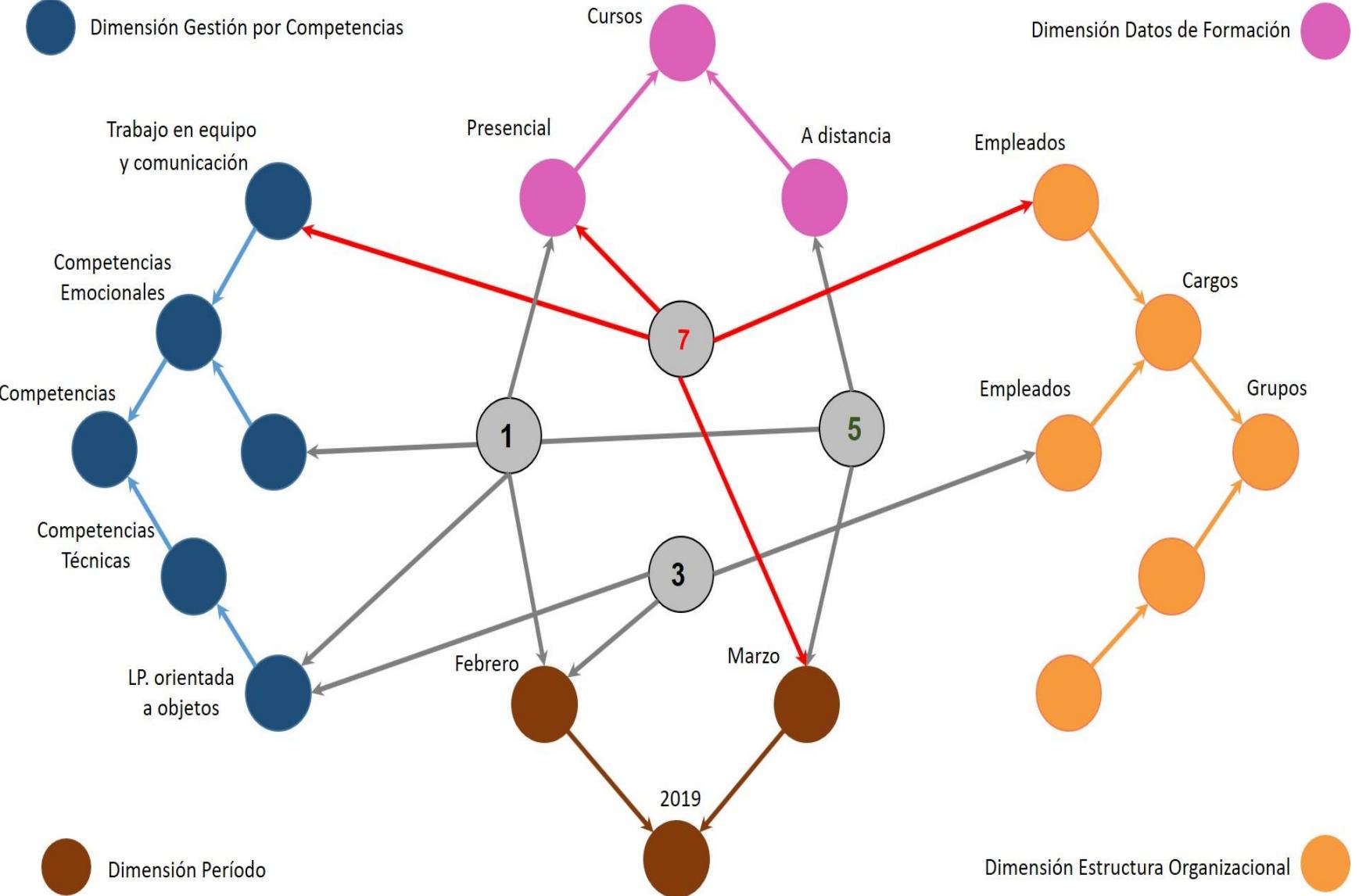


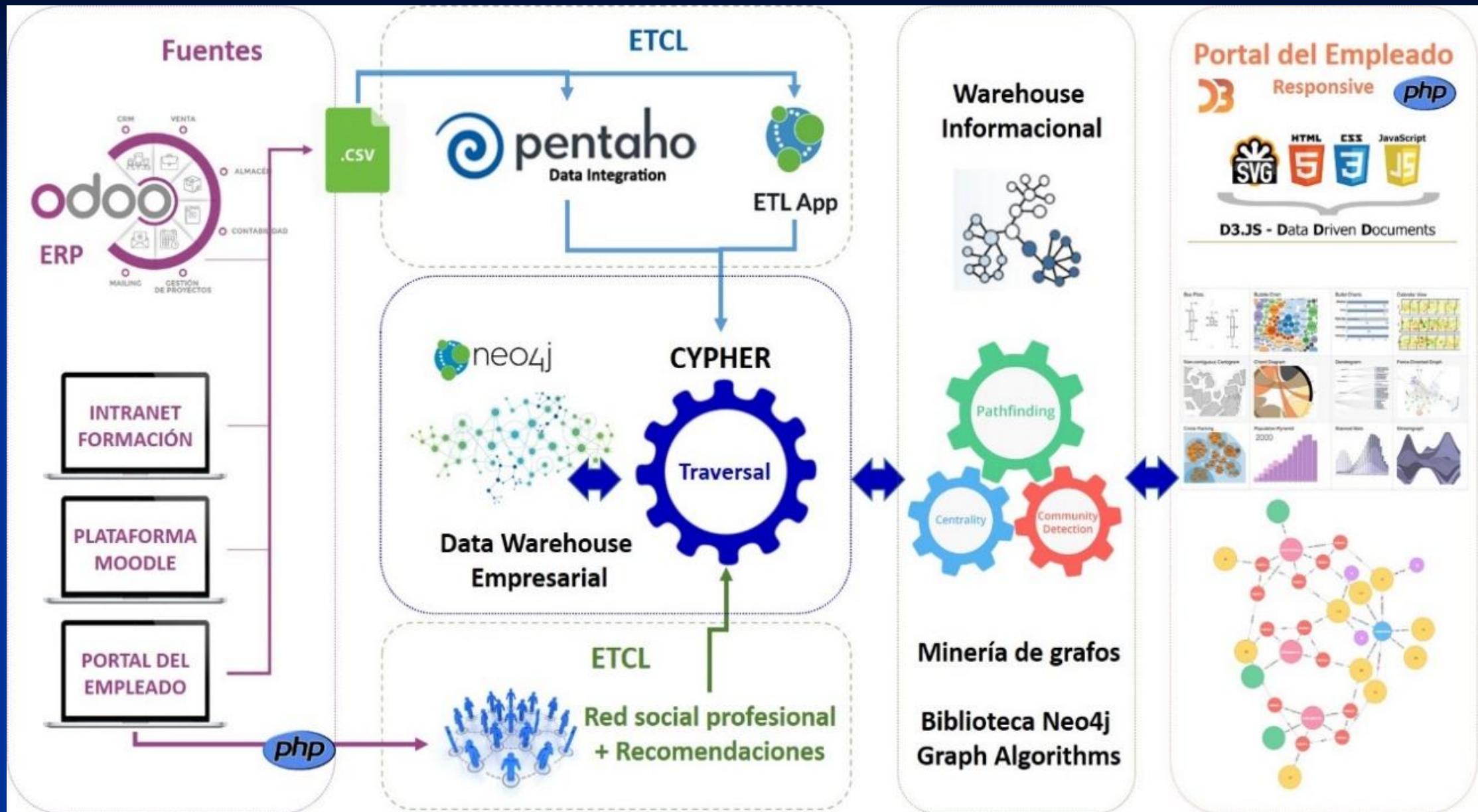
# Solución de inteligencia de negocios con orientación a grafos para la gestión del conocimiento en Desoft.

Instancia del  
modelo de datos

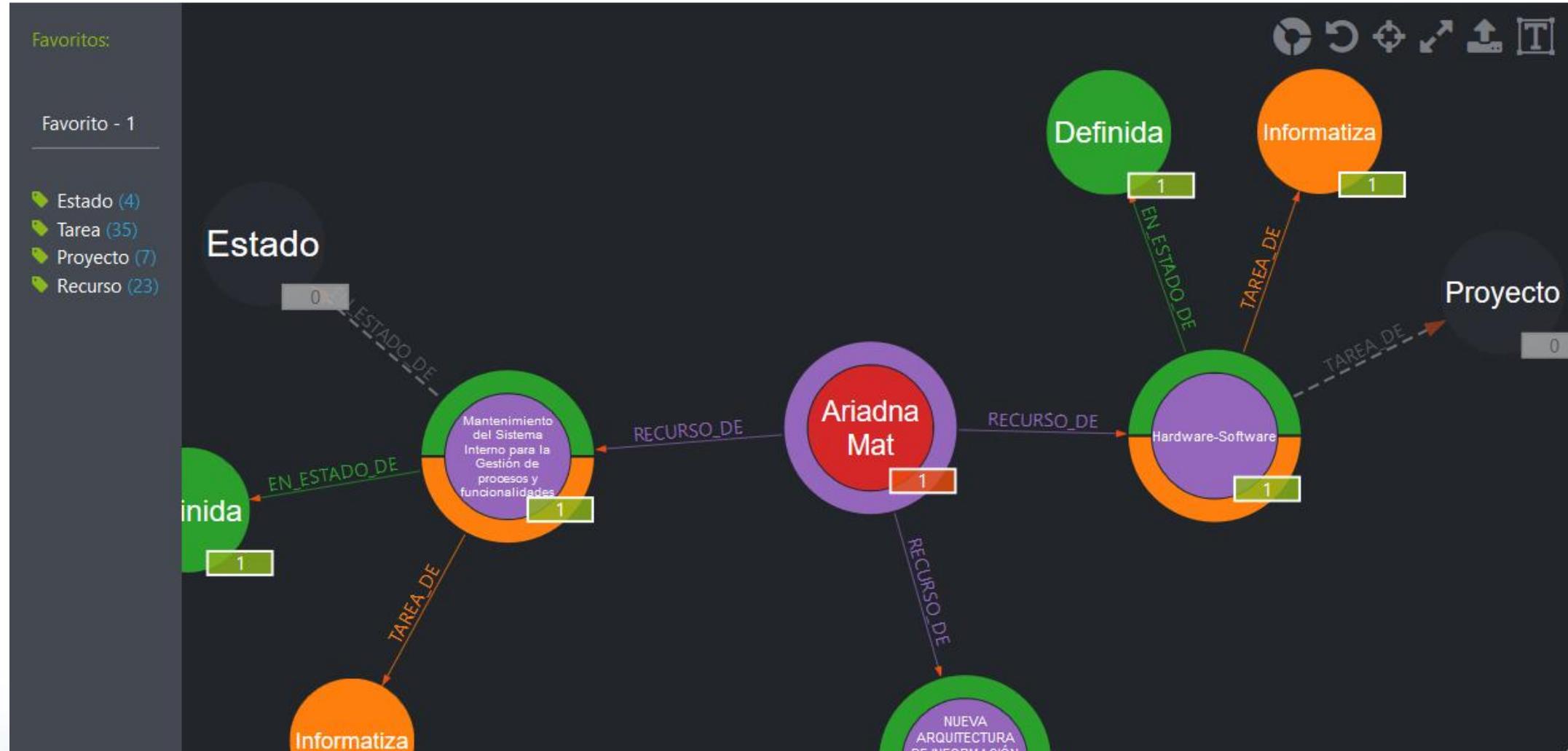


# Instancia de consulta de un cuboide





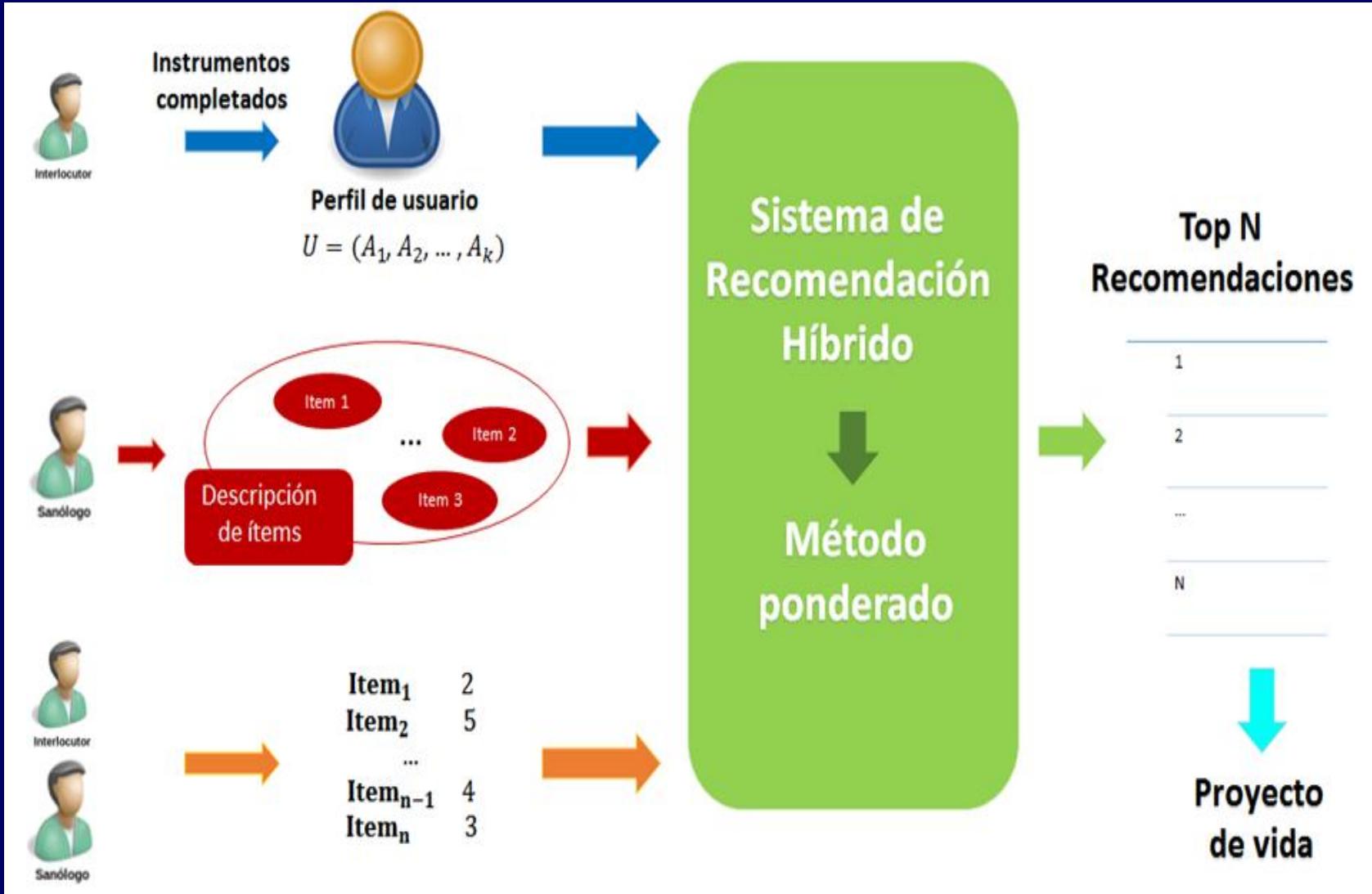
# Diversidad visual y de autoservicio



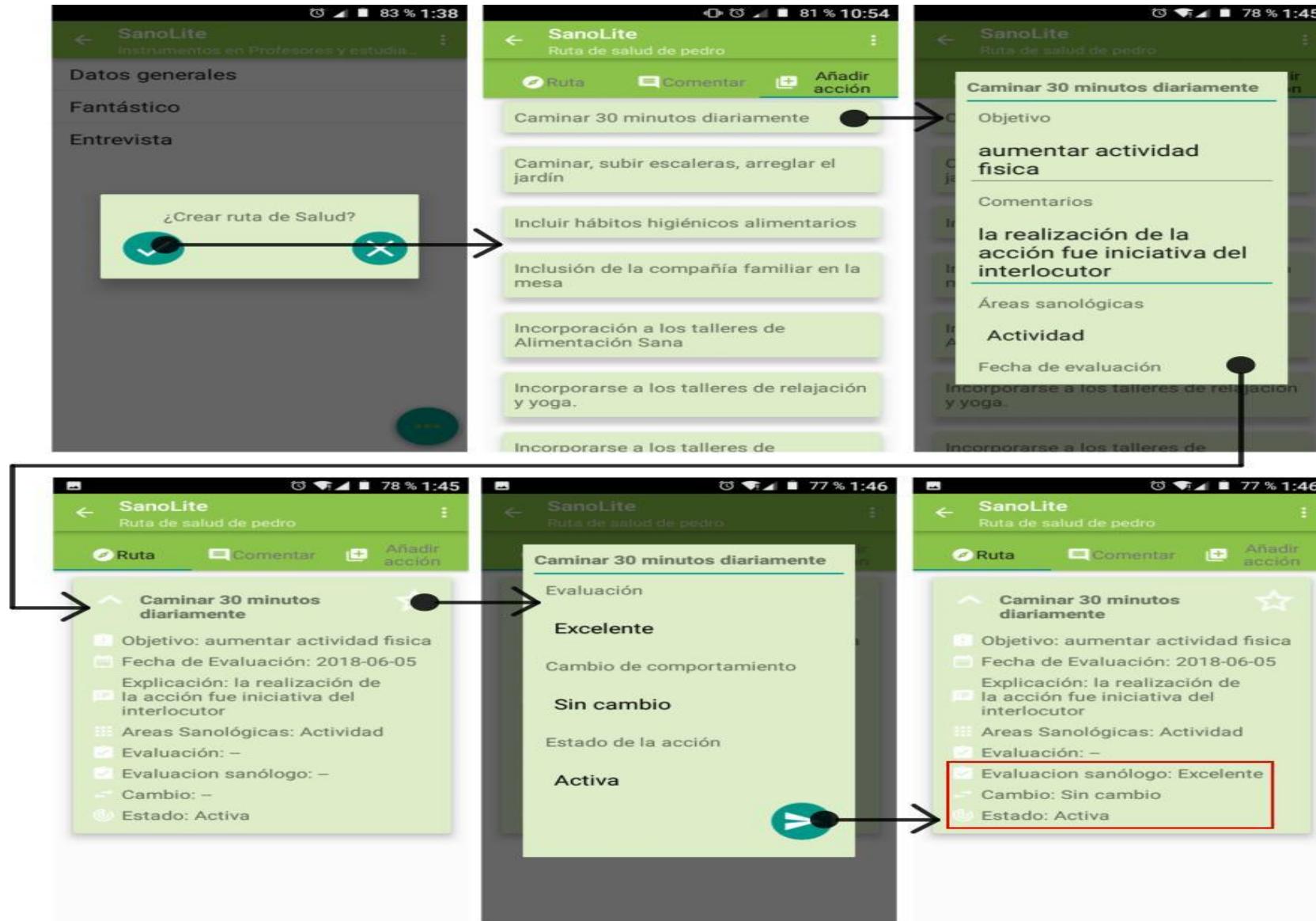
Mapa topológico con enfoque de autoservicio e interactividad en la presentación de los resultados, para el análisis topológico entre Tareas, Recursos, Estados y Proyectos

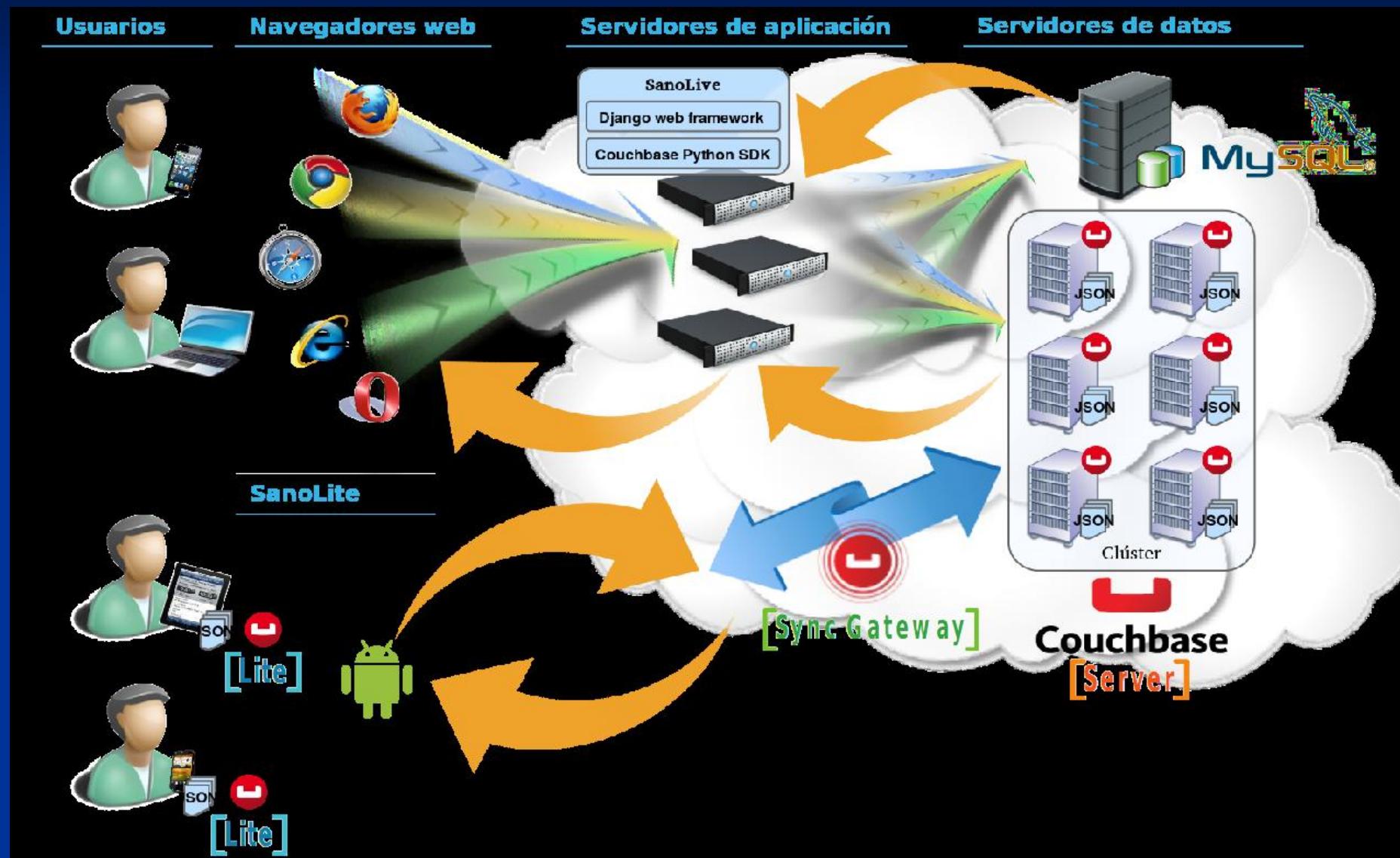
# Solución computacional distribuida para la promoción de la salud y el bienestar humanos (orientación a documentos)





# Interacción de un sanólogo con SanoLite







# Esto sucede en Internet en un minuto

Estimación de una selección de actividades y datos generados online en un minuto en 2021



Fuente: Lori Lewis vía AllAccess



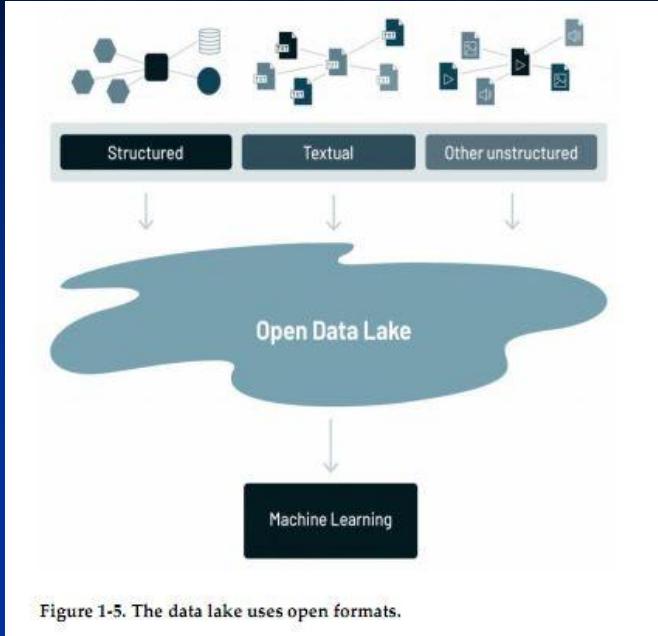
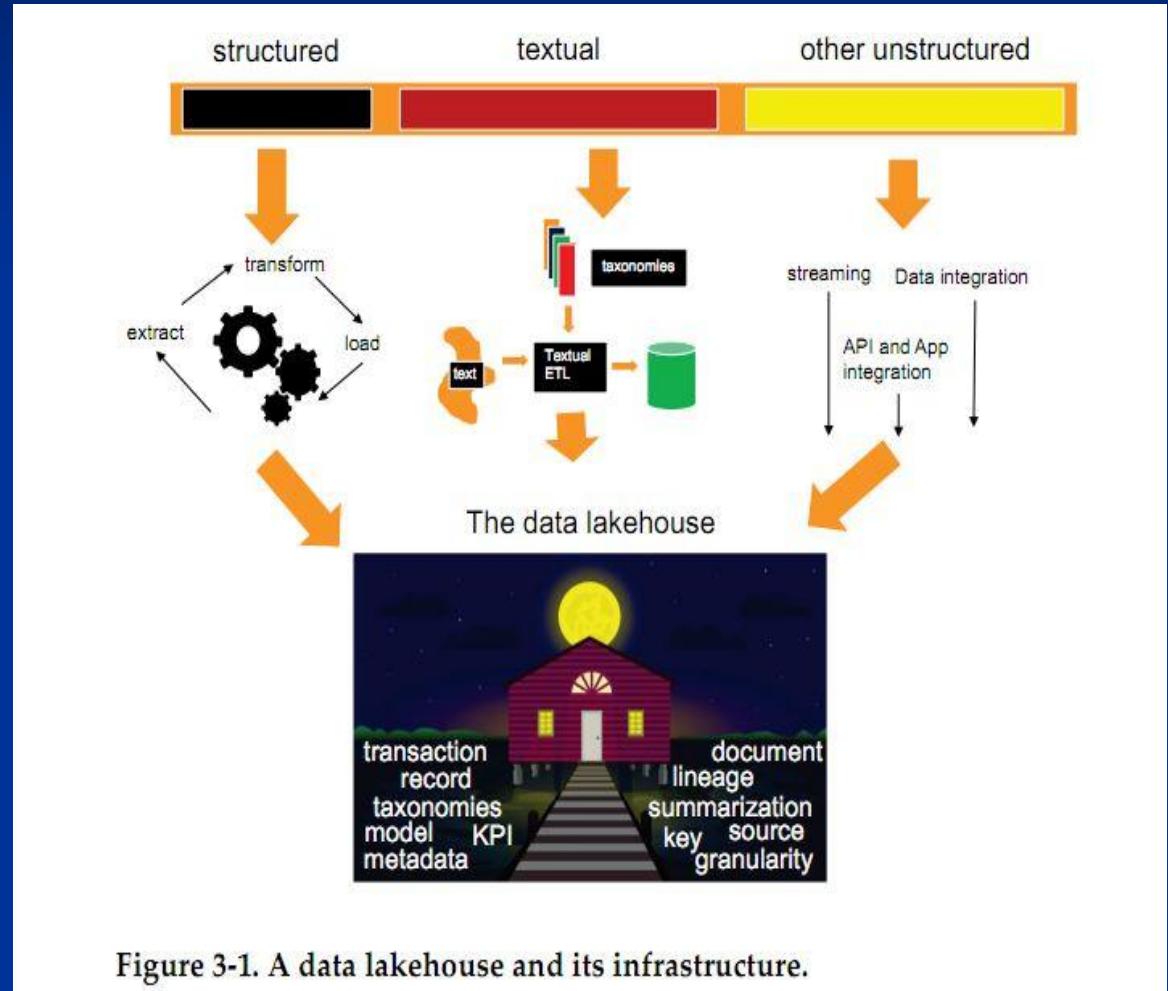
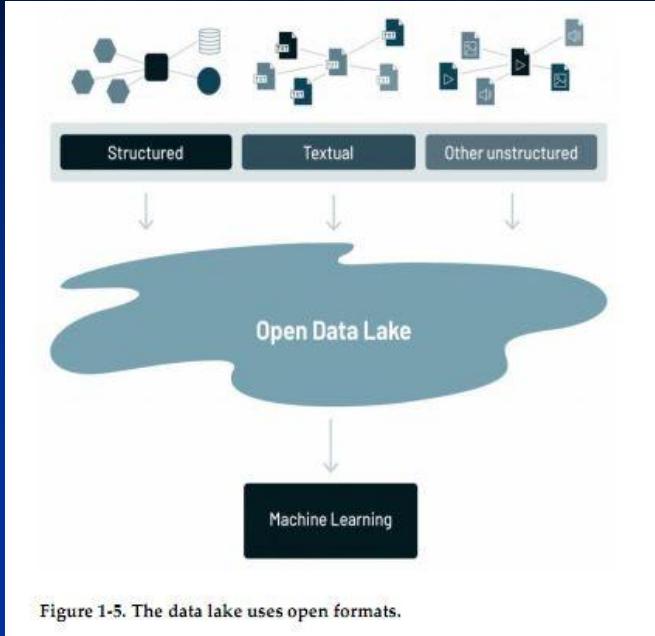


Figure 1-5. The data lake uses open formats.



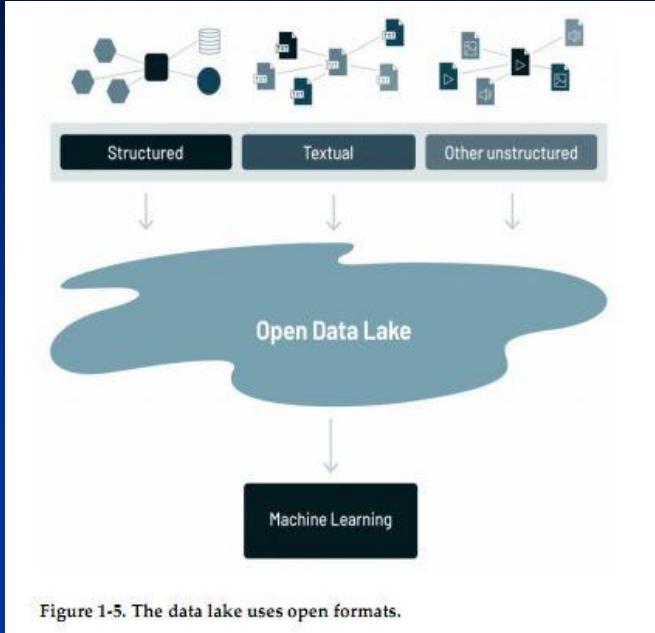
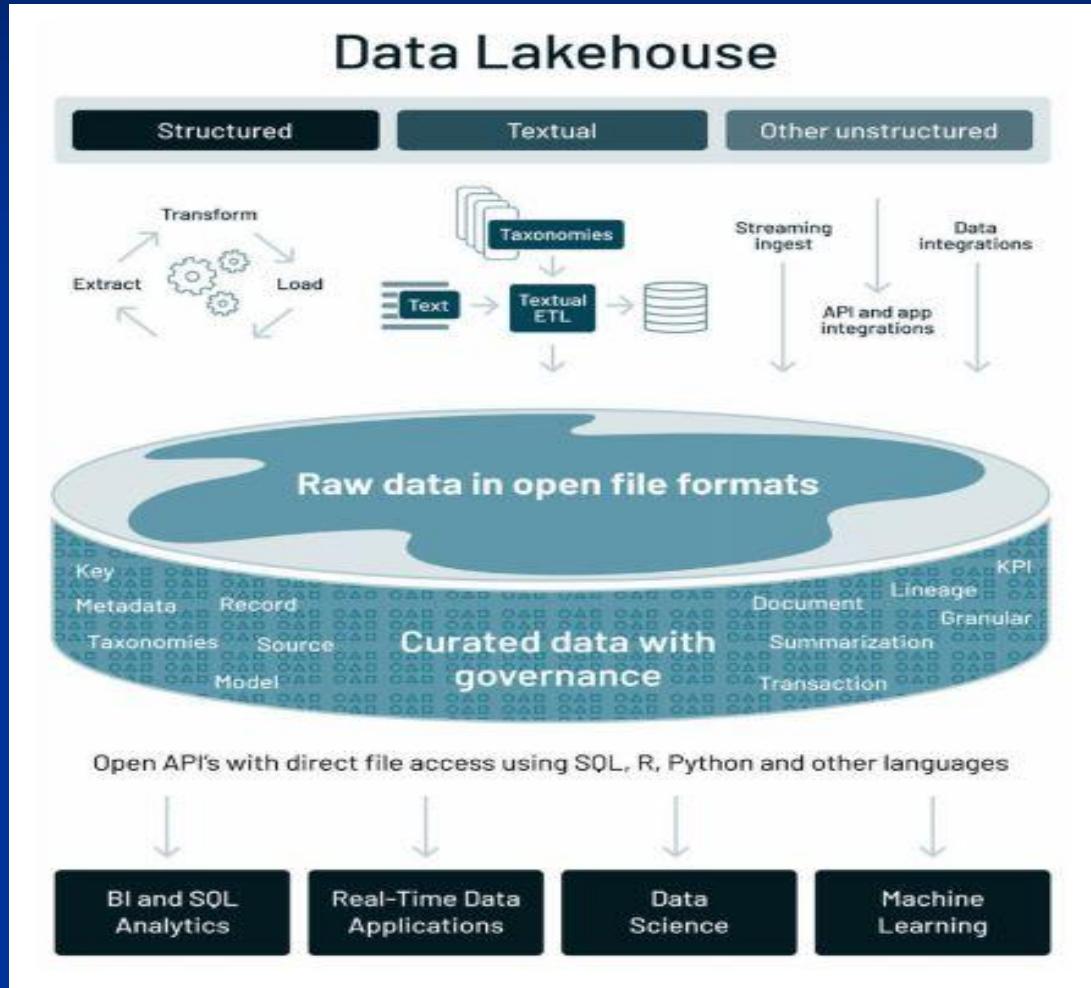


Figure 1-5. The data lake uses open formats.





- Bases de datos autogestionadas
- 
- Más demanda de bases de datos de transmisión (series de tiempo)
- 
- Mayor adopción de Big Data en tiempo real (grafos)
- Clústeres de múltiples nubes
- Seguridad inherente (blockchain)



Relational DB

Distributed SQL

Serverless DB

# Tomando decisiones



# Bases de Datos I

Más allá de las Bases de Datos Relacionales

Lic.

Dra. C. Lucina García Hernández  
Lic. Carlos León González

Lic.

Departamento de Computación  
Facultad de Matemática y Computación  
Universidad de La Habana

5 de diciembre de 2023