

Análisis y Diseño de Bases de Datos

Técnicas avanzadas de gestión
de la información

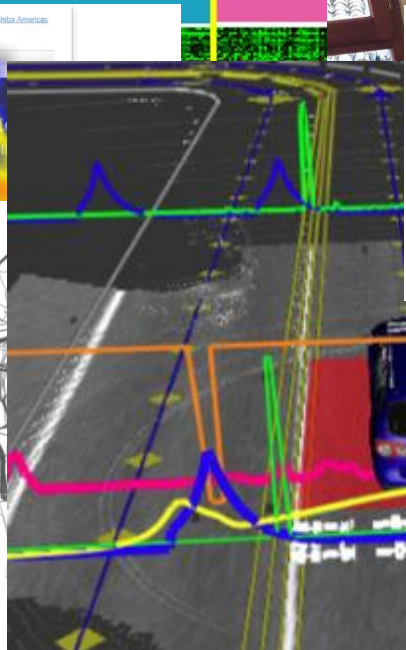
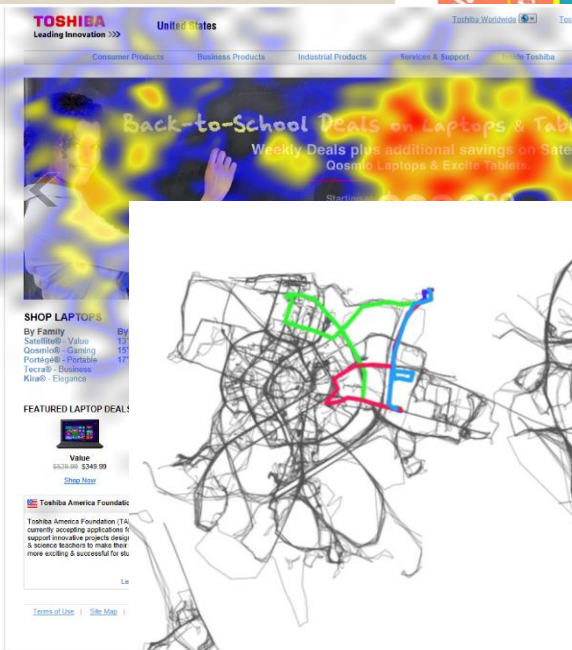
Datos estructurados

- Datos organizados en entidades (*muy*) similares
- Estructura = <entidades, atributos, relaciones>
- Esquema relacional: tablas (relación), filas (entidad), columnas (atributo), claves foráneas (asociaciones)
- Manipulación de datos: formalizada (álgebra/cálculo relacional), lenguaje (SQL)
- Esquema predefinido
- Consistencia estricta
- Escalado *central* (*scaling up*)
- No es flexible / versátil



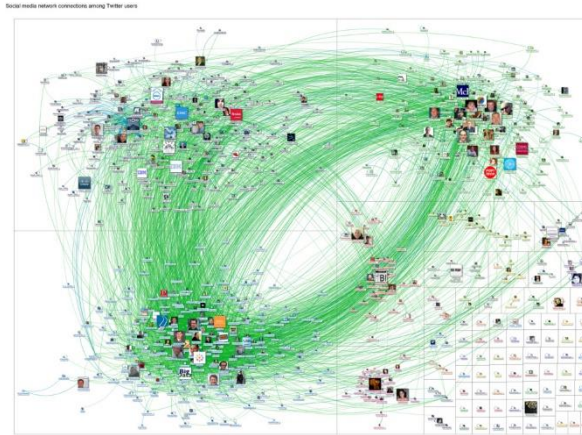
Rank	Site	Both			Site
		Incidence rate (1/10 ⁵)	%	ASIRC* (1/10 ⁵)	
1	Lung (C33-34)	53.57	18.74	25.34	Lung (C33-34)
2	Stomach (C16)	36.21	12.67	17.85	Stomach (C16)
3	Colon, rectum	29.44	10.30	14.21	Liver (C22)

Rank	Site	Both			Site
		Incidence rate (1/10 ⁵)	%	ASIRC* (1/10 ⁵)	
1	Lung (C33-34)	53.57	18.74	25.34	Lung (C33-34)
2	Stomach (C16)	36.21	12.67	17.85	Stomach (C16)
3	Colon, rectum	29.44	10.30	14.21	Liver (C22)



La revolución de los datos

- Big data



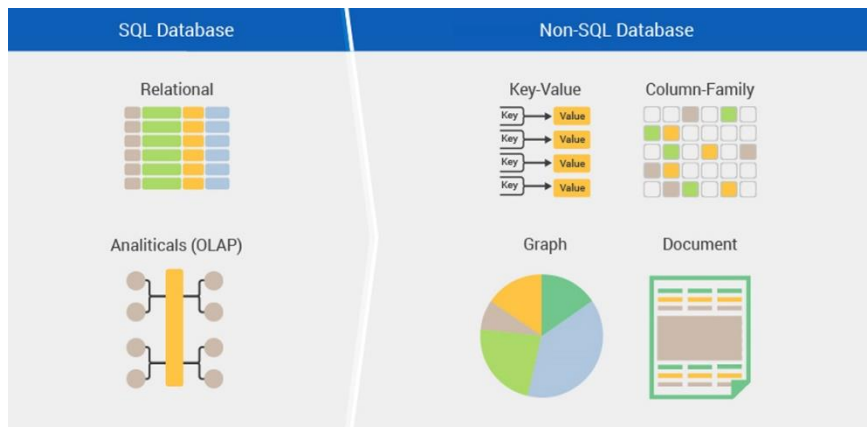
BIG DATA

ANALYTICS

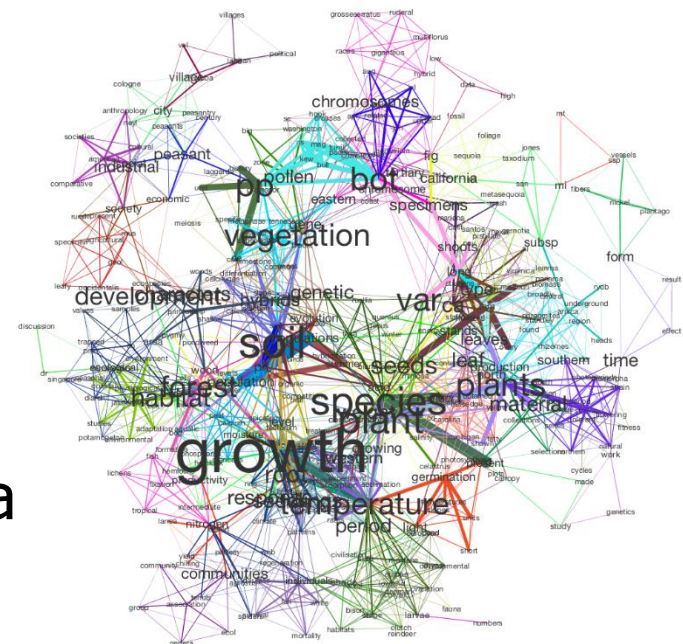
DECISIONS



- NoSQL



- Web semántica / linked data



Big data

■ Big data

- grandes colecciones de datos complejos
- procesos complejos:
 - captura, integración, limpieza, anotación
 - almacenamiento, disponibilidad, acceso
 - 50-80% trabajo antes de que puedan utilizarse
 - análisis, visualización, modelos de conocimiento
- ámbitos de aplicación: economía, sociología, salud, industria, administraciones públicas, ...
- alinear big data con objetivos específicos: qué datos capturar, cómo tratarlos, para obtener qué
- Vs: *volume, velocity, variety* —*veracity, value*

5 Vs



Velocity

Frequency of data generation



300 hours

of video uploaded to YouTube every minute (estimate)



2,400,000 search queries

per minute on Google

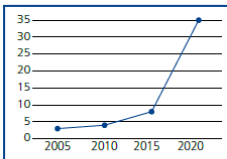
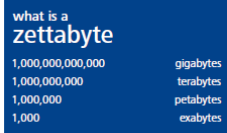


4,170,000 posts liked

on Facebook per minute

Volume

The growth of world data



1 terabyte holds the equivalent of roughly 210 single sided DVDs.

Variety

Structured and unstructured data - types of Big Data



Web and social media



Machine to machine



Big transaction data



Biometric



Human-generated

Veracity

Establishing trust in data



1 in 3

business leaders don't trust the information they use



Uncertainty

due to inconsistency, ambiguity, latency and approximation

Viability

Relevance and feasibility

"Can we use mobile phone data to monitor cross-border tourism?" **Hypothesis**

Validation to determine if the data will have a meaningful impact



Long-term

rewards and better outcomes from hidden relationships in data

Value Return on investment



Costs

Risk of simply creating Big Costs without creating the value



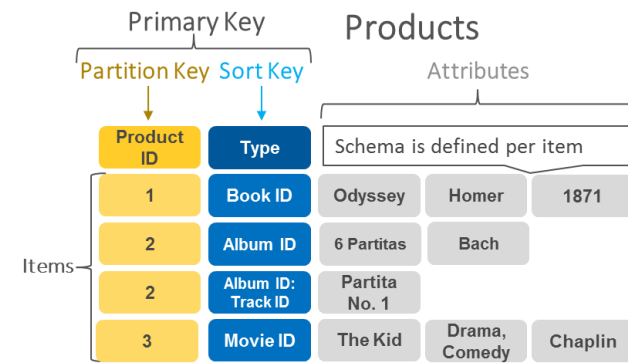
Insights

Sophisticated queries, counterintuitive insights and unique learning

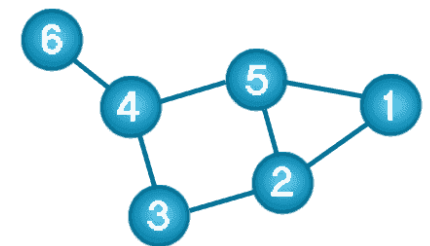
NoSQL

■ Tipos

- *key-value*: datos como colección indexada de registros / campos / datos (diccionario); sencillos modelos y accesos por clave
- *document*: datos semi-estructurados en formato documento; atributos/cabeceras con grandes volúmenes de datos documentales
- *column*: tratamiento orientado a columna (= tipo datos) rendimiento en agregaciones, datos derivados, business intelligence
- *graph*: entidades conectadas aplicaciones con representación de conectividad, interacción, etc.



1167	Ale C	Miller	570
3424	Beerio	Ians	340
5612	Amstel	Amstel	121
2409	Colt's	BeerCo	98



■ ■ ■

Key Value



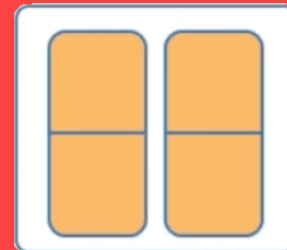
Example:
Riak, Tokyo Cabinet, Redis
server, Memcached,
Scalaris

Document-Based



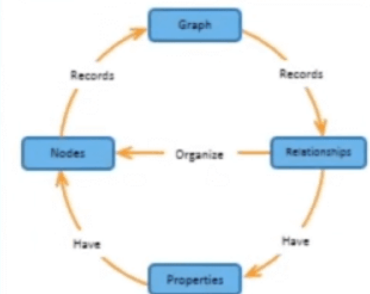
Example:
MongoDB, CouchDB,
OrientDB, RavenDB

Column-Based



Example:
BigTable, Cassandra,
Hbase,
Hypertable

Graph-Based



Example:
Neo4J, InfoGrid, Infinite
Graph, Flock DB

Semántica

■ Web semántica

- contenido semántico, procesable por máquinas
- web de los datos —vs web de los documentos
- componentes (habituales)
 - XML, sintaxis básica de estructuración de contenidos
 - XML Schema, proporciona y restringe la estructura de contenidos
 - RDF, modelo de datos en forma de <sujeto, propiedad, objeto>; normalmente en sintaxis XML
 - RDF Schema, vocabularios de propiedades
 - OWL, ontologías / conocimiento
 - SPARCQL, lenguaje de consulta

■ Linked data

- reformulación de la web semántica

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE recipe PUBLIC "-//Happy-Monkey/DTD RecipeBook//EN"
"http://www.happy-monkey.net/recipebook/recipebook.dtd">
```

```
<recipe>
```

```
<title>Peanut-butter On A Spoon</title>
```

```
<ingredientlist>
```

```
<ingredient>Peanut-butter</ingredient>
```

```
</ingredientlist>
```

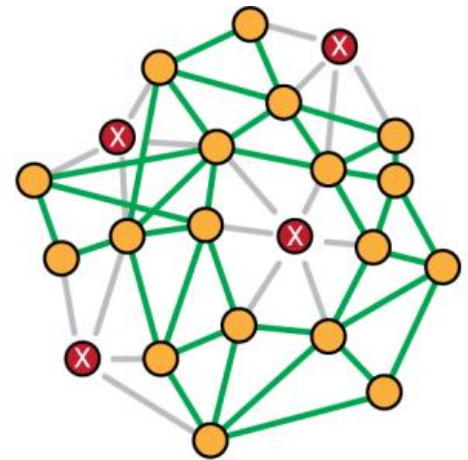
```
<preparation>
```

```
Stick a spoon in a jar of peanut-butter,
scoop and pull out a big glob of peanut-butter.
```

```
</preparation>
```

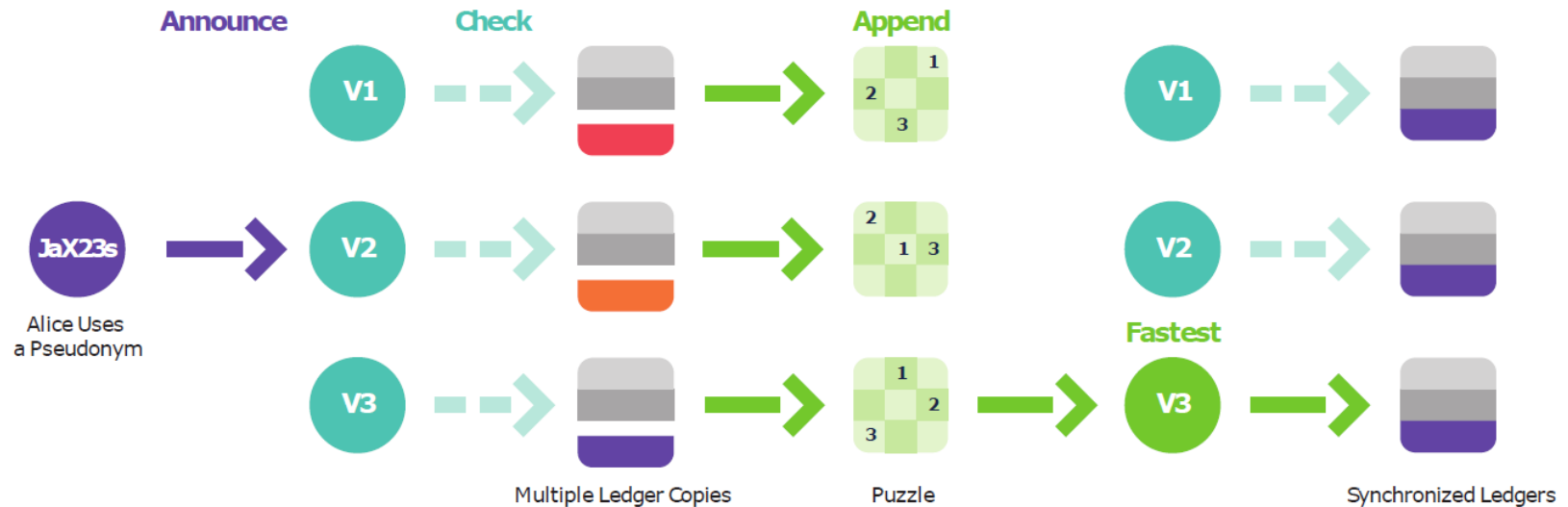
```
</recipe>
```

Blockchain



- BD distribuida
 - no depende de un nodo central
 - nodos no *fiables*; globalmente fiable por el protocolo de funcionamiento —*consenso*
- Estado / transacciones
 - el estado cambia por transacciones anotadas en un registro —*libro anotaciones (ledger)*
- Criptografía
 - aseguramiento de privacidad, autenticación, integridad, inmutabilidad, etc.
- Lógica programada / contratos inteligentes

Distributed ledger technology



- Operativa descentralizada
- Copia distribuida
- Prueba de participación
- Convergencia por copia más larga