

# Bases de dades i Big Data.

## Activitat 1

Javier Pedragosa Lozano

### Què és una base de dades relacional? Com s'hi estructuren les dades?

Una base de dades relacional és un tipus de base de dades que emmagatzema i dona accés a dades que estan relacionades les unes amb les altres. Les bases de dades estan basades en el model relacional, una manera intuïtiva i directa de representar dades en taules. En una base de dades relacional, cada fila de la taula es un registre amb una identificació única anomenada clau (key). Les columnes de la taula contenen atributs de les dades, i cada registre normalment té un valor per a cada atribut, facilitant establir les relacions entre les dades.

### Què és una base de dades jeràrquica? Com s'hi estructuren les dades?

Una base de dades jeràrquica és un tipus de base de dades en què les dades s'emmagatzemen en forma de registres i organitzades en una estructura d'arbre, o de pares-fills (parent-child), a on cada node pare pot tenir diversos nodes fill connectats a través d'enllaços.

### Què és Big Data?

Big Data fa referència a la increïble quantitat d'informació estructurada o no que les persones i les màquines generen (de l'ordre dels petabytes diaris).

Són tots aquells conjunts extremadament llargs i complexos de dades que no poden ser fàcilment gestionats o analitzats amb eines de processament de dades tradicionals, particularment fulls de càlcul.

### Quines són les seves característiques?

Les 5 «Vs» del Big Data són:

1. **Volum.** La quantitat de dades importa. Amb Big Data s'han de processar grans quantitats de dades de baixa densitat i desestructurades. Aquestes dades poden ser de valor desconegut, com poden ser les fonts de dades d'una xarxa social, les visites d'un vídeo, o lectures d'una màquina provista de sensors. Per algunes empreses podriem estar parlant de desenes de terabytes de dades, per unes altres de centenars de petabytes.
2. **Velocitat.** La velocitat és el ritme en que es reben i (potser) es tracten les dades. Hi ha aplicacions en temps real, les quals han d'enviar i processar directament les dades

a memòria, i d'altres que poden permetre's el ritme d'escriure aquestes dades en un disc dur.

3. **Varietat.** La varietat fa referència a la quantitat de tipus de dades de les que disposem. Els tipus de dades tradicionals estaven estructurats i cabien perfectament en una base de dades relacional. Amb l'auge del Big Data, les dades venen donades en nous tipus de dades desestructurades. Els tipus de dades desestructurades i semiestructurades, com ara text, audio i vídeo, requereixen un preprocessament adicional per a tenir sentit i suportar metadades.
4. **Veracitat.** Com de fiables i reals son les dades. La idea de veracitat en les dades està connectada a altres conceptes funcionals, com ara la qualitat de les dades i la integritat de les dades. Finalment, totes aquestes se superposen i dirigeixen la organització a un repositori de dades que lliura dades d'alta qualitat, acurades i de fiar, que faciliten les idees i la presa de decisions.
5. **Valor.** Les dades tenen un valor intrínsec als negocis. Però no serveix de res fins que aquest valor és descobert. Pel fet que el Big Data reuneix tant l'amplitud com la profunditat dels coneixements, en algun lloc d'entre tota aquesta informació es poden trobar coneixements que poden beneficiar una empresa. Aquest valor pot ser intern, com ara processos operacionals que poden ser optimitzats, o externs, com ara suggerències de perfils de clients que poden maximitzar la fidelització.

## Posa exemples de la utilització de Big Data.

Transport. El Big Data és la base de les aplicacions de GPS de les quals la majoria de nosaltres depenem per anar d'un lloc a un altre. Les fonts de dades de GPS inclouen imatges per satèl·lit, agències del govern, rutes d'usuaris...

Publicitat i màrqueting. Els anuncis sempre tenen uns segments de població com a objectius. Avui dia els anunciants compren o acumulen grans quantitats de dades per identificar què cliquen els usuaris.

## Com es guarden les dades en Big Data?

Normalment es guarden en Data Lakes, pensats per a afegir i consultar dades, i en Data Warehouses, pensats només per a la consulta de les dades, normalment dades històriques.

## Explica què són, les dades estructurades, les dades semiestructurades i les dades no estructurades.

Les dades estructurades són aquelles dades els elements de les quals son referenciables per a poder analitzar-les de manera efectiva. S'organitzen en un repositori formatat (normalment una base de dades).

Les dades semiestructurades són tota aquella informació que no resideix a una base de dades relacional però que té certes propietats d'organització que en faciliten l'anàlisi.

Les dades no estructurades són aquelles dades que no estan organitzades de cap manera predefinida o no tenen un model de dades predefinit, per la qual cosa no encaixen gaire per una base de dades relacional convencional.

## Què és NoSql? Posa un exemple on sigui necessari utilitzar-ho.

Les bases de dades NoSql emmagatzemen les dades de manera diferent que les taules relacionals. Venen donades en una varietat de tipus basades en el seu model de dades. Els tipus principals són per document, per valor-clau, wide-column i grafs. Proporcionen esquemes flexibles i escalen fàcilment amb grans quantitats de Big Data i grans càrregues d'usuaris.

És necessari utilitzar-lo en casos d'emmagatzematge de dades estructurades i semiestructurades, en casos que es tractin grans volums de dades, quan es requereixi la possibilitat d'escala, i en paradigmes d'aplicacions modernes com els microserveis i l'*streaming* en temps real.