

ORIGEN I UTILITAT DE LES DADES

En aquest article parlarem dels tipus de dades que existeixen en l'univers del *big data* i quines són les seves principals fonts. Mostrarem exemples perquè ens anem familiaritzant a poc a poc amb formats, sintaxis i perquè anem comprenent com afecta cada cas d'ús al negoci i com pot millorar la nostra qualitat de vida.

Tipus de dades

En primer lloc, analitzarem els diferents tipus de dades amb els quals es treballa. Com a primera aproximació, imaginarem que treballem amb molts fitxers independents i que cadascun d'ells pot comprendre un dels formats que es descriuen a continuació, els quals es categoritzen en funció de la seva estructura.

- **Estructurades**
 - Aquest tipus de dades tenen la seva longitud, format i mida ben definits. S'emmagatzemen en taules, fulls de càlcul o bases de dades. A la taula que es mostra a continuació, es poden apreciar les últimes transaccions bancàries d'una persona particular. La informació ve organitzada pels atributs "Tipus", "Origen" i "Concepte", i cada registre té una mètrica, corresponent a l'"Import" de l'operació:

TIPUS	ORIGEN	CONCEPTE	IMPORT
Pagament amb targeta	Compte corrent	Gasolinera	55
Transfarència	Compte d'estalvi	Balanç de despeses	500
Pagament amb targeta	Compte corrent	Supermercat	30

- **Semiestructurades**
 - Els registres d'un origen semiestructurat no segueixen una definició estàndard com els estructurats, és a dir, no tenen assignat un format comú, però sí que presenten una organització basada en metadades. Gràcies a ella, s'estableixen objectes i les seves relacions, que moltes vegades estan acceptats per convenció en formats coneguts com a HTML, JSON o XML. En l'exemple que es mostra a continuació, es pot observar un document tipus JSON, en el qual s'ha creat un objecte "Operació" i els seus elements niats, que representen transaccions bancàries i les seves propietats:

```
{
  "Operació": [
    {
      "Tipus": "Pagament targeta",
      "Origen": "Compte corrent",
      "Concepte": "Gasolinera",
      "Import": "55 €"
    },
    {
      "Tipus": "Transferència",
      "Origen": "Compte estalvi",
      "Concepte": "Balanç despeses",
      "Import": "500 €"
    },
    {
      "Tipus": "Pagament targeta",
      "Origen": "Compte corrent",
      "Concepte": "Supermercat",
      "Import": "30 €"
    },
  ],
}
```

- No estructurades

- Tal com indica el seu nom, els conjunts de dades que comprenen informació no estructurada són aquells que no segueixen un format específic. Solen associar-se a documents de text, imatges, vídeos o *e-mails*, entre d'altres. Seguint amb els exemples anteriors, la taula que descriu les transaccions bancàries podria estar emmagatzemada bé en una fulla de càlcul (.xlsx), en format de text avançat (.docx) o documents PDF. La seva explotació és més complexa que fitxers de text pla i requereix de programari específic, moltes vegades sota llicència, o d'algoritmes avançats, com pot ser el processament d'imatges o fitxers d'àudio.

Tipus de dades per origen

Els sistemes *big data* han arribat per donar resposta al gran volum d'informació disponible que gestionem, de manera que es pugui processar de manera ràpida i crear un valor que ofereixi solucions a qualsevol tipus d'escenari. Fins ara, les bases de dades tradicionals (Oracle, MySQL, SQL Server, SAP, etc.) han pogut donar una solució parcial al processament de la informació, però, ara, els sistemes moderns, gràcies a les noves arquitectures i capacitats de processament, són capaços d'adquirir, transformar i analitzar les dades de múltiples orígens en temps real. Aquestes fonts poden provenir de l'app de missatgeria del nostre *smartphone*, dels sensors que monitoren la geolocalització, temperatura, posició, etc. de diferents dispositius, de correus electrònics o de *logs* de servidors. Tots aquests nous focus d'informació formen un nou i complex univers de dades, que, combinats, multipliquen el seu valor, pel fet de ser emmagatzemats de manera conjunta al mateix lloc.

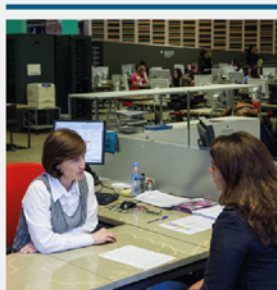
Podríem considerar alguns dels exemples que es descriuen a continuació com els orígens més comuns dels sistemes de *big data*.

- **Generats per persones.** Són aquelles fonts que generem de manera o no conscient, bé sigui amb l'ús de les eines quotidianes o amb diversos registres sanitaris, bancaris, etc.
 - Missatgeria instantània
 - Notes de veu i enregistraments d'àudio
 - Correus electrònics
 - Registres electrònics (mèdics, Seguridad Social, bancaris, sistemes privats, etc.)
 - Documents electrònics (DNI, passaport, permís de conduir, targetes de fidelització, etc.)
- **Xarxes socials i fonts web.** Avui dia, gran part de la població publica imatges, redacta històries o anècdotes, interactua amb altres persones a través de comentaris o reaccions i, per descomptat, utilitza motors de cerca per accedir a qualsevol tipus d'informació. Tot això s'emmagatzema en les famoses *cookies*, en l'historial del nostre navegador i, sobretot, en els servidors dels portals als quals accedim. Aquesta informació és la més valuosa per a les empreses, perquè amb ella poden traçar tendències de mercat, recomanacions personalitzades de pel·lícules, roba o productes d'interès per generar publicitat o registres d'activitat. Per tant, trobem:
 - Qualsevol tipus de xarxa social
 - Motors de cerca (Google, Bing, Yahoo!, etc.)
 - Informació sobre clics a vincles i elements
 - RRSS (fonts de dades de Twitter, publicacions a Facebook, xarxes socials de blogs, premsa...)
 - Contingut web (pàgines, imatges, enllaços, etc.)
- **Comunicació entre màquines (*machine-to-machine*, M2M).** També conegut per molts com a IoT (*internet of things*). Es tracta de fonts físiques i automatitzades que utilitzen

sistemes de radiofreqüència, protocols com Bluetooth, ZigBee, WiFi, RDIF o GPS per monitorar una certa activitat.

- Targetes d'accés i dispositius de seguiment RFID
 - Dispositius de geolocalització a través de senyals GPS
 - Altres sensors (parquímetres, màquines expenedores, caixers, etc.)
- **Transaccions.** Esdeveniments generats per dispositius de telecomunicacions (mòbils, senyals de radiocomunicació...) o esdeveniments únics de pagaments bancaris o vendes.
 - Registres de comunicacions (crides, missatgeria, VoIP, etc.)
 - Registres de facturació (pagaments amb targeta, pagament *online*, registre de comptador de llum intel·ligent, etc.)
 - Registres de vendes o comandes *online* (utilització de comerç electrònic i d'altres aplicacions com Glovo o Wallapop)
 - **Biomètrics.** Són aquells que ajuden a la identificació unívoca d'una persona en funció de la base de trets físics o de conducta. Empleats principalment en sistemes d'accés i seguretat.
 - Reconeixement facial
 - Informació genètica (ADN)
 - empremtes dactilars
 - Escaneig de retina

Descobreix tot el que Barcelona Activa pot fer per a tu



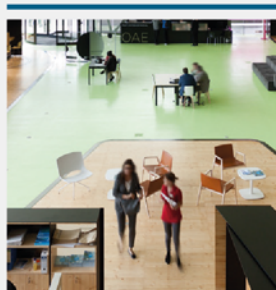
Acompanyament durant tot el procés de recerca de feina

barcelonactiva.cat/treball



Suport per posar en marxa la teva idea de negoci

barcelonactiva.cat/emprenedoria



Serveis a les empreses i iniciatives socioempresarials

barcelonactiva.cat/empreses

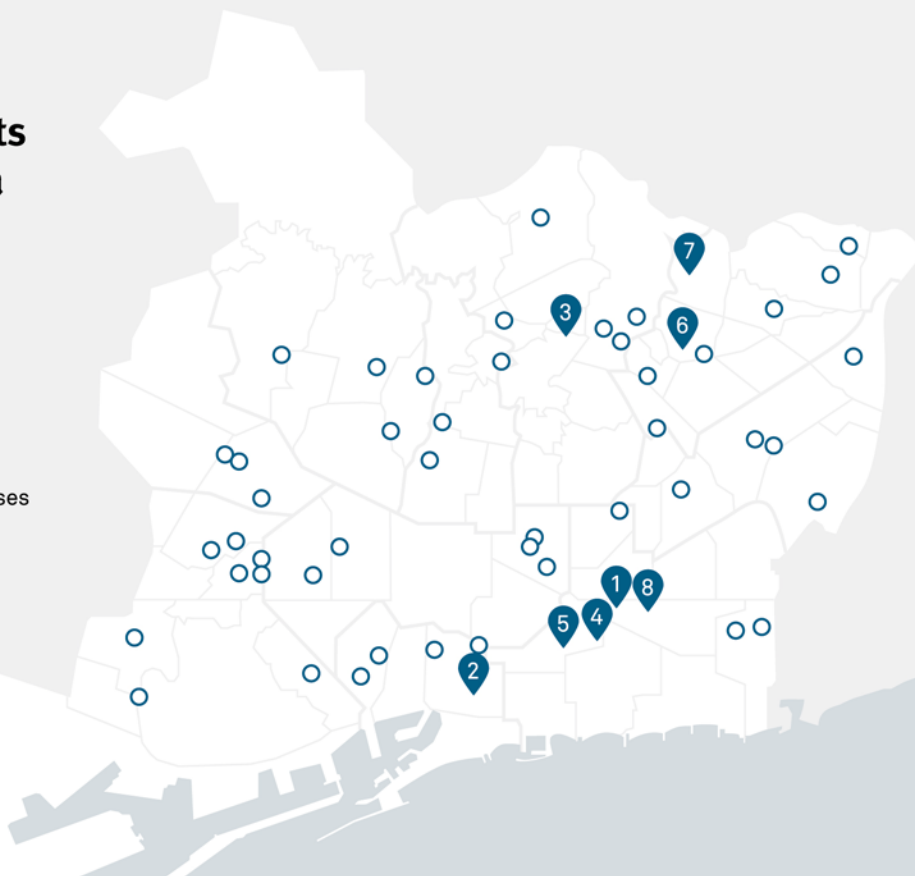


Formació tecnològica i gratuïta per a la ciutadania

barcelonactiva.cat/cibernarium

Xarxa d'equipaments de Barcelona Activa

- 1 Seu Central Barcelona Activa
Porta 22
Centre per a la Iniciativa
Emprenedora Glòries
Incubadora Glòries
- 2 Convent de Sant Agustí
- 3 Ca n'Andalet
- 4 Oficina d'Atenció a les Empreses
Cibernàrium
Incubadora MediaTIC
- 5 Incubadora Almogàvers
- 6 Parc Tecnològic
- 7 Nou Barris Activa
- 8 innoBA
- Punts d'atenció a la ciutat



© Barcelona Activa
Darrera actualització 2020

Cofinançat per:



UNIÓ EUROPEA
Fons Europeu de Desenvolupament Regional

Segueix-nos a les xarxes socials:



barcelonactiva.cat/cibernarium



[barcelonactiva](#)



[barcelonactiva](#)



[company/barcelona-activa](#)