

UNITAT 6

**WEBS D'INTERÈS I CONSELLS ÚTILS EN L'ÚS DEL NÚVOL
COM A EINA PER TREBALLAR I COMPARTIR**

CONTINGUTS

- 1. Com funciona el núvol**
- 2. Importància de les dades**



Autors: Simó Calafat / Víctor Herrera/ José Luis Guerrero

Data d'elaboració: Juny de 2020

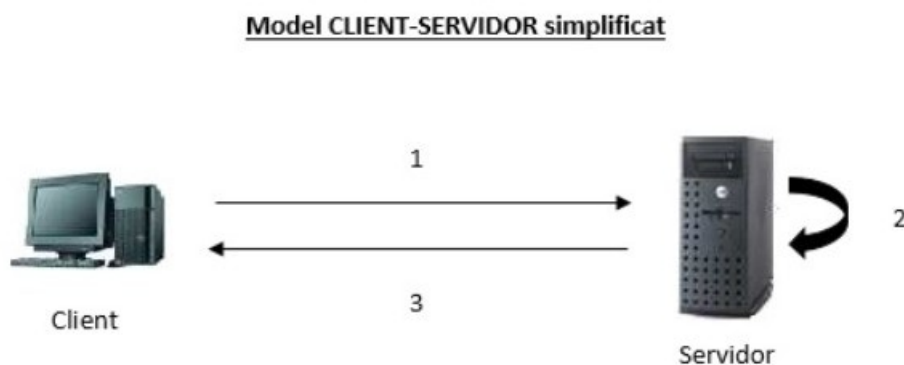
Aquesta obra es difon mitjançant la llicència [Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional License](https://creativecommons.org/licenses/by-nc-sa/4.0/).



1. 1. COM FUNCIONA EL NÚVOL

Per entendre com funciona el núvol, és necessari **abans comprendre com funciona internet** avui en dia.

Internet es basa en intercanvis de dades que segueixen un model **client-servidor**, en què es diferencien dos tipus d'actors: el **client**, que demana un recurs determinat a un altre actor que ell coneix, i el **servidor**, que li «serveix» el recurs demanat, si el té.



- | |
|---|
| <ol style="list-style-type: none">1. Vull mostrar https://ebapenlinia.caib.es . Servidor, tens aquesta URL?2. Cercaré a la meua base de dades l'URL <https://ebapenlinia.caib.es>.3. Client, aquí la tens |
|---|

Aquest model tan bàsic es complica molt avui en dia. Els problemes que sorgeixen, per exemple, si el client no sap a quin servidor demanar el recurs, s'acaben arreglant amb la incorporació de molts sistemes intermedis (p. ex.: DNS, CDN, etc.) que faciliten que l'intercanvi de dades es realitzi d'una manera correcta i eficaç.

El **núvol** és una **evolució del model client-servidor**, possible gràcies al desenvolupament de les noves tecnologies, que han permès un abaratiment dels costos i han aconseguit altes capacitats de processament de les màquines. Tot això fa que sigui més accessible, per a les grans empreses, disposar d'un gran nombre de servidors molt potents.

La informàtica en núvol (en anglès, *cloud computing*) és un paradigma que **ofereix com a servei tot el que permet un sistema informàtic** sense que l'usuari final en tingui coneixement. És, per tant, un paradigma que ofereix serveis de computació a través d'internet.



L'accés a aquests serveis es pot fer des de qualsevol dispositiu (ordinadors, mòbil, tauleta tàctil...) **sempre que estigui connectat a internet**. És així perquè el núvol acaba sent un conjunt de servidors connectats a internet, distribuïts per molts llocs del món, que processen les peticions que reben i donen uns serveis preconfigurats als usuaris o clients.

Els serveis que ofereixen es poden classificar en diferents categories, però sobretot en destaca un, el «**software com a servei**» (*SaaS*, o *Software as a Service*), per ser el més comú per als usuaris. El SaaS és un model de distribució de programari en què les aplicacions es troben allotjades a una companyia o proveïdor de servei i es posen a disposició dels usuaris a través d'internet. Per tant, **ofereix un software o programa informàtic a través d'internet**.

Altres dels serveis destacats són la «**plataforma com a servei**» (*PaaS* o *Platform as a Service*), que ofereix sistemes operatius, emmagatzematge i eines per a desenvolupadors, per exemple, i la «**infraestructura com a servei**» (*IaaS* o *Infrastructure as a Service*), que ofereix infraestructura de processament, com ara allotjament de pàgines web en servidors virtuals.

Un exemple de SaaS o d'aplicació al núvol és el cas del correu de Google. Gmail és una aplicació web que ofereix un servei de client de correu electrònic disponible a través d'internet. L'usuari accedeix al servei sense preocupar-se de quin servidor li respon, ni d'on està situat, l'únic que sap és que si accedeix a <https://www.gmail.com> tindrà a la seva disposició un servei de client de correu electrònic.

Segons la gestió, el núvol es pot categoritzar en tres tipus:



- **Públic:** està mantingut i gestionat per terceres persones, en el qual totes les dades i processos dels clients es mesclen als servidors, sistemes d'emmagatzematge i altres infraestructures, sense que els usuaris ho sàpiguen, i fa la impressió que l'usuari treballa tot sol a l'aplicació o servidor.
- **Privat:** l'usen les organitzacions que necessiten una protecció de dades alta. Es tracta d'una infraestructura gestionada per un sol client que controla quines aplicacions s'executen i on, i que els permeten controlar la privacitat de la informació.
- **Híbrid:** combina els models dels núvols públics i privats, en què un usuari és propietari o propietària d'unes parts i en comparteix unes altres, de manera controlada. Un exemple en serien els sistemes de correu electrònic empresarial, com GSuite, que dona un servei de client de correu electrònic basat en Gmail, però per a l'organització.

1.1. Computació a demanda

Tot i ser sobretot reconegut com el model que permet donar solucions de programari informàtic, el fet és que el núvol ha instaurat un nou **model de prestació de serveis** al sector informàtic.

Era habitual que una organització desitjés tenir un programa determinat que l'ajudés a fer la seva activitat diària. Per això, l'organització es posava en contacte amb una altra d'especialitzada en el desenvolupament de programari a mida i acordaven la contractació del projecte a partir d'una **negociació amb un pressupost** previ.

Ara el model ha canviat. Gràcies a les característiques del núvol, les organitzacions productores donen solucions genèriques i la contractació es fa mitjançant subscripcions mensuals o anuals basades en el **pagament per ús**, i així estableix un model d'obtenció de **computació a demanda** del client. Ex: el client ara necessita un programari per a 100 usuaris, doncs se li cobra X €/mes per 100 usuaris; un mes després vol 1.000 usuaris, doncs se li augmenta immediatament la capacitat de les màquines fins als 1.000 usuaris sense que la clientela noti res a 2X €/mes.

Aquest model de computació a demanda **recorda, en certa manera, els sistemes de subministrament d'electricitat, aigua i gas**, ja normalitzats a les vides diàries de les persones com a servei públic. Per poder viure, les persones, normalment, ja no es preocupen d'aconseguir un poc d'electricitat, un poc d'aigua o un poc de gas, sinó que es preocupen del que poden fer, donant per fet



que són recursos ja disponibles. El subministrament no es talla mai, sempre que no hi hagi incidències.

El concepte de la computació a demanda vol seguir el mateix exemple: una organització o una persona ja no s'ha de preocupar de com està muntat el sistema *software* que necessita, si és o no compatible amb la seva infraestructura o quanta gent qualificada necessita per poder tenir i mantenir una solució informàtica.

Al contrari, ara només cerca les solucions que se li ofereixen, paga una subscripció adaptada a les seves necessitats i es despreocupa totalment del manteniment i dels problemes que hi pugui haver per dedicar-se, pràcticament des del primer dia, a fer feina amb l'aplicació, de manera que guanya en temps i doblers, i de forma **molt més accessible** que abans.

1.2. Avantatges i inconvenients

El núvol representa un canvi respecte del model tradicional de fer feina. Tot i que té avantatges, s'han de tenir en compte els aspectes nous que també introdueix i amb els quals abans no es comptaven.

Els principals **avantatges** dels serveis que s'ofereixen a través del núvol són:

- **Escalabilitat:** permet modificar el tipus de subscripció o la *mida* del servei sol·licitat de manera quasi immediata i sense que el client se n'adoni.
- **Independència del lloc de feina físic:** com que la informació i el *software* es troben a Internet i s'hi pot accedir des de qualsevol dispositiu amb connexió, amb un portàtil o tauleta tàctil, qualsevol lloc es converteix en oficina, cosa que **fa més factible el teletreball**.
- **Estalvi en infraestructura:** tant a nivell econòmic com a nivell laboral, com que la informació i els sistemes informàtics ja no depenen de la mateixa organització, tota avaria, cost de personal, obsolescència de sistemes, etc., corre a càrrec de l'empresa proveïdora. Se segueix necessitant un dispositiu mòbil o un ordinador, però ja no importa tant la potència d'aquest mentre es pugui connectar a internet.
- **Despreocupació i recuperació davant caigudes de sistema:** la contractació d'un servei a través del núvol fa que el manteniment no caigui en mans de la mateixa organització, sinó de l'empresa proveïdora. Normalment, aquestes empreses tenen uns serveis de còpies de seguretat molt complets i els permeten recuperar-les ràpidament en cas de fallada del sistema. En qualsevol cas, es recomana sempre revisar quin servei s'està contractant per evitar sorpreses de darrera hora.



- **Actualitzacions ràpides i incloses dins el servei:** si el proveïdor del servei actualitza els seus sistemes, aquesta actualització s'aplica de manera immediata a tots els seus clients. D'aquesta manera, sempre es té el servei actualitzat a la darrera versió i funcionant per a tots amb independència de l'ordinador o dispositiu amb què es fa feina.
- **Fomenta el treball col·laboratiu:** a diferència dels programes informàtics antics, en què cada programa s'instal·lava (i es distribuïa als companys mitjançant CD, DVD, USB o disquets) i s'emprava de manera individual a cada ordinador per després posar les coses en comú, amb el núvol ho podem fer tot alhora. Al núvol només existeix un punt d'entrada, a través d'internet, des d'on tothom accedeix a les mateixes dades del mateix programa, de manera que tot canvi que faci una persona el pot veure un altre col·laborador a l'instant. A més, es pot facilitar la comunicació amb la inclusió de xats.



En canvi, els principals **problemes** que apareixen amb aquest model són els següents:

- **Total dependència en el proveïdor:** com ja s'ha comentat, **l'organització depèn del proveïdor**. Aquest fet provoca molts problemes que també podrien succeir si es tinguessin tots els sistemes en propietat (sense comptar amb el núvol), la diferència és que al cas del núvol **no hi podem fer res**, més que exigir una compensació econòmica gràcies als **acords de nivell del servei** (SLA, *Service Level Agreements*) o contractes amb les empreses proveïdores, que normalment no serveixen per arreglar els danys, però ja és una solució. Alguns d'aquests problemes són:
 - **On són, de qui són i com s'empren les dades:** tot i existir la reglamentació de protecció de dades, **confiar les dades a un extern** provoca una falta de confiança i una preocupació que serien inexistents si es tinguessin sistemes propietaris o propietàries. A més, si aquesta empresa proveïdora finalment es dissol, es fusiona o una altra la compra, s'hi suma el fet que es pot desconèixer qui és la nova propietària de les dades, què s'hi ha fet o què s'hi farà.



- **Seguretat:** a part que la informació es troba a Internet i no hi tenim control total, el fet de transmetre i rebre dades constantment per fer feina no deixa de preocupar pels temes de seguretat. És **informació que surt dels nostres ordinadors**, del nostre control, i que viatja a través de la xarxa. A més, un forat de seguretat en l'empresa proveïdora deixa els clients totalment desprotegits. Són coneguts els casos de grans empreses com eBay, Paypal, Playstation i fa molt poc temps Google, amb la xarxa social Google+.
- **Errada humana o desastre natural** que esborri dades o provoqui l'aturada del funcionament de totes les màquines del proveïdor sense una bona política de recuperació.
- **Falta de maduresa del servei ofert:** si un servei ofert, per exemple, el programari en línia, està constantment rebent actualitzacions gràfiques, la corba d'aprenentatge augmenta en quantitat considerable perquè una vegada t'acostumes a unes pantalles, canviar sol donar problemes. Una aplicació més madura veu reduïts el nombre de canvis i modificacions.
- **Total dependència de la connexió a internet:** com que tot és a internet, si no hi ha internet, no se sol poder actuar. És a dir, tot i que hi ha versions de SaaS que permeten fer feina fora connexió, el cas és que si no tenim una connexió a internet les possibilitats de fer feina es redueixen molt, ja que no es disposaran de les dades ni dels programes per treballar.



2. IMPORTÀNCIA DE LES DADES

Darrerament només es parla d'això. Dades, dades i més dades. Quin és el motiu perquè Google sempre tengui informació tan precisa? Quin és el motiu pel qual es dona programari gratuït? Per què Facebook té problemes amb el govern dels Estats Units? La resposta ve donada per les **dades massives** o **big data**, que es troben dins el marc del web 3.0 o web semàntic, que es va comentar a la unitat 1.

2.1. *Big Data*

El *Big data* descriu el **gran volum de dades**, estructurades, semiestructurades (programari, informes, fulls de càlcul) i no estructurades (documents, vídeos, àudios...) que s'han generat i que es troben a disposició de les empreses. Es tracta de conjunts de dades (o combinacions d'aquests) amb una mida, complexitat i velocitat de creixement tals, que es fa difícil capturar-los, gestionar-los, processar-los o analitzar-los mitjançant eines convencionals.

La utilitat principal de les dades massives és que proporcionen respostes a preguntes que les organitzacions ni s'havien plantejat. És a dir, donen un punt de referència que els permet **identificar els problemes d'una manera més comprensible, actuar més ràpid i trobar noves oportunitats**.

La principal problemàtica de les dades massives és l'extracció de dades reals i d'alta qualitat, de conjunts de dades tan massius, canviants i complicats, que es resumeix en les cinc ves que caracteritzen el *big data*: volum, velocitat, varietat, veracitat i valor.

L'aparició del *big data* ha significat un canvi molt important en el funcionament intern de les empreses i ha aconseguit que les grans entitats de diferents sectors, com a mínim, hagin estudiat l'adaptació a aquest nou model de dades. Tant és així, que és habitual que les grans empreses incorporin perfils de treballadors que comptin amb experiència o formació d'anàlisi i explotació de dades.

La recopilació de la informació adequada els permet l'obtenció de dades que poden tenir, o no, caràcter personal amb les quals poder fer negoci o, fins i tot, reconèixer patrons de comportament dels clients que els aporten un gran valor. Es poden veure exemples de l'explotació de dades massives en empreses turístiques, tecnològiques, equips de futbol, agències de màrqueting i, fins i tot, administracions públiques.

Fins l'arribada de les dades massives, aquesta informació s'emmagatzemava i recopilava a través de sistemes ERP i CRM, coneguts al món empresarial. Ara es permet obtenir informació addicional que es troba fora del món empresarial, com la que prové de les xarxes socials, dades estadístiques de tercers, etc.



D'on surten les dades, doncs? De les aplicacions webs o de les aplicacions i sistemes operatius mòbils. Quan una organització ofereix el seu sistema de manera gratuïta, es pot deduir que allò que no es paga econòmicament es paga d'una altra manera. Aquesta altra manera és amb informació, donada d'una manera voluntària o involuntària. Fins fa uns anys, cada versió del sistema operatiu de Microsoft, Windows, es podia aconseguir a través del pagament d'una llicència. Amb l'entrada de Windows 10, Microsoft va decidir regalar-lo de manera gratuïta durant un any.

Cada aplicació en fa un ús diferent, d'aquestes dades. Normalment, depenent del tipus d'aplicació, aquesta informació no permet identificar un usuari en particular més que per un número d'usuari il·legible i **s'aprofita per donar un servei molt útil** als usuaris. És el cas de Google Maps, que pot saber si hi ha un embús de trànsit a determinades carreteres, sense saber si hi ha en Josep o en Pepito. Simplement sap que allà hi ha molta gent per la quantitat de dispositius mòbils Android amb la ubicació activada i la velocitat amb què es mouen. Això, per desgràcia, no implica que fent una recerca exhaustiva no s'aconseguís saber la identitat de les persones; tot depèn de l'empresa proveïdora, Google en aquest cas, i de la reglamentació de protecció de dades actual, el Reglament general de protecció de dades (GDPR, de *General Data Protection Regulation*) de la Unió Europea.

Una anècdota curiosa que demostra el funcionament de Google Maps es va donar a principis de l'any 2020 quan l'artista Simon Weckert va idear el seu [Google Maps Hacks](#), que consistia en ajuntar noranta-nou telèfons mòbils en una carreta per modificar el comportament de l'aplicació.

Confiar, o no, doncs, en l'organització a la qual es confien les dades és un tema que sorgeix a causa de l'[escàndol de Facebook](#), acusada de filtrar dades d'usuaris a una consultora que va fer feina per a Donald Trump durant la campanya per ser elegir president dels EUA. Moltes empreses (entre les quals, Facebook) asseguren no vendre mai les dades a entitats externes, però no diuen si hi existeix un lloguer o no, cosa que augmenta la desconfiança.



2.2. Explotació de dades a les Administracions

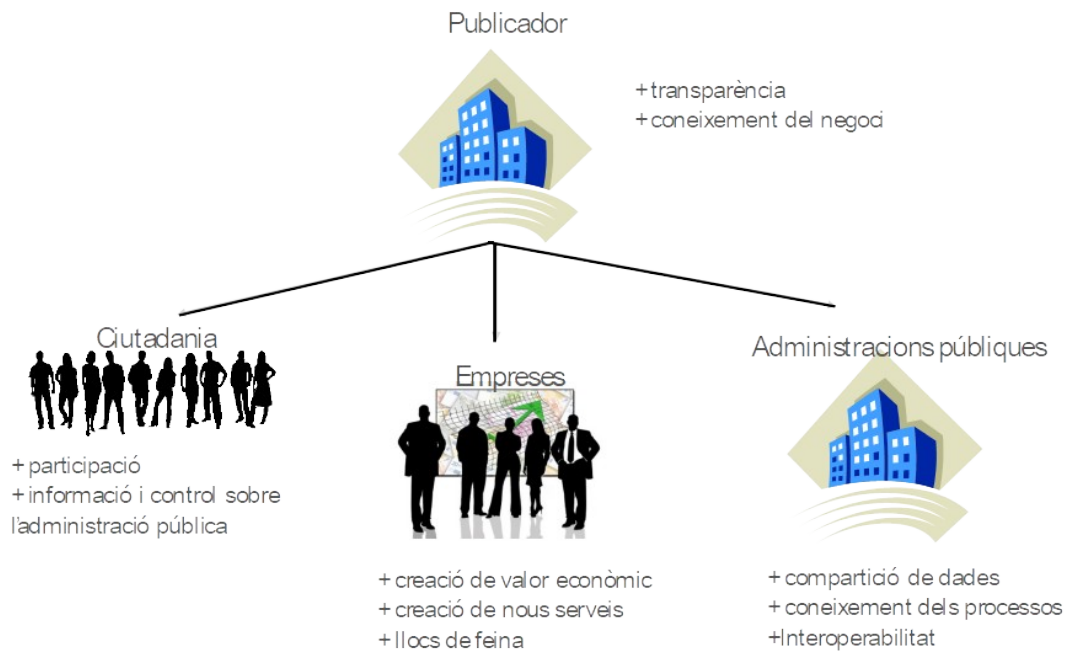
El mateix concepte de l'obtenció, processament i ús de les dades ha estat analitzat per les Administracions Públiques, donant peu al concepte de **dades obertes** (*Open data* en anglès).

Tal com indica la Llei 37/2007, sobre la reutilització de la informació del sector públic, la informació generada des de les instàncies públiques posseeix un gran interès per les empreses a l'hora d'operar en els seus àmbits d'actuació, contribuir al creixement econòmic i la creació de llocs de feina, i pels ciutadans com element de transparència i guia per a la participació ciutadana.

Les dades obertes, doncs, són les dades produïdes o recopilades per una administració pública o privada que **es posen a l'abast de la societat** perquè puguin ser utilitzades amb llibertat de manera senzilla i còmoda. Es consideren essencials per a la transparència de les administracions públiques i es reconeixen com a la matèria primera per a l'impuls de la innovació i el desenvolupament social i econòmic.

Els **beneficiaris** d'aquestes dades obertes poden ser:

- **El publicador de les dades:** És a dir, aquella entitat que és propietària de les dades (conselleria, ajuntament, organisme autònom, institut...) i decideix publicar-les. Aquesta actuació permet guanyar en transparència i l'ús i anàlisi de les seves dades que li ajuden a tenir un coneixement més profund del seu negoci.
- **Ciutadania:** Aquest guany en transparència permet a la ciutadania conèixer millor les seves institucions públiques i fer un control i seguiment de la seva activitat. A més, aconsegueix la seva participació, que apropa l'administració a la ciutadania.
- **Empreses:** Amb l'explotació de les dades publicades, les empreses poden crear nous serveis, desenvolupar productes i, alhora, permet la creació de nous llocs de feina i generació de valor econòmic.
- **Administracions Públiques:** Comparteixen i empen dades d'altres departaments dins la mateixa administració, de manera que ajuda a millorar la interoperabilitat entre les administracions i a afavorir l'eficiència de les mateixes.



Totes aquestes dades són publicades mitjançant conjunts de dades (o **datasets**), que representen les dades de manera **estructurada** i en format **reutilitzable** (csv, xls, etc.) i han de complir un requisits determinats:

- Han de ser de **qualitat**, és a dir, actualitzades. Les dades desactualitzades no són útils per a la societat.
- Han de ser **d'interès** per a la societat. De res serveix publicar dades que no importen a ningú o quasi ningú.
- Han de ser dades anomenades '**en brut**', és a dir, han de ser dades bàsiques. No poden ser el resultat de càlculs simples ni complexos.
- Han de garantir la **privacitat** de les persones, segons la legislació corresponent.

Els Portals de Dades Obertes

Els **datasets** de les entitats públiques o privades es poden trobar en els anomenats **Portals de Dades Obertes**, que són les pàgines web on es publiquen els catàlegs de dades que contenen tots els **datasets**, disponibles per ser consultats.

Les administracions fan servir aquests portals com a nodes centralitzadors de dades i permeten als publicadors dur una millor gestió i control de les dades publicades per tal d'estalviar en temps i esforç.



Es poden trobar portals en l'àmbit [europeu](#), [estatal](#) i en l'àmbit de les Comunitats Autònomes. A les Illes Balears, el portal de Dades Obertes el tenim a l'adreça <http://dadesobertes.caib.cat/> o <http://dadesobertes.caib.es/>

Des d'aquest Portal, hi ha disponible l'accés al [catàleg de dades](#), on es poden trobar molts *datasets* ja publicats que es poden consultar.

Cas d'ús: les dades COVID

La pandèmia del COVID-19, a l'Administració del Govern de les Illes Balears, ha impulsat un exemple clar d'un procés d'extracció, transformació i càrrega (de l'anglès ETL, *Extract, Transform, Load*) de dades per a una visualització efectiva.

D'aquesta manera, totes les dades obtingudes (*extretes*) per l'IBSalut, en relació als contagis del COVID-19 i publicades en el Portal de Dades Obertes [en un dels catàlegs](#), han estat processades (*transformades*) per després ser carregades (*load*) en un portal de Bussiness Intelligence com Power BI.



Aquestes eines permeten una millor visualització de les dades carregades i una major interactivitat, el que aporta una gran valor afegit a partir de les dades 'en brut' ofertades pel Portal de Dades Obertes.

