

Distribució de dades y *Map Reduce*

PAC2

Exercici 1 (25%)

Es desitja dissenyar una aplicació que implementa un generador d'alertes sobre epidèmies d'al·lèrgies de gramínies. Aquesta aplicació està orientada cap a qualsevol ciutadà amb aquesta malaltia. L'aplicació consulta diferents fonts d'informació, processa les dades i genera alertes. Un usuari pot establir les fonts d'informació que vol utilitzar, la importància que hauria de tenir cadascuna d'elles, i altres opcions de configuració com format de sortida, notificació de les alertes, notificacions, etc. Una alerta consta d'una llista de les fonts d'informació utilitzades, una llista de les notícies que s'han recuperat de cada font i que s'han utilitzat per generar l'alerta, i enllaços a les notícies de manera que l'usuari pugui consultar-les (per assegurar l'accés a la notícia, l'aplicació guarda sempre una còpia de seguretat de cada notícia utilitzada). El nombre de fonts d'informació i usuaris que haurà de gestionar el sistema és enorme.

Es demana analitzar:

1. El model de dades. En particular, estudiar la conveniència d'un model relacional o un de NoSQL i raonar quin model seria el més adequat.
2. A partir de l'arquitectura de distribució escollida:
 - a. Indicar els avantatges i inconvenients que té l'estratègia de fragmentació (partició) escollida respecte a altres.
 - b. Indicar els avantatges i inconvenients que té l'estratègia de replicació escollida respecte a altres.
3. El model transaccional més adequat per als requisits de l'aplicació descrita.

Exposeu la solució de forma argumentada en una pàgina com a màxim.

Exercici 2 (25%)

A partir de la lectura del llibre NoSQL Distilled indiqueu què us semblen les següents afirmacions.

Afirmació 1

El model de replicació mestre-esclau és una solució molt útil per escalar un sistema si s'han de realitzar lectures intensives sobre les dades.

Afirmació 2

El model de replicació peer-to-peer té com a principal avantatge assegurar la consistència de la informació.

Afirmació 3

La consistència de replicació consisteix en que dins d'una sessió d'usuari hi hagi una consistència de tipus "read-your-writes".

Afirmació 4

Una funció reductor combinable és una funció de reducció que té com a peculiaritat que la seva sortida coincideix amb la seva entrada.

Per a cadascuna de les afirmacions:

- Indiqueu si és certa o falsa. **No seran vàlides les respostes que no indiquin si l'afirmació és certa o és falsa.**
- Justificar breument la vostra resposta fent referència al llibre bàsic de l'assignatura o als apunts.

Es valorarà la concisió (una pàgina com a màxim per les 4 afirmacions juntes).

Exercici 3 (25%)

Un professor de la UOC vol explotar la informació procedent de la participació dels seus estudiants de l'assignatura del campus virtual. La informació disponible és la següent:

- Cada vegada que ha entrat al campus virtual.
- Cada vegada que s'ha descarregat un recurs de l'assignatura.
- Cada vegada que s'ha visualitzat un vídeo de l'assignatura.
- Cada missatge que s'ha enviat dins el campus virtual.
- Cada vegada que ha participat en un fòrum de l'assignatura.

Les preguntes que li agradaria consultar al professor són les següents:

1. De cada estudiant es vol conèixer informació agrupat per mesos sobre el nombre de vegades que ha entrat al campus virtual, nombre de missatges que ha enviat des del campus virtual, nombre de descàrregues realitzades de cada recurs de l'assignatura, i nombre de visualitzacions realitzades de cada vídeo. Un estudiant s'identifica per nom, cognoms i adreça electrònica.
2. De cada fòrum que té l'assignatura es vol conèixer l'activitat que s'ha produït en cada data (dia): nombre de fils oberts en el fòrum, nombre de rèpliques realitzades als fils, nombre d'estudiants diferents que han intervingut, i nom de

l'estudiant que més vegades ha intervingut en aquesta data. Un fòrum s'identifica pel seu nom i un id alfanumèric.

3. De cada recurs que té l'assignatura es vol conèixer informació sobre les descàrregues realitzades cada mes: nombre de descàrregues realitzades, nombre d'estudiants diferents que han descarregat el recurs, darrera data de descàrrega, i dia en el que més descàrregues es van realitzar. Un recurs s'identifica pel seu nom i un id alfanumèric.

Es demana: Dissenyar una solució per a una base de dades NoSQL orientada a documents que garanteixi la atomicitat de les operacions inserts i updates per a cada un dels casos plantejats. S'espera com a solució un agregat independent per a cadascuna de les consultes. Els agregats s'han d'escriure en format JSON (<https://jsonlint.com/>).

La resposta al problema consisteix en una descripció argumentada (màxim una pàgina) de la solució que es proposa a cada consulta i la il·lustració amb un exemple.

Exercici 4 (25%)

Considereu una taula d'una base de dades relacional d'una llibreria que emmagatzema informació sobre els llibres que han estat venuts a les llibreries de cadascuna de les editorials. Els camps de la taula Vendes són: isbn, títol, escriptor, editorial, exemplars, i guanys. Tots els camps són cadenes llevat dels camps exemplars i guanys que són enters positius.

Suposant que es tinguéssim una llista que contingués les tuples de la taula descrita, es demana simular utilitzant Map-Reduceix la següent consulta SQL:

```
SELECT escriptor, editorial, SUM(exemplars), SUM(guanys)
FROM vendes
GROUP BY escriptor, editorial
```

Descriviu els càlculs realitzats per les funcions de Map y Reduce, utilitzant el següent exemple que es mostra en format tabular

ISBN	TÍTOL	ESCRITOR	EDITORIAL	EXEMPLARS	GUANYS
234534J	El Quijote	Miguel de Cervantes	Anaya	45	546
134444H	Novelas ejemplares	Miguel de Cervantes	Anaya	34	234
244553I	Noche de reyes	William Shakespeare	Santillana	23	256

267787T	El mercader de Venècia	William Shakespeare	Alianza	15	345
234567R	Dràcula	Bram Stoker	Alianza	34	456
545545E	El Lazarillo de Tormes	Anónimo	Càtedra	35	245
987685E	El conde Lucanor	Don Juan Manuel	Càtedra	27	346

Criteris de valoració

Cadascun dels apartats tindrà un pes del 25%. Es valorarà, per a cada apartat, la validesa de la solució i la claredat de l'argumentació.

Format i data de lliurament

Heu d'enviar la PAC a la bústia de Lliurament i registre d'AC disponible a l'aula (apartat Avaluació). El format de l'arxiu que conté la vostra solució pot ser .pdf, .odt, .doc i .docx. Es recomana el format .pdf. Per a altres opcions, si us plau, contacteu prèviament amb el vostre professor col·laborador. El nom del fitxer ha de contenir el codi de l'assignatura, el vostre cognom i el vostre nom, així com el nom de l'activitat (PAC2). Per exemple nomcognom1_nosql_pac2.pdf

La data límit per lliurar la PAC2 és el **24 d'abril**.

Propietat intel·lectual

En presentar una pràctica o PAC que faci ús de recursos aliens, s'ha de presentar juntament amb ella un document en què es detallin tots ells, especificant el nom de cada recurs, el seu autor, el lloc on es va obtenir i el seu estatus legal: si l'obra està protegida pel copyright o s'acull a alguna altra llicència d'ús (Creative Commons, llicència GNU, GPL etc.). L'estudiant haurà de assegurar que la llicència que sigui no impedeix específicament seu ús en el marc de la pràctica o PAC. En cas de no trobar la informació corresponent haurà d'assumir que l'obra està protegida pel copyright. Caldrà, a més, adjuntar els fitxers originals quan les obres utilitzades siguin digitals, i el seu codi font, si així correspon.