

Tema 2-DATC 2018

Partiționarea datelor pentru o aplicație de tip cloud

În multe domenii, datele se partitionează. Acest lucru, se dorește a fi bun în următoarele cazuri:

1. Optimizarea performanței
2. Reducerea riscului de a pierde toate datele unui utilizator
3. Îmbunătățește securitatea.

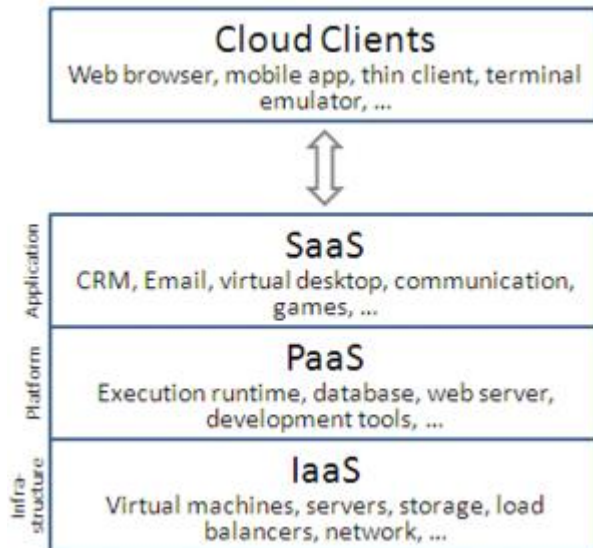
Se cunosc 3 strategii de partitionare: orizontală, verticală și funcțională.

Orizontală(sharding): fiecare partiție este stocată separat, dar toate partițiile au aceeași schemă. Fiecare conține un subset de date bine cunoscut.

Verticală: fiecare partiție păstrează un subset de câmpuri pentru articole în baza de date. Câmpurile se deosebesc prin frecvența lor de utilizare. Cele mai folosite sunt separat de cele accesate mai rar.

Funcțională: datele sunt împartite în funcție de cum sunt utilizate. De obicei, cele stocate într-o anumită parte tind spre același scop.

Găzduirea datelor într-o manieră centralizată poate fi un catalizator pentru sharing-ul de date dintre diferite domenii.



Infrastructure as a Service(IaaS):

Închirierea virtualizată și construirea unui sistem IT, pe care îl poți controla în totalitate. Acestea pot include: monitorizarea, securitatea, facturarea detaliată, rezistența la stocare, replicarea, recuperarea. Un utilizator poate implementa politici care să conducă la echilibrarea încărcării pentru a menține disponibilitatea și performanța aplicațiilor.

Platform as a Service(PaaS):

O platformă unde clienții pot dezvolta, rula și administra aplicații fără complexitatea construirii sau a întreținerii. Poate fi livrată ca: un cloud public de la furnizor, privată sau ca software implementat pe o infrastructură publică ca serviciu.

Software as a Service(SaaS):

Permite utilizatorilor să se aboneze și să utilizeze software-ul aplicațiilor în cloud. De asemenea utilizatorii pot să personalizeze software-ul pentru rolul pe care îl dețin. Prin suita SaaS modernă, vă puteți conecta cu ușurință întreaga afacere, de oriunde și de pe orice dispozitiv. Acest lucru elimină cheltuielile de achiziție, furnizare și întreținere hardware, precum și licențierea, instalarea și suportul de software.

Principiu în arhitectura software în care o singură instanță software rulează pe un server și deserveste clienți multipli. Într-o arhitectură multi-tenant, aplicațiile sunt proiectate astfel încât să existe o partiționare virtuală a datelor și configurațiilor => fiecare client (e.g. companie, individ,...) lucrează cu o instanță de aplicație virtuală personalizată.

Serviciile în cloud oferă găzduire pentru pagini web care au la bază o rețea mare de servere fizice. Clientul folosește un serviciu și plătește ce consumă.

Un API face posibilă exploatarea serviciilor furnizorului.

Proiectarea partițiilor pentru disponibilitate:

Datele partiționate trebuie împărțite astfel încât să poată fi gestionate independent.

Identificarea datelor cu o mare importanță și a celor cu mică importanță. Se stabilesc proceduri separate pentru monitorizarea diferitelor seturi de date.

În funcție de cât de des se folosesc datele, ele se pot grupa diferit sau împreună. Spre exemplu: partițiile care găzduiesc tranzacțiile e necesar să li se facă back-up mai frecvent decât cele care țin datele de logare.

Dacă o partiție se strică, datele trebuie să fie recuperate independent fără aplicațiile care țin de alte partiții.

Partiționare Azure:

Tabelele reprezintă unități de date stocate în tabel. Entitățile sunt definite de 3 proprietăți:

- i. PartitionKey: stochează string-urile unde sunt identificate partițiile pe care sunt entitățile.
- ii. RowKey: stochează string-ul unic unde sunt localizate partițiile.
- iii. Timestamp: ține transabilitatea unei entități. Ajută dezvoltatorul să știe când a fost modificată ultima dată.

Entity Group Transactions= un set de operații stocate care sunt implementate cu aceeași cheie de partiție.

- nu comprimate mai mult de 100 de stocări și nu poate avea o dimensiune mai mare de 4GB.
- furnizează un tabel limitat în consistență, izolație.
- avantaj economic: o tranzacție este facturată ca o singură operațiune de stocare, indiferent de numărul de operațiuni pe care le conține.

Range Partitions:

- utilizarea valorilor unice pentru entități face ca fiecare entitate să aibă o partiție. Dacă acestea cresc sau scad, Azure e posibil să creeze o gamă de partiții.
- îmbunătățesc performanța interogărilor din interval.

În concluzie, e nevoie ca datele să fie partiționate pentru o bună administrare și pentru o recuperare rapidă în caz de pierdere.