

DEPOZITE DE DATE SQL

PAVLOV Alexandr

Universitatea Tehnică a Moldovei

Abstract: În acest articol este analizat serviciului oferit de către compania Microsoft - depozitele de date SQL AZURE. Sînt descrise punctele forte ale serviciului, este descrisă structura sistemului, sînt arătate aplicațiile care pot să interacționeze cu acest serviciu. Este arătată importanța depozitelor de date în procesarea datelor de volum extrem de mare.

Cuvinte cheie: SQL, depozite de date, AZURE, Transact-SQL, blob, performanță, cloud, noduri, scalabilitate, calcule, stocare.

1 Introducere

Depozitul de date SQL AZURE este o bază de date distribuită de clasa enterprise, capabilă să proceseze un volum extrem de mare de date relaționale și non-relaționale. Este prima soluție industrială care combină capabilitățile SQL cu posibilitatea de a crește, a se compacta, și a face pauză în câteva secunde. Acest depozit este bine integrat în AZURE, care permite de a implementa soluții în câteva secunde. În plus, serviciul conține soluții de reducere a timpului consumat pentru patch-uri, backup-uri și mentenanță. Depozitul de date SQL efectuează backup-uri automat.

2 Aspecte forte și caracteristici ale SQL AZURE

Printre aspectele forte importante ale depozitelor de date SQL AZURE putem enumera următoarele:

1. procesarea paralelă a volumelor de petabyte's,
2. scalarea independentă pentru calcule și stocarea datelor,
3. suportul limbajului Transact-SQL pentru date relaționale cât și non-relaționale,
4. lucrează perfect cu instrumentele: Power BI, Machine Learning, HDInsight și Data Factory.

Prima soluție pentru întreprinderi în vederea stocării datelor pe cloud, care prevăd creșterea și descreșterea volumului de date instantaneu. Din punctul de vedere istoric, depozitele de date au nevoie de combinații de depozitare și calcule, de multe ori suprautilizînd resursele hardware costisitoare. Cu depozitele de date SQL a serviciului AZURE, datele și calculele se pot scala independent. Serviciul dat permite lansarea dinamică, creșterea, descreșterea și chiar pauza în calcule, obținînd cel mai bun avantaj din punct de vedere a performanței și prețului. De asemenea, depozitele de date SQL utilizează puterea limbajului T-SQL, care permite ușor integrarea rezultatelor interogărilor asupra datelor relaționale obținute în depozitul d-stră, și a datelor non-relaționale în depozitele AZURE blob.

Livrarea soluțiilor performante calitative și la preț rezonabil. Depozitele de date SQL utilizează arhitectura Microsoft pentru procesarea paralelă masivă (MPP - massively parallel processing), unde se plătește doar pentru servicii, nu și pentru hardware. Performanța este punctul forte pentru depozitul bazelor de date din punct de vedere a scalării și a calculelor.

Plățiți pentru aceia ce aveți nevoie atunci când aveți nevoie. Depozitele de date SQL scalează independent calculele și spațiul de stocare, astfel încât se plătește pentru performanța interogărilor, și numai atunci când aveți nevoie de ele. Spre deosebire de alte depozite de date, care necesită zile întregi pentru redimensionare, depozitele de date SQL oferă posibilitatea de scalare doar în câteva minute. Se obțin toate avantajele stocării pe cloud și interogări de calcul în baza cerințelor variate de performanță. Când calculele sunt oprite, se plătește doar pentru stocarea datelor, și astfel de oportunitate oferă doar AZURE.

Stocarea hibridă pe cloud. Depozitele de date SQL sunt bazate pe motorul bazelor de date relaționale a MS SQL Server și include atfel de caracteristic deja bine cunoscute ca proceduri, funcții definite de utilizator, partiționarea tabelară, indexarea și sortarea. Depozitele de date SQL utilizează același nucleu al BD ca și SQL Server și Platforma Analitică a companiei Microsoft, care asigură coerența dintre cloud și stocarea locală a datelor, oferind alegerea perfectă pentru stocarea datelor în bază de performanță, necesități de securitate și scalare.

Accesul la interogări prin intermediul a 2 tipuri de date: relaționale și non-relaționale. Utilizînd PolyBase și Transact-SQL interogările acoperă ambele tipuri de date conform formatelor Hadoop. O singură comandă Transact-SQL combină datele non-relaționale din depozitele AZURE Bblobs cu tabelele relaționale din depozitul de date. Acest lucru permite data mining-ul, construirea rapoartelor și analiza complexă utilizînd posibilitățile cunoscute ale limbajului Transact-SQL.

Dezvoltarea infrastructurii securizate cu acces rapid fără plată pentru mentenanță. Depozitele de date SQL oferă posibilitatea de implementare a soluțiilor doar în câteva secunde. Serviciul este complet menținut și oferă reducerea timpului pentru actualizarea sistemului și mentenanța lui. Depozitele conțin backup-uri automate care restabilesc datele în spațiul de stocare AZURE, revenind la punctele de restaurare instantanee a bazei de date.

3 Arhitectura Depozitului de date SQL AZURE

Depozitul de date SQL rulează utilizînd arhitectura companiei Microsoft pentru procesarea paralelă masivă a datelor MPP. Această arhitectură profită de îmbunătățiri de performanță de depozitare a datelor, cât și posibilitatea de a

scala ușor și a paraleliza operațiile de calcul a interogărilor SQL complexe. Nucleul este constituit din 4 componente (figura1):

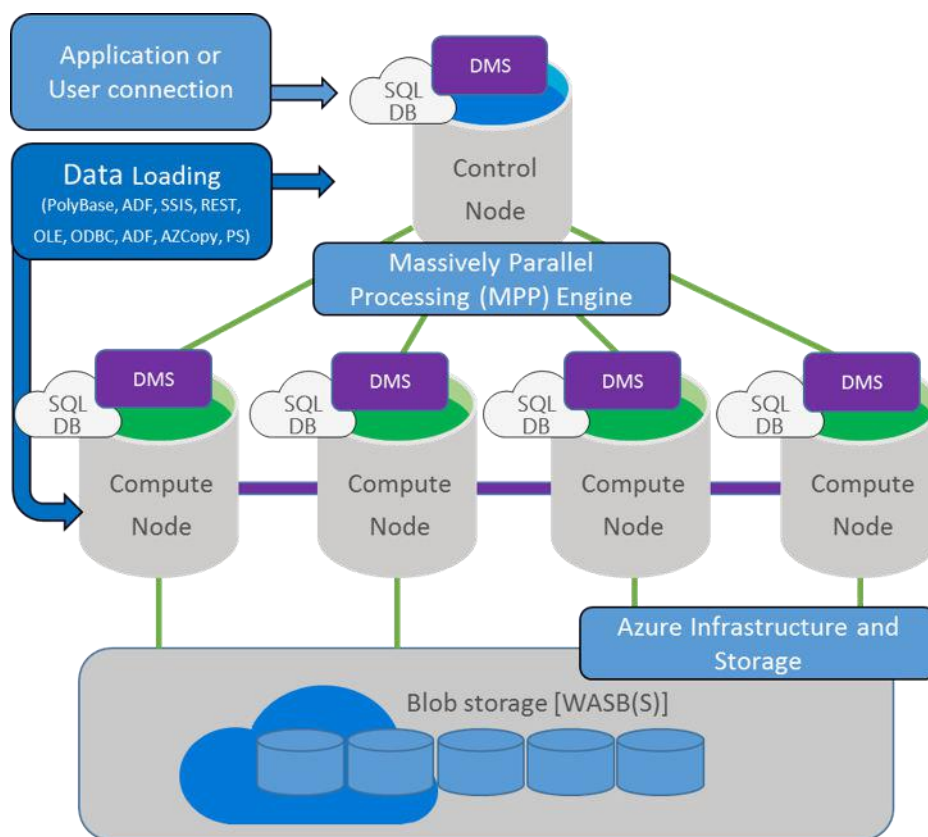


Figura 1 – Arhitectura MPP

1. Nodurile de control (Control Nodes): permite conectarea la nod utilizând depozitul bazelor de date cu orice instrumente de dezvoltare, încărcare sau business intelligence. Nodul de control este SQL Database, care se aseamănă și se simte exact așa ca o bază de date SQL standardă. Totuși, dincolo de interfață grafică se ascunde acel mecanism de coordonare a datelor și calcule, care are loc în sistem.

2. Nodurile de calcul (Compute Nodes): ca și nodurile de control, ele sunt propulsate utilizând SQL Database. Lucrul lor constă în efectuarea calculelor în sistem. De fiecare dată când datele sunt încărcate în depozit, ele sunt distribuite pe toate nodurile. După ce nodul de control primește comanda, el o desparte în părți mici și le trimite fiecărui nod de calcul. După finisarea calculelor, nodurile de calcul transmit datele nodului de control, unde are loc agregarea rezultatelor parțiale înainte de a întoarce rezultatul.

3. Spațiul de stocare (Storage): toată memoria de stocare pentru depozitul bazei de date reprezintă AZURE Storage Blobs standard. Aceasta înseamnă datele din nodurile de calcul sunt citite și scrise direct din/în Blob-uri.

4. Serviciile de transfer a datelor (Data Movement Services): permit de a controla nodurile la nivel de comunicare între ele. De asemenea, ele permit nodurilor de calcul să transfere informația dintr-un nod în altul cu scopul de efectuarea instrucțiunilor cât mai rapid.

4 Optimizările motorului

Procesarea paralelă masivă oferă posibilitatea de divizare a problemelor. Atunci când operațiile sunt executate de diferite noduri de calcul, fiecare nod procesează informația, apoi o transmite nodului de control unde are loc agregarea, după finisarea căreia ea este transmisă înapoi utilizatorilor. Acest proces poate fi optimizat prin următoarele metode:

1. Depozitul de date SQL utilizează un optimizator avansat de interogări și o statistică complexă. Utilizând informația despre dimensiunile și mărimea și distribuția datelor, serviciul este capabil să optimizeze interogări distribuite în baza evaluării costurilor operațiilor de interogare specifice.

2. Pentru a crea planuri optimizate pentru interogări, depozitul de date SQL încorporează algoritmi și tehnici avansate, care distribuie eficient datele printre resursele de calcul pentru realizarea interogărilor. Aceste operații sunt incluse în serviciul de transfer a datelor al depozitului și optimizările au loc automat.

5 Scalabilitate

Arhitectura depozitelor introduce stocarea datelor separat de operațiile de calcul, astfel scalabilitatea are loc independent pentru fiecare. Structura de implementare simplă și rapidă a bazei de date permite calculul suplimentar

care este disponibil la momentul potrivit. Complementar acestui fapt este utilizarea AZURE Storage Blobs. Aceasta ne dă nu numai o stocare stabilă și replicată, dar și o infrastructură pentru extindere fără efort la un preț scăzut. Utilizând această combinație dintre stocarea scalabilă pe cloud și calculele pe AZURE, depozitul SQL permite de a plăti pentru o stocare pentru interogări performante atunci când este nevoie de ea. Volumul calculelor este măsurat în unități specifice depozitului DWUs (DWUs - Data Warehouse Units). DWUs este o măsură a puterii de bază pe care o posedă depozitul de date, care este proiectat pentru a vă asigura că aveți o valoare standardă a performanței asociată depozitului la un moment de timp. DWU-urile sunt utilizate pentru a asigura următoarele:

- posibilitatea de a scala eficient datele fără grijă de software-ul sau hardware-ul utilizat;
- înțelegerea nivelului performanței în DWU până la eventuala scalare a depozitului;
- ajustarea arhitecturii de bază a serviciului fără a afecta performanța volumului dumneavoastră de muncă;
- deoarece am îmbunătățit rapid performanța în depozitul SQL, putem fi siguri că facem asta astfel ca scalarea sa afecteze uniform sistemul.

6 Hibrid

Utilizarea depozitelor în combinație cu PolyBase oferă posibilitatea de acoperire a întregii ecosisteme, facilitând abilitatea de setare a scenariilor hibride cu date non-relaționale. PolyBase asigură transferarea ușoară de la diferite surse, utilizând aceleași comenzi ca și limbajul Transact-SQL. PolyBase utilizează tabelele externe pentru accesarea datelor non-relaționale. Definițiile tabelor sînt stocate în depozitul SQL și pot fi accesate prin SQL sau instrumente ca și datele relaționale. Polybase este agnostic despre integrarea sa. El oferă aceleași funcționalități la toate sursele ce îl suportă. Datele citite de PolyBase pot fi de diferite formate, incluzând fișiere delimitate sau fișiere de tip ORC.

7 Integrare

Depozitele de date SQL sunt bazate pe motorul bazelor de date relaționale SQL Server, verificat deja, care aduce o mulțime de funcționalități corespunzătoare unui depozit de date industrial. Dacă cunoșteți limbajul Transact-SQL, transferul cunoștințelor la depozitul SQL nu este o problemă:

Pentru cei avansați dar și pentru începători, este pusă la dispoziție documentația cu exemple. Depozitele de date SQL utilizează sintaxa limbajului Transact-SQL pentru majoritatea operațiilor și mențin un set larg de construcții SQL tradiționale, precum proceduri stocate, funcții definite de utilizator, de partiționare tabelor, indecsarea. În plus, depozitele de date SQL oferă funcționalități de ultima generație, așa ca indecși de grup, integrarea cu PolyBase și Auditul de Date.

Instrumentarul integrat în depozitele de date SQL include instrumentele tradiționale din SQL Server și instrumentele pentru cloud: SQL Server Analysis Services, Integration Services, Reporting Services, AZURE Data Factory, Stream Analytics, Machine Learning, Power BI.

8 Instrumente de Management

Portalul AZURE reprezintă un portal de management, care permite crearea, actualizarea și ștergerea bazelor de date, cât și resurselor bazelor de date. Acest instrument este efectiv atunci când numărul de baze de date nu este mare și operațiile necesită operativitate în timp. SQL Server Data Tools în Visual Studio este un instrument client, care permite gestionarea și crearea aplicațiilor utilizând conexiunea cu depozitele de date SQL. De asemenea, el permite de a vizualiza conținutul bazelor de date, de a aplica diferite script-uri pentru configurare. Linia de comandă face parte din PowerShell sau sqlcmd, care permite gestionarea cu resursele AZURE. Se recomandă pentru gestionarea unui număr mare de servere, implementarea modificărilor în resursele mediului de producție, implementarea și scriptarea acestor resurse la nivel automat.

9 Concluzii

Depozitul de date Microsoft SQL AZURE este un serviciu de baze de date relaționale, construit pe tehnologiile SQL Server. Acesta este caracterizat printr-un grad ridicat de disponibilitate și scalabilitate, alcătuit din mai multe entități, care este găzduit de Microsoft în cloud. AZURE servește la implementarea mai multor baze de date și la asigurarea accesului la acestea. Astfel, dezvoltatorii nu sunt nevoiți să instaleze programe software, să le configureze, să le gestioneze sau să le aplice corecții. Serviciul beneficiază de un grad mare de disponibilitate și de o toleranță ridicată la erori, ceea ce face ca administrarea fizică să nu fie necesară. Acest serviciu este recomandat companiilor mari, care au de gestionat un volum mare de informație, deoarece scalabilitatea paralelă asigură acest lucru independent pentru calcule și operații aritmetice. În plus toată informația este stocată pe servere internaționale, reducând riscul de pierdere a datelor la minim.

Bibliografie

1. SQL Data Warehouse: An elastic data warehouse as a service with enterprise-class features [Resursă electronică]. -Regim de acces: <https://AZURE.microsoft.com/en-us/services/sql-data-warehouse/>
2. AZURE, SQL Data Warehouse Documentation [Resursă electronică]. -Regim de acces: <https://AZURE.microsoft.com/en-us/documentation/services/sql-data-warehouse/>
3. Drew DiPalma. What is AZURE SQL Data Warehouse? [Resursă electronică]. -Regim de acces: <https://AZURE.microsoft.com/en-us/documentation/articles/sql-data-warehouse-overview-what-is/>