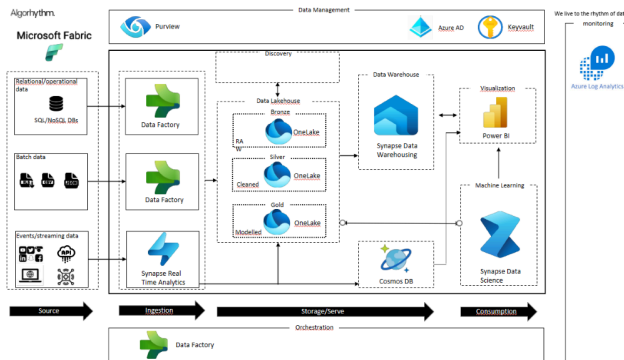


Data Architecture

Tuesday, March 19, 2024 10:33 AM

Gelieve aanpassingen maken in de [Referentiearchitectuur ppt](#) en deze screenshot & beschrijvingen te updaten.



Source: Fabric kan zowel gestructureerde, semi-gestructureerde als ongestructureerde datatypes aan. Van tabellen uit een SAP HANA database, tot CSV's en JSON als streaming data van IOT-devices.

Ingestion: De ingestion layer in Fabric bestaat uit 2 grote services: Data Factory en Synapse Real Time Analytics. Voor relationele/operationele data en batch data gebruiken we Data Factory terwijl we voor de streaming Synapse Real Time Analytics gebruiken. We gebruiken de real time analytics omdat deze een gemakkelijke integratie heeft met Azure Cosmos DB, deze is gemaakt voor het snel werken en opslaan van streaming data.

Storage/Serve:

- OneLake** zorgt voor de opslag van alle data binnen Fabric. Het heeft overeenkomsten met OneDrive, maar is specifiek ontworpen voor een cloudomgeving, met ondersteuning voor tabellen, semi-gestructureerde data, parquet-bestanden, en meer. Ook maakt OneLake het makkelijk om data te managen en te delen binnen je organisatie. Wat OneLake ook kenmerkt is het principe "single copy of data". Data kan in meerdere analytische engines (Spark, PowerBI en meer gebruikt worden zonder redundante kopies te maken. OneLake kan ook gebruikt worden om streaming data op te slaan, maar dit hoeft niet. Data kan rechtstreeks gevisualiseerd worden zonder opgeslagen te worden.
- Synapse Data Warehousing:** Na alle gewenste ETL-processen kan gestructureerde data ingeladen worden in het data warehouse. Binnen Fabric wordt Synapse Data Warehousing hiervoor gebruikt. Synapse Data Warehousing werkt naadloos samen met PowerBI en Azure Data Factory. Synapse Data Warehousing schaalbaar en support serverloze queries, wat wil zeggen dat je queries kan uitvoeren zonder hier vooraf resources hoeft voor toe te wijzen.

Consumption:

- PowerBI:** PowerBI is de visualisatietool van Microsoft die volledig geïntegreerd is in Fabric. Wat deze integratie betekent is dat PowerBI gebruik kan maken van de kracht van Fabric, data hoeft niet geconnecteerd worden met Import mode of Direct Query, dit zorgt voor redundante kopies van data en dure, trage refreshes van je dashboard. Door de integratie in Fabric kan data rechtstreeks gevisualiseerd worden uit het Data Warehouse.
- Synapse Data Science:** Data kan rechtstreeks uit OneLake ingeladen worden en adhv notebooks wordt data exploratie en snelle inzichten gefaciliteerd.

Monitoring:

- Azure Log Analytics:** is een krachtige tool binnen het Azure-portal waarmee je query's kunt bewerken en uitvoeren op gegevens die zijn opgeslagen in de Azure Monitor Logs. Via deze monitor kan je ze gemakkelijk bekijken en opvolgen.

Data Management:

- Purview:** Purview is een data management tool binnen Fabric dat het toelaat een 'plattegrond' te maken van van alle opgeslagen data in de cloud en on premises. Purview maakt het eenvoudig om security compliance te verzekeren waar gevoelige data is opgeslagen. Purview zorgt voor manieren om het risico op verlies van data te verminderen en biedt tools voor risk management aan. Wat interessant is voor een data engineer is dat het automatisch een data catalog aanlegt met metadata, dit is vergelijkbaar met de voor ons bekende Glue Data Catalog in AWS.
- Keyvault:** Keyvault is een cloudservice waarmee je geheimen kan opslaan en ophalen. Denk aan wachtwoorden, API-keys, etc... Er zijn vaults waarin je deze geheimen kan opslaan die tenants kunnen ophalen, een tenant is een organisatie bv Microsoft services gebruikt. Binnen de tenant heb je verschillende rollen en groepen om zo de toegang te verdelen en zodat iedereen alleen toegang heeft tot hetgeen wat zij nodig hebben.
- Microsoft Entra ID (Vroeger Azure AD):** is net zoals Keyvault een cloudservice met gelijkaardige focus toch verschillen ze een beetje. Microsoft Entra ID legt de focus vooral op gebruikersidentiteiten, OAuth-tokens en toegang terwijl Keyvault de focus legt op geheimen en andere gevoelige informatie maar ook wat toegangbeheer.

