Architecture deep dive: Azure Synapse Analytics

Azure Synapse Analytics brings together enterprise data warehousing and big data analytics in a unified environment, enabled by a flexible and scalable architecture. Azure Synapse Analytics is highly extensible and allows for the easy integration of various data sources and provides a powerful analytics platform.

Deeply integrated Azure Spark and SQL engines

Azure Synapse Analytics connects various analytics runtimes (such as Apache Spark and SQL) through a single platform to enhance collaboration among data professionals working on advanced analytics solutions.

Flexible to bring together relational and non-relational data

Easily query files in the data lake with the same service used to build data warehousing solutions. Orchestrate pipelines to perform common analytics scenarios without writing a line of code. By defining a pipeline, a data source can be linked from the Orchestrate hub and copied into an Azure Data Lake Storage account without any coding.

Powerful performance

SQL pools can process high concurrent and complex T-SQL queries across petabytes of data to serve BI tools and applications. Cloud elasticity enables Azure Synapse Analytics to quickly increase and decrease its capacity according to demand with no impact to infrastructure availability, performance, or security. Best of all, you only pay for your actual usage.

Flexible Spark integration

The Apache Spark engine simplifies the use of big data by removing the complexity of setup and cluster tuning. The power of Spark with built-in support for Azure Machine Learning addresses the full range of analytics needs, from data engineering to data science, using PySpark (Python), Spark (Scala), .NET Spark (C#), and Spark SQL. This enables enhanced collaboration, as you can now use T-SQL on both your data warehouse and embedded Spark engine.

Fast, easy to explore and analyze data

The serverless endpoint in Synapse SQL makes it fast and easy to explore and analyze over data in a data lake —with no infrastructure to set up or manage. With T-SQL, you can run serverless queries over the data lake without provisioning or managing any infrastructure. By eliminating the overhead of data center management and operations for the data warehouse, you can reallocate resources to where value is produced and focus on using the data warehouse to deliver the best information and insight. This lowers overall total cost of ownership and provides better cost control over operating expenses.

Highly scalable, hybrid data integration capability

Data ingestion and operationalization are accelerated through automated data pipelines. While the volume of data in a data warehouse typically grows with the age of the establishment, the scalability of Azure Synapse matches this by incrementally adding resources as data and workloads increase.

Get more details in a free technical e-book from Packt.