**Understand Azure Synapse serverless SQL pool capabilities and use cases**

Completed100 XP

* 5 minutes

Azure Synapse Analytics is an integrated analytics service that brings together a wide range of commonly used technologies for processing and analyzing data at scale. One of the most prevalent technologies used in data solutions is SQL - an industry standard language for querying and manipulating data.

**Serverless SQL pools in Azure Synapse Analytics**

Azure Synapse SQL is a distributed query system in Azure Synapse Analytics that offers two kinds of runtime environments:

* **Serverless SQL pool**: on-demand SQL query processing, primarily used to work with data in a data lake.
* **Dedicated SQL pool**: Enterprise-scale relational database instances used to host data warehouses in which data is stored in relational tables.

In this module, we'll focus on serverless SQL pool, which provides a pay-per-query endpoint to query the data in your data lake. The benefits of using serverless SQL pool include:

* A familiar Transact-SQL syntax to query data in place without the need to copy or load data into a specialized store.
* Integrated connectivity from a wide range of business intelligence and ad-hoc querying tools, including the most popular drivers.
* Distributed query processing that is built for large-scale data, and computational functions - resulting in fast query performance.
* Built-in query execution fault-tolerance, resulting in high reliability and success rates even for long-running queries involving large data sets.
* No infrastructure to setup or clusters to maintain. A built-in endpoint for this service is provided within every Azure Synapse workspace, so you can start querying data as soon as the workspace is created.
* No charge for resources reserved, you're only charged for the data processed by queries you run.

**When to use serverless SQL pools**

Serverless SQL pool is tailored for querying the data residing in the data lake, so in addition to eliminating management burden, it eliminates a need to worry about ingesting the data into the system. You just point the query to the data that is already in the lake and run it.

Synapse SQL serverless resource model is great for unplanned or "bursty" workloads that can be processed using the always-on serverless SQL endpoint in your Azure Synapse Analytics workspace. Using the serverless pool helps when you need to know exact cost for each query executed to monitor and attribute costs.

**Note**

Serverless SQL pool is an analytics system and is not recommended for OLTP workloads such as databases used by applications to store transactional data. Workloads that require millisecond response times and are looking to pinpoint a single row in a data set are not good fit for serverless SQL pool.

Common use cases for serverless SQL pools include:

* **Data exploration**: Data exploration involves browsing the data lake to get initial insights about the data, and is easily achievable with Azure Synapse Studio. You can browse through the files in your linked data lake storage, and use the built-in serverless SQL pool to automatically generate a SQL script to select TOP 100 rows from a file or folder just as you would do with a table in SQL Server. From there, you can apply projections, filtering, grouping, and most of the operation over the data as if the data were in a regular SQL Server table.
* **Data transformation**: While Azure Synapse Analytics provides great data transformations capabilities with Synapse Spark, some data engineers might find data transformation easier to achieve using SQL. Serverless SQL pool enables you to perform SQL-based data transformations; either interactively or as part of an automated data pipeline.
* **Logical data warehouse**: After your initial exploration of the data in the data lake, you can define external objects such as tables and views in a serverless SQL database. The data remains stored in the data lake files, but are abstracted by a relational schema that can be used by client applications and analytical tools to query the data as they would in a relational database hosted in SQL Server.

**Khả năng và Trường hợp sử dụng của Serverless SQL Pool trong Azure Synapse Analytics**

**Azure Synapse Analytics** là một dịch vụ phân tích tích hợp, kết hợp nhiều công nghệ phổ biến được sử dụng để xử lý và phân tích dữ liệu ở quy mô lớn. Một trong những công nghệ phổ biến nhất được sử dụng trong các giải pháp dữ liệu là SQL - ngôn ngữ chuẩn của ngành để truy vấn và thao tác dữ liệu.

**Serverless SQL pool trong Azure Synapse Analytics**

**Azure Synapse SQL** là một hệ thống truy vấn phân tán trong Azure Synapse Analytics, cung cấp hai loại môi trường runtime:

* **Serverless SQL pool:** Xử lý truy vấn SQL theo yêu cầu, chủ yếu được sử dụng để làm việc với dữ liệu trong data lake.
* **Dedicated SQL pool:** Các phiên bản cơ sở dữ liệu quan hệ dành cho doanh nghiệp được sử dụng để lưu trữ các kho dữ liệu, trong đó dữ liệu được lưu trữ trong các bảng quan hệ.

Trong phần này, chúng ta sẽ tập trung vào **serverless SQL pool**, cung cấp điểm cuối trả tiền theo lượt truy vấn để truy vấn dữ liệu trong data lake của bạn. Những lợi ích của việc sử dụng serverless SQL pool bao gồm:

* **Ngôn ngữ Transact-SQL quen thuộc:** Cho phép truy vấn dữ liệu tại chỗ mà không cần sao chép hoặc tải dữ liệu vào một kho lưu trữ chuyên biệt.
* **Kết nối tích hợp:** Kết nối từ nhiều công cụ phân tích kinh doanh và truy vấn ad-hoc, bao gồm các trình điều khiển phổ biến nhất.
* **Xử lý truy vấn phân tán:** Được xây dựng cho dữ liệu quy mô lớn và các hàm tính toán, dẫn đến hiệu suất truy vấn nhanh.
* **Khả năng chịu lỗi thực thi truy vấn tích hợp:** Đảm bảo độ tin cậy và tỷ lệ thành công cao ngay cả đối với các truy vấn chạy lâu liên quan đến bộ dữ liệu lớn.
* **Không cần thiết lập cơ sở hạ tầng hoặc quản lý cụm:** Một điểm cuối tích hợp cho dịch vụ này được cung cấp trong mọi không gian làm việc Azure Synapse, vì vậy bạn có thể bắt đầu truy vấn dữ liệu ngay khi không gian làm việc được tạo.
* **Không tính phí cho tài nguyên được đặt trước:** Bạn chỉ bị tính phí cho dữ liệu được xử lý bởi các truy vấn bạn chạy.

**Khi nào nên sử dụng serverless SQL pool**

Serverless SQL pool được thiết kế để truy vấn dữ liệu nằm trong data lake, do đó ngoài việc loại bỏ gánh nặng quản lý, nó còn loại bỏ nhu cầu lo lắng về việc đưa dữ liệu vào hệ thống. Bạn chỉ cần trỏ truy vấn vào dữ liệu đã có trong data lake và chạy nó.

Mô hình tài nguyên serverless của Synapse SQL rất phù hợp cho khối lượng công việc không được lên kế hoạch trước hoặc theo kiểu "bùng nổ" (bursty) có thể được xử lý bằng điểm cuối serverless luôn bật trong không gian làm việc Azure Synapse Analytics của bạn. Sử dụng serverless pool hữu ích khi bạn cần biết chi phí chính xác cho mỗi truy vấn được thực thi để theo dõi và phân bổ chi phí.

Lưu ý:

Serverless SQL pool là một hệ thống phân tích và không được khuyến nghị cho khối lượng công việc OLTP như cơ sở dữ liệu được các ứng dụng sử dụng để lưu trữ dữ liệu giao dịch. Khối lượng công việc yêu cầu thời gian phản hồi bằng mili giây và tìm kiếm một hàng duy nhất trong một tập dữ liệu không phù hợp với serverless SQL pool.

**Các trường hợp sử dụng phổ biến cho serverless SQL pool:**

* **Khám phá dữ liệu:** Khám phá dữ liệu bao gồm việc duyệt qua data lake để có được những hiểu biết ban đầu về dữ liệu và có thể dễ dàng đạt được với Azure Synapse Studio. Bạn có thể duyệt qua các file trong kho lưu trữ data lake được liên kết và sử dụng serverless SQL pool tích hợp để tự động tạo một tập lệnh SQL để chọn 100 hàng TOP từ một file hoặc thư mục giống như cách bạn làm với bảng trong SQL Server. Từ đó, bạn có thể áp dụng chiếu (projections), lọc (filtering), nhóm (grouping) và hầu hết các hoạt động trên dữ liệu như thể dữ liệu nằm trong bảng SQL Server thông thường.
* **Chuyển đổi dữ liệu:** Mặc dù Azure Synapse Analytics cung cấp các khả năng chuyển đổi dữ liệu tuyệt vời với Synapse Spark, một số kỹ sư dữ liệu có thể thấy chuyển đổi dữ liệu dễ dàng hơn bằng cách sử dụng SQL. Serverless SQL pool cho phép bạn thực hiện chuyển đổi dữ liệu dựa trên SQL; tương tác hoặc như một phần của pipeline dữ liệu tự động.

**Lưu trữ dữ liệu**: Serverless SQL pool có thể được sử dụng để lưu trữ dữ liệu theo kiểu "data lakehouse", kết hợp các khía cạnh của kho dữ liệu và data lake. Dữ liệu được lưu trữ trong các file trong data lake, nhưng được truy cập bằng các bảng và truy vấn SQL quen thuộc.

* **Phân tích thời gian thực**: Serverless SQL pool có thể được sử dụng để phân tích dữ liệu thời gian thực bằng cách kết hợp nó với các dịch vụ như Azure Stream Analytics.
* **Kết luận**: Serverless SQL pool là một công cụ mạnh mẽ và linh hoạt có thể được sử dụng cho nhiều trường hợp sử dụng phân tích dữ liệu.