**How Azure Synapse Analytics works**

Completed100 XP

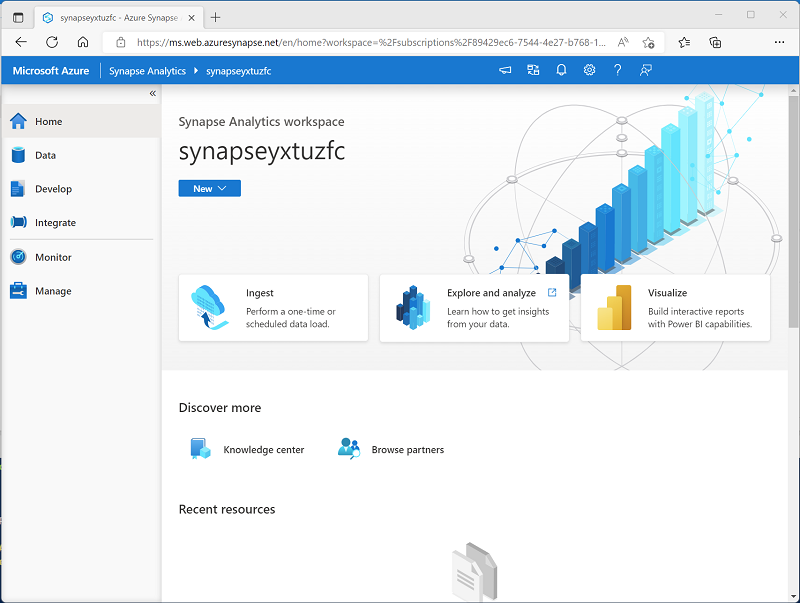
* 8 minutes

To support the analytics needs of today's organizations, Azure Synapse Analytics combines a centralized service for data storage and processing with an extensible architecture through which *linked services* enable you to integrate commonly used data stores, processing platforms, and visualization tools.

**Creating and using an Azure Synapse Analytics workspace**

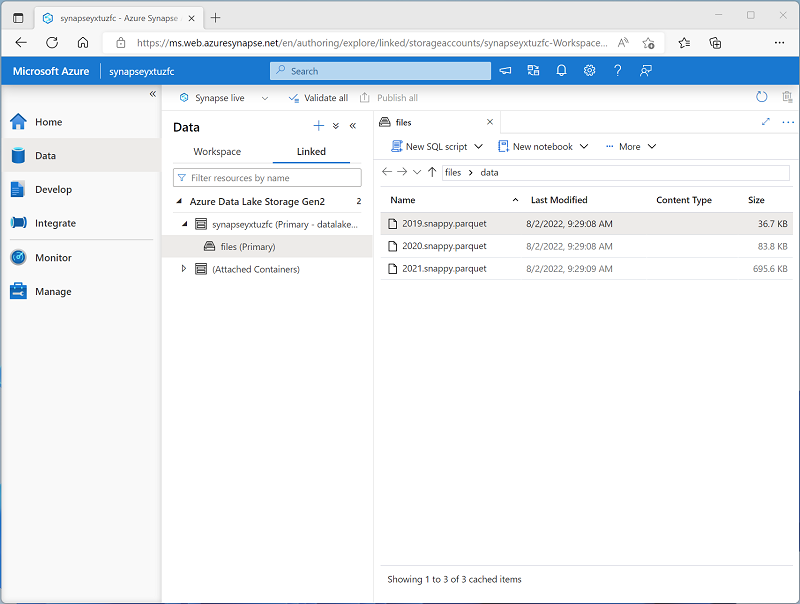
A Synapse Analytics *workspace* defines an instance of the Synapse Analytics service in which you can manage the services and data resources needed for your analytics solution. You can create a Synapse Analytics workspace in an Azure subscription interactively by using the Azure portal, or you can automate deployment by using Azure PowerShell, the Azure command-line interface (CLI), or with an Azure Resource Manager or Bicep template.

After creating a Synapse Analytics workspace, you can manage the services in it and perform data analytics tasks with them by using *Synapse Studio*; a web-based portal for Azure Synapse Analytics.

[](https://learn.microsoft.com/en-us/training/wwl-data-ai/introduction-azure-synapse-analytics/media/synapse-studio.png#lightbox)

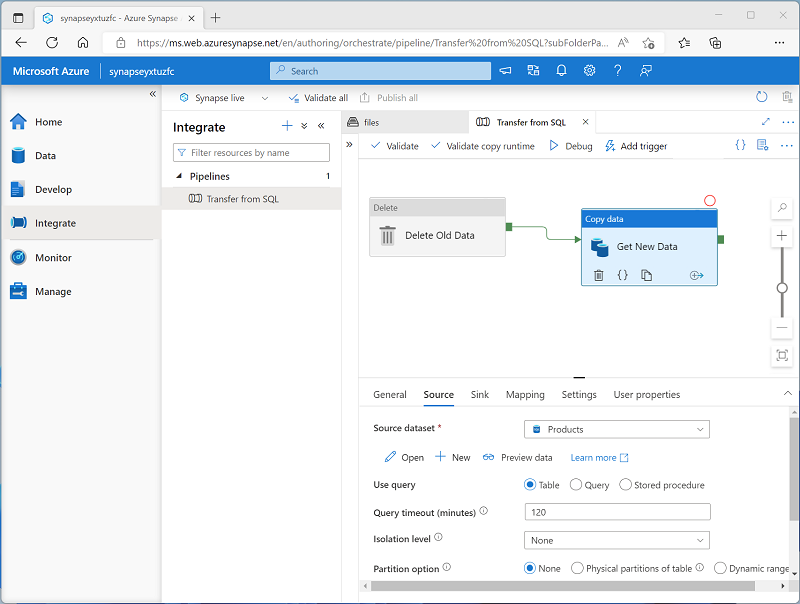
**Working with files in a data lake**

One of the core resources in a Synapse Analytics workspace is a *data lake*, in which data files can be stored and processed at scale. A workspace typically has a default data lake, which is implemented as a linked service to an Azure Data Lake Storage Gen2 container. You can add linked services for multiple data lakes that are based on different storage platforms as required.

[](https://learn.microsoft.com/en-us/training/wwl-data-ai/introduction-azure-synapse-analytics/media/data-lake-store.png#lightbox)

**Ingesting and transforming data with pipelines**

In most enterprise data analytics solutions, data is extracted from multiple operational sources and transferred to a central data lake or data warehouse for analysis. Azure Synapse Analytics includes built-in support for creating, running, and managing *pipelines* that orchestrate the activities necessary to retrieve data from a range of sources, transform the data as required, and load the resulting transformed data into an analytical store.

[](https://learn.microsoft.com/en-us/training/wwl-data-ai/introduction-azure-synapse-analytics/media/synapse-pipeline.png#lightbox)

**Note**

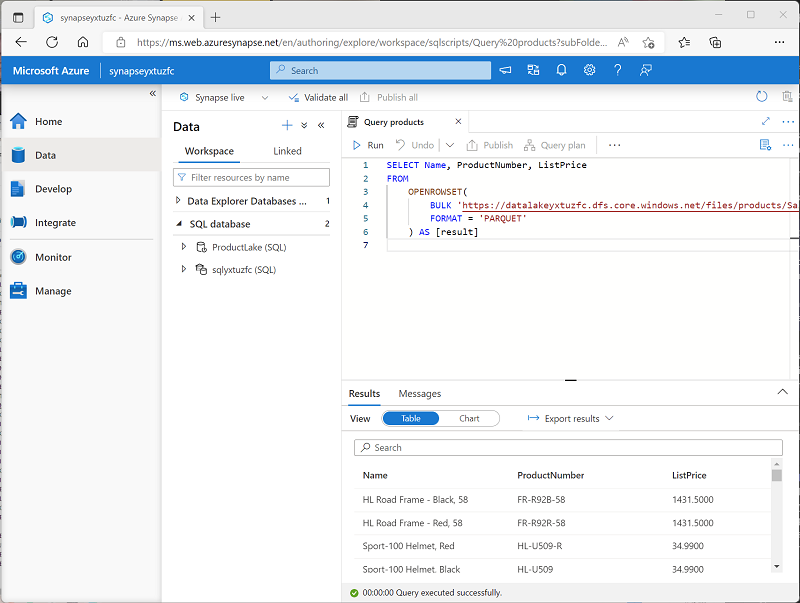
Pipelines in Azure Synapse Analytics are based on the same underlying technology as Azure Data Factory. If you are already familiar with Azure Data Factory, you can leverage your existing skills to build data ingestion and transformation solutions in Azure Synapse Analytics.

**Querying and manipulating data with SQL**

Structured Query Language (SQL) is a ubiquitous language for querying and manipulating data, and is the foundation for relational databases, including the popular Microsoft SQL Server database platform. Azure Synapse Analytics supports SQL-based data querying and manipulation through two kinds of SQL *pool* that are based on the SQL Server relational database engine:

* A built-in *serverless* pool that is optimized for using relational SQL semantics to query file-based data in a data lake.
* Custom *dedicated* SQL pools that host relational data warehouses.

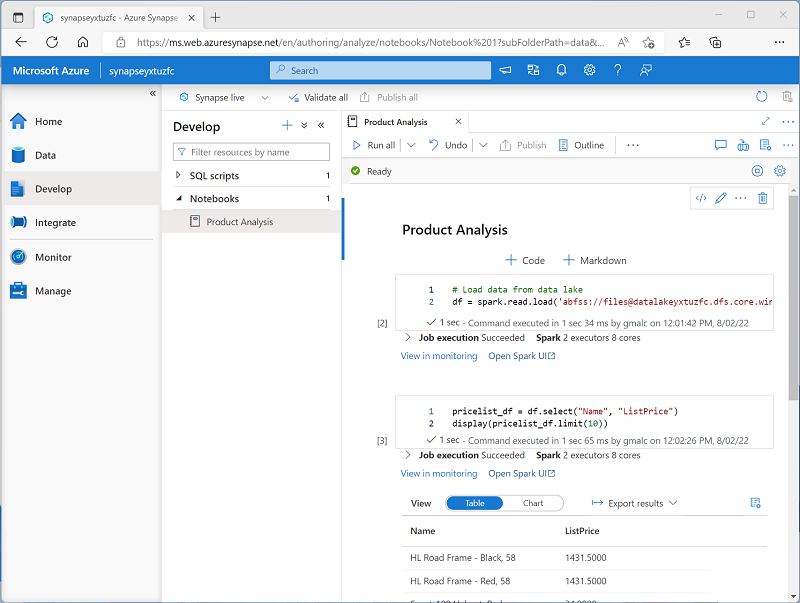
The Azure Synapse SQL system uses a distributed query processing model to parallelize SQL operations, resulting in a highly scalable solution for relational data processing. You can use the built-in serverless pool for cost-effective analysis and processing of file data in the data lake, and use dedicated SQL pools to create relational data warehouses for enterprise data modeling and reporting.

[](https://learn.microsoft.com/en-us/training/wwl-data-ai/introduction-azure-synapse-analytics/media/synapse-sql.png#lightbox)

**Processing and analyzing data with Apache Spark**

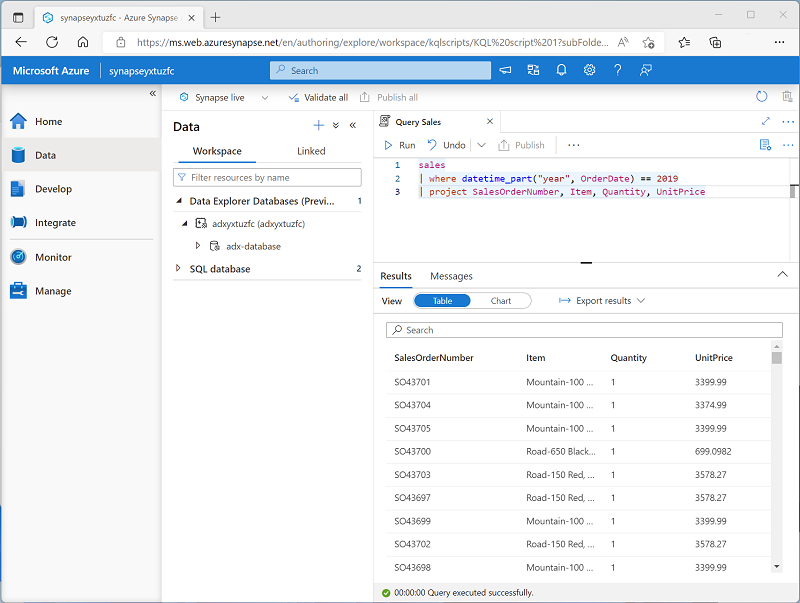
Apache Spark is an open source platform for big data analytics. Spark performs distributed processing of files in a data lake by running jobs that can be implemented using any of a range of supported programming languages. Languages supported in Spark include Python, Scala, Java, SQL, and C#.

In Azure Synapse Analytics, you can create one or more Spark pools and use interactive *notebooks* to combine code and notes as you build solutions for data analytics, machine learning, and data visualization.

[](https://learn.microsoft.com/en-us/training/wwl-data-ai/introduction-azure-synapse-analytics/media/synapse-spark.png#lightbox)

**Exploring data with Data Explorer**

Azure Synapse Data Explorer is a data processing engine in Azure Synapse Analytics that is based on the Azure Data Explorer service. Data Explorer uses an intuitive query syntax named Kusto Query Language (KQL) to enable high performance, low-latency analysis of batch and streaming data.

[](https://learn.microsoft.com/en-us/training/wwl-data-ai/introduction-azure-synapse-analytics/media/synapse-data-explorer.png#lightbox)

**Integrating with other Azure data services**

Azure Synapse Analytics can be integrated with other Azure data services for end-to-end analytics solutions. Integrated solutions include:

* **Azure Synapse Link** enables near-realtime synchronization between operational data in Azure Cosmos DB, Azure SQL Database, SQL Server, and Microsoft Power Platform Dataverse and analytical data storage that can be queried in Azure Synapse Analytics.
* **Microsoft Power BI** integration enables data analysts to integrate a Power BI workspace into a Synapse workspace, and perform interactive data visualization in Azure Synapse Studio.
* **Microsoft Purview** integration enables organizations to catalog data assets in Azure Synapse Analytics, and makes it easier for data engineers to find data assets and track data lineage when implementing data pipelines that ingest data into Azure Synapse Analytics.
* **Azure Machine Learning** integration enables data analysts and data scientists to integrate predictive model training and consumption into analytical solutions.

**Azure Synapse Analytics hoạt động như thế nào**

Để hỗ trợ các nhu cầu phân tích của các tổ chức ngày nay, Azure Synapse Analytics kết hợp một dịch vụ tập trung cho lưu trữ và xử lý dữ liệu với kiến trúc mở rộng thông qua các **dịch vụ được liên kết** (linked services) cho phép bạn tích hợp các kho lưu trữ dữ liệu, nền tảng xử lý và công cụ trực quan hóa thông dụng.

**Tạo và sử dụng không gian làm việc Azure Synapse Analytics**

**Không gian làm việc Synapse Analytics** (Synapse Analytics workspace) định nghĩa một phiên bản của dịch vụ Synapse Analytics, nơi bạn có thể quản lý các dịch vụ và tài nguyên dữ liệu cần thiết cho giải pháp phân tích của mình. Bạn có thể tạo không gian làm việc Synapse Analytics trong đăng ký Azure באופן tương tác bằng cách sử dụng cổng Azure hoặc tự động triển khai bằng Azure PowerShell, Giao diện dòng lệnh Azure (CLI) hoặc với mẫu Azure Resource Manager hoặc Bicep.

Sau khi tạo không gian làm việc Synapse Analytics, bạn có thể quản lý các dịch vụ trong đó và thực hiện các tác vụ phân tích dữ liệu với chúng bằng cách sử dụng **Synapse Studio**, một cổng web dành cho Azure Synapse Analytics.

**Làm việc với các file trong data lake**

Một trong những tài nguyên cốt lõi trong không gian làm việc Synapse Analytics là **data lake**, nơi các file dữ liệu có thể được lưu trữ và xử lý ở quy mô lớn. Thông thường, một không gian làm việc có một data lake mặc định, được triển khai dưới dạng dịch vụ được liên kết với một vùng chứa Azure Data Lake Storage Gen2. Bạn có thể thêm các dịch vụ được liên kết cho nhiều data lake khác nhau dựa trên các nền tảng lưu trữ khác nhau theo yêu cầu.

**Truyền và chuyển đổi dữ liệu với pipelines**

Trong hầu hết các giải pháp phân tích dữ liệu của doanh nghiệp, dữ liệu được trích xuất từ nhiều nguồn vận hành khác nhau và được chuyển đến một data lake hoặc data warehouse trung tâm để phân tích. Azure Synapse Analytics bao gồm hỗ trợ tích hợp để tạo, chạy và quản lý **pipelines**, điều phối các hoạt động cần thiết để truy xuất dữ liệu từ nhiều nguồn khác nhau, chuyển đổi dữ liệu theo yêu cầu và tải dữ liệu đã được chuyển đổi vào kho lưu trữ phân tích.

**Lưu ý:**

Pipelines trong Azure Synapse Analytics dựa trên cùng nền tảng cơ bản như Azure Data Factory. Nếu bạn đã quen thuộc với Azure Data Factory, bạn có thể tận dụng các kỹ năng hiện có để xây dựng các giải pháp thu thập và chuyển đổi dữ liệu trong Azure Synapse Analytics.

**Truy vấn và thao tác với dữ liệu bằng SQL**

Ngôn ngữ truy vấn cấu trúc (SQL) là một ngôn ngữ phổ biến để truy vấn và thao tác dữ liệu, và là nền tảng cho các cơ sở dữ liệu quan hệ, bao gồm nền tảng cơ sở dữ liệu Microsoft SQL Server phổ biến. Azure Synapse Analytics hỗ trợ truy vấn và thao tác dữ liệu dựa trên SQL thông qua hai loại **SQL pool** dựa trên công cụ cơ sở dữ liệu quan hệ SQL Server:

* **Serverless pool** tích hợp được tối ưu hóa để sử dụng ngữ nghĩa SQL quan hệ để truy vấn dữ liệu dựa trên file trong data lake.
* **Dedicated SQL pool** tùy chỉnh lưu trữ các kho dữ liệu quan hệ.

Hệ thống Azure Synapse SQL sử dụng mô hình xử lý truy vấn phân tán để song song hóa các hoạt động SQL, mang lại giải pháp có khả năng mở rộng cao cho xử lý dữ liệu quan hệ. Bạn có thể sử dụng serverless pool tích hợp để phân tích và xử lý dữ liệu file trong data lake một cách tiết kiệm chi phí và sử dụng dedicated SQL pool để tạo các kho dữ liệu quan hệ cho mô hình hóa dữ liệu và báo cáo của doanh nghiệp.

**Xử lý và phân tích dữ liệu với Apache Spark**

Apache Spark là một nền tảng mã nguồn mở cho phân tích dữ liệu lớn. Spark thực hiện xử lý phân tán các file trong data lake bằng cách chạy các tác vụ có thể được triển khai bằng bất kỳ ngôn ngữ lập trình được hỗ trợ nào. Các ngôn ngữ được hỗ trợ trong Spark bao gồm Python, Scala, Java, SQL và C#.

Trong Azure Synapse Analytics, bạn có thể tạo một hoặc nhiều **Spark pool** và sử dụng các **sổ ghi chép** (notebooks) tương tác để kết hợp mã và ghi chú khi bạn xây dựng các giải pháp cho phân tích dữ liệu, học máy và trực quan hóa dữ liệu.

**Khám phá dữ liệu với Data Explorer**

Azure Synapse Data Explorer là một công cụ xử lý dữ liệu trong Azure Synapse Analytics dựa trên dịch vụ Azure Data Explorer. Data Explorer sử dụng cú pháp truy vấn trực quan có tên Kusto Query

**Azure Synapse Analytics không hoạt động đơn độc!** Nó có thể kết nối với các dịch vụ dữ liệu khác của Microsoft Azure để tạo ra các giải pháp phân tích dữ liệu toàn diện.

**Các dịch vụ tích hợp chính bao gồm:**

* **Azure Synapse Link:** Cho phép đồng bộ gần như theo thời gian thực giữa dữ liệu hoạt động (operational data) và dữ liệu phân tích. Dữ liệu hoạt động có thể đến từ các nguồn như Azure Cosmos DB, Azure SQL Database, SQL Server và Microsoft Power Platform Dataverse.
* **Microsoft Power BI:** Cho phép các nhà phân tích dữ liệu kết hợp Power BI với Synapse Analytics để tạo các hình ảnh biểu đồ (visualization) tương tác cho dữ liệu.
* **Microsoft Purview:** Giúp các tổ chức quản lý dữ liệu trong Azure Synapse Analytics dễ dàng hơn. Purview giúp theo dõi nguồn gốc của dữ liệu (data lineage) và giúp các kỹ sư dữ liệu tìm thấy dữ liệu cần thiết.
* **Azure Machine Learning:** Cho phép các nhà phân tích dữ liệu và các nhà khoa học dữ liệu tích hợp việc xây dựng và sử dụng các mô hình dự báo vào các giải pháp phân tích.