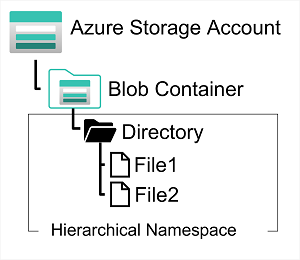
**Explore Azure DataLake Storage Gen2**

Completed100 XP

* 3 minutes

Azure Data Lake Store (Gen1) is a separate service for hierarchical data storage for analytical data lakes, often used by so-called *big data* analytical solutions that work with structured, semi-structured, and unstructured data stored in files. Azure Data Lake Storage Gen**2** is a newer version of this service that is integrated into Azure Storage; enabling you to take advantage of the scalability of blob storage and the cost-control of storage tiers, combined with the hierarchical file system capabilities and compatibility with major analytics systems of Azure Data Lake Store.



Systems like Hadoop in Azure HDInsight, Azure Databricks, and Azure Synapse Analytics can mount a distributed file system hosted in Azure Data Lake Store Gen2 and use it to process huge volumes of data.

To create an Azure Data Lake Store Gen2 files system, you must enable the **Hierarchical Namespace** option of an Azure Storage account. You can do this when initially creating the storage account, or you can upgrade an existing Azure Storage account to support Data Lake Gen2. Be aware however that upgrading is a one-way process – after upgrading a storage account to support a hierarchical namespace for blob storage, you can’t revert it to a flat namespace.

**khám phá Azure Data Lake Storage Gen2**

**Azure Data Lake Storage Gen2** là phiên bản mới hơn của dịch vụ lưu trữ dữ liệu cho các giải pháp phân tích Big Data. Nó tích hợp với Azure Blob Storage, mang lại cho bạn những lợi ích:

* **Khả năng mở rộng của lưu trữ blob:** Lưu trữ lượng lớn dữ liệu một cách linh hoạt.
* **Kiểm soát chi phí với các tầng lưu trữ:** Chọn tầng lưu trữ phù hợp với nhu cầu để tiết kiệm chi phí.
* **Hệ thống tệp phân cấp:** Dễ dàng tổ chức dữ liệu với cấu trúc thư mục.
* **Tương thích với các hệ thống phân tích chính của Azure:** Sử dụng các công cụ phân tích Big Data quen thuộc của Azure để xử lý dữ liệu.

**Azure Data Lake Storage Gen2 hoạt động như thế nào?**

1. **Hệ thống phân tích Big Data:** Các hệ thống như Apache Hadoop trong Azure HDInsight, Azure Databricks và Azure Synapse Analytics có thể kết nối với Data Lake Storage Gen2.
2. **Hệ thống tệp phân tán:** Data Lake Storage Gen2 hoạt động như một hệ thống tệp phân tán, cho phép các hệ thống phân tích xử lý song song dữ liệu trên nhiều máy chủ.
3. **Lưu trữ Blob với không gian tên phân cấp:** Data Lake Storage Gen2 được xây dựng trên nền tảng Azure Blob Storage, nhưng kích hoạt tính năng "Hierarchical Namespace" (Không gian tên phân cấp) để cung cấp cấu trúc thư mục cho dữ liệu.

**Cách tạo Azure Data Lake Storage Gen2:**

1. **Tạo tài khoản lưu trữ Azure:** Bật tùy chọn "Hierarchical Namespace" khi tạo tài khoản lưu trữ Azure mới.
2. **Nâng cấp tài khoản lưu trữ hiện có:** Nếu đã có tài khoản lưu trữ Azure, bạn có thể nâng cấp nó để hỗ trợ Data Lake Gen2. **Lưu ý:** Đây là quá trình một chiều, không thể khôi phục về lưu trữ blob dạng phẳng sau khi nâng cấp.

**Tóm tắt:**

Azure Data Lake Storage Gen2 là giải pháp lưu trữ dữ liệu lý tưởng cho các dự án phân tích Big Data trên nền tảng đám mây Azure. Nó kết hợp khả năng mở rộng, tiết kiệm chi phí và tích hợp với các hệ thống phân tích hàng đầu, giúp bạn xử lý và phân tích dữ liệu hiệu quả.