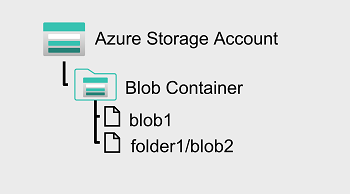
**Compare Azure Data Lake Store to Azure Blob storage**

Completed100 XP

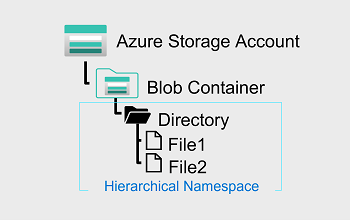
* 5 minutes

In Azure Blob storage, you can store large amounts of unstructured ("object") data in a flat namespace within a blob container. Blob names can include "/" characters to organize blobs into virtual "folders", but in terms of blob manageability the blobs are stored as a single-level hierarchy in a flat namespace.



You can access this data by using HTTP or HTTPs

Azure Data Lake Storage Gen2 builds on blob storage and optimizes I/O of high-volume data by using a hierarchical namespace that organizes blob data into *directories*, and stores metadata about each directory and the files within it. This structure allows operations, such as directory renames and deletes, to be performed in a single atomic operation. Flat namespaces, by contrast, require several operations proportionate to the number of objects in the structure. Hierarchical namespaces keep the data organized, which yields better storage and retrieval performance for an analytical use case and lowers the cost of analysis.



**Tip**

If you want to store data *without performing analysis on the data*, set the **Hierarchical Namespace** option to **Disabled** to set up the storage account as an Azure Blob storage account. You can also use blob storage to archive rarely used data or to store website assets such as images and media.

If you are performing analytics on the data, set up the storage account as an Azure Data Lake Storage Gen2 account by setting the **Hierarchical Namespace** option to **Enabled**. Because Azure Data Lake Storage Gen2 is integrated into the Azure Storage platform, applications can use either the Blob APIs or the Azure Data Lake Storage Gen2 file system APIs to access data.

**So sánh Azure Blob Storage và Azure Data Lake Storage Gen2**

**Azure Blob Storage** là dịch vụ lưu trữ đối tượng của Microsoft Azure, cho phép bạn lưu trữ một lượng lớn dữ liệu phi cấu trúc. Dữ liệu này được lưu trữ trong các thùng chứa blob (blob container) với không gian tên phẳng (flat namespace). Tên blob có thể chứa ký tự "/" để phân cấp thành các "thư mục" ảo, nhưng về mặt quản lý blob, chúng được lưu trữ như một hệ thống phân cấp một cấp.

**Azure Data Lake Storage Gen2** được xây dựng dựa trên Blob Storage, tối ưu hóa hoạt động đầu vào/đầu ra (I/O) của dữ liệu khối lượng lớn bằng cách sử dụng không gian tên phân cấp. Không gian tên này tổ chức dữ liệu blob thành các thư mục và lưu trữ siêu dữ liệu về từng thư mục và các file bên trong. Cấu trúc này cho phép thực hiện các hoạt động như đổi tên và xóa thư mục chỉ trong một thao tác nguyên tử (atomic operation). Ngược lại, không gian tên phẳng yêu cầu nhiều thao tác tương ứng với số đối tượng trong cấu trúc. Không gian tên phân cấp giúp dữ liệu được tổ chức khoa học, mang lại hiệu suất lưu trữ và truy xuất tốt hơn cho các phân tích, đồng thời giảm chi phí phân tích.

**Mẹo:**

* Nếu bạn muốn lưu trữ dữ liệu mà không phân tích, hãy đặt tùy chọn **Không gian tên phân cấp** thành **Tắt** để thiết lập tài khoản lưu trữ như tài khoản Azure Blob Storage. Blob Storage cũng hữu ích để lưu trữ dữ liệu ít sử dụng hoặc các nội dung trang web như hình ảnh và media.
* Nếu bạn cần phân tích dữ liệu, hãy thiết lập tài khoản lưu trữ thành Azure Data Lake Storage Gen2 bằng cách đặt tùy chọn **Không gian tên phân cấp** thành **Bật**. Vì Data Lake Storage Gen2 được tích hợp với nền tảng Azure Storage, các ứng dụng có thể sử dụng API Blob hoặc API hệ thống file của Data Lake Storage Gen2 để truy cập dữ liệu.

**Tóm tắt:**

|  |  |  |
| --- | --- | --- |
| Tính năng | Azure Blob Storage | Azure Data Lake Storage Gen2 |
| Không gian tên | Phẳng (flat namespace) | Phân cấp (hierarchical namespace) |
| Tổ chức dữ liệu | Thùng chứa blob | Thư mục |
| Hiệu suất phân tích | Thấp | Cao |
| Phù hợp cho | Lưu trữ dữ liệu không phân tích | Phân tích dữ liệu khối lượng lớn |
| Truy cập dữ liệu | HTTP(S) API | Blob API hoặc API hệ thống file |