**Polybase** là một tính năng quan trọng trong SQL Server và Azure Synapse Analytics, đóng vai trò thiết yếu trong việc **lấy dữ liệu đầu vào** (ingestion data) cho kho dữ liệu. Nó cho phép bạn truy cập và xử lý dữ liệu từ nhiều nguồn khác nhau, bao gồm:

* **Hệ thống lưu trữ dữ liệu phi quan hệ:** Hadoop, Spark, Azure Blob Storage.
* **Hệ quản trị cơ sở dữ liệu khác:** Cosmos DB, Oracle, Teradata, MongoDB.

**Ví dụ cụ thể:**

Giả sử bạn có một kho dữ liệu Azure Synapse Analytics lưu trữ thông tin bán hàng. Bạn muốn bổ sung dữ liệu về hành vi khách hàng từ hệ thống Hadoop. Polybase giúp bạn thực hiện điều này bằng cách:

1. **Tạo bảng bên ngoài:** Sử dụng Transact-SQL để định nghĩa bảng ảo đại diện cho dữ liệu Hadoop. Bảng ảo này có cấu trúc tương tự như bảng SQL thông thường, giúp bạn dễ dàng truy cập dữ liệu.
2. **Truy vấn dữ liệu:** Sử dụng Transact-SQL để truy vấn dữ liệu Hadoop thông qua bảng ảo. Polybase sẽ tự động chuyển đổi truy vấn sang định dạng phù hợp với Hadoop và thực hiện truy vấn trên hệ thống Hadoop.
3. **Kết hợp dữ liệu:** Kết hợp dữ liệu từ bảng ảo Hadoop với dữ liệu từ các bảng SQL khác trong cùng một truy vấn. Việc này giúp bạn có được thông tin tổng quan toàn diện về hoạt động kinh doanh.

**Lợi ích của Polybase:**

* **Truy cập dữ liệu đa dạng:** Polybase giúp bạn truy cập dữ liệu từ nhiều nguồn khác nhau mà không cần di chuyển dữ liệu.
* **Dễ sử dụng:** Sử dụng Transact-SQL quen thuộc để truy vấn dữ liệu từ các nguồn khác nhau.
* **Hiệu suất cao:** Polybase tối ưu hóa truy vấn để đảm bảo hiệu suất cao.
* **Tiết kiệm chi phí:** Polybase giúp bạn tiết kiệm chi phí lưu trữ và quản lý dữ liệu.

**Polybase là một công cụ mạnh mẽ giúp bạn:**

* **Tích hợp dữ liệu từ nhiều nguồn khác nhau**
* **Thực hiện phân tích dữ liệu toàn diện**
* **Cải thiện hiệu quả hoạt động kinh doanh**

**Ngoài Polybase, Azure Data Factory cũng hỗ trợ nạp dữ liệu bằng cách:**

* **Gọi trực tiếp Polybase** nếu dữ liệu của bạn nằm trong kho dữ liệu tương thích với Polybase.
* **Sử dụng các hoạt động tích hợp sẵn** cho các nguồn dữ liệu khác nhau.

**Kết luận:**

Polybase là một công cụ thiết yếu cho việc lấy dữ liệu đầu vào trong kho dữ liệu. Nó cho phép bạn truy cập và xử lý dữ liệu từ nhiều nguồn khác nhau một cách dễ dàng và hiệu quả