**1. Ingest (thu thập):** Quá trình thu thập dữ liệu từ nhiều nguồn khác nhau vào một hệ thống hoặc kho lưu trữ.

**2. Raw data (dữ liệu thô):** Dữ liệu chưa được xử lý hoặc phân tích.

**3. ETL (Extract, Transform, Load):** Quá trình trích xuất dữ liệu từ một nguồn, chuyển đổi dữ liệu sang định dạng phù hợp và tải dữ liệu vào một hệ thống khác.

**4. ELT (Extract, Load, Transform):** Quá trình trích xuất dữ liệu từ một nguồn, tải dữ liệu vào một hệ thống khác và sau đó chuyển đổi dữ liệu sang định dạng phù hợp.

**5. Pipeline (luồng công việc):** Một tập hợp các hoạt động được kết nối với nhau để thực hiện một nhiệm vụ.

**6. Dataflow (luồng dữ liệu):** Một cách trực quan để xây dựng các quy trình ETL trong Azure Data Factory.

**7. Compute services (dịch vụ điện toán):** Các dịch vụ cung cấp khả năng xử lý dữ liệu, chẳng hạn như Azure HDInsight, Azure Databricks và Azure SQL Database.

**8. Business Intelligence (BI):** Trí tuệ kinh doanh là tập hợp các phương pháp, quy trình, kiến trúc và công nghệ được sử dụng để thu thập, lưu trữ và phân tích dữ liệu để đưa ra quyết định kinh doanh tốt hơn.

**9. Blob (đối tượng blob):** Một đơn vị dữ liệu được lưu trữ trong Azure Blob Storage.

**10. Partition (phân vùng):** Quá trình chia dữ liệu thành các phần nhỏ hơn để xử lý hiệu quả hơn.

**11. U-SQL:** Ngôn ngữ lập trình được sử dụng để viết các tác vụ trong Azure Data Lake Analytics.

**12. CSV (Comma-separated values):** Định dạng tập tin văn bản sử dụng dấu phẩy để phân cách các trường dữ liệu.

**13. Catalog (danh mục):** Một bộ sưu tập metadata (dữ liệu về dữ liệu) mô tả các dữ liệu được lưu trữ trong Azure Data Lake Store.

**14. Outputter (bộ xuất ra):** Một thành phần trong Azure Data Lake Analytics được sử dụng để ghi dữ liệu vào một đích cụ thể.

**15. INSERT, UPDATE, DELETE:** Các câu lệnh SQL được sử dụng để thêm, sửa đổi và xóa dữ liệu trong bảng.

**Ngoài ra:**

* **HDInsight:** Dịch vụ phân tích dữ liệu dựa trên Hadoop được quản lý trên Microsoft Azure.
* **Cosmos DB:** Cơ sở dữ liệu NoSQL được quản lý, phân tán và có khả năng mở rộng cao trên Microsoft Azure.
* **Hive:** Kho dữ liệu SQL được xây dựng trên Hadoop.
* **Spark:** Nền tảng xử lý dữ liệu mã nguồn mở được sử dụng để xử lý dữ liệu lớn.

**Hy vọng những giải thích trên giúp bạn hiểu rõ hơn về các từ mới trong đoạn văn**