**1. Công cụ phân tích (Analytics engine):** Chương trình phần mềm được thiết kế để xử lý và phân tích lượng dữ liệu lớn. (Ví dụ: Azure Synapse Analytics)

**2. Thu thập dữ liệu (Ingest data):** Mang dữ liệu vào hệ thống từ một nguồn bên ngoài.

**3. Các tệp phẳng (Flat files):** Tệp văn bản đơn giản không có cấu trúc phân cấp, thường được sử dụng để lưu trữ dữ liệu trong bảng.

**4. Tổng hợp (Aggregate):** Kết hợp nhiều phần dữ liệu thành một giá trị tóm tắt duy nhất.

**5. Kiến trúc xử lý song song hàng loạt (Massively parallel processing - MPP architecture):** Kiến trúc máy tính được thiết kế để phân phối các tác vụ xử lý trên nhiều bộ xử lý đồng thời để thực thi nhanh hơn.

**6. Nút điều khiển (Control node):** Nút trung tâm trong kiến trúc MPP quản lý và phối hợp công việc của các nút tính toán.

**7. Nút tính toán (Compute node):** Nút trong kiến trúc MPP thực hiện các tác vụ xử lý dữ liệu thực tế.

**8. SQL pool (SQL pool):** Loại nhóm tính toán trong Azure Synapse Analytics sử dụng Azure SQL Database để xử lý dữ liệu.

**9. Spark pool (Spark pool):** Loại nhóm tính toán trong Azure Synapse Analytics sử dụng Apache Spark để xử lý dữ liệu.

**10. PolyBase:** Công nghệ trong Azure Synapse Analytics cho phép truy xuất dữ liệu từ nhiều nguồn khác nhau như cơ sở dữ liệu quan hệ, nguồn phi quan hệ và tệp văn bản.

**11. Spark job (Spark job):** Đơn vị công việc được thực thi bởi Apache Spark, thường được viết bằng ngôn ngữ lập trình như Python hoặc Scala.

**12. Sổ tay (Notebook):** Môi trường tương tác trong Azure Databricks nơi người dùng có thể viết và thực thi mã, thường được sử dụng cho phân tích dữ liệu.

**13. Xử lý trong bộ nhớ (In-memory processing):** Xử lý dữ liệu được lưu trữ trong bộ nhớ chính (RAM) của máy tính để truy cập và thao tác nhanh hơn so với xử lý dựa trên đĩa.

**14. Tự động điều chỉnh kích thước (Autoscaling):** Khả năng của dịch vụ đám mây tự động thêm hoặc xóa tài nguyên (ví dụ: nút tính toán) dựa trên nhu cầu.

**15. Xử lý phân tích trực tuyến (Online analytical processing - OLAP):** Loại xử lý dữ liệu tập trung vào phân tích dữ liệu đa chiều cho mục đích trí tuệ kinh doanh.

**16. Mô hình bảng (Tabular model):** Mô hình dữ liệu được sử dụng trong Analysis Services biểu diễn dữ liệu dưới dạng bảng với mối quan hệ giữa các bảng.

**17. Nguồn dữ liệu (Data source):** Vị trí lưu trữ dữ liệu, chẳng hạn như cơ sở dữ liệu, hệ thống tệp hoặc dịch vụ web.

**18. Tính đồng thời (Concurrency):** Số lượng người dùng có thể truy cập và tương tác với hệ thống cùng một lúc.

**19. ETL (Extract, Transform, Load - Trích xuất, Chuyển đổi, Tải):** Quá trình trích xuất dữ liệu từ nhiều nguồn, chuyển đổi nó thành định dạng nhất quán và tải vào kho dữ liệu.

**20. Tiền xử lý (Preprocessing):** Quá trình chuẩn bị dữ liệu cho phân tích sâu hơn, chẳng hạn như dọn dẹp, lọc và chuyển đổi dữ liệu.

**21. Dịch vụ xử lý dữ liệu lớn (Big data processing service):** Dịch vụ đám mây được thiết kế để xử lý lưu trữ, xử lý và phân tích các tập dữ liệu lớn và phức tạp.

**22. Mô hình cụm (Clustered model):** Mô hình điện toán phân phối các tác vụ trên nhiều máy tính trong một cụm để xử lý song song.

**23. Apache Kafka:** Nền tảng mã nguồn mở cho các đường ống dữ liệu luồng.

**24. Mô hình xử lý Apache Hadoop:** Khung mã nguồn mở để xử lý phân tán các tập dữ liệu lớn trên các cụm máy tính.

**25. Hive:** Dự án phần mềm kho dữ liệu được xây dựng trên Apache Hadoop cho phép truy vấn dữ liệu sử dụng cú pháp giống SQL.

**26. Bảng ngoài (External tables):** Bảng trong Hive trỏ đến dữ liệu được lưu trữ ở các vị trí bên ngoài như HDFS (Hadoop Distributed File System) hoặc Azure Data Lake Storage.