* **Ingested data (n):** Dữ liệu thô được đưa vào hệ thống, chưa qua xử lý. (Ex: The data warehouse stores ingested data from various sources.)
* **Data store (n):** Kho lưu trữ dữ liệu. (Ex: A data store is a system designed to hold and manage data.)
* **Data warehouse (n):** Kho dữ liệu (thường chứa dữ liệu được xử lý và tổ chức theo cấu trúc phục vụ cho việc phân tích). (Ex: Businesses use data warehouses to store and analyze large amounts of data.)
* **Apache Spark (n):** Apache Spark là một nền tảng mã nguồn mở dùng để xử lý dữ liệu song song và phân tích Big Data. (Ex: Spark is a popular tool for large-scale data processing.)
* **Transact-SQL (T-SQL) (n):** Ngôn ngữ truy vấn mở rộng của SQL được dùng trong Microsoft SQL Server. (Ex: T-SQL allows you to perform complex operations on data in a relational database.)
* **Notebook (n):** Sổ tay điện tử, thường được dùng để ghi chép và chạy code. (Ex: Data scientists use notebooks to write and execute code for data analysis.)
* **Clustered architecture (n):** Kiến trúc cụm, là một nhóm máy tính được kết nối với nhau để hoạt động như một hệ thống duy nhất. (Ex: A clustered architecture allows for parallel processing of data.)
* **Control node (n):** Nút điều khiển, là nút trung tâm chịu trách nhiệm điều phối các hoạt động trong cụm. (Ex: The control node receives requests and distributes tasks to other nodes in the cluster.)
* **Compute node (n):** Nút tính toán, là nút thực sự thực hiện các tác vụ xử lý dữ liệu. (Ex: Compute nodes perform the actual calculations on the data.)
* **Driver (n):** Trình điều khiển, là một thành phần phần mềm cho phép kết nối với một nguồn dữ liệu cụ thể. (Ex: A driver enables you to read and write data from a specific data source.)
* **Extensible architecture (n):** Kiến trúc mở rộng, là kiến trúc có thể dễ dàng tích hợp thêm các thành phần mới. (Ex: An extensible architecture allows you to add new features and functionality as needed.)
* **Massive quantities of data (n):** Lượng dữ liệu khổng lồ. (Ex: Spark is designed to handle massive quantities of data efficiently.)
* **Machine learning (n):** Học máy, là một lĩnh vực của trí tuệ nhân tạo cho phép máy tính học hỏi từ dữ liệu. (Ex: Machine learning algorithms can be used to make predictions based on data.)
* **Statistical analysis (n):** Phân tích thống kê, là quá trình thu thập, tổ chức, phân tích, diễn giải và trình bày dữ liệu. (Ex: Statistical analysis can be used to identify trends and patterns in data.)
* **Linear and non-linear modeling (n):** Mô hình tuyến tính và phi tuyến tính, là các phương pháp để mô tả mối quan hệ giữa các biến. (Ex: Linear models assume a straight-line relationship between variables, while non-linear models can capture more complex relationships.)
* **Predictive analytics (n):** Phân tích dự báo, là việc sử dụng dữ liệu để dự đoán các sự kiện tương lai. (Ex: Predictive analytics can be used to forecast sales trends or identify potential customer churn.)
* **Repository (n):** Kho lưu trữ, là một nơi tập trung để lưu trữ dữ liệu. (Ex: The results can be saved to a data repository for future reference.)
* **Managed analytic service (n):** Dịch vụ phân tích được quản lý, là một dịch vụ đám mây cung cấp các công cụ để xử lý dữ liệu. (Ex: Azure HDInsight is a managed analytic service that simplifies the process of running Hadoop clusters.)
* **Open source tools and utilities (n):** Công cụ và tiện ích mã nguồn mở, là các phần mềm được phát triển cộng đồng và có sẵn miễn phí. (Ex: Apache Hadoop is a collection of open-source tools for distributed data processing.)
* **Framework (n):** Khung phần mềm, là một tập hợp các thư viện và công cụ được thiết kế để hỗ trợ việc xây dựng các ứng dụng. (Ex: Spark is a framework that can be used to build large-scale data processing applications.)
* **MapReduce (n):** MapReduce là một mô hình lập trình song song được sử dụng để xử lý các tập dữ liệu lớn. (Ex: MapReduce jobs are split into smaller tasks that can be run in parallel.)
* **Apache Hive (n):** Apache Hive là một công cụ mã nguồn mở