**1. SQL Server Integration Services (SSIS):**

* **Nền tảng:** Là một nền tảng để xây dựng các giải pháp tích hợp dữ liệu và chuyển đổi dữ liệu ở cấp độ doanh nghiệp.
* **Chức năng:**
  + Giải quyết các vấn đề kinh doanh phức tạp bằng cách:
    - Sao chép hoặc tải xuống các tập tin.
    - Nạp dữ liệu vào kho dữ liệu.
    - Làm sạch và khai thác dữ liệu.
    - Quản lý đối tượng và dữ liệu trong database SQL.
* **Nguồn dữ liệu:** Có thể trích xuất và chuyển đổi dữ liệu từ nhiều nguồn khác nhau, chẳng hạn như:
  + Tệp dữ liệu XML.
  + Tệp phẳng (flat files).
  + Nguồn dữ liệu quan hệ.
* **Điểm đến dữ liệu:** Nạp dữ liệu vào một hoặc nhiều đích đến.
* **Công cụ:**
  + Bao gồm một bộ phong phú các tác vụ và chuyển đổi được tích hợp sẵn.
  + Cung cấp các công cụ đồ họa để xây dựng các gói (package).
  + Bao gồm Integration Services Catalog, một database lưu trữ nơi bạn chạy và quản lý các gói.
* **Gói (Package):**
  + Là một tập hợp được tổ chức bao gồm:
    - Kết nối.
    - Các thành phần điều khiển luồng (control flow).
    - Các thành phần luồng dữ liệu (data flow).
    - Trình xử lý sự kiện (event handler).
    - Biến.
    - Tham số.
    - Cấu hình.
  + Có thể được xây dựng bằng:
    - Công cụ thiết kế đồ họa của SSIS.
    - Lập trình.
* **Lưu trữ:**
  + Gói hoàn thành có thể được lưu vào:
    - SQL Server.
    - Integration Services Package Store.
    - Hệ thống tập tin.
* **Lập trình:**
  + Có thể sử dụng các công cụ đồ họa SSIS để tạo giải pháp mà không cần viết một dòng code nào.
  + Ngoài ra, bạn có thể lập trình mô hình đối tượng Integration Services mở rộng để tạo các gói theo chương trình và mã hóa các tác vụ tùy chỉnh và các đối tượng gói khác.
* **Giới hạn:** SSIS là một tiện ích cục bộ (on-premises).
* **Azure Data Factory:**
  + Cho phép bạn chạy các gói SSIS hiện có của bạn như một phần của pipeline trên cloud.
  + Điều này giúp bạn bắt đầu nhanh chóng mà không cần viết lại logic chuyển đổi hiện có.
* **SSIS Feature Pack for Azure:**
  + Là một tiện ích mở rộng cung cấp các thành phần kết nối với các dịch vụ Azure.
  + Chuyển dữ liệu giữa Azure và các nguồn dữ liệu cục bộ.
  + Xử lý dữ liệu được lưu trữ trong Azure.
  + Các thành phần trong Feature Pack hỗ trợ truyền dữ liệu đến hoặc đi từ:
    - Azure Storage.
    - Azure Data Lake.
    - Azure HDInsight.
  + Bằng cách sử dụng các thành phần này, bạn có thể thực hiện xử lý dữ liệu quy mô lớn trên dữ liệu đã được nạp (ingested data).

**2. Azure Data Factory:**

* **Giới thiệu:** Dịch vụ nạp và chuyển đổi dữ liệu cho phép bạn tải dữ liệu thô từ nhiều nguồn khác nhau (cả trong và ngoài môi trường cloud).
* **Đã được đề cập trước đó trong bài đọc**

**3. Azure Databricks:**

* **Nền tảng phân tích:** Được tối ưu hóa cho các dịch vụ đám mây Microsoft Azure.
* **Dựa trên Apache Spark:** Spark là một engine xử lý song song hỗ trợ phân tích dữ liệu quy mô lớn và được tích hợp với Microsoft Azure để đơn giản hóa quy trình làm việc.
* **Không gian làm việc tương tác:** Cung cấp một không gian làm việc tương tác cho phép cộng tác giữa các nhà khoa học dữ liệu, kỹ sư dữ liệu và nhà phân tích kinh doanh.
* **Nguồn dữ liệu:** Có thể xử lý dữ liệu được lưu trữ trong nhiều loại lưu trữ khác nhau, bao gồm:
  + Azure Blob Storage.
  + Azure Data Lake Store.
  + Lưu trữ Hadoop.
  + Tệp phẳng.
  + Database SQL.
  + Kho dữ liệu.
  + Các dịch vụ Azure khác như Cosmos