**Azure Synapse Analytics - Các Thành Phần**

**Synapse Pipelines**

* Pipeline là một nhóm hoạt động logic cùng nhau thực hiện một tác vụ.
* Ví dụ: Một pipeline có thể chứa một tập hợp các hoạt động trích xuất và làm sạch dữ liệu nhật ký, sau đó kích hoạt luồng dữ liệu ánh xạ để phân tích dữ liệu nhật ký.
* Pipeline cho phép bạn quản lý các hoạt động như một tập hợp thay vì từng hoạt động riêng lẻ.
* Bạn triển khai và lên lịch cho pipeline thay vì triển khai và lên lịch cho từng hoạt động độc lập.
* Các hoạt động trong pipeline xác định các hành động thực hiện trên dữ liệu của bạn.
* Ví dụ: Bạn có thể sử dụng hoạt động sao chép để sao chép dữ liệu từ Azure Blob Storage vào Azure Synapse bằng cách sử dụng pool Sequel.
* Sau đó, sử dụng hoạt động luồng dữ liệu hoặc hoạt động sổ ghi chép sử dụng pool Spark để xử lý và tạo mô hình học máy.
* Synapse Pipelines sử dụng cùng một công cụ tích hợp dữ liệu được sử dụng bởi Azure Data Factory.
* Điều này cung cấp cho bạn sức mạnh trong Synapse Studio để tạo các pipeline có thể kết nối với hơn 90 nguồn từ các tệp phẳng, cơ sở dữ liệu hoặc dịch vụ trực tuyến.
* Bạn có thể tạo các luồng dữ liệu không cần mã cho phép bạn thực hiện ánh xạ và chuyển đổi phức tạp trên dữ liệu khi nó chảy vào các giải pháp phân tích của bạn.
* Ví dụ: Bạn có thể có một pipeline với ba hoạt động. Đầu tiên, pipeline trích xuất dữ liệu, sau đó nó sử dụng sổ ghi chép Spark để tạo mô hình học máy. Cuối cùng, pipeline chứa một hàm Azure ở cuối để kiểm tra và xác nhận mô hình học máy.

**Azure Synapse Link**

* Azure Synapse Link for Azure Cosmos DB là xử lý giao dịch lai và xử lý phân tích đám mây (HTAP) hoặc khả năng HTAP cho phép bạn chạy phân tích gần thời gian thực trên dữ liệu hoạt động được lưu trữ trong Azure Cosmos DB.
* Synapse Link sử dụng một tính năng của Cosmos DB có tên là Kho lưu trữ phân tích Cosmos DB.
* Kho lưu trữ phân tích Cosmos DB chứa một bản sao của dữ liệu trong vùng chứa Cosmos DB nhưng được tổ chức như một kho lưu trữ theo cột.
* Kho lưu trữ theo cột nhóm dữ liệu theo cột thay vì theo hàng, kho lưu trữ theo cột là một định dạng tối ưu hơn đối với việc chạy các khối lượng công việc phân tích cần tổng hợp dữ liệu xuống một cột thay vì trên một hàng.
* Ví dụ như tạo tổng, trung bình, giá trị tối đa hoặc tối thiểu cho một cột.
* Cosmos DB tự động đồng bộ hóa dữ liệu trong các vùng chứa của nó với các bản sao trong kho lưu trữ theo cột.
* Azure Synapse Link cho phép bạn chạy các khối lượng công việc lấy dữ liệu trực tiếp từ Cosmos DB và chạy các khối lượng công việc phân tích bằng Azure Synapse Analytics.
* Dữ liệu không cần phải trải qua quy trình trích xuất, chuyển đổi và tải (ETL) vì dữ liệu không được sao chép vào Synapse Analytics. Nó vẫn nằm trong Kho lưu trữ phân tích Cosmos DB.
* Giờ đây, các nhà phân tích kinh doanh, kỹ sư dữ liệu và nhà khoa học dữ liệu có thể sử dụng pool Spark của Synapse hoặc pool SQL của Synapse để chạy phân tích kinh doanh thông minh gần thời gian thực và các pipeline học máy.
* Bạn có thể đạt được điều này mà không ảnh hưởng đến hiệu suất của khối lượng công việc giao dịch của bạn trên Azure Cosmos DB.
* Synapse Link có nhiều ứng dụng rộng rãi bao gồm phân tích và dự báo chuỗi cung ứng, báo cáo hoạt động, tích hợp và dàn nhạc dữ liệu hàng loạt, cá nhân hóa theo thời gian thực và bảo trì IOT.

**Synapse Studio**

* Synapse Studio là giao diện web cho phép bạn tạo pool và pipeline tương tác.
* Với Synapse Studio, bạn có thể phát triển, kiểm thử và gỡ lỗi các sổ ghi chép Spark và tác vụ Transact SQL.
* Bạn có thể theo dõi hiệu suất của các hoạt động đang chạy và bạn có thể quản lý các tài nguyên không máy chủ hoặc được cung cấp sẵn.
* Tất cả các khả năng này đều được truy cập thông qua Synapse Studio gốc web cho phép quản lý mô hình, giám sát, mã hóa và bảo mật.
* Bạn có thể truy cập Synapse Studio trực tiếp từ Azure Portal.