

Data Transform Load (DTL)

Nền tảng thu thập, chuyển hóa dữ liệu

Nền tảng thu thập, chuyển hóa dữ liệu giải pháp được Bkav phát triển giúp thu thập dữ liệu nguồn, hút dữ liệu từ các CSDL, các Sensor hoặc API của các phần mềm / dịch vụ nhằm mục đích thu thập dữ liệu, từ đó đưa chuyển về Trung tâm dữ liệu. Khi nhận được dữ liệu, nền tảng sẽ phân tích, chuẩn hóa và làm sạch dữ liệu nhận được, sau đó đóng gói thành dữ liệu có cấu trúc và lưu trữ vào kho dữ liệu lớn (BigData) để phục vụ cho phân tích, báo cáo, dự báo.

Giải pháp hỗ trợ nhận dữ liệu từ trực liên thông dữ liệu. Dữ liệu từ Nền tảng tích hợp chia sẻ dữ liệu đẩy về sẽ được nhận ở các định dạng như JSON, XML... Sau khi nền tảng nhận được dữ liệu này, hệ thống sẽ bắt đầu thực hiện quy trình chuyển đổi dữ liệu. Trong quy trình này người dùng có thể cấu hình các tập luật để lọc dữ liệu với mục đích làm sạch bộ dữ liệu này.

Dữ liệu sau khi được chuyển đổi và làm sạch, hoặc những dữ liệu thuộc định dạng bán cấu trúc/phi cấu trúc/có cấu trúc nhưng chưa có nhu cầu sử dụng, Nền tảng thu thập, chuyển hóa dữ liệu sẽ kết nối đến dữ liệu lớn BigData (Data Warehouse hoặc Data Lake). Bộ dữ liệu đã được chuyển đổi sẽ được đưa vào trong bảng của kho dữ liệu. Bảng dữ liệu này là bảng đã được tạo trước hoặc là sẽ do hệ thống chuyển hóa tự động tạo bảng tương ứng với bộ dữ liệu chuyển đổi.



The background of the central image features a hand holding a globe, with a large gear shape overlaid. The scene is filled with binary code (0s and 1s) in various colors (blue, green, orange) against a dark background. The text is centered over this image.

CHANGE DATA CAPTURE (CDC)

Thu thập dữ liệu nguồn

Data Transform Load

Nền tảng thu thập, chuyển hóa dữ liệu



Chức năng

Kết nối Datasource

- Cấu hình module listener lựa chọn các database muốn phần mềm lấy dữ liệu bằng cách điền địa chỉ kết nối đến CSDL cần lấy, điền thông tin xác thực cho phần mềm kết nối với CSDL.
- Hỗ trợ điền các thông tin kết nối, hệ thống sẽ khởi tạo lấy toàn bộ dữ liệu trong CSDL nguồn, để làm dữ liệu lịch sử. Sau khi lấy xong dữ liệu, hệ thống sẽ định kỳ lấy dữ liệu thay đổi của CSDL.
- Hỗ trợ khởi tạo và lấy dữ liệu về tùy thuộc vào nguồn dữ liệu mà DTL sẽ có các cách đưa dữ liệu về để phân tích khác nhau.

Thu thập dữ liệu

- Cấu hình định kỳ lấy dữ liệu (theo Realtime/Near-Realtime), kiểm tra xem có dữ liệu mới hơn không để lấy dữ liệu.



- Hệ thống định kỳ theo thời gian quy định thực hiện kiểm tra nguồn dữ liệu xem có dữ liệu mới không, nếu có dữ liệu mới sẽ thực hiện thu thập, nếu không sẽ chuyển sang trạng thái đợi pha thực hiện tiếp theo.

Lựa chọn lấy dữ liệu

- Quản trị có thể lấy toàn bộ dữ liệu thay đổi của CSDL hoặc chỉ lấy một bảng, một trường trong bảng của CSDL.
- Hỗ trợ sử dụng các luật được cấu hình để lấy dữ liệu cần thiết theo cấu hình. Dữ liệu này sẽ được đóng gói và chuyển về lưu trữ trong CSDL lớn BigData.

Mapping & packaging dữ liệu nguồn với đích

- Hỗ trợ cấu hình Mapping những trường dữ liệu lấy được từ source với các trường dữ liệu đích để có thể truyền được dữ liệu về đích.
- Khi dữ liệu được map với các trường dữ liệu đích, hệ thống sẽ đóng gói theo quy chuẩn của đích để truyền dữ liệu về đích có thể hiểu được.

Cấu hình Dest và giám sát quy trình

- Hỗ trợ lựa chọn các đích nhận dữ liệu khác nhau để sau khi mapping, packaging thì truyền dữ liệu về nguồn tương ứng.
- Hệ thống giám sát sẽ tự động đưa ra cảnh báo để có thể kịp thời xử lý các vấn đề.

Ghi log hệ thống và quản trị farm

- Hỗ trợ đặt các loại log khác nhau để theo dõi và debug hệ thống bao gồm đặt log warning, log alert...
- Hiện thị toàn bộ các agent CDS trên trang quản trị tập trung, quản trị viên có thể xem được trạng thái của các agent, thông số kết nối.



EXTRACT - TRANSFORM - LOAD (ETL)
Chuyển hóa dữ liệu

Data Transform Load

Nền tảng thu thập, chuyển hóa dữ liệu



Chức năng

Lọc, làm sạch dữ liệu

- Cấu hình kiểu của trường dữ liệu, cấu hình, loại bỏ các dữ liệu không đúng định dạng cấu trúc.
- Hỗ trợ sử dụng các luật được cấu hình để lấy đúng dữ liệu cần thiết. Dữ liệu này sẽ được đóng gói và chuyển về lưu trữ trong BigData.

Giám sát các quy trình

- Nếu bất kỳ hệ thống nào bị lỗi hoặc kiểm tra cho thấy dữ liệu không chính xác, hệ thống giám sát sẽ tự động đưa ra cảnh báo để có thể kịp thời xử lý các vấn đề.

Mapping & packaging dữ liệu nguồn với đích

- Quản trị có thể cấu hình mapping những trường dữ liệu lấy được từ source với các trường dữ liệu đích để có thể truyền được dữ liệu về đích
- Khi dữ liệu được map với các trường dữ liệu đích, hệ thống sẽ đóng gói theo quy chuẩn của đích để truyền dữ liệu về đích có thể hiểu được.

Kết nối đến đích

- Quản trị có thể lựa chọn các đích nhận dữ liệu khác nhau để sau khi Mapping, packaging thì truyền dữ liệu về nguồn tương ứng.

Ghi log hệ thống, chạy thử nghiệm

- Hỗ trợ đặt các loại log khác nhau để theo dõi và debug hệ thống bao gồm đặt log warning, log alert...



- Kiểm tra độ chính xác của dữ liệu cũng như việc các quy trình hoạt động chính xác, DTL cung cấp chức năng chạy thử nghiệm cho phép người dùng đánh giá một cách chính xác nhất.

Data Transform Load

Nền tảng thu thập, chuyển hóa dữ liệu



Giải pháp & Thông số kỹ thuật



Ngôn ngữ lập trình

Java, Python,
Scala



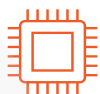
Cơ sở dữ liệu

MySQL, PostgreSQL,
Mongodb, Cassandra



Hệ điều hành

Linux server



CPU

Central Processing Unit
16 Cores



RAM

Random Access Memory
32 GB



HDD

Hard Disk Drive
500 GB

Thông tin liên hệ

Trụ sở chính: Tòa nhà Bkav, Khu đô thị Yên Hòa, Cầu Giấy, Hà Nội

Điện thoại: (024) 3763 2552 Số fax: (024) 3868 4755

Website: www.bkav.com.vn Email: DuAn@bkav.com

Bkav TP. HCM: Số 67, Đường số 3, Khu dân cư City Land, P. 7, Q. Gò Vấp, TP HCM

Điện thoại: (028) 6296 6626 Số fax: (028) 2253 6103