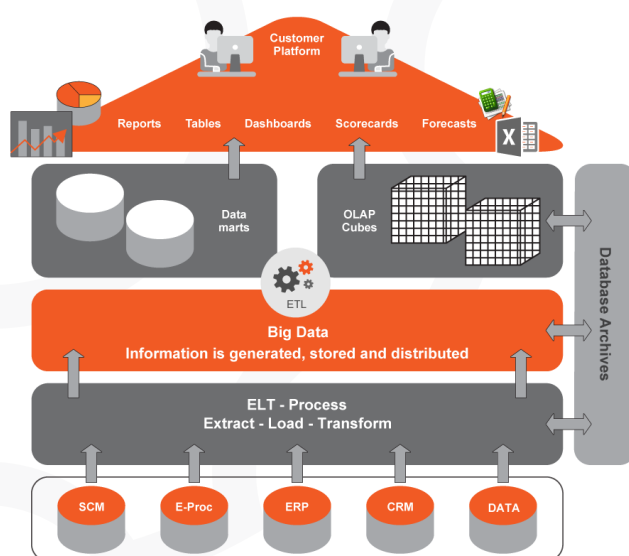


Big Data

Nền tảng dữ liệu lớn

Nền tảng dữ liệu lớn (Big Data) là kho dữ liệu tập trung (Data Warehouse, Data Lake) chuyên dùng cho việc phân tích, thống kê, báo cáo, dự báo.

Dữ liệu lớn Big Data hỗ trợ các truy vấn phức tạp, vừa là điểm tập trung dữ liệu từ nhiều nguồn khác nhau để phân tích đầy đủ nhất, xoay nhiều chiều dữ liệu. Dữ liệu trong Big Data hướng chủ đề, toàn vẹn, không bị rò rỉ và có giá trị lịch sử.



Kiến trúc nền tảng

Ưu điểm của giải pháp:

- ✓ Lấy dữ liệu làm trung tâm (Data-Centric), tập trung mọi dữ liệu về Big Data.
- ✓ Có dữ liệu phát sinh sẽ đẩy về Big Data. Tất cả dữ liệu nằm trong 01 Big Data chung với tầm nhìn dữ liệu lớn.
- ✓ Khi cần sử dụng tới dữ liệu nào chỉ cần cấu hình các CSDL riêng là đáp ứng được ngay.
- ✓ Sẵn sàng cho việc bùng nổ hiệu năng để đáp ứng các nhu cầu vô hạn, chưa biết trước.

Big Data Platform

Nền tảng dữ liệu lớn



Chức năng

CSDL lưu trữ dữ liệu lớn

- Lưu trữ nhiều dạng CSDL dữ liệu khác nhau: có quan hệ, không có quan hệ.
- Hỗ trợ tính toán trên từng CSDL hoặc trên nhiều CSDL cùng lúc.
- Phân quyền truy cập từng CSDL cho từng account và theo địa chỉ mạng.
- Hỗ trợ nhiều công cụ xử lý dữ liệu dưới dạng SQL, NoSQL, Stream, Data science phân tích tính toán trên tập dữ liệu lớn được lưu trữ trên HDFS.
- Chuyển đổi hiệu quả dữ liệu từ RDBMS sang HDFS và ngược lại.
- Bảo mật toàn diện cho các thành phần trong cụm, kiểm soát truy cập vào các thành phần trong cụm.



Khai thác và xử lý dữ liệu

- Hỗ trợ query trên mô hình phân tán, tính toán trên nhiều node và tổng hợp kết quả.
- Hỗ trợ câu lệnh SQL để truy vấn vào DWH thông qua JDBC.

- Hỗ trợ cài đặt tự động index dữ liệu bản ghi mới, phân quyền dữ liệu theo user, theo domain, theo địa chỉ ip được truy cập và khai thác dữ liệu trên DWH.
- Tăng hiệu suất tính toán sử dụng map-reduce.
- Hỗ trợ Machine Learning và Deep Learning để xử lý dữ liệu.
- Cấu hình tài nguyên cho các thành phần trong cụm tối ưu đối với hiện trạng phần cứng hiện có.

Quản trị DataWarehouse (DWH)

- Cấu hình tùy chỉnh (cài đặt từng thành phần/gỡ bỏ từng module) của DWH trên giao diện đồ họa.
- Hỗ trợ quản trị các node của DWH trên giao diện đồ họa bao gồm:
 - ✓ Trạng thái các node đang Active hoặc inActive.
 - ✓ Dung lượng các node.
 - ✓ Số các tiến trình đang chạy trên từng node.

Big Data Platform

Nền tảng dữ liệu lớn



Chức năng

Quản trị DataWarehouse (tiếp)

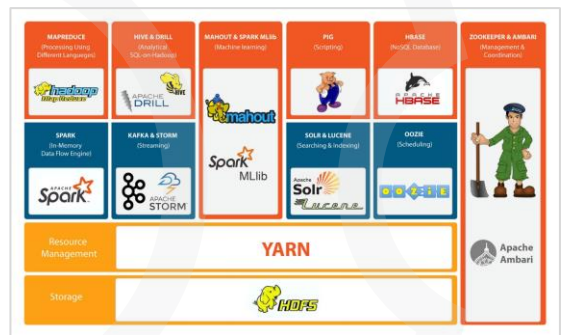
- Hỗ trợ xem trạng thái chi tiết của từng node:
 - ✓ Dung lượng đã sử dụng.
 - ✓ Bộ nhớ đã sử dụng.
 - ✓ Dung lượng mạng đã sử dụng.
 - ✓ Bộ xử lý đang sử dụng.
 - ✓ Thời gian đã sử dụng của node.
- Thiết lập hạn ngạch (quota) cho số lượng file HDFS sử dụng và dung lượng cho các thư mục.
- Quản lý các luồng công việc được xử lý trong cụm.

Triển khai

- Hỗ trợ tạo hàng ngàn các node và hàng tỷ các file phân tán.
- Hỗ trợ tích hợp lưu trữ với bên thứ 3 như: ADLS, WASB, S3, GCP. Nền tảng hỗ trợ triển khai trên nhiều server khác nhau hoặc chỉ trên 1 server.
- Hỗ trợ cấu hình backup trong một cluster, sao lưu cả cụm theo kiến trúc HDFS cluster.

- Nền tảng hỗ trợ backup lên cloud hoặc nhà cung cấp cloud.
- Xây dựng Data model cho chuyên ngành phục vụ báo cáo, thống kê, phân tích số liệu.

Công nghệ sử dụng



Công cụ SQL- on-Hadoop

Big Data Platform

Nền tảng dữ liệu lớn



Giải pháp & Thông số kỹ thuật



Ngôn ngữ lập trình

Python, Java,
Scala



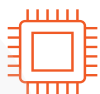
Cơ sở dữ liệu

Hive



Hệ điều hành

Linux Server



CPU

Central Processing Unit
64 Cores



RAM

Random Access Memory
128 GB



HDD

Hard Disk Drive
2 TB

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