Big data analytics architecture

Overview

SDS Cloud provides DW (Data Warehouse)-type **DB** Service by utilizing Vertica DB, which is a database dedicated to analysis. Vertica service allows you to easily and conveniently build and manage Vertica clusters and helps the UI to manage information and status of the cluster.

Vertica is designed in masterless Pure-MPP (Massively Parallel Processing) architecture and suitable for fast parallel analysis of large data. It offers in-DB machine learning and advanced analytics, storing various information from basic data to finance, health, monitoring, and event data. These diverse analysis functions help users to quickly extract and analyze desired information.

Various data storage, such as **Object Storage** and HDFS, is expected to be available later, enabling easy extraction and analysis of data anytime, anywhere

Architecture Diagram

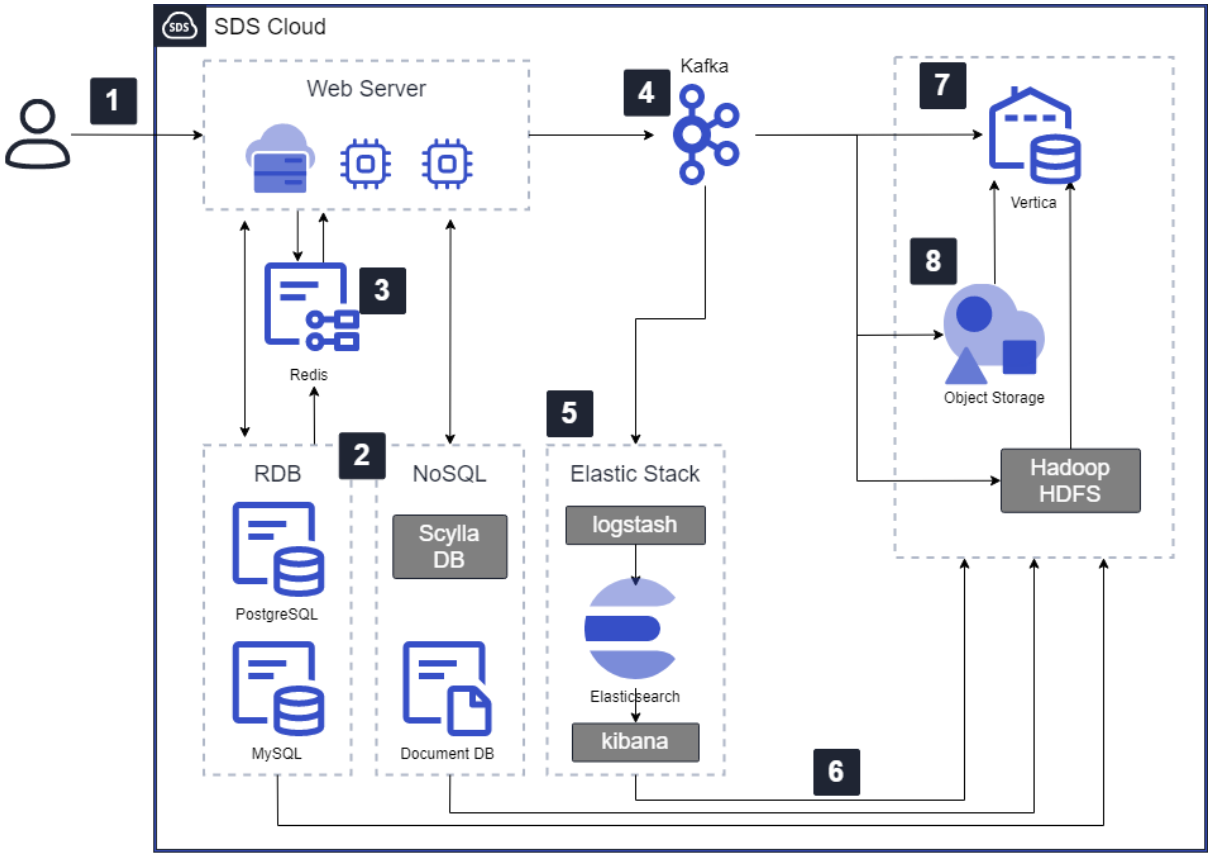

Figure 1. Big Data Analytics Architecture Diagram

1. Users can use various services (finance, healthcare, events, etc.).
2. For service continuity, RDB is composed of HA and NoSQL is composed of clusters to store a variety of data desired by users.
3. By using the Cache (Redis) service, contents are quickly provided to users, and session information is saved to shorten the processing time.
4. Using the messaging service (**Kafka**), data is transmitted to the target storage in real time or batch.
5. Using the Elastic Stack, data can be collected from multiple systems (Logstash), analyzed and searched (Elasticsearch), and various information can be provided by utilizing the visualization function (Kibana).
6. Data saved in RDB and NoSQL are transferred to Vertica through a batch and saved.
7. Analyze using various data stored in Vertica.
8. Data stored in Object Storage or HDFS can also be analyzed by linking with Vertica.

Use Cases

A. Building monitoring system

When it is necessary to check more than dozens of servers to analyze abnormal symptoms regularly, various information (server configuration, system log, SW inventory, security information, etc.) of the server is collected and stored using the collection agent in each server. You can check a problem object by mapping it with a pre-defined abnormal symptom pattern.

B. Data-driven hospital

Patients' treatment history (e.g. the name of disease, affected area, general symptoms, details, treatment status, etc.) from multiple hospitals in a contract is stored in one place. When a patient visits the hospital, a diagnosis is quickly made based on the symptoms and the treatment status and methods are provided to the patient.

Pre-requisites

Customer license(BYOL) is required to build Vertica services.

Limitations

Backup for Vertica provides a recovery function to a specific point in time (Snapshot) by performing incremental snapshot after the first full backup (not Transaction Log method). Backup by Object (Schema or Tables) will be provided in the second half of 2021. Services using **Object Storage** or HDFS are planned to be launched in 2022.

Considerations

According to Vertica license, data capacity is limited, and it can be a problem if data suddenly increases while using the service. You need to purchase a license after calculating the amount of data you want to store in advance.

Related Products

- DB Service

- Object Storage
- Kubernetes Apps
- Kafka
- Cloud Hadoop (To be available in 2022)