Cloud Hadoop-based data platform

Overview

For big data processing and analysis, a platform that can efficiently manage data collected in various forms is required.

SDS Cloud utilizes Cloud Hadoop to configure and manage Hadoop Clusters. Cloud Hadoop provides major components of the Hadoop ecosystem in a package and offers the Manager feature for management functions.
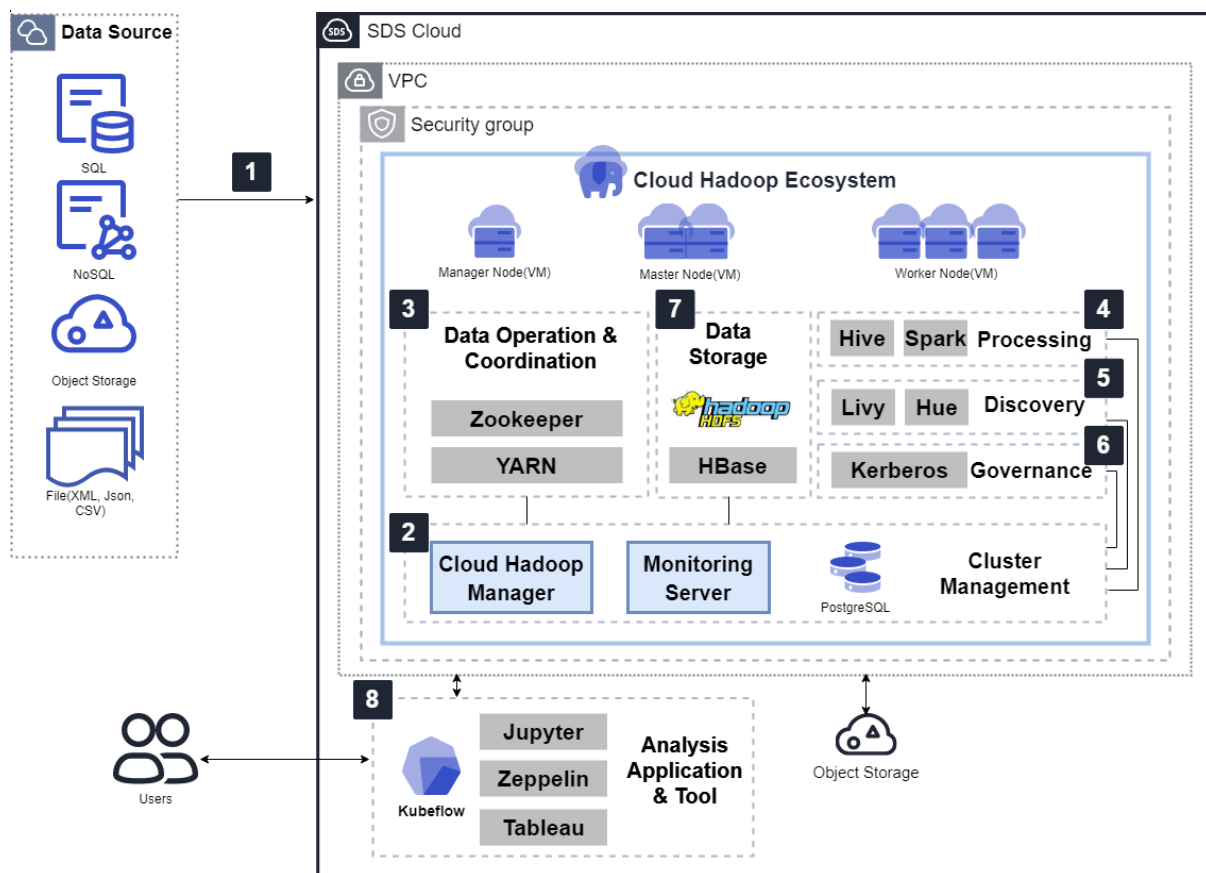
Architecture Diagram

Figure 1. Cloud Hadoop-based data platform configuration

1. Store data received from various data sources, such as RDB and NoSQL, in **Cloud Hadoop**

2. Manage Hadoop service configuration with Hadoop Manager and monitor Hadoop service through Monitoring Servers

3. Share information between distributed applications and check status

4. Data Processing: Distributed processing, refinement, and preparation of large data using Hive and Spark

5. Data Discovery: Explore data to gain new insights by evaluating large amounts of data
   ※ For Data Processing and Discovery, users may install additional S/W if necessary.

6. Data Governance: Create and change data according to data standards and policies, and manage the availability, usability, integrity and security of the generated data

7. Support various data storage for loading and utilizing large-capacity data:
   Warm data (HDFS), hot data (Kudu), and archived data (Object Storage)

8. Users may access, explore and analyze data in **Cloud Hadoop** using analytics solutions.

Use Cases

A. Building a big data integration platform

A big data platform can be built in a system that requires a big data platform integration offering for structured and unstructured data analysis, as well as real-time processing of large amounts of data and integrated monitoring for each node.

B. Building a data lake

By standardizing customer business data from the data collection/generation stage, you can systematically manage data and build a data lake platform for data-driven tasks.

C. Building a data platform

You can collect and load big data from various data sources and build an environment for distributed big data exploration and analysis tailored to your business environment.

Related Products

● Virtual Server

● Object Storage

● Kubeflow