

Cloud Hadoop 기반 데이터 플랫폼

개요

빅데이터 처리와 분석을 위해서는 여러가지 형태로 수집되는 데이터를 효율적으로 관리할 수 있는 플랫폼이 필요합니다.

SDS Cloud는 Cloud Hadoop을 활용하여 Hadoop Cluster를 구성하고 관리할 수 있습니다. Cloud Hadoop은 Hadoop 에코시스템의 주요 컴포넌트들을 패키징하여 제공하며, 관리 기능을 위한 Manager를 제공합니다.

아키텍처 다이어그램

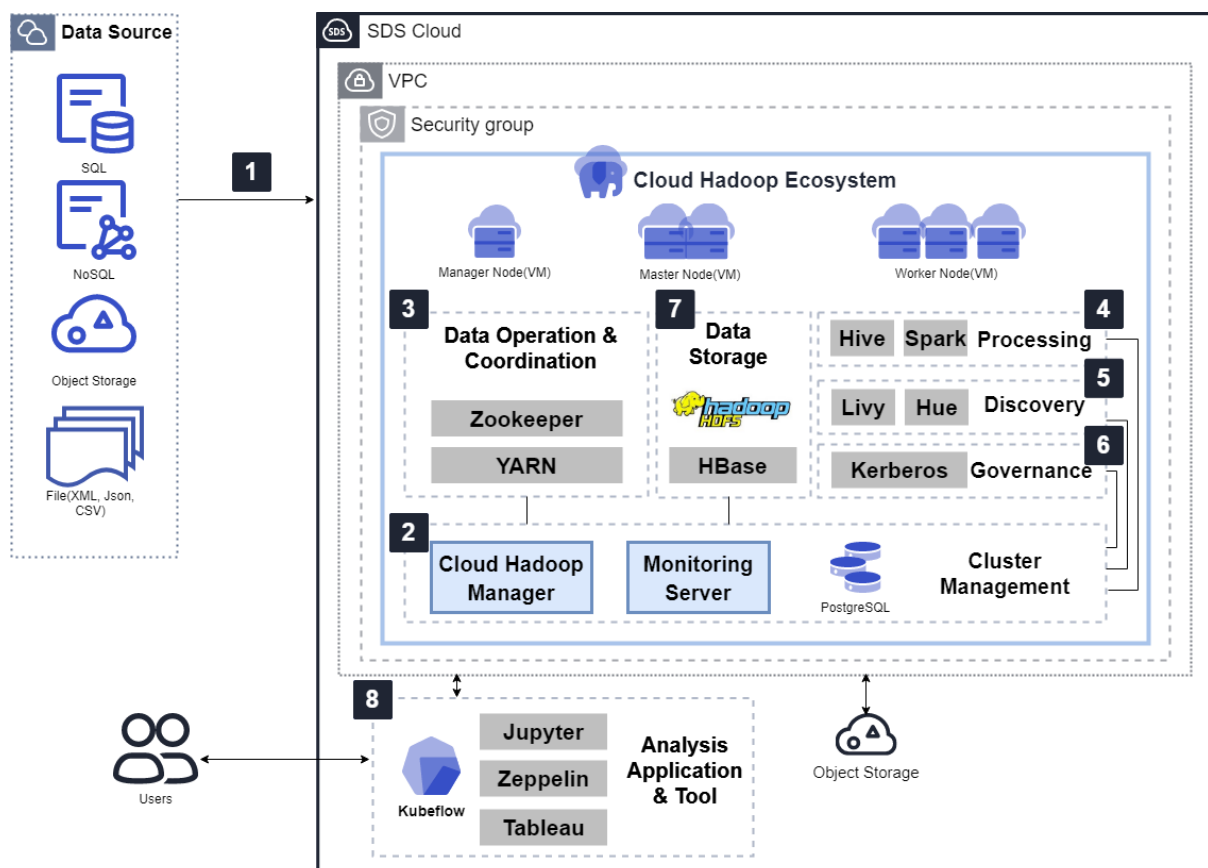


Figure 1. Cloud Hadoop 기반 데이터 플랫폼 구성

1. 다양한 데이터 소스(RDB, NoSQL, File등)로부터 받은 데이터를 Cloud Hadoop에 저장한다.
2. Hadoop Manager로 Hadoop 서비스 구성을 관리하고, Monitoring Server를 통해 Hadoop서비스를 모니터링한다
3. 분산 애플리케이션간의 정보를 공유하고 상태를 체크한다.
4. Data Processing: Hive와 Spark를 이용하여 대용량 데이터를 분산 처리, 정제(Refine), 준비(Preparation)한다.
5. Data Discovery: 대용량 데이터를 평가하여 새로운 통찰(Insight)을 얻기 위해 데이터를 탐색한다.
※ Data Processing, Discovery 부문은 사용자가 필요한 S/W를 추가 설치 가능
6. Data Governance: 데이터 표준·정책에 따라 데이터를 생성·변경하고 생성된 데이터의 가용성·유용성·무결성과 보안을 관리한다.
7. 대용량 데이터 적재·활용을 위한 다양한 Data Storage를 지원한다.
Warm Data(HDFS), Hot Data(Kudu), Archive Data(Object Storage)
8. 사용자는 분석 솔루션을 통해 Cloud Hadoop의 데이터에 접근하여 데이터를 탐색하고 분석한다.

사용 사례

A. 빅데이터 통합 플랫폼 구축

정형 및 비정형 데이터 분석을 위한 빅데이터 플랫폼 통합 오퍼링이 요구되고, 대량 데이터의 실시간 처리 및 노드 별 통합 모니터링 지원이 필요한 시스템에 빅데이터 플랫폼을 구축할 수 있습니다.

B. 데이터 레이크 구축

고객 비즈니스 데이터 수집/생성 단계부터 데이터를 표준화하여 데이터를 체계적으로 관리하고 데이터 기반 업무를 수행할 수 있는 데이터 레이크 플랫폼을 구축할 수 있습니다.

C. 데이터 플랫폼 구축

다양한 데이터 소스로부터 빅데이터를 수집 및 적재하고 고객의 비즈니스 환경에 맞춘 분산 환경의 빅데이터 탐색 및 분석 환경을 구축할 수 있습니다.

관련 상품

- Virtual Server
- Object Storage
- Kubeflow