대용량 데이터처리 아키텍처

개요

SDS Cloud에서는 분석 전용 Database인 Vertica DB를 활용하여 DW형 DB Service를 제공합니다. Vertica서비스는 Vertica 클러스터를 쉽고 간편하게 구축하고 관리할 수 있는 서비스이며, 클러스터에 대한 정보 및 상태를 관리할 수 있는 UI를 제공합니다.

Vertica는 Masterless의 Pure-MPP(Massively Parallel Processing) 아키텍처로 설계되어 대용량 데이터를 병렬로 빠르게 분석하기에 적합하며, In-DB 머신 러닝 및 고급 분석 기능을 포함하고 있습니다. 기준 정보는 물론 금융, 헬스, 모니터링, 이벤트 발생 정보 등을 저장하며, 다양한 분석기능을 활용하여 사용자가 원하는 정보를 빠르고, 신속하게 추출 및 분석 할 수 있습니다.

향후에는 **Object Storage** 또는 HDFS등 다양한 데이터저장소를 사용하여 언제 어디에서든 쉽게 추출하여 분석할 수 있습니다.

아키텍처 다이어그램

Copyright 2021. Samsung SDS Co., Ltd. All rights reserved.

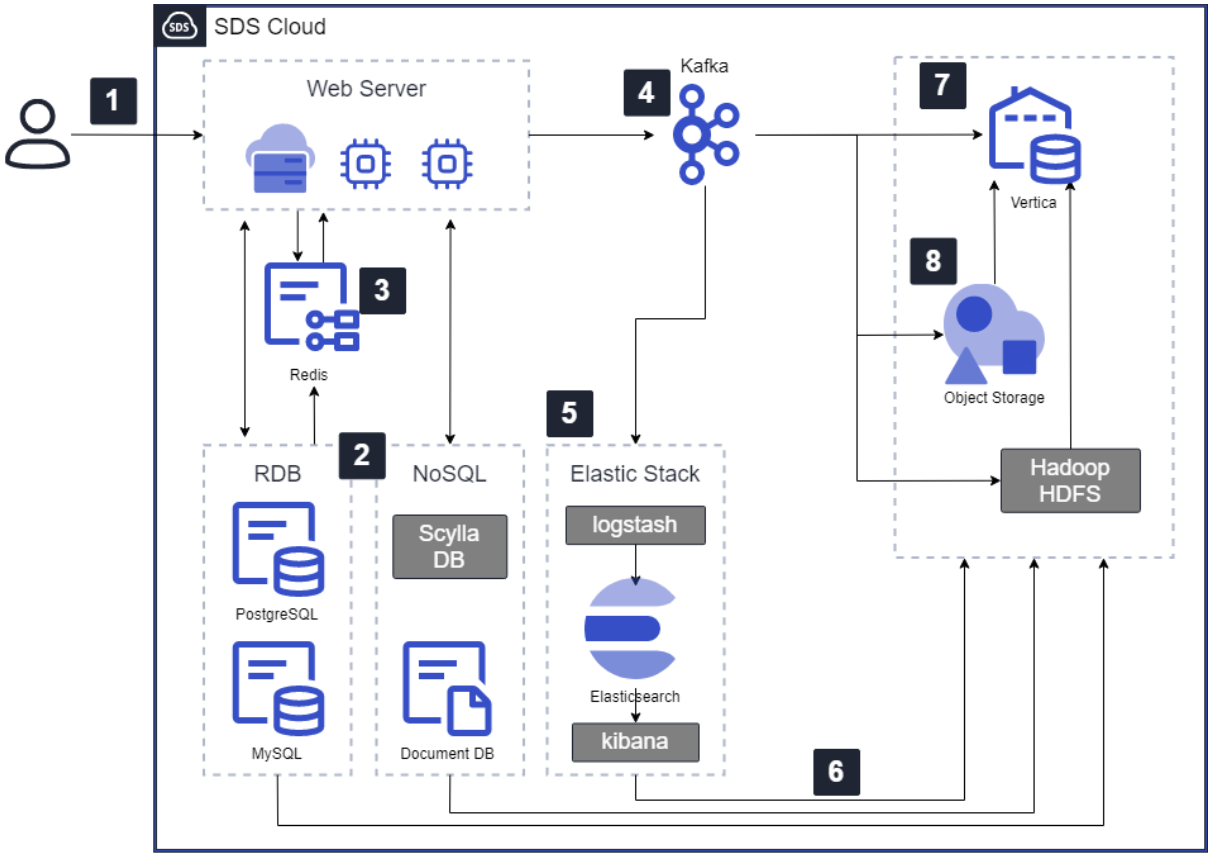

Figure 1. 대용량 데이터처리 아키텍처

1. 사용자가 다양한 서비스(금융, 병원, 이벤트)를 이용한다.
2. 서비스 지속성을 위해서 RDB는 HA, NoSQL은 Cluster로 구성하여 사용자가 원하는 다양한 데이터를 저장한다.
3. Cache(Redis)서비스를 이용하여 콘텐트를 사용자에게 빠르게 제공하고, 세션 정보를 저장하여 처리시간을 단축한다.
4. 메시지서비스(**Kafka**)를 이용하여 실시간 또는 배치로 목적에 맞는 Target 저장소로 데이터를 전송한다.
5. Elastic Stack를 활용하여 여러 시스템에서 데이터를 수집(Logstash)하고, 분석 및 검색하며(Elasticsearch), 시각화 기능(Kibana)을 활용하여 다양한 정보를 제공 받을 수 있다.
6. RDB 및 NoSQL등에 저장된 데이터를 배치를 통해서 Vertica로 전송하여 저장한다.
7. Vertica에 저장된 다양한 데이터를 활용하여 분석한다.
8. **Object Storage** 또는 HDFS에 저장된 데이터 또한 Vertica로 연계하여 데이터 분석이 가능하다.

Copyright 2021. Samsung SDS Co., Ltd. All rights reserved.

사용 사례

A. 모니터링 시스템 구축

주기적으로 수십 대 이상의 서버를 점검해서 이상징후를 분석해야 할 경우 각 서버 내 수집 Agent를 활용하여 서버의 다양한 정보(서버 설정 정보, 시스템 로그, SW설치정보, 보안정보 등)를 수집하여 저장합니다. 사전에 정의한 이상징후 패턴과 mapping하여 문제가 되는 대상을 선정할 수 있습니다.

B. 데이터 중심 병원

계약관계에 있는 여러 병원의 환자의 진료 내역(병명, 발생 부위, 일반적인 증상, 특이사항, 치료 현황 등)을 한 곳에서 저장합니다. 환자가 내원하였을 경우 증상을 기반으로 진단을 빠르게 할 수 있고, 치료 현황과 방법 등을 환자에게 제공할 수 있습니다.

선결 사항

Vertica 서비스 구축을 위해서는 고객 라이선스 사용(BYOL)이 필요합니다.

제약 사항

Vertica에 대한 백업은 최초 Full 백업 이후 증분(Incremental Snapshot) 백업하는 방식으로 특정 시점(Snapshot)으로 복구 기능을 제공합니다(Transaction Log 방식 아님). Object(Schema 또는 Tables)단위의 백업은 21년 하반기에 제공될 예정입니다. Object Storage 또는 HDFS를 활용한 서비스는 22년에 계획하고 있습니다.

고려 사항

Vertica의 라이선스에 따라 데이터 용량에 제약이 있으며, 서비스 이용 중 갑작스럽게 데이터가 증가 될 경우 문제가 될 수 있습니다. 사전에 저장하고자 하는 데이터 용량을 충분히 산정 한 후에 라이선스 구매가 필요합니다.

관련 상품

Copyright 2021. Samsung SDS Co., Ltd. All rights reserved.

- DB Service
- Object storage
- Kubernetes Apps
- Kafka
- Cloud Hadoop (‘22년 출시 예정)

Copyright 2021. Samsung SDS Co., Ltd. All rights reserved.