Data Warehouse Database-based Analysis

Overview

When it comes to enterprise-class data analysis, data is collected from various data sources to build a Data Warehouse(hereinafter DW) data analysis is performed based on this.

Samsung Cloud Platform provides a Data Warehouse-type database service based on **VMWare Greenplum**. **VMWare Greenplum(DBaaS)** is a parallel processing analysis platform for large data sets that is based on PostgreSQL. It supports enhanced transaction processing and streaming collection to handle analysis workload of various data sets.

Architecture Diagram

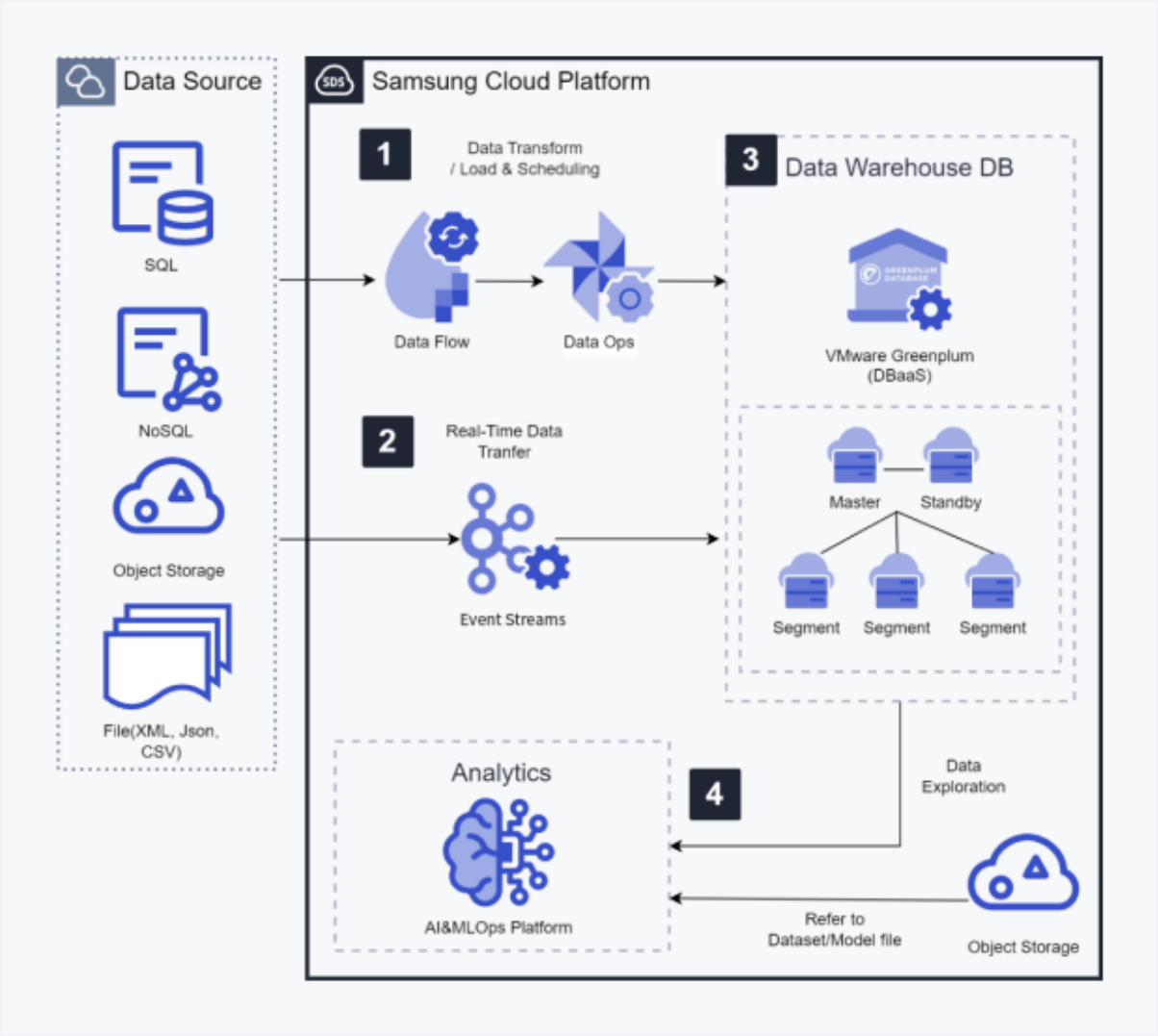

Figure 1. Data Warehouse Database-based analysis

1. Perform extraction/transformation/loading and scheduling for necessary data other than the Data Warehouse Database(hereinafter DW DB) data. (Data flow/Data Ops can be used)
2. Real-time data that are required (e.g., log data) is linked to the DW DB. (**Event Streams** can be used)
3. Retrieves and analyzes data within the DW DB(**Greenplum**). **Greenplum** consists of a master node and a segment node, and the master node adds a standby node to redundancy to ensure service continuity even in the event of a failure.
4. Other analysis tools such as **AI&MLOps Platform** conduct data search from DW DB and use it for analysis.

Use Cases

A. Data analysis in the manufacturing sector

Users can collect manufacturing-related log data in real time, load large amounts of data, and view and analyze large amounts of data through a high-performance query engine.

B. Data analysis in the financial sector

Based on the ETL solution, data can be loaded from multiple related systems, viewed/analyzed within the DW, or be used for data analysis in conjunction with the Analytics Platform.

Prerequisites

None

Limitations

None

Considerations

None

Related Products

- VMWare Greenplum(DBaaS)
- Data Flow
- Data Ops
- Event Streams
- AI&ML Ops Platform

Related Documents

None