BRIEF REPORT OF APACHE DORIS

I. SUMMARY OF THE TOPIC

Apache Doris is a high-performance, real-time analytical database based on MPP architecture, known for its extreme speed and ease of use. It only requires a sub-second response time to return query results under massive data and can support not only high-concurrent point query scenarios but also high-throughput complex analysis scenarios. All this makes Apache Doris an ideal tool for scenarios including report analysis, ad-hoc query, unified data warehouse, and data lake query acceleration.

II. SCOPE OF RESEARCH

- Understand the concept, architecture, and how Apache Doris works.
- What situations require Apache Doris application.
- Advantages and disadvantages of Apache Doris and compared it with other similar products.
- Know how to install and configure Apache Doris and use Apache Doris to query.

III. TECHNICAL OVERVIEW

The Apache Doris architecture is simple and neat, with only two types of processes.

- Frontend (FE): user request access, query parsing and planning, metadata management, node management, etc.
- Backend (BE): data storage and query plan execution

Both types of processes are horizontally scalable, and a single cluster can support up to hundreds of machines and tens of petabytes of storage capacity. And these two types of processes guarantee high availability of services and high reliability of data through consistency protocols. This highly integrated architecture design greatly reduces the operation and maintenance cost of a distributed system.