1. **BigQuery is a serverless model**

**Serverless: no servers to manage or database software to install.**

1. **BigQuery Basic Aritecture: Dremel, Jupiter, Colossus and Brog**

**Dremel: the execution engine for the BigQuery, a highly scalable system. → Dremel implements a multi-level serving tree to execute queries, has the capability to handle terabytes of data in seconds**

**BigQuery client: Normally would be the BigQuery Web UI or bg command-line tool or REST APIs → Client interacts with Dremel engine via a client interface**

**Borg: Google’s large-scale cluster management system → It allocates the compute capacity for the Dremel jobs**

**Colossus: Google’s latest generation distributed file system → Colossus handles replication, recovery (when disks crash), and distributed management**

**Jupiter Network: It is the internal data center network that allows BigQuery to separate storage and compute**

**Dremel jobs then read data from Google’s Colossus file systems using Jupiter network, perform various SQL operations and return results to the client.**

1. **BigQuery uniquely incorporates a massively parallel query engine (Dremel) that enables it to efficiently analyze large-scale datasets**

**Dremel allows BigQuery to process complex queries with the help of multiple servers in parallel which significantly improves the processing speed.**

**这种 BigQuery 架构允许它在多个服务器的帮助下并行处理复杂的查询，从而显着提高处理速度。**

**Dremel Detailed Process:** Root receives incoming queries, distributed it to all its leaf nodes, leaf nodes works in parallel, final results are sent back to the root by the mixer after accumulating the results of all the leaf node

1. **Big Query extremely high cost-effectiveness and full-scan performance for data queries**

**Explain: BigQuery stores data in a columnar format known as Capacitor, each column of data stored in a separate capacitor file enable the BigQuery to achieve high compression ratio and scan throughput.**

**Capacitor enabled BigQuery to directly operate on compressed data**

1. **BigQuery’s high scalability and high availability**

**Scalability:** BigQuery relies on massively parallel computing and a highly scalable and secure storage engine to offer users true scalability and consistent performance. A Complex software stack manages the entire infrastructure that runs into thousands of machines per region

**Availability:** Due to the use of the Colossus system, BigQuery doesn’t have to worry about the crash problem for its ability to handle replication, recovery, and distributed management