BDA5-5

1

Hello everyone, I am Haiying Che, from Institute of Data Science and knowledge Engineering

School of Computer Science, in Beijing Institute of Technology,

in this session, we discuss Massively Parallel Processing model for Structured Data.

2

The data processing system provides big data computing and processing capabilities and an application development platform.

From the perspective of computing architecture, the data processing system is divided into data algorithm layer, computing model layer, computing platform layer, computing engine layer, etc.

Computing models are the way that different kinds of big data is processed in different scenarios,

which include batch processing, stream computing, **Large-scale concurrent processing (MPP) model** for structured data, In-memory Computing model, and Data Flow Graph models.

Now let’s look at the third computing model **Massively Parallel Processing.**

3

In Massively Parallel Processing (MPP) databases data is partitioned across multiple servers or nodes with each server/node having memory/processors to process data locally.

All communication is via a network interconnect — there is no disk-level sharing or contention to be concerned with (i.e. it is a ‘shared-nothing’ architecture).

4

Big data may be semi-structured or unstructured.

The **massively parallel processing** (**MPP**) architecture structures big data to enable easy querying for reporting and analytic purposes.

MPP systems are sometimes referred to as shared nothing systems.

This means that data is partitioned across many servers (otherwise known as nodes) and each server processes queries locally.

In the diagram , 4 nodes are using to replace one single server and reduce the query time from one hour to 15 minutes.

5

The process begins by the Client issuing a query that is then passed to the Master Node.

The Master Node contains information, such as the data dictionary and session information,

which it uses to generate an execution plan designed to retrieve the needed information from each underlying Node.

Parallel Execution represents the implementation of the execution plan through the parallel computing of Node 1 to Node n.

And the query results return to master node

6

**Massively Parallel Processing is the coordinated processing of a single task by multiple processor,**

**Each processor using its own OS and memory and communicating with each other using some form of messaging interface**

**MPP can be setup with a shared nothing or shared disk architecture**

**In the shared nothing architecture, there is no single point of contention across the system and nodes do not share memory or disk storage.**

**data is horizontally partitioned across nodes such that each node has a sub set of rows from each table in the DB**

**Each node then processes only rows on its own disks.**

The “shared nothing” systems with distributed databases need a lot of coordination to complete a common task.

Each node owns slices of the database.

Managing this database could be very difficult. Shared nothing systems with the replicated database are not suitable for applications with tremendous data requirements.

If the computation needs a lot of data modification operations like data insertion and join, then the “shared nothing” architecture may not be viable.

7

Performance through segment instance parallelism

Master host and standby master Host

Master coordinates work with segment host

Segment host with one or more segment instances

Segment instances process queries in parallel

Segment hosts have their own CPU disk memory(shared nothing)

High speed interconnect for continuous pipelining of data processing

8

Architecture of MPP databases. There are a master and 4 slaves, with 2 segments on each slave and a mirror for each segment on another machine.

MPP Database Architecture

Typical MPP databases usually adopt a shared-nothing architecture , composed of one master node and *n* slave nodes.

The master node is responsible for interacting with clients, managing the whole cluster and coordinating the query processing.

Each of the *n* salve nodes is responsible for storing a partition of the data and performing query processing on its partition.

Each slave node hosts *d* database instances, which will be referred to as *segments* subsequently.

Most MPP databases provide fault tolerance at storage level. A mirror scheme, i.e., replication, is commonly used to ensure durability and availability of data.

As Fig shows, each segment (*primary* segment) is allocated with a mirror (*mirror* segment) in another node.

The master node detects node failures by monitoring heartbeats of slave nodes. If a slave node stops responding for a certain amount of time, known as a system delay time (normally around 1 min),

the master will treat it as a failed node. Once a failure is detected, the corresponding mirror will be activated to replace the failed primary segment.

Similarly, a *standby* works as the replication/mirror of the master node.

Through the mirror scheme, the system’s availability can be significantly enhanced.

However, such a mirror scheme does not support intra-query fault tolerance automatically.

When a node failure occurs, the running query’s state on the failed node will be lost.

After the corresponding mirror is activated, the whole query has to be rerun.

If it is a long running query, response to the client will be severely delayed. In the worst case, a query will run indefinitely, if the probability of failure is high.

9

in Massively Parallel Processing (MPP) databases data is partitioned across multiple servers or nodes with each server/node having memory/processors to process data locally. All communication is via a network interconnect — there is no disk-level sharing or  
contention to be concerned with (i.e. it is a ‘shared-nothing’ architecture).

1. Master Host – Separate physical server with its own OS/CPU/storage/memory.
2. Hosts master database. There is no user data in master database but stores **metadata about segments – think in terms of system tables**.

3、 2, 3, 4 Segment hosts – Individual physical servers with their own OS/CPU/storage/memory.

4、Hosts segment database. Each database stores portion of user data.

5、 Interconnect switch – Segment server databases communicate through an interconnect switch

the main characteristic of MPP database is data distribution.

Data is distributed across each segment database to achieve data and processing parallelism.

This is achieved by creating a database table with **DISTRIBUTED BY** clause.

By using this clause data is automatically distributed across segment databases.

In Greenplum you can either use hash or round-robin distribution.

10

In this session we learned **Large-scale concurrent processing (MPP** Massively Parallel Processing) **model** for structured data. Thank you for your attention, if you have any question, feel free to contact me.