Aster Data SQL and MapReduce

Class Outline

CoffingDW education has been customized for every customer for the past 20 years. Our classes can be taught either on site or remotely via the internet.

Education Contact: Thomas Coffing
513-300-0341
Tom.Coffing@coffingdw.com
CoffingDW.com
**Description:** In this course, participants will learn Aster Data SQL and MapReduce beginning at a basic level and moving to more advanced functions.

**Objectives:** At the completion of this course, participants will have the skills to write and understand Aster Data SQL and MapReduce.

**Topics:**
- The Aster Data Architecture
- Four Options for Aster Data Table Design
- How Joins Work Inside the Aster Engine
- Temporary and Analytic Tables
- Tera-Tom’s Top Tips
- Indexes
- SQL-MapReduce
- Time and Date
- Aster Windows Functions
- The Fundamental SQL Commands That Work on Aster

**Audience:** This course is designed for all users of Aster Data to help give them the skills to write and understand Aster Data SQL and MapReduce.

**Prerequisites:** None
Duration: 1-2 Days

Outline:

Chapter 1 – The Aster Data Architecture
What is Parallel Processing?
Aster Data is a Parallel Processing System
Each vworker holds a Portion of Every Table
The Rows of a Table are Spread Across All vworkers
Aster Tables are defined as Fact or Dimension when Created
Fact Table
A More Detailed Look at the Fact Table Distribution
Dimension Table are Replicated
A Dimension Table is often Replicated across vworkers
Aster Data has Fact and Dimension Tables
Aster Tables are defined as Fact or Dimension when Created
Fact and Dimension Tables can be Hashed by the same Key
Distribution Key Rules
Aster Data Uses a Hash Formula
The Hash Map Determines which vworker will own the Row
The Hash Formula, Hash Map and vworker
Placing rows on the vworker
Placing rows on the vworker Continued
A Review of the Hashing Process
Like Data Hashes to the Same vworker
Distribution Key Data Types
Run ANALYZE to COLLECT STATISTICS on a Table
Some Examples of ANALYZE
What Columns to Analyze
Chapter 2 - Four Options for Aster Data Table Design

There are Four Options to Aster Table Design

Straight up Distribute by Hash

Straight up Distribute by Hash - Problems

Straight up Distribute by Replication

Partition the Table with Logical Partitioning

This Partitioned Table Sorts Rows by Month of Order_Date

An All vworkers Retrieve By Way of a Single Partition

You can Partition a Table by Range or by List

A Partitioned By List Example with Three Tactical Queries

Aster Data Multi-Level Partitioning

Aster Allows for Multi-Level Partitioning

SQL Commands for Logical Partitioning as One Table

What Partitions are on my Table?

What does a Columnar Table look like?

A Comparison of Data for Normal Vs. Columnar

A Columnar Table is best for Queries with Few Columns

Quiz – How Many Blocks Move to vworker Memory?

Answer – How Many Containers Move to vworker Memory?

When to use a Columnar Table

Chapter 3 - How Joins Work Inside the Aster Engine

Aster Join Quiz

Aster Join Quiz Answer

The Joining of Two Tables

Aster Moves Joining Rows to the Same vworker

Because of the Join Rule – Dimension Table are Replicated

The Two Different Philosophies for Table Join Design

What Could You Do If Two Tables Joined 1000 Times a Day?
Fact and Dimension Tables can be Hashed by the same Key
Joining Two Tables with the same PK/FK Distribution Key
A Join With Co-Location
A Performance Tuning Technique for Large Joins
The Joining of Two Tables with an Additional WHERE Clause
Aster Performs Joins Using Three Different Methods
The Hash Join
The Merge Join
Nested Loop Joins

Chapter 4 - Temporary and Analytic Tables
Aster has Three Types of Data
Create a Permanent Table Using Create Table AS (CTAS)
Create a Logically Partitioned Table and Populate It
Create a Temporary Table with using Create Table AS (CTAS)
A Temporary Table in Action
A Temporary Table That Uses an Insert/Select
Create an Analytic Table Using an Insert/Select
Create an Analytic Table Using CREATE TABLE AS (CTAS)
Operations that Invalidate an Analytic Table
If an Analytic Table is Invalid
Tera-Tom History

Chapter 5 – Tera-Tom’s Top Tips
Tera-Tom's Top Tips
Tera-Tom's Top Tips # 2
Tera-Tom's Top Tips #3
Tera-Tom's Top Tips # 3 Rewritten
Tera-Tom's Top Tips #4
When the GROUP BY Column is NOT the Distribution Key
Example of GROUP BY Column is NOT the Distribution Key

Tera-Tom's Top Tips #5

Tera-Tom's Top Tips #6 – Use EXPLAIN Query Plan and Estimates

Explain Plan Showing a Hash Join

Explain Plan Showing a Merge Join

Explain Plan Showing a Nested Loop Join

**Chapter 6 - Indexes**

There are Only Three Types of Scans

Guidelines for Indexes

An Index Syntax Example

The B-Tree Index

Which Columns Might You Create an Index?

Answer - Which Columns Might You Create an Index?

A Visual of an Index (Conceptually)

A Query Using an Index Uses All vworkers

Multicolumn indexes

A NUSI BITMAP Theory

A NUSI Bitmap in Action

Indexes on Expressions

Indexes on Extracts of Dates

GiST Indexes

Five Operational Tips for Efficient Indexing

REINDEX

createCompressedIndexOnCompressedTableByDefault Flag

**Chapter 7 – SQL-MapReduce**

MapReduce History

What is MapReduce?
What is SQL-MapReduce?
SQL-MapReduce Input
SQL-MapReduce Output
Subtle SQL-MapReduce Processing
Aster Data Provides an Analytic Foundation
Path Analysis
Text Analysis
Statistical Analysis
Segmentation (Data Mining)
Graph Analysis
Transformation of Data
Sessionize
Tokenize
SQL-MapReduce Function… nPath
nPath SELECT Clause
nPath ON Clause
nPath PARTITION BY Expression
nPath DIMENSION Expression
nPath ORDER BY Expression
nPath MODE Clause has Overlapping or NonOverlapping
nPath PATTERN Clause
Pattern Operators
Pattern Operators Order of Precedence
Matching Patterns Which Repeat
nPath SYMBOLS Clause
nPath RESULTS Clause
Adding an Aggregate to nPath Results
Adding an Aggregate to nPath Results (Continued)
SQL-MapReduce Examples - Use Regular SQL
SQL-MapReduce Examples - Create Objects
SQL-MapReduce Examples - Subquery
SQL-MapReduce Examples - Query as Input
SQL-MapReduce Examples - Nesting Functions
SQL-MapReduce Examples - Functions in Derived Tables
SQL-MapReduce Examples - SMAVG
SQL-MapReduce Examples - Pack Function
SQL-MapReduce Examples - Pack Function (Continued)
SQL-MapReduce Examples - Pivot Columns
Workshop: Create This Table
Login to your GNOME Terminal
Login to your Linux
Using the GNOME Terminal Unzip the bank_web_data.zip
Use the Function ncluster_loader to Load the Bank Data
Run this nPath Map Reduce Function on your Table
nPath in Action
Operators at their Simplest
Pattern
Accumulate
Accumulate With All Pages
Accumulate – nPath with a WHERE Clause
SQL-MapReduce Examples - Path Generator
SQL-MapReduce Examples - Path Generator (Continued)
SQL-MapReduce Examples - Path Generator (Continued)
SQL-MapReduce Examples - Path Generator (Continued)
SQL-MapReduce Examples - Path Generator (Continued)
SQL-MapReduce Examples - Linear Regression
SQL-MapReduce Examples - Linear Regression (Continued)
SQL-MapReduce Examples - Linear Regression (Continued)
SQL-MapReduce Examples - Naive Bayes
SQL-MapReduce Examples - Naive Bayes (Continued)
Join Aster, Teradata and Hadoop Tables; feed into MapReduce
Run Both of these Examples Together and Compare
Run this nPath Map Reduce Function
nPath in Action
Another nPath Example
Finding Out What Functions You Have Installed
Workshop 1 – Fill in the x's
Answer Workshop 1 - Fill in the x's
Workshop 2 – Fill in the x's
Answer Workshop 2 – Fill in the x's
Answer Workshop 2 – You Could Have Used a GROUP BY
Workshop 3 – Add to the Query
Workshop 3 – Answer to Add to the Query
Workshop 4 – Fill in the x's
Answer to Workshop 4 – Fill in the x's
Workshop 5 – Find that Customer
Answer to Workshop 5 – Find that Customer
Workshop 6 – Change the MapReduce Function
Answer to Workshop 6 – Change the MapReduce Function
Workshop 7 – Build the MapReduce Function
Answer to Workshop 7 – Build the MapReduce Function
Best Answer to Workshop 7 – Build the MapReduce Function
Workshop 8 – Build the Accumulate in the Result
Answer to Workshop 8 – Build the Accumulate in the Result
SQL-MapReduce Examples - Linear Regression (Continued)

Workshop 9 – Build the Subquery
Answer to Workshop 9 – Build the Subquery

Workshop 10 – Do Your First Join
Answer to Workshop 10 – Do Your First Join
Answer to Workshop 10 – Do the Join Using a New Syntax

Workshop 11 – Super Join the Tables
Answer to Workshop 11 – Super Join the Tables
Answer to Workshop 11 – Super Join the Tables

Workshop 12 – Sessionize the Data
Answer to Workshop 12 – Sessionize the Data

Workshop 13 – What is this Query Doing?
Answer to Workshop 13 – What is this Query Doing?

Workshop 14 – Using ilike
Answer to Workshop 14 – Using ilike
Answer to Workshop 14 – Using ilike

Workshop 15 – What are the First Two Pages Visited?
Workshop 15 – What are the First Two Pages Visited?
Workshop 16 – Advanced - First Two Pages Visited?
Answer to Workshop 16 Advanced - First Two Pages Visited?

Workshop 17 – Can You Clean Up the Results?
Answer to Workshop 17 – Can You Clean Up the Results?
Answer to Workshop 17 – Format the Date

Workshop 18 – Build a Churn Table
Workshop 18 – Run the Query Before Building to Test
Workshop 18 – A Better Example
Answer to Workshop 18 – Build a Basic Churn Table

Workshop 18 – Create the Churn Table with a Better Example

Multi-Case

The Multi-Case Function
The Multi-Case Function in Nexus
The Multi-Case Function Mixing and Matching
The Multi-Case Function Mixing and Matching
SQL-MapReduce Examples - cFilter
SQL-MapReduce Examples - cFilter (Continued)
SQL-MapReduce Examples - Linear Regression (Continued)
SQL-MapReduce Examples - cFilter (Continued)
SQL-MapReduce Examples - Linear Regression (Continued)
SQL-MapReduce Examples - cFilter (Continued)
SQL-MapReduce Examples - cFilter (Continued)
SQL-MapReduce Examples - cFilter (Continued)
SQL-MapReduce Examples - cFilter (Continued)
SQL-MapReduce Examples - cFilter (Continued)
SQL-MapReduce Examples - cFilter (Continued)
SQL-MapReduce Examples - cFilter (Continued)
SQL-MapReduce Examples - cFilter (Continued)
CFILTER in Action with Bank_Web_Clicks
CFILTER in Action
CFILTER using Nexus
nPath Error

Chapter 8 – Time and Date
Date, Time, and Timestamp Keywords
Add or Subtract Days from a date
The to_char command
A Summary of Math Operations on Dates
Using a Math Operation to find your Age in Years
Find What Day of the week you were Born
Date Related Functions
The EXTRACT Command
EXTRACT from DATES and TIME
EXTRACT with DATE and TIME Literals
EXTRACT of the Month on Aggregate Queries
A Side Title example with Reserved Words as an Alias
Implied Extract of Day, Month and Year
DATE_PART Function
DATE_TRUNC Function
DATE_TRUNC Function using TIME
Aster NOW() Function

Chapter 9 – Aster Windows Functions
Cumulative Sum
Cumulative Sum - Major and Minor Sort Key(s)
The ANSI CSUM – Getting a Sequential Number
The ANSI OLAP – Reset with a PARTITION BY Statement
PARTITION BY only Resets a Single OLAP not ALL of them
ANSI Moving Sum is Current Row and Preceding n Rows
How ANSI Moving SUM Handles the Sort
Quiz – How is that Total Calculated?
Answer to Quiz – How is that Total Calculated?
Moving SUM every 3-rows vs. a Continuous Sum
Moving Average
Quiz – How is that Total Calculated?
Answer to Quiz – How is that Total Calculated?
Quiz – How is that 4th Row Calculated?
Answer to Quiz – How is that 4th Row Calculated?
Partition By Resets an ANSI OLAP
Moving Average Using BETWEEN
Moving Difference using ANSI Syntax
Moving Difference using ANSI Syntax with Partition By
RANK Defaults to Ascending Order
Getting RANK to Sort in DESC Order
You can use Window Functions in Expressions
RANK() OVER and PARTITION BY
DENSE_RANK() OVER
PERCENT_RANK() OVER
PERCENT_RANK() OVER with 14 rows in Calculation
PERCENT_RANK() OVER with 21 rows in Calculation
RANK With ORDER BY SUM()
COUNT OVER for a Sequential Number
Quiz – What caused the COUNT OVER to Reset?
Answer to Quiz – What caused the COUNT OVER to Reset?
The MAX OVER Command
MAX OVER with PARTITION BY Reset
The MIN OVER Command
Quiz – Fill in the Blank
Answer to Quiz – Fill in the Blank
The Row_Number Command
Quiz – How did the Row_Number Reset?
Quiz – How did the Row_Number Reset?
NTILE
NTILE Using a Value of 10
NTILE With a Partition
CUME_DIST
CUME_DIST With a Partition
LEAD
LEAD With Partitioning
LAG
LAG with Partitioning
FIRST_VALUE
FIRST_VALUE After Sorting by the Highest Value
FIRST_VALUE with Partitioning
LAST_VALUE
NTH_VALUE
NTH_VALUE With Partition
SUM(SUM(n))

Chapter 10 – The Fundamental SQL Commands That Work on Aster

BETWEEN is Inclusive
BETWEEN Works for Character Data
LIKE uses Wildcards Percent ‘%’ and Underscore ‘_’
LIKE command Underscore is Wildcard for one Character
GROUP BY Vs. DISTINCT – Good Advice
The Five Aggregates of Aster Data
GROUP BY when Aggregates and Normal Columns Mix
GROUP BY Delivers one row per Group
GROUP BY Dept_No or GROUP BY 1 the same thing
Limiting Rows and Improving Performance with WHERE
WHERE Clause in Aggregation limits unneeded Calculations
Keyword HAVING tests Aggregates after they are Totaled
Keyword HAVING is like an Extra WHERE Clause for Totals
Getting the Average Values per Column
Getting the Average Values per Column
Average Values per Column for All Columns in a Table
A two-table join using Non-ANSI Syntax
A two-table join using Non-ANSI Syntax with Table Alias
Aliases and Fully Qualifying Columns
A two-table join using ANSI Syntax
Both Queries have the same Results and Performance
Quiz – Can You Finish the Join Syntax?
Answer to Quiz – Can You Finish the Join Syntax?
Quiz – Can You Find the Error?
Answer to Quiz – Can You Find the Error?
Quiz – Which rows from both tables Won’t Return?
Answer to Quiz – Which rows from both tables Won’t Return?
LEFT OUTER JOIN
LEFT OUTER JOIN Brings Back All Rows in the Left Table
RIGHT OUTER JOIN
RIGHT OUTER JOIN Brings Back All Rows in the RIGHT Table
FULL OUTER JOIN
FULL OUTER JOIN Brings Back All Rows in All Tables
Which Tables are the Left and which are the Right?
Answer - Which Tables are the Left and which are the Right?
INNER JOIN with Additional AND Clause
ANSI INNER JOIN with Additional AND Clause
ANSI INNER JOIN with Additional WHERE Clause
OUTER JOIN with Additional WHERE Clause
OUTER JOIN with Additional AND Clause
Results from OUTER JOIN with Additional AND Clause
Quiz – Why is this considered an INNER JOIN?
The DREADED Product Join
Result Set of the DREADED Product Join
The Horrifying Cartesian Product Join
The ANSI Cartesian Join will ERROR
Quiz – Do these Joins Return the Same Answer Set?
Answer – Do these Joins Return the Same Answer Set?
How would you Join these two tables?
How would you Join these two tables? You Can’t Yet!
An Associative Table is a Bridge that Joins Two Tables
Quiz – Can you Write the 3-Table Join?
Answer to Quiz – Can you Write the 3-Table Join?
Quiz – Can you Write the 3-Table Join to ANSI Syntax?
Answer – Can you Write the 3-Table Join to ANSI Syntax?
Quiz – Can you Place the ON Clauses at the End?
Answer – Can you Place the ON Clauses at the End?
The 5-Table Join – Logical Insurance Model
Quiz - Write a Five Table Join Using ANSI Syntax
Answer - Write a Five Table Join Using ANSI Syntax
Quiz - Write a Five Table Join Using ANSI Syntax
Answer - Write a Five Table Join Using ANSI Syntax
Quiz - Write a Five Table Join Using Non-ANSI Syntax
Answer - Write a Five Table Join Using Non-ANSI Syntax
Quiz –Re-Write this putting the ON clauses at the END
Answer –Re-Write this putting the ON clauses at the END
The Nexus Query Chameleon Writes the SQL for Users.
An IN List is much like a Subquery
An IN List Never has Duplicates – Just like a Subquery
An IN List Ignores Duplicates
The Subquery
How a Basic Subquery Works
The Final Answer Set from the Subquery
Quiz- Answer the Difficult Question
Answer to Quiz- Answer the Difficult Question
Should you use a Subquery or a Join?
Quiz- Write the Subquery
Answer to Quiz- Write the Subquery
Quiz- Write the More Difficult Subquery
Answer to Quiz- Write the More Difficult Subquery
Quiz- Write the Subquery with an Aggregate
Answer to Quiz- Write the Subquery with an Aggregate
Quiz – Write the Triple Subquery
Answer to Quiz – Write the Triple Subquery

CHARACTER_LENGTH AND OCTET_LENGTH

The TRIM Command trims both Leading and Trailing Spaces

Trim and Trailing is Case Sensitive

Trim and Trailing works if Case right

The SUBSTRING Command

How SUBSTRING Works with NO ENDING POSITION

Using SUBSTRING to move Backwards

How SUBSTRING Works with a Starting Position of -1

How SUBSTRING Works with an Ending Position of 0

An Example using SUBSTRING, TRIM and CHAR Together

SUBSTRING and SUBSTR are equal, but use different syntax

The POSITION Command finds a Letters Position

Concatenation

The Basics of CAST (Convert and Store)

Some Great CAST (Convert and Store) Examples

Some Great CAST (Convert and Store) Examples

Combining Searched Case and Valued Case

A Trick for getting a Horizontal Case

Nested Case

Put a CASE in the ORDER BY