ProTech protechtraining.com *"Charting the Course ... ... to Your Success!"*

Aster Data SQL and MapReduce

Course Summary

Description

In this course, students will learn the Aster Data SQL and MapReduce starting at the most basic level and going to the most advanced level with many examples.

Objectives
At the end of this course, students will be able to:

- Have a deeper knowledge and understanding of the Aster Data Architecture, SQL and MapReduce functions and how to write them.

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 anyone who has a desire to learn the Aster Data SQL and MapReduce from beginners to an advanced audience. This course is completely customizable by the client.

Prerequisites

There are no prerequisites for this course.

Duration

Two to three days

ProTech protechtraining.com *"Charting the Course ... ... to Your Success!"*

Aster Data SQL and MapReduce

Course Outline

I. The Aster Data Architecture

- A. What is Parallel Processing?
- B. Aster Data is a Parallel Processing System
- C. Each vworker holds a Portion of Every Table
- D. The Rows of a Table are Spread Across All vworkers
- E. Aster Tables are defined as Fact or Dimension when Created
- F. Fact Table
- G. A More Detailed Look at the Fact Table Distribution
- H. Dimension Table are Replicated
- I. A Dimension Table is often Replicated across vworkers
- J. Aster Data has Fact and Dimension Tables
- K. Aster Tables are defined as Fact or Dimension when Created
- L. Fact and Dimension Tables can be Hashed by the same Key
- M. Distribution Key Rules
- N. Aster Data Uses a Hash Formula
- O. The Hash Map Determines which vworker will own the Row
- P. The Hash Formula, Hash Map and vworker
- Q. Placing rows on the vworker
- R. Placing rows on the vworker Continued
- S. A Review of the Hashing Process
- T. Like Data Hashes to the Same vworker
- U. Distribution Key Data Types
- V. Run ANALYZE to COLLECT STATISTICS on a Table
- W. Some Examples of ANALYZE
- X. What Columns to Analyze

II. Four Options for Aster Data Table Design

- A. There are Four Options to Aster Table Design
- B. Straight up Distribute by Hash
- C. Straight up Distribute by Hash - Problems
- D. Straight up Distribute by Replication
- E. Partition the Table with Logical Partitioning
- F. This Partitioned Table Sorts Rows by Month of Order_Date
- G. An All vworkers Retrieve By Way of a Single Partition
- H. You can Partition a Table by Range or by List
- I. A Partitioned By List Example with Three Tactical Queries
- J. Aster Data Multi-Level Partitioning
- K. Aster Allows for Multi-Level Partitioning
- L. SQL Commands for Logical Partitioning as One Table
- M. What Partitions are on my Table?
- N. What does a Columnar Table look like?
- O. A Comparison of Data for Normal Vs. Columnar
- P. A Columnar Table is best for Queries with Few Columns
- Q. When to use a Columnar Table

III. How Joins Work Inside the Aster Engine

- A. The Joining of Two Tables
- B. Aster Moves Joining Rows to the Same vworker
- C. Because of the Join Rule – Dimension Table are Replicated

Aster Data SQL and MapReduce

Course Outline (cont'd)

Aster Data SQL and MapReduce

Course Outline (cont'd)

N. GiST Indexes
O. Five Operational Tips for Efficient Indexing
P. REINDEX
Q. createCompressedIndexOnCompressedTableByDefault Flag

VII. SQL-MapReduce

A. MapReduce History
B. What is MapReduce?
C. What is SQL-MapReduce?
D. SQL-MapReduce Input
E. SQL-MapReduce Output
F. Subtle SQL-MapReduce Processing
G. Aster Data Provides an Analytic Foundation
H. Path Analysis
I. Text Analysis
J. Statistical Analysis
K. Segmentation (Data Mining)
L. Graph Analysis
M. Transformation of Data
N. Sessionize
O. Tokenize
P. SQL-MapReduce Function… nPath
Q. nPath SELECT Clause
R. nPath ON Clause
S. nPath PARTITION BY Expression
T. nPath DIMENSION Expression
U. nPath ORDER BY Expression
V. nPath MODE Clause has Overlapping or NonOverlapping
W. nPath PATTERN Clause
X. Pattern Operators
Y. Pattern Operators Order of Precedence
Z. Matching Patterns Which Repeat
AA. nPath SYMBOLS Clause
BB. nPath RESULTS Clause
CC. Adding an Aggregate to nPath Results
DD. Adding an Aggregate to nPath Results (Continued)
EE. SQL-MapReduce Examples - Use Regular SQL
FF. SQL-MapReduce Examples - Create Objects
GG. SQL-MapReduce Examples - Subquery
HH. SQL-MapReduce Examples - Query as Input
II. SQL-MapReduce Examples - Nesting Functions
JJ. SQL-MapReduce Examples - Functions in Derived Tables
KK. SQL-MapReduce Examples - SMAVG
LL. SQL-MapReduce Examples - Pack Function
MM. SQL-MapReduce Examples - Pack Function (Continued)
NN. SQL-MapReduce Examples - Pivot Columns
OO. Workshop: Create This Table
PP. Login to your GNOME Terminal
QQ. Login to your Linux
RR. Using the GNOME Terminal Unzip the bank_web_data.zip
SS. Use the Function ncluster_loader to Load the Bank Data
TT. Run this nPath Map Reduce Function on your Table
UU. nPath in Action
VV. Operators at their Simplest
WW. Pattern
XX. Accumulate
YY. Accumulate With All Pages
ZZ. Accumulate – nPath with a WHERE Clause
AAA. SQL-MapReduce Examples - Path Generator
BBB. SQL-MapReduce Examples - Path Generator (Continued)

Aster Data SQL and MapReduce

Course Outline (cont'd)

CCC. SQL-MapReduce Examples - Path Generator (Continued)
DDD. SQL-MapReduce Examples - Path Generator (Continued)
EEE. SQL-MapReduce Examples - Path Generator (Continued)
FFF. SQL-MapReduce Examples - Linear Regression
GGG. SQL-MapReduce Examples - Linear Regression (Continued)
HHH. SQL-MapReduce Examples - Linear Regression (Continued)
III. SQL-MapReduce Examples - Naive Bayes
JJJ. SQL-MapReduce Examples - Naive Bayes (Continued)
KKK. SQL-MapReduce Examples - Naive Bayes (Continued)
LLL. SQL-MapReduce Examples - Naive Bayes (Continued)
MMM. SQL-MapReduce Examples - Naive Bayes (Continued)
NNN. SQL-MapReduce Examples - Naive Bayes (Continued)
OOO. SQL-MapReduce Examples - Naive Bayes (Continued)
PPP. SQL-MapReduce Examples - Naive Bayes (Continued)
QQQ. Join Aster, Teradata and Hadoop Tables; feed into MapReduce
RRR. Run Both of these Examples Together and Compare
SSS. Run this nPath Map Reduce Function
TTT. nPath in Action
UUU. Another nPath Example
VVV. Finding Out What Functions You Have Installed
WWW. Workshop 1 – Fill in the x's
XXX. Answer Workshop 1 - Fill in the x's
YYY. Workshop 2 – Fill in the x's
ZZZ. Answer Workshop 2 – Fill in the x's
AAAA. Answer Workshop 2 – You Could Have Used a GROUP BY
BBBB. Workshop 3 – Add to the Query
CCCC. Workshop 3 – Answer to Add to the Query
DDDD. Workshop 4 – Fill in the x's
EEEE. Answer to Workshop 4 – Fill in the x's
FFFF. Workshop 5 – Find that Customer
GGGG. Answer to Workshop 5 – Find that Customer
HHHH. Workshop 6 – Change the MapReduce Function
IIII. Answer to Workshop 6 – Change the MapReduce Function
JJJJ. Workshop 7 – Build the MapReduce Function
KKKK. Answer to Workshop 7 – Build the MapReduce Function
LLLL. Best Answer to Workshop 7 – Build the MapReduce Function
MMMM. Workshop 8 – Build the Accumulate in the Result
NNNN. Answer to Workshop 8 – Build the Accumulate in the Result
OOOO. SQL-MapReduce Examples - Linear Regression (Continued)
PPPP. Workshop 9 – Build the Subquery
QQQQ. Answer to Workshop 9 – Build the Subquery
RRRR. Workshop 10 – Do Your First Join
SSSS. Answer to Workshop 10 – Do Your First Join
TTTT. Answer to Workshop 10 – Do the Join Using a New Syntax

Aster Data SQL and MapReduce

Course Outline (cont'd)

UUUU. Workshop 11 – Super Join the Tables
VVVV. Answer to Workshop 11 – Super Join the Tables
WWWW. Answer to Workshop 11 – Super Join the Tables
XXXX. Workshop 12 – Sessionize the Data
YYYY. Answer to Workshop 12 – Sessionize the Data
ZZZZ. Workshop 13 – What is this Query Doing?
AAAAA. Answer to Workshop 13 – What is this Query Doing?
BBBBB. Workshop 14 – Using ilike
CCCCC. Answer to Workshop 14 – Using ilike
DDDDD. Answer to Workshop 14 – Using ilike
EEEEE. Workshop 15 – What are the First Two Pages Visited?
FFFFF. Workshop 15 – What are the First Two Pages Visited?
GGGGG. Workshop 16 – Advanced - First Two Pages Visited?
HHHHH. Answer to Workshop 16 Advanced - First Two Pages Visited?
IIIII. Workshop 17 – Can You Clean Up the Results?
JJJJJ. Answer to Workshop 17 – Can You Clean Up the Results?
KKKKK. Answer to Workshop 17 – Format the Date
LLLLL. Workshop 18 – Build a Churn Table
MMMMM. Workshop 18 – Run the Query Before Building to Test
NNNNN. Workshop 18 – A Better Example
OOOOO. Answer to Workshop 18 – Build a Basic Churn Table
PPPPP. Workshop 18 – Create the Churn Table with a Better Example
QQQQQ. Multi-Case
RRRRR. The Multi-Case Function
SSSSS. The Multi-Case Function in Nexus
TTTTT. The Multi-Case Function Mixing and Matching
UUUUU. The Multi-Case Function Mixing and Matching
VVVVV. SQL-MapReduce Examples - cFilter
WWWWW. SQL-MapReduce Examples - cFilter (Continued)
XXXXX. SQL-MapReduce Examples - Linear Regression (Continued)
YYYYY. SQL-MapReduce Examples - cFilter (Continued)
ZZZZZ. SQL-MapReduce Examples - Linear Regression (Continued)
AAAAAA. SQL-MapReduce Examples - cFilter (Continued)
BBBBBB. SQL-MapReduce Examples - cFilter (Continued)
CCCCCC. SQL-MapReduce Examples - cFilter (Continued)
DDDDDD. SQL-MapReduce Examples - cFilter (Continued)
EEEEEE. SQL-MapReduce Examples - cFilter (Continued)
FFFFFF. SQL-MapReduce Examples - cFilter (Continued)
GGGGGG. SQL-MapReduce Examples - cFilter (Continued)
HHHHHH. CFILTER in Action with Bank_Web_Clicks
IIIIII. CFILTER in Action
JJJJJJ. CFILTER using Nexus
KKKKKK. nPath Error

Aster Data SQL and MapReduce

Course Outline (cont'd)

VIII. Time and Date

- A. Date, Time, and Timestamp Keywords
- B. Add or Subtract Days from a date
- C. The to_char command
- D. A Summary of Math Operations on Dates
- E. Using a Math Operation to find your Age in Years
- F. Find What Day of the week you were Born
- G. Date Related Functions
- H. The EXTRACT Command
- I. EXTRACT from DATES and TIME
- J. EXTRACT with DATE and TIME Literals
- K. EXTRACT of the Month on Aggregate Queries
- L. A Side Title example with Reserved Words as an Alias
- M. Implied Extract of Day, Month and Year
- N. DATE_PART Function
- O. DATE_TRUNC Function
- P. DATE_TRUNC Function using TIME
- Q. Aster NOW() Function

IX. Aster Windows Functions

- A. Cumulative Sum
- B. Cumulative Sum - Major and Minor Sort Key(s)
- C. The ANSI CSUM – Getting a Sequential Number
- D. The ANSI OLAP – Reset with a PARTITION BY Statement
- E. PARTITION BY only Resets a Single OLAP not ALL of them
- F. ANSI Moving Sum is Current Row and Preceding n Rows
- G. How ANSI Moving SUM Handles the Sort
- H. Moving SUM every 3-rows vs. a Continuous Sum
- I. Moving Average
- J. Partition By Resets an ANSI OLAP
- K. Moving Average Using BETWEEN
- L. Moving Difference using ANSI Syntax
- M. Moving Difference using ANSI Syntax with Partition By
- N. RANK Defaults to Ascending Order
- O. Getting RANK to Sort in DESC Order
- P. You can use Window Functions in Expressions
- Q. RANK() OVER and PARTITION BY
- R. DENSE_RANK() OVER
- S. PERCENT_RANK() OVER
- T. PERCENT_RANK() OVER with 14 rows in Calculation
- U. PERCENT_RANK() OVER with 21 rows in Calculation
- V. RANK With ORDER BY SUM()
- W. COUNT OVER for a Sequential Number
- X. The MAX OVER Command
- Y. MAX OVER with PARTITION BY Reset
- Z. The MIN OVER Command
- AA. The Row_Number Command
- BB. NTILE
- CC. NTILE Using a Value of 10
- DD. NTILE With a Partition
- EE. CUME_DIST
- FF. CUME_DIST With a Partition
- GG. LEAD
- HH. LEAD With Partitioning
- II. LAG
- JJ. LAG with Partitioning
- KK. FIRST_VALUE

Aster Data SQL and MapReduce

Course Outline (cont'd)

LL. FIRST_VALUE After Sorting by the Highest Value
MM. FIRST_VALUE with Partitioning
NN. LAST_VALUE
OO. NTH_VALUE
PP. NTH_VALUE With Partition
QQ. SUM(SUM(n))

X. The Fundamental SQL Commands That Work on Aster

A. BETWEEN is Inclusive
B. BETWEEN Works for Character Data
C. LIKE uses Wildcards Percent '%' and Underscore '_'
D. LIKE command Underscore is Wildcard for one Character
E. GROUP BY Vs. DISTINCT – Good Advice
F. The Five Aggregates of Aster Data
G. GROUP BY when Aggregates and Normal Columns Mix
H. GROUP BY Delivers one row per Group
I. GROUP BY Dept_No or GROUP BY 1 the same thing
J. Limiting Rows and Improving Performance with WHERE
K. WHERE Clause in Aggregation limits unneeded Calculations
L. Keyword HAVING tests Aggregates after they are Totaled
M. Keyword HAVING is like an Extra WHERE Clause for Totals
N. Getting the Average Values per Column
O. Getting the Average Values per Column
P. Average Values per Column for All Columns in a Table
Q. A two-table join using Non-ANSI Syntax
R. A two-table join using Non-ANSI Syntax with Table Alias
S. Aliases and Fully Qualifying Columns
T. A two-table join using ANSI Syntax
U. Both Queries have the same Results and Performance
V. LEFT OUTER JOIN
W. LEFT OUTER JOIN Brings Back All Rows in the Left Table
X. RIGHT OUTER JOIN
Y. RIGHT OUTER JOIN Brings Back All Rows in the RIGHT Table
Z. FULL OUTER JOIN
AA. FULL OUTER JOIN Brings Back All Rows in All Tables
BB. Which Tables are the Left and which are the Right?
CC. Answer - Which Tables are the Left and which are the Right?
DD. INNER JOIN with Additional AND Clause
EE. ANSI INNER JOIN with Additional AND Clause
FF. ANSI INNER JOIN with Additional WHERE Clause
GG. OUTER JOIN with Additional WHERE Clause
HH. OUTER JOIN with Additional AND Clause
II. Results from OUTER JOIN with Additional AND Clause
JJ. The DREADED Product Join
KK. Result Set of the DREADED Product Join
LL. The Horrifying Cartesian Product Join
MM. The ANSI Cartesian Join will ERROR

Aster Data SQL and MapReduce

Course Outline (cont'd)

NN. How would you Join these two tables?
OO. How would you Join these two tables? You Can't Yet!
PP. An Associative Table is a Bridge that Joins Two Tables
QQ. The 5-Table Join – Logical Insurance Model
RR. The Nexus Query Chameleon Writes the SQL for Users.
SS. An IN List is much like a Subquery
TT. An IN List Never has Duplicates – Just like a Subquery
UU. An IN List Ignores Duplicates
VV. The Subquery
WW. How a Basic Subquery Works
XX. The Final Answer Set from the Subquery
YY. Should you use a Subquery or a Join?
ZZ. CHARACTER_LENGTH AND OCTET_LENGTH
AAA. The TRIM Command trims both Leading and Trailing Spaces
BBB. Trim and Trailing is Case Sensitive
CCC. Trim and Trailing works if Case right
DDD. The SUBSTRING Command
EEE. How SUBSTRING Works with NO ENDING POSITION
FFF. Using SUBSTRING to move Backwards
GGG. How SUBSTRING Works with a Starting Position of -1
HHH. How SUBSTRING Works with an Ending Position of 0
III. An Example using SUBSTRING, TRIM and CHAR Together
JJJ. SUBSTRING and SUBSTR are equal, but use different syntax
KKK. The POSITION Command finds a Letters Position
LLL. Concatenation
MMM. The Basics of CAST (Convert and Store)
NNN. Some Great CAST (Convert and Store) Examples
OOO. Some Great CAST (Convert and Store) Examples
PPP. Combining Searched Case and Valued Case
QQQ. A Trick for getting a Horizontal Case
RRR. Nested Case
SSS. Put a CASE in the ORDER BY