



Building Big Data Applications with SQL-MapReduce and Aster Developer Express

Title of Presenter | Date

Peter Pawlowski, Software Engineer

Big Data: The Trend

Many companies have an ever-increasing amount of data.

Big Data: The Challenge

How do we enable all these companies to build useful applications using this data?

“Big Data Applications”

Not Everyone is Google



Our Goal: Big Data Applications Made Easy

1. We enable storing terabytes to petabytes of data in our scalable, parallel database.
2. We provide a framework for running rich, interesting analytics on this data.

Our Goal: Big Data Applications Made Easy

1. We enable storing terabytes to petabytes of data in our scalable, parallel database.
2. We provide a framework for running rich, interesting, and fast analytics on this data.

Reports, Advanced Analytics, Applications
(SQL, SQL-MapReduce, ODBC, JDBC)

Queries/Results

Aster Data nCluster

Queen

Queries

Workers (Parallel
Processes)

Data

Data

Loading/Export

Backup

We started with a SQL interface

- Great fit for many data problems
- Many people know it
- Plugs into many applications & tools

... but SQL ...

- Is a bad fit for many kinds of analytic applications
- Can be hard to write & maintain
- Can be hard to execute efficiently

Enter SQL-MapReduce

- SQL-MapReduce is our framework for pushing analytic code into the database
- Enables much richer analytics to happen on the data in the database

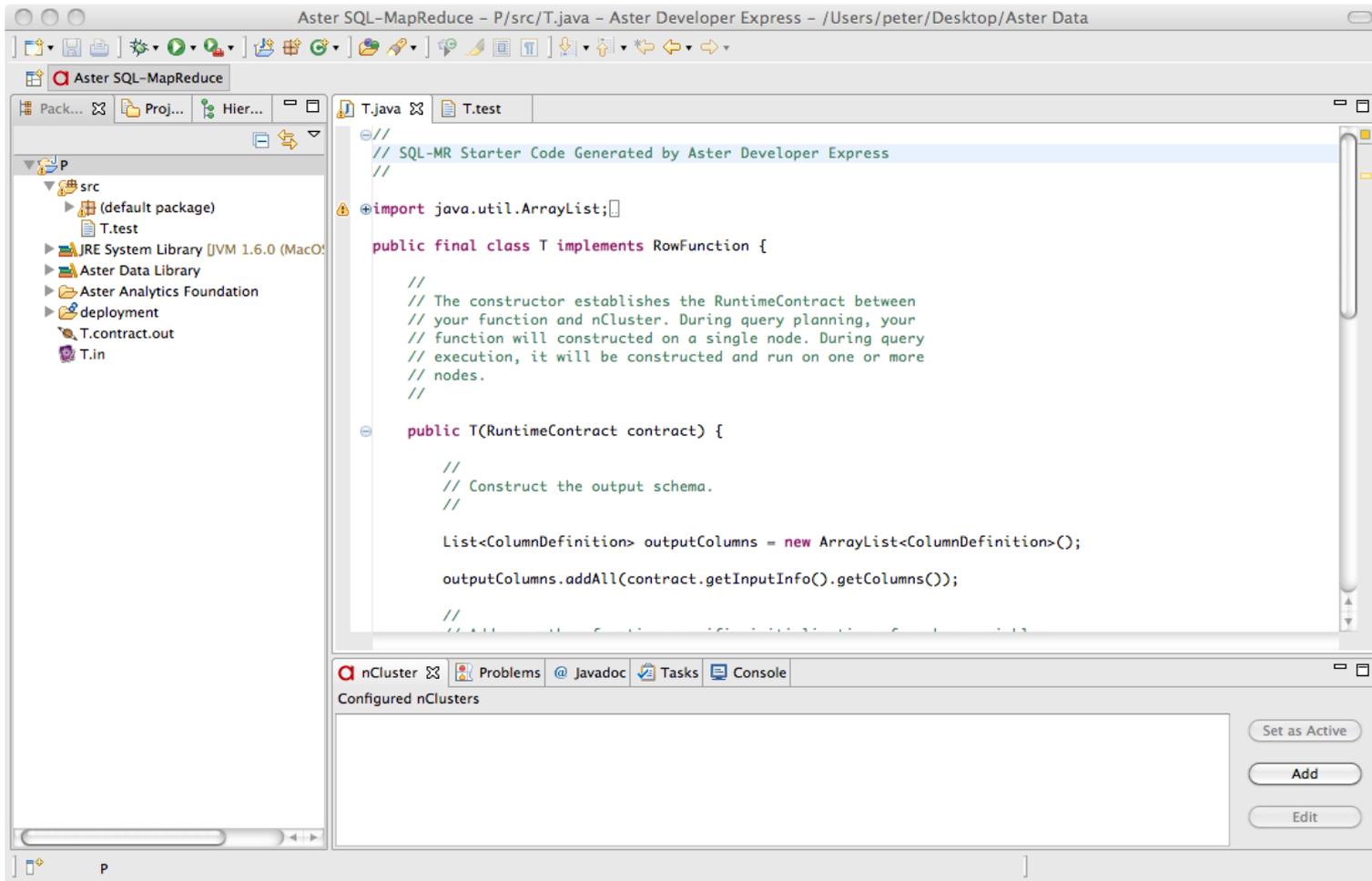
Two Steps to SQL-MapReduce

- Write a SQL-MapReduce function in Java*
- Invoke it from SQL

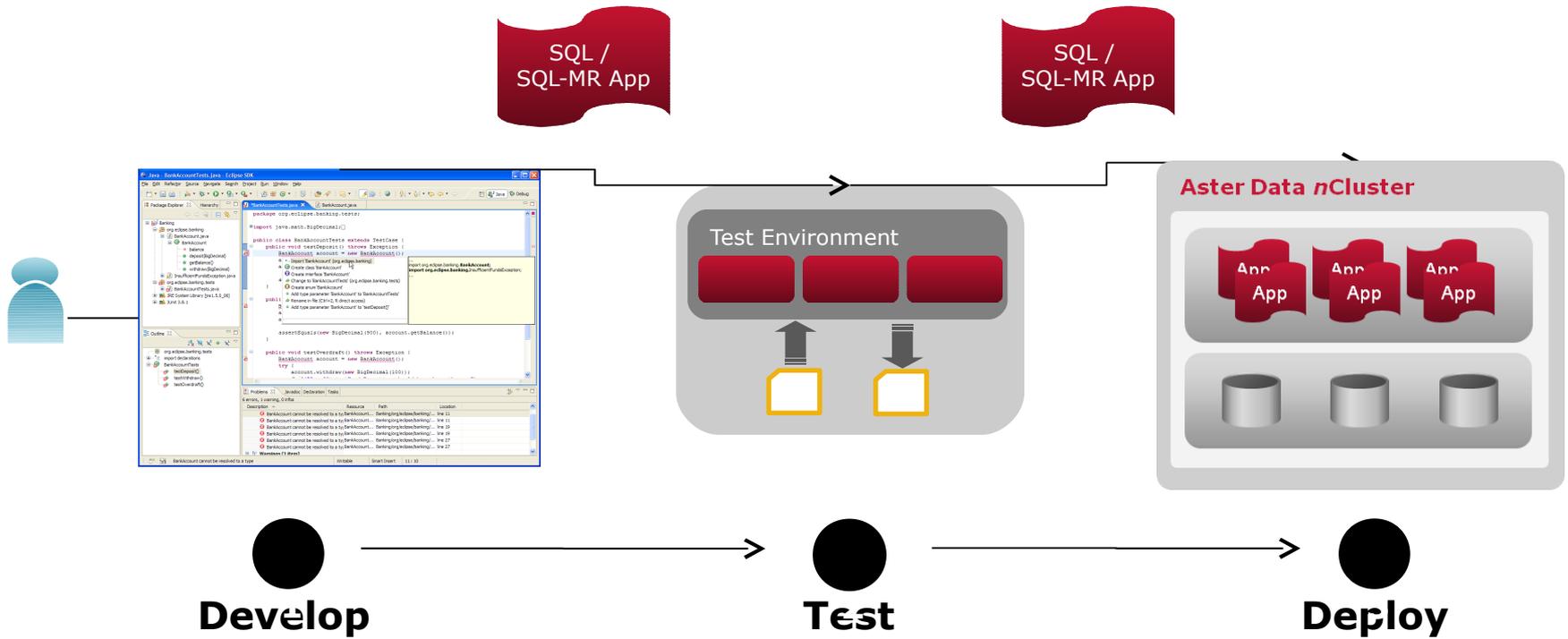
* Other languages supported, too. See our website.

aster data

Developer Express



Aster Data Developer Express



Step 1: Develop

New Aster Data Function

SQL-MapReduce Function
Create a new Aster SQL-MapReduce function.

Source folder:

Package: (default)

Name:

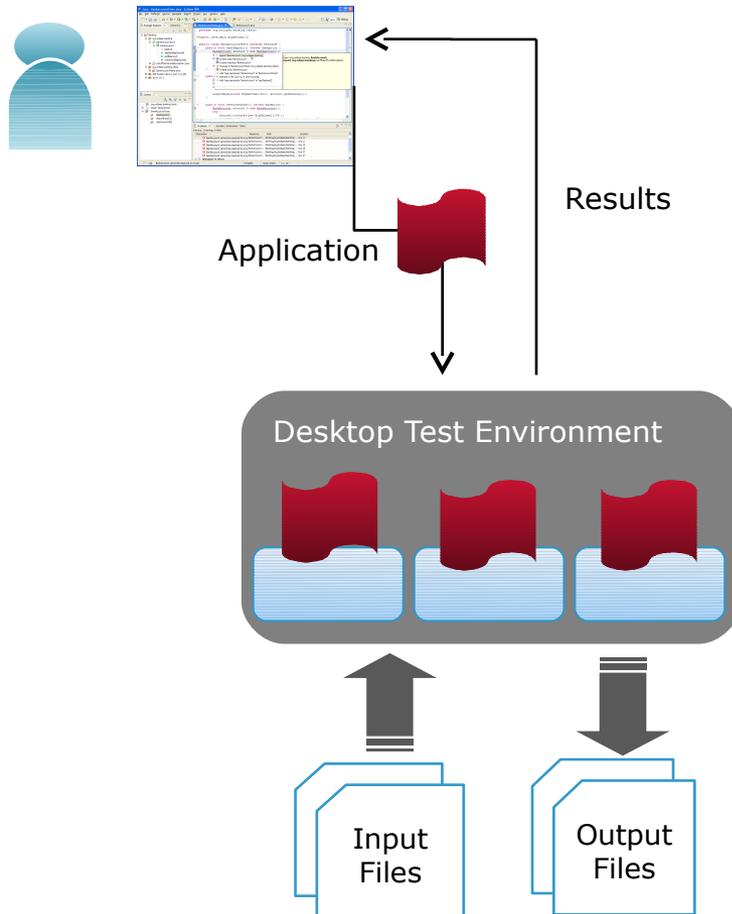
Which interfaces should this SQL-MapReduce function implement?

- RowFunction (like a "map")
- PartitionFunction (like a "reduce")

Advanced features

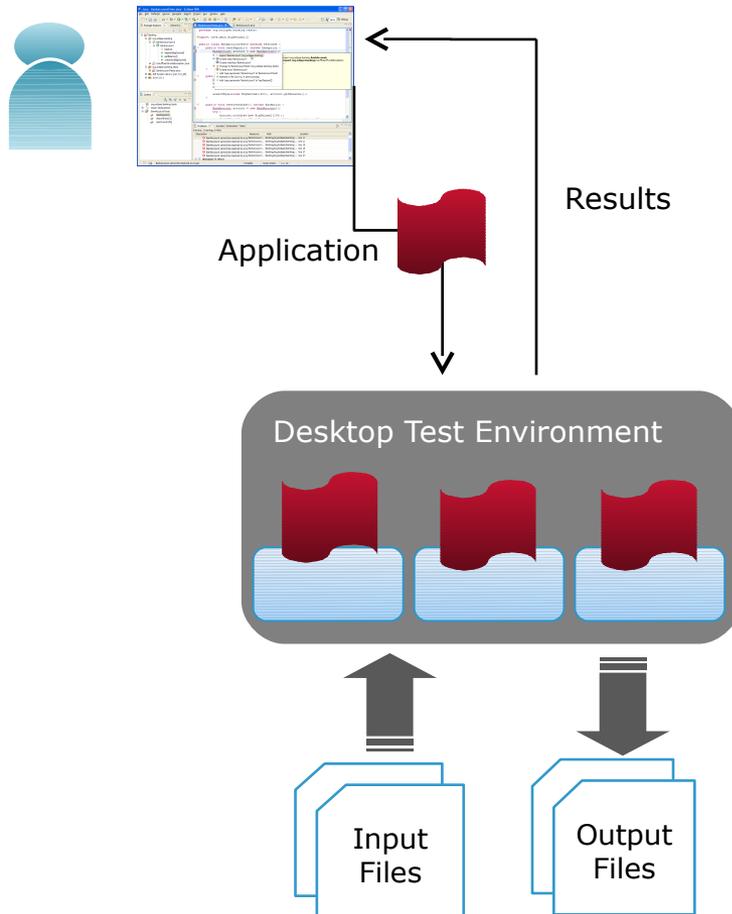
- Implement Drainable Interface

Step 2: Test Locally



- Run tests on your own machine without requiring an *nCluster* instance.
- Configure with JSON.

Step 2: Test Locally



- Run tests on your own machine without requiring an *nCluster* instance.
- Configure with JSON.

Step 2: Test Locally

New Aster SQL-MapReduce Test Configuration

Input
Define the input to be used in the test

Use File:

File Type: comma-separated values (CSV)
 tab-separated values (TSV)

Schema:

Name	Type

Input File:

Step 3: Deploy

Add a new nCluster connection

 The connection information was successfully validated.



Name:

Specify the hostname and database to connect:

Hostname:

Database:

Specify the user authentication information:

Username:

Password:

Re-type Password:

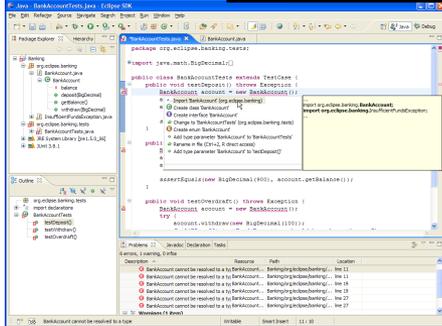


Check connection

Cancel

OK

Step 3: Deploy



SQL /
SQL-MR App

Aster Data *n*Cluster

App
App

App
App

App
App



- Push application into Aster database directly from within IDE
- Initiate and validate application with actual data stored inside *n*Cluster

Aster Data Developer Express is freely available

Download at
www.asterdata.com

Questions?
peter@asterdata.com