Processing Large Data sets for the National Broadband Map with FME

FME 2011 WORLD TOUR

Reed Whittington
Principal

SAFE SOFTWARE

National Broadband Map
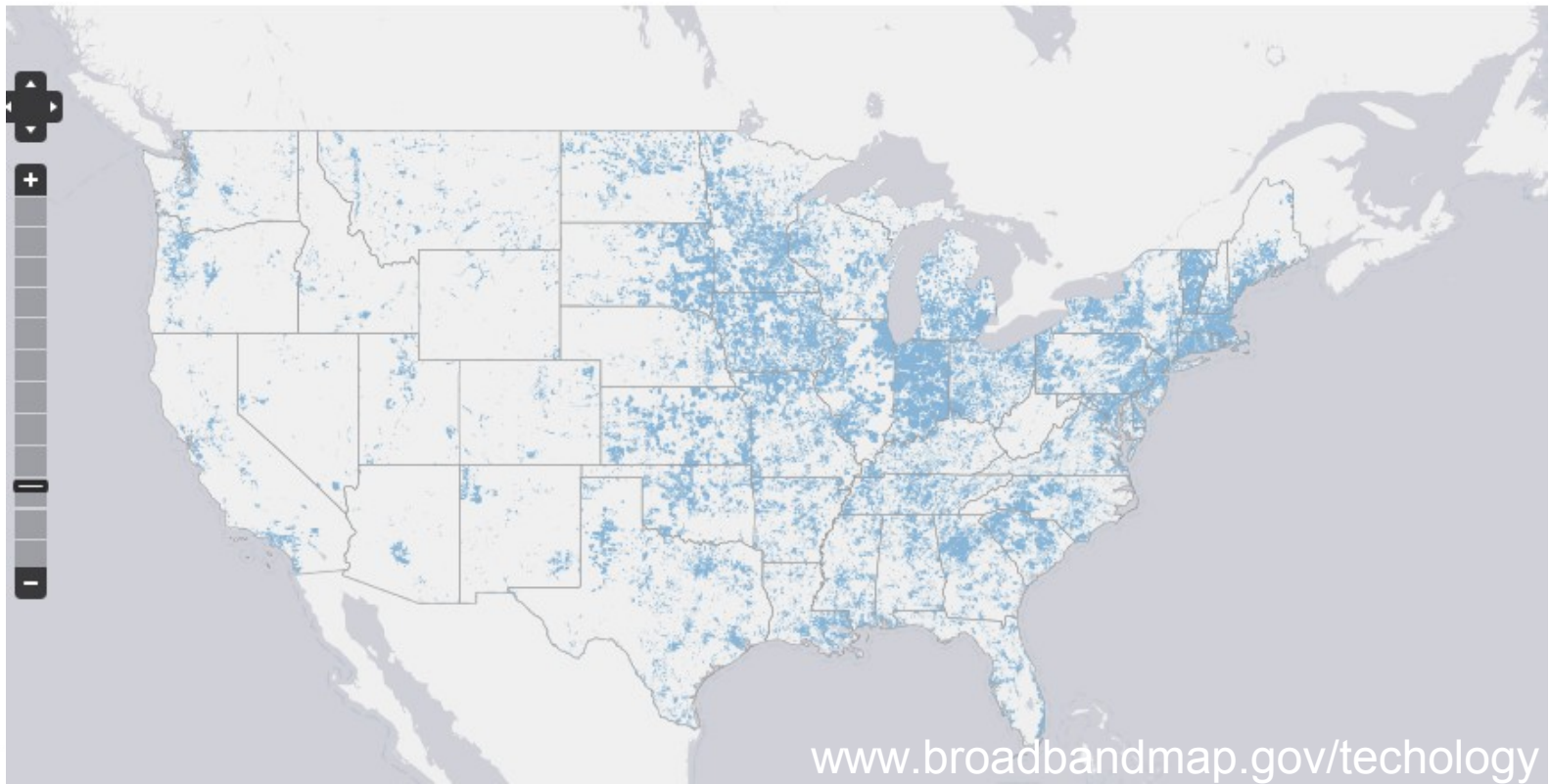
- Massive amounts of data for all states all providers wireline and wireless

- Coverage statistics needed for each census block for all counties and all states

- Process Must be Automated so processing can be repeated for data updates

- Interoperability with Oracle and ESRI GDB

- Short 1 month project cycle

    Start Mid Dec 2010 working on site

GISINT

Interoperability Solution

- FME was chosen for factors including:

  Interoperability with Oracle and ESRI

  Scalable Automated processing

  Fast overlay and dissolve operations

  End customer suggested FME

- Professional Services provided by GIS INT

  10+ Years FME Integration Experience

  Responsiveness

  Recommended by Safe

- Modular Design

    One-Function Workflow with SQL query published so that workflow can be called to iterate across states or states and counties

    Oracle Spatial index was used to read just the polygons required

    - Coverage that cross Census Blocks returned from SQL SELECT

GISINT

Where Clauses allow processing distinct parts of data set.

Translation Parameters

Destination Oracle Non-spatial Service:	
Source Oracle Spatial Object Service:	
Log File:	
WHERE Clause Blocks	STATEFP00 = 33 AND COUNTYFIPS = 017
WHERE Clause Wireless	STATE_CODE = 'NH' AND objectid=30001633
Username:	
Password:	••••••

Help   OK   Cancel

GISINT

- The actual iteration is done in a separate FME workspace that reads the parts to process

- We realized halfway through processing a need to start processing over but not reprocess any parts that had succeed.

  - Manual tracking initially

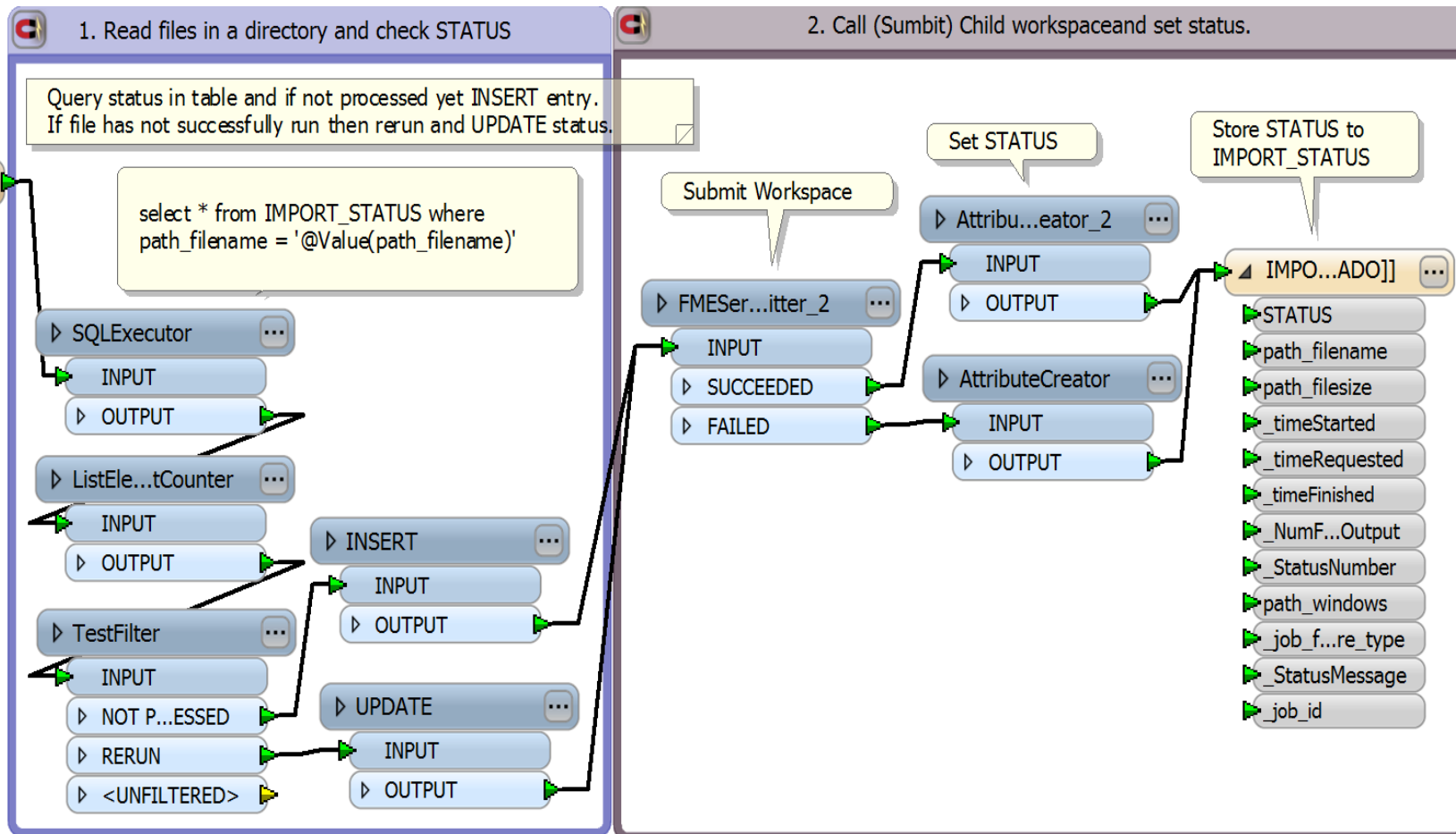- Calling a workspace from parent workspace

  WorkspaceRunner

  - Used without FME Server

  - A manual process would be used to distribute processing over multiple CPUs with a command line call passing in State to process.

- The return status of each iteration (workspace call) can be stored and used later to reprocess fails.

- FME Server can be utilized by replacing

  - WorkspaceRunner

    - Used to run another workspace on the same Desktop box with

  - FMEServerWorkspaceRunner

    - Used to queue FME Server hosted workspace or

  - FMEServerJobSubmitter

    - Also used to queue FME Server job

- Workflow Submitter Iterator Improvements

  - Optimization – faster database

  - Better Logging – Retrieve child log when failed with FMEServerLogFileRetriever

  - Metrics - estimate percent completed

  - Job Control – Allow iterator to be killed or halted and started up on another engine

  - More Parameter driven

- Backfit/Upgrade to replace existing installations

Thank You!

- Questions?

- For more information:

  - Reed Whittington

  - GIS Interop Solutions

  - reedwhit@gisint.com

  - www.gisint.com

  - www.linkedin.com/in/reedwhit