Storage and Retrieval of Spatially-Qualified Data from NASA’s EOSDIS Data Pool

K. Carr, R. Meyer, C. Duma, K. Bryant, R. Hartranft, R. Bergman, S. Fox
Raytheon Information Solutions
Corresponding email: kcarr@eos.hitc.com
Data Pool Overview

- Developed under Raytheon contract with NASA to improve and promote end user access to EOSDIS data
- Data Pool is primary Data Accessibility enhancement
  - Large online cache of ECS data
  - User access via web and ftp
- Goals for implementation
  - Rapid-response, easy-to-use web drill down interface
  - Real time updates of database inventory data
  - Ability to scale insert and drill down performance to expected volumes
Data Pool Web Drill Down

• Six options (dimensions) for search refinement
  • Data product
  • Spatial coverage
  • Temporal/date coverage
  • Time of Day
  • Day/Night/Both
  • Data Quality

• Data Warehouse design chosen, to facilitate rapid data retrieval
  • Each dimension covered by static grid
  • Spatial, Temporal, and Time of Day grids arranged hierarchically
  • Intersecting grid elements calculated for each dimension at granule insert time and stored as granule “facts”
Spatial Drill Down

- Spatial search area specified by:
  - Rectangle (shown)
  - Polygon
  - Tile
- Spatial drill down by rectangle and polygon use Spatial Query Server (SQS)
- Concern about SQS performance led to spatial drill down by tile
Spatial Grid

• “Tile” drill down maps directly to spatial grid
• Lowest level grid is 1° x 1° tiles covering entire earth
• 6 “parent” overlay grids
  • 3° x 1°
  • 9° x 2°
  • 9° x 6°
  • 18° x 12°
  • Hemisphere
  • Whole globe
Spatial Data Retrieval

- At insert time, SQS is used to intersect granule with spatial grid tiles to determine spatial “facts”
- At retrieval time, Web Access map applet translates mouse pointer map coordinates to corresponding tile
- Tile selection results in direct query on spatial facts table
• Pruning algorithm eliminates storage of derivable spatial facts
• Two steps
  • Compute set of intersects between granule coverage and spatial grid
  • Eliminate children if all children of immediate parent are in intersection set
  • Otherwise, eliminate immediate parent and preserve children
• Average 100 - 150 spatial facts per granule, after pruning
 Spatial Facts - Lessons Learned

- Large numbers of spatial facts for:
  - Orbit-based granules (average of 3000 spatial facts per MISR granule)
  - Large polar granules (1°x1° tiles are “smaller” at the poles)
  - Near-global granules

- Operational impacts
  - Insert performance for orbit granules
  - Large spatial facts table defeats attempt to keep inventory database in cached memory

- Operational tuning
  - Spatial facts no longer populated for orbit data; instead, additional join performed during drill down using pre-computed swath to tile cross reference table
Next Steps

• Eliminate need for spatial grid by tuning SQS searches
  • Performance benchmark results are inconsistent
• Experiment with customized spatial grids
  • Align spatial grid with spacecraft orbit
  • Use hierarchical tessellation of the earth
  • Tune spatial grid to each data product
• Use pre-calculated “summary maps” for estimating spatial search results