

ORACLE®

S P A T I A L

May 2012
Oracle Spatial User Conference

A horizontal banner with a red background featuring a faint, stylized map pattern. The text "Oracle Spatial User Conference" is written in large, white, sans-serif font across the center of the banner.

Oracle Spatial User Conference

May 23, 2012
Ronald Reagan Building and International Trade Center
Washington, DC USA



Atri Kalluri

Assistant Division Chief

Spatial Data Systems & Database Management

Geography Division

US Census Bureau



Experiences with Exadata and Oracle Spatial



Program Agenda

- MAF/TIGER Introduction
- Oracle Spatial as a solution
- Processing Constraints
- Future needs
- Oracle Spatial on Exadata
- Benefits realized so far



MAF/TIGER Introduction

- Master Address File
 - An inventory of all known living quarters in the US
 - Contains
 - Mailing Address, if one exists
 - Descriptive Address, when no city-style address exists
 - Census Geographic Location
 - Source and History data
- Topologically Integrated Geographic Encoding and Referencing System
 - A digital map of US, Puerto Rico & Island Areas
 - Contains
 - Roads (Streets), Rail, Hydro
 - Geographic Entity Boundaries
 - Housing Unit Locations
 - Address Ranges and Zip codes
 - Key Geographic Locations



Oracle Spatial as a solution

- Oracle Spatial Topology Data Model
- Persistent topology
- Oracle TopoMap > MTTopoMap
- Business Rules
- Delayed Calculations
- Continuous updates (interactive & batch)
- Map Viewer
- Benchmarks & MAF Updates (DSF Refresh, Geocoding)

```
graph TD; A[MAF/TIGER Update Applications] --> B[Security Implementation - Access Roles, System Protections]
```





GATRES

Navigation

Geographic Acquis-Based Topological
Real-time Editing System (GATRES)

Production Version 3.1.1 - PRODTRAN



Data Source: PRODTRAN

☒ Stored Geometry

☐ Calculate Geometry

Quick Launch

Outline

Plot

Change Area

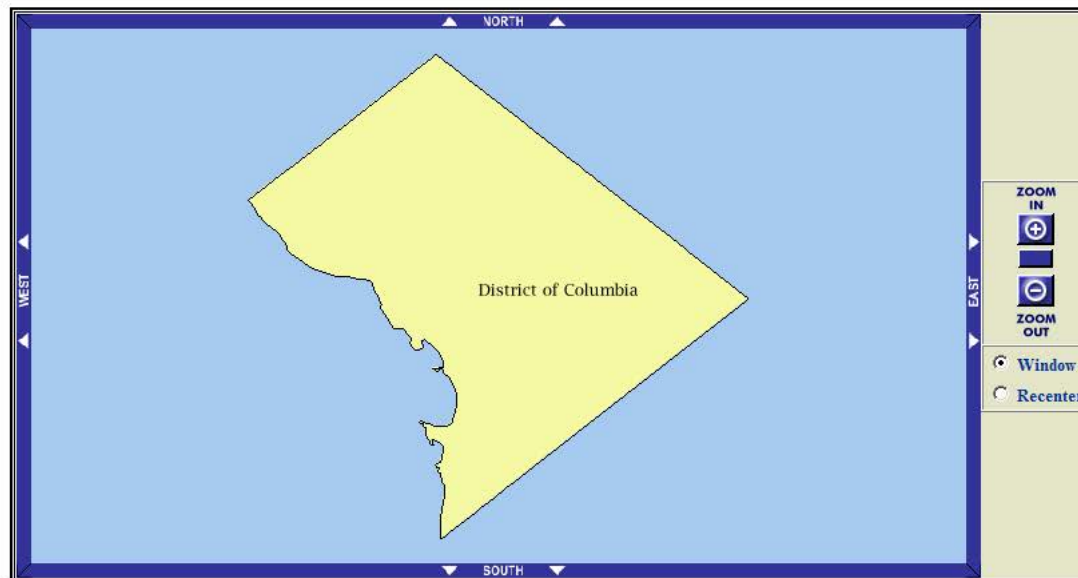
Geography
Type: County

ST/COU FIPS Code:

Vintage: 90 - Current

State: 11 - District of Columbia

County: 001 - District of Columbia

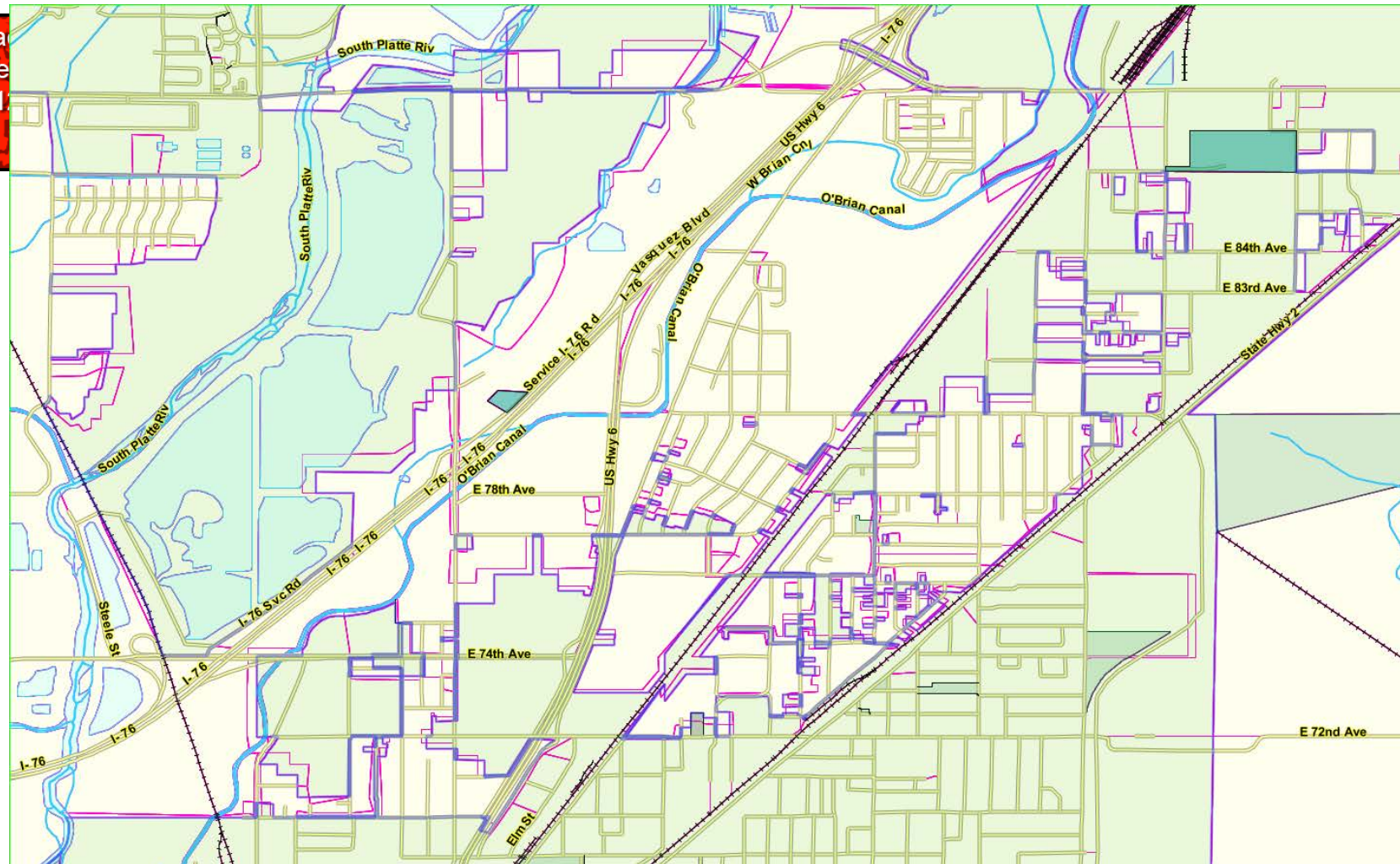


Oracle Spatial
User Conference
2012

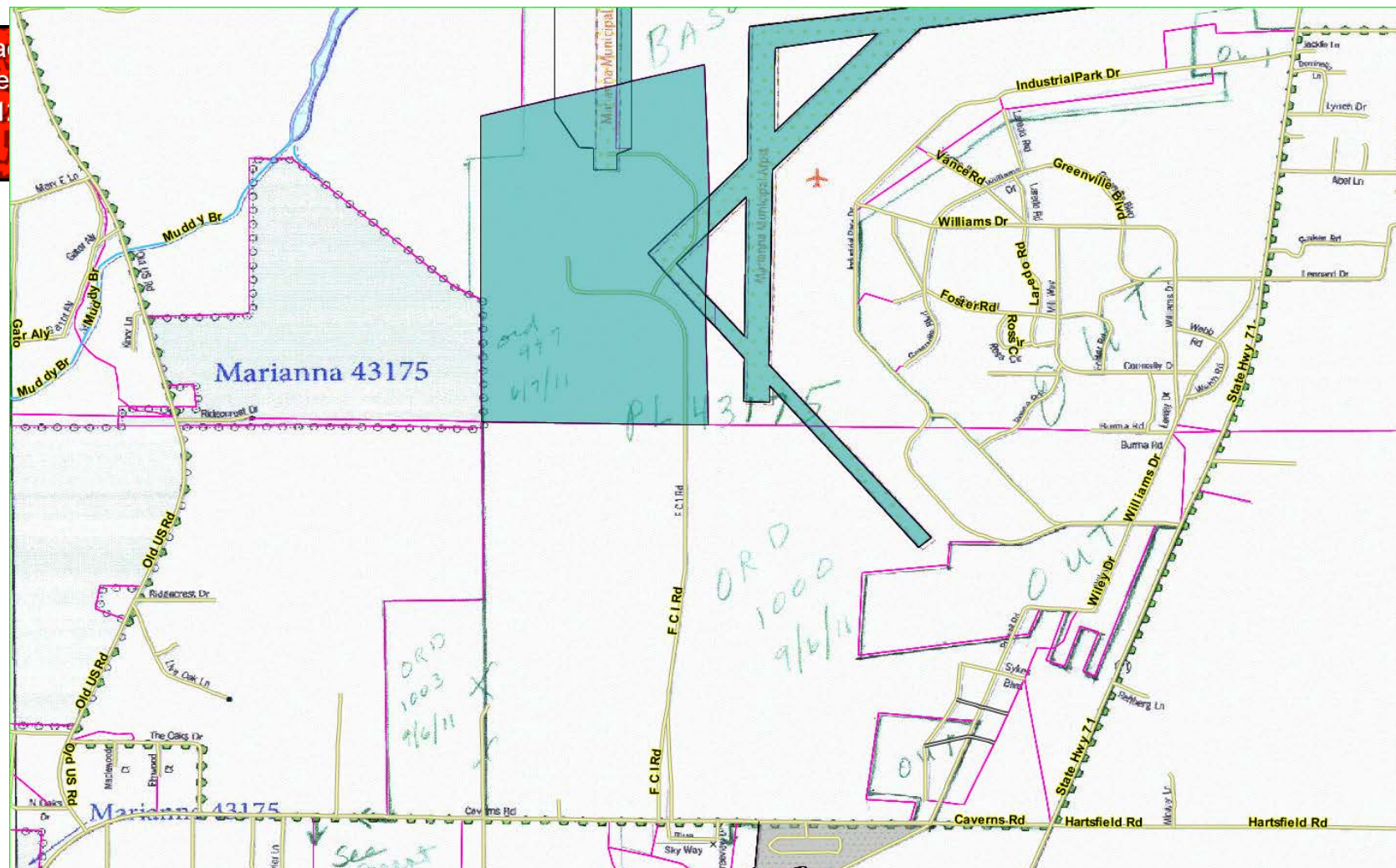
GATRES with Imagery



Oral
Use
201



Oral
Use
2011



TIGERweb (beta) - Windows Internet Explorer

http://tigerweb.geo.census.gov/tigerweb/default.htm

File Edit View Favorites Tools Help

Convert Select

TIGERweb (beta)

U.S. Census Bureau

People | Business | Geography | Newsroom | Subjects A to Z | Search@Census

United States Census 2010

TIGERweb (beta)
Geography Division

Pan Legend Help

200km
100mi
1:9,244,648

Privacy Policy | 2010 Census | Data Tools | Information Quality | Product Catalog | Contact Us | Home

Map Extent: [-11717898.35, 3789283.99] - [-9320833.15, 5034290.30]

Trusted sites 100%



Launch **TIGERweb Viewer** (beta)

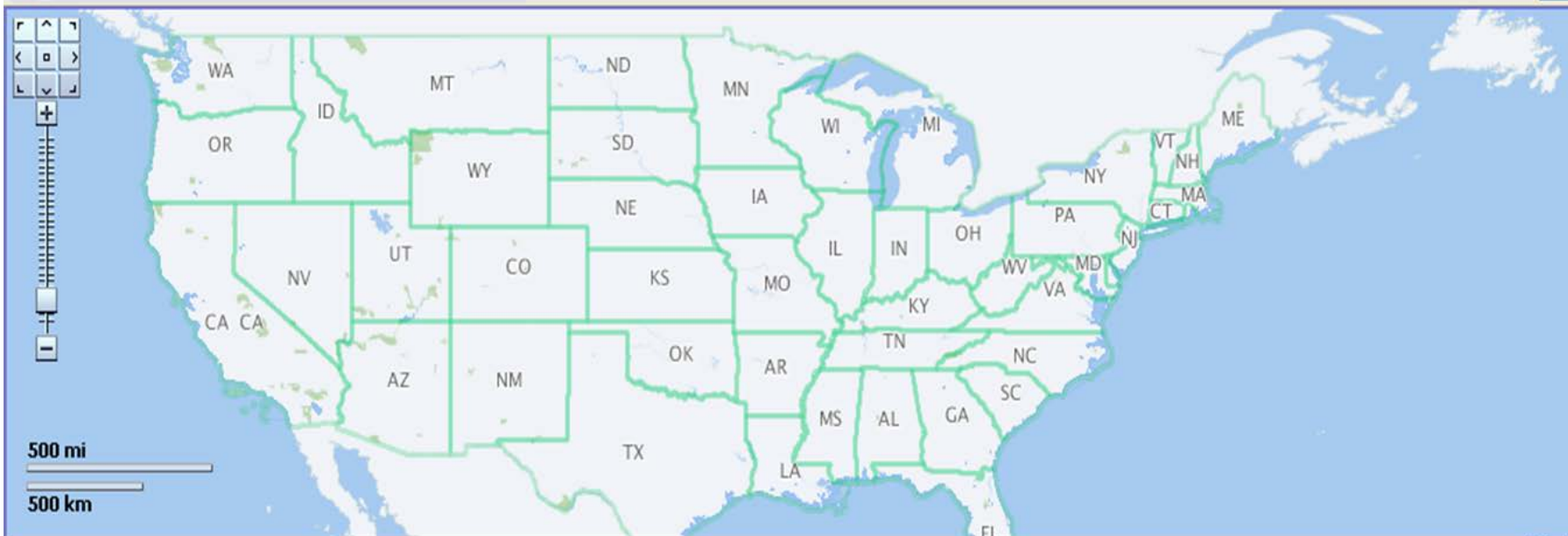
(Note: This link will open a new browser window.)

TIGERweb WMS

TIGER data on the web is released following the Open Geospatial Consortium, Inc. (OGC) Web Map Service (WMS) standard. Users that have a client that supports the WMS standard may access the TIGERweb service using the following URL:

`//tigerweb.geo.census.gov/ArcGIS/services/tigerWMS/Map Server/WMSServer`

Display ▼ Street Map ▼ Zoom To Report an Issue ▼



My Issues

Nearby Issues

Recent Issues

			Date Reported	Short Description	Issue Type
+		19001	11/16/2011	Name is wrong	Incorrect Name
+		16002	10/28/2011	Street Alignment issue	Incorrect Road Shape
+		16001	10/28/2011	This is a new issue	Incorrect Name



Batch Processing

- Multiple simultaneous batch runs
- Designated a branch for workflow control
- Spatial locking enables multiple updates
- Nightly runs for DTD
- Linux Batch nodes (Java, Perl, Shell, C, etc.)
- Job scheduler (Cronacle)

Production Queues Status

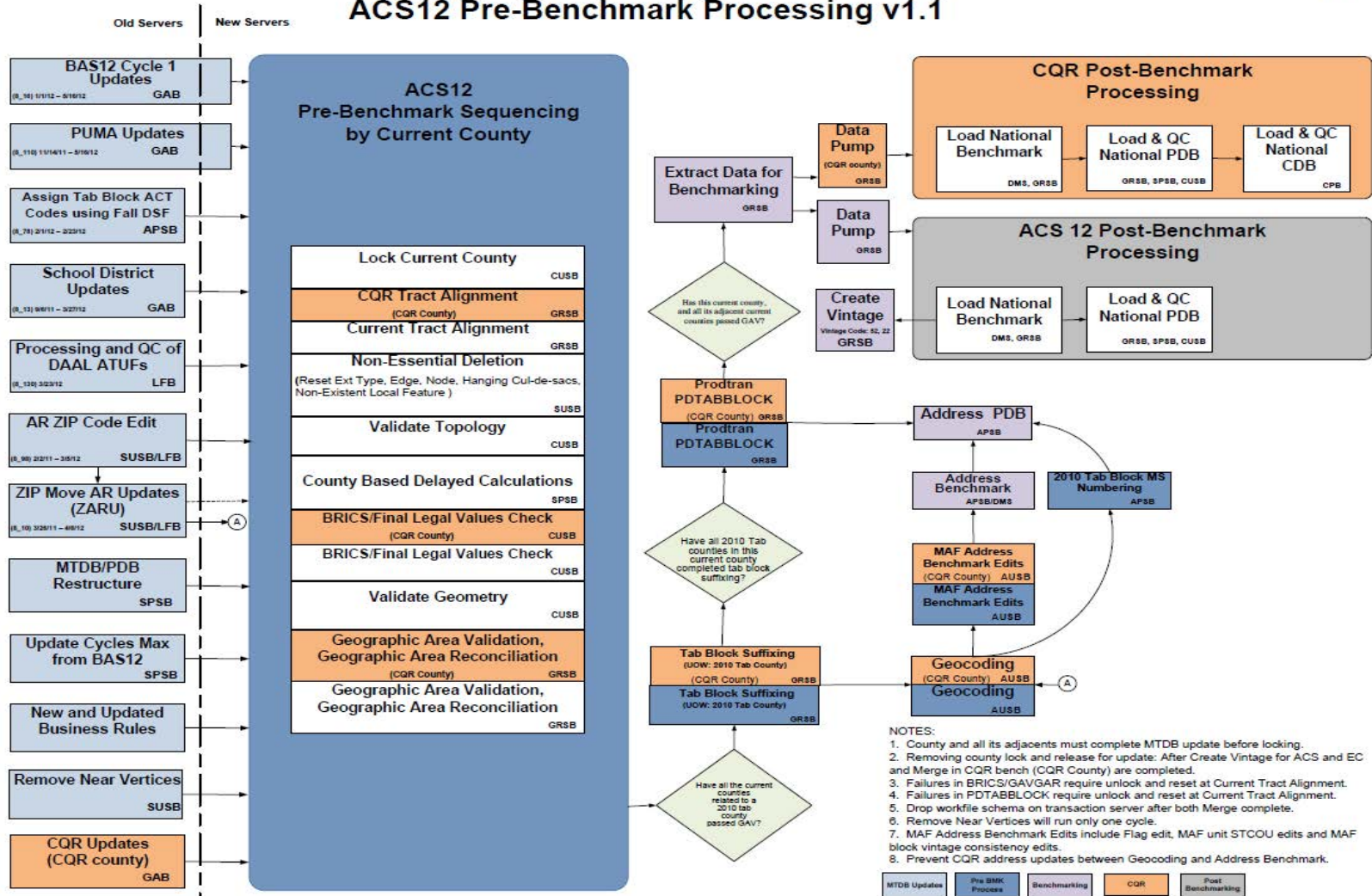
05/03/12 15:25:04

Queue Name	Status	Job Count
ACS12_CQR_RELEASE	HOLD	1
BENCHEXT	RUNNING	7
BENCHMRG	RUNNING	2
BGTRTADJ	RUNNING	1
BLKSUF	RUNNING	9
BMKPUMP	RUNNING	1
BRICSACS	RUNNING	4
CQRBGTRT	QUEUEHOLD	13
CQRBLKSF	RUNNING	1
CQRPCS_JOBS	HOLD	1
CQRPDTAB	QUEUED	1
CQRPDTAB	RUNNING	1
DAALMAFU_HOURLY	QUEUEHOLD	1
DSFREF	QUEUED	41
DSFREF	RUNNING	30
DSFREF_04013	RUNNING	1
DSFREF_06037	QUEUED	2
DSFREF_06037	RUNNING	1
DSFREF_17031	QUEUED	1
DSFREF_17031	RUNNING	1
DSFREF_48201	RUNNING	1
DSFREF_ACSTEST	QUEUED	6
DSFREF_ACSTEST	RUNNING	1
DTDBENCH	RUNNING	1
DTD_PRODBNCH	RUNNING	1
LISRD_TRANSACTION	RUNNING	1
NEEDALON	QUEUED	23
NEEDALON	RUNNING	60
VINTAGE41	SCHEDULED	2

Oracle
User C
2012

17

ACS12 Pre-Benchmark Processing v1.1



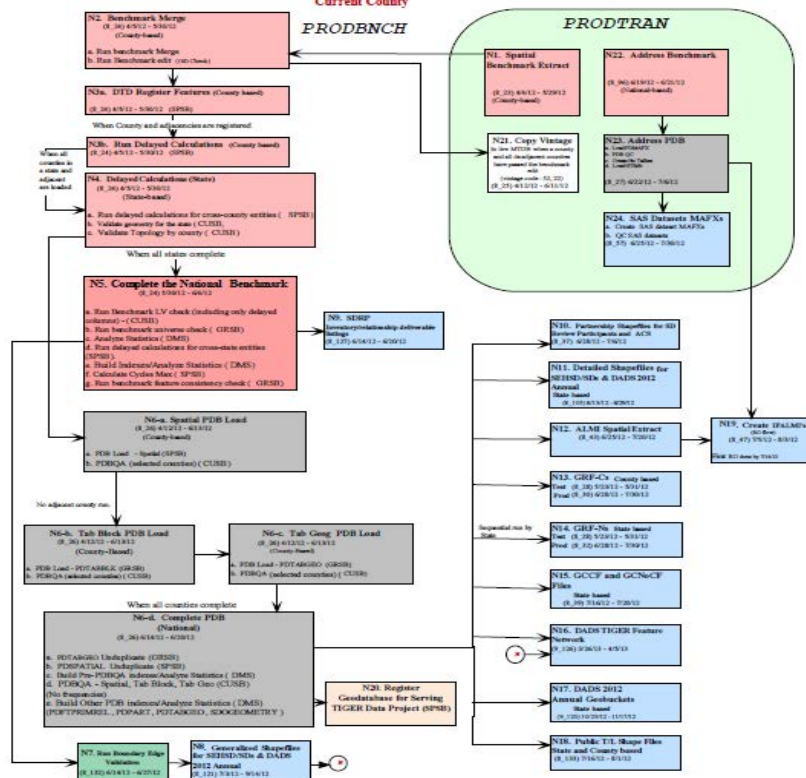
UNIVERSE: US Stateside, Puerto Rico, Island Areas (exclude state 74)
SCHEMAS: ACS12
AC75132400

ACS12 Post-Benchmark Processing v1.1

(x_xx) = Geographic Programs Flowchart Box #

03/12/2012

County-Based
means
Current Count



CONSTRAINTS:

- CONSTRAINTS:
1. No adjacent county run.
 2. PDR Spatial load for a county must be done when all adjacent counties have been loaded into the benchmark.
 3. Drop workfile schema after national merge completes.
 4. Delayed calculations (N3b, N4 and N5d) include Boundary Edges and copy to PDR if record exists.
 5. GRP/Ns run sequentially by State.

Benchmarking	TVM Load	CDB
New Products	Products	Other



Processing Constraints

- Complex Spatial Database, quite large, mission critical
- Growing at 10-15% annually
- Demands from user community for spatial and temporal accuracy and quality
- Stringent processing deadlines remain, so GEO is processing more data in shorter time
- Oracle database on >100 nodes, scores of applications
- Cache fusion



Processing Constraints

- IT Expectation
 - Consolidation
 - Databases
 - Servers
 - Virtualization
 - Elasticity, Agility
 - Service Oriented Architecture
 - Reduction in storage



Future Needs

- GSS Initiative
- Ability to handle larger loads on systems
 - Data Visualization
 - Data Caching now a default expectation
 - Change Detection
 - Quality assessments before update
 - Frequent data exchange with partners
 - Conflation
 - Services including those for Corporate Listing Device



Future Needs

- Ability to handle larger loads on systems (address related)
 - Geocoding
 - Address Matching
 - Address Standardization
 - Unduplication
 - Address Extracts
 - MAF Structure Points
 - Uploads of Address Lists from partners



Future Needs

- Twice a year product generation to be converted to immediate availability after updates as desired and appropriate
- Eliminate months of MAF related large table processing
- Interactive (GATRES) users desire to see the benefit of a better performing database
- Elimination/reduction of cache fusion, our production bottleneck



Oracle Spatial on Exadata - POC

GEO's Exadata Testing

- Oracle coordinated about a week's access
- Access to Exadata X2-8 and a batch node
 - Dedicated Oracle Spatial expert, DBA, architect
 - Daily status meetings with Census staff
- 2 ADC areas, 5 branches involved at Census



Oracle Spatial on Exadata - POC Data

- Data for Testing
 - GEO prepared a spatial dataset on an 11g RAC
 - Maryland and Virginia
 - Scrubbed Title 13
 - Simulated MSPs
 - Topologically valid, complex geometries
 - Data delivered on DVD
 - Disk sanitized at the end of testing



Oracle Spatial on Exadata - POC Apps

- Applications for Testing
 - GEO chose CPU intense, I/O intense applications
 - Spatial, Topological, Relational
 - Very large queries to large number of queries
 - Reads and updates, Oracle cache and JVM
 - Parallel processing, concurrent execution
 - DBA related activities (backup, exports, imports)
 - Workflow control, tracking, logging, job status
 - VALGEO, VALTOPO, DTD, TOLCHECK, LISRDS



Oracle Spatial on Exadata - POC Install

- Installation and Configuration
 - All Census applications could be installed and configured quickly
 - Minor modifications to the drivers to use JDBC on their batch node
 - Their DBA loaded the data into Exadata
 - Driver scripts in PERL and Java libraries bundled
 - Data refreshed from master copy to start fresh
 - Performed same actions at Census



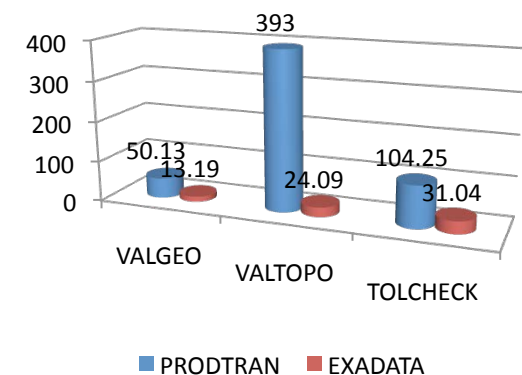
Oracle Spatial on Exadata - POC Test Execution

- Execution per a test plan
 - Identified tests for every day
 - Jobs submitted on Oracle and Census servers
 - Jobs captured results into a control table
 - Overnight runs
 - Daily evaluation of results
 - Additional ad hoc queries during day



Oracle Spatial on Exadata – POC Results

- Test Results
 - **Functional testing had no issues, all applications and DBA activities performed as expected.**
 - Biggest improvement was seen when jobs were submitted in parallel.
 - VALGEO submitted as 84 counties at a time and it finished the whole 158 counties in 13:19 min.
 - Similarly VALTOPO finished in 24:09 min.
 - TOLCHECK finished in 31:04 min.
 - Individual jobs were faster
 - Outperformed PRODTRAN (10 node RAC).



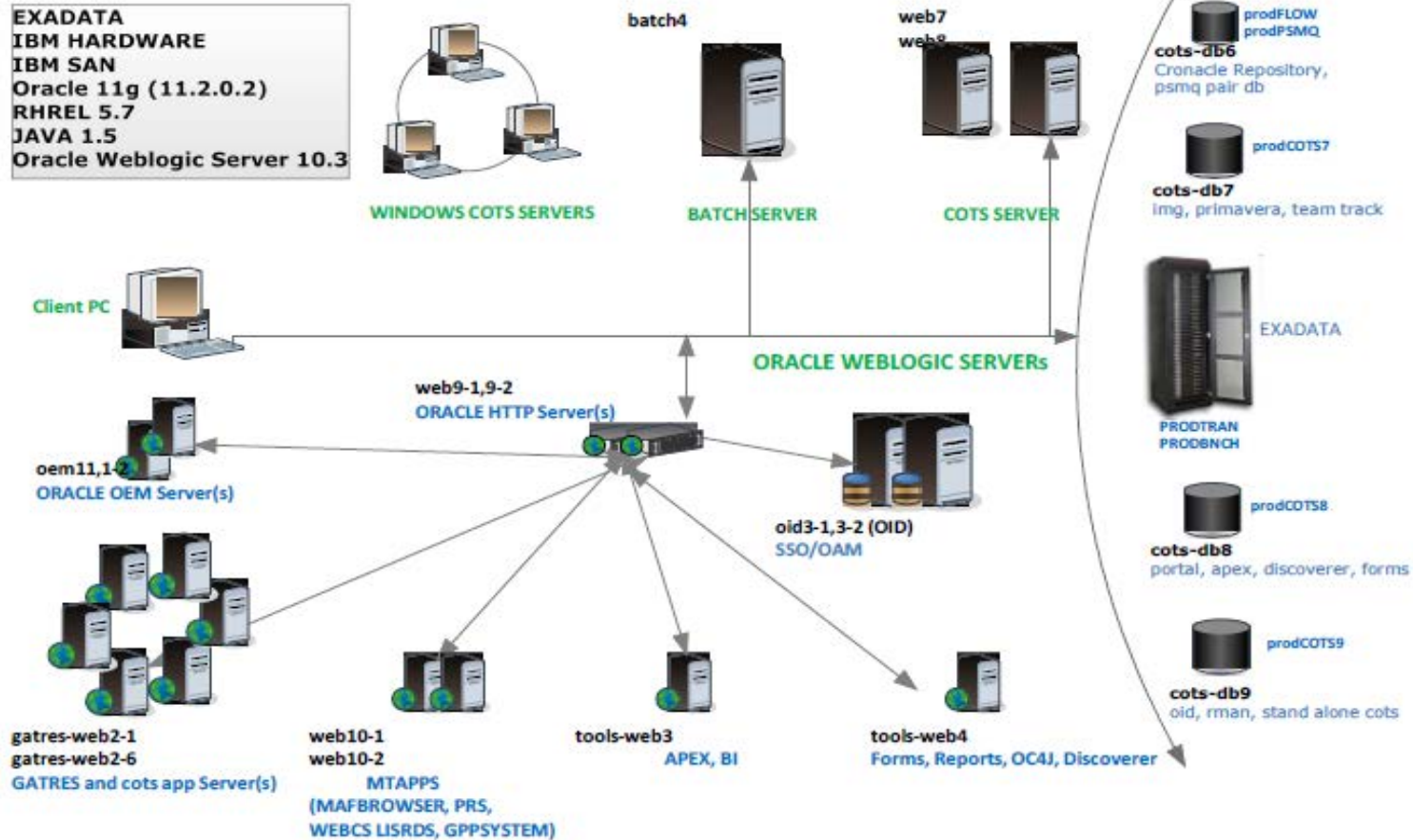


Oracle Spatial on Exadata - Transition

- Preparation for transition, installation, and implementation
 - 11g Spatial database for testing, prototype environment
 - 11g new features – Expert presentations
 - Exadata presentations, technical discussions, solutions design
 - Logistics, planning at data center, tools
 - Training
 - DEV, TEST and Production schedules
 - Exadata as a component of Hardware Refresh & an **Enterprise** solution
 - Integrated Exadata, Oracle spatial, GEO, IT team

GEO's 11g PRODUCTION Environment

EXADATA
IBM HARDWARE
IBM SAN
Oracle 11g (11.2.0.2)
RHREL 5.7
JAVA 1.5
Oracle Weblogic Server 10.3





Migration

- Delete old vintages, identify stable and volatile schemas
- Identify patch sets based on testing (11.2.0.3)
- Identify candidates for compression
- All applications were re-released
- Minimized downtime
- GoldenGate POC
- Timeline of activities and assignment of responsibility



Benefits Realized so far

Activity	Legacy	Exadata																												
MAF/TIGER Schema Refresh (extract, staging, import, topology registration, index rebuild)	4 Days	2 Days																												
Index Rebuilds	2 Days	< 1 Day																												
Table Rebuilds	6-10 Hours for FEATMAFUNITREL (1.2 billion rows)	1.5 Hours																												
Vintage Deletion	3 weeks (indexes were critical for deletion)	< 1 Day, without indexes (able to delete 189,585,759 rows in 08:31:43.01 from 1.2 Billion row table)																												
Table Compression	NA	<table><tr><th>OWNER</th><th>SEGMENT_NAME</th><th>Size in Gig</th><th></th></tr><tr><td colspan="4">-----</td></tr><tr><td>MAFTIGER</td><td>TABBLOCK</td><td>63</td><td>-- no compression</td></tr><tr><td>SJM</td><td>TABBLOCK_QL</td><td>24</td><td>Query Low</td></tr><tr><td>SJM</td><td>TABBLOCK_AH</td><td>16</td><td>Archive High</td></tr><tr><td>SJM</td><td>TABBLOCK_QH</td><td>16</td><td>Query High</td></tr><tr><td>SJM</td><td>TABBLOCK_AL</td><td>15</td><td>Archive Low</td></tr></table>	OWNER	SEGMENT_NAME	Size in Gig		-----				MAFTIGER	TABBLOCK	63	-- no compression	SJM	TABBLOCK_QL	24	Query Low	SJM	TABBLOCK_AH	16	Archive High	SJM	TABBLOCK_QH	16	Query High	SJM	TABBLOCK_AL	15	Archive Low
OWNER	SEGMENT_NAME	Size in Gig																												

MAFTIGER	TABBLOCK	63	-- no compression																											
SJM	TABBLOCK_QL	24	Query Low																											
SJM	TABBLOCK_AH	16	Archive High																											
SJM	TABBLOCK_QH	16	Query High																											
SJM	TABBLOCK_AL	15	Archive Low																											
Index Compression	TABBLOKK_UK; Space Used: 2,377; Space Allocated: 2,386	TABBLOKK_UK; Space Used: 1,455; Space Allocated: 1,519																												

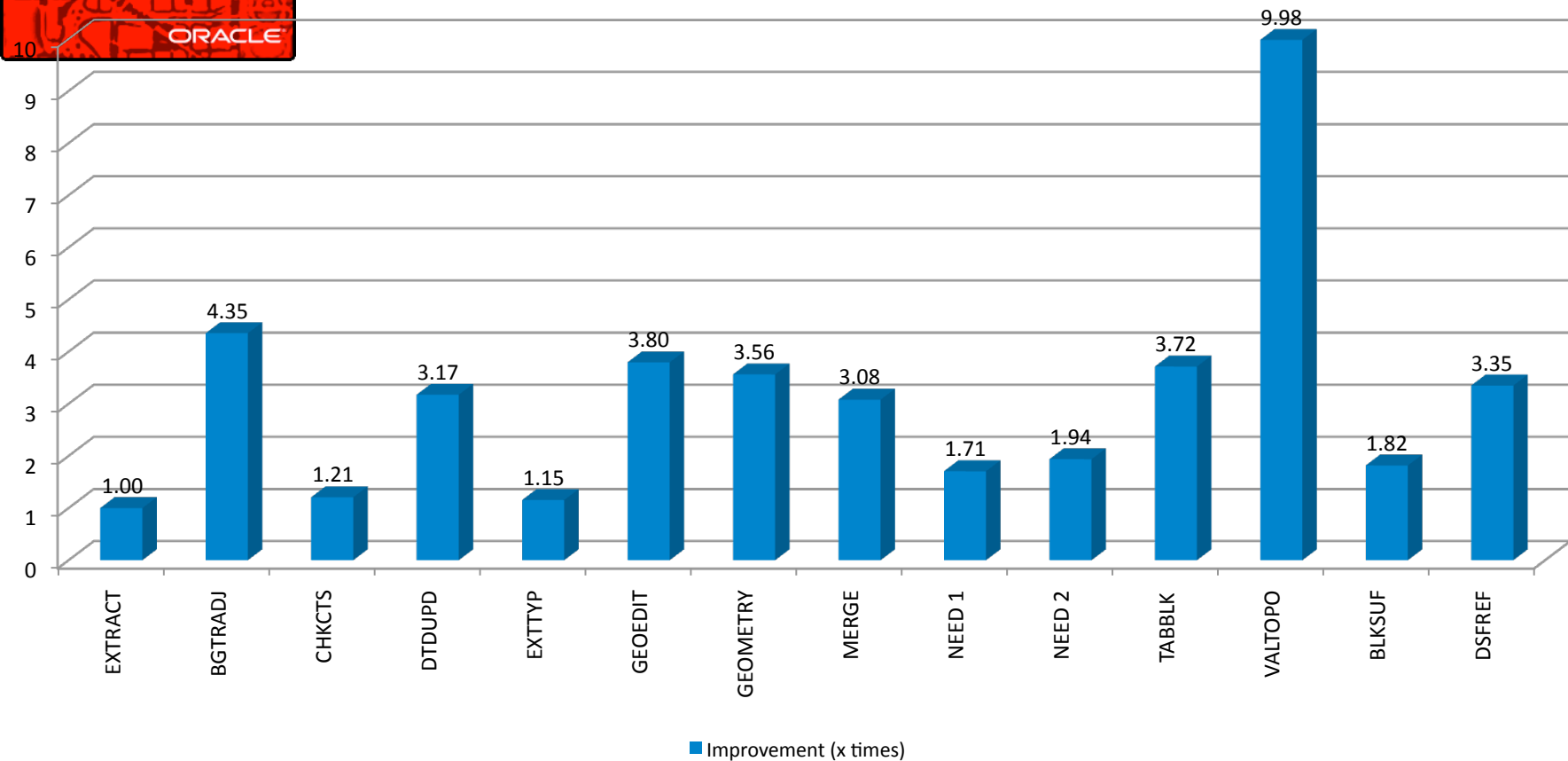


Benefits Realized so far

- Out of the box solution helped advance schedule by months
- One vendor, facilitated one comprehensive solution
- Larger queue sizes for batch jobs without cache fusion
- Reduction in overall calendar time for projects
- DSF Refresh: 98% completed in 6 days versus 3 weeks
- Benchmarking progressing at <50% legacy time

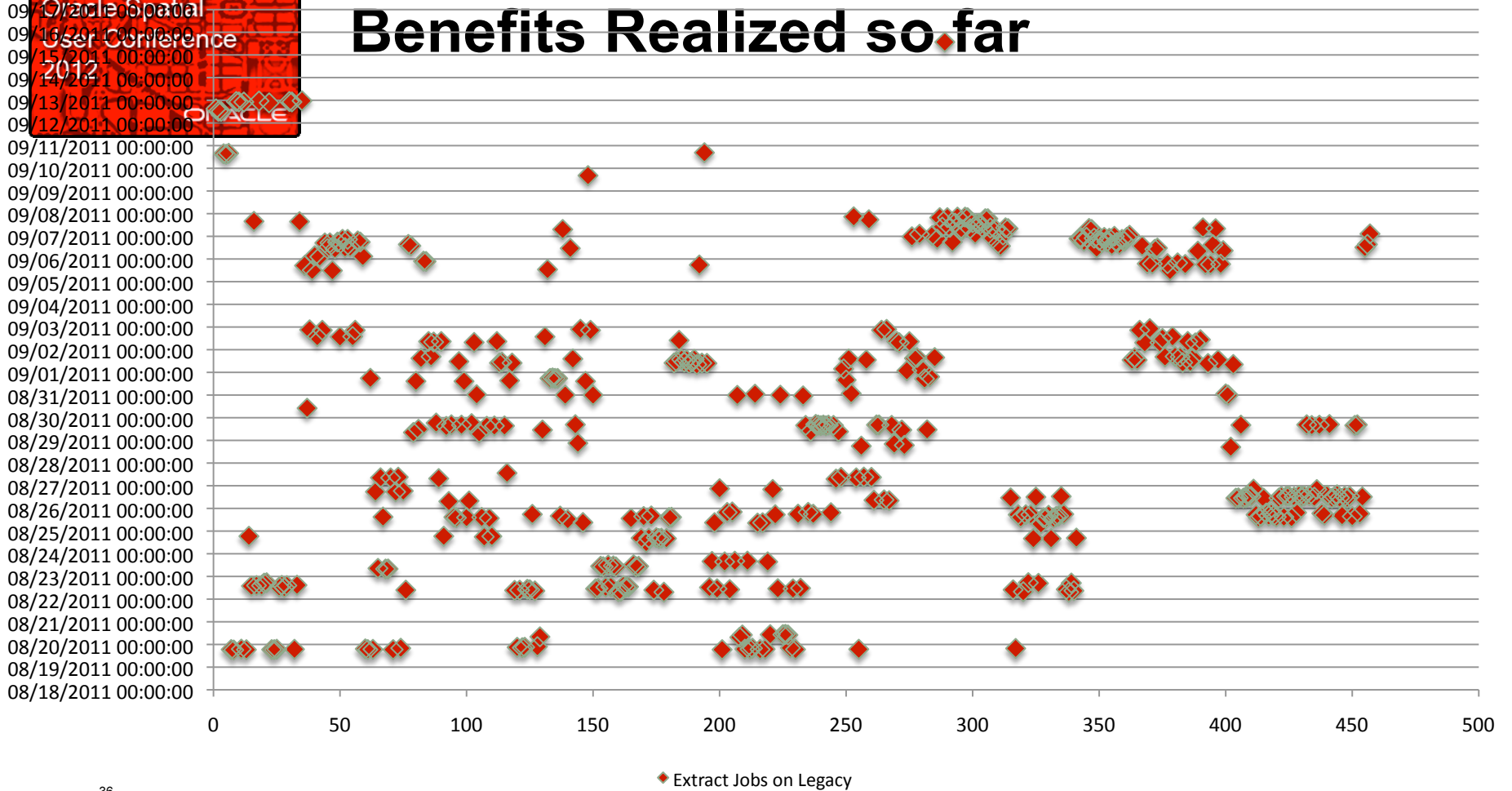


Benefits Realized so far



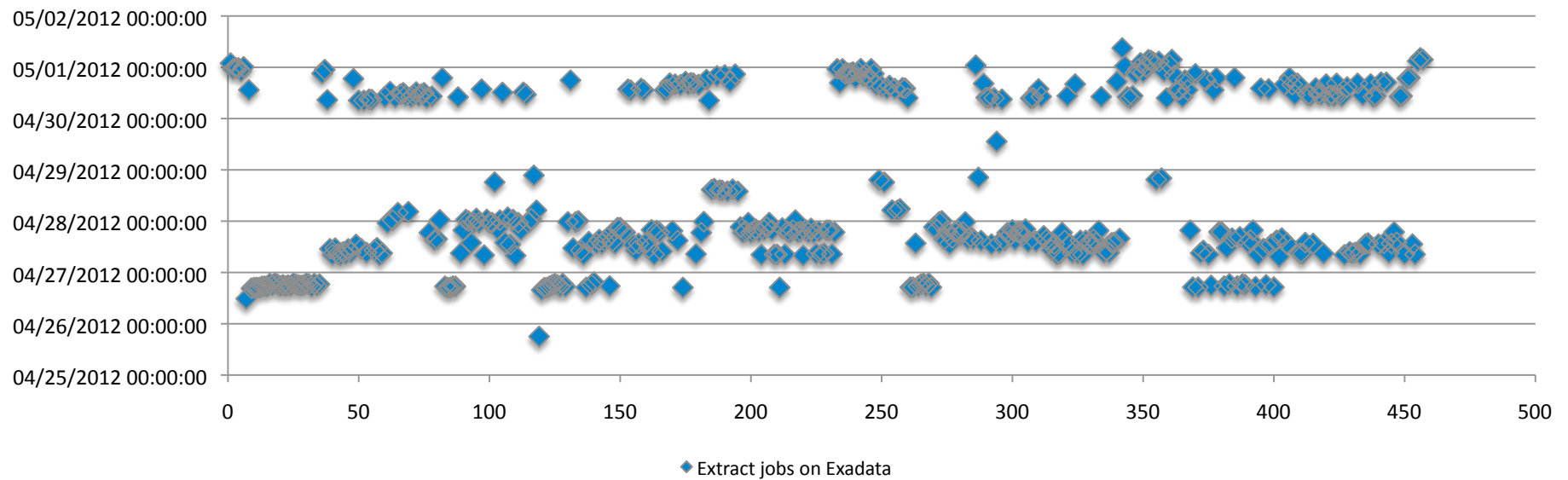
Oracle Spatial
User Conference
2012

Benefits Realized so far



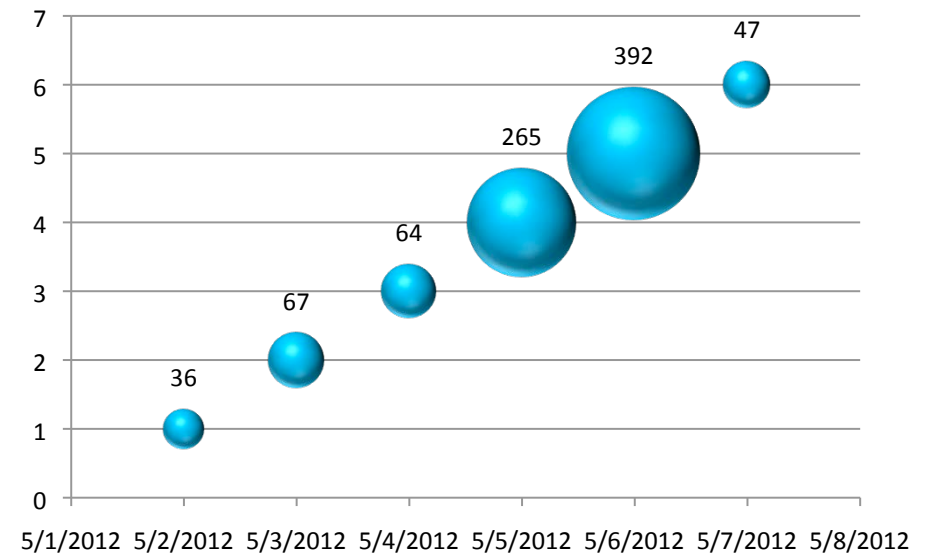
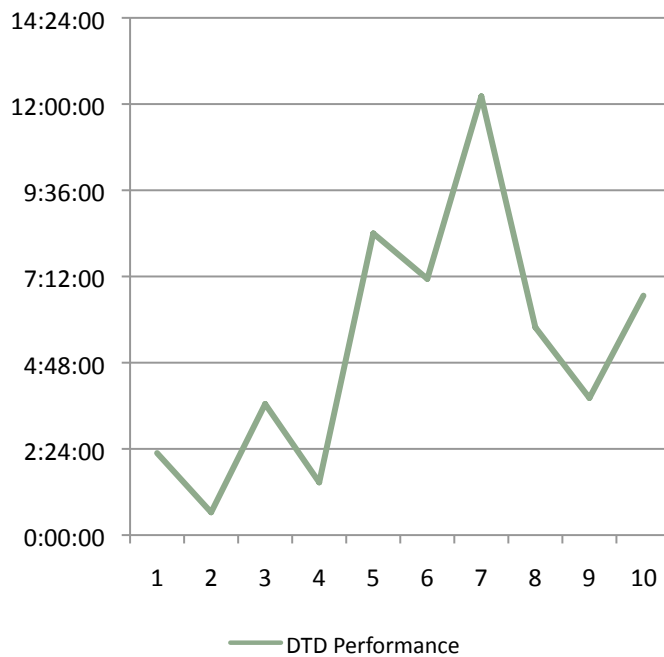


Benefits Realized so far





Benefits Realized so far



DSF Refresh Progress

Q&A