

Data Access White Paper – Food for Thought –

DRC

10/05/2015

Introduction

- This is not a comprehensive review of the data access white paper
- I read/wrote this from the point of view of the SUI/T and the needs for data access and visualization
- This is mostly a list of questions – all of these thoughts are from the point of view of the user
- Date of this document is 2007 with operations starting in 2014

DataBase Queries

- Two Levels: “Large” and “Small”
- Large: 6GB results
 - 20 simultaneous queries
 - 20 per hr → 120 GB/hr
- Small: 0.1GB results
 - 50 simultaneous queries
 - 18,000 per hr (10 queries per sec) → 1800 GB/h

Storage Resources

- 4 service levels
 - Levels 1 – 3 special applications
 - Level 4 by default

Service Level	Community-available Network Bandwidth to storage (total)*	Community-available Disk/tape Storage (total)	Deployment Schedule for Storage	Facility Location
1	6 Gbps	10 PB	1 PB in 2015 plus 1 PB/yr 2015 - 2023	DAC at Archive Center
2	2 Gbps	900 TB	100 TB/yr 2015 - 2023	DAC at Archive Center
3	2 Gbps	100 TB	100 TB 2015	DAC at Archive Center
4	6 Gbps	1 PB	1 PB 2015	U.S. and Chilean Data Access Centers
TOTAL	16 Gbps	12 PB		

Table 1 Open Access Network and Storage Resources by Service Level for U.S. and Chile

Computing Resources

- 4 service levels

Service Level	Community-available Computing Services (total)	Deployment Schedule	By Facility
1	25 TFLOPS	5 TFLOPS/yr from 2015 - 2019	DAC at Archive Center
2	5 TFLOPS/ DAC	1 TFLOPS/yr/DAC from 2015 - 2019	U.S. and Chilean DACs, each
3	1 TFLOPS/ DAC	1 TFLOPS/DAC in 2015	U.S. and Chilean DACs, each
4	1 TFLOPS	1 TFLOP in 2015	U.S. and Chilean DACS, total
TOTAL	44 TFLOPS	See above	See above

Table 2 Open Access Computing Resources by Service Level for U.S. and Chile

Questions ?

- These resources were scoped out 10 years ago
- Are these resource numbers still relevant?
- Are these resource numbers still the 'requirements'?

Questions ?

- Lifetime requirements of DB results?
 - Large: ~1 PB/yr, Small: ~15 PB/yr
- Visualization/interactivity requirements of DB results?
- No requirements for image retrieval resources
 - Are there tiered services here as well?
 - What is the break-point between repeated image retrieval and stored images?
 - Resources needed for all-sky images vs calibrated image access?
 - How many all-sky image browsers are supported?

Allocation of Resources ?

- How are tiered services granted, allocated, monitored?
- Are storage quotas built solely around workspace sizes?
- Are computing resources limited to the resources within a single workspace for a user?
 - Connecting workspaces and batch computing resources?
 - Computing resources within workspaces (e.g., MySQL)?
- “Oversold” model vs. Strict quotas

Access: Policy Issues

- Policy issues not entirely firm or clear but need to implement a system – likely prior to any “final” policies
- What is the current data access policy? US, Chile, France?, Brazil?
- What are expiration rights? (e.g., moving scientists)?
- Alerts are public
 - Does that include all the data associated with alert?