Southern Analysis Region

Performance of an Operating High Energy Physics Data Grid: DØSAR-Grid

Horst Severini

DPF 2004

August 30, 2004

Outline

- **Introduction: Problems to be addressed**
  - Large High Energy Physics projects
  - Lots of data to be analyzed
  - Collaborators spread geographically
- **Solution: DØSAR-Grid (DØ Southern Analysis Region)**
  - Who's involved
  - How the Grid works
  - What's currently underway
  - Outlook for the future
- **Conclusions**

Introduction

Southern Analysis Region

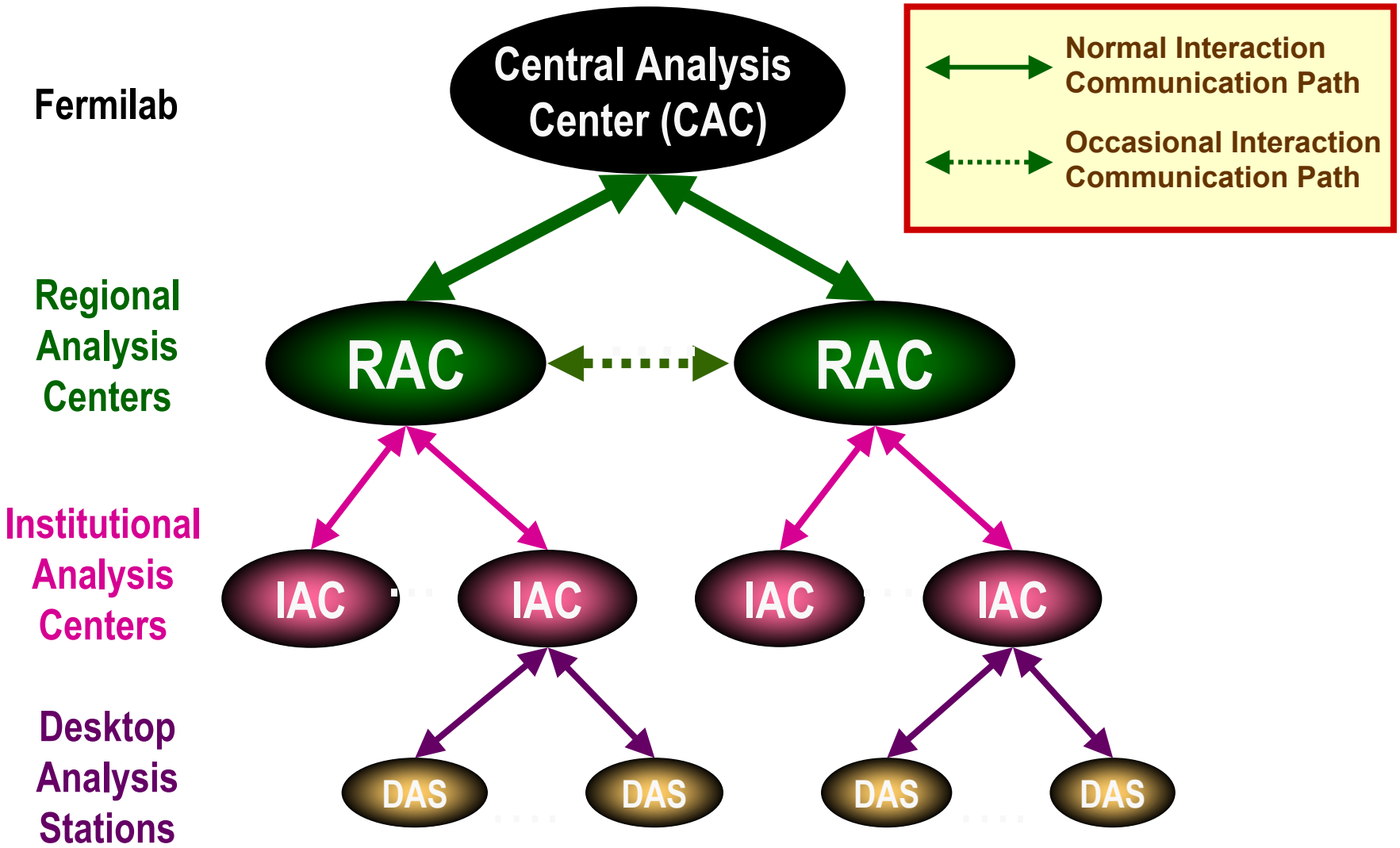
- **Particle Physics employs accelerators and detectors to probe smallest distance scales**
- **Current highest energy accelerators**
  - Fermilab's Tevatron
  - CERN's Large Hadron Collider (LHC)
  - Further into the Future: the Linear Collider (LC)
- **Problems:**
  - ***Petabytes*** of Data per year
  - Ever more computing power required
  - HEP Collaborators are geographically spread
- **One Solution: Data Grids**
  - Distributed computing = ability to cope with large amounts of data
  - Geography not a problem.

Glossary

- **SAM: Sequential data Access via Meta-data**
  - Data handling system to store and retrieve files and associated metadata
- **SamGrid:**
  - Fermilab's Grid Middleware tools system
- **VDT (Virtual Data Toolkit): Globus, GridFTP, Condor, …**
  - Generic Grid Middleware, High Throughput batch system
- **Grid3**
  - International Data Grid Collaboration operated jointly by the U.S. Grid projects iVDGL, GriPhyN, and PPDG
- **VO: Virtual Organization**
  - Flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources

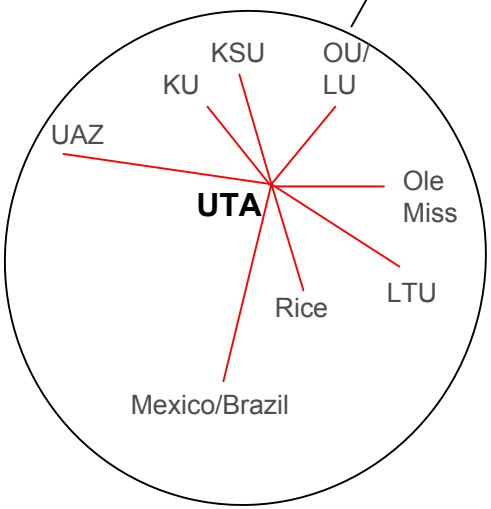
DØ Remote Analysis Model (DØRAM)

Southern Analysis Region



DØSAR Geography

Southern Analysis Region



The first US Region in DØ to adopt hierarchical "Remote Analysis Model"

DØSAR Institutions

Southern Analysis Region

- **First Generation IAC's**
  - University of Texas at Arlington
  - Louisiana Tech University
  - Langston University
  - University of Oklahoma
  - Tata Institute (India)
- **Second Generation IAC's**
  - Cinvestav, Mexico
  - Universidade Estadual Paulista, Brazil
  - University of Kansas
  - Kansas State University

Each 1st generation institution is paired with a 2nd generation institution to help expedite implementation of D0SAR capabilities

- **Third Generation IAC's**
  - Ole Miss, MS
  - Rice University, TX
  - University of Arizona, Tucson, AZ

Both 1st and 2nd generation institutions can then help the 3rd generation institutions implement D0SAR capabilities

DØSAR Computing Resources

Site	CPUs	RAM (GB)	Disk (TB)
LTU	19	11	0.6
LU	16	10	1.3
OU	19	11	1.5
SPRACE	48	24	4.5
TATA	52	26	1.6
UTA-RAC	192	192	51.8
OSCER	270	270	8.5
Total	346/616	274/544	61.3/69.8

Not all Farms are dedicated; many are shared with other Experiments

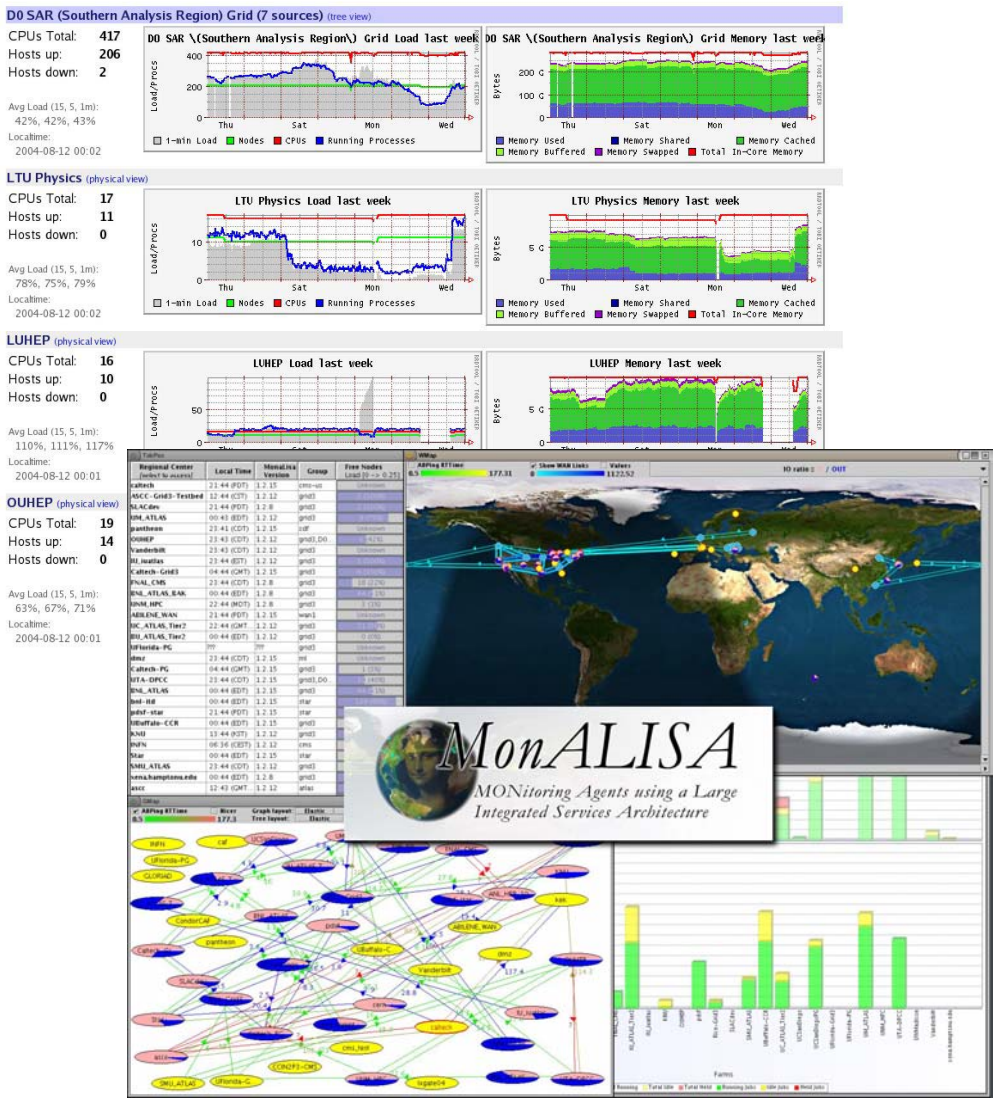
DØSAR Pre-Grid Solution: McFarm

- **Hi-level production manager layered on mc_runjob, SAM, and optional batch system**
- **Set of daemon and utility scripts on job server and worker nodes**
  - ✓ Fetch input files from SAM
  - ✓ Parallelize request and Distribute jobs
  - ✓ Store metadata and merged output in SAM
  - ✓ Archive request jobs locally
  - ✓ Inform operator of status and problems (and solutions)
- **McFarm Data Production to Date:**

Site	LU	OU	UTA	LTU	SPRACE	TATA
Evts	1.25 mil	1 mil	12.5 mil	0.25 mil	1 mil	3.25 mil

Monitoring Software at DØSAR

- **McFarmGraph, McPerM**
  - MC production Job Monitoring systems
  - Use Globus and GridFTP
  - McFarmGraph provides Graphical MC request info
  - McPerM monitors MC production Performance
- **Ganglia resource monitoring**
  - Mid-level System and Process Information Monitoring tool
- **MonALISA**
  - High-level Performance and Job Monitoring tool



Grid-Enabling DØSAR: Status

Southern Analysis Region

- **Continue current production → SamGrid/McFarm**
  - McFarm requires local pre-staged DØ release, minbias, cardfiles, mc_runjob, and root access
  - McFarm production initiated via SamGrid framework
  - Requests submitted via SamGrid to *McFarm exec sites* (UTA-RAC, LTU, LU, OU)
- **Prepare for SamGrid/runjob stable operation**
  - McFarm will be replaced by SamGrid/runjob
  - SamGrid/runjob will stage everything in a remote sandbox
  - Requests will be submitted via SamGrid to *runjob exec sites*
- **Installing SamGrid/runjob exec site on OSCER**
  - Generic Cluster without root access
  - OS Compatibility, internal Network and Firewall issues

Grid-Enabling DØSAR: Status

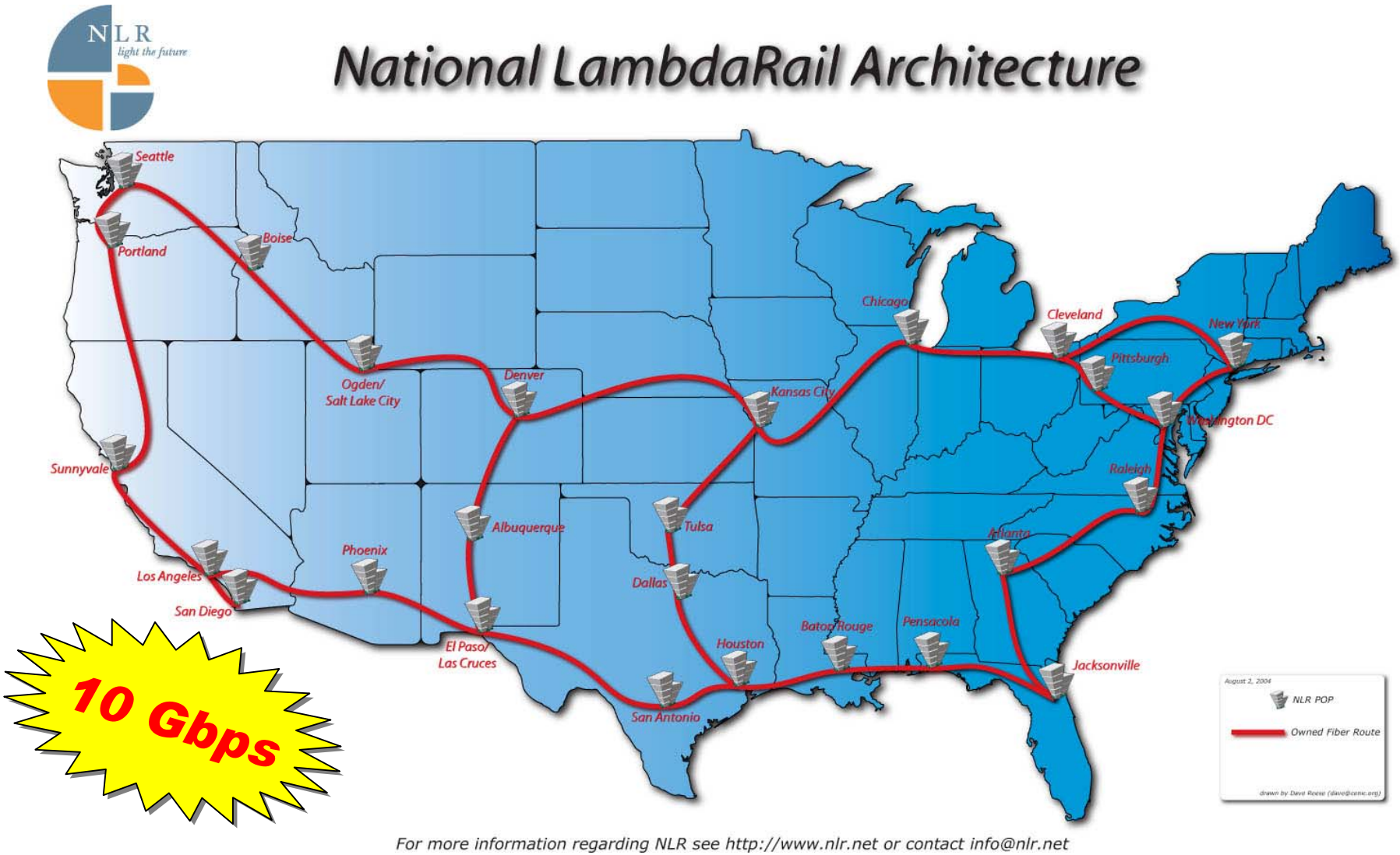
- **"Dial-A-Job" user front end for Grid Job Submission**
  - List datasets, minbias, etc in SAM
  - Create datasets; list releases, runjob, cardfiles, etc in SAM
  - Submits jobs to the Grid
- **Establishing Interoperability between SamGrid and VDT/Grid3 Tools**
  - SamGrid can use Grid3's VDT Installation
  - One GateKeeper handles both SamGrid & Grid3
- **Creating Proxy Database at RAC**
  - Installed but not tested
- **Participating in MC production, Reprocessing, Data reduction, Analysis**

Grid-Enabling DØSAR: Plans

- **Establish SAR Virtual Organization (VO)**
  - Authorization and Resource Coordination Management in Region
- **Establish Client and Execution sites**
  - SPRACE, KSU and KU, CSF, Miss, etc.
  - Scheduler site at LU already exists
- **Migrate SamGrid/McFarm ⇒ SamGrid/runjob**
- **At OSCER:**
  - Bring up current Linux Cluster
  - Deploy 64 CPU Itanium2 Cluster (NSF MRI Grid Research Grant)
- **Obtain and deploy 288 CPU Opteron cluster at OU**
  - DOE EPSCOR Grant: OCHEP OU,OSU,LU
- **Utilize MonALISA for System and Job Monitoring**
- **Develop Brokering Tools for Automatic Site Selection**
- **Improve Cluster & System fault tolerance, fail-over, redundancy**

5 Year Plan: National LambdaRail

Southern Analysis Region



Conclusions

- **SAR has considerable resources effectively used**
  - *… And more to come!*
- **SAR moving to SamGrid/runjob**
  - McFarm ⇒ SamGrid/McFarm ⇒ SamGrid/runjob
  - Migrate SAR monitoring tools to SamGrid/runjob
- **Need enhanced tools for seamless job creation, execution, and data storage**
- **5 Year Plan: NLR – 10 Gbps**
- **SAR demonstrates: *The Grid Empowers!***