

DØ Regional Analysis Center Concepts

Roadmap of Talk

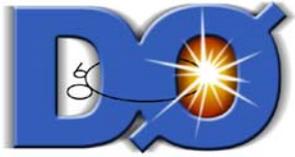
CHEP 2003

UCSD

March 24-28, 2003

Lee Lueking

- **The Mission**
- **The Resource Potential**
- **DØ Regional Strategy**
- **RAC Details**
- **RAC progress**
- **Summary and Future**



DØ Offsite Analysis Task Force

Official members (and other participants)

Iain Bertram – Lancaster University, UK

Chip Brock, Dugan O'Neil – Michigan State University

John Butler – Boston University

Gavin Davies, (Rod Walker) – Imperial College, United Kingdom

*Amber Boehnlein, David Fagan, Alan Jonckheere, Lee Lueking, Don Petravik,
Vicky White, (co-chair) - Fermilab*

Nick Hadley (co-chair) - University of Maryland

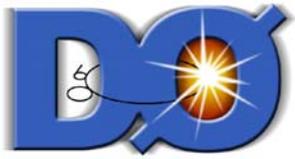
Sijbrand de Jong - University of Nijmegen, The Netherlands

*Peter Maettig, (Daniel Wicke, Christian Schmitt) – Wuppertal, Germany
(Christian Zeitnitz) – Mainz, Germany*

*Pierre Petroff (co-chair) - Laboratoire de l'Accélérateur Lineaire, France
(Patrice Lebrun) – ccin2p3 in Lyon, France*

Jianming Qian – University of Michigan

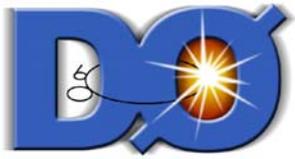
Jae Yu – University of Texas Arlington



A Complex Mission!

We have a very complex physics mission:

- Billions of recorded triggers
- Dozens of physics analysis areas
- Complex analyses, Precision measurements, Minute signal searches, subtle systematics
 - Understand the underlying event consistent with 5 MeV/c² statistical precision on M_W
 - Understand the jet energy scale to more precisely measure M_{top}
 - Tag and vertex B mesons in an environment of 5-10 overlapping interactions
- Estimated R2a (through 2004) computing needs for MC, Reconstruction, and Analysis. Needs beyond 2004 are larger still.
 - 4 THz CPU
 - 1.5 PB total data archive



Many Potential Resources, But...

- **We have many potential resources**
 - **Technology and Computing Resources abound.**
 - **CPU and memory are inexpensive**
 - **Networking is becoming more pervasive**
 - **Disk and tape storage is affordable**
 - **An army of Physicists, Over 600 collaborators, are “available”**

- **But, they are not all in one place anymore, and they are not really “ours”**
 - **The resources are distributed around the world at 80 institutions in 18 countries on 4 continents.**
 - **In most places, the resources are shared with other experiments or organizations**
- **Management, Training, Logistics, Coordination, Planning, Estimating needs, and Operation are real hard**
- **Infrastructure and tools needed to pull this all together are essential.**

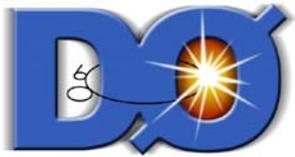
The Good News is ...

There are \$\$\$, €€, and £££ for computing.

The Rub is...

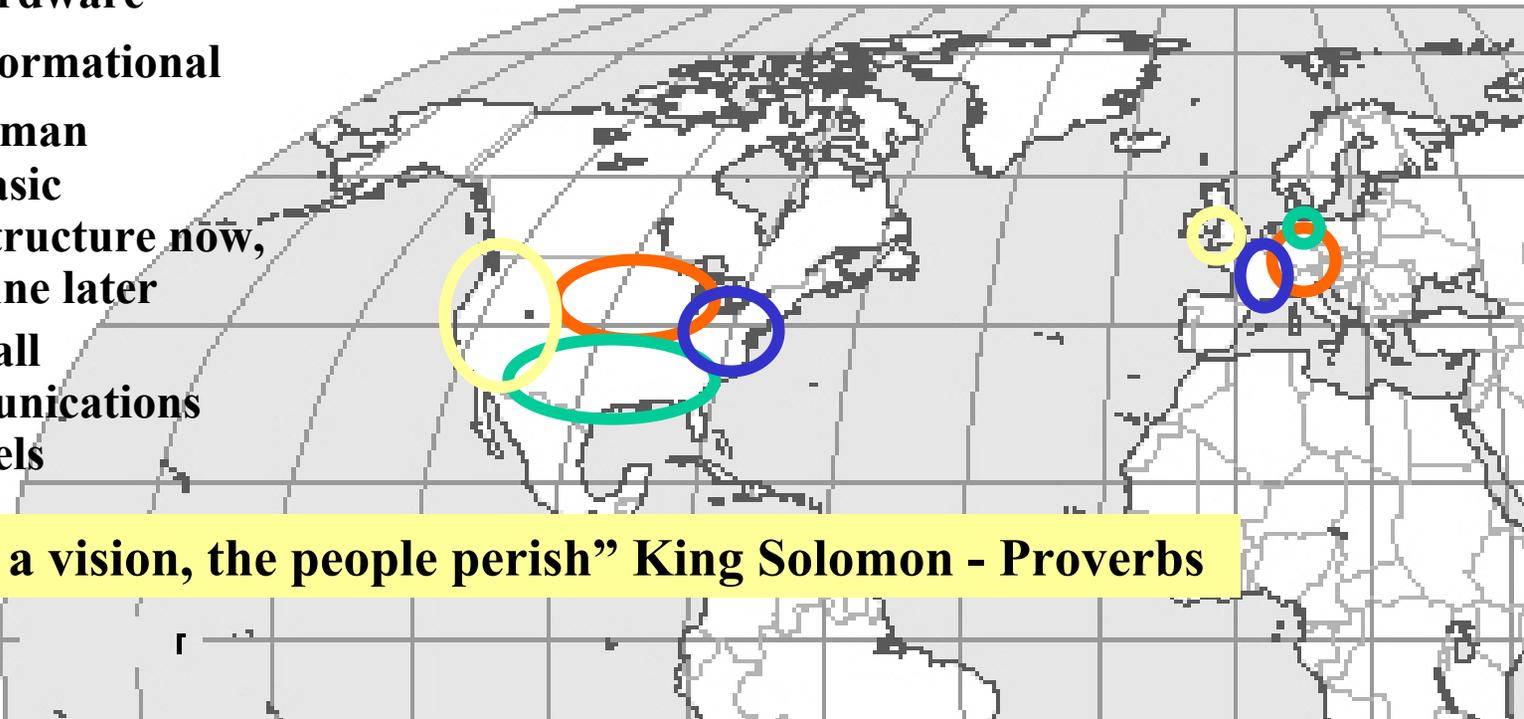
It is for many projects, LHC, Grid, and multi-disciplinary...

so we need to share and be opportunistic

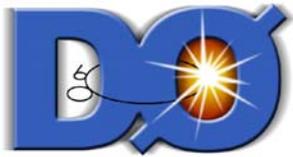


The Overall Game Plan

- **Divide and conquer**
 - Establish 6-10 geographical/political regions.
 - Establish a **Regional Analysis Center (RAC)** in each area.
 - Define responsibilities for each region.
- **Enable the effective use of all resources**
 - Hardware
 - Informational
 - Human
- **Lay basic infrastructure now, fine-tune later**
- **Open all communications channels**



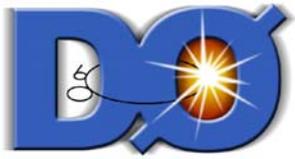
“Without a vision, the people perish” King Solomon - Proverbs



The DØ Process

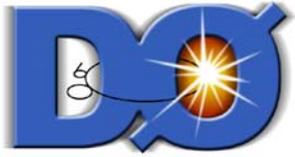
- **1998: DØ Computing Model-** The distributed computing concepts in SAM were embraced by the DØ management. All of DØ 's Monte Carlo was produced at remote centers. **DØ DH in section 8.**
- **2001: DØRACE – Remote Analysis Coordination Effort** team helped to get the basic DØ infrastructure to the institutions. With this effort, 60% of the DØ sites have official analysis code distributions and 50% have SAM stations.
- **2002: RAC grassroots team – Met throughout spring and summer to write a formal document outlining the concepts.***
- **2002: OATF - Offsite Analysis Task Force – Charged by the Spokespersons to further study the needs of offsite computing and analysis**
- **DØ Finance committee – decides how the collaboration as a whole will contribute remote computing resources to the experiment.**
- **Plans for MOU's are being made.**

***Bertram, *et al.*, "A Proposal for DØ Regional Analysis Centers", DØ Internal Note # 3984, Unpublished(2002)**



Why Regions are Important

- 1. Opportunistic use of ALL computing resources within the region**
- 2. Management for resources within the region**
- 3. Coordination of all processing efforts is easier**
- 4. Security issues within the region are similar, CA's, policies...**
- 5. Increases the technical support base**
- 6. Speak the same language**
- 7. Share the same time zone**
- 8. Frequent Face-to-face meetings among players within the region.**
- 9. Physics collaboration at a regional level to contribute to results for the global level**
- 10. A little spirited competition among regions is good**

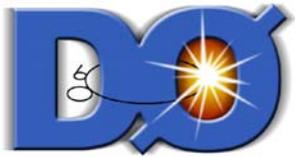


Deployment Model

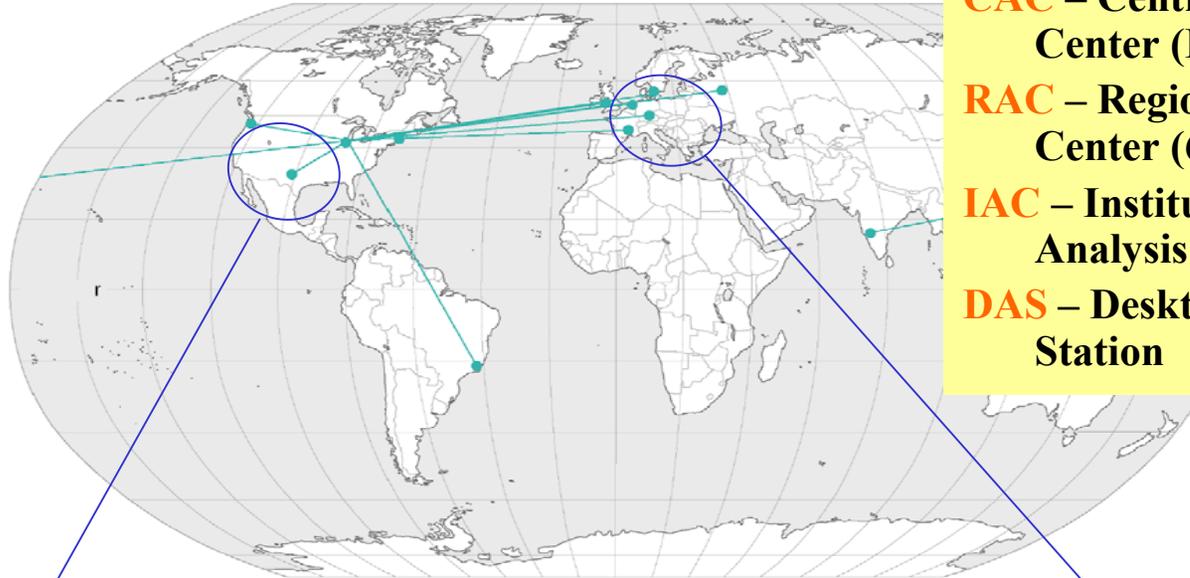
- **Fermilab-centric SAM infrastructure is in place, ...**



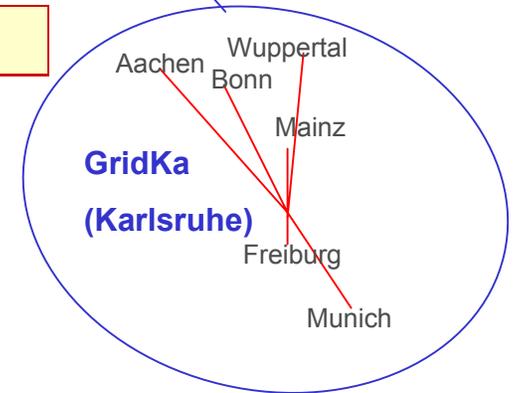
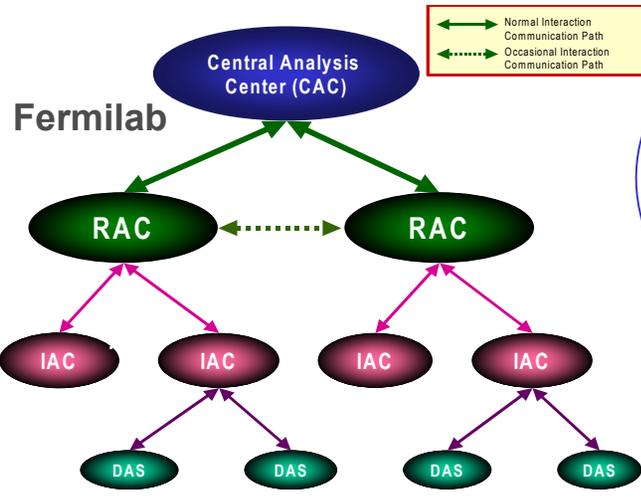
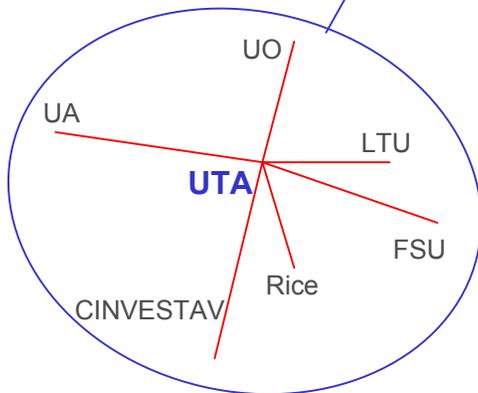
...now we transition to more hierarchical Model →

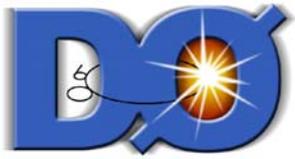


Hierarchical Model



- CAC** – Central Analysis Center (Fermilab)
- RAC** – Regional Analysis Center (6-10)
- IAC** – Institutional Analysis Center
- DAS** – Desktop Analysis Station





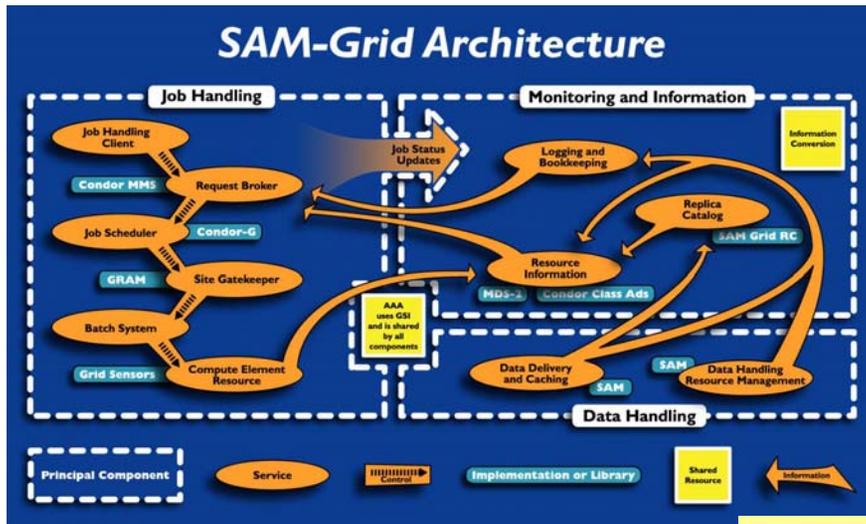
RAC Functionality

- **Preemptive caching**
 - **Coordinated globally**
 - **All DSTs on disk at the sum of all RAC's**
 - **All TMB files on disk at all RACs, to support mining needs of the region**
 - **Coordinated regionally**
 - **Other formats on disk: Derived formats & Monte Carlo data**
 - **On-demand SAM cache: ~10% of total disk cache**
- **Archival storage (tape - for now)**
 - **Selected MC samples**
 - **Secondary Data as needed**
 - **CPU capability**
 - **supporting analysis, first in its own region**
 - **For re-reconstruction**
 - **MC production**
 - **General purpose DØ analysis needs**
 - **Network to support intra-regional, FNAL-region, and inter-RAC connectivity**

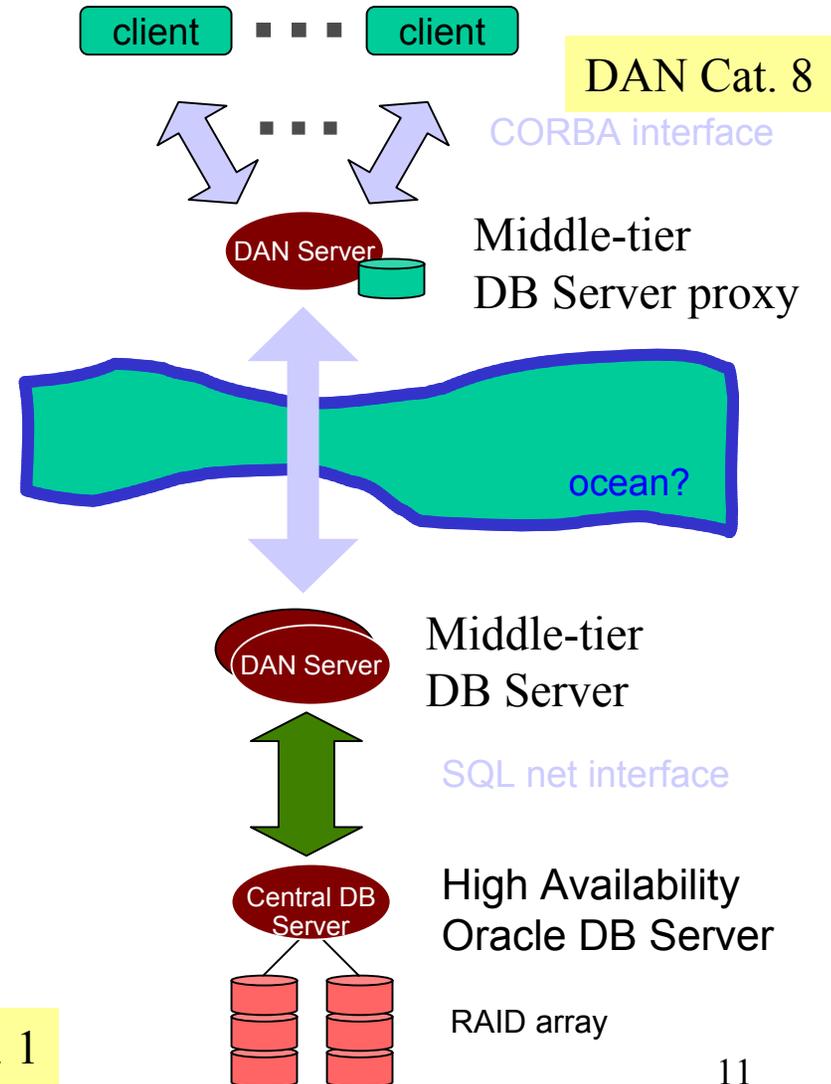


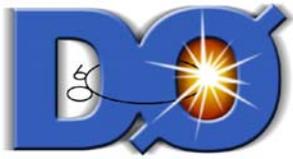
Required Server Infrastructure

- SAM-Grid (SAM + JIM) Gateway
- Oracle database access servers (DAN)
- Accommodate realities like:
 - Policies and culture for each center
 - Sharing with other organizations
 - Firewalls, private networks, et cetera



JIM Cat. 1





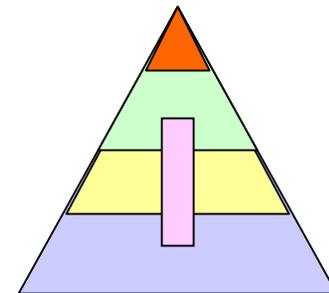
Data Model

Fraction of Data Stored

Data Tier	Size/event (MB)	FNAL Tape	FNAL Disk	Remote Tape	Remote Disk
RAW	0.25	1	0.1	0	0
Reconstructed	0.50	0.1	0.01	0.001	0.005
DST	0.15	1	0.1	0.1	0.1
Thumbnail	0.01	4	1	1	2
Derived Data	0.01	4	1	1	1
MC D0Gstar	0.70	0	0	0	0
MC D0Sim	0.30	0	0	0	0
MC DST	0.40	1	0.025	0.025	0.05
MC TMB	0.02	1	1	0	0.1
MC PMCS	0.02	1	1	0	0.1
MC root-tuple	0.02	1	0	0.1	0
Totals RIIa/RIIb		1.5PB/ 8 PB	60TB/ 800 TB	~50TB	~50TB

per Region

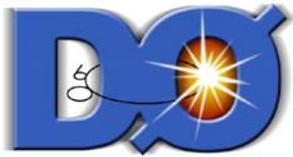
Data Tier Hierarchy



▲ Metadata
~0.5TB/year

Numbers are rough estimates

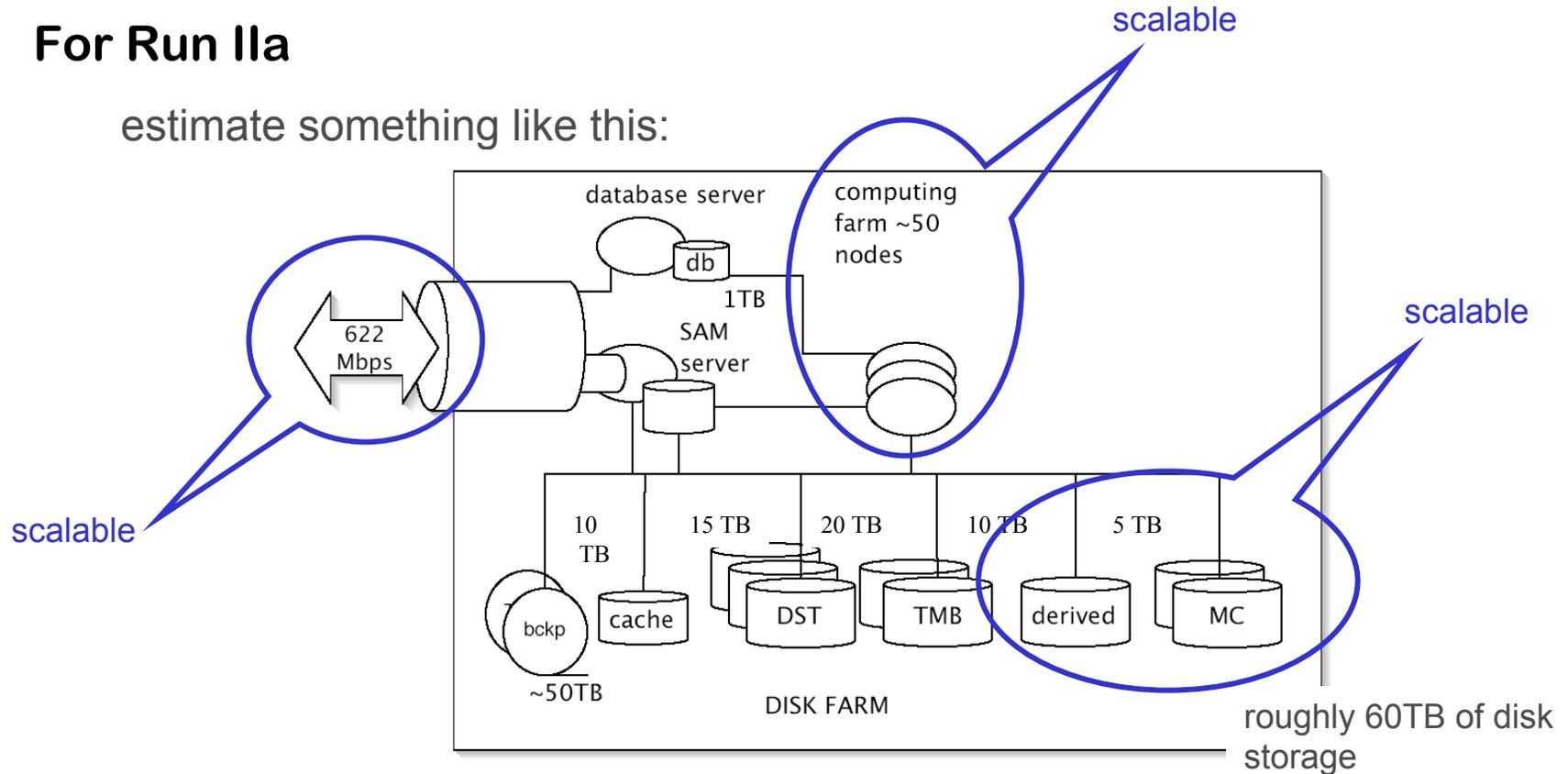
the cpb model presumes:
25Hz rate to tape, Run IIa
50Hz rate to tape, Run IIb
events 25% larger, Run IIb



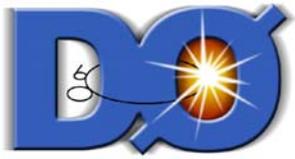
Summary of the *minimum* RAC

For Run IIa

estimate something like this:



- This alone adds > 500 cpu's, deployed in an efficient way - where the physicists are
- IAC's should have have considerable additional capability
- All in host countries.

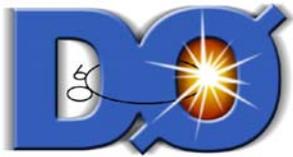


Characterizing RAC's

Hardware needed to achieve various levels of RAC utility

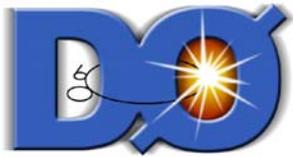
Hardware	Good	Better	Best
Network Connectivity	1 Gbps	1 Gbps	10 Gbps
Disk Cache	60 TB	80 TB	100 TB
Archival Storage	0	100 TB	500 TB
HA Servers	1	2	4
Processing CPU's	50 x (Clock Rate de Jour)	100 x (Clock Rate de Jour)	200 x (Clock Rate de Jour)
Estimated Cost	\$250k	\$500k	\$1M

This is the Run Iia investment



Challenges

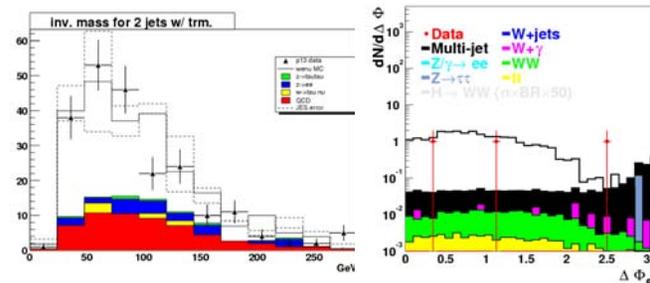
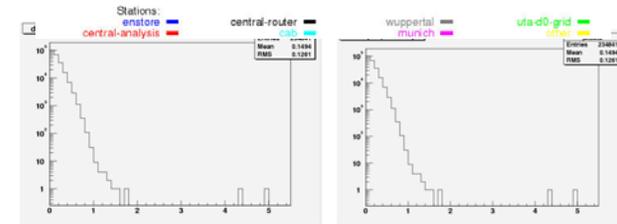
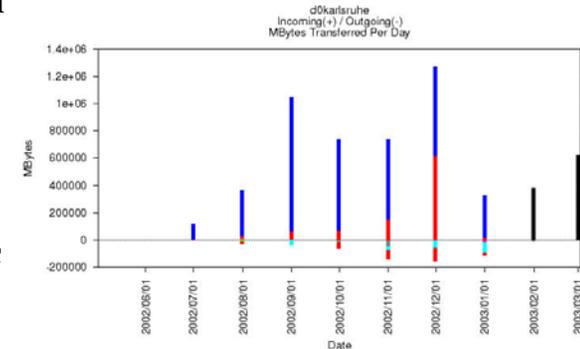
- **Operation and Support**
 - **Ongoing shift support: 24/7 “helpdesk” shifters (trained physicists)**
 - **SAM-Grid station administrators: Expertise based on experience installing and maintaining the system**
 - **Grid Technical Team: Experts in SAM-Grid, DØ software + technical experts from each RAC.**
 - **Hardware and system support provided by centers**
- **Production certification**
 - **All DØ MC, reconstruction, and analysis code releases have to be certified**
- **Special requirements for certain RAC’s**
 - **Forces customization of infrastructure**
 - **Introduces deployment delays**
- **Security issues, grid certificates, firewalls, site policies.**



RAC Prototype: GridKa



- **Overview:** Aachen, Bonn, Freiburg, Mainz, Munich, Wuppertal
 - **Location:** Forschungszentrum Karlsruhe (FZK)
 - **Regional Grid development, data and computing center. Established: 2002**
 - **Serves 8 HEP experiments:** Alice, Atlas, BaBar, CDF, CMS, Compass, DØ, and LHCb
- **Political Structure:** Peter Mattig (wuppertal) FNAL rep. to Overview Board, C. Zeitnitz (Mainz), D. Wicke (Wuppertal) Tech. Adv. Board reps.
- **Status:** Auto caching Thumbnails since August
 - Certified w/ physics samples
 - Physics results for Winter conferences
 - Some MC production done there
 - Very effectively used by DØ in Jan and Feb.



Resource Overview: (summarized on next page)

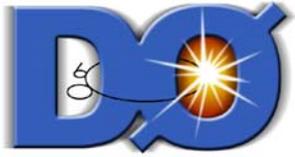
- **Compute:** 95 x dual PIII 1.2GHz, 68 x dual Xeon 2.2 GHz. DØ requested 6%. (updates in April)
- **Storage:** DØ has 5.2 TB cache. Use of % of ~100TB MSS. (updates in April)
- **Network:** 100Mb connection available to users.
- **Configuration:** SAM w/ shared disk cache, private network, firewall restrictions, OpenPBS, Redhat 7.2, k 2.418, DØ software installed.



Summary of Current & Soon-to-be RACs

RAC	IAC's	CPU ΣHz (Total*)	Disk (Total*)	Archive (Total*)	Schedule
GridKa @FZK	Aachen, Bonn, Freiburg, Mainz, Munich, Wuppertal,	52 GHz (518 GHz)	5.2 TB (50 TB)	10 TB (100TB)	Established as RAC
SAR @UTA (Southern US)	AZ, Cinvestav (Mexico City), LA Tech, Oklahoma, Rice, KU, KSU	160 GHz (320 GHz)	25 TB (50 TB)		Summer 2003
UK @tbd	Lancaster, Manchester, Imperial College, RAL	46 GHz (556 GHz)	14 TB (170 TB)	44 TB	Active, MC production
IN2P3 @Lyon	CCin2p3, CEA-Saclay, CPPM-Marseille, IPNL-Lyon, IRES-Strasbourg, ISN- Grenoble, LAL-Orsay, LPNHE-Paris	100 GHz	12 TB	200 TB	Active, MC production
DØ @FNAL (Northern US)	Farm, cab, clued0, Central- analysis	1800 GHz	25 TB	1 PB	Established as CAC

*Numbers in () represent totals for the center or region, other numbers are DØ's current allocation.



From RAC's to Riches

Summary and Future

- **We feel that the RAC approach is important to more effectively use remote resources**
- **Management and organization in each region is as important as the hardware.**
- **However...**
 - **Physics group collaboration will transcend regional boundaries**
 - **Resources within each region will be used by the experiment at large (Grid computing Model)**
 - **Our models of usage will be revisited frequently. Experience already indicates that the use of thumbnails differs from that of our RAC model.**
 - **No RAC will be completely formed at birth.**
- **There are many challenges ahead. We are still learning...**