

# User Analysis Computing at U.S. Tier-2 Centers

F. van Lingen

F. Würthwein

R. Cavanaugh

Caltech

U.C. San Diego

U. Florida

# Goals

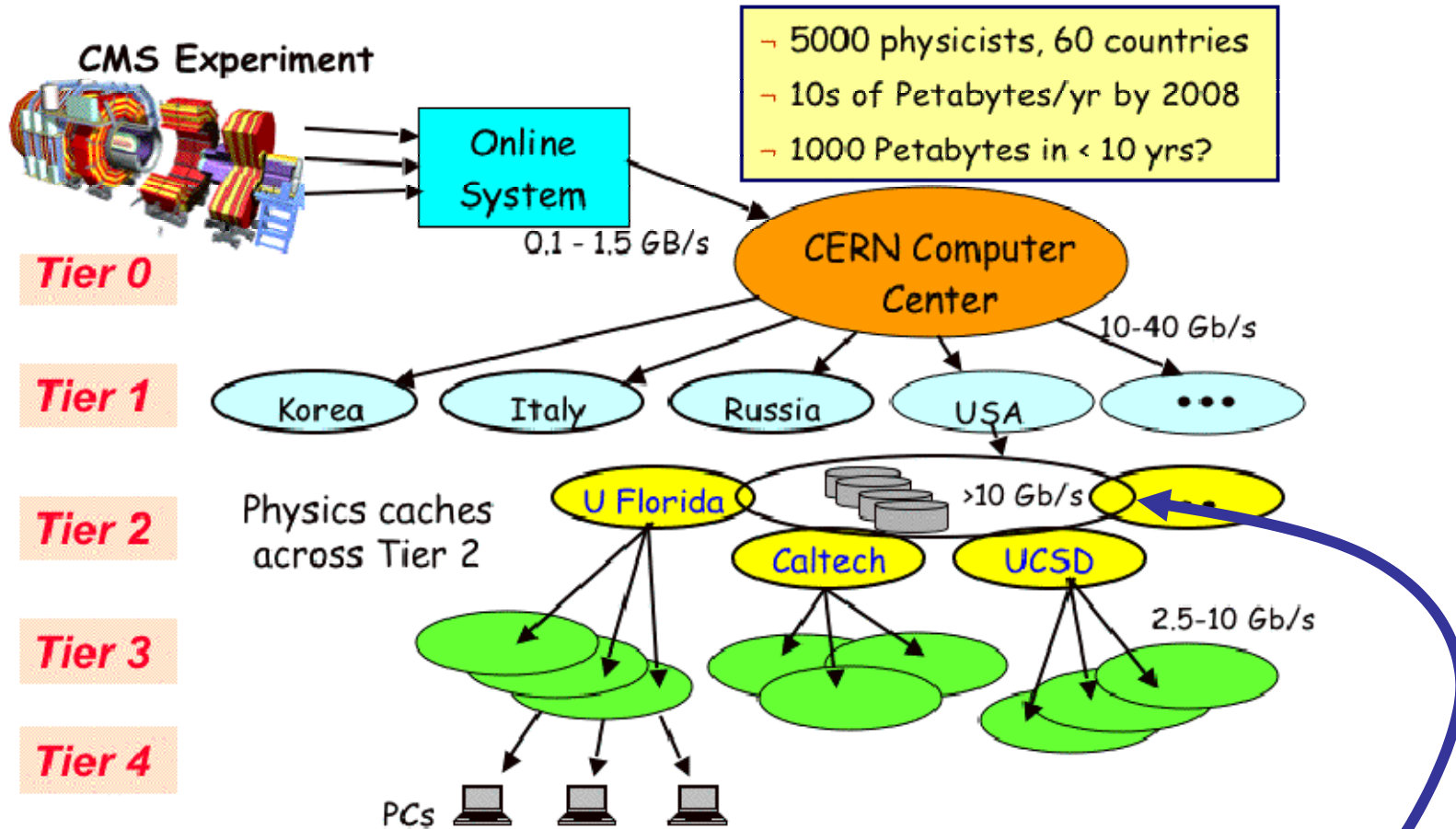
- Bring the U.S. Tier-2s up to speed :
  - Enable local data analysis (first)
  - Enable remote data analysis (second)
- Contribute to CMS Grid Integration
  - LCG integration exists
  - Provide OSG integration
  - Interface to common CMS tools

# Stitch Together Several Funded Projects

- USCMS S&C
  - Provides leadership
  - Basic U.S. Tier-2 Funding
- UltraLight
  - Focuses on network & CMS integration
- GriPhyN/iVDGL
  - Established prototype U.S. Tier-2s
- DISUN (anticipated)
  - Focuses on U.S. T2 User Analysis

Provide tactical and strategic effort to enable and sustain analysis at U.S. Tier-2s

# CMS Global Data Grid



- Advanced UltraLight network connecting T1 & T2s
  - Provides distributed data cache amongst U.S. Tier-2 centers
  - Creates a single “virtual” Tier-2 cyber-infrastructure ⇒ “DISUN”
- Exploit the network as additional Tier-2 resource

# UltraLight Considerations

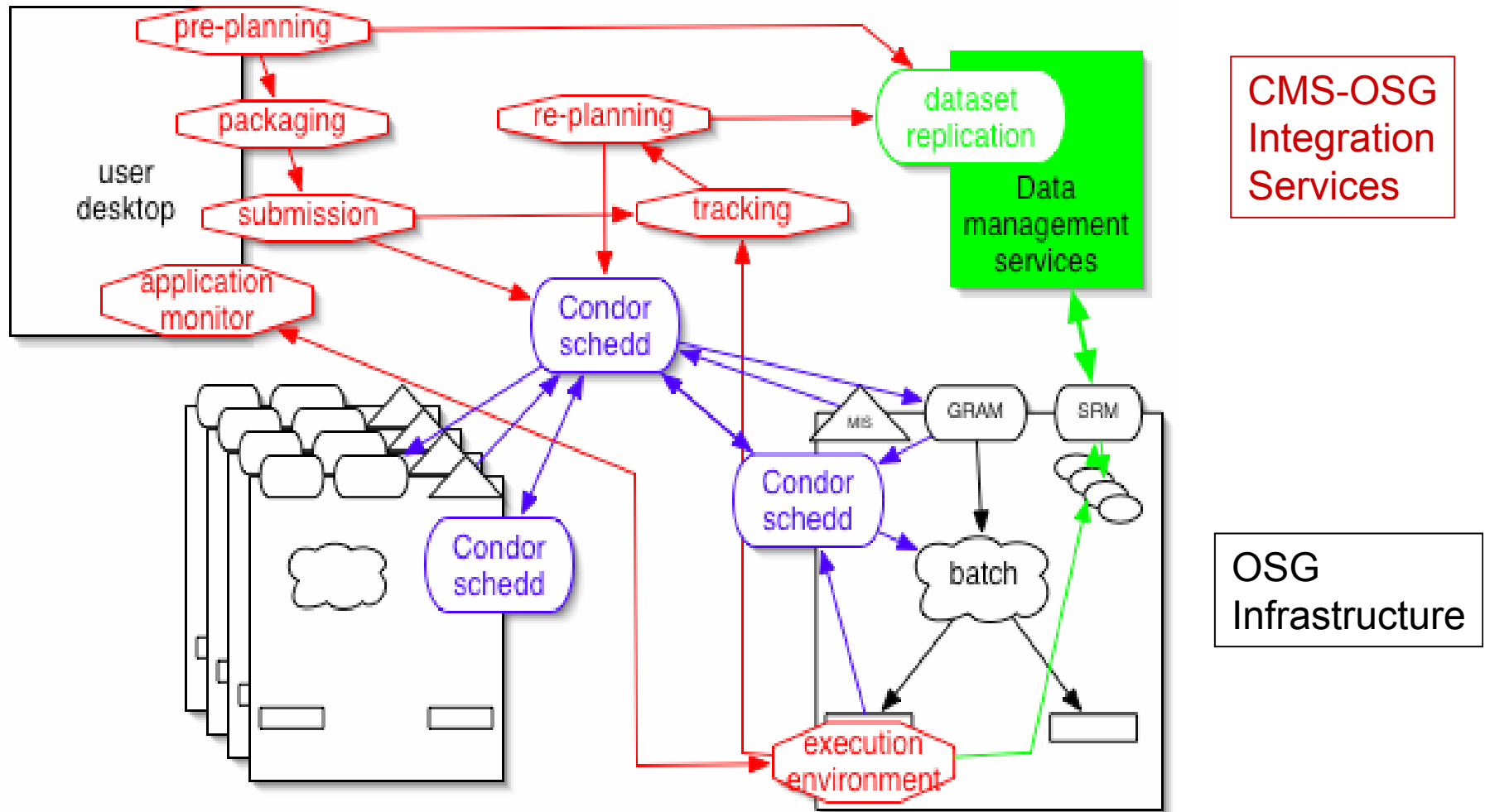
- Provides U.S. Network backbone
  - Infrastructure currently in deployment phase
  - Services currently in design phase
- Facilitates Terabyte-scale transactions
  - Movement initiated by individual users
  - Controlled via VO prioritisation/policies
- Distributes load of hosting analysis data
  - U.S. Tier-1  $\Rightarrow$  Primary storage facility
  - U.S. Tier-2  $\Rightarrow$  Analysis data cache (*a la* dCache)
- Strongly affects the U.S. T1-T2-T3 computing model

# Grid3/OSG Considerations

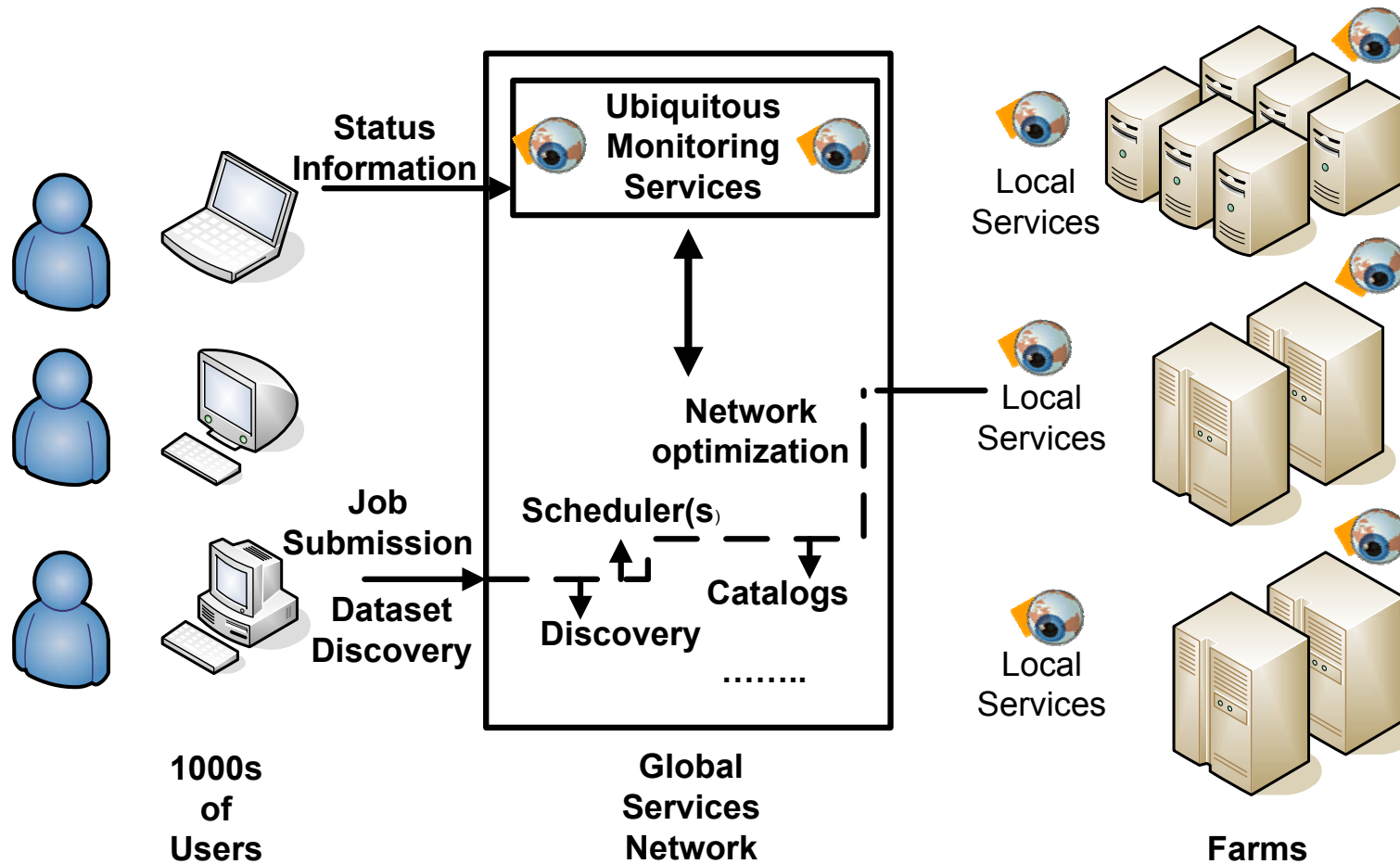
- Current Grid3 Picture
  - VOMS
  - Client side
    - Condor-G, Globus
  - CE/SE
    - GRAM, GridFTP
      - public
    - Condor/PBS WNs
      - Private (NFS)
  - No UI or RB
- Evolving OSG Picture
  - Privilege Project
    - Uses VOMS, replace gridmap file
  - Client side
    - Condor-G/C, Globus
  - CE
    - GRAM, schedd
  - SE
    - SRM
  - No official UI or RB

Grid3/OSG quite different from LCG/EGEE Picture ⇒ Requires dedicated CMS integration

# Scientific Distributed Analysis Framework (Single User)



# SDAF (Multi User)



**Resource Constraint Environment!**



# SDAF Model

- Users develop their analysis at their desktop.
- Large scale processing across the Tier-2s and on the grid “as if it were local”.
  - Read-only access to running batch environment.
  - PROOF Enabled Analysis Center
- Data distributed across Tier-1 & Tier-2s
  - Initially static, later data movement on demand

# DISUN Project

- Fully integrated with the central UAF at Fermilab
- First production quality infrastructure Summer 2005.
- Incrementally increase functionality over a 5 year period.
- Expect  $10^3$ - $10^4$  parallelization per workload
- Collaboration between:
  - Caltech
  - University of California at San Diego
  - University of Florida
  - University of Wisconsin (HEP & Condor)
- Based on success of Grid3 & CDF Analysis Facility-CAF

# Summer 2005

- Deploy PhEDEx on U.S. Tier-2s
  - Populate U.S. Tier-2s with P-TDR MC data
  - Baseline  $\Rightarrow$  local analysis
- Survey existing CMS tools
  - Understand precisely relation to Grid3/OSG/UltraLight
  - Integrate in a deliberate fashion
- Application monitor (based on Clarens and MonALISA)
- Analysis job submission on OSG
  - GRAM
  - SRM/dCache
  - Dynamic accounts via Privilege Project
  - Clarens “services gateway”

# SDAF Effort

- Working on integration of CMS tools (e.g. PhEDEx, PHYSH, CRAB, etc)
  - Integrate to work on OSG
  - Integrate to take full advantage of UltraLight
- Interested to *accept contributions* to tools that SDAF develops
  - Make generic for all of CMS (e.g. on LCG)
- GOAL:
  - Contribute to common CMS Analysis Environment
    - Try to provide interface homogeneity if possible, even if underlying computing grid infrastructure is different

# SDAF Integration Issues

- Several different CMS technologies
  - Sometimes a lot of overlap in functionality
  - Which one's to choose? (Whole tool? Part of tool?)
  - Not necessarily bad...
    - Pragmatic: allows rapid prototyping and deployment
    - Genetics: different solutions, same problem  $\Rightarrow$  evolution
- Contribute to Common CMS Analysis Environment
  - Stay flexible, choose US-Tier2 deployment baselines deliberately, carefully
  - Support CMS users early on  $\Leftrightarrow$  feedback

# Areas of Interest to PhEDEx

- Any U.S. User
  - Needs to be able to initiate data movement within the U.S.
- If necessary, special (designated) user
  - Needs to be able to initiate data movement from CERN to U.S.
- Would like to help address any :
  - Technical issues
  - Organisational issues

# User Analysis Jobs

- Examine existing CMS tools
  - CRAB, GROSS, RunJob, Physh
  - Determine how best to proceed on Grid3/OSG/UltraLight
- Settle on an interface,
  - SDAF can integrate Grid3/OSG/UltraLight services to that interface
  - Our preferred solution