PEAC: Proof Enabled Analysis Cluster

M. Ballintijn, K.Gulbrandsen, J. Letts, E. Lipeles, F. van Lingen, A. Rana, C. Steenberg, F. Wuerthwein

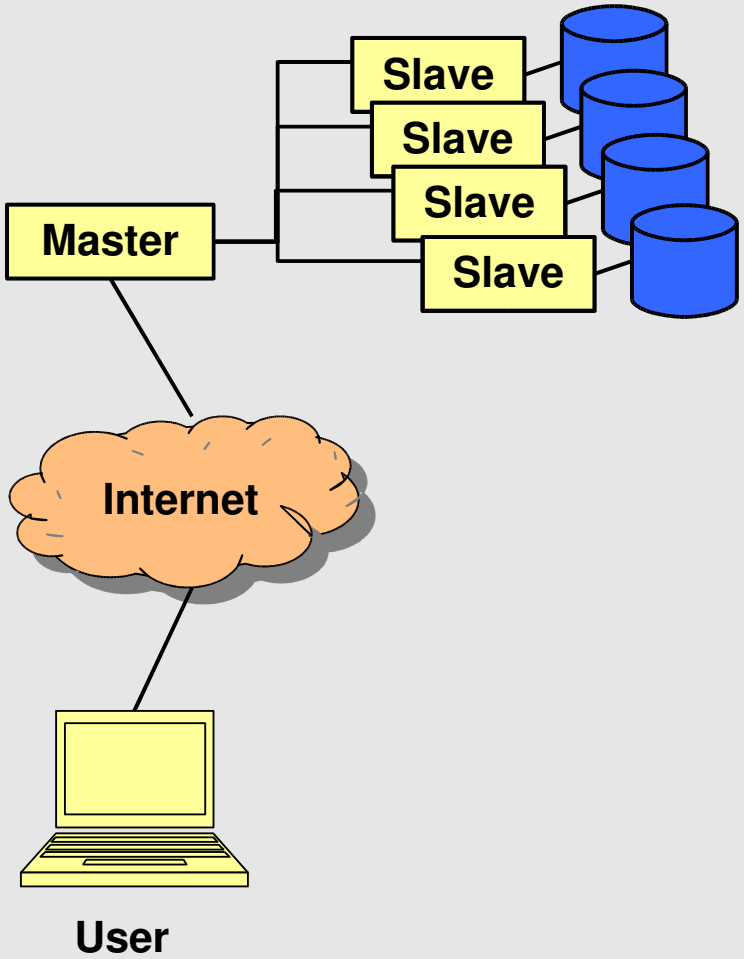
GRID computing for the human attention span

- Idea:
  - Parallel Interactive Computing
  - Resource Allocation with Data Collocation
- Technologies:
  - ROOT: common physics analysis tool
  - Proof: parallelized ROOT facility
  - Clarens: GRID enabled web services
  - Dcache: data storage and service
  - SAM: file metadata and replica catalog

Example:Computing @ CDF in 2004

Primary Data	*"skim"*	*"ntuple"*	*Physics Results*
batch	batch	interactive → PEAC ←	*Results are a product of sequence of data analysis steps with varying degrees of coordination among the 767 CDF collaborators.*
O100) TB, O(1e9) evts, O(1) GHz-sec/evt,	*O(1) TB, O(1e7) evts, O(0.1) Ghz-sec/evt,*	*O(1) GB, O(1e6) evts, O(1e-3) GHz-sec/evt,*	*Final processing requires interactive analysis.*
~30 GHz years	*~10 GHz days*	*~20 GHz min*	
1 in 3 months	*100 per day*	*100 in 10s*	*Needs outgrow Moore's law for 2004-2006.*
0.3THz @ 100% duty cycle	*~4 THz @ 100% duty cycle*	*~0.1 THz @ 1% duty cycle*	

PROOF in a Slide

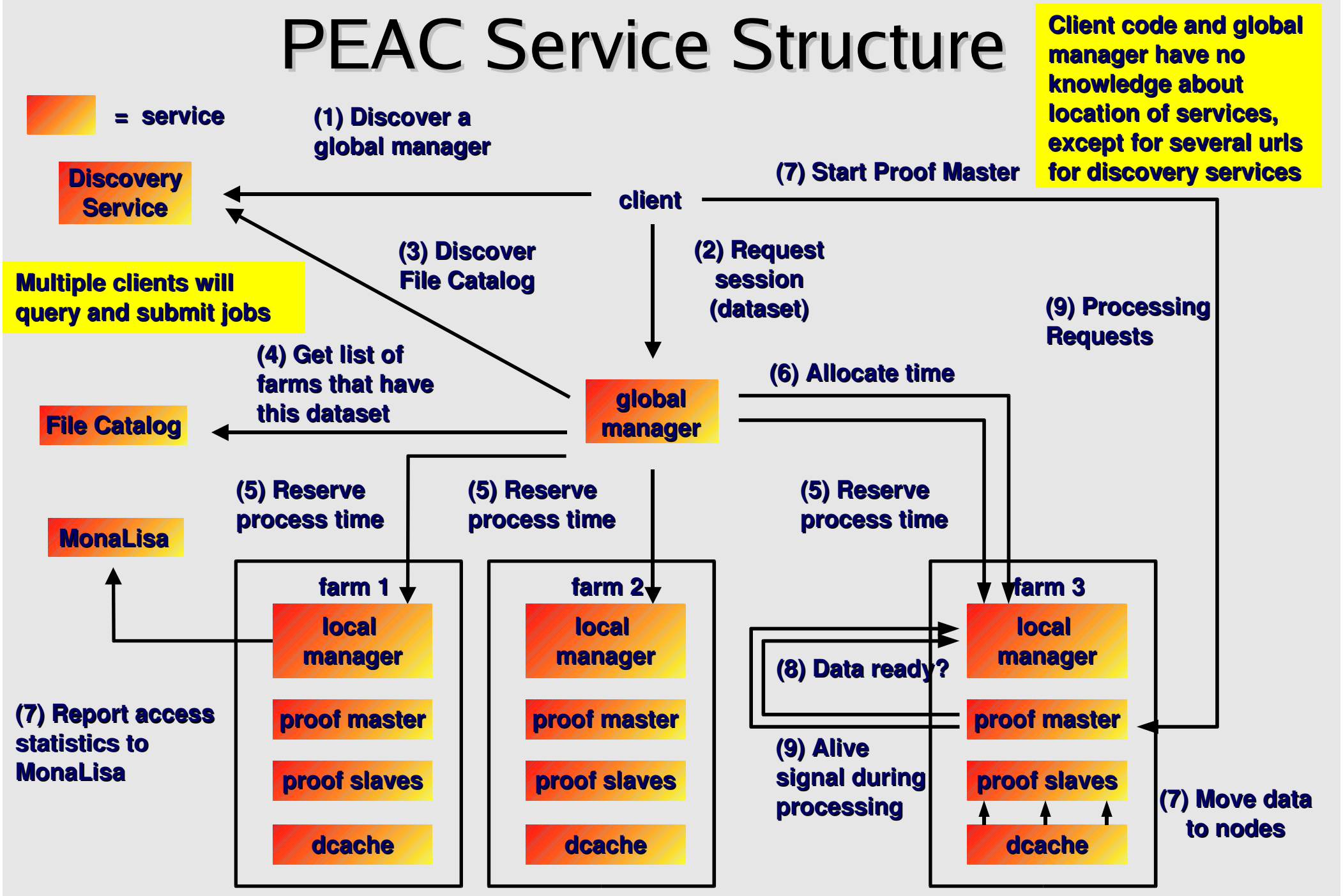


- Interactive Physics Analysis of Large Data Sets using Parallel Processing and I/O
- Transparency, Adaptability, Scalability
  - Integrated into ROOT, Known Paradigm
  - Adapt to Network and System Performance, support Heterogeneous systems
  - Efficiency and Reliability
- Multi Tier architecture, Data Locality
  - WAN Ready and GRID compatible

PROOF Features

- Selector framework
  - Transparent way to program analysis
  - Universally applicable, local, PROOF, batch
  - Extensible, combine with TTask ?
- TDSet data description
  - Lightweight, easily created from CLI
  - Versatile

PROOF Features

- Packages
  - Extend and customize environment
  - Efficient and heterogeneous
- Feedback
  - Simple progress bar
  - Flexible feedback mechanism

PEAC Service Structure