HEP-CG Workshop
27./28. April 2006
Darmstadt

dCache

A scalable storage element

Martin Radicke
Patrick Fuhrmann

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

dCache in a nutshell

involved people

Responsibility, dCache

Patrick Fuhrmann Rob Kennedy

Responsibility, SRM

Timur Perelmutov

Core Team (Desy and Fermi)

Jon Bakken
Alex Kulyavtsev
Dmitri Litvintsev
Vladimir Podstavkov
Birgit Lewendel
Neha Sharma

Patrick Fuhrmann
Michael Ernst
Tigran Mrktchyan
Martin Radicke
Mathias de Riese

External Development

Nicolo Fioretti, BARI

Abhishek Singh Rana, SDSC

Support and Help

Maarten Lithmaath, CERN

Owen Synge, RAL

dCache in a nutshell

- dCache is a LCG storage element (SE)
  - full Storage Resource Manager (SRM) support
  - variety of data access protocols
    - local area: dCap, (xRootd)
    - wide area: gsiFtp, HTTP(s)
  - information providing: GIP (LCG), JClarens (OSG)
- dCache is an Tertiary Storage-enabling SE
  - supported: OSM, TSM, Enstore, HPSS

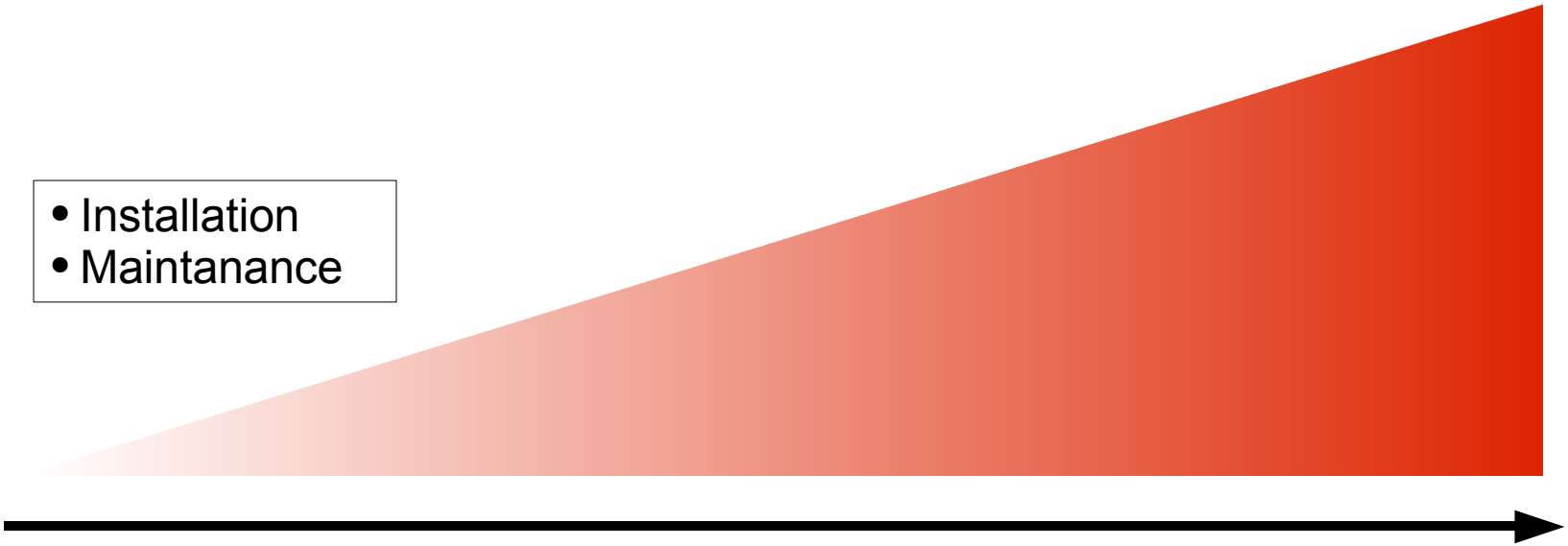
dCache: features

- combines hundreds of commodity disk servers to get a huge data store
- strictly separates between namespace and data repositiories → increased fault tolerance
- allows several copies of a single file for distributed data access
- automatic load-balancing on hotspot detection

dCache scaling

Issues related to SE scaling

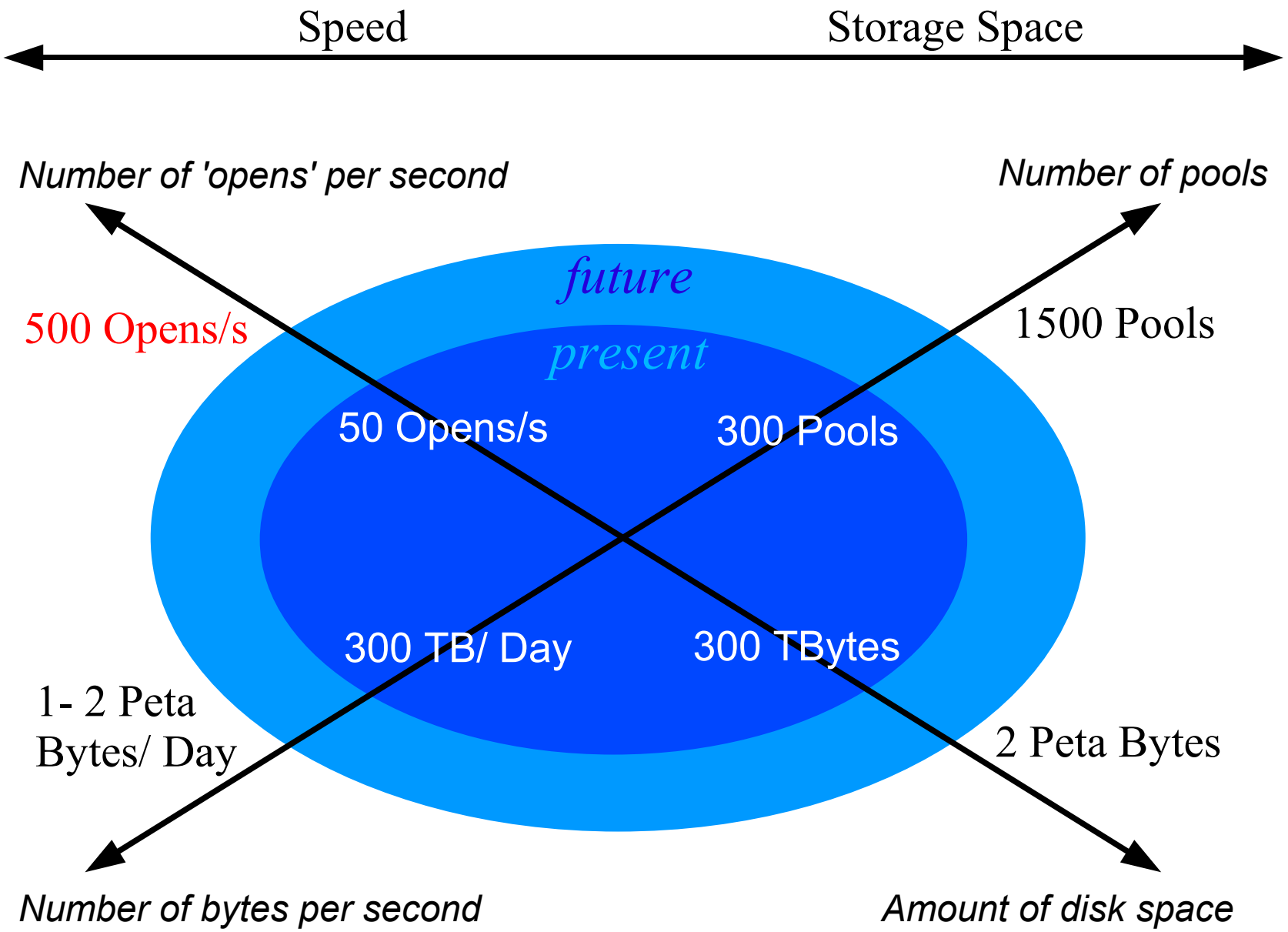
- file-system access frequency
- Tertiary Storage optimization
- SE partitioning
- Addressing hardware issues



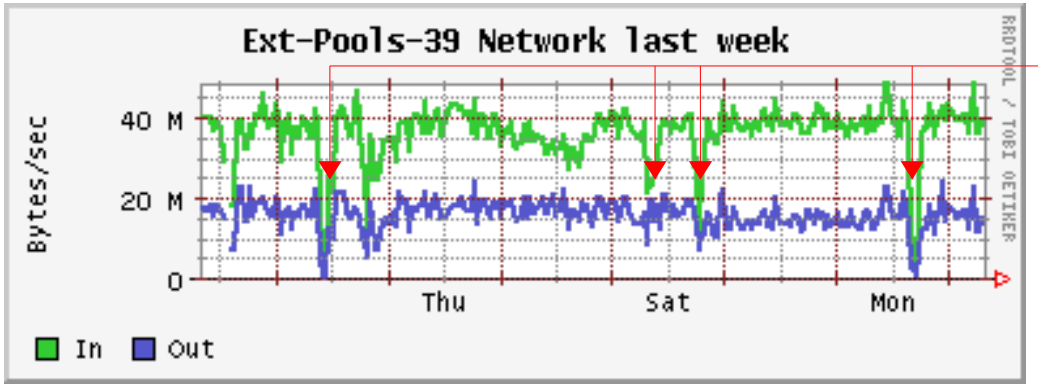
- Installation
- Maintanance

Storage Element Size

SE scaling in the near future

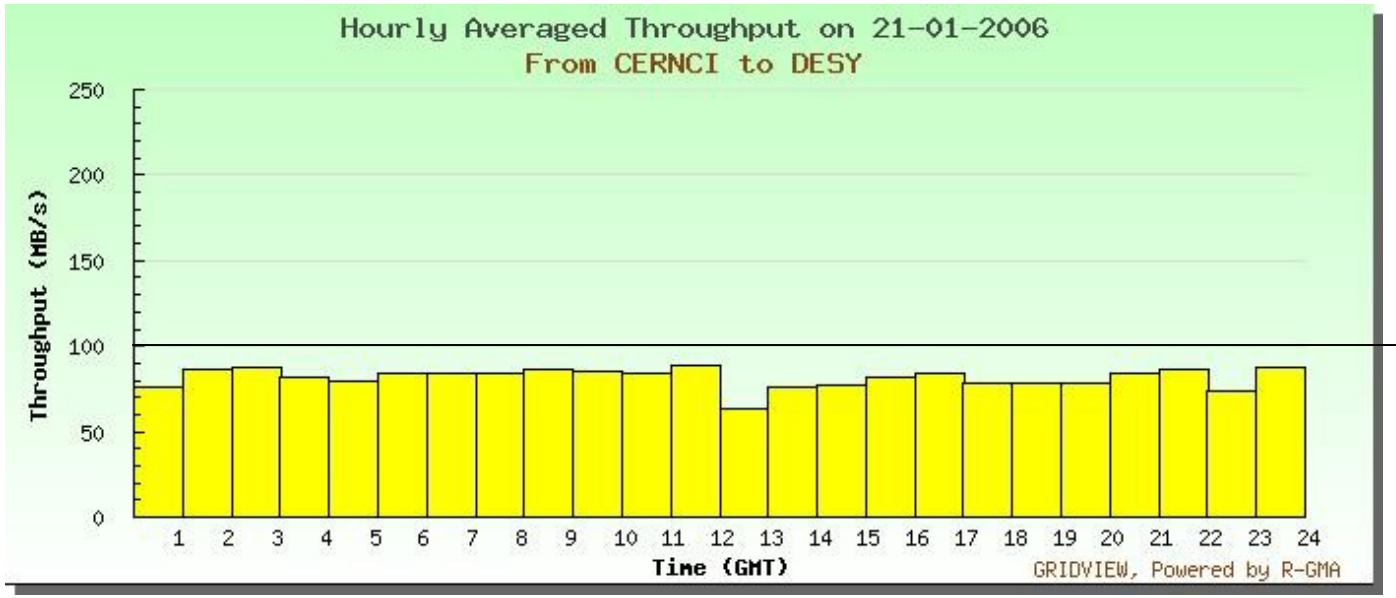


Throughput at DESY



FTS Problems at CERN

Service Challenge 3 (Jan. 06)



Limit of DESY's WAN link

dCache Storage Element improvement

Information Service

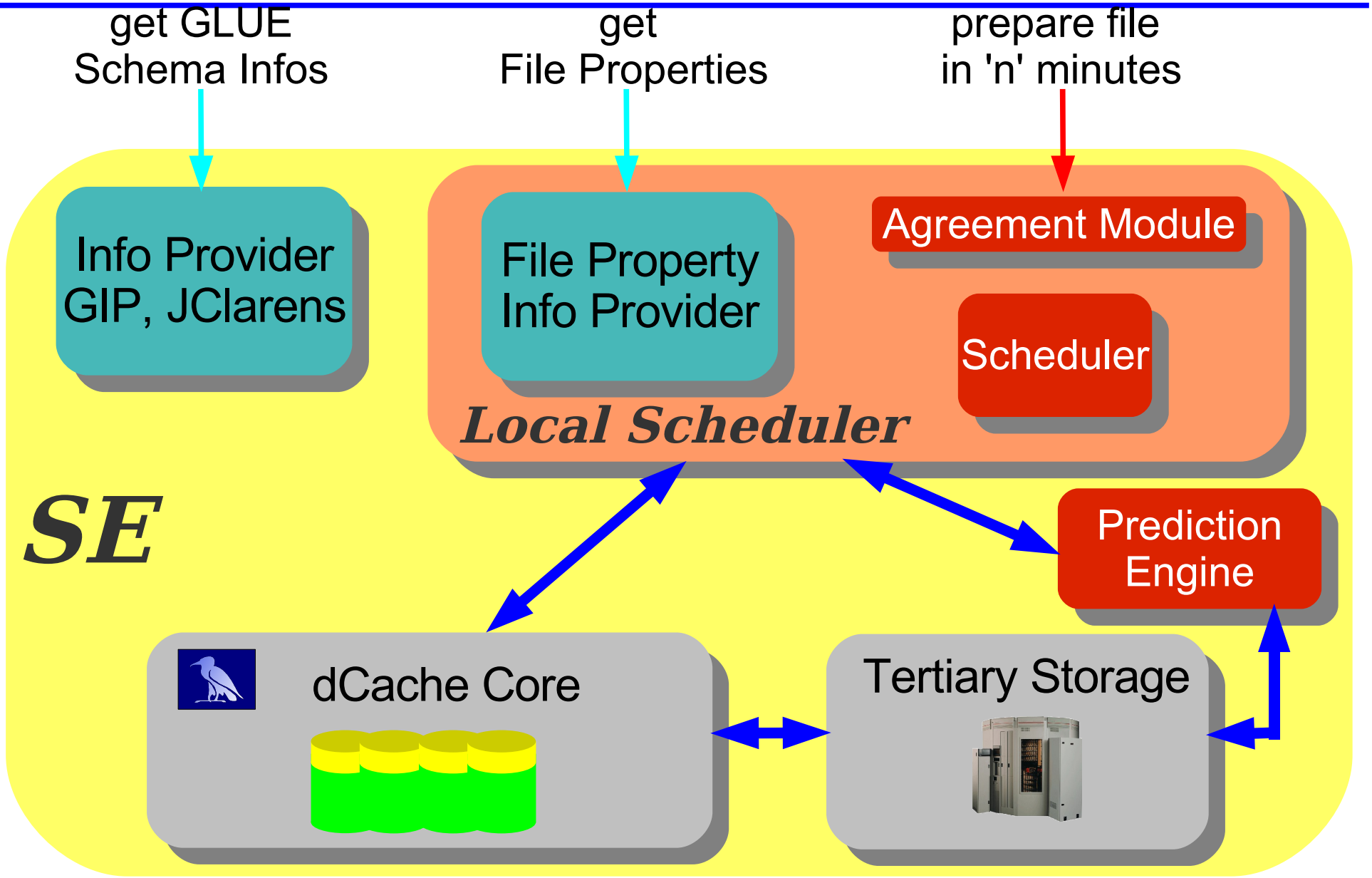
- Information Provider (SE-specific)
  - dCache gets grid-enabled by publishing its status according to the GLUE schema
  - already deployed in LCG and OSG installations
- File Property Infoprovider (file-specific)
  - Extendend Information Service, WS-based
  - file is cached/ on tape
  - time to get file ready for transfer

  → important requirement for co-scheduling

Co-scheduling of jobs and data

- More precise information provided by SE's will help the Resource Broker to do better matchmaking
  - RB is querying SE for specific file metadata using File Property Infoprovider
- more details in „HEP-CG Scheduling Architecure“, held by Lars Schely, Uni Dortmund

Extended SE overview



dCache SE deployment

dCache Packaging

- approved shell-based build system
  - requires lots of manual interaction
  - not very handy for developers
- new build system based on Apache Ant
  - will power next major release
  - XML scripts perform CVS checkout, source compilation and rpm creation
  - provides faster and more automated builds
  - automatic installation in preparation
- RPMs delivered for SL 3/4

SE Deployment

- last major version 1.6.6. (released Nov. 05) has shown stable and is in production on many LCG sites
- active support via different channels
  - Request Tracker support@dcache.org
  - User Forum user-forum@dcache.org
  - „dCache, the Book“ www.dcache.org/manuals/Book
  - direct phone support
- next major release expected in June

Sites using dCache 1.6.6

- Germany
  - LCG: gridKa (Tier 1), Aachen, DESY, Freiburg, Dortmund
  - d-Grid: Juelich (ZAM), Berlin (ZIB), Darmstadt (GSI)
- UK
  - 8 LCG Sites: Manchester, Edinburgh, Liverpool
- France
  - Lyon: IN2P3 (Tier 1)
- US
  - FERMI (Tier 1), BNL (Tier 1) CMS: 7 sites, ATLAS: 7 sites in preparation

.. and much more

D-Grid Integration Project (DGI)

- collaboration with DGI
  - providing dCache-Software as well as installation support
- mid-size dCache-installation at Research Centre Jülich
  - tape backend in use: Tivoli Storage Manager
  - becoming resource of the Core-D-Grid

Further development

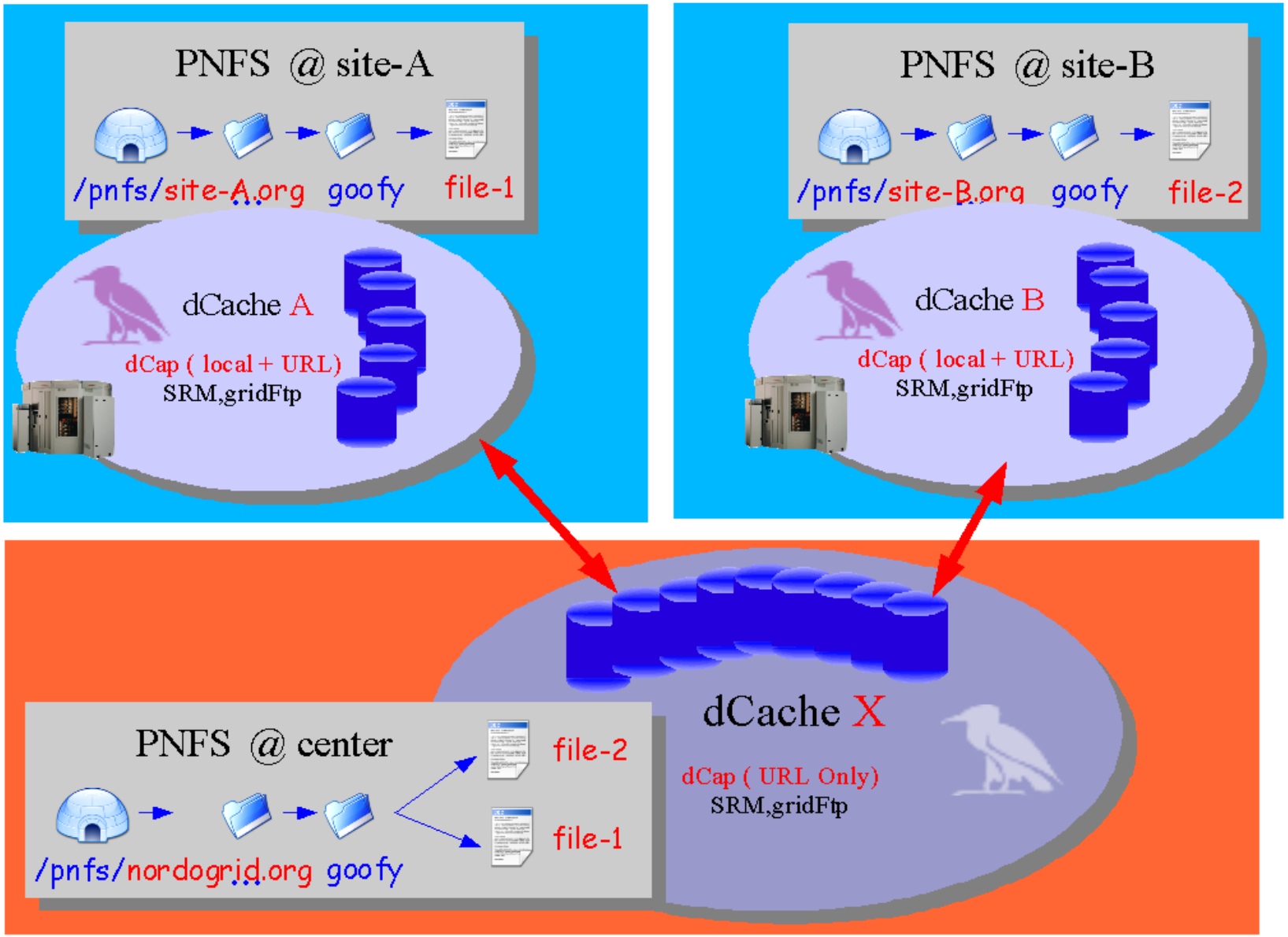
SE improvements

- Local scheduler of the dCache SE
  - Implementation of Prediction Engine (already in progress), Scheduler and Agreement Module
- File Property Infoprovider
  - Extended file metadata
    - Availability
    - Media Quality (Disk/Tape, Raid-Level,..)
    - Persistence (permanent, volatile)

SE improvements (II)

- xrootd
  - allowing ROOT-users to access dCache-files transparently
  - basic read/write mode already implemented
  - work to be done: authentication, async file transfer

Complex scenarios: Nordo-Grid



Complex scenarios: Partitioning

- dCache allows definition of partitions (groups of diskservers)
- Extended properties for partitions
  - Allowed data transfer protocols
  - Behavior on Hotspot Detection
  - Pool-2-Pool-Copy behavior
  - Tertiary Storage Connectivity
  - Impact on cost module

Contact

dCache, the Book

www.dCache.org

need specific help for your installation or help in designing your dCache instance.

support@dCache.org

dCache user forum

user-forum@dCache.org