The Work Package 2 experience

Data Management on the Grid

Peter.Kunszt@cern.ch

Outline

- Objectives and how they were met

- Achievements

- Lessons learned

- Future & Exploitation

- Questions

Objectives

DataGrid Technical Annex:

✓ Enable secure access to massive amounts of data in a universal global name space.

✓ Move and replicate data at high speed from one geographical site to another.

✓ Interface to heterogeneous mass storage management systems.

✎ Manage synchronisation of remote copi Write-only versions.

✎ Automate data caching and distribution according to dynamic usage patterns.

Network monitoring considerations.

Achievements

Delivering Middleware

- Many existing Grid components were included in the first release (like GDMP, Globus replica catalog)

- Based on first experience and on user feedback, the EDG2 services have been designed and developed

- Pioneering role in the usage of J2EE-based web services (long time before OGSI)

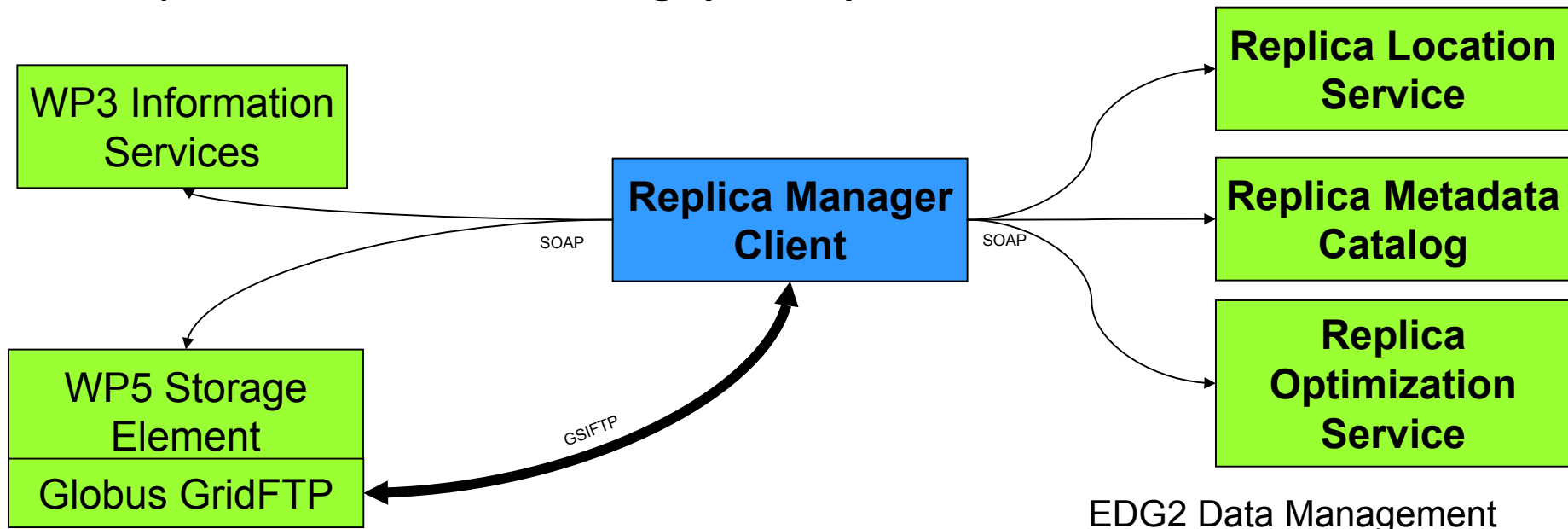- EDG2 is a complete set of data management solution – but should still be considered first generation.

Collaboration

- Close interaction with the Globus Project on replica location service

- Collaboration with CrossGrid on storage resource metrics

- Strong participation in GGF and the various groups therein

Achievements per Task

Data Replication Task

- Replica Manager (EDG1 and EDG2)

- Replica Catalog (EDG1) and Replica Location Service (EDG2)

- Grid Data Mirroring Package (EDG1)

- Replica Metadata Catalog (EDG2)

WP3 Information Services

SOAP

Replica Manager Client

SOAP

WP5 Storage Element

Globus GridFTP

GSIFTP

Replica Location Service

Replica Metadata Catalog

Replica Optimization Service

EDG2 Data Management

Achievements per Task

Optimization Task

◆ Replica Optimization Service (EDG2)

◆ Active research: OptorSim Simulation package

- Simulation of optimal replica placement strategies
- Includes simulation of network conditions

Metadata Access Task

◆ Spitfire – Grid Access to Relational DB (Demo)

- Served as a prototype web service application for EDG2 services

◆ Replica Location Service metadata

◆ Replica Metadata Catalog metadata

Achievements per Task

Security Task

- Secure java-based web services:
    - TrustManager for authentication (EDG2). In use by WP3 and WP5 as well
    - AuthorizationManager for authorization (In operation). It can apply VOMS authorization information to the service.

- Secure clients in java and c++ to web services (EDG2)

- Strong participation in EDG Security Group

Lessons Learned

Development Cycle

- In EDG it was not possible to do a proper requirements gathering, prototyping, testing, development fast enough for the lifetime of the project. A faster release cycle to the end-users will be possible from now on since future projects won't start from zero.

- We focussed on core services in EDG. The much needed end-to-end capabilities can now be added easier since the users also know better what they want and how they want it.

- User interface and documentation are important and difficult to get right first time

Less is more

- Focusing on the basics: stability and usability paid off

- Extra features good, but should be pluggable because not all users want them

Lessons Learned

Security is key

- Can't 'add security later' – horizontal through all services

- Security mechanisms are deeply reflected in the design

- Lots of open issues: Performance, delegation, site buy-in...

Web Services work well

- Modular web service structure

- Pluggable QoS (deployable in open source or commercial environments)

- Based on standards: well supported by industry and open source community

Future & Exploitation

Products

- ◆ The LHC Computing Grid is running WP2 services (except Replica Optimization and Spitfire) and will maintain and support them for at least this year for their community.

- ◆ Spitfire has served as an example for other projects already.

- ◆ The security infrastructure will serve as one of the bases for java-based web service infrastructures over SSL for the next projects.

- ◆ The optimization work has enriched the computer science community with many valuable insights through its many publications.

- ◆ The code base will be made available through GridForge.

- ◆ The publications, documentation and tutorials serve as a reference for future projects.

Future & Exploitation

People

◆ All members of WP2 have gained valuable experience while working on EDG. Their expertise will be very useful to their future projects.

Processes

◆ The lessons learned in EDG will help improve the processes of the future projects that EDG members participate in.

A lot of work remains to be done

◆ Data sets and virtual data

◆ Application metadata bindings into the low-level services

◆ End-to-end integration with user applications

◆ …

BIG THANKS

To all people who have contributed to WP2.

CERN: Diana Bosio, Akos Frohner, Leanne Guy, Wolfgang Hoschek, Javier Jaen-Martinez, Marcin Kania, Arnaud Lacroix, Erwin Laure, Levi Lucio, Ben Segal, Heinz Stockinger, Kurt Stockinger

INFN: Giuseppe Andronico, Federico Di Carlo, Andrea Domenici, Flavia Donno, Livio Salconi, Marco Serra

PPARC: William Bell, David Cameron, Gavin McCance, Paul Millar, Caitriana Nicholson

HIP/CSC: Joni Hahkala, Niklas Karlsson, Juho Karppinen, Ville Nenonen, Marko Niinimäki, Tuomas Nissi, Henri Mikkonen, Olli Serimaa, Mika Silander, John White

KDC: Olle Mulmo, Bjorn Torkelsson, Gian Luca Volpato

ITC-IRST: Paolo Busetta, Luigi Capozza, Mark Carman, Ruben Carvajal-Schiaffino, Luciano Serafini, Floriano Zini

LCG: Itzhak Ben-Akiva, James Casey, Radovan Chytracek, Kálmán Kövári, Sophie Lemaitre

PPDG: Andrew Hanushevsky, Shahzad Muzaffar, Asad Samar

And: Brian Tierney(BNL), Aleksandr Konstantinov(NorduGrid/CrossGrid)