UNIVERSITY OF PISA

ABSTRACT

STORAGE MANAGEMENT AND ACCESS IN LHC COMPUTING GRID

by Dr. Flavia Donno

Supervisors: *Prof. Gigliola Vaglini*
Dr. Andrea Domenici
Department of Computer Engineering

One of the big challenges in Grid computing is storage management and access. Several solutions exist to store data in a persistent way. In this work we describe our contribution within the Worldwide LHC Computing Grid project. Substantial samples of data produced by the High Energy Physics detectors at CERN are shipped for initial processing to specific large computing centers worldwide. Such centers are normally able to provide persistent storage for tens of Petabytes of data mostly on tapes. Special physics applications are used to refine and filter the data after spooling the required files from tape to disk. At smaller geographically dispersed centers, physicists perform the analysis of such data stored on disk-only caches. In this thesis we analyze the application requirements such as uniform storage management, quality of storage, POSIX-like file access, performance, etc. Furthermore, security, policy enforcement, monitoring, and accounting need to be addressed carefully in a Grid environment. We then make a survey of the multitude of storage products deployed in the WLCG infrastructure, both hardware and software. We outline the specific features, functionalities and diverse interfaces offered to users. We focus in particular on StoRM, a storage resource manager that we have designed and developed to provide an answer to specific user request for a fast and efficient Grid interface to available parallel file systems. We propose a model for the Storage Resource Management protocol for uniform storage management and access in the Grid. The black box testing methodology has been applied in order to verify the completeness of the specifications and validate the existent implementations. an extension for storage on the Grid. We finally describe and report on the results obtained.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I wish to express sincere appreciation to Prof. Gigliola Vaglini and Dr. Andrea Domenici for their assistance in the preparation of this manuscript and for their supervision during my PhD studies. In addition, special thanks to Prof. Mirco Mazzucato and Dr. Antonia Ghiselli from INFN (Italian National Institute of Nuclear Physics) who made possible my participation to several Grid projects. I am especially grateful to the following people who are a big source of inspiration and from whom I learned everything I know about SRM and much more: Jean-Philippe Baud and Maarten Litmaath from CERN (Switzerland), Arie Shoshani, Alex Sim and Junmin Gu from LBNL (U.S.), Timur Perelmutov from FNAL (U.S.). I can never stop admiring the code written by Jirí Mencák from RAL (UK) for the S2 interpreter: it is really a must for all students approaching object oriented and C++ programming. A big thank you to the colleagues of the StoRM project from INFN and ICTP (Luca Magnoni, Riccardo Zappi, Ezio Corso and Riccardo Murri) for the many discussions and for sharing with me the difficult steps involved in the design and development of a software product for a production Grid environment. Finally I would like to thank my bosses at CERN, Les Robertson, Ian Bird and Jamie Shiers who always encouraged and supported me in the difficult task of "making things work"!

A special thank goes to the father of my son, Heinz, who has shared with me the difficult moments and has strongly supported me discussing many technical issues, while providing Kilian with many "Unterhosenservices". A big hug and a huge "Bussi" to my little son Kilian, who allowed me to work on my PhD project during the first months of his life.

Finally, I would like to thank my mother and my father who have always been by my side, encouraging me in pursuing this career and giving me the possibility to live this wonderful and exciting experience.

GLOSSARY

AC Attribute Certificate

ACL Access Control List

AFS Andrew File System

API Application Programming Interface

ARC Advanced Resources Connector

BDII Berkeley Database Information Index

BESTMAN Berkeley Storage Manager

CE Computing Element

CA Certification Authority

CAS Community Authorization Server

CASTOR CERN Advanced STORage Manager

CERN Conseil Européen pour la Recherche Nucléaire

CLI Command Language Interface

DAP Data Access Point

DAS Direct Attached Storage

DCACHE Disk Cache

DESY Deutsches Elektronen-SYnchroton

DIT Directory Information Tree

DLI Data Location Interface

DM Data Management

DMF Data Migration Facility

DN Distinguished Name

DPM Disk Pool Manager

DRM Disk Resource Manager

DRS Data Replication Service

DST Distributed Storage Tank

EDG European Data Grid

EGEE Enabling Grid for E-SciencE

FNAL Fermi National Accelerator Laboratory

FQAN Fully Qualified Attribute Name

FTS File Transfer Service

GAA Generic Authorization and Access

GASS Global Access to Secondary Storage

GDMP Grid Data Mirroring Package

GFAL Grid File Access Library

GG Grid Gate

GIIS Grid Information Index Service

GLUE Grid Laboratory for a Uniform Environment

GOC Grid Operation Center

G-PBOX Grid Policy Box

GPFS General Parallel File System

GRAM Globus Resource Allocation and Management

GRIP Grid Resource Information Protocol

GRIS Grid Resource Information Service

GSI Grid Security Infrastructure

GSM-WG Grid Storage Management – Working Group

GSS-API Generic Security Service Application Programming Interface

GT Globus Toolkit

GUID Grid Unique IDentifier

HEP High Energy Physics

HPSS High Performance Storage System

ICTP Abdus Salam International Center for Theoretical Physics

INFN Istituto Nazionale Fisica Nucleare

IS Information Service

JDL Job Description Language

LB Logging and Bookkeeping Service

LBNL Lawrence Berkeley National Laboratory

LCAS Local Centre Authorization Service

LCG LHC Computing Grid

LCMAPS Local Credential Mapping Service

LDAP Lightweight Directory Access Protocol

LDPM LCG Disk Pool Manager

LHC Large Hadrons Collider

LFC LCG File Catalogue

LFN Logical File Name

LOA Legion Object Address

LOID Legion Object IDentifier

LSF Load Sharing Facility

LRMS Local Resource Management System

LRU Least Recently Used

MDS Monitoring and Discovery Service

MSS Mass Storage System

NAS Network Attached Storage

NDGF Nordic Data Grid Facility

NFS Network File System

OGF Open Grid Forum

OGSA Open Grid Service Architecture

OPR Object Persistent Representation

OSG Open Science Grid

PAT Policy Administration Tool

PDP Policy Decision Point

PEP Policy Enforcement Point

PKI Public Key Infrastructure

PR Policy Repository

PVFS Parallel Virtual File System

RB Resource Broker

RFIO Raw File I/O

RFTS Reliable File Transfer Service

RGMA Relational Grid Monitoring Architecture

RLS Replica Location Service

RSL Resource Specification Language

SAM Storage Access Manager

SAN Storage Area Network

SE Storage Element

SOAP Simple Object Access Protocol (original meaning)

SQL Structured Query Language

SRB Storage Resource Broker

SRM Storage Resource Manager

SSE Smart Storage Element

STORM STOrage Resource Manager

SURL Storage URL

SUT System Under Test

TDR Technical Design Report

TLS Transport Layer Security

TSM Tivoli Storage Manager

TURL Transport URL

UI User Interface

VDC Virtual Data Catalog

VDL Virtual Data Language

VDT Virtual Data Toolkit

VOMS Virtual Organization Management System

WLCG Worldwide LHC Computing Grid

WMS Workload Management System

WN Worker Node

WSDL Web Service Description Language

WSRF Web Service Resource Framework

PREFACE

Grid Computing is one of the emerging research fields in Computer Science. The Grid aims at providing an infrastructure that enables the sharing, selection, and aggregation of geographically distributed "autonomous" resources dynamically depending on their availability, capability, performance, cost, and users' quality-of-service requirements. Users belonging to a "Virtual Organization" can establish policies of usage, requirements, a working environment and even a set of virtual resources for operation. From the time of the first proposal made by Ian Foster and Carl Kesselman with the publication of the book "The Grid: Blueprint for a new computing infrastructure" and the development of the Globus Toolkit, the Grid has gone through major evolutions, attracting industry partners as well. In Europe, the projects European DataGrid (EDG), Worldwide Large Hadrons Collider Computing Grid (WLCG) and Enabling Grid for E-SciencE (EGEE) have promoted the development of Grid middleware and the creation of a worldwide computing infrastructure available for science and research.

Even though current middleware solutions are much more complete than the first prototype proposed by the Globus Toolkit, there are many areas that still need investigations and development. One of these is certainly storage management and access. The are many research challenges: applications running on the Grid need to transparently access data on the specific local storage device, exploiting a set of needed features such as space and quota management, POSIX file access, security and policy enforcement, reliable file movement, without being aware of the specific hardware/software solutions implemented at a site.

At the time of writing a complete, self-contained and coherent solution to storage management is missing in many of the existing Grid research infrastructures today: WLCG in Europe, Nordic Data Grid Facility (NDGF) in the Northern European countries, Open Science Grid (OSG) in USA, etc. One of the issues that complicate the task is the heterogeneity of storage solutions used in computing centers around the world. This work aims at providing a proposal for v2.2 Storage Resource Manager (SRM) protocol, a Grid protocol for storage systems that provides for uniform storage management capabilities and flexible file access. In particular a formal model has been designed and used to check the consistency of the specification proposed. The test modeling approach has been used to generate a test suite to validate the implementations made available for the storage services deployed in the WLCG infrastructure. The study of the black box testing methodology applied to SRM has allowed us to find

many inconsistencies in the specifications and to deeply test the behavior of the SRM systems. In order to converge toward a first real implementation of SRM in version 2 we left uncovered issues and features that will be dealt with in version 3 of the SRM protocol. The first deployment in production of SRM v2.2 based storage services is foreseen in June 2007. Among the solutions SRM v2.2 based there are CASTOR developed at CERN and Rutherford Appleton Laboratory (RAL), dCache developed at Deutsches Elektronen-Synchroton (DESY) and Fermi National Accelerator Laboratory (FNAL), LDPM developed at CERN, BeStMan developed at LBNL and StoRM developed in Italy by Istituto Nazionale di Fisica Nucleare (INFN) and International Centre of Theoretical Physics (ICTP). StoRM is a disk-based storage resource manager designed to work over native parallel filesystems. It provides for space reservation capabilities and uses native high performing POSIX I/O calls for file access. StoRM takes advantage of special features provided by the underlying filesystem like ACL support and file system block pre-allocation. Permission management functions have also been implemented. They are based on the Virtual Organization Management System (VOMS) and on the Grid Policy Box Service (G-PBox). StoRM caters for the interests of the economics and finance sectors since security is an important driving requirement.