

**600-WP-001-001**

# **RDS Prototype Grid Technologies, Access and Applications Checkpoint**

**White Paper**

**May 2004**

Prepared Under Contract NAS5-60000

## **RESPONSIBLE AUTHOR**

<u>Bruce Clark /s/</u>	<u>5/12/2004</u>
Bruce Clark, Senior Systems Engineer EOSDIS Core System Project	Date

## **RESPONSIBLE OFFICE**

<u>Don Chewning /s/</u>	<u>5/12/2004</u>
Don Chewning, RDS Project Manager EOSDIS Core System Project	Date

Raytheon Company  
Upper Marlboro, Maryland

# Abstract

---

This document defines, at a high-level, current grid terminologies and technologies, reviews current commercial and scientific perspectives on grid technologies and discusses the current and potential Remote Data Storage (RDS) project application of these technologies.

**Keywords:** Remote Data Storage, grid, SRB

This page intentionally left blank.

# Contents

---

## Abstract

## Contents

### 1. RDS Prototype Checkpoint Summary

1.1	Purpose.....	1-1
1.2	Background.....	1-1

### 2. Definitions Review

2.1	Grid Terminology .....	2-1
-----	------------------------	-----

### 3. Current Industry and Scientific / Academic Perspectives

3.1	Commercial and Industry Perspectives.....	3-1
3.1.1	Corporate Perspectives Today .....	3-1
3.1.2	Lifecycle Analysis .....	3-2
3.1.3	Commercial and Industry Perspectives Summary .....	3-3
3.2	Scientific and Academic Perspectives .....	3-3
3.2.1	Scientific and Academic Perspectives Today.....	3-3
3.2.2	Lifecycle Analysis .....	3-4
3.2.3	Multiple Grid Usage Models .....	3-4
3.2.4	Scientific and Academic Community Market Summary .....	3-5

### 4. Applicability of Grid technologies to RDS

4.1	Current RDS Storage Architecture and Application.....	4-1
-----	---	-----

## **5. RDS Collaborative Initiatives**

## **6. Conclusions**

### **Attachment A: Select Major International Grid Computing Projects**

#### **List of Figures**

Figure 3-1. Hype Cycle for Emerging Technologies, Gartner 2003.....	3-2
Figure 4-1. RDS Hardware Architecture .....	4-1

# 1. RDS Prototype Checkpoint Summary

---

## 1.1 Purpose

The purpose of this document is to define grid terminologies and technologies, review current commercial and scientific perspectives on grid technologies and discuss the current and future Remote Data Storage (RDS) project application of these technologies.

## 1.2 Background

The RDS Prototype Statement of Work (SOW) includes a requirement for a Technical Assessment Report to be provided at the end of the RDS Prototype period of performance. That RDS Prototype report will contain the RDS Prototype Assessment results along with recommendations for the RDS applicability to NASA Earth Science activities.

The RDS Prototype project scope covers the following:

1. Developing a prototype backup and restore system for the GSFC Earth Science (GES) DAAC
2. Deploying the prototype at the NASA IV&V facility in Fairmont, West Virginia and the GES DAAC
3. Procuring and integrating a firewall for the OC12 connection at the NASA IV&V facility
4. Establishing an ongoing backup activity flowing data from the GSFC Earth Science (GES) DAAC to RDS
5. Enabling grid-based connectivity to RDS storage resources
6. Assessing applicability of grid technologies RDS for future NASA implementation
7. Assessing applicability of emerging storage system architectures for future NASA research.

The following items are addressed or proposed for further clarification:

1. Definitions review relative to Grids
2. Current Industry and Scientific / Academic perspectives on Grid Technologies
3. Applicability of Grid technologies to RDS.

This page intentionally left blank.

## 2. Definitions Review

---

It is possible that there is no need to review definitions for this audience. However, throughout the many documents and correspondences, including the PID SOW for the RDS Prototype, there is at least a loose usage of terms that suggests a review of basic definitions would be useful. In addition, some of the “forward view” components envisioned by the RDS project early on are now near enough that much of the insight that has been gained should allow further clarification and refinement.

### 2.1 Grid Terminology

The word and concepts associated with the term “Grid” describe architectures, systems, processes and applications that support the concept of a loosely coupled, multi-node, multi-path system. Within an IT environment, some of the topics that are often being referred to by the word “Grid” are:

- Utility Computing
- Utility Services
- Grid Computing
- Storage Grids
- Computational Grids
- Data Grids
- Grid Technologies
- Grid Services
- Distributed Grid Resources across Wide Area Networks (WAN).

Since the true intended meaning of these phrases is important for a clear common understanding, the high level definitions for many of these concepts are provided here.

**Utility Computing.** Utility Computing is a network delivered technology, which is provisioned and billed primarily on a metered basis. In effect, a pay-as-you-go model, that is often compared to how consumers use electricity.

Utility computing generally refers to the ability of companies to access computing services, business processes, and applications from a utility-like service over a network. The idea being that if a company has to pay only for what it is using then it will be able to save money. The company offering utility computing services can benefit from economies of scale by using the same infrastructure to service multiple clients.

Utility computing is a service provisioning model in which a service provider makes computing resources and infrastructure management available to the customer as needed, and charges them for specific usage rather than a flat rate. The utility computing model seeks to maximize the efficient use of resources and/or minimize associated costs.

Another version of utility computing is carried out within an enterprise. In a shared pool utility model, an enterprise centralizes its computing resources to serve a larger number of users without unnecessary redundancy.

**Utility Services.** Utility services is a broader model in which an IT product or service exists in a virtual grid environment so that it is standardized, shared across multiple users and scalable on demand, enabling the client to buy on a pay-per-use basis. Includes Business Utility Services, IT Utility Services and Infrastructure Utility Services

Utility Services is the umbrella model and strategy – more than utility computing –, which is primarily infrastructure. Utility services also includes applications and business processes.

**Grid Computing.** Generally defined as the ability to locate, request and use computing, network, storage and application resources and services to form a “virtual IT infrastructure” for a particular function, and then release them for other uses upon completion.

Grid computing is a form of utility computing. One can think of the components of a virtual computer being organized from component (resource) pools across the Internet, then assembled on-demand for use and released back to their component pools when no longer needed.

Grid computing (or the use of a computational grid) is applying the resources of many computers in a network to a single problem at the same time - usually to a scientific or technical problem that requires a great number of computer processing cycles or access to large amounts of data.

Grid computing requires the use of software that can divide and farm out pieces of a program to as many as several thousand computers. Grid computing can be thought of as distributed and large-scale cluster computing and as a form of network-distributed parallel processing. It can be confined to the network of computer workstations within a corporation or it can be a public collaboration (in which case it is also sometimes known as a form of peer-to-peer computing).

Grid computing appears to be a promising trend for three reasons: (1) its ability to make more cost-effective use of a given amount of computer resources, (2) as a way to solve problems that can't be approached without an enormous amount of computing power, and (3) because it suggests that the resources of many computers can be cooperatively harnessed and managed as a collaboration toward a common objective. In some grid computing systems, the computers may collaborate rather than being directed by one managing computer. One likely area for the use of grid computing will be pervasive computing applications - those in which computers pervade our environment without our necessary awareness.

**Storage Grid.** A Storage Grid is a resource network primarily focused on data storage. While any storage purposed network could qualify, Storage Grids are most often thought of as high-performance, high-capacity distributed storage architectures – often with tiered performance/cost trade-offs.

**Computational Grid.** Computational Grids, as the name suggests, are resource networks primarily focused on the pooling and dispatching of distributed compute resources or CPU cycles.

**Data Grid.** The term “Data Grid” has two commonly used definitions, which often leads to miscommunication and confusion. The first definition below is the recommended usage and is the definition used throughout this document.

The most common and generally used meaning is generically another name for Grid Computing – often applied to very large-scale databases or archives (terabyte and petabyte scale). A Data Grid architecture may utilize a Storage Grid, but it is much more than a Storage Grid. Access to analyzed data in addition to raw data is also often implied.

Another somewhat less commonly used and less accepted meaning effectively equates the “Data Grid” with a “Storage Grid” or a network-level file system (NFS). For consistency with the broadest cross section of the general Grid Technologies community, the terms “Storage Grids” and “NFS” should be used in these cases – not Data Grid.

**Grid Technologies.** The broader umbrella of all Grid Computing services, including infrastructures, applications and tools.

**Grid Services.** Grid Services represent an emerging model for developing and deploying software applications that fundamentally change not only the ways companies build and deploy software but also the way they communicate with their employees, partners and customers. Grid Services not only defines “how to access and use the resources of a grid network” but also includes the notion of web services that rely on an underlying grid infrastructure.

Web services (sometimes called application services) are services (usually including some combination of programming and data, but possibly including human resources as well) that are made available from a business's Web server for Web users or other Web-connected programs. Web services range from such major services as storage management and customer relationship management (CRM) down to much more limited services such as the furnishing of a stock quote and the checking of bids for an auction item. The accelerating creation and availability of these services is a major Web trend.

Users can access some Web services through a peer-to-peer arrangement rather than by going to a central server. Some services can communicate directly with other services. This “service to service” exchange of procedures and data is generally enabled by a class of software referred to as middleware. Services previously possible only with the older standardized service known as Electronic Data Interchange (EDI) increasingly are likely to become Web services. Besides the standardization and wide availability to users and businesses of the Internet itself, Web services are also increasingly enabled by the use of the Extensible Markup Language (XML) as a means of standardizing data formats and exchanging data. XML is the foundation for the Web Services Description Language (WSDL).

As Web services proliferate, concerns include the overall demands on network bandwidth and, for any particular service, the effect on performance as demands for that service rise. A number of new products have emerged that enable software developers to create or modify existing applications that can be "published" (made known and potentially accessible) as Web services.

**Distributed Grid Resources across Wide Area Networks (WAN).** Grid technologies are not generally constrained by a need to have all the available computing and storage resources

co-located. The components that make up a Grid are often geographically distributed and connect to each other over Wide Area Network (WAN) telecommunications links. Tools have evolved in the Grid community to support the access and transfer of data between distributed Grid components including the globus alliance Globus Toolkit support for GridFTP and the SRB client support for connectivity to remote SRB storage resources.

## 3. Current Industry and Scientific / Academic Perspectives

---

There are two major communities to consider with respect to views and acceptance of Grid Technologies – each with significantly different views and levels of acceptance of these technologies. These are:

1. Commercial and Industry community
2. Academic and Scientific community

The Government aligns with each of these communities depending on the particular agency. NASA would align most often with the Academic and Scientific Community. There is increasing interest within the Commercial and Industry Community in how Grid Technologies may give them a competitive advantage. In addition, many companies have scientific, engineering and research organizations, which cause the distinction to be less pronounced.

### 3.1 Commercial and Industry Perspectives

Based on information gathered from grid Independent Software Vendor (ISV) suppliers, market analysts and industry groups, commercial organizations will initially focus on applications that are compute intensive. Such applications lend themselves more easily to a client-specific grid, which can improve performance and show a strong Return On Investment (ROI).

Grid Computing will continue to mature, as more commercial and mainstream applications are adapted into grid computing architectures. However, based on niche commercial use seen today, the next 18 to 24 months are expected to see limited large-scale implementations.

#### 3.1.1 Corporate Perspectives Today

The general corporate perspective of Grid Technologies:

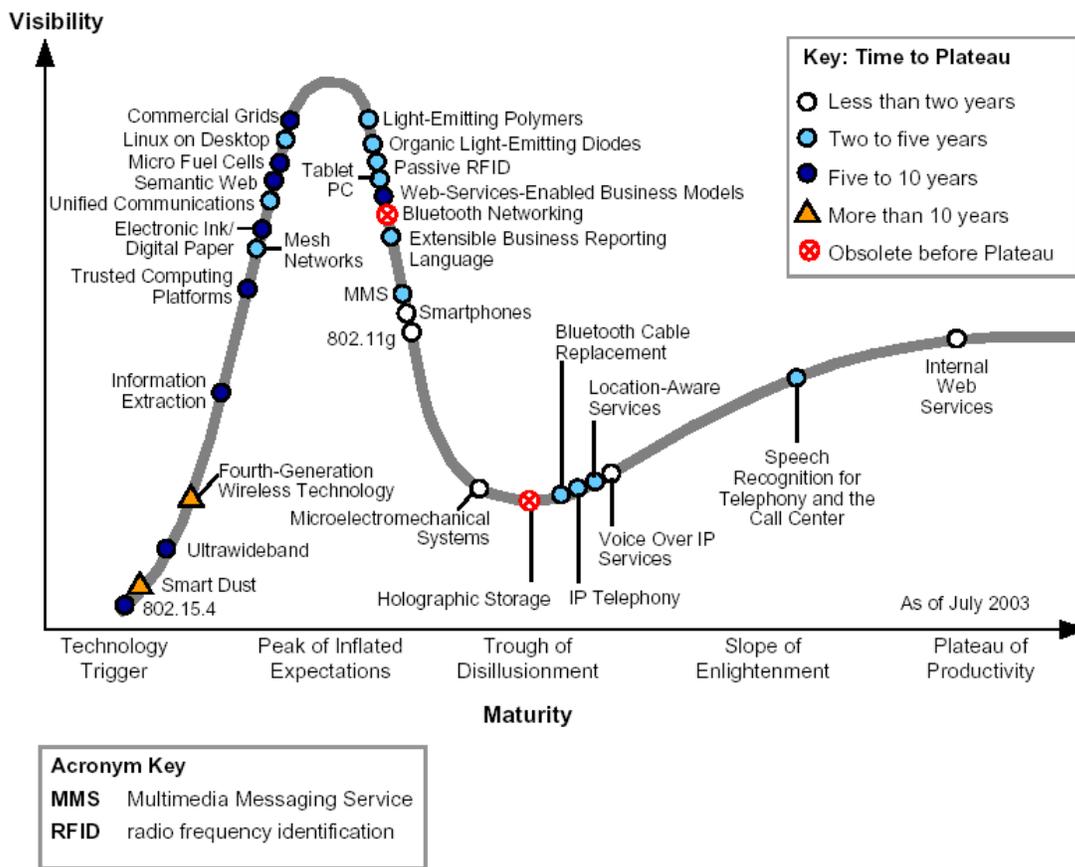
- General commercial acceptance of Grid Computing has more significant ‘political’ and emotional barriers than technical barriers.
- Corporate use of Intra-Grid services (within an Intranet) is increasing.
- Current corporate demand for Inter-Grid (multiple independent Intranets) compute and storage capacities is not sufficient to drive a corporate paradigm shift.
  - By implementing simpler Intra-Grid services today, the existing capacity within most corporate intranets provides improvements that are significant, sufficient and safe in terms of security, investment, and politics.
  - Transaction-based business does not lend itself well to Grid Computing, though recent advances in distributed database technology could eventually facilitate the implementation of transaction-based processing within tightly-coupled computing clusters. Distributed database processing executing on a computing cluster within

a data center may be possible in support of a “utility computing” or “pay per CPU cycle” processing model.

- Corporate Inter-Grid Computing will happen, but ***not*** at increased risk or cost (lower ROI).

### 3.1.2 Lifecycle Analysis

The figure below shows the Gartner 2003 “Hype Cycle for Emerging Technologies” in the commercial market space. It highlights Grid Computing at the top of the peak of inflated expectations and suggests that maturity in this space is still 5 to 10 years away. The positioning of Grid Computing on this Hype Cycle curve *does not* apply to Academic and Scientific community for which Grid Computing is a much more mature technology.



Source: Gartner Research (July 2003)

**Figure 3-1. Hype Cycle for Emerging Technologies, Gartner 2003**

There is a lot of ‘hype’ about Grid Technologies in the IT world, but there is very little independent market assessment information available. Industry Analysts such as Gartner, IDC,

Forrester, etc. are all still very quiet on all of the topics related to Grid Technologies. Early “high-level” information from Gartner includes the following projections:

- In 2006, 5 percent of large enterprise will be using Grids internally and will start with shared resources and dynamic provisioning
- By 2007, ISP's will have resource rental models where clients could rent out of a pool when overflow is needed (seasonal, events, etc) but renting on an ad-hoc basis
- By 2008, 15 percent of Fortune 100 enterprises will use that service and 50 percent of Fortune 100 enterprises will use some form of a grid internally
- By 2008/2009, the market demand for grid services and expertise is expected to expand exponentially.

### **3.1.3 Commercial and Industry Perspectives Summary**

The general technology perspective of the commercial and industry market today:

- Grid computing will not replace existing infrastructure for 5-10 years.
- Grid computing will have a long adoption curve.
- Grid computing tools are starting to have an influence on commercial software suppliers.

## **3.2 Scientific and Academic Perspectives**

### **3.2.1 Scientific and Academic Perspectives Today**

Very few people in the scientific, research or engineering community need an Industry Watchdog analysis to inform them that the market, the need and the acceptance of Grid Technologies is here today within this community. Many public Grids and many more private Grids are already in heavy use. Refer to Attachment A: “[Select Major International Grid Computing Projects](#)” for a large but partial list of Grid projects. Grid computing is already having a tremendous impact in many fields such as:

- Earth Sciences
- Life Sciences
- Space and Astronomy
- Medical Research and Drug Research
- Integrated Chip Design
- Engineering and Manufacturing

Although computer power, data storage, and communication continue to improve exponentially, computational resources are failing to keep up with what scientists demand of them. Designers of major life sciences and physics experiments are planning petabyte data archives. Scientists who create sequences of high-resolution simulations are also planning petabyte archives.

Such large data volumes demand increasingly greater analysis capabilities. Tremendous improvements in microprocessor performance means desktops and laptops are now powerful

computation engines. Nevertheless, computational (analysis) power is falling behind storage capabilities and the massive data stores in need of analysis.

Along with large-scale data analysis capabilities, grid technologies are making it possible to perform data analysis on very nearly a real-time basis allowing computation and analysis itself to be designed into data collection systems. The implications and uses of Grid Technologies goes far beyond sharing and distributing data and computing resources. For the community of scientists, researchers, engineers and academics the Grid offers new and more powerful ways of approaching problems.

*“In a future in which computing, storage, and software are no longer objects that we possess, but utilities to which we subscribe, the most successful scientific communities are likely to be those that succeed in assembling and making effective use of appropriate Grid infrastructures and thus accelerating the development and adoption of new problem solving-methods within their discipline.” — Ian Foster*

### **3.2.2 Lifecycle Analysis**

If we re-examine the placement of Grid Computing on Gartner’s Technology Hype curve in the context of the scientific / engineering / research community, we are significantly further along the maturity path. In many cases, this community would place Grid Computing in the “slope of enlightenment” phase and even in the “Plateau of Productivity” phase. This suggests that there is anywhere from a 4 to 10 year difference in the acceptance and adoption of Grid technologies between these general communities. Certainly some difficult challenges still remain, especially in the area of Grid security regarding authentication, authorization and access policies, and will need to be addressed before Grid technologies can be considered even minimally mature at the Commercial or Industry level.

As the Grid matures, standard technologies are emerging for basic Grid operations. As an example, the community-based, open-source Globus Toolkit is being adopted and applied by most major Grid projects today. With over 1000 people on its mailing lists, the Global Grid Forum (<http://www.gridforum.org>) is a significant and active force for setting standards in the scientific community.

### **3.2.3 Multiple Grid Usage Models**

Formidable challenges remain in evolving both Data and Grid Technologies into a truly ubiquitous, easily used and readily accessible resource. The broad adoption of open standards is an important prerequisite, which will allow multiple resource utilization models to eventually be available to the end user. Examples of such models are:

- User dispatches a data set in a compute request and receives an answer from the grid
- User dispatches both a data set and an application and receives an answer from the grid
- User receives analyzed results without interacting with the raw data (“In-line Computing”)

- User pulls data from a data store or data grid to process locally.

Data Mining on massive databases is an example of an application that could be run locally or dispatched to compute grids through a Utility Computing service. Many other models are utilized as required by their application, which even today is most often a custom design for the intended task.

### **3.2.4 Scientific and Academic Community Market Summary**

Grid Computing is commonly considered to be the next “big thing” in high-performance computing. Tremendous investments have been made in scientific communities and governments throughout the world. Institutions and some countries and even some states are in a high-stakes race to see which one of them can build the best grid for life-science research, such as genomic and bio-informatics research. The institutions and/or governments that can build the best infrastructure to support the needed research and science are the institutions and countries that are going to draw the best intellectual capital to solve the problems.

An example of such a grid is the Extensible Terascale Facility, which some experts anticipate will be the world’s largest and fastest grid for open scientific research. The National Science Foundation is spending \$88 million on this grid, whose computing power will be on the order of 20 trillion operations per second. The network for the terascale grid will move data at peak speeds of 40 billion bits per second (four times faster than the 10 billion bit/second Internet2 backbone!)

In another example, North Carolina may be the first state to build its own terascale-computing grid for life-science research called the BioGrid, centered at Research Triangle Park, North Carolina State University and the University of North Carolina at Chapel Hill. The BioGrid is expected to accelerate the pace of drug discovery and agricultural research resulting in a tremendous advantage to the state’s economy.

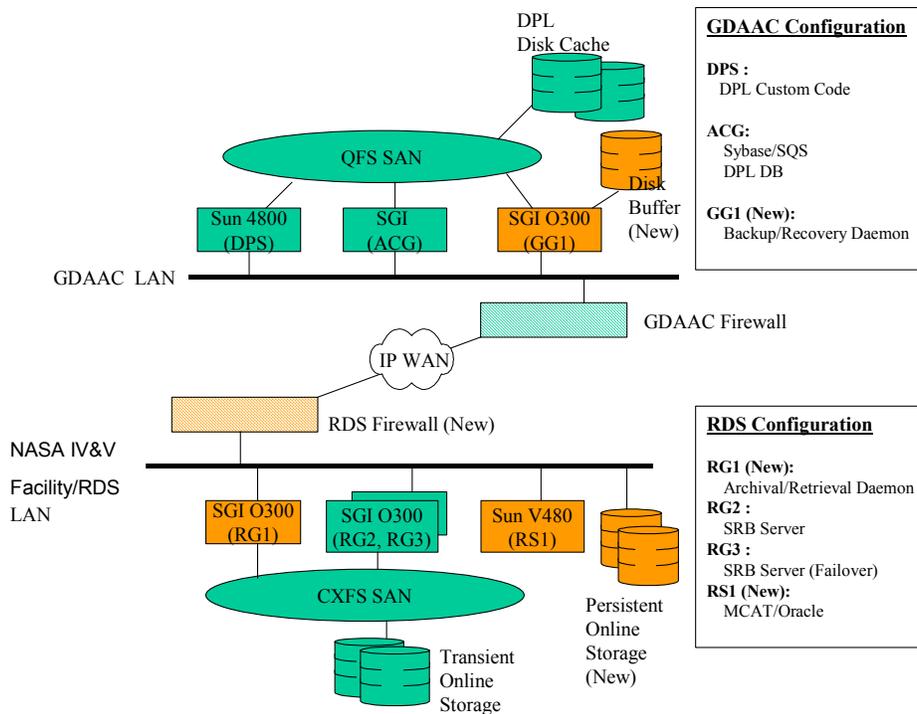
Universities that in the past few years acquired computing clusters are now moving on the next stage – Grid Computing. Grids enable different universities to pool their computing resources for solving large scientific and research problems. Many of these university and research grids are listed at the end of this document (see [Select Major International Grid Computing Projects](#)). One of the largest is a massive grid built by the Chinese Ministry of Education, which will put 100 of China’s universities onto this grid within the next two years.

This page intentionally left blank.

## 4. Applicability of Grid technologies to RDS

### 4.1 Current RDS Storage Architecture and Application

The RDS project is already implementing a prototype using a storage architecture based on the Data Grid model of grid computing (*Figure 2 – RDS Hardware Architecture*). Through the use of the Nirvana Storage's Storage Resource Broker (SRB) middleware software, RDS is able to backup, recover and share data originating from the Goddard Space Flight Center (GSFC) Earth Sciences (GES) DAAC ECS Data Pool and ECS Archive in Greenbelt, Maryland and stored in the NASA IV&V Facility in Fairmont, West Virginia. Using an application developed by the RDS project team, the selected data is staged to a SGI system in the GES DAAC and then transferred using the SRB client software through a firewall, over an OC-12 wide area network connection, and into the SRB-managed storage architecture in the IV&V Facility Data Center.



**Figure 4-1. RDS Hardware Architecture**

The storage architecture is made up of a cluster of SGI host systems running the SRB server software managing approximately 60 Terabytes (TB) of SGI TP9500 RAID disk and about 12TB

of EMC Centera Content Addressed Storage (CAS). RDS is exploring the addition of other storage technologies in ongoing prototype activities, including Advanced Technology Architecture (ATA) RAID disk and robotic tape libraries.

The data managed by SRB is accessible via the SRB client GUI, a web browser client or an API, under the control of a single-logon security environment, using multiple user profiles to grant or restrict access to the RDS data. The data appears to the RDS user(s) as a single storage resource, effectively abstracting the underlying storage hardware resources from the user's view.

## 5. RDS Collaborative Initiatives

---

The RDS project has begun to pursue a collaborative processing initiative with the Institute for Scientific Research (ISR) based in Fairmont, WV and the GES DAAC Earth Science team to explore the use of the RDS data in a Grid computing environment. The ISR is working with the GDAAC science team to create a Virtual Data Product (VDP) derived from ECS Level 0 satellite data by extracting data from the RDS storage in the IV&V data center and processing it on a Linux cluster-computing node. The ISR systems will access the data as an external SRB user over a local area network connection and process the data on their own compute hardware.

The ISR team is working jointly with RDS and GES DAAC Earth Science teams to explore the option of using the Pegasus and Chimera cluster processing software to extract data from the RDS storage architecture and processing the data into a higher level Virtual Data Product. The GES DAAC Earth Science team is working to create a more granular (“chunked”) version of the larger (6-7GB) MODIS data sets that would be used by the ISR team to execute a multi-threaded processing stream on their Linux cluster to produce the Virtual Data Product. The ISR Virtual Data Product experiment will utilize data from the GES DAAC, transferred and stored in the RDS data storage resources, extracted and processed on the ISR Linux cluster system located in the IV&V data center in Fairmont, WV in an effort to demonstrate the viability of using shared RDS data in a collaborative processing environment.

The RDS project has also initiated discussions with the NASA Advanced Supercomputing Division (NAS) located at the NASA Ames Research Center about the potential for joint Grid technology research activities. The potential for collaboration with the elements of NAS associated with the NASA Information Power Grid (IPG) project continues to be investigated by the RDS and IPG project teams.

This page intentionally left blank.

## 6. Conclusions

---

To date, the RDS project has successfully implemented a Grid-based data storage architecture that will allow a diverse set of ECS science data users to access data stored in the RDS storage resources. The RDS project continues to explore the expanding range of storage technologies to determine which of the emerging technologies might be applicable to future RDS prototype activities. The instantiation of the RDS prototype storage environment in the NASA IV&V data center provides the opportunity for NASA to explore and experiment with a variety of Grid-based storage technologies, with the goal of determining potential for application of those same technologies in future NASA Earth Science research activities.

The RDS project team will continue to work to cultivate relationships and collaborative research opportunities within the NASA Earth Science user community in order to determine how the technologies implemented in the RDS prototype can benefit and support further experimentation in the areas of Grid-based storage and Grid computing.

This page intentionally left blank.

# Attachment A: Select Major International Grid Computing Projects

---

*[Note: If you are online while viewing this, you can follow the links to any of these sites.]*

- **Globus**: This project provides a unifying framework for work on high-performance distributed computing; it includes investigations of security, resource management, communication protocols, data management mechanisms, and other issues, funded by a number of sources, in particular DOE Office of Science MICS (including its SciDAC program), the NSF PACI program, NASA IPG, IBM, and Microsoft, and with early support provided by DARPA.
- **GriPhyN** (Grid Physics Network) and PPDG (Particle Physics Data Grid): These projects funded under the NSF ITR and DOE SciDAC programs, respectively, plan to implement the first Petabyte-scale computational environments for data intensive science in the 21st century.
- **iVDGL** (International Virtual Data Grid Laboratory) is creating an international Data Grid infrastructure.
- **Earth Systems Grid**: This project funded under the DOE SciDAC program is creating technology for the collaborative and distributed analysis of environmental data.
- **GRIDS Center**: Part of NSF Middleware Initiative integrating, deploying, supporting Grid middleware.
- **NEESgrid**: A Distributed Virtual Laboratory for Advanced Earthquake Experimentation and Simulation. The NEESgrid will augment existing experimental methods used by the earthquake research community with computation approaches, which will include the development of numerical models that can predict the responses of buildings, various construction materials, or specific structural members under a variety of loadings.
- **GrADS** (Grid Application Development Software). This project is investigating fundamental issues relating to the developing of applications for heterogeneous, dynamic computing environments.
- **NCSA Alliance**. A partner in the National Computational Science Alliance, which is developing advanced infrastructure for computational science.
- NASA's Advanced Supercomputing Division Projects (**NAS**)
- **Information Power Grid**. (NASA) Aerospace sciences research support. Globus technologies are being used to build this advanced distributed computing environment.
- **ASCI FLASH**. This project aims to solve the long-standing problem of thermonuclear flashes on the surfaces of compact stars such as neutron stars and white dwarf stars, and in the interior of white dwarfs (i.e., Type Ia supernovae).

- The **DOE Science Grid** and **DISCOM Grid** that link systems at DOE laboratories
- The **TeraGrid** being constructed to link major U.S. academic sites.
- **Biomedical Informatics Research Network** (U.S. National Institutes of Health) is focused on the integration and analysis of biomedical data
- **MyGrid** (U.K. e-Science program) Workbench for bio-informatics applications
- **National Fusion Grid** (U.S. DOE Office of Science) Computational fusion research
- **National Research Grid Initiative** (Japanese Ministry of Education) Japanese national research grid project
- **National Virtual Observatory (NVO)** (U.S. National Science Foundation) Integration and analysis of astronomical data
- **GEANT** (European Union). Reaching over 3,000 research and education institutions in over 30 countries through 28 national and regional research and education networks, GÉANT provides the highest capacity (nine circuits at the core of the network operate at speeds of 10 Gbps, while eleven others run at 2.5 Gbps) and offers the greatest geographic coverage of any network of its kind in the world. It has dual roles of providing an infrastructure to support researchers, as well as providing an infrastructure for (network) research itself.
- **Dante** (European Union). Dante’s mandate is: "... to rationalize the management of otherwise fragmented, uncoordinated, expensive and inefficient transnational services and operational facilities.”
- GRIDSTART – A consortium of European Union funded projects – is an initiative sponsored by the European Commission with the specific objective of consolidating technical advances in Europe, encouraging interaction amongst similar activities both in Europe and the rest of the world and stimulating the early take-up by industry and research of Grid-enabled applications. These projects are:
  1. **AVO** which will combine astronomical databases and processing capabilities in a virtual observatory;
  2. **CROSSGRID** which will develop techniques for large-scale grid-enabled real-time simulations and visualizations;
  3. **DAMIEN** which will develop essential software supporting the GRID infrastructure;
  4. **DATAGRID** which will develop techniques supporting the processing and data-storage requirements of next generation scientific research providing intensive computation and analysis of shared large-scale databases, from hundreds of Terabytes to Petabytes, across widely distributed scientific communities (applications in high-energy physics, environmental science and bio-informatics);
  5. **DATATAG** which will develop techniques to support reliable and high-speed collaboration across widely distributed networks and to create a large-scale intercontinental test bed for data-intensive Grids;

6. **EGSO** (European Grid of Solar Observatories) which will lay the foundations of a virtual solar observatory;
  7. **EUROGRID** which will develop core GRID software components;
  8. **GRIA** which will devise business models and processes that make it feasible and cost-effective to offer and use computational services securely in an open GRID marketplace;
  9. **GRIDLAB** which will develop software able fully to exploit dynamic resources (grid technologies and applications);
  10. **GRIP** which will realize the interoperability of Globus and UNICORE, two leading software packages central to the operation of the GRID.
- The Distributed ASCI Supercomputer (DAS-2) system that links clusters at five Dutch universities
  - **Particle Physics Data Grid (PPDG)** (U.S. DOE Office of Science) Data analysis in high-energy and nuclear physics experiments
  - **Singapore BioGrid** (Singapore government) Bio-informatics research
  - **WorldGrid** (U.S. DOE, National Science Foundation, European Union) International infrastructure for data-intensive science
  - **SETI@home** A scientific experiment that uses Internet-connected computers in the Search for Extraterrestrial Intelligence (SETI). Human Genome Project uses similar approach.
  - IBM IntraGrid. (IBM) An internal grid test bed linking IBM laboratories.
  - **Global Grid Forum** Community and standards organization for Grid computing
  - **AstroGrid**. (U.K. eScience Program) Build a grid infrastructure that will allow a Virtual Observatory, unifying interfaces to astronomy databases and providing remote access as well as assimilation of data
  - **Grid Enabled Optimisation & Design Search for Engineering (GEODISE)** (U.K. eScience Program) Provide Grid-based access to a knowledge repository, optimization and search tools, industrial strength analysis codes, and distributed computing and data resources. Also see overview: [www.gridoutreach.org.uk/docs/pilots/geodise.htm](http://www.gridoutreach.org.uk/docs/pilots/geodise.htm).
  - **Grid Research Integration Development & Support (GRIDS)** Center (NSF) Integration, deployment, support of the NSF Middleware Infrastructure for research & education
  - **Southern California Earthquake Center 2** (NSF) Full geophysics modeling using Grids and knowledge-based systems
  - **UK Grid Center** (U.K. eScience Program) Support center for Grid projects within the U.K.

This page intentionally left blank.