



The Management of Distributed Data in a Grid

Khalid A. Fakeeh, PhD
King Abdullaziz University,
Jeddah, Saudi Arabia

ABSTRACT

A data grid is an architecture or assortment of services that bestow individuals/users or groups of users the capability to entrée, amend and shift tremendously an outsized amount of geographically distributed data for investigating or research reasons [24]. It bestows a data virtualization organization, making it possible to deal with appropriated records or distributed data into a canny amassing that has all the reserves of being predominantly accessible. From research point of view, this area is getting the opportunity to be continuously vital for giving/sharing sweeping data gatherings in communitarian circumstances. We depict the usage of the Storage Resource Broker (SRB), a data grid base (infrastructure), in building collaborative circumstances for considerable-scale data. We portray the projected experiences with delivering Digital Simulation Videos for the Hayden Planetarium, Distributed Data Management for the project of Visible Embryo, sorting and securing far reaching number of sky pictures for the 2MASS Digital Sky undertaking and the contribution with building the Biomedical Informatics Research Network.

Keywords

Grid Computing, Virtualization, Data Management, Storage

1. INTRODUCTION

The term Grid is brought into play to depict an infrastructure that connects diverse computational resources, for instance, people, computers, data and sensors. The declaration Data Grid now mean an arrangement of appropriated/distributed stockpiling resources, from archival systems, to saves, to databases, that are joined employing a savvy/logical name space to make around the all inclusive decided identifiers. It offer practices to the virtualization of data. Data grids make it possible to grant records that are appropriated transversely over remote stockpiling files and even controlled by particular association ranges. The distributed files appear to be, in every way, to be adjacent that can be controlled, well-organized, and controlled as if they were roosted on the local disk. The talent to bestow straightforwardness to remote file operations and to orchestrate the scattered/distributed records into an authentic logical assortment is called data virtualization. Interestingly, Storage Area Networks (SAN) give stockpiling virtualization. Its make possible to organize the stockpiling (Storage) resources utilized by distinctive PC systems into a single area that can be managed a fundamental plan of programming tools. Data grid moreover offers methods to depicting data employing metadata and annotations, and bringing into play this data about data to discover data of leisure activity. It furthermore give an expects to portray mechanisms and virtual data – data that does not stay alive and is ready or construed on leisure pursuit or demand. Data grids are getting the chance to be growing useful for directing far reaching scale circled data that are of excitement to various disciplines.

Each of these gatherings has made essentials for data offering, data generation, and data preservation and constructs such data grids to realize them. Delineations of data grids can be found in the physics areas [1][2][4][20], in field of biomedical [3] and in environmental and earth system sciences [7][23], in astronomy [8][21] and, oceanographic structures, sensors, seismology [9][22].

2. STRUCTURAL DESIGN

Data grids outline the key and imperative structural design (infrastructure) for constructing data management systems (DMS) that compass various association/administration ranges, different sorts of storage systems, and diverse sorts of data access circumstances. Data management circumstances/environments can be easily classified as appropriated data collections, automated libraries and tenacious annals. Distributed data aggregations give a lone name space to referencing data set away on distinctive stockpiling systems, customarily inside the same association or administration region. Modernized libraries coordinate remote archival stockpiling systems into a collection of data or data get-together, while giving disclosure and control organizations. Tireless records/archives reinforce the development of data aggregations onto new progressions, while sparing the ability to form, discover, and access data. Each of these systems develops the limits gave by the lower level system, and all develop data systems for directing passed on resources as mentioned below:

A) Propensity of Distributed Data Collection [17]

- Incorporate or put together data collections and associated metadata. As a significant part of any DGA, it is acknowledged that the electronic (Digital) entities inside a data gathering/collection will be depicted by qualities that depict structural, administrative, provenance/origin, and discipline-based/restricted information.
- Hold an assortment of Platforms, Resource & Data Types. The DGA framework (network) should grip distinctive computational and stockpiling/storage resources. In the framework, one should have the ability to get to reports (files) on a super computer, for instance, an IBM SP-2 or a desktop system or a convenient workstation with installed Linux or O.S.
- Faultless or unspoiled access to data and information set away inside the DGA. The data from diverse collections at sharing districts will be secured in archival stockpiling/storage systems, (for instance, DMF, HPSS, ADSM, UniTree), file systems (UNIX, NTFS, Linux), and databases (DB2, Oracle, Postgres, Sybase). Investigators at remote destinations should have the aptitude to get to these data generally as they were getting to an adjacent dataset, including sponsorship for



scrutinizing and forming (reading & writing) records.

B) Propensity of Digital Library [18]

- Leverage faultless Authentication and Authorization. An electronic or Digital library ordinarily administers propelled/digital components under an aggregation or gathering ID. To get to data from a remote record, the DGA should have the ability to manage the confirmation of a customer to the data dealing with environment, the support of the customer for access to a mechanized/digital component (entity), and the affirmation of the data dealing with structure to the remote document. The DGA should have the ability to offer access to the customer to all the stockpiling structures (storage systems) with a lone sign on check.
- Effective affiliation structure for data and information considering a modernized library framework. Notwithstanding the way that data will be secured at different destinations, it would help customers if the data are dealt with according to some sensible structure with a straightforward navigational backing. Therefore, the DGA needs to offer means to assembling data into aggregations and give management workplaces to the same.
- Grip dataset scaling in size and number. The sizes and amounts of datasets included in a DGA will keep on growing in the nearing years. Hence any response for the data grid should be adaptable to handle a colossal number of datasets, a few Terabytes and also broad records that are numerous Gigabytes in size. Support is also needed for gathering little data archives into physical pieces called holders for limit into annals, and for decreasing idleness through overhauled parallel data trade over wide area frameworks/networks.

C) An importunate faculty of Archive [15][16][19]

- For reasons of adjustment to inside disappointment (fault-tolerance), calamity recovery and load balancing, it will be functional for data to be imitated transversely over distributed resources. The uniformity of the multiplications should be stayed aware of no effort from the customers.
- In view of the fact that datasets may progress after eventually, giving passed on adjustment control will help in collaborative situated data advertising (sharing). This consolidates workplaces for locking and taking a gander at records (files).
- Leverage access control and bestow auditing qualifications. In a couple of gatherings, data need to be observed so that access to them is offered just to picked and germane people. Likewise, the determination should be done by the proprietor of the data. The DGA should have the ability to control access at various levels (collections, datasets, resources, et cetera) for customers and customer clusters past that offered by file systems. Likewise, on occasion, it may be vital to survey usage of the aggregations/datasets. Hence, analyzing workplaces will be needed as a major aspect of the structure.

Data grids is able to offer hold up to each of the above limits, fashioning it achievable to pull out contemporary data gathering (collection), Digital library, and determined archive into distributed data management circumstances/settings.

3. STORAGE RESOURCE BROKER:

A regular data management structure has been delivered at the San Diego Supercomputer Center (SDSC) that addresses the issues of various if not most of the gatherings. The advancement is known as the Storage Resource Broker or SRB in other words. The system operational at SDSC in the blink of an eye reinforces more than 50 TeraBytes of data, typifying very nearly 9 million records/files, on storage facilities/repositories at the San Diego Supercomputer Center and diverse destinations. Other SRB structures/systems moreover hand round data in the solicitation of TeraBytes world-wide. The SRB is brought into play by endeavors bolstered by the National Science Foundation, the National Aeronautics and Space Agency, the Department of Energy, the National Archives and Records Administration, the National Institute of Health, and the National Historical Publications and Records Commission. The “SRB” is a client server based middleware that gives an office to assembling building, administering, addressing, getting to, and securing data in a distributed data grid archetype [10][11]. Fundamentally, the SRB gives the going with capacities [6][10].

- i. Global constant identifiers for naming records.
- ii. Hold up for metadata to portray the region and obligation regarding files ownership.
- iii. Hold up for particular metadata to sustain disclosure through automated/digital library request/query approaches.
- iv. Standard methods of access through Web programs, Unix shell charges, Windows programs, Python scripts, Java, C library calls, Linux I/O redirection, WSDL, et cetera.
- v. Storage files/repositories pondering for chipping in with various sorts of storage structures.
- vi. Inter-space affirmation structure for secure access to remote stockpiling/storage structures.
- vii. Hold up for replication of records/files between areas.
- viii. Support for saving copies of records/files onto an adjacent stockpiling/storage structure and sponsorship for getting to archives in a narrative.
- ix. Support for amassing records into containers.
- x. Entrée/access controls and audit trails to have power over and track data usage.
- xi. Hold up for the carrying out of remote operations for data sub-setting, metadata pulling out, indexing, 3rd parties, untouchable data advancement, et cetera.



- xii. Hold up for the selection or registration of records into the structure/system, and mass data trade of archives.

The SRB has been in progress usage the length of four years. The going hand in hand with summary gives a look of the broadness of undertakings that at this time use the SRB:

- a) Astronomy: 2-Micron All Sky Survey picture amassing (2MASS), Digital Palomar Observatory Sky Survey picture get-together, Hayden Planetarium visualizations, National Virtual Observatory organizations/services.
- b) Earth-systems and ecological Sciences: HyperLTER Project data cross section, Land Data Assimilation System (LDAS), CEED: Caveat Emptor Ecological Data Repository, ROADNet ceaseless sensor data management system, SEEK: Science Environment for Ecological Knowledge.
- c) Medical Sciences: Visible Embryo Project automated library.
- d) Molecular Sciences: Joint Center for Structural Genomics data cross section, Alliance for Cell Signaling (AfCS) propelled/digital library.
- e) Neurosciences: NPACI Brain Data Archiving Project, Biomedical Informatics Research Network data grid.
- f) Physics and Chemistry: Grid Physics Network Project, GAMESS computational science data aggregations, Babar high imperativeness material science data system.
- g) Digital Libraries and Archives: National Science Digital Library (NSDL), National Archives and Records Administration model consistent report, Univ. of Michigan Digital Library Archive.
- h) Data Grids: NASA Information Power Grid, NPACI Grid Portal Project, UK eScience Data Grid.

4. RELATED WORK

Around there an assortment of experiences has been depicted in utilizing the SRB as a piece of four arranged applications. We picked these endeavors for trade not only for their broad scale data joining needs furthermore for their disparities in highlighting the capacities of the SRB structure/system. Particularly they exhibit the gigantic data dealing with capacity, data accumulation and improvement highlights, parallel data trade value, group/collaborative arranged qualities and metadata organization/management workplaces natural in the SRB. The endeavors consolidate usages of the SRB for usage in data systems for data offering, automated/digital libraries for data generation, and persistent reports (archives) for data assurance.

4.1. The Project of the Hayden Planetarium

This visualization undertaking for the Hayden Planetarium at the American Museum of Natural History (AMNH) associated 3D volume rendering of multiplication respect make a fly-through of the course of action of an emission nebula provoking the making of our adjacent solar system. The generation of the nearby solar framework advancement was performed at NCSA, AMNH and SDSC with the greatest

retributions being done at NCSA using 512 processors of a SGI Origin2000. The diversion at AMNH used 64 processors of a SGI Origin2000 and that at SDSC brought into play the Blue Horizon, a 1000 processor IBM SP2. The last diversion made more than 10,000 reports holding more than 2.5 TeraBytes of data. The data at NCSA was at first secured in the UniTree archival limit structure, and a while later traded to SDSC utilizing the SRB over a period of nine days by method for parallel data streams over the Internet2 framework. Offering the data over the different destinations obliged the usage of overall identifiers that were site self-sufficient. It has been satisfied this by enrolling the records/files into the SRB astute/logical name space. Replication of records was directed by the SRB by accomplice distinctive physical zones with each sensible record/file name. All assistants (partners) in the facilitated exertion referenced the records by their predictable names. The data at SDSC was at first orchestrated on an 800GB store system besides mostly rehashed on the IBM SP2's GPFS record structure or file system. The data was moreover repeated onto the HPSS document at SDSC. Since the data size was around 2.5TB and the store was only 800GB, the SRB was brought into play to preferably put the data at different stores, comprising the HPSS system at CalTech and disks fitting in with diverse exercises. The SRB made this data improving and data plans steady to the application since the applications see simply the logical names and not the changing physical route names of the reports/files. The rendering of the film was done at SDSC on the IBM Blue Horizon utilizing the SDSC 3D Volume Renderer. The transitional ventures in the cleaning and rendering system made more than 5TBytes of data of which 3TBytes were needed in the midst of the rendering scheme. The rendering achieved a movie run on 7 highlight projectors utilized as a piece of full-curve projections at the Hayden Planetarium. In the midst of rendering the SRB was brought into play to stir data all through the IBM GPFS from the distinctive regions where data sets were rehashed. Data needed for rendering runs were put on the GPFS and emptied when it was done with new data taking its place. In the midst of the whole process, the instinctive dialog between the diverse players at AMNH, NCSA and SDSC was uncommon to make the multiplication/simulation as close to physical reality as could sensibly be normal. The SRB was not simply utilized for data advancement/movement and data circumstance furthermore as a collaborator for passing on data pieces over the areas.

4.2. The project of the Mass Digital Sky

The 2-Micron All Sky Survey [5] conveyed pictures at the 2-micron wavelength of the entire sky, and separated the photos to describe ranges of both stars and galaxies. The 2MASS study brought into play two extraordinarily automated 1.3-m telescopes, one at Mt. Hopkins, AZ, USA and one at CTIO, Chile. The rough records from the telescopes were secured in the FITS sort out on tapes. The purpose of the Digital Sky endeavor was to load the records that were on logged off tapes into a nearby line structure/system where astronomers could get to the photos through the web. The tapes were examined at CalTech and transmitted to SDSC bringing into play the CalRen2 framework. The data was moved in two determined streams. The ingestion practice was created by personnel at IPAC, and made 'issue tolerant' by a SRB expert at SDSC. The data advancement was confined by a couple of components at distinctive times in the midst of the method:



the measure of space open at CalTech for scrutinizing tapes, store space hindrances at SDSC, obliged tape drive openness and HPSS archival limit trade issues. Despite these inadequacies and unpredictable framework disillusionments, the data trade was done effortlessly by the SRB. Before ingestion into the HPSS stockpiling structure (storage System) at SDSC, it was probable to oversee another many-sided nature. Archive, for instance, HPSS are planned to regulate unlimited reports/files. At the time, the structure held 250 Terabytes of data addressing 16 million records. The development of 5 million more records including only 10 Terabytes of data was seen as a disgraceful use of the HPSS namespace. Moreover, securing 5 million reports freely on tape would procure far reaching tape latencies. To evade these issues, the SRB uses compartments/containers to aggregate little records into a singular far reaching record [6]. It moreover brought into play the holder to re-orchestrate the data into a more supportive structure/system. The data on tapes at IPAC are secured transitorily, i.e., in the appeal in which they were taken. Since the telescopes pass the sky every day each tape contained pictures from tremendous swaths of the sky. In any case, the cosmologists generally make use of the data by domain, i.e., they look at specific areas of the sky and access objects in a region. To make the data access perfect, it has been decided to store the data in compartments/containers by the spatial locale instead of putting out of this world from the tape. This temporal-to-spatial sorting recommended that the photos from a lone tape could go into upwards of 3,000 compartments/containers. In view of the nonattendance of store space, in the midst of the later times of the ingestion, the SRB was moving containers all through HPSS to oblige filling them from the tapes. This incited some whipping, however was unavoidable by virtue of the method for the sorting that were carrying out. The whole data ingestion system took around 18 months and was essentially limited by the rate at which data could be scrutinized from the first tapes. The photos in the SRB are right away being brought into play by the cosmology bunch through electronic access: we have more than 1000 hits consistently for pictures that were involved just two years back. As a part of the National Virtual Observatory expand, the ability to reinforce considerable scale planning of the entire picture social occasion is being made. A specimen is the advancement of mosaics of pictures. No less than two social occasions need to make full-sky mosaics utilizing most of the photos as a piece of the review. A predictable/logical name space was brought into play to perceive pictures that were put on different stockpiling structures (storage systems), the advancement of pictures between limit systems was trailed by the SRB and the region information was hence upgraded, holders were used to aggregate records before limit into the record, and data access was maintained through dynamic holding of containers on disks. Facets of data grids for data bestowing amidst districts and digital libraries for data creation were utilized by the undertaking.

4.3. The Project of Evident Embryo

It was closely coupled with the Next Generation Internet Initiative and upheld by the National Library of Medicine. The explanation behind the undertaking was to display utilization of driving edge information propels in data organization/management, joint exertion, and frameworks organization to support developmental studies, clinical work and teaching in embryology. The endeavor had three focuses:

a) send high-audit workstations for facilitated exertion, a) digitize the Carnegie Collection of embryo pictures and spot it in an digital library setting utilizing the SRB, and c) exhibit the system in annotation and illustrating, embryology preparing/education and planning's for clinical management. The National Museum of Health and Medicine (AFIP/NMHM) in Maryland was responsible for digitization and data getting. SDSC maintained data stockpiling/storage, replication and volume rendering of the photos. Remedial authorities at additional destinations made embryology course material from the photo gathering. A data structure was relied upon to direct transmission of data from Maryland to California for development of a modernized digital library. Electronic access (Internet/www) was then needed by the authorities to combine the photos into 3D volumes that could be rotated and controlled in a classroom setting. The SRB was brought into play for management of data and metadata organization/management all through the endeavor. Each photo was enrolled/registered into the SRB astute/logical name space. The measure of each interesting picture was generally 15 Megabytes. Lower determination pictures were produced bringing into play the firsts to reinforce scrutinizing and natural control and were enrolled into the SRB sensible/logical name space. The lower determination pictures were made and continued forward a relentless reason into the HPSS archive at SDSC. A digital library was made at SDSC by mapping unmistakable metadata onto the sensible/logical name space. The unmistakable metadata were made as a highlight of the changing endeavors at the NMHM. The metadata were secured as archives in the SRB and in a pilot undertaking differentiated and set into an Oracle database that was under the control of the SRB. The metadata file maintained disclosure through inquiries on attributes values. A critical predictable record/file name was perceived, the physical zone of the archive was dictated by the SRB, fitting stockpiling/storage vault was gotten to, and the photo was returned. All accomplishes the photos were through the predictable/logical name space. A subsequent increase for the endeavor was to give negligible exertion organization of tape systems for long term backup [14]. A SRB driver was created that could interface with tape robots for mounting tapes, and that could issue tape drive orders. To overhaul execution, the thought of a compound resource was delivered that addresses the mix of a data cache and a tape drive. The compound resource was managed as a singular resource. Staying in contact with the compound resource realized development of a copy on the disk cache. Synchronization arranges then obliged the replication of the report/file onto tape. Requests for access to the file hence achieved development of the copy on tape back onto the disk cache.

4.4. The Project of Biomedical Informatics

The BIRN undertaking [3] is a NCRR/NIH movement strove for making a testbed for biomedical researchers to get to and inspect data at a variety of levels of gathering arranged at contrasting districts all through the country. Various data grids dilemmas, for instance, customer confirmation and looking at, data dependability, security, and data ownership are required as a highlight of the BIRN project. Both data grids for data giving/sharing, digital libraries for data dispersion, and unswerving archives for data protection were looked-for. A central rundown/catalog was made at SDSC for directing data that was reliably dealt with into an aggregation. The principal data stayed at sharing areas. Records/files were



copied into a report/archive at SDSC for a backup. The BIRN Coordinating Center, arranged at the University of California, San Diego, holds up the BIRN framework and joining over diverse activities: 1) The Mouse BIRN Project fuses practices at Duke University, UCLA, Caltech, and UCSD and 2) The Brain Morphology BIRN endeavor consolidates two investigation clusters at Harvard Medical School, one at Duke University, and two social events at UCSD. An indispensable aspect of the BIRN undertaking is to send mostly administrated custom hardware that is extensible in both breaking point and site region. A linux-based structure/system, called the BIRN RACK [12], was sent at each site to construct scattered/distributed data holds (caches). Which was a key system part anticipated that would manage access to hard controlled data resources over a wide region framework. The BIRN Racks are revamped to run the SRB data framework. The Coordinating Center at UCSD coordinates the SRB data system. Data is imitated between destinations under SRB control. The SRB furthermore holds up data access by visualization programs that are starting now under control by the diverse BIRN accessories. It may be said, the uniform SRB access parts make it possible to apply various techniques to the same data to see the relative merits. The major purpose of the BIRN undertaking is to vanquish the troubles and issues in getting to endless datasets across over destinations while meeting the strict appropriateness needed by the restorative/medical data offering or sharing regulations. A particular need of the BIRN undertaking was the ability to apply get to controls on both the photos and the illustrative and definitive metadata enrolled onto the local name space. All passageway controls were joined direct to the sensible/logical name. This suggested that when a photo was moved, the passageway controls or access controls thusly were joined at the new region. In the SRB environment, all records/files are secured under the control of the SRB data dealing with structure/system. By designating a UNIX identifier under which the SRB servers run at each storage system, all gets the opportunity to can be limited to requests issued through the SRB programming. Customers approve themselves to the SRB data dealing with system, the SRB affirms itself to the remote stockpiling structure /storage system, and a while later utilizes access controls to pick whether access is permitted. The SRB system brings into play the Generic Security Service API to correspond with the acceptance structure at the remote site, and uses Grid Security Infrastructure to hold up between inter-realm affirmations, based upon certificates of open key. Access controls on metadata were additionally puzzling. Access repression was requested such that customers would simply see explaining metadata for reports that they would be allowed to get to. All records for which they didn't have assent would not be seen on an inquiry. Access repression was similarly requested administrative metadata, such that customers would not see any qualities for picked properties. Thus the heads could use property estimations for managing the amassing that would not be seen by customers. The proposal for the SRB was that additional get to parts were required. Despite the Unix parts for make, read, and create, parts were in like manner needed for curation (metadata control) and audit trail dispatch. One of the outcomes of the BIRN undertaking was the need and change of insignificant exertion Grid Bricks for the limit of extensive data aggregations [13]. Experts need to keep data on swinging disks to minimize access latencies. With the evolvement and improvement of commodity disks and

commodity CPUs, it is instantly possible to execute disk-based storage to the cost of about \$3,000 each Terabyte, comprising framework/network access. Network Bricks are specific systems that solidify a 1.7 Ghz CPU, a Gigabyte of memory, 1 Terabyte of disk, a RAID controller, and a Gigabit-Ethernet framework/network affiliation running under the Linux working structure/system. To develop the measure of disk stockpiling (storage), additional system squares are added to the structure. The different structure pieces are supervised by the Storage Resource Broker data framework. The SRB reinforces customer confirmation, the smart/logical name space for records set away on the disks, access controls, honest to goodness relationship of the logical name space, and archive scattering over the grid bricks. This minimizes the association required for each system piece to the foundation of the Unix ID under which the SRB runs, the extension of security patches to the Linux working structure, and the substitution of failed disks. All other administrative assignments are managed by the SRB data system, making it possible to realize an extensible structure. The BIRN endeavor is in progression and a couple data fuse and collaboration attempts are a work in advancement. The BIRN is giving a testbed that is being brought into play by neuroscience masters from a couple of recuperating offices and examination centers.

5. NON PARTICULAR INFRASTRUCTURE

Over most of the endeavors, the SRB expected to administer reliability necessities on upgrade of the metadata properties that were selected onto the logical name space. Each new-fangled framework service achieved the development of new attributes that were required to depict the delayed consequences of the grid service. Tests are taking after which holder holds a record, which site holds a proliferation, which get to controls, apply, and which clear metadata can be associated with the report/file. It has been observed that an expansive bit of the operations could achieve a clashing state, and that legitimate metadata attributes must be made to manage reliability. Right when staying in contact with a holder/container, write locks are required to verify that two experts don't overwrite data. Right when staying in contact with an imitated record/file, interchange copies need to be checked as invalid, or then again the changed duplicate must be stamped as the critical copy. Synchronization approaches are then anticipated that would overhaul most of the duplicates. Right when compartments/containers are reproduced, both of these game plans of consistency goals need to be at the same time joined. The overhaul of edifying metadata attributes needs to be associated dependably over the duplicates. This speaks to a test as every duplicate has uncommon legitimate metadata attributes for territory and creation time. Likewise

some metadata characteristics for impersonations must have particular qualities. A second consistency challenge on metadata was the usage of fragile associations inside the logical name space to make it workable for a record/file to appear as a member from different collections, while keeping minimally a particular physical copy. These consistency necessities are executed inside the SRB servers, guaranteeing that all operations issued all the way through the SRB servers can be taken after for summit or completion status, and the



appropriate metadata redesigns self-dependably with the delayed consequence of the SRB maneuver.

6. CONCLUSION

SRB is brought into play to reinforce data systems, digital libraries, and enduring archives. The qualifications needed by each of these sorts of data management systems can be executed or incorporated through the practice of a logical name space for supervising overall identifiers, the mapping of definitive attributes onto the logical name space for access controls, territory, and replication, the exercise of a storage repository reflection for getting to data in a wide blend of storage structures, and the utilization of a passageway/access pondering that can be mapped onto a favored access approach.

7. REFERENCES

- [1] "TheParticlePhysicsDataGrid", (<http://www.ppdg.net/http://www.cacr.caltech.edu/ppdg/>).
- [2] "TheGridPhysics Network", (<http://www.griphyn.org/proj-desc1.0.html>).
- [3] "BIRN:BiomedicalInformaticsResearchNetwork", (<http://www.nbirn.net>).
- [4] "NetworkforEarthquakeEngineeringSimulation", (<http://www.eng.nsf.gov/news/>).
- [5] "2MicronAllSkySurvey", <http://www.ipac.caltech.edu/2mass/>.
- [6] "Rajasekar, M. Wan, & R. Moore", "MySRB & SRB - Components of a Data Grid," The 11th International Symposium on High Performance Distributed Computing (HPDC-11) Edinburgh, Scotland, July 24-26, 2002.
- [7] "The Knowledge Network for Biocomplexity", (<http://knb.ecoinformatics.org/>).
- [8] "National Virtual Observatory", (<http://www.srl.caltech.edu/nvo/>).
- [9] "ROADNet: Real-time Observatories, Applications and Data Management", (<http://roadnet.ucsd.edu/>).
- [10] "Storage Resource Broker, Version 2.0", SDSC (<http://www.npaci.edu/dice/srb>).
- [11] "Moore., & Rajasekar", "Data and Metadata Collections for Scientific Applications", High Performance Computing and Networking, Amsterdam, NL, June 2001.
- [12] "Rajasekar., & M. Wan", "SRB & SRBRack - Components of a Virtual Data Grid Architecture", Advanced Simulation Technologies Conference (ASTC02) San Diego, April 15-17, 2002.
- [13] "Rajasekar, A., Michael Wan, Reagan Moore, George Kremenek, Tom Guptil", "Data Grids, Collections, and Grid Bricks", Proceedings of the 20th IEEE Symposium on Mass Storage Systems and Eleventh Goddard Conference on Mass Storage Systems and Technologies, San Diego, April 2003.
- [14] "Wan, M., Arcot Rajasekar, Reagan Moore, Phil Andrews", "A Simple Mass Storage System for the SRB Data Grid", Proceedings of the 20th IEEE Symposium on Mass Storage Systems and Eleventh Goddard Conference on Mass Storage Systems and Technologies, San Diego, April 2003.
- [15] "Moore, R., C. Baru, A. Gupta, B. Ludaescher, R. Marciano, A. Rajasekar", (1999), "Collection-Based long-Term Preservation," GA-A23183, report to National Archives and Records Administration, June, 1999.
- [16] "Moore, R., C. Baru, A. Rajasekar, B. Ludascher, R. Marciano, M. Wan, W. Schroeder, & Gupta", (2000), "Collection-Based Persistent Digital Archives – Parts 1& 2", D-Lib Magazine, April/March 2000,.
- [17] "Moore, R.", (2001a) "Knowledge-based Grids," Proceedings of the 18th IEEE Symposium on Mass Storage Systems and Ninth Goddard Conference on Mass Storage Systems and Technologies, San Diego, April 2001.
- [18] "Moore, R.," (2001b) "Knowledge-Based Data Management for Digital Libraries", NIT2001, Beijing, China, May 2001.
- [19] "Rajasekar, A., R. Marciano, R. Moore", (1999), "Collection Based Persistent Archives," Proceedings of the 16th IEEE Symposium on Mass Storage Systems, March 1999.
- [20] "Hoschek, W., Jaen-Martinez, J., Samar, A., Stockinger, H., & Stockinger", K. (2000) "Data Management in an International Data Grid Project," IEEE/ACM International Workshop on Grid Computing Grid'2000, Bangalore, India 17-20 December 2000.
- [21] "Hammond, S.," (1999). "Prototyping an Earth System Grid", at the Workshop on Advanced Networking Infrastructure Needs in Atmospheric and Related Sciences, National Center for Atmospheric Research, Boulder CO, 03 June 1999.
- [22] "EarthScope", (<http://www.earthscope.org/>).
- [23] "SEEK: Science Environment for Ecological Knowledge", (<http://seek.ecoinformatics.org/>).
- [24] Allcock, Bill; Chervenak, Ann; Foster, Ian; et al. Data Grid tools: enabling science on big distributed data.