Information Technology Architectures for Grid Computing and its Applications

Abinaya. P[1], Pavithra. P[2], Nandhini. R[3] and Manikandan. S[4]

Abstract— **This paper presents an information technology infrastructure enabled by grid computing as a solution to todays changing business needs. The paper introduces the concepts of grid computing by highlighting its evolution, business and technological benefits. Since the Information Technology has taken the center stage in the operation of business appropriate investment is essentially required. The paper provides a survey of various organizations that have successfully employed grid computing as a tool. The study has revealed that participating organization haves been able to reduce costs by moving to an enterprise grid computing environment.**

Index Terms— Information Technology, Virtual Realization, IGT, Open Grid Services Architecture.

I. . INTRODUCTION

Over the last few years we have seen grid computing evolve from a niche technology associated with scientific and technical computing, Into a business-innovating technology that is driving increased commercial adoption. Grid deployments accelerate application related function performance , improve productivity and collaboration, and optimize the level resiliency of the IT infrastructure. By accelerating application performance, companies can more quickly deliver business results; achieving greater productivity, faster time to market, and increased customer satisfaction. Grid technology also provides the ability to store, share and analyze large volumes of data, ensuring that people have access to information at the right time, which can improve decision making, employee productivity and collaboration. Grid technology improves resource utilization and reduces costs, while maintaining a flexible infrastructure that can cope with the changing business demands, yet remain reliable, resilient and secure. At its core, grid is about virtualization, of both information and workload. In non-grid environments, existing and valid infrastructures are very much "siloed;" resources are dedicated to applications and information. Many such dedicated infrastructures exist for common applications such as HR, Payroll, etc. and for data/information mining purposes. System response is limited by server capacity – and access to the data stored. It is very difficult to dynamically respond to new requirements, as a new cross infrastructure would be required, inefficiencies would predominate, and full utilization across the many silos would be difficult to achieve.

In a grid environment, resources are virtualized to create a pool of assets. Workload is spread across servers and data can be seamlessly retrieved. By separating applications and information from the infrastructure they run on, and providing this abstract, "virtualized" view, a new level of infrastructure flexibility can be achieved. Infrastructures can now dynamically adapt to business requirements, instead of the other way around. Resources are more fully utilized, by resulting in decreased infrastructure costs, reduced processing time, increased responsiveness and faster to time-to-market. IBM has been a strong advocate and practitioner in facilitating the commercial adoption of grid computing, even before the topic became a focus of media hype and analyst attention. Over the years, IBM has made investments in all aspects of the grid domain: standards definitions, technical development, open-source contributions, deployment of the innovative and business solutions, and nurturing of a robust ecosystem that extends to software developers and business partners. We view grid as a game-changing technology that challenges basic operation assumptions of ownership, access, usage, operating efficiency, utilization of assets and total operating costs. A technology that fosters innovation and collaboration while helping customers establish a competitive advantage in their market.

With particular emphasis on the future of the grid marketplace, this paper investigates the roots and evolution of grid computing germane to customer needs for associated business solutions that span the grid spectrum. Starting with a quick view into the origins of grid computing, we proceed to analyze its value proposition with respect to innovation and perform collaboration. Furthermore, we present a holistic view of technologies associated with the grid and virtualization domain, and

Manuscript received June, 2014. Abinaya. P[1], Pavithra. P[2], Nandhini. R[3] 1,2,3 UG Student, Department of Information Technology, EGS Pillay Engineering College, Nagapattinam, Tamilnadu, India
Manikandan. S[\] Assistant Professor , Department of Information Technology, EGS Pillay Engineering College, Nagapattinam, Tamilnadu, India

ISSN: 2278 – 1323

focus upon the importance of developing and adopting industry standards. We also explore significant market trends and dynamics and investigate critical value challenges, especially those that extend beyond the compute elements of grid into data an object information management. Throughout this paper, we define IBM's vision for grid computing and position grid relative to other IBM initiatives.

II. Design Considerations and Variations

One feature of distributed grids is that they can be formed from computing resources belonging to multiple individuals or organizations (known as multiple administrative domains). This can facilitate commercial transactions, as in utility computing, or make it easier to assemble volunteer computing networks.

One disadvantage of this feature is that the computers which are actually performing the calculations might not be entirely trustworthy. The designers of the system must thus introduce measures to prevent malfunctions or malicious participants from producing false, misleading, or erroneous results, and from using the system as an attack vector. This often involves assigning work randomly to different nodes (presumably with different owners) and checking that at least two different nodes report the same answer for a given work unit. Discrepancies would identify malfunctioning and malicious nodes.

Due to the lack of central control over the hardware, there is no way to guarantee that nodes will not drop out of the network at random times. Some nodes (like laptops or dialup Internet customers) may also be available for computation but not network communications for unpredictable periods. These variations can be accommodated by assigning large work units (thus reducing the need for continuous network connectivity) and reassigning work units when a given node fails to report its results as expected. The impacts of trust and availability on performance and development difficulty can influence the choice of whether to deploy onto a dedicated computer cluster, to idle machines internal to the developing organization, or to an open external network of volunteers or contractors.

In many cases, the participating nodes must trust the central system not to abuse the access that is being granted, by interfering with the operation of other programs, mangling stored information, transmitting private data, or creating new security holes. Other systems employ measures to reduce the amount of trust "client" nodes must place in the central system such as placing applications in virtual machines.

Public systems or those crossing administrative domains (including different departments in the same organization) often result in the need to run on heterogeneous systems, using different operating systems and hardware architectures. With many languages, there is a tradeoff between investment in software development and the number of platforms that can be supported (and thus the size of the resulting network). Cross-platform languages can reduce the need to make this tradeoff, though potentially at the expense of high performance on any given node (due to run-time interpretation or lack of optimization for the particular platform).

Various middleware projects have created generic infrastructure, to allow diverse of the scientific and commercial projects to harness a particular associated grid, or for the purpose of setting up new grids. BOINC is a common one for academic projects seeking public volunteers; more are listed at the end of the article

III. CPU Scavenging

CPU-scavenging, **cycle-scavenging**, **cycle stealing**, or **shared computing** creates a "grid" from the unused resources in a network of participants (whether worldwide or internal to an organization). Typically this technique uses desktop computer instruction cycles that would otherwise be wasted at night, during lunch, or even in the scattered seconds throughout the day when the computer is waiting for user input or slow devices.Volunteer computing projects use the CPU scavenging model almost exclusively.

In practice, participating computers also donate some supporting amount of disk storage space, RAM, and network bandwidth, in addition to raw CPU power. Since nodes are apt to go "offline" from time to time, as their owners use their resources for their primary purpose, this model must be designed to handle such contingencies.

A. History

The term Grid computing originated in the early 1990s as a metaphor for making system computer power as easy to access as an electric power grid in Ian Foster and Carl Kesselmans seminal work, "The Grid: Blueprint for a new computing infrastructure".

CPU scavenging and volunteer computing were popularized beginning in 1997 by distributed.net and later in 1999 by http://en.wikipedia.org/wiki/SETI@home to harness the power of networked PCs worldwide, in order to solve CPU-intensive research problems.

The ideas of the grid (including those from distributed computing, object oriented method programming, cluster computing, web services and others) were brought together by Ian Foster, Carl Kesselman and Steve Tuecke, widely regarded as the "fathers of the grid[1]." They led the effort to create the Globus Toolkit incorporating not just computation management but also have storage management, security provisioning, data movement, monitoring and a toolkit for the developing additional services based on the same infrastructure including agreement negotiation, notification mechanisms, trigger services and information aggregation. While the Globus Toolkit remains the defacto standard for building grid solutions, a number of other tools have been built that answer some subset of services needed to create an enterprise or global grid.

IV. Challenges and Solutions

The information grid solves the problem of managing information, which may include databases, files, storage spanning across heterogeneous resources and software and hardware. The following are some common computing challenges and their solutions.

Challenge No. 1:

Accessing "heterogeneous data" stored in different formats across multiple "business" areas. The application must perform multiple I/O requests to retrieve the data that slows down the execution of the job. Programmers that build and maintain these applications must be aware of different formats and determine how to transform and join the data within their applications.

Solution:

Data-access virtualization technology across diverse data formats is instrumental in helping solve the challenge of reading data stored in different formats. Programmers can simplify access to data that are stored in mixed formats (e.g., multivendor rational databases, flat files) by enabling these data to be accessed with a single structured query language instruction. Such access also helps reduce the need to move remote files. The data's virtual view is also known as federated access to the data. That is, making the data appear as one source even though the data are distributed and stored in mixed formats. In the event that large volumes of data need to be tailored for an application, specific extraction, transformation and loading functions can be preformed on the grid. Once the data has been prepared into a proper format for an application, the data can be temporally "transported" to and cashed at a location where the processing will take place.

Challenge No. 2:

Data discovery and information delivery in a grid environment with mixed file system types. It is difficult for application developers and end-users to locate and access data because data are stored under multiple directories that are associated with each file system type.

Solution:

Poor storage-resource utilization can be resolved through the use of SAN technology. Optimal solutions would include SAN software that enables system administrators to create a virtual view of all of the SAN storage, making them appear to be one homogeneous set of files with a common name space. Also, it is necessary to move large data volumes across a network to facilitate remote processing. A software solution that addresses this challenge should enable data to be cached close to where distributed processing occurs.An ideal solution would include global naming, secure wide-area access to consistent, current data and distributed data access including POSIX/NFS interface, access control and remote-data caching

Challenge No. 3:

Mixed vendor storage is common within any given enterprise. It is costly for the level administrators to manually manage data placement across heterogeneous sets of storage devices. In many cases, bottlenecks occur in retrieving data from these devices due to congestion of over utilized devices even though there may be space available on underutilized storage resources.

Solution:

Frequently, customers have heterogeneous (multi-vendor) storage devices installed. Each vendor's storage device comes with its own management console, which makes it difficult to efficiently manage data placement across the various devices and ensure that there isn't uneven data loading. Uneven data distribution could cause some of the devices to be over utilized while others remain underutilized. This unbalanced condition could lead to bottlenecks when attempting to retrieve data, thus slowing the application processing. A virtualization portal that consolidates the view across all of the SAN devices allows a single administrator to see how data is being loaded on these devices. Administrators are then able to shift data from over utilized the devices to those that are underutilized without disrupting how applications access the data. IBM has the ability to deliver the building blocks for such solution having industry-leading storage of virtualization products such as SAN Volume Controller and SAN File System. Other system grid considerations in a SAN environment include error detection and data resiliency. It is important that data are protected and secure while providing the right data to applications.

V. Higher Level of Grid Specific Standards

A. Grid Standards

Beyond these fundamental and very general purpose web services specifications, is to another group of standards that builds on top of them, to define more functional protocols and operations (e.g. scheduling and workload management, application deployment, resource, value, provisioning, data movement, and data access). The Global Grid Forum (GGF) is involved in the definition of a number of important standards at this level. Most notable is a broad architectural focus called the Open Grid Services Architecture (OGSA) which defines a very rich vision of the execution, management and data services to enable the creation of grids. OGSA is not a single specification, but rather a set of related standards in several areas. Some important OGSA-related
specifications include:

- OGSA Basic Profile
- OGSA Security Profile

ISSN: 2278 – 1323

- Basic Execution Services (OGSA-BES)
- Job Submission Description Language (JSDL)
- Data Access and Integration Services (DAIS)
- Configuration Description, Deployment, and Lifecycle Management (CDDLM)
- OGSA Byte I/O (ByteIO)

B. *Information Model:*

While several existing standards have embedded within them some form of resource representation, it has become apparent that a truly common framework for an abstract and information model is needed. To this end, the Global Grid Forum (GGF) and other standards bodies have increasingly turned to the Distributed Management Task Force (DMTF) Common Information Model (CIM) as a touchstone. CIM has been developed over a number of years to describe all kind of IT resources, from very high level conceptual capabilities to very specific low level components. While the GGF and OGSA working groups have not yet formally that identified DMTF CIM as the information model that they will use for grid computing, they are working towards that direction.

C. *Management Standards*

Along with the push to develop a common modeling approach for web services based standards, there has been a strong motivation to establish common high level protocols and operations for managing resources that are exposed as web services. OASIS' Web Services Distributed Management (WS-DM) is an industry-wide standard for management both using web services and managing web services. WS-DM attempts to exploit web services technology to create a universal and consistent abstraction for management and manageability interfaces that leverage key features of web services protocols. The specific types of management capabilities exposed by WS-DM include:

- Monitoring the quality of a service associate with a service
- Enforcing a service level agreement (quality of service)
- Querying or controlling the basic operational state of a resource
- Managing a resources lifecycle (create / destroy)

As in the case of the foundational web services standards, there is an ongoing effort to develop a "convergence" plan for overlapping management standards such as WS-DM and WS Management. In summary, there is a growing collection of related standards and architecture being developed in open standards bodies like IETF, W3C, OASIS, DMTF and GGF that are all based on web services and can be composed to help develop interoperable grid middleware and infrastructure. IBM has been a leader in the definition and development of these open standards and is driving its important implementations along this standards roadmap.

VI. CONCLUSION

The foresaid study has revealed that participating organizations have been able to reduce costs by moving to an enterprise grid computing environment. The above discussion has given the foresight into how grid computing is accomplished in finding the ideal solutions in modern business enterprises and beyond by taking advantage of low cost servers and low cost operating systems. Organizations have been able to drive more efficient scaling and better over using SMP servers. Their price/performance ability to add hardware incrementally, as needed, for increased capacity and performance in a mixed environment has allowed for overall reduced hardware costs, as well as the ability to protect their IT investment.

REFERENCES

[1] Antony Davies "Computational Intermediation and the Evolution of Computation as a Commodity, "http:/business duq. Edue/faculty davies/ research/Economics of computation/pdf)" (pdf) applied Economics June,2004.

[2] Ian Foste and Cari Kesselman, "The Grid: Blueprint for a New Computing Infrastructure " (http:/www.mkp.com/grids/)

[3] Plaszezak, Pawel and Rich Wellner, Jr. Grid Computing "The Savvy Manager's Guide" (http//savvygrid.com/). Morgan Kaufmann Publisher. ISBN 0-12-742503-9.

[5] Fran Berman, Geoffrey Fox and Tony Hey, "Grid Computing: Making the global infrastructure a reality", Published by Wiley.

[6] Luis Ferreira and Non Keung, " Grid Computing Products and Services" by IBM Redbooks Publications, ISBN 0738491780

[7] Rich Wolski and Todd Bryan, "Grid Resource Allocation and Control Using Computational Economies", University of Tennessee, Wheaton College

