24

GLOSSARY OF TERMS

For the reader, the following glossary is provided for terms used throughout this book. The vocabulary words range from business-specific, financial/accounting terms, information technology in origin, grid computing jargon, and other terms introduced in this book that may or may not make their way into the grid technology jargon mainstream.

API See definition for *Application Programmatic Interface*.

Application Programmatic Interface Application Programmatic Interface (API) is a standard programming language (e.g., C++, Java) that allows access to the information from the system with which it is communicating. "Standard" in this context refers only to the product or system.

C# This is an object-oriented language used in a Microsoft.Net environment for building applications and systems.

compute grid This is a grid that enables sharing, management, and distribution of tasks based on configurable service-level policies. It provides the core resource and task management services for grid computing.

compute utility This provides for customer resource consolidation and location independence, extending their capability to optimize the center for their business needs and for a particular pattern of workload, thus making the data center actually platform-architecture-transparent and providing a much more simple and convenient "high-level" virtual architecture (see Ref. 35, p. 2).

Distributed Data Management for Grid Computing, by Michael Di Stefano

data affinity This is a key data management feature or objective of any data grid that describes data locality or "physical closeness" of data in the grid to compute nodes where a running task is accessing that data.

data atom As described in this book, the data atom is the smallest element of data for which a data set can be broken down. The data grid will apply the management policies of synchronization, distribution, and replication to the data atoms of a data region.

data grid This provides the data management functions that are required for data access, synchronization, and distribution of a grid.

data marshaling A process of packaging the data in formats that other applications understand. Often, the format of the marshaled data is common and machine-independent.

data region See the definition of *regionalization*.

distributed file system In a client/server environment, this is a collection of files, physically distributed across any number of machines on a network that are logically structured into a hierarchical organization by one or more coordinating servers. Clients of the logical file system access it via the distributed file system servers.

distribution policy The data distribution policy describes the distribution pattern of data within a data grid. The scope of the data distribution policy can range from a subset of data atoms within a data region to the distribution of data regions within the entire data grid space. The finest data granularity point on which a data distribution policy can operate is a data atom.

document type definition Document type definition (DTD) is a schema specification method for SGML and XML documents where the definition for the data is represented.

DTD See definition for *document type definition*.

EAI See definition for *enterprise application integration*.

EII See definition for *enterprise information integration*.

enterprise application integration Enterprise application integration (EAI) is a best-practices method of architecture and implementation using technology tools (typically termed *middleware*) enabling the integration of applications across an enterprise (data integration). It integrates information at the application level.

enterprise information integration Enterprise information integration (EII) is a best-practices method of architecture and implementation using technology tools enabling the integration of information at the business level across an enterprise.

event notification policy The event notification policy describes events and how they are to be managed with the data grid. Subscribers of events can range from other data grid management policies (e.g., synchronization policies) to external user programs. The event itself indicates that a data atom has changed state and an action must be taken.

eXtensible Markup Language eXtensible Markup Language (XML) consists of self-describing data with customized user tags for data definition, data processing, and data parsing among applications, systems, and any organizations. The programs written to process XML data structures normally obtain the data structure from the document type definition (DTD).

eXtensible Style Language eXtensible Style Language (XSL) is a language used to create a style sheet for specifying the style of an XML document.

Globus Project This is an organization that conducts research to create fundamental technologies for grid computing and offers a toolkit called the *Globus Toolkit* (more information can be found on the Web site www.globus.org).

grid computing Grid computing is any distributed cluster of compute resources that provide an environment for the sharing and managing of the resource for the distribution of tasks based on configurable service-level policies.

GridFTP This is an attempt by the Globus Project to establish a universal data transfer protocol for grid computing through the use of a common File Transfer Protocol (FTP).

high availability High availability (HA) is the ability of a resource or service to withstand failure, typically through resource duplication in a hot-standby, monitoring, failure detection, and finally automated failover to the hot-standby resource in such a fashion that the user of the resource detects only a reduction in resource response time during the failover process.

Java Database Connection Java Database Connection (JDBC) is a set of Java APIs that allow access to any database that supports SQL. The API executes the appropriate SQL command to perform the respective operation.

JDBC See the definition for *Java Database Connection*.

LAN See the definition for *local-area network*.

level 0 data grid These data grids are optimal for data sets that are static in nature.

level 1 data grid These data grids are optimal for data sets that are dynamic in nature.

local-area network A local-area network (LAN) consists of local computers networked together and confined to a limited geographic space such as a floor or a building.

Monte Carlo simulation This is a mathematical method that uses statistical techniques (e.g., randomness) to model complex systems in a variety of disciplines such as physics, biopharmaceuticals, and finance.

object–relational database management system Object–relational database management (ORDBMS) is a data management system that enhances the relational data model by supporting arrays, inheritance, and functions that represent some of the basic concepts of object-oriented programming.

OLAP Online analytical processing (OLAP) is the collection, management, process, and reporting of multidimensional data.

ORDBMS See definition of *object–relational database management system*.

Parallelize This is a unit of work that can be subdivided into smaller atomic subunits of work called "worklets" in such a way that each subunit of work can be run in parallel across physically dispersed computers (compute grid).

PostgreSQL PostgreSQL is an object–relational database management system (ORDBMS).

QoS See definition of *quality of service*.

quality of service Quality of service (QoS) is the level of service as defined by the business unit that the grid architecture needs to meet.

regionalization Also referred to as a *data region*, this is the logical grouping of data atoms within the data grid space. A simple analogy would be a data region in a distributed data management system or a database in a relational data management system.

replication policy Data replication policy describes exactly how the data atoms are to be replicated within a data grid.

SGML See definition for *Standard Generalized Markup Language*.

SQL See definition for *Structured Query Language*.

SQL3 See the definition for *SQL99*.

SQL99 Also referred to as SQL3, this is an ANSI/ISO standard that replaces SQL92 addressing advanced topics such as object–relational database concepts, call level interfaces, and integrity management not found in SQL92.

Standard Generalized Markup Language Standard Generalized Markup Language (SGML) is a standard metalanguage, a description of how to specify a document markup language or tag set; for example, XML is a SGML-based language.

Standard Template Library The Standard Template Library (STL) is a C++ library of container classes, algorithms, and iterates.

STL See definition for *Standard Template Library*.

STP See definition for *straight-through processing*.

straight-through processing Straight-through processing (STP) is a best-practices method of architecture and implementation using technology tools (typically termed *middleware*) that automates end-to-end processing of transactions from a business perspective.

Structured Query Language Structured Query Language (SQL) is a data management and query language for databases. There is no standard SQL today, but there are many extensions to the ANSI-SQL.

synchronization policy The data synchronization policy describes how the data atoms within a data region are to synchronize with each other. Types of synchronization are optimistic (showing complete trust in the data grid to synchronize the data atoms in a best-faith method), pessimistic (complete end-to-end transactional behavior), or somewhere in between these two extremes.

WAN See description for *wide-area network*.

wide-area network A wide-area network (WAN) is a collection of LANs typically spanning vast geographic distances.

worklet This is a unit of work that has counterparts, all of which are atomic with respect to each other and contribute to a larger work unit.

XML See the definition for *eXtensible Markup Language*.

XSL See definition for *eXtensible Style Language*.