System integration of heterogeneous complexes for scientific computing, based on the use of DB2 technology

Professor A.V.Bogdanov
Institute of High Performance Computing and Integrated Systems,
St.Petersburg, Russia
e-mail: bogdanov@csa.ru

Thurein Kyaw Lwin, Myo Tun Tun, La Min Htut
St.Petersburg State Marine Technical University
St.Petersburg, Russia
e-mails: phothar83@gmail.com, bonge21@gmail.com, laminhtut@mail.ru

ABSTRACT
This paper is the problem of consolidation of large distributed computational complexes on the basis of DBMS is considered. Possibility of using for these purposes of DB2 and some problems of its integration with Sun Grid Engine (SGE) are described.

Keywords
DBMS(database management system), SGE (Sun Grid Engine), SOA(Service Oriented Architecture), NAS (Network Attach Storage), SAN (Storage Area Network), UDB DB2 (Universal Database), MPP (Massive Parallel Processing),SMP(Symmetric Multiprocessing).

1. INTRODUCTION
The organization of access to the distributed computing resources, especially in the solution of the large complex problems, is a challenge for any general purpose computer centre. At the all advantages of the new distributed approaches based on metacomputing [1] with the application of SOA(Service Oriented Architecture) [2], an overhead of such approaches is very huge. At the same time consolidation problem arising on this way could be efficiently solved only for the limited number of applications. We offer this approach as a base for accessing to distributed resources and consolidation of applications. In this case we can use distributed databases for any purposes. DB2 is planned as base DBMS [3]. It could be combined with Grid portal and SGE (Sun Grid Engine) which is the necessary as resource manager. DB2 Universal Database is a basic of the whole family of software to process the heterogeneous information. Use of DB2 Universal Database in a combination with such software provides very effective realization of corporate information systems, including the heterogeneous ones. In this paper the basic requirements for DB realization on SGE (Sun Grid Engine), basic characteristics of safety of SGE and DB2technology for the organization of production Grids are considered.

SGE is designed for networks of a class of Cluster Grid and it is accessible free of charge. The package allows to unite some servers or workstations in a uniform computing resource which can be used both for complex problems, and for high-performance package calculations. The manager of the computer network can obtain data of monitoring and statistics, and on their basic optimize level of use of resources. The administrative interface allows to set various parameters of computing problems, such as the priorities, demanded service lives, licenses for the software, a time window of performance, the right of users to access to those or other resources [4].

2. WHY WE CHOOSED DB2?
DB2 is a scaled objective-relational control system of database with the integrated support of multimedia and Web, working from personal computers and servers on Intel processors with operational systems OS/2, Windows NT to Unix stations under control of various versions of UNIX, from uniprocessor systems to symmetric multiprocessing systems (SMP) and the systems with mass parallelism (MPP) and clusters [3].

The possibilities of expansion included in a kernel database, allow DB2 to include the extensions to some key modern technologies [4]. First of all: support of complex objects - focused and multimedia types of data, providing access to data through Internet, difficult transformations and the analysis of data together with maintenance of high reliability, productivity and scale in a range from uniprocessor systems to other systems with mass parallelism [4].

Configuration of communication functions, and functions of network interaction has been simplified, as well as adjustment of optimum productivity of a database. Besides closely integrated control facilities databases, the control of productivity and tuning of applications, in DB2 Universal Database, the schedulers of problems and the navigator/browser options, which considerably facilitate a life to managers of databases [4], have been included.

Data Storage procedures, triggers and functions of DB2 are defined by the user to promote collective and repeated use of functions, that considerably simplifies work of developers of the applications focused on databases.

3. GRID AND DB2
IBM had declared already for a long time intention to develop the products in a direction of wider application of Grid-technologies. Creating new versions of DBMS and a server of applications, it was expected with the help of "screen" calculations to raise productivity, scaling and reliability of applications, to provide the big safety of data at flexible resource management (including automatic) and at the same time to reduce equipment costs. For the solution of these problems following four basic technological directions have been defined.

One of the basic ways of decrease the expenses is to use the standard devices for calculations and data storage. It means, in particular, orientation to application of Web-Sphere, Tivoli, Lotus, Rational products and IBM Power PC, modular (blade) servers, network systems of a data storage (NAS and SAN), high-speed network connections, and also Linux platform.

Concentration on an IT infrastructure in one place or within the limits of a small number of computer centers allows raising essential efficiency of processing at the expense of application of larger clusters solutions.

In classical variant the Grid should provide dynamic distribution of all resources. Moving to the given direction, it is necessary to raise level of automation of these processes.

4. FEATURES OF DB2 GRID

Grid servers can be grouped together and they can be started on one copy of DB2. Thus each server in cluster receives the subset of a full data set and operates independently, and after association of results from all systems then the problem will be solved. Such paradigm, however, is not so effective; as it is badly scaled and also all system can lose productivity while adding of several servers to a group.

Architecture of DB2 UDB is completely parallel and supports parallelized execution of the majority of operations, including inquiries, inserting, updating and removal of data, creation of indexes, loading and export of data. And functionality of DB2, the transition from the standard, not parallel environment of execution on parallel, does not meet any limitations at the increase of efficiency [5]. DB2 UDB has been specially developed for successful work in a number of parallel environments including systems MPP, SMP and MPP clusters from SMP nodes [5].

Often enough it happens that the necessary information is stored in several absolutely various databases or in the information part of file system [5]. Such essential characteristics of DB2 UDB as objective - relational structure and corresponding expansions of DB2, integration with external sources of data by means of technology Data-Links, and also, in particular, algorithms of complex search in a structural types of data makes DB2 an ideal server for realization of the concept of the federal database including a variety of distributed sources and variety of types of data [5].

The demands for intellectual information systems always were high, but only recently relational DBMS have found possibilities in a sufficient measure to support such systems [5]. The analysis of data demands the use of the consolidated data sets from numerous operative sources of the data united in a data storage. The storage of data becomes a platform to support the diverse analytical applications [5]. As the storage of data urged to store higher and higher volumes of data with high efficiency and scaling requirements, they become the cores for DBMS.

5. CONSTRUCTION STORAGE OF DATA

In the most cases of the problems of the analysis of data on distributed systems the storage of data [1] is local. For designing, construction and maintenance of storage of data it is necessary to solve a number of problems of sampling of data, their preliminary processing, clearing and correction of incorrect values, loading of data in storage according to the designed model for the analysis and, at last, the whole group of the problems connected with construction of the automated environment of loading and management of data storage.

In a view of necessary performance support it is necessary to launch DB2 on SGE. The process of launching of DB2 can be divided into three basically stages:

- Code installation.
- Creation of additional copies of the data connected with each established to copy DB2.
- Formation of containers of databases.

In the majority of applications all parts are loaded and adjusted on local nodes with several copies of sections of data bases. DB2 supports physical creation of a copy by one machine and sharing of the house catalog to the owner of a copy through SGE for all participating nodes. In this case the loaded code, files of a database and containers all the same are local for all participating machines.

After installation and adjustment of guest operational system it is possible to load DB2 on virtual machine. Installation of DB2 on the virtual machine does not differ from installation on physical section SGE. The Setup can be started from a command line by the way of a command db2setup.

After installation of DB2 and successful creation of a copy of the program it is possible to pass the following stages. At an input point in the system, the owner of a copy, in this case db2inst1, it is possible to start the scenario db2profile for the task of variables of an environment. The database example can be created by the way of the scenario db2sampl.

6. ADMINISTRATION OF DB2

The main tool of administration is DB2 Control Center which helps to create, delete, and modify databases, tabular spaces, tables, indexes and triggers and to receive the information on their condition and parameters [4]. Except creation and administration of objects, Control Center is integrated into the uniform environment with Smart Guides, helping the managers step by step to carry out the problem of preservation and restoration of a database, creation of database, tables, tabular spaces and indexes, productivity adjustment. A number of components (Performance Monitor, Event Monitor and Event Analyzer) allow to carry out within adjustment and monitoring of work of DB2, are used for monitoring of productivity and the analysis of activity of various objects of a database (the table, tabular space, etc.) [4]. Visual tools allow investigating the plan of execution of inquiry to analyze what DB2 UDB addresses to data and processes them.

7. CONCLUSION

Integration of architectures of the Grid and DB2 has already produced some experimentally used software. Many real projects had already used components of developed middleware. Nowadays, DB2 is one of the leading software tools among manufacturers of DBMS and potentially the individual leader and possesses ability to give the solutions for a wide spectrum of business problems. In respect to the constant long-term tendency for DB2 orientation to corporate solutions corporate standards of productivity and reliability remains. Therefore the offered solution has the excellent prospects in the problem of consolidation of resources for large data centers.

8. ACKNOWLEDGEMENT

Initially, I would like to say thank you to Professor A.V.Bogdanov and Professor Alexander Degtyarev for

supporting me to do such a wonderful research. I am also really appreciate to all the people as well as the Saint Petersburg State University and State Marine Technical University Of St.Petersburg for giving me the great chance, great time and all the supporting.

REFERENCES

[1] http://www.ibm.com/
[2] Susan Visser , Bill Wong “Sams Teach Yourself DB2 Universal Database in 21 Days”, Sams Publishing, 2003
[3] http://www.citforum.ru/seminars/
[4] George Baklarz, Bill Wong DB2 Universal Database v8.1 for Linux, UNIX, and Windows database administration certification guide. IBM press, 2003.
[5] Nikolay Ignatovich “DB2 v 7 – integration of technology for data management”, *http://www.olap.ru/desc/ibm/db2ee7.asp* (in Russian)