Oracle Database 10*g*: Information Integration

An Oracle Business White Paper
May 2004

ORACLE

Oracle Database 10*g*: Information Integration

Oracle Database 10*g*: Information Integration

EXECUTIVE OVERVIEW

As a company's business priorities change, they are often faced with the challenge of aligning their resources to meet changing business needs. Oracle Database 10*g* provides a robust and complete grid computing solution that enables companies to easily align their resources as required. Information integration is a critical component of these solutions, as it enables companies to access information when and where it's needed in a distributed environment.

Companies who successfully implement information integration solutions will realize reduced costs, increased revenue, faster time-to-market, and increased customer satisfaction.

Oracle10g offers the most complete and the most comprehensive platform for information integration. As is demonstrated by its extensive history running critical business applications for the most demanding solutions, Oracle10*g* provides a robust set of features critical for integration, including high availability, security, scalability, and flexibility. It offers secure and standard communication mechanisms that enable communication between applications/users on the Oracle database using queues, data replication and distributed access in both homogeneous and heterogeneous environments.

INTRODUCTION

Over time, business priorities change. Enterprises reorganize themselves, acquire other businesses, grow some applications, and downsize others. Throughout these changes, companies are faced with the challenge of effectively and economically provisioning their resources to align them with their business priorities. To get your data where you want it, when you need it, requires sophisticated information integration technologies.

Oracle Database 10*g* provides robust and complete solutions to address all your information integration requirements. These solutions provide access to information when and where you need it, optimizing access to that information regardless of the physical location. They integrate information across a distributed environment, whether within a grid, across multiple standalone systems, or across some combination of these.

This paper covers information integration using distributed SQL, Oracle Streams, Oracle Transparent Gateways, and other data movement features. Oracle Streams provides a single unified environment for data sharing, including message queuing and replication. Oracle Transparent Gateways allow transparent access to non-Oracle systems from an Oracle environment. This paper further describes how these features enable not only efficient information integration, but efficient data provisioning for grid computing, as well.

Information Integration Requirements

There are several reasons why a company might be considering an information integration solution. A company may have recently undergone a merger or acquisition, or it may have chosen to exchange data with a business partner. Information integration solutions allow these companies to make the most efficient use of their computing resources by:

- Exchanging data between homogeneous and heterogeneous databases
- Enabling communication between applications
- Exchanging information with customers, partners, and suppliers
- Replicating data between databases

An information integration solution improves decision making by providing an integrated view of data, regardless of where that data is actually located. It provides improved operational efficiency by allowing multiple applications to operate cooperatively. Ultimately, this can result in improved operational efficiency for the IT department, as well, by reducing the number of data sources and improving communication and cooperation among those that remain.

Information Integration Solutions

Oracle Database 10*g* provides a variety of information integration solutions to address these varied requirements.

- Consolidation—All data is moved into a single database and managed from a central location.
- Federation—Data appears to be integrated in a single virtual database, while actually remaining in its current location.
- Sharing—Multiple copies of information are maintained in multiple databases and application data stores.

CONSOLIDATION

Oracle Database 10*g* lets you consolidate heterogeneous information into a single database with proven scalability. Oracle supports all of your data types,

including audio, video, XML, e-mail, messages, and others, and runs on nearly any platform, from Windows to Linux to Tru64.

Although consolidation provides the simplest form of information integration, it is quite often the most difficult to achieve. Certain departments may be either unwilling, or in the case of certain legacy systems, unable to consolidate their data. Most companies instead prefer to consolidate their data into a small number of databases. Reducing the total number of databases allows these companies to lower their administrative costs by reducing their total number of servers and databases to manage. By deploying applications only once (or a small number of times) companies can achieve faster application deployment. Finally, because queries run faster against centralized data, these companies will have faster access to global data, making consolidation an ideal strategy for data warehousing and business intelligence applications.

Oracle Database 10*g* provides a variety of features for customers wishing to consolidate their information, including Oracle Real Application Clusters (RAC), Virtual Private Database (VPD), partitioning, loading and migration tools, and many more.

Real Application Clusters (RAC)--RAC enables a single database to run across multiple clustered nodes, allowing a collection of inexpensive computers or blades to equal the performance of an expensive SMP box.

Virtual Private Database—VPD enables consolidation across security domains by providing fine-grained access to rows within a database.

Partitioning—Oracle's partitioning capabilities improve the performance, availability, and manageability of large tables, without requiring an application re-write.

FEDERATION

Federation leaves information in its original location, where it is normally maintained and updated. Multiple data sources appear to be integrated into a single virtual database, masking the number and different kinds of databases behind the consolidated view. Oracle's federated database solution allows access to Oracle and non-Oracle data, as well as structured and unstructured data, while hiding the actual physical location of the data from the application.

Because it is not necessary to move any information, federation provides the fastest path to information integration; minor changes yield immediate results. Federation also provides support for information, such as legacy applications or applications requiring local ownership, that simply cannot be consolidated. You should use federation to support ad hoc integration of infrequently accessed information.

Oracle Database 10*g* supports multiple data sources, as well as multiple methods of accessing these data sources.

Accessing Data with Oracle Distributed SQL

Oracle Distributed SQL makes a distributed database system containing both Oracle and non-Oracle databases appear as though it is a single Oracle database, by masking the physical location of any object from end-users. A company can use this feature to make all its databases look like one, and thus reduce some of the complexity of the distributed system. Oracle uses database links to enable users on one database to access objects in a remote database. A local user can access a link to a remote database without having to be a user on the remote database.

Oracle Distributed SQL provides the following benefits:

- Location transparency—the physical location of database objects can be changed as needed, with no impact on end-users or existing applications.
- SQL and COMMIT transparency—Using two-phase commit, Oracle automatically insures that standard SQL statements work the same in a federated architecture as they do in a standalone environment.
- Distributed query optimization—Oracle insures optimal performance, by minimizing the amount of data transfer required when a transaction references a remote table.

Accessing Data with XQuery

Oracle is actively participating in the W3C XML Query Working Group to develop XQuery, a standard for querying XML. XQuery provides a standard way to search and manipulate XML data. XML data has characteristics that cannot be represented by relational tables—for example, the order of elements. XQuery was designed to access XML data, and thus has some capabilities that are not inherent in SQL. XQuery can reduce application complexity when dealing with XML data. In addition to querying XML data, XQuery can also access relational data.

Heterogeneous Data Access

Many companies run several different database systems. Each of these systems stores data and has a set of applications that run against it. Oracle offers a variety of connectivity solutions that enable a company to seamlessly integrate the different systems and provide a consolidated view of the company as a whole.

- Generic Connectivity—is a database feature that uses an ODBC or OLEDB driver to transparently access any ODBC or OLEDB compliant non-Oracle system.

- Oracle Transparent Gateways—are certified, optimized solutions, specifically coded for the non-Oracle system, including Sybase, DB2, Informix, Microsoft SQL Server, Ingres, and Teradata.

Both Generic Connectivity and Oracle Transparent Gateways provide the ability to transparently access data in non-Oracle systems from an Oracle environment. This transparency eliminates the need for application developers to customize their applications to access data from different non-Oracle systems, thus decreasing development efforts and increasing the mobility of the application.

Accessing External Files

You can query native XML files using XQuery or, if they can be mapped to rows and columns, using SQL. Additionally, Oracle provides query access to over 100 different file types through the use of filters, which transform these files to appear as XML. The External Table feature allows for flat files, which reside outside the database, to be accessed just like relational tables within the database: the flat-file data can be queried and joined to other tables using standard SQL. Data access can be serial or parallel for maximum scalability. From a user's point of view, the main difference between an external table and a regular table is that the external table is read-only.

Accessing Web Services as a SQL Data Source

Web services have recently become an important feature for exchanging data between systems. A web service can be nearly any type of application that can not only expose to other applications what it does, but also perform that action for authorized applications or parties.

Oracle Database 10*g* can be a web services consumer; that is, it can call external web services from Java Classes, PL/SQL procedures, and triggers. In addition, Oracle can make a web service look like a SQL row source.

Oracle Database 10*g* can also be a web services provider. Oracle exposes many database operations as web services.

Locating Information with UltraSearch

Oracle Ultra Search provides uniform search-and-locate capabilities over multiple repositories. You can search the document metadata, as well as the content. UltraSearch indexes documents by crawling through the sources you want indexed. You can crawl Oracle databases, and other any sources that support the ODBC standard.

UltraSearch enables a portal search across the content assets of an organization, bringing to bear Oracles core capabilities of platform scalability and reliability.

SHARING

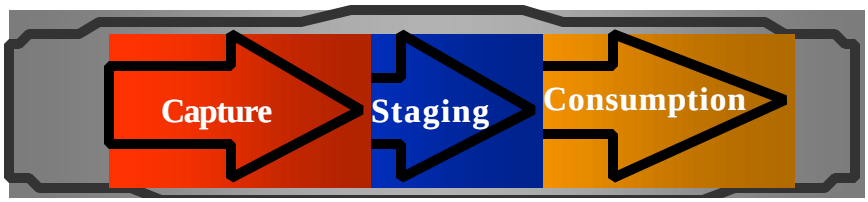
As a company grows and expands, it becomes increasingly important for that company to be able to share information among multiple databases and applications. Traditionally, a company may have selected from among a variety of information sharing technologies, each aimed at solving a specific business problem. While these targeted solutions may have initially appeared easier to use and implement, they fail to deliver once the needs of the company grow beyond the capabilities of the simple targeted solution. Suddenly, developers must implement multiple tools to build a solution, and complexity grows exponentially.

Oracle provides a variety of information sharing features including Oracle Streams, materialized views, transportable tablespaces, and data pump. These features allow you to locate the data where it is most commonly accessed, improving response time and eliminating dependencies on remote data stores.

Oracle Streams

At the heart of any integration is the sharing of data among various applications in the enterprise. Oracle Streams enables the propagation and management of data, transactions and events in a data stream either within a database, or from one database to another. The stream routes published information to subscribed destinations. The result is a feature that provides greater functionality and flexibility than traditional solutions for capturing and managing events, and sharing the events with other databases and applications. As users' needs change, they can simply implement a new capability of Oracle Streams, without sacrificing existing capabilities.

As shown in the following illustration, Oracle Streams provides a set of elements that allow users to control what information is put into a stream, how the stream flows or is routed from node to node, what happens to events in the stream as they flow into each node, and how the stream terminates. By specifying the configuration of the elements acting on the stream, a user can address specific requirements.



- Capture—Oracle can automatically capture DML and DDL events at a source database. Additionally, applications can explicitly generate events and place them in the staging area.

- Staging—The staging area provides a holding area with security, as well as auditing and tracking of captured data. Events can be propagated between staging areas, allowing for networked routing.
- Consumption—Events in a staging area can be automatically applied to a database or explicitly dequeued by an application. Data can be transformed when necessary.

Oracle Streams is an open information sharing solution. Each element supports industry standard languages and standards. Streams supports capture and apply from Oracle to non-Oracle systems. Changes can be applied to a non-Oracle system via a transparent gateway or generic connectivity. Streams also includes an API to allow non-Oracle data sources to easily submit or receive change records, allowing for heterogeneous data movement in both directions.

By specifying the configuration of the elements acting on the stream, a user can address specific requirements, such as message queuing or data replication.

Message Queuing

Oracle Streams Advanced Queuing provides a unified framework for processing events. It allows user applications to enqueue messages into the staging area, propagate to subscribing staging areas, notify user applications that messages are ready for consumption, and dequeue the message at the destination.

It supports all the standard features of message queuing systems including multi-consumer queues, publish and subscribe, content-based routing, internet propagation, and transformations Additionally, the message gateway functionality of Oracle integrates Oracle database applications with other message queuing systems such as Websphere MQ (formerly called MQ Series) and Tibco.

Unlike a traditional queue, the Oracle Streams staging area can hold messages of different types, so it is possible to enqueue different types of messages into a single staging area. Streams also supports notifications to user applications, merging the near real-time benefits of a push model with the scalability and manageability benefits of a pull model.

Data Replication

Replication provides a solution to the scalability, availability, and performance issues facing companies today. Oracle Streams determines what information is relevant and shares that information with those who need it.

Oracle Streams replication automatically captures changes from a source database, propagates those changes to one or more remote databases, and then applies the changes at each target database. Oracle Streams can replicate not

just changes to data (DML changes), but changes to the structure of a table (DDL changes), as well.

Database administrators can choose which changes are propagated to each destination database, and can specify how these messages are routed. This directed network approach enables complex configurations including n-way and hub-and-spoke, as well networked configurations that propagate changes once to a single site for later fan-out to other destinations. In the event that a change is made to the same data at multiple locations, Oracle Streams will automatically detect the conflict and resolve it as desired.

Materialized Views

A common requirement of many companies is to periodically disseminate a product catalog to regional offices and to enable the sales force to place orders from customer sites. To address these needs, Oracle offers a replication type referred to as materialized view replication. A materialized view contains a complete or partial copy of a table from a single point in time, which can be incrementally maintained.

Materialized views can be either read-only or updateable with conflict resolution. Because of their support for disconnected computing and easy mass deployment, materialized views are especially suited to mobile computing applications. Materialized views are also commonly used in data warehousing environments to improve performance by computing and storing aggregated data, such as sums and averages.

Hybrid Configurations

Customers can utilize the full power of Oracle Streams, along with Oracle Transparent Gateways and Oracle materialized views, to create configurations that seemingly span multiple markets, enabling new classes of applications. In addition, all deployments and their associated meta-data are compatible. For example, a replication installation can easily be extended to load a data warehouse, enable bi-directional replication, or even send notifications of DML changes to an application—a complete reconfiguration is not required.

Information Sharing with Oracle Transportable Tablespaces

Transportable tablespaces provides a mechanism for bulk information sharing. You can unplug a table space, copy or move it, and then plug it into another database, even if that database is on a different platform. Unplugging and plugging operations are very fast and do not relate to the size of a tablespace. You can also mount a single tablespace in read only mode on multiple databases simultaneously, essentially allowing two databases to share information and process that data independently and in parallel.

Information Sharing with Oracle Data Pump

Oracle Data Pump also supports bulk information sharing by providing high-speed parallel bulk data and metadata movement of Oracle database contents. Data Pump lets users move or copy subsets of a tablespace or database, providing more granularity than Transportable Tablespaces. Users can specify specific objects to transfer, can remap schemas, data files and tablespaces.

Data Pump can also operate in an interactive mode that lets users add additional files to the dump file. Additionally, because Data Pump is compatible with Oracle Streams, users can export objects to a new database and use streams to keep those objects in synch with a production database.

Bulk Data Movement with Oracle Streams

Oracle Database 10*g* Streams can move any operating system file, BFILE, and database, including those comprising a table space. This allows you to create a replica with a single command. All you have to do is specify the table spaces to replicate and a dB link pointing to the replica database. Streams will automatically create a copy of the tablespaces you specify, move that copy to the destination database, and synchronize that copy with the production source. You also have the option of specifying you want bi-directional replication set up as part of the operation. Once the replica is instantiated, you can easily modify the configuration—for example you can add data transformations.

USING ORACLE'S INFORMATION INTEGRATION TECHNOLOGIES

Companies use Oracle's information integration technologies for a variety of different purposes, from providing access to legacy applications to improving the performance of critical business applications.

One such customer, a large institutional investment firm, provides a typical usage scenario. In order to provide high scalability and availability, with improved performance, this company replicates three of its databases to secondary servers. The replication for each database is such that either replica can service the functions of both, but for performance (load balancing) and administrative reasons, each database typically serves different operational needs.

Call center agents access one replica of the BackOffice database, which houses the company's proprietary CRM and authentication systems, while another replica is used to monitor client activity. One replica of the Search database is used to perform bulk loading of documents in hourly batches, while another replica is used to receive users' queries, which can then be executed on either Search database server. Finally, the replica of the Document database, which contains metadata about these documents, as well as their actual, physical locations, is used primarily as a standby for the production system.

This Oracle Streams-based solution allows this investment firm to execute queries in half the time or better, and provides instantaneous switchover in the event of a failure.

USING INTEGRATION TECHNOLOGIES FOR PROVISIONING DATA IN A GRID ENVIRONMENT

Many of the same technologies described in this paper that are useful for information integration are useful for provisioning data in a grid environment, as well. You can use the fine-grained provisioning features of Oracle Streams, to provision information a record at a time. You can also use features like Distributed SQL to leave data in place until it's needed, thus provisioning on demand. Bulk data movement tools such as Transportable Tablespaces and Data Pump are an efficient way to provision large amounts of data in a grid.

CONCLUSION

Today's rapidly changing business environment often drives companies to a distributed environment. Looking at the information integration needs that arise for companies in such situations it is obvious that Oracle offers a complete and robust solution to address each of these needs in a reliable, and secure environment.

Distributed SQL is used to transparently communicate between the Oracle databases. Generic Connectivity and Transparent Gateways extend that capability to non-Oracle systems. Oracle Streams Advanced Queuing extends the communication capability beyond just databases to applications. Additionally the replication capabilities of Oracle Streams can significantly improve the availability and scalability of these applications.

Oracle Database 10*g*: Information Integration
May 2004
Author: Maria Pratt
Contributing Authors: Vira Goorah, Brajesh Goyal, Pat McElroy, Bob Thome
Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
www.oracle.com

Oracle Corporation provides the software
that powers the internet.