Introduction to Databases, Fall 2003

IT University of Copenhagen

Lecture 12, part I: Overview of some commercial DBMSs

November 18, 2003

Lecturer: Rasmus Pagh

Today's lecture, part I

We look at the three leading commercial DBMSs:

- Oracle

- SQL Server (Microsoft)

- DB2 (IBM)

Disclaimer

This lecture to a large degree is based on material written by employees of Oracle Corporation, Microsoft, and IBM.

While it is believed that this material is fairly objective, there is the possibility that some of the information is "sales talk".

No serious effort has been made to criticize and control this information – in particular the lecturer has no experience with two of the three products.

SQL variants

Every RDBMS has its own variant of SQL:

- Similar to recent SQL standards (SQL-92 or SQL:1999).

- Sometimes small deviations from the standards for historical reasons (can't suddenly change the semantics) or implementation reasons.

- Non-standard (or de facto standard) features to attract customers.

Standards compliance:

- Oracle supports all core SQL:1999 features "fully or partially".

- SQL Server complies with the entry level of SQL-92.

- DB2 has no stated compliance.

Object oriented features

A trend in the last 10-15 years has been to add features of object oriented programming in RDBMSs. This is sometimes called the **object-relational model**.

All three DBMSs have support for:

- User-defined types.

- Persistent modules (i.e., methods in the database schema).

- ... and other things you would expect if you know about OO.

There seems to be no major differences in this respect.

Databases on several machines

Very large databases often need to be distributed on a number of computers to achieve the desired performance.

- Oracle and IBM seem to have made the largest effort in distributed and parallel databases, each claiming to be the most scalable.

- SQL Server has had limited support for parallelism, but is making an effort to catch up.

SQL Server and Oracle offer the creation of a "standby database", which is a read-only copy of the database, which may take over if the main database crashes.

Transaction processing

A major difference in the implementation of Oracle vs DB2 and SQL Server is how transactions are handled.

- In Oracle, a read-only transaction never needs to wait for other transactions to avoid a dirty read (because data overwritten by an uncommitted transaction can be looked up).

- In SQL Server, a read-only transaction may have to wait for another transaction (using pessimistic concurrency control), or may be rolled back as the result of a dirty read (using optimistic concurrency control).

- In DB2, a read-only transaction may have to wait for another transaction (uses pessimistic concurrency control).

Transaction processing II

Locking tables and rows is the mechanism RDBMSs use to assure that transactions have the desired isolation level:

- Parts of a table "write locked" by a transaction cannot be read by other transactions.

- Parts of a table "read locked" by a transaction cannot be written to by other transactions.

Oracle sometimes locks smaller parts of tables than DB2/SQL Server (avoiding so-called escalating locks).

Altogether, Oracle seems to have an advantage in settings with many read-only transactions concurrently with (large) database updates.

Data warehousing

All three DBMSs come with tools for data warehousing.

- Oracle seems to go for a unified OLTP/OLAP solution using **materialized views**, which are views that are physically stored on disk, to maintain precomputed summaries.

- SQL Server also offers materialized views, but the OLAP data is stored in separate tables.

- DB2 offers separate data analysis and data mining products.

Performance

The performance of the different DBMSs seems to depend a lot on the computer system it runs on.

Usually the difference is small or modest (less than 30%). Trend:

- Oracle is fastest,

- closely followed by DB2,

- and with SQL Server sometimes lacking behind.

Both Oracle and IBM have whitepapers showing their DBMS to be fastest in specific scenarios.

Market share

According to an independent analyst (Gartner), DB2 is the leading DBMS in terms of 2002 market share (new license revenues):

- DB2 had 36% market share.

- Oracle had 34% market share.

- SQL Server market share was roughly 15%.

However, Oracle publicly questions the data provided by IBM to analysts, and provide data suggesting that Oracle is ahead in terms of number of users.

There is a battle on words regarding which DBMS is best in terms of "total cost of ownership".

Open source competitors

Recently, several free, open source RDBMSs have had some success, e.g.:

- MySQL.

- PostGreSQL.

Until now it seems that these have mostly seen use in non-critical applications like web servers, but this may change as more features are added.

Most important point in this lecture

- That you have a good chance of knowing at least a little bit about the DBMS used by your future employer(s).

The material covered by these slides (and the slides themselves) is **not** curriculum for the exam.