

UNIVERSITY OF TURKU  
Department of Information Technology

LASSE LAHTINEN

Performance Issues in Data Warehouses

Master of Science in Technology Thesis, 68 p.  
Embedded software systems  
Jan 2010

---

Working with data warehouses means working with a huge amount of data against a user requirement for a fast response time. Every developer or database administrator faces performance problem issues: how to handle huge amounts of data and at the same time how to provide the data with a fast response time. There are a number of things which can be considered to respond to these challenges. A data warehouse design is one important task. When the design has been accomplished well, the rest is much easier. Against slow queries, indexes are the common problem solving method, and if the indexes do not give enough power, then, parallelism can be used in different ways. Partitions help to split a big table into smaller parts and then, an optimizer can take advantage of them.

In this thesis work, Oracle and Teradata data warehouses are investigated, and the purpose is to find common problem solving methods for both systems. Also, the differences between Oracle and Teradata data warehouses are under investigation.

It has been found that indexes are a common problem solving method in both data warehouses in slightly different ways. Where Teradata uses parallelism basically everywhere the Oracle problem solving method is more sophisticated with indexes, hints, statistics and histograms.

Keyword: Oracle, Teradata, index, data warehouse, partition