

PERFORMANCE COMPARISON OF INTERSYSTEMS CACHE AND ORACLE IN A DATA MART APPLICATION

Abstract

A global provider of mobile telecommunications software, Airwide Solutions, tested the performance of InterSystems Caché® and Oracle 9i as the database in a simulated data mart application. They found Caché to be 41% faster than Oracle at building a data mart. When testing the response time to SQL queries of the data mart, Caché's performance ranged from 1.8 times to 513 times faster than Oracle.

a white paper by

Mark Massias, PhD
Sales Engineer
InterSystems Corporation

PERFORMANCE COMPARISON OF INTERSYSTEMS CACHÉ IN A DATA MART APPLICATION

Introduction

Telecommunications companies, because they generate and must analyze enormous amounts of information, are among the most demanding database users in the world. In order to make practicable business intelligence solutions, telecommunications firms typically select key pieces of raw data to be loaded into a “data mart”, where it is indexed and aggregated in various ways before being made available for analysis. Even so, the data marts in question may be hundreds of gigabytes in size. Database performance, both in the creation of the data mart, and in the query response time of the data mart, is critical to the timely analysis of information, and ultimately to the ability of the enterprise to identify and act upon changes in their business environment.

This paper presents the results of comparative performance tests between InterSystems Caché and Oracle 9i. They were performed by the mobile telecommunications software company Airwide Solutions, as they evaluated database technology for incorporation into a new business intelligence application relating to mobile phone usage.

The Test

For the test, Airwide simulated a typical data mart scenario. In the first part of the test, a large amount of mobile phone data was loaded, indexed, and summarized. The times to complete each step and the resulting database size were measured. In the second part of the test, Airwide measured the response times to several different SQL queries.

All testing was performed on a dedicated HP DL380 server running 32-bit Linux Redhat A.S. 3.0. The server had 4GB RAM with 2GB shared memory, and a single 3.4GHz XEON processor.

Part 1 – Load, Index, and Aggregate Data

The simulated data mart used in the test consisted of a main “fact table” containing the following elements:

- Manufacturer – mobile phone manufacturer
- Model – mobile phone model
- MSISDN (Mobile Station International ISDN Number, i.e.: the phone number)
- Roaming (Yes/No) – whether a subscriber is roaming
- Changed (Yes/No) – whether a subscriber has changed between regions
- Region – What region the subscriber is in (providers may cover multiple regions)

The fact table was indexed, using a combination of standard and bit-map indexes, to speed query response times. In addition, several summary tables, which aggregated the data by manufacturer, model, region, and permutations thereof, were created.

The system was loaded with 10 files of 500 MB each, for a total of 5 GB of data.

Figure #1 shows the time in seconds to load, index, and aggregate the data. Overall, it took Caché 2,819 seconds to complete the full process. This is 41% faster than the 3,967 seconds required by Oracle.

Figure #2 compares the sizes of the resulting databases and indexes. Caché used slightly less disc space, being 7.8% smaller.

Figure #1: Data Load, Index, and Consolidation Times

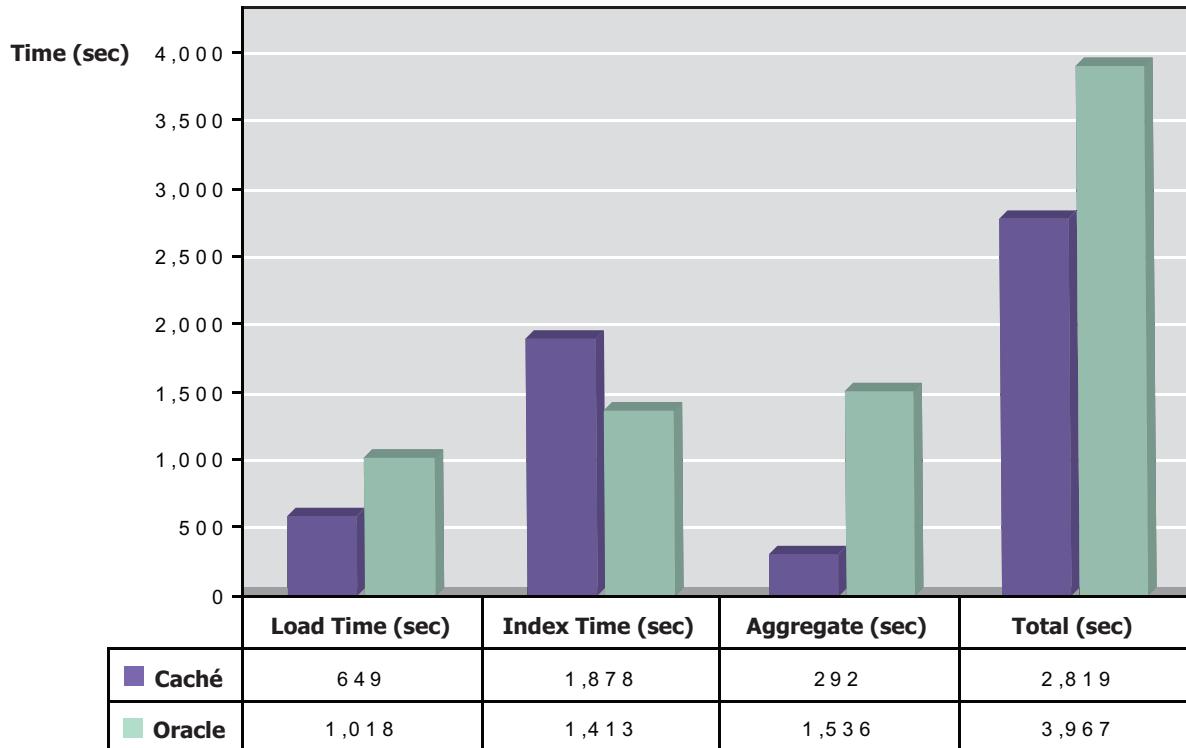
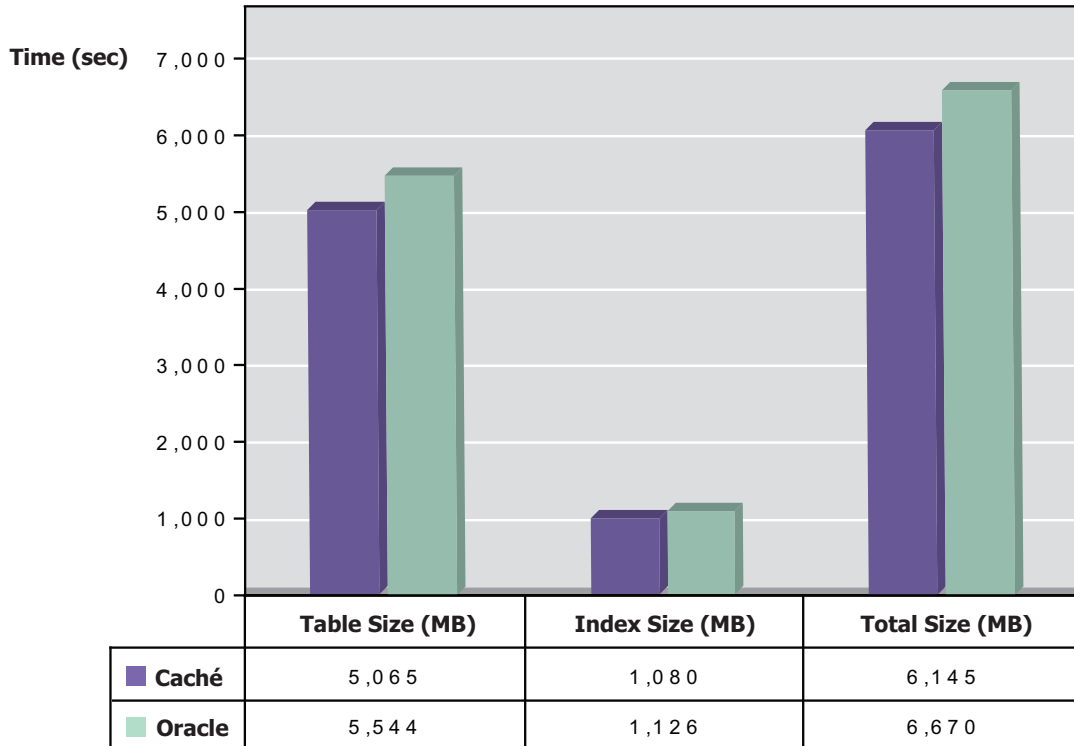


Figure #2: Database Sizes



Part 2 – Query Performance

Five different SQL queries were run against the Caché and Oracle databases:

Query 1: `select count(*) from fact_table where MSISDN like '%579%'`

Query 2: `select count(*) from fact_table where MSISDN like '%356%' and Manufacturer='Nokia'`

Query 3: `select * from fact_table where MSISDN like '%59421%'`

Query 4: `select * from fact_table where MSISDN like '%21%46%' and ISRoamer = 0 and ISChanged = 1`

Query 5: `select * from fact_table where MSISDN = '161323202273' and ISRoamer = 0`

All queries were run without “warming” the databases beforehand. This avoided the use of cached data and provided a more rigorous basis for the performance comparison.

Several indexing strategies were tested when building the Oracle database. One replicated the strategy used with the Caché database. Others were specifically designed to optimize Oracle’s performance. The response times given for the Oracle database are the best that were achieved.

Figure #3 shows the response times in milliseconds for each of the five SQL queries. Caché was faster in every case.

Table #1 demonstrates how much faster Caché was, by showing the ratio of the Caché/Oracle response times for each query.

Figure #3: Query Performance (in mS)

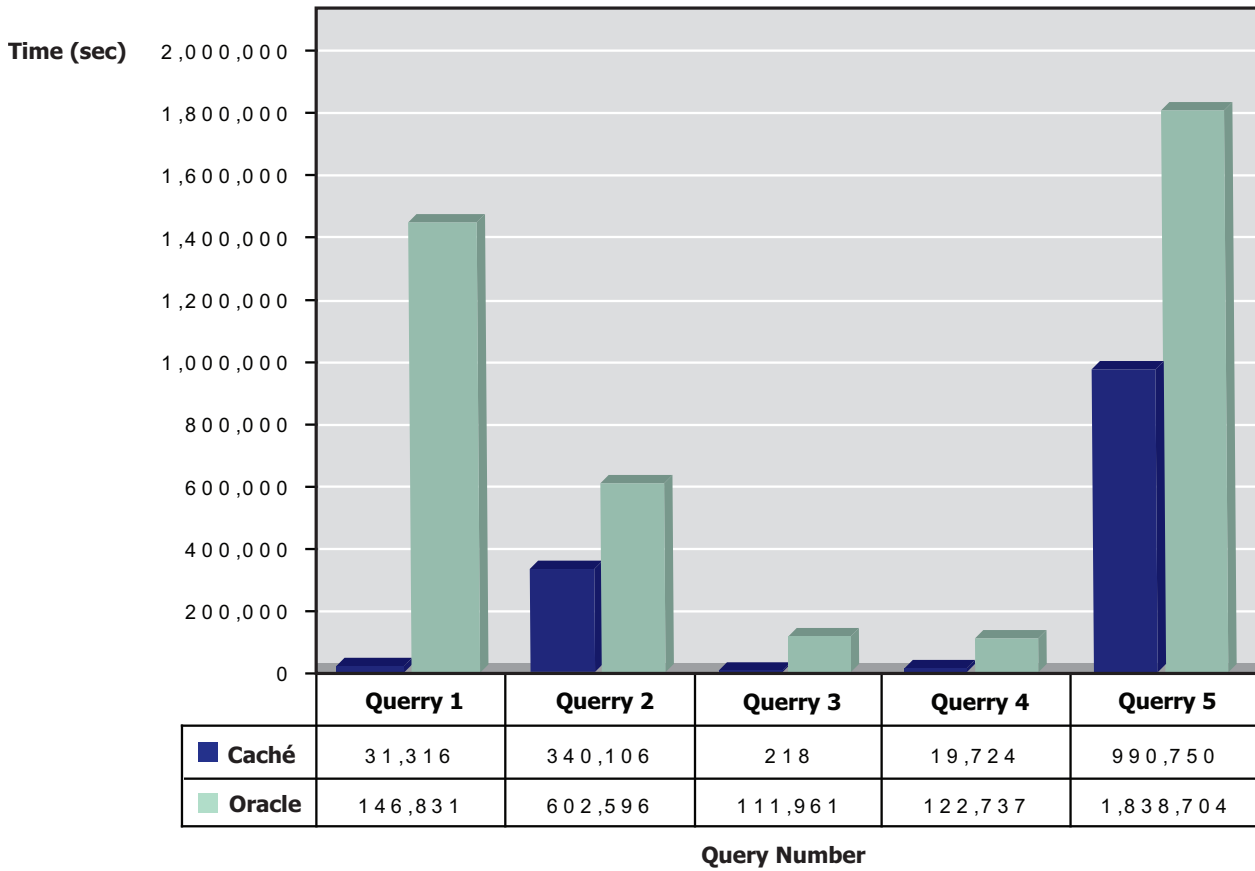


Table #1: Caché/Oracle Ratio

Query #	Caché faster than Oracle by
1	4.7 X
2	1.8 X
3	513 X
4	6.2 X
5	1.9 X

Conclusion

In tests simulating a data analysis application typical for a telecommunications software firm, Caché was 41% faster than Oracle 9i when creating a data mart of mobile phone information. When the resulting data mart was queried using SQL, Caché's response times ranged from 1.8 to 513 times faster. *Clearly, Caché's unique multidimensional data engine make it a good choice for applications that require rapid analysis of large amounts of data.*

InterSystems Corporation

World Headquarters

One Memorial Drive

Cambridge, MA 02142-1356

Tel: +1.617.621.0600

Fax: +1.617.494.1631

InterSystems.com

INTERSYSTEMS