

REAL-WORLD BENCHMARK: CACHE VS. ORACLE IN A DATA WAREHOUSING APPLICATION

A technical white paper by:

Mark Ramsay

Senior Sales Engineer

InterSystems Corporation

REAL-WORLD BENCHMARK: CACHÉ VS. ORACLE IN A DATA WAREHOUSING APPLICATION

Abstract

A benchmark of a real-world application, which loads data into a data warehouse for subsequent analysis, was performed. To conduct the benchmark, one module of the Oracle-based application was replicated in Caché ObjectScript. Only about 40 person-hours of work was required to duplicate the functionality of the original module in Caché.

The time required to complete certain tasks was measured running the Caché-based module, and was compared to existing data. **The Caché application was five times faster**, taking only 286 minutes to complete the data loading and filtering tasks that took 1441 minutes when using Oracle.

Introduction

Meralco Corporation, a large power supply company in the Philippines, currently uses Oracle for their corporate data warehouse. Every month a large amount of text-based data is loaded into the data warehouse, filtered, and made available for analysis. In an effort to improve their data warehousing operations, Meralco approached Digital Dimensions, Inc., a distributor of the high-performance Caché DBMS from InterSystems Corporation. Meralco agreed to let Digital Dimensions and InterSystems replicate the Field Order module of their Data Warehouse ETL (Extract, Transform, Load) application in a Caché-based system, in order to test Caché's performance compared to Oracle.

The Field Order Module

Meralco's Field Order module, based on Oracle 8i, is designed to take input from text files generated in the field and generate two tables - a field order dimension table and a field order fact table - that will subsequently be used for data analysis and reporting. To accomplish this task, it performs the following operations:

- 1) Multiple text files (ordenes.txt and gcahorde.txt) generated in the field are loaded into two staging tables, ORDENES and GCAHORDE. In the Oracle-based application, this is accomplished using Oracle SQL*Loader.
- 2) The GCAHORDE table is filtered, and the GCAHORDE_FILTERED table is generated.
- 3) The field order dimension table, DD_FIELD_ORDER is generated from GCAHORDE_FILTERED and ORDENES, using Oracle packages and stored procedures.
- 4) The field order fact table, DF_FIELD_ORDER is generated from GCAHORDE_FILTERED and ORDENES, using Oracle packages and stored procedures.

Replicating the Field Order Module in Caché

Caché allows ODBC and JDBC access to data, so many relational applications can run against the Caché database with minimal changes. The exceptions are those applications that, like Meralco's Field Order module, rely heavily on stored procedures. Stored procedures often need to be reincarnated as Caché ObjectScript routines when converting a relational application to Caché. For that reason, it was determined that it would be quicker to replicate the functionality of the Field Order module in Caché ObjectScript, than it would be to adapt the existing code.

This approach had the added benefit that it allowed the new Caché-based module to take advantage of Caché's ability to pre-filter text files on import. Because data from the original gcahorde.txt files was filtered as it was loaded into the database, there was no need for the Caché-based module to perform a separate filtering step.

Working with advice from Meralco, InterSystems created a Caché-based version of the Field Order module. The task required approximately 40 person-hours of work by developers familiar with Caché.

The Results

Both the Caché and Oracle 8i versions of the Field Order module were run starting with the same set of text-based data files. The time it took each version to complete various tasks was measured. The Run-Times for Caché were between four and twenty-three times faster than Oracle, depending on the task. The results are summarized below:

Task	Oracle		Caché		Run-Time Ratio (Oracle : Caché)
	Elapsed time	# of records	Elapsed time	# of records	
Load ORDENES	138 min.	927,857	6 min.	927,857	23:1
Load & Filter GCAHORDE	245 min.	90,349 *	23 min.	90,349	10:1
Generate FO dimension table	168 min.	1,018,216	24 min.	1,018,216	7:1
Generate FO fact table	890 min.	1,018,216	233 min.	1,018,216	4:1
Total time to load and filter data	1441 min.		286 min.		5:1

* 13,396,510 source records filtered, resulting in 90,349 records loaded.

Conclusion

In this real-world benchmark of a data warehousing application, Caché was 5 times faster than Oracle when performing the data loading and filtering operations.

Only 40 person-hours of work was required to replicate the original Oracle-based application in Caché.

InterSystems Corporation

World Headquarters
One Memorial Drive
Cambridge, MA 02142
Tel: 1.617.621.0600
Fax: 1.617.494.1631

www.InterSystems.com

