Database Storage & Retrieval of Seafloor Observatory Data using DBXten

Prepared for the Annual Meeting of the Gulf of Mexico Hydrates Research Consortium in Oxford, Mississippi

February 26-27, 2008

by

Cedric Zala and Ian Barrodale

Barrodale Computing Services Ltd. (BCS)

http://www.barrodale.com

BCS

Storing Scientific Data in an Object Relational Database Management System (ORDBMS)

The Seafloor Observatory will generate large volumes of data, such as hydrophone and geophone array output. Storing these data sequences in an ORDBMS has the advantages of:

- ***Improved concurrency*** - concurrent users can safely access (read/write) the same data repository.
- ***Improved security*** – access can be controlled through table, row, and/or column permissions.
- ***Composite data types*** - data is bundled with its metadata.
- ***Database extensibility*** - easy addition of data types and operations.
- ***Uniform treatment of data items*** - the SQL interface can perform complex queries based on **any** of these data items, e.g., metadata as well as data; hence there is less need for custom programming by users.
- ***Custom data access/search methods*** - e.g., R-tree indexes.
- ***Point-in-time recovery*** of data is possible.
- ***Powerful SQL functions*** can be provided for data operations - e.g., aggregating, slicing, subsetting, reprojecting, etc.
- ***Improved integrity*** - ability to reject bad data before it is stored in an ORDBMS.

BCS

Data Sequences

Much scientific data, including most of the data to be generated by the Seafloor Observatory, consist of data sequences:

- The data naturally occur in the form of tuples: e.g., (x, y, z, t, pressure).
- The data consist of temporal and/or spatial sequences (e.g., time series, trajectories).
- There are often strong internal correlations between the data values: e.g., grid coordinates are perfectly correlated, trajectory (x, y, z) values are highly correlated, and pressure values are significantly correlated.

Some examples are:

- CMRET hydrophone/geophone arrays.
- Sea-surface temperature data from buoys.
- Meteorological/oceanographic gridded data.
- GIS data - points, lines, polygons.
- Financial data - stock prices.
- Oil and Gas – prospecting and modeling data.

BCS

The BCS DBXten Database Extension

- is designed to store tuple data.
- is suitable in cases where the tuples represent some object that has a parent – all relationships are through that parent.
- supports timestamps, floats, integers, and strings.
- supports very efficient content-based indexing.
- works particularly well when there is a correlation between tuple rows.
- can be used as a general replacement for more specialized types.
- can handle any dimensionality.
- provides compression.
- handles both gridded and non-gridded data.
- provides for convenient and fast database loading and extraction.
- provides C, Java, and SQL application programming interfaces.
- is supplied with full user/programmer documentation.

BCS

Example of DBXten Performance

The table in the next slide shows the performance improvements obtained when DBXten was used within a commercial ORDBMS to store a large oceanographic dataset. This dataset consisted of approximately 50,000,000 rows with 26 columns, representing two days' worth of NOAA sea surface temperature readings and associated data.

BCS

Example of DBXten Performance

Task	Conventional Approach	BCS DBXten	Improvement Ratio
Size of table	15.6 GB	1.4 GB	x 11
Size of index	6,605 MB	6.8 MB	x 971
Index creation time	5 hours, 15 minutes	5 seconds	x 3,780
Insertion time	1 minute, 39 seconds	1.2 seconds	x 83
Retrieval time	14.7 seconds	3.8 seconds	x 4

The results show the dramatic improvements in speed and size afforded by DBXten.

BCS

DBXten Product Development

- DBXten is already available for IBM Informix, PostgreSQL, and SQL Server, and we are currently porting it to other ORDBMSs, including Oracle.
- We are marketing DBXten for use in meteorology, oceanography, geophysical prospecting, financial analysis, GIS, and resource management.
- DBXten is a patent pending technology; use the link below to read the patent application:

http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=%2Fnetahtml%2FPTO%2Fsrchnum.html&r=1&f=G&l=50&s1=%2220080021914%22.PGNR.&OS=DN/20080021914&RS=DN/20080021914

BCS

SUMMARY

The BCS DBXten Database Extension

…..is designed for applications where some of the following apply:

1. **the data consists of high-volume data sequences of tuples.**
2. **there is a correlation between tuple rows.**
3. **efficient content-based indexing is required.**
4. **multiple dimensions are involved.**
5. **compression may be required.**

BCS

CONTACT INFORMATION:

Dr. Ian Barrodale, President

Dr. Cedric Zala, Projects Manager

Barrodale Computing Services Ltd. (BCS)

P.O. Box 3075 STN CSC

Victoria BC V8W 3W2 Canada

(250) 472 4372 voice, (250) 472 4373 fax

e-mail: info@barrodale.com

For more information about BCS projects, experience and capabilities, please visit:

http://www.barrodale.com

BCS