



An Introduction to BCS products for “Making Databases Scientist-Friendly”

Ian Barrodale

Barrodale Computing Services Ltd. (BCS)

www.barrodale.com

BCSsales@barrodale.com

BCS



Barrodale Computing Services Ltd. (BCS)

- Founded in 1978, internally financed, debt free, develops customized software under contract to scientists, engineers and other technologists
- Completed 450+ software projects to date
- Is an IBM Advanced Business Partner
- Is now going to market with database extension products



Typical Data Sources: Gridded Data and Sensor Data

- **Gridded data occurs in meteorology**, life sciences, oceanography, non-destructive testing, exploration for oil, natural gas, diamonds,...
- **Gridded datasets** range from simple, uniformly spaced grid points along a single dimension (e.g., **time series**) to multi-dimensional grids with many values (e.g., **4D cubes** of meteorological attributes)
- The **\$10 billion US sensor industry continues to expand**, with imaging, proximity/positioning and chemical property sensors experiencing the most growth through aerospace, automotive, and military expenditures
- **Sensor webs are being deployed worldwide**, utilizing web-connected devices such as flood gauges, air pollution monitors, stress gauges on bridges, mobile heart monitors, webcams, and robots, as well as oceanographic, space and airborne earth imaging devices



The Problem – Making Databases Work for Scientists

BCS has interacted with many scientists and engineers who have large amounts of complex data, sometimes kept in databases, but often just in files

Some problems with using flat files are

- files introduce artificial seams
- hard to find the needle in a large, complex haystack of varying format files

Our suggestion: use a database

- provides a uniform treatment of data
- no seams
- built-in support for indexing and parallelization

What's holding scientists back?

- perception that databases are expensive, complicated, and inefficient
- databases are business-oriented, not science-oriented



The Problem – Making Databases Work for Scientists

Our Approach

Modern databases are extensible, and efficient specialized features can be added to DBMSs by developing database extensions (DataBlades, Cartridges)

Our Solutions

The **DaL** client - for loading scientific data files into database tables

The **Grid DataBlade** - for efficiently managing multi-dimensional gridded datasets in a database

The **DBXten DataBlade** - for improving query performance and reducing storage of large quantities of measurement/sensor data in a database

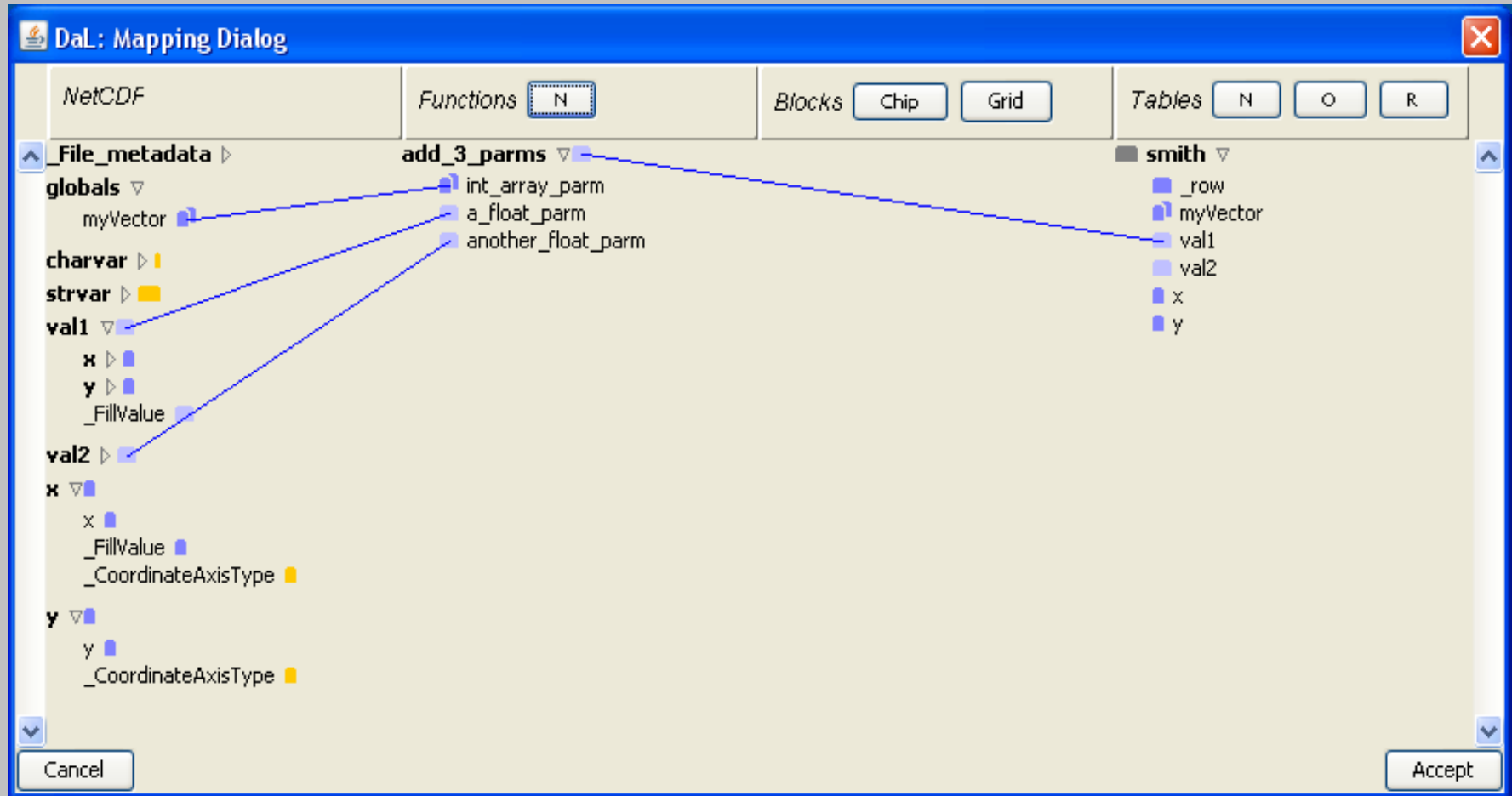
Speed of DBXten or Grid DataBlade versus flat files?

Fastest time is **green**, next fastest is “**amber**”, and the slowest time is **red**

Nature of Extraction	Extraction Size	DBXten DataBlade	Grid DataBlade	NetCDF flat files
Time profile	81x1x1x1	1.03 sec	2.24 sec	1.30 sec
Depth profile	1x66x1x1	0.07 sec	0.17 sec	0.72 sec
Planar slices	1x1x50x50	0.06 sec	0.09 sec	0.03 sec
	1x1x100x100	0.09 sec	0.13 sec	0.03 sec
	1x1x200x200	0.17 sec	0.23 sec	0.04 sec
Depth-Latitude slices	1x33x25x1	0.07 sec	0.16 sec	0.37 sec
	1x33x50x1	0.09 sec	0.20 sec	0.38 sec
	1x66x100x1	0.14 sec	0.38 sec	0.80 sec
	1x66x200x1	0.20 sec	0.59 sec	0.68 sec
4D Volumes	10x5x20x40	0.19 sec	0.46 sec	0.69 sec
	10x10x40x80	0.34 sec	0.99 sec	1.24 sec
	10x20x80x160	1.12 sec	3.75 sec	2.49 sec
Average time		0.30 sec	0.78 sec	0.73 sec

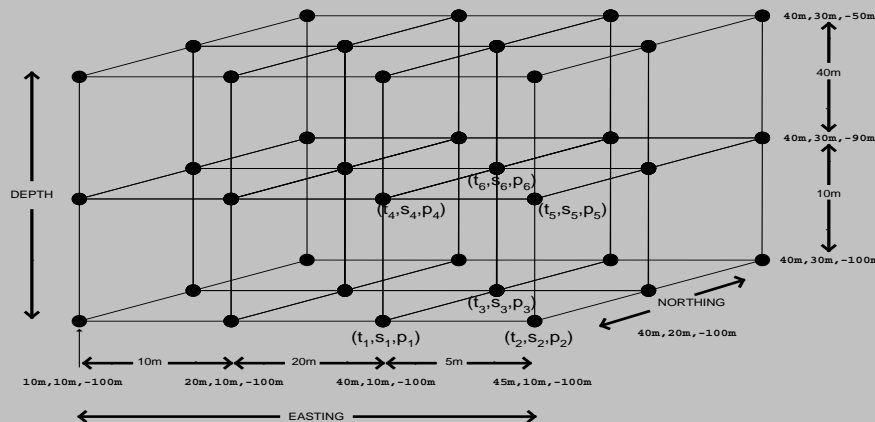
The BCS Draw and Load (DaL) Tool

DaL is a Java-based database client that provides the ability to load scientific data files (in a wide variety of formats) into a database



The BCS Grid DataBlade

- Handles 1D-4D gridded data
- Can load and extract data in a variety of formats
- Re-projects data on the fly
- Provides C, Java, and SQL APIs
- Full user/programmer documentation freely downloadable
- Versions currently available for Oracle, IBM Informix, and PostgreSQL





The BCS DBXten DataBlade

- Ideal for data generated by instruments/sensors
- Stores data sequences in compressed form
- Uses content-based indexing
- Able to exploit known precision of data
- Equipped with C, Java, and SQL APIs
- Full user/programmer documentation freely downloadable
- Versions currently available for Oracle, IBM Informix, and PostgreSQL
- BCS filed US Patent Application #11/779,791 for DBXten on 21/07/2006



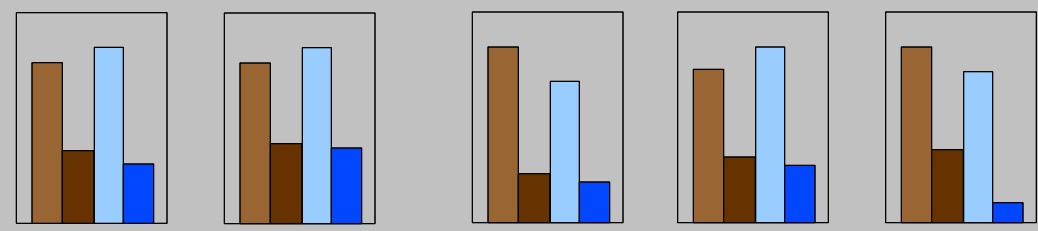
Example of DBXten performance

- The next slide shows, for a specific dataset, the performance of IBM Informix and Oracle, both with and without DBXten
- This benchmark test consisted of loading a table, with and without indexes, with 50,000,000 rows and 26 columns of NOAA sea surface temperature readings and associated data – and then querying it
- Four representative queries were run against the 50 million records



DBXten Benchmark

		Unindexed		Indexed		
		Space MB	Load secs	Space MB	Load secs	Query secs
ORACLE						
Conventional	■	1,894	2,836	3,172	3,574	205
with DBXten	■	848	1,410	860	1,576	85
INFORMIX						
Conventional	■	2,024	3,036	2,556	4,069	176
with DBXten	■	733	1,318	736	1,367	23



Performance Improvement with DBXten

	Unindexed		Indexed		
	Space	Load	Space	Load	Query
ORACLE	2.2	2.0	3.7	2.3	2.4
INFORMIX	2.8	2.3	3.5	3.0	7.7



Summary

A DBMS equipped with BCS extensions can be faster and easier than using a flat file format, with all of the added benefits of using a DBMS

BCS **DaL** makes it very easy to load scientific data files into a database

BCS **Grid DataBlade** easily meets the sophisticated requirements of meteorologists, atmospheric scientists, oceanographers, aviation authorities and others to 'slice and dice' 4D gridded datasets

BCS **DBXten DataBlade** table-in-table column compression extension gives impressive performance improvements over the performance of the most recent versions of DBMSs using row compression

Thank you for attending this presentation,
Ian Barrodale (ian@barrodale.com)