



BCS产品介绍

--建立研究人员易用的数据库

Ian Barrodale

Barrodale 计算服务有限公司 (BCS)

www.barrodale.com

ian@barrodale.com

BCS



Barrodale 计算服务有限公司 (BCS)

- 成立于**1978**年，内部资金，没有债务，以合同的方式为科研人员、工程师、及其他技术专家们开发软件
- 到今天已完成了**450**个以上的软件项目
- 是 **IBM** 高级合作伙伴
- 数据库扩展产品正在上市中

典型的数据来源：格子式的数据和传感器的数据

- 格子式的数据来源于气象学，生命科学，海洋学，非破坏性的试验，原油，天然气，砖石的勘查...
- 格子式的数据集合从简单,等距离的一维点(比如时间系列)到多维的格子(比如4维体的气象属性)
- 美国**100亿美元**的传感工业继续在扩展。通过图像,最接近的/定位和化学性质传感器，在航空，汽车行业和军队经费中快速发展
- 传感网在全球被广泛使用：利用网络连接的设备，像洪水监测仪表、空气污染监测器、桥梁上的压力表、移动心脏监测器、网络照相和机器人, 同时还有海洋的、航空和空中的地球图像设备

当前问题 – 如何让数据库为科研人员服务

BCS 跟许多科研人员和工程师们打交道，这些研究人员有大量的复杂的数据，有些数据是保存在数据库中，可多数只是保存在简单的文件中。

使用简单文件的问题在于

- 文件产生人为的误差
- 在不同格式的文件中找像大海捞针

我们的建议：使用数据库

- 对数据统一对待
- 无人为的误差
- 数据库本身具有的索引和比较

为什么科研人员不愿用数据库

- 直觉认为数据库很贵，很复杂但低效
- 认为数据库是为商业服务的，不是为科研服务的

当前问题 – 如何让数据库为科研人员服务

我们的处理方法

当前的数据库是可扩展，
通过开发数据库扩展产品

(比如**DataBlades**,
Cartridges), 有效的,
专门的性能可加入到数据
库管理系统

我们的解决方法

DaL 客户端 - 载入科研数据文件到数据表中

Grid DataBlade – 有效的管理数据库中的
多维格子式的数据集合

DBXten DataBlade – 对有大量的测量/传
感器数据的数据库，可提高查询性能并减少
存储空间

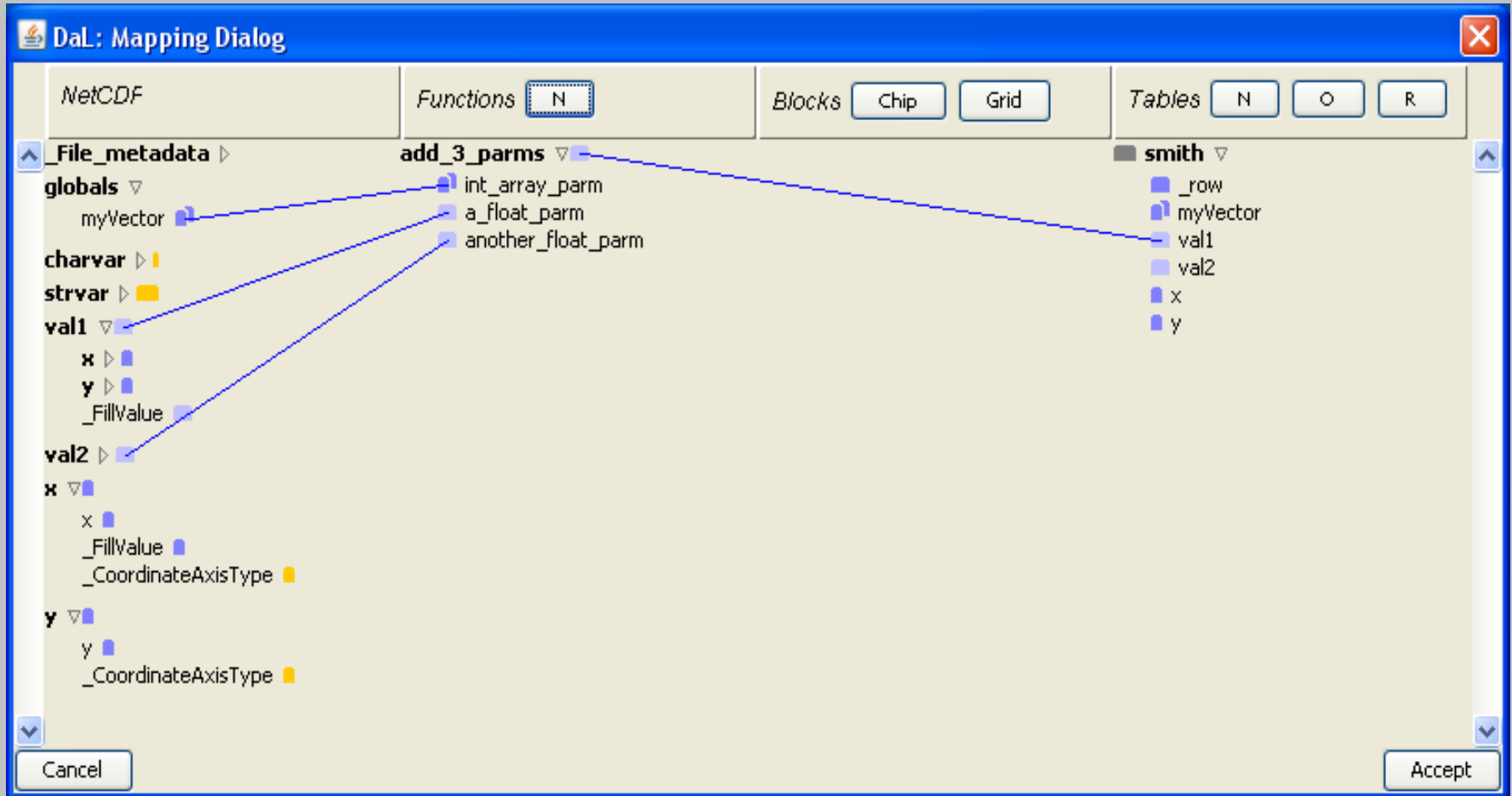
用DBXten或者Grid DataBlade与普通文件的速度比较?

最快的是**绿色**, 次快的是“**琥珀色**”, 最慢的是**红色**

提取数据的种类	提取大小	DBXten DataBlade	Grid DataBlade	NetCDF 普通文件
一维时间	81x1x1x1	1.03 sec	2.24 sec	1.30 sec
一维深度	1x66x1x1	0.07 sec	0.17 sec	0.72 sec
二维平面	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
深度-纬度部分	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
4维体积	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
平均时间		0.30 sec	0.78 sec	0.73 sec

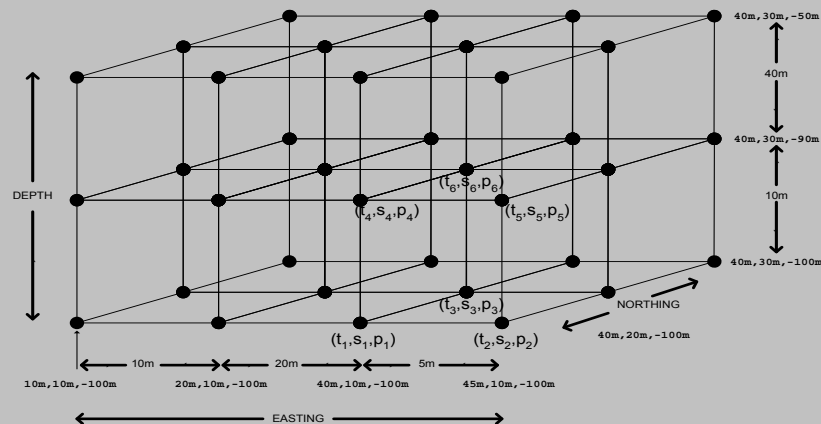
BCS 画和载入工具 (DaL)

DaL 是基于Java的数据库客户端，它能载入不同格式的科研数据文件到数据库中



BCS Grid DataBlade

- 能处理一维--四维格子式的数据
- 能载入多种格式的数据和提取数据到多种格式的文件中
- 可提取一种格式的数据然后转换成另一种格式的数据
- 提供C, Java, 和SQL应用程序接口
- 可免费下载用户和程序员文件
- 目前提供的板本与 Oracle, IBM Informix, PostgreSQL兼容





BCS DBXten DataBlade

- 适合于仪器产生的数据
- 把数据压缩后储存
- 使用基于内容的索引
- 利用已知的数据的精确位数减少储存空间
- 备有C, Java, 和 SQL应用程序接口
- 可免费下载用户和程序员文件
- 目前提供的板本与 Oracle, IBM Informix 和 PostgreSQL兼容

US Patent 8077059



DBXten 性能举例

- 以下几张幻灯片是用一组数据来说明，在IBM Informix 和 Oracle 中有和没有DBXten 的区别
- 这个基准试验包括在有和没有索引的情况下载入和查询 一组数据，这组数据是美国海洋和空气管理机构（NOAA）的海表面温度以及一些相关数据，共有50,000,000 行，26列。
- 四个有代表性的查询在5千万个数据中试验



DBXten 基准

没有索引

有索引

ORACLE

普通的

有 DBXten



空间 MB	载入 secs
1,894	2,836
848	1,410

空间 MB	载入 secs	查询 secs
3,172	3,574	205
860	1,576	85

INFORMIX

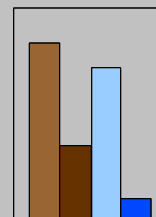
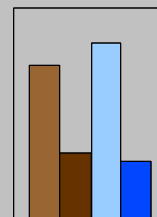
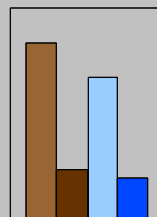
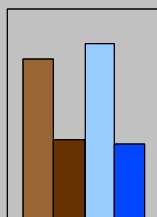
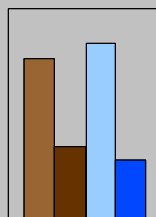
普通的

有 DBXten



2,024	3,036
733	1,318

2,556	4,069	176
736	1,367	23





DBXten可提高运行性能

	没有索引		索引		
	空间	载入	空间	载入	查询
ORACLE	2.2	2.0	3.7	2.3	2.4
INFORMIX	2.8	2.3	3.5	3.0	7.7

总结

数据库管理系统 加上 **BCS** 扩展产品比简单文件格式快速和容易, 并有数据库管理系统的优点

BCS DaL 使得研究数据文件很容易载入到数据库

BCS Grid DataBlade 符合气象学家、空气学家、海洋学家、航空权威的四维格子式的数据集合的复杂要求

BCS DBXten DataBlade 表中表、列向压缩比最新的数据库管理系统版本用行向压缩有明显的提高

谢谢你来参加这个演示,

Ian Barrodale (ian@barrodale.com)