UNB

UNIVERSITY OF
NEW BRUNSWICK

Introduction to DW/OLAP segment

Segment by Owen Kaser and Daniel Lemire.

Presented by Owen Kaser, UNB Saint John.

Overview

✔ Data Warehousing and OLAP ▮

✔ Tie to web services ▮

✔ Course structure

Idea of Data Warehouse

Collect and clean data, typically from several sources.

Reason: to enable discovery of information/patterns within.

Corporate world: often collecting old operational/transaction data.

Needs of data analysis $\neq$ operational needs. Affects schema etc.

Who analyzes?

Machine analysis via "data mining" algorithms. Not our focus.

There is a grad course at UNBSJ this term about this!

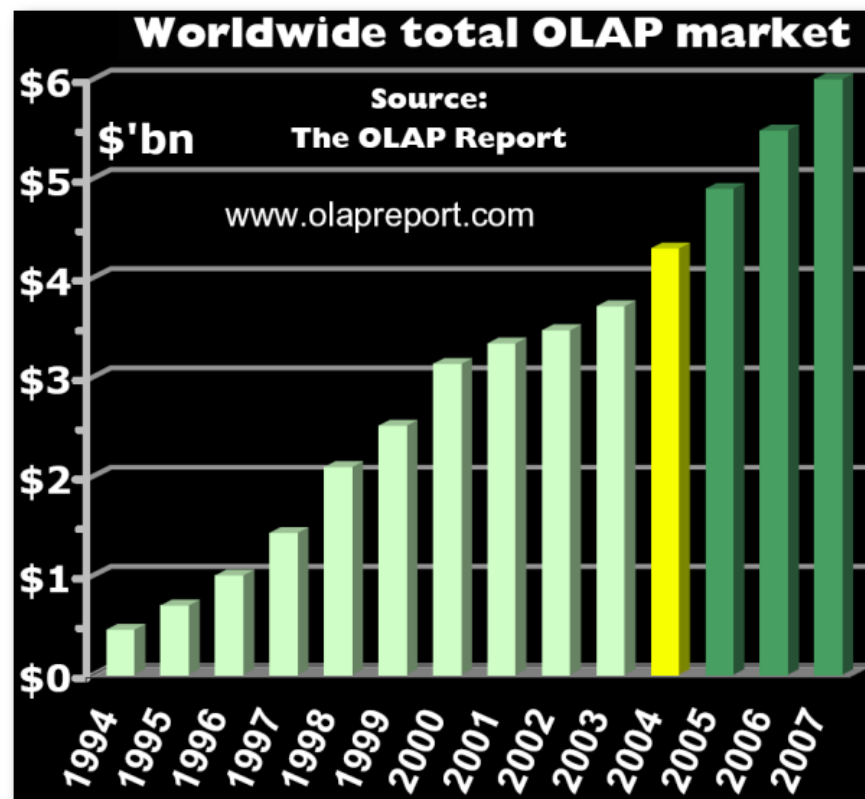*Human* exploration via *OLAP*: our focus.

OLAP

On-Line Analaytical Processing: tools to support (human) analysis.

Typical analysis queries are awkward in SQL.

Queries over huge data warehouse. Typical SQL analysis query will *not* return results in a few seconds, as required.

Use specialized query languages/ database engines. (Often an "add on" to conventional RDBMS).

OLAP is important?

15.7% growth in 2004 (most recent result)

Source: OLAP Report of 28 June 2005. (http://www.olapreport.com)

OLAP and Web Services

Provide OLAP services over the Web.

Natural fit: huge databases, database engines, specialized software need not be replicated at each site.

Recent trend, eg Microsoft's XMLA (XML for Analysis) using SOAP.

Course structure

DW and OLAP portion starts Feb. 22, after the Web Services portion.

3 assignments worth 5%, 10% and 15%.

A test, worth 20% in total. Held during undergrad reading period or exam period (April).

Course details

We'll probably use WebCT to return grades/comments.

Syllabus is on WebCT now. OLAP+DW segment uses other (offsite) Web pages. Syllabus has a link.

Mostly, lecture slides will be on line.

There are various readings for OLAP/Data Warehousing segment, but no textbook.

Three assignments

Probably 2006 will be similar to 2004, when we had

▷ small assignment: views in a RDBMS (postgreSQL)

▷ medium assignment: OLAP-type queries in SQL (using MySQL)

▷ larger assignment: MDX query language and XML/A (using Mondrian).

We provided a database server with postgreSQL, MySQL, Mondrian last time.

Some installed postgreSQL and MySQL themselves.

Is Daniel Lemire real?

Daniel is real. But he is in Montréal at UQAM.

He designed some lectures/assignments, helped design others.

He grades some things.

Technology permitting, he may lecture some.

Videoconferencing technology

The UNBSJ and NRC teleconference facilities are in some demand.

We seek alternatives. Next Monday, UNBSJ room is busy…

Class locations and the equipment used for teleconference will change during the term.

Please check your email shortly before coming to any class!