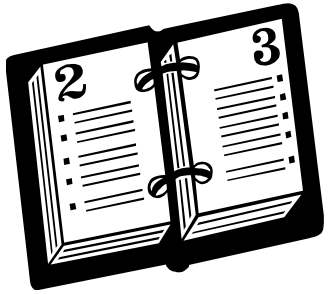
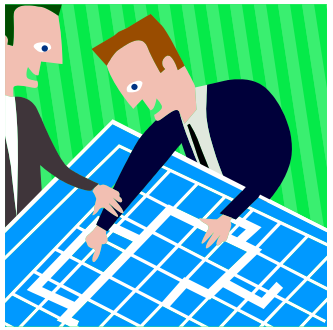
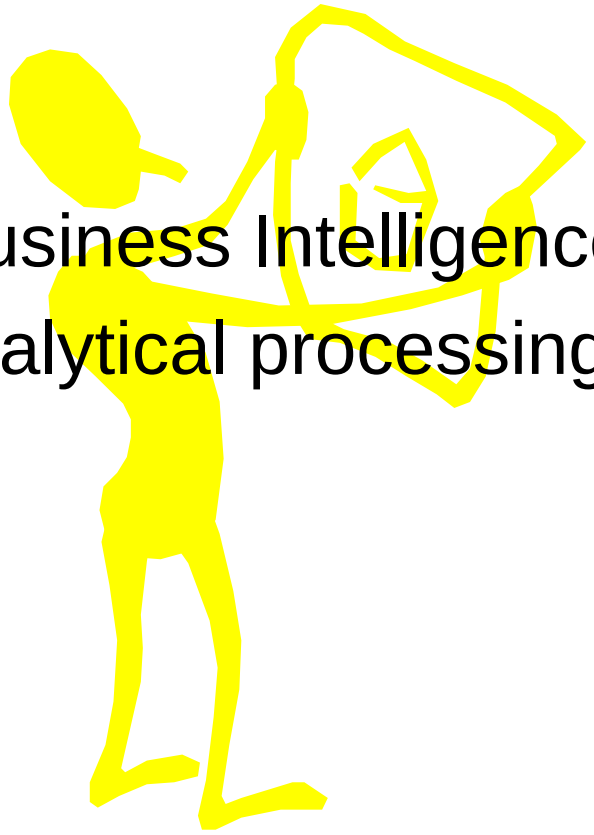
THE UNIVERSITY OF SYDNEY

INFO5990 Professional Practice in IT

Lecture 12B





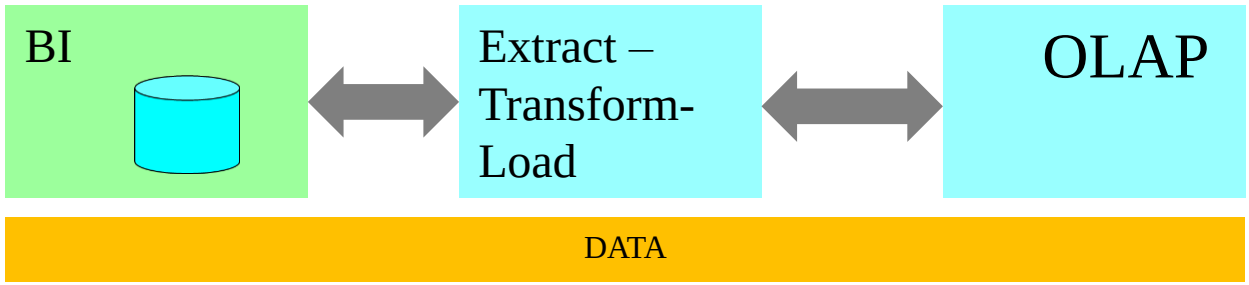


Business Intelligence

Online analytical processing (OLAP)

By the end of this lecture you will be able to:

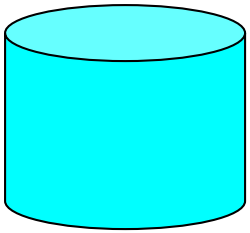
- Understand what business intelligence is
- Explain the benefits of using business intelligence tools (ETL)
- Understand the nature of online analytical processing (OLAP) Cube



What is Business Intelligence?

Business Intelligence enables the business to make intelligent, fact-based decisions

Aggregate Data



Database, Data Mart, Data Warehouse, ETL Tools, Integration Tools

Present Data

Reporting Tools, Dashboards, Static Reports, Mobile Reporting, OLAP Cubes

Enrich Data

Add Context to Create Information, Descriptive Statistics, Benchmarks, Variance to Plan or LY

Inform a Decision

Decisions are Fact-based and Data-driven

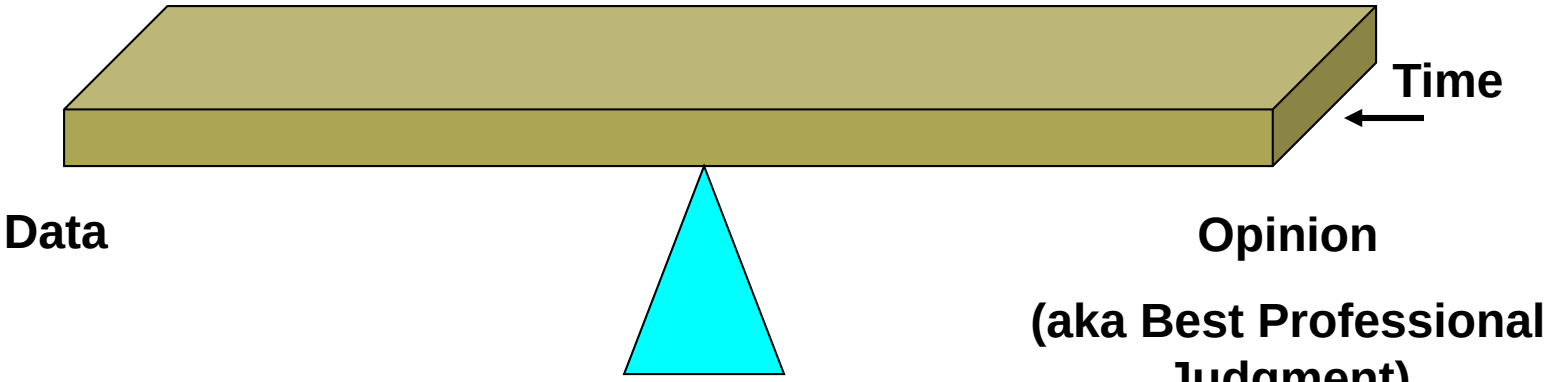
How Important is BI?

Top 10 Business and Technology Priorities for 2011:

1. Cloud computing
2. Virtualization
3. Mobile technologies
4. IT Management
5. **Business Intelligence**
6. Networking, voice and data communications
7. Enterprise applications
8. Collaboration technologies
9. Infrastructure
10. Web 2.0

Source: Gartner's 2011 CIO Agenda (aka "Reimagining IT: The 2011 CIO Agenda").

Why is Business Intelligence So Important?



Making Business Decisions is a Balance

In the absence of data, business decisions are often made by the HiPPO.

With Business Intelligence, we can get data to you in a timely manner.

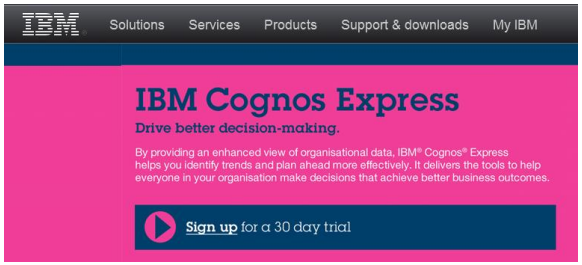
Current & Major BI Trends

- Mobile
- Cloud
- Social Media
- Advanced Analytics

Unstructured Text Processing

Example of BI in business ?

Business intelligence tools



pentaho™
open source business intelligence™

eAlliance
Customized Business Intelligence Solutions

hexarus
Business Intelligence: Transforming Your Business

ORACLE
BUSINESS INTELLIGENCE

bakerhill

Four key components of a business intelligence system

1. data warehouse containing both internal and external data
2. business analytic tools for manipulating, mining, and analyzing data
3. a set of business performance indicators for monitoring and analyzing performance
4. user interface

Benefits of business intelligence tools

- ‘One version of the truth’ – a single, reliable presentation of corporate information
- Alignment of an organization around a consistent set of Key Performance Indicators (KPIs) and Metrics
- Integrated access to multiple data sources (ERP, CRM, Spreadsheets, Budgets, etc.)
- Faster collection and dissemination of information.
- Simplified graphical presentation of KPIs and metrics
- Quicker, better, fact-based decision making

Perceived benefits

A survey of 1047 companies in Nov 2012 by Nigel Pendse showed that they thought benefits were real.

Benefit	**% of Companies Realizing Benefit**
Faster, more accurate reporting	81
Improved decision making	78
Improved customer service	56
Increased revenue	49
Savings in non-IT costs	50
IT savings	40

Q1. The Business Intelligence system in any organisation is intended to provide information for:

(A) making strategic and tactical decisions
(B) forecasting demand for goods & services
(C) optimising operational decisions
(D) facilitating effective human resource management
(E) ALL of the above

Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Score / 6
A B C D E	A B C D E	A B C D E	A B C D E	A B C D E	A B C D E	

Q2. Which of the following is NOT thought of as an essential component of a Business Intelligence system

(A) a set performance indicators
(B) a data warehouse
(C) a strategic business plan
(D) a set business analytic tools
(E) a user interface

Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Score / 6
A B C D E	A B C D E	A B C D E	A B C D E	A B C D E	A B C D E	

To summarise - BI

A Business intelligence system provides accurate information when needed, about the organisation and its environment, including a (nearly) real-time view of corporate status and performance

ETL tools

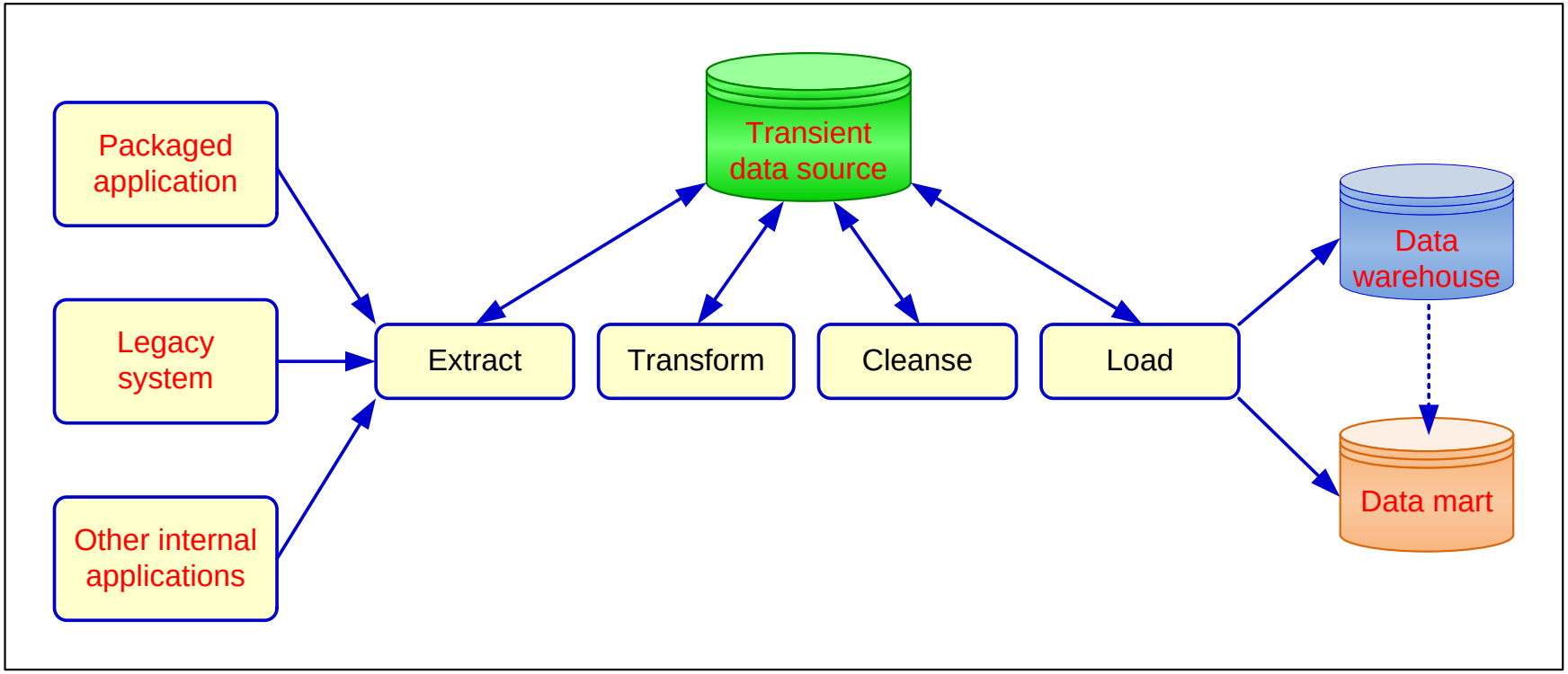
Extract-Transform-Load

- Extract data from multiple diverse data sources including those outside the organisation
- Transform data to fit operational needs, including ‘cleansing’ (quality)
- Load data into target database, data mart or data warehouse
- ETL ‘World Record’:
  5.4 TB data loaded in Under 1 Hour (Syncsort)

Click here for Intricity ETL tool (4:59)

Data Integration

The Extract, Transform and Load (ETL) Process



Data Warehouse

- A physical repository where relational data are specially organized to provide enterprise-wide, cleansed data in a standardized format
- “The data warehouse is a collection of integrated, subject-oriented databases design to support DSS functions, where each unit of data is non-volatile and relevant to some moment in time.”

Characteristics of a data warehouse

- Subject oriented
- Integrated
- Time-variant (time series)
- Nonvolatile
- Summarized
- Non-normalized
- Metadata
- Web based, relational/multi-dimensional
- Client/server
- Real-time and/or right-time (active)

Benefits of a data warehouse

- One view of the corporate data
- Allows end users to perform extensive analysis more efficiently
- Allows a consolidated view of corporate data
- Better quality data
- More timely information
- Enhanced system performance
- Simplified data access

Factors affecting the effectiveness of a data warehouse

- Suitability of dimensions as defined by IT specialists
- Quality of data
- Frequency of update
- It is difficult to suit everyone

Q3. Which of the following is NOT usually a function of an ETL tool?

(A) Removing anomalies from transaction data
(B) Capturing data at point of sale activities
(C) Accessing external data sources
(D) Reformatting data according to fixed rules
(E) ALL of these are usual ETL functions

Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Score / 6
A B C D E	A B C D E	A B C D E	A B C D E	A B C D E	A B C D E	

Q4. Which of the following statements concerning a data warehouse is FALSE?

(A) It contains cleansed data in a standardized format

(B) It represents an integrated, subject-oriented database

(C) Each unit of data is non-volatile

(D) It reflects the current status of the organisation

(E) NONE of these is false

Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Score / 6
A B C D E	A B C D E	A B C D E	A B C D E	A B C D E	A B C D E	

Online Analytical Processing (OLAP)

- Provides advanced tools for decision making
- An approach to answering ad hoc multi-dimensional analytical queries
- Part of the broader field of ‘business intelligence’
- Incorporates reporting and data mining

The OLAP cube

- ‘OLAP cube’
  - a generalisation of a two-dimensional spreadsheet
  - an array of data of three or more dimensions
  - *multidimensional dataset*
  - sometimes then called *hypercube*
- World’s largest cube 1.4 terabytes
  - Average response time 1.2 seconds

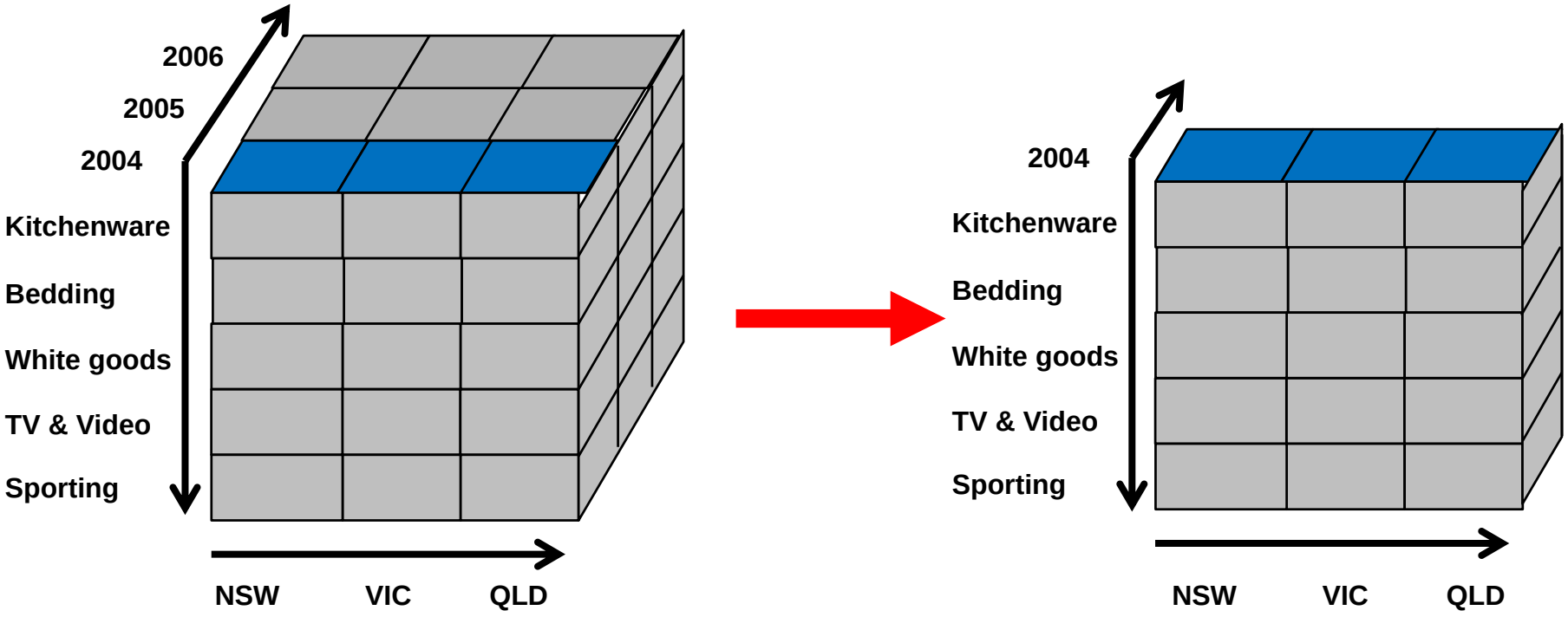
Dimensions, hierarchies & measures

- Dimensions:
  - an aspect of the business: product, customer, store, time-period, actual and budget expenses
  - described by a set of attributes, e.g. product has attributes category, industry, model number, year of introduction
- Hierarchies:
  - the attributes of a dimension can be in hierarchies: each attribute being a ‘child’ of the previous parent level, e.g. Year, Quarter, Month, Day, time of day
- Measures:
  - each cell of cube holds a number, some fact about the business, e.g. sales, profits, expenses, budget, forecast
- Grain:
  - A question of how finely grained to store data

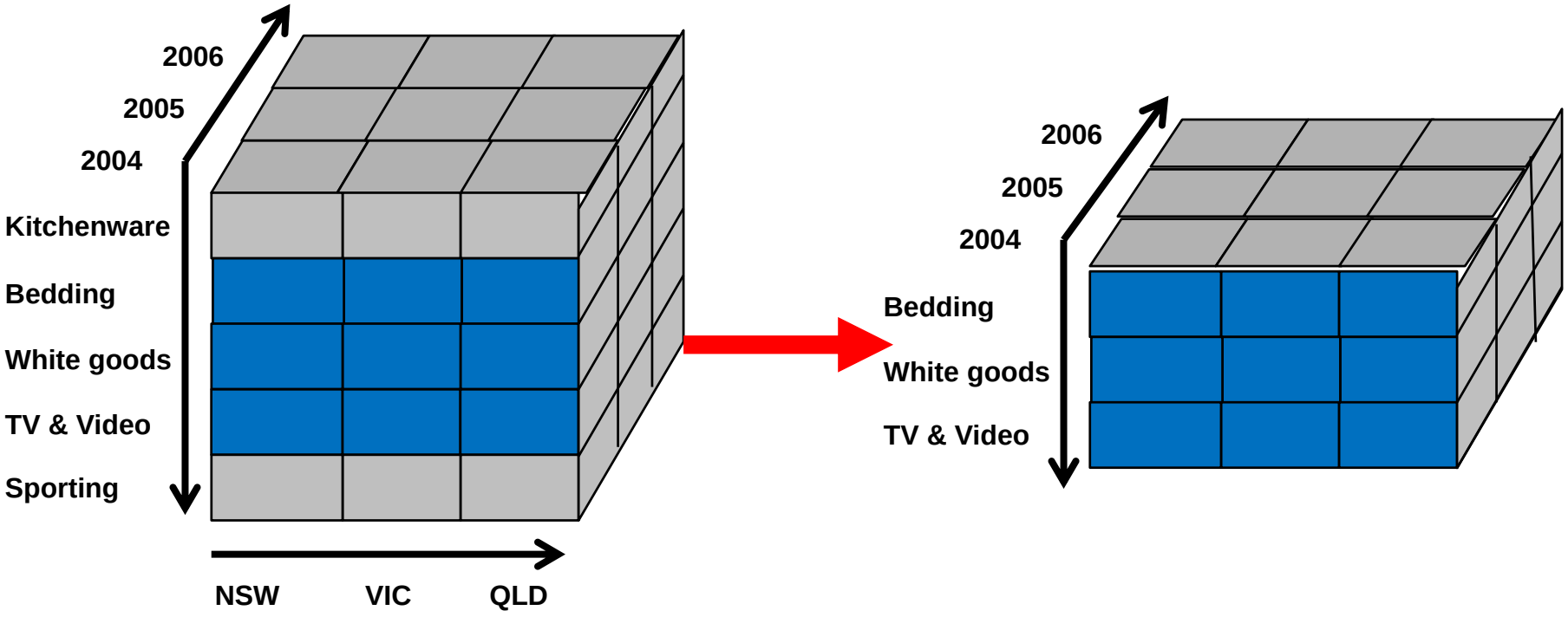
OLAP operations

1. Slicing
2. Dicing
3. Drill down
4. Roll up
5. Pivot

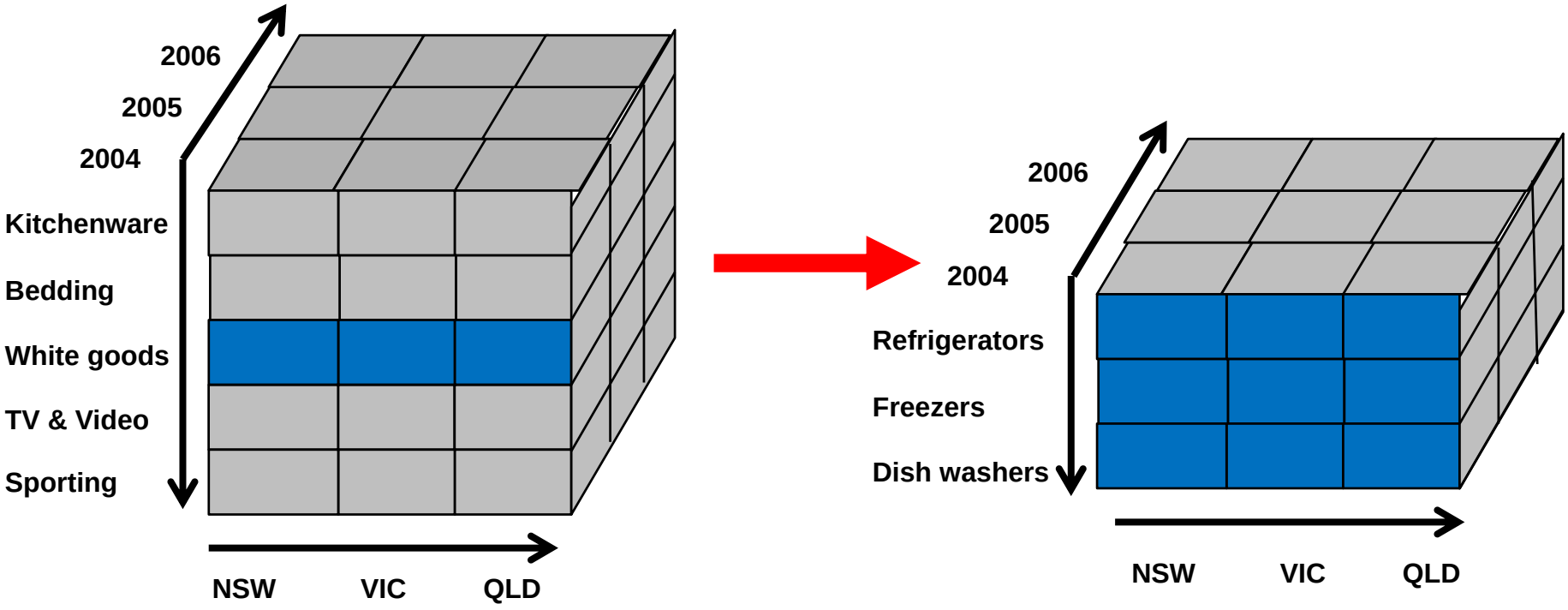
Slicing


2006
2005
2004
Kitchenware
Bedding
White goods
TV & Video
Sporting
NSW
VIC
QLD
2004
Kitchenware
Bedding
White goods
TV & Video
Sporting
NSW
VIC
QLD

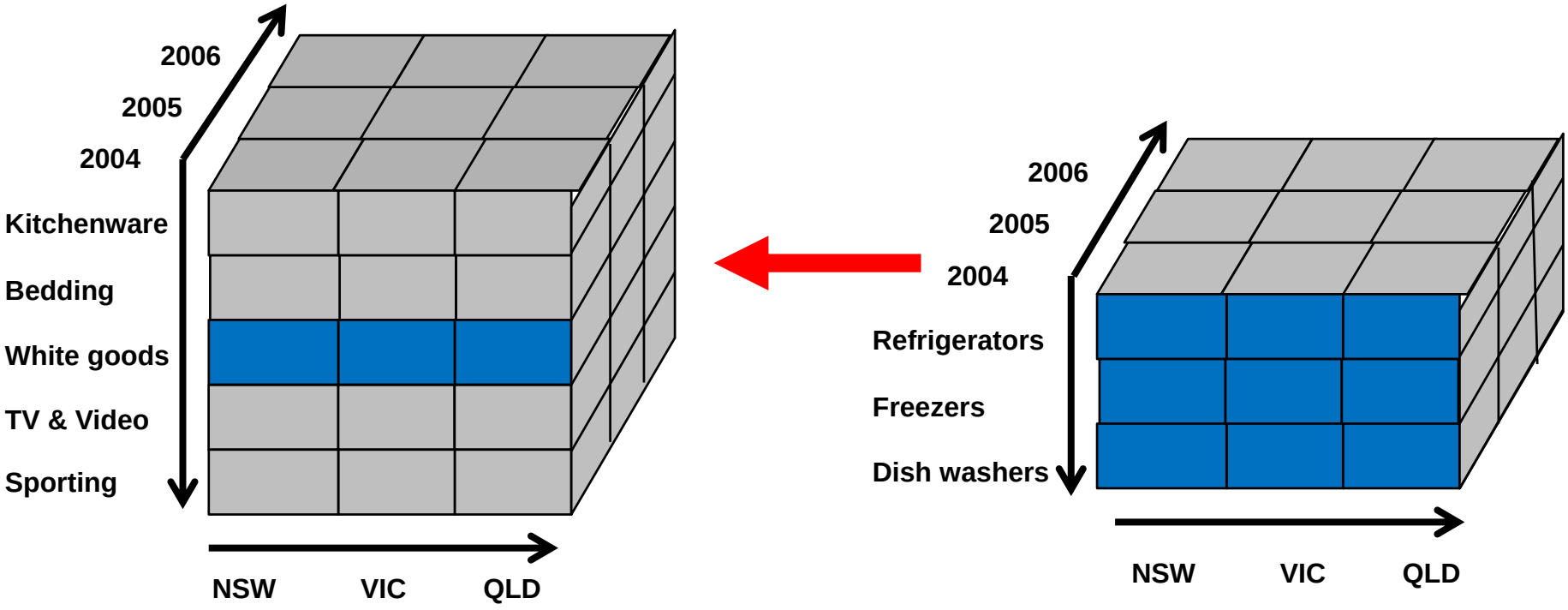
Dicing


2006
2005
2004
Kitchenware
Bedding
White goods
TV & Video
Sporting
NSW
VIC
QLD
2006
2005
2004
Bedding
White goods
TV & Video

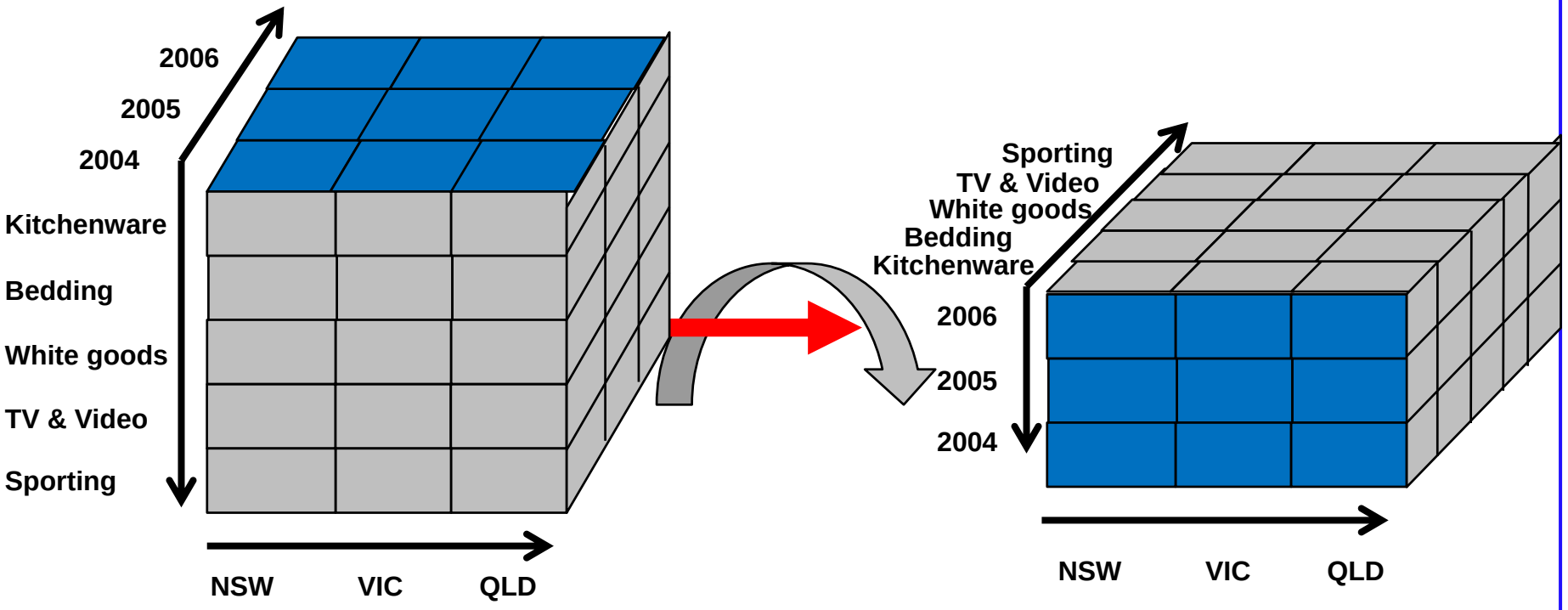
Drilling down



Rolling up


2006
2005
2004
Kitchenware
Bedding
White goods
TV & Video
Sporting
NSW
VIC
QLD
2006
2005
2004
Refrigerators
Freezers
Dish washers
NSW
VIC
QLD

Pivoting


2006
2005
2004
Kitchenware
Bedding
White goods
TV & Video
Sporting
NSW
VIC
QLD
Sporting
TV & Video
White goods
Bedding
Kitchenware
2006
2005
2004
NSW
VIC
QLD

OLAP operations

- Slicing – selecting a ‘rectangular’ subset by choosing a single value of one dimension
- Dicing – selecting one or more specific values from a dimension
- Drill down – access the ‘child’ values of a parent
- Roll up (consolidation) – summarize data along a dimension, or calculate a derived value, say profit
- Pivoting – rotate cube in space to see a different ‘face’

Q5. Which of the following statements concerning the OLAP data cube is FALSE?

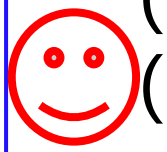
(A) The cube is a copy of the organisation data base

(B) Each view of the data in the cube is called a dimension

(C) Dimensions are organised into hierarchies

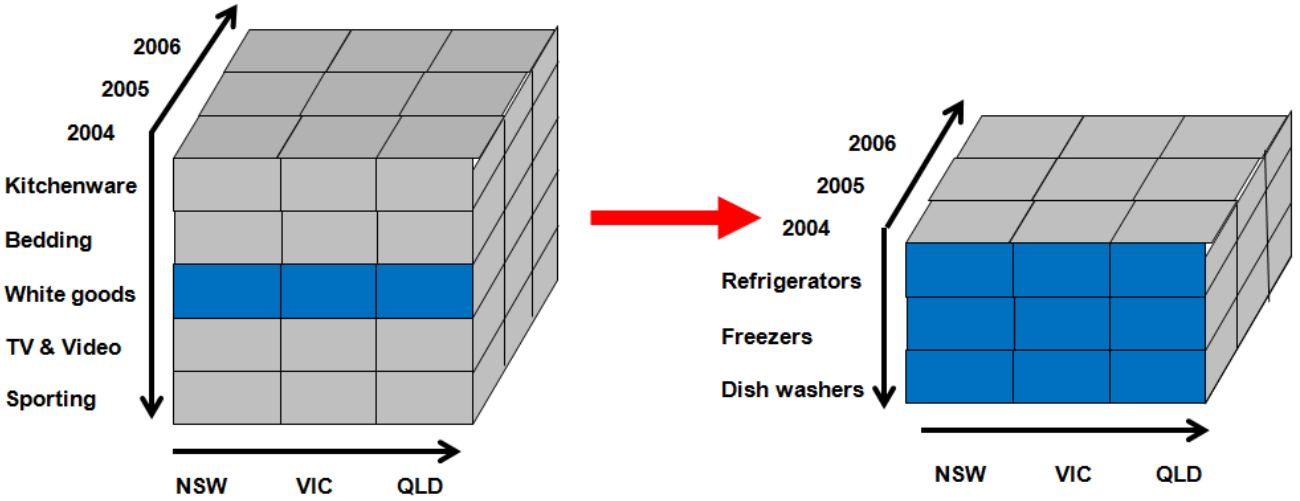
(D) Data in the OLAP must be updated regularly

(E) NONE of these is false



Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Score / 6
A B C D E	A B C D E	A B C D E	A B C D E	A B C D E	A B C D E	

Q6. Which of the OLAP operations is represented by the following diagram?



(A) Slicing
(B) Dicing
(C) Drilling down
(D) Rolling up
(E) NONE of these

Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Score / 6
A B C D E	A B C D E	A B C D E	A B C D E	A B C D E	A B C D E	

What's next?

Unstructured data
Big data

Unstructured data

- 85 percent of all business information exists as unstructured data (Merrill Lynch, 2003)
  - news, reports, letters, e-mails, memos, notes from call centres or customer support, user groups, chats, surveys, white papers, marketing material, research, PowerPoint presentations, Web pages
- white collar workers spend 30-40% of their time managing documents (Gartner, 2003)
- 2 billion Web pages created since 1995
- 200 million pages being added every month*
- “Semi-structured”: In many cases meta-data exists which can be used to classify and correlate, e.g. HTML, XML

* International Data Corporation (IDC) An organisation providing market research, analysis and advice on information technology and telecommunications

The next move: Content intelligence

- Advanced search tools
  - draw on context and meta-data to specify more precisely
- Classification
  - placing unstructured documents within a taxonomy
  - automated tools for building and maintaining taxonomies
- Discovery software
  - generates meta data from documents and classifies the documents
  - Example: analyzing product defect information for heavy equipment to tailor parts and services

Bonus Question 7

Write down your score /7

Which of the following is/are an e
"unstructured data"?

(A) Data representing discontinued products
(B) Free text comments in the 'comment' column of an order screen
(C) Data included in tags in an HTML document
(D) Price and availability data in an email
(E) NONE of these

Question 1	Question 2	Question 3	Bonus question	Question 5	Question 6	Score / 6
A B C D E	A B C D E	A B C D E		A B C D E	A B C D E	

Next week

- Exam Tips –tell your friends
- Discussion & Debate
- Guest Lecture
- Practice the sample exam questions for next week's class

Last class – hurray !