Le management d’une immense banque de données

Stéphane Mouton
SST Department manager
16ème rencontre EVELINE M ARKIEWICZ
Brussels, 14/11/2015
In the beginning...

First compare quality. Then compare cost.

Morrow Designs’ 10 megabyte hard disk system: $3,695.

MORE MEMORY LESS MONEY.
Compare Morrow Designs’ DISCUS™ M26™ hard disk systems to any system available for $100 or Crayson machines. First, compare features. Then, compare cost per megabyte. The M26 works out to under $100 a megabyte, and the M10 is about half the cost of competing systems.

COMPLETE SUBSYSTEMS.
Both the M10 (8”) and the M26 (14”), are delivered complete with disk controller, cables, fan, power supply, cabinet, and CP/M™ operating system. Pick your choice: 10 Mb at $3695 or 20 Mb at $5,995. That’s single unit. Quantity prices are available.

BUILD TO FOUR DRIVES.

Sausalito, CA (415) 332-4443.

extensys corporation

64K for $1495
The most cost effective products for your microcomputer.

AND NOW, MULT-X/O
Mult-X/O is an I/O controller that allows multi-terminal and multi-purpose use of " Intelligence Controlled" terminals.
Then IT resource became affordable
But data production explodes
What is the status now?

• Network lags behind

“Never underestimate the bandwidth of a station wagon full of tapes hurtling down the highway.”


Fiber network is crucial
What is the status now?

• Most of “usual” commercial offering don’t stack up seamlessly
Need for scalability

• The capability of IT systems to adapt to system demand at a predictable, ideally linear, pace.
Scalable storage to cope with volume

• When a single IT resource can’t handle all the data anymore
  • Distributed solutions, based on many cloned (and possibly inexpensive) nodes
Scalable databases

- New ways to structure data storage
  - NoSQL (Not Only SQL databases)
Process and analyse data in a scalable way

- What if you have possibly an infinite number of computers to process your data?
  - Your data analysis doesn’t change, the way you express it does.
Process and analyse data in a scalable way

- New programming frameworks for data analysis
Big Data

- A relative concept, linked to data
  - Volume
  - Variety
  - Velocity
Volume: data integration approaches

• 3 main models ensuring data bases autonomy
  • Full Data Integration (Data Warehousing)
    • Not suitable for data with a very short-term relevance
    • Can lead to improperly formed data warehouses, not focused on the analysis topic
  
  • Federated databases
    • One global schema
    • One main data format/model
    • Mostly static

  • Mediator/Wrapper
    • Support dynamic sets of data sources
    • High flexibility: schema can evolve at query time
    • Wrappers to handle heterogeneous data format

Tightly coupled
  • Administrator builds a unique FDB schema

Loosely coupled
  • Component DBs out of control
  • Import schemas as view over data sources
Data integration leads to data variety

• Variety in content type
Data integration leads to data variety

- Variety in data format
Data integration leads to data variety

• Variety in data model

Latest Semantic Analysis of Textual Data

Latest Semantic Analysis of Textual Data: The paper presents an overview of the current state of research in data integration and the mathematical apparatus required to perform it effectively. The focus is on data models, which are the foundation for the development of data integration systems. The paper discusses the problems of data integration and presents an overview of the existing research in the field.

Variety in data model

Latest Semantic Analysis of Textual Data: The paper presents an overview of the current state of research in data integration and the mathematical apparatus required to perform it effectively. The focus is on data models, which are the foundation for the development of data integration systems. The paper discusses the problems of data integration and presents an overview of the existing research in the field.

Variety in data model
Data integration leads to data variety

• Variety in data source
Challenges of data variety

• Identify relevant data sources

• Extract, clean and store data

• Understand the data (semantics, meta-data)

• Deliver information
Data management also applies

• Not only in terms of cost
Aéropole de Charleroi-Gosselies
*Bâtiment Éole*
Rue des Frères Wright, 29/3
B-6041 Charleroi
Tel: +32.71.490.700
Fax: +32.71.490.799

**www.cetic.be**
info@cetic.be

*linkedin.com/company/cetic*
*twitter.com/@CETIC*

---

**Questions?**

**Stéphane Mouton**
*SST Department manager*
Tel : +32 71 490 726
Mob : +32 475 76 78 50
stephane.mouton@cetic.be