Building a Data Warehouse For Scalability and Flexibility

Ray Welsh

Business Intelligence Marketing Manager

Informix Software Ltd.

Inform*i*x

Agenda

- Informix examples of VLDB
- Drivers of growth & evolution
- Failure of traditional methods to meet requirements of growth and evolution
- Technology and methodology directions
- What should research community be working on

Inform*i*x

Very Large Informix Databases

- First Union Corporation
  - planning for 27 terabytes
- Catalina Marketing
  - 95m records/night
- Telecom Italia
  - 2.5 terabytes
- AOL
  - 4 terabytes & 1000 users
- MCI Colorado
  - 6 terabytes
- Longs Drug Stores
  - 10,000 queries per day

- Kraft
  - 3,000 users
  - 150,000 queries per week
  - 95% under 1 second
- Fleet Bank
  - 4 terabytes
- SmarTone
  - one of Hong Kong's largest DW
- Dominos Pizza
  - 1.2 terabytes
- Sears Roebuck and Co
  - 3.1 terabytes

Informix

Very Large Data Warehouses

- Large web site will get 10,000,000 hits per day
- As many as PacificBell (telco) will have calls per day
- A web log contains as much, if not more information, as a telco CDR
- There will be thousands of web sites of this size

Informix

Growth in Capacity

- Capacity as defined by TPC-H
  - query throughput, complex queries, large data set
- 5 aspects of growth
  - volume of data
  - number of users
  - complexity of analysis
  - reduction in load times
  - faster response times

Inform*i*x

Growth in Capacity

- Business drivers of growth
    - realisation that all transaction data is important
    - ability to store and process it all within budget
    - more customers, more products, more sales channels, more customer service
    - decision making pushed down
    - benefits of centralised storage become apparent
    - improved understanding of ROI
        - CRM is the killer app DW needed
    - market pressure to compete in i.Economy

Informix

Growth in Capacity

- Technical drivers of growth
  - skills and technology become available to store and process larger volumes
  - the web is generating volumes of data that grow at increasing rates
    - Pre-web: data doubled in 50 years, then 10 years
    - Post-web: data doubled in 1 year.
    - Tomorrow, data volumes could double in 1 hour
  - move from IT push to business pull

Inform*ix*

Why Detail Data is Important

- Push decision making down the organization's hierarchy
  - Information is needed to make decisions
  - Local autonomy will save money
    - local ordering is more accurate
- Better customer service
  - Transaction history
  - Product shipment tracking
- Deeper understanding of the business
  - Strategic data mining

Inform*i*x

Evolution

- In addition to the volume growth, business evolution is driving requirements for greater flexibility
- The business environment is ever changing
  - now approaching web speed
- Changing business environment alters how a successful organisation competes
- Examples of changing environment, leading to greater demand for management and utilization of data;

Informix

Evolution

- Changes to business environment
  - Business velocity is increasing
    - shrinking product cycles
    - faster decision making
  - More power to customers
    - 'Pro-sumers'
      - more choice
      - more information
      - empowered to act
  - Need to offer added value - relationship

Informix

Evolution

- New competitors
  - With the web, they are everywhere
  - Web provides low cost to start-up
  - Business intelligence shows niche markets not served
    - coupled with short development times, this speeds up the rate of change within a market

Informix

Evolution

Example

1. Mobile phone operator identifies a gap in the market by analyzing customers and their usage
2. Design a new product (tariff)
3. Market to selected audience
4. Monitor sign-up rate and profitability

- All these steps require rapid processing of large volumes of detailed data.
- Also valid in other markets
  - Finance, retail, utilities, government, manufacturing

Inform*i*x

What Doesn't Work

- Single, monolithic data warehouse
  - big bang approach
    - too expensive
    - too long
    - too complex

Inform*ix*

What Doesn't Work

- Multiple, independent data marts
  - individual point of pain solutions
  - no re-use of data & processes
  - no shared definitions
  - 'Islands of information', in a new format

Informix

What Does Work

- New technologies and methodologies that manage greater growth & flexibility
- Managing large and rapidly growing volumes of data
- Providing flexibility to serve business environment in web speed
- Combine best of both approaches
  - Rapid deployment of data marts
  - Coordination from data warehouse

Informix

What Does Work

- Data Warehouses
  - Improving ETL from sources
    - From ERP, web, other known sources
    - Easier, quicker, cheaper
  - Meta data management
  - More cost effective
    - Use of hardware
    - Proven ROI figures

Informix

What Does Work

- Data Marts
  - Responsiveness to user requirements
    - Rapid deployment
    - Flexibility
    - Volume
  - Powerful analysis
  - Some will have a shorter lifetime
  - Dependant on data warehouse
    - Conformed definitions
    - Consistent answers
    - Lower maintenance

Informix

An Architecture That Works

Data Staging

Data Marts
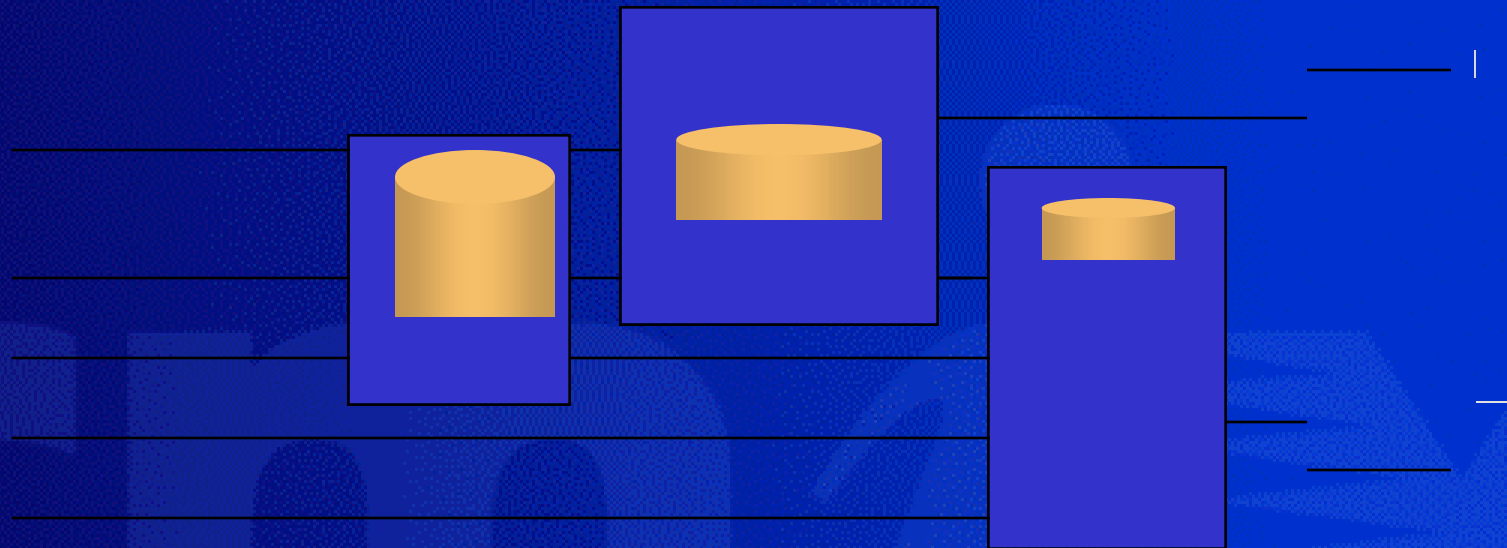
Warehouse Management

VISA

Informix

What Would Help

- What the research community should be working on
- Business modeling methodologies specific to business intelligence
- More power
  - Indexes
  - Analysis techniques
  - ETL performance
  - High Availability
  - Data Storage
    - Partitioning of data
    - Near-line storage

Informix

Summary

- Growth in volume is inevitable
- Growth occurs in several dimensions
- Business has become more dynamic, so have the demands on business intelligence
- Flexibility is as important as scalability

Informix

Contact Details

Ray Welsh

Informix Software Ltd.

rayw@informix.com

www.informix.com

Inform*i*x