Business Intelligence and Datawarehousing on iSeries

A presentation for **common** Europe Luxembourg

DB2 Information Management Software

29/04/2004
Luxembourg

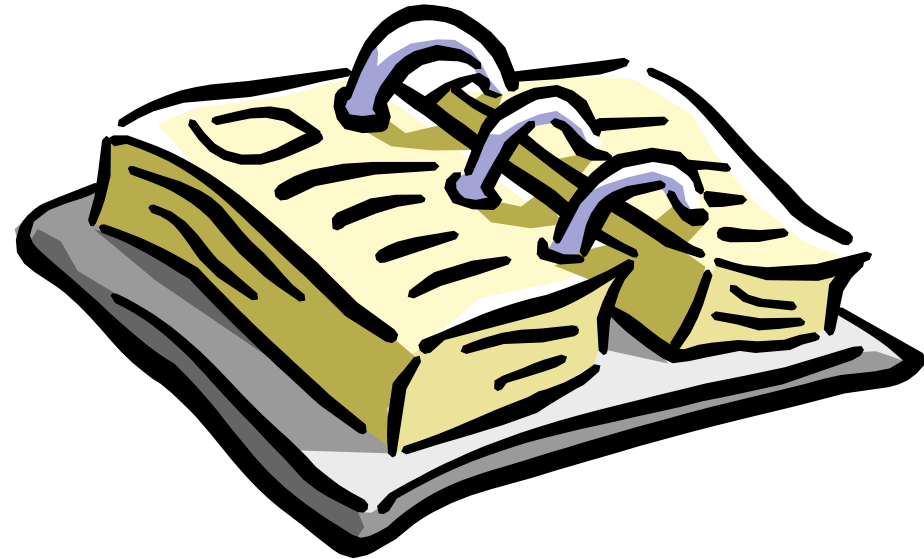**Business Intelligence**

Laurent ZELLER

Information Management Software Technical Sales Specialist

laurent_zeller@be.ibm.com

@business on demand software

Agenda

- **Concepts and General Best Practices**

- **History – Evolution – Trends**

- **BI/DW on iSeries**

- **IBM value proposition for BI**
  - ▸ On iSeries
  - ▸ On UWL systems

- **Demos**

Concepts and General Best Practices

DB2 Information Management Software

Laurent ZELLER

Information Management Software Technical Sales Specialist
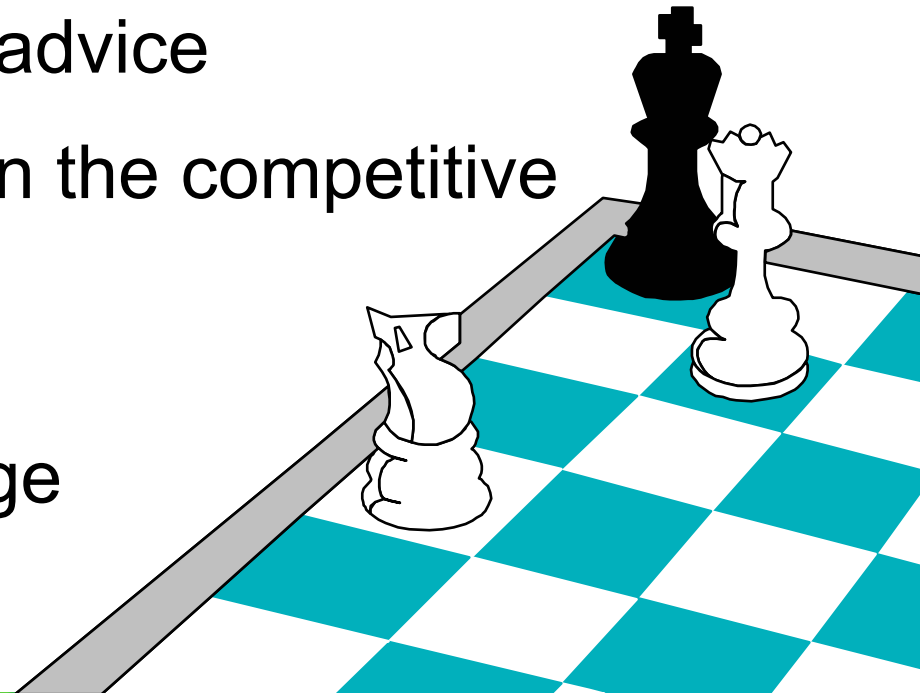
laurent_zeller@be.ibm.com

@business on demand software

Terminology

- Business Intelligence ?

The Intelligence function offers:

▶ strategic decision support

▶ objective analysis and advice

▶ forecasting of change in the competitive environment

▶ options for action – positioning for advantage
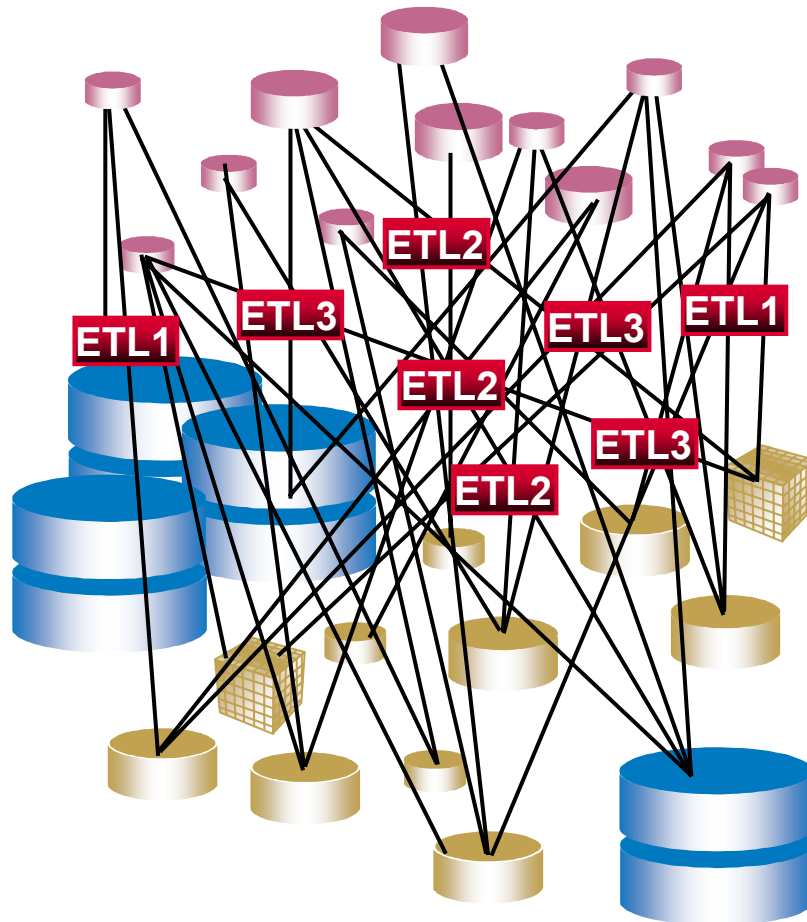
Terminology

- **Datawarehouse / Datawarehousing**

  - Concept of extracting data from different databases, consolidating, cleansing and storing that data in a format optimal for reporting or analytical purposes

  - Subject oriented, integrated, time variant, non-volatile collection of information in support of decision making process (Bill Inmon)

  - a repository of business information used for analysis, separate from the "active" business system files
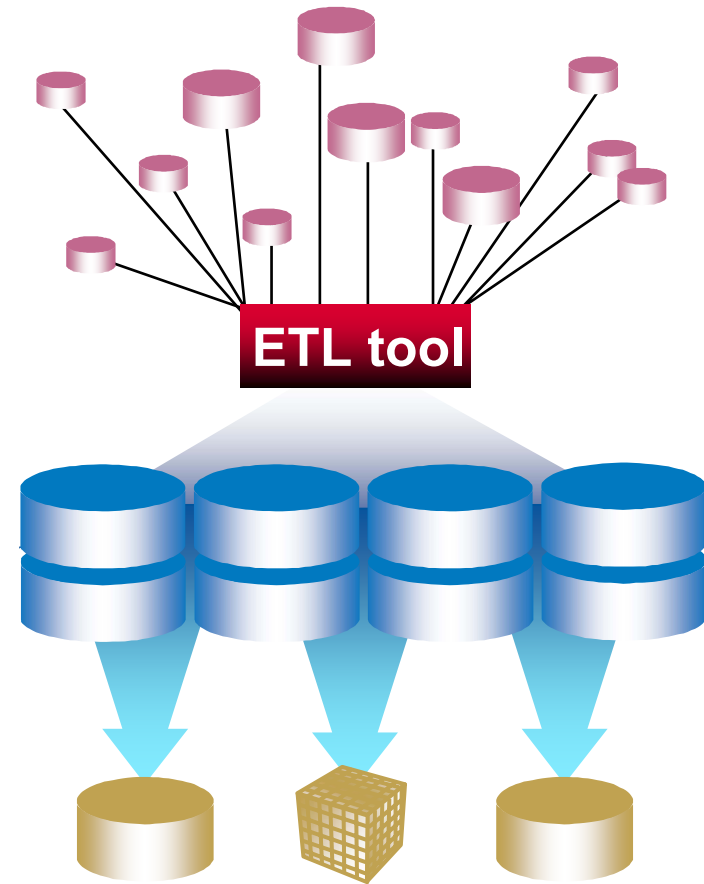
- **Datamart**

  - business-focused data warehouse.

  - Usually focuses on a particular business issue

  - Often aggregated

  - Dependant or independant

Data Marts & Data Warehouse proliferation

ETL2

ETL3

ETL1

ETL3

ETL1

ETL2

ETL3

ETL2

ETL tool

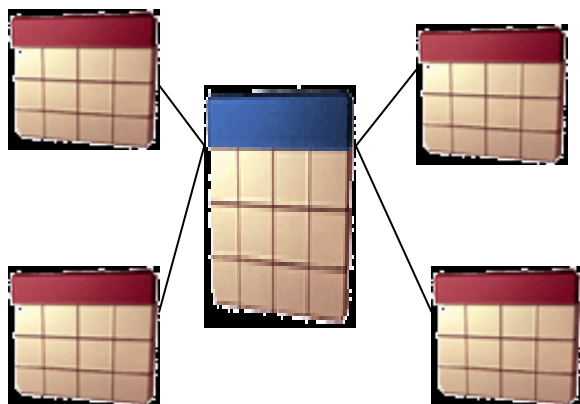**Multiple Sources; "Stovepipe" Data Marts; No Central Metadata**

Dependent Marts

Terminology
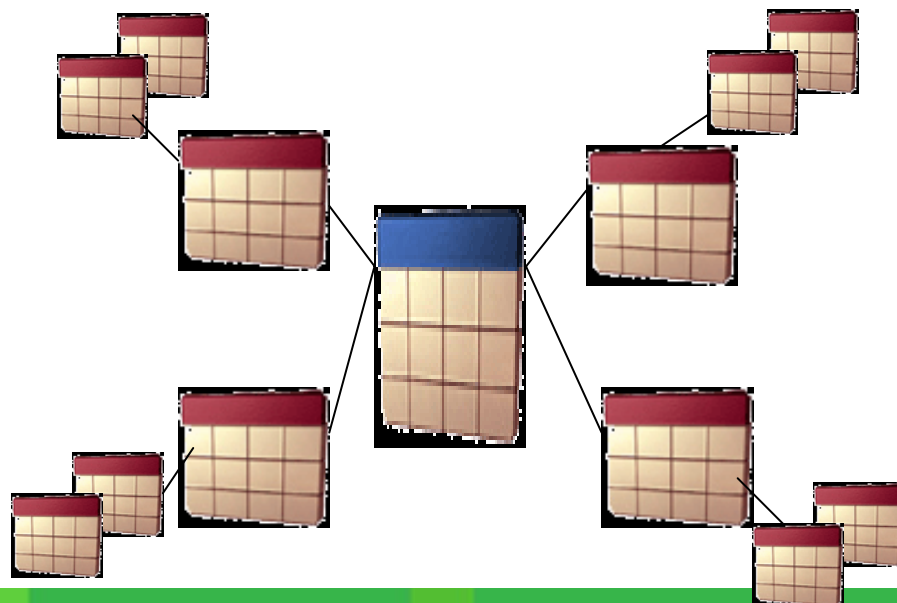
- Database model : Star
  - A.k.a. "dimensional model"
  - Dimensions surrounds a fact
  - Dim. contains descriptive information
  - Fact contains basic numeric transac. level info
  - Why use a star schema for BI ?
    - efficiency : query speed
    - schema matches user perception of business

- Database model : Snowflake
  - Extended/stretched star model
 - Normalization is introduced here

Terminology

- **Operational Data Store (ODS)**
    - an integrated database of operational data. Its sources include legacy systems and it contains current or near term data. An ODS may contain 30 to 60 days of information, while a data warehouse typically contains years of data.
    - Unlike a datawarehouse, an ODS may well
        - have updates applied to the contents either directly by users or by receiving transactions from the source systems
        - Only contain detailed data at the lowest level of granularity

- **RDBMS vs MDDB**

Characteristics	RDBMS DW	MDDDB
Typical Operations	Report	Analyze
Level of Analytical requirements	Medium	High
Screens	User-defined	User-defined
Amount of data per transaction	Small to large	Large
Data level	Detail and summary	Mostly summary
Age of data	Historical and current	Historical, current and projected
Orientation	Record	Arrays
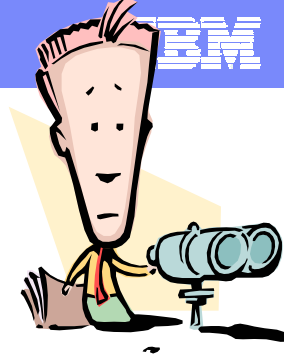
Terminology

- Metadata

  ▸ Data about data

  ▸ It keeps track of what is where in the datawarehouse

  ▸ It is with these metadata that the most effective use of the datawarehouse is made

  ▸ Of 2 types :

    • Technical to ease consistency of use by DBA's and developers
    • Business to help find, understand and use information provided by end-users

  ▸ Typically, metadata stores

    • The structure of data as known to the programmer
    • The stucture of data as known by the analyst
    • The source feeding the datawarehouse
    • The transformations of data as it passes into the datawarehouse
    • The history of extracts

Terminology

- OLAP (MOLAP, ROLAP, DOLAP, HOLAP, …schmolap =;-))
  - ▶ On-line Analytical Processing
    - As opposed to OLTP
    - Tools and applications that support interactive analysis of multidimensional data.

  - ▶ MOLAP – Multidimensional OLAP – use of a proprietary array-based storage
    - An application (or datamart) is built as a cube, which resides on a server.
    - Access is made via clients using OLAP Specific functions

  - ▶ ROLAP – Relational OLAP
    - Provides OLAP type processing by generating SQL against relational tables.

  - ▶ DOLAP – Desktop OLAP - Data is extracted from relational or other sources and built into a 'micro-cube' against which OLAP can be performed.

  - ▶ HOLAP – Hybrid OLAP - provide multidimensional analysis simultaneously of data stored in a multidimensional database and in an RDBMS
    - combine MOLAP's fast response time and
    - analytical capabilities with ROLAP's dynamic access to detailed data.

Terminology

- Dimension

  ▸ A point of view on the data available in the datawarehouse/datamart.

  ▸ A list of members, all of which are of a similar type in the user's perception of the data.

    - all months, quarters, years, etc., make up a time dimension
    - all cities, regions, countries, etc., make up a geography dimension

  ▸ Dimensions offer a very concise, intuitive way of organizing and selecting data for retrieval, exploration and analysis

- Facts / Measures / KPI

  ▸ A data that describes specific events within a business, such as bank transactions or product sales

  ▸ Organized into one or several fact tables, the central table in a data warehouse schema that contains these numerical measures and keys relating facts to dimension tables

Terminology

- Hierarchies
  - ▶ Organization of dimension members based on parent-child relationships
  - ▶ Typically a parent member represents the consolidation of the members which are its children

- Drill down/up
  - ▶ A method of exploring detailed/summary data using the defined dimension hierarchies
  - ▶ Drill levels depend on the granularity (detail) of the data in the datawarehouse

- Slice & Dice
  - ▶ A method of exploring the available information to view different perspectives

by January
by all products
accounts

by Product A
by all time
accounts
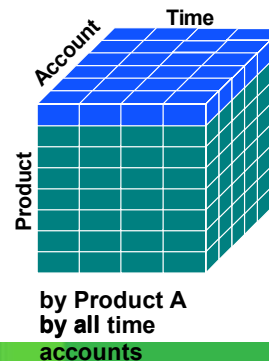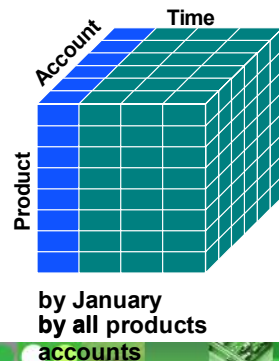
BI : **Who** – What – Why ?

- **The Sales/Product Manager:**
  - ▶ track top performing sales by region, sales rep, product, and time, et.al.?
  - ▶ track performance on sales, costs, and PROFIT?
  - ▶ localize merchandising decisions to account for trends at the store level?
  - ▶ ensure product availability WHEN the customer wants it?
  - ▶ What products are selling over the net vs. brick and mortar and why?

- **The CFO:**
  - ▶ monitor company performance against the annual budget
  - ▶ reduce time required to build corporate reports for shareholders?
  - ▶ determine profitability of company, divisions, etc.?
  - ▶ compare revenue growth rates to 'best in industry'?

- **The Marketing Department:**
  - ▶ measure the success of promotions?
  - ▶ optimize funding for promotions by creating better targeted marketing programs?
  - ▶ understand most profitable customers, and which customers ARE NOT?
  - ▶ understand the response rate of direct marketing programs?
  - ▶ How are customers navigating through our websites?
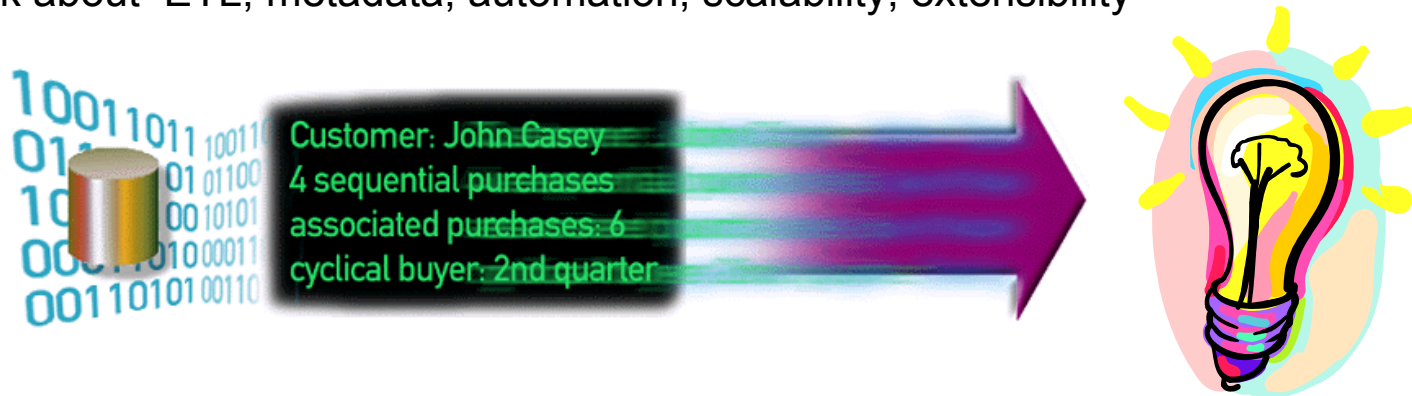 - ▶ How can we personalize this?

- **The Director of Operations:**
  - ▶ ..purchase the right products and the right amount of materials from the right supplier?
  - ▶ ..track trends by monitoring open orders to forecast short term product demand?
  - ▶ ..review key performance indicators each morning to see trouble spots

BI : Who – **What** – Why ?

- "Business Intelligence is the process of transforming data into information and through discovery transforming that information into knowledge" (Gartner Group)

- It means using your data assets to make better business decisions. It's about access, analysis, and uncovering new opportunities.

- It is about providing business users with a 360-degree view of their business environment

- more than just a front-end tool
  - ▶ think about ETL, metadata, automation, scalability, extensibility

Customer: John Casey
4 sequential purchases
associated purchases: 6
cyclical buyer: 2nd quarter

BI : Who – What – **Why** ?

- Massive amounts of Data Accumulated
- Operational systems collecting TONS of transactional data about the business

    BUT...

- Data Inaccessible by Business Users because
    ▸ Accessing the data was too slow or too intrusive
    ▸ Tools to access the data were scarce or too complex for users
    ▸ The data was NOT stored in presentable formats for analysis
    ▸ The timing of the data was not appropriate for making business decisions
    ▸ Limited historical data was kept "on-line" - no trend analysis
    ▸ Data was stored in separated incompatible islands...

- No single version of the TRUTH!

Some Fun – BI cheese

Best Practices

- Business Sponsorship of the project
  - ▶ Create a business case and make sure it is sponsored by business lines

- True and close collaboration between Analysts/Business users and IT
  - ▶ Requirement gathering / analysis
  - ▶ Functions needed (Finance: plan, budget, reporting / Sales : forecast / Marketing : Mining)
  - ▶ Solution mapping

- Coordination effort between people, process, technology and data
  - ▶ BI center of excellence (BI CoE)
  - ▶ Provide a common set of resources, models and tools meeting the broadest set of business needs while taking advantage of existing organizational expertise, standards and infrastructure.

Best Practices

- Limit scope of first implementation then grow it through new subject cycles
  - ▶ Start small
    - With a focused project quickly implemented and yet deliver substantial benefits
    - Which area provides the greatest impact ? + aim at delivering short-term, incremental results

  - ▶ Think global
    - Create a high-level roadmap tied to corporate objectives
      - − Define the org. priorities and what it wants to measure
    - Identify the right metrics
      - − Which metrics best serve as KPI's
      - − Who owns them, how are they calculated, what are the thresholds for alerts

  - ▶ Use Incremental deployment

- Then enlarge the BI reach
  - ▶ Don't limit the BI solution to analysis of operational efficiencies
    - = scorecard, dashboard on what has happened / in terms of cost
  - ▶ Also think about the external forces impacting the organization
    - Focus on what can be done to improve / increase the revenue pipeline

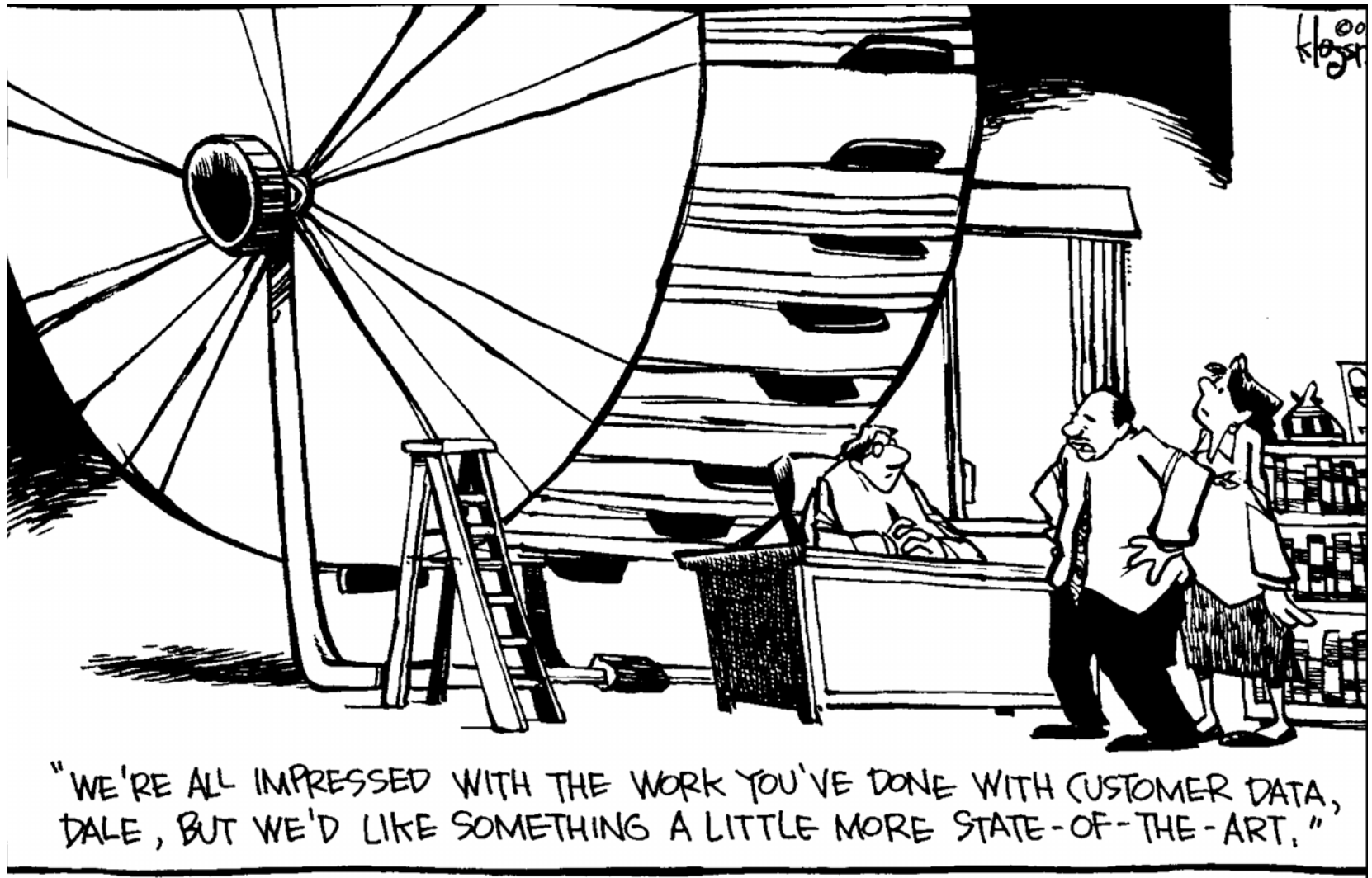# History – Evolution - Trends

DB2 Information Management Software

Laurent ZELLER

Information Management Software Technical Sales Specialist

laurent_zeller@be.ibm.com

@business on demand software

Some fun … First generation datawarehouse ?

Business Intelligence
The Telescope of the Past

Business intelligence started out as a **means to use historical data collected over a period of time** to **predict trends**. Analysts would spin through a mountain of data, and use their business knowledge to determine future strategies.
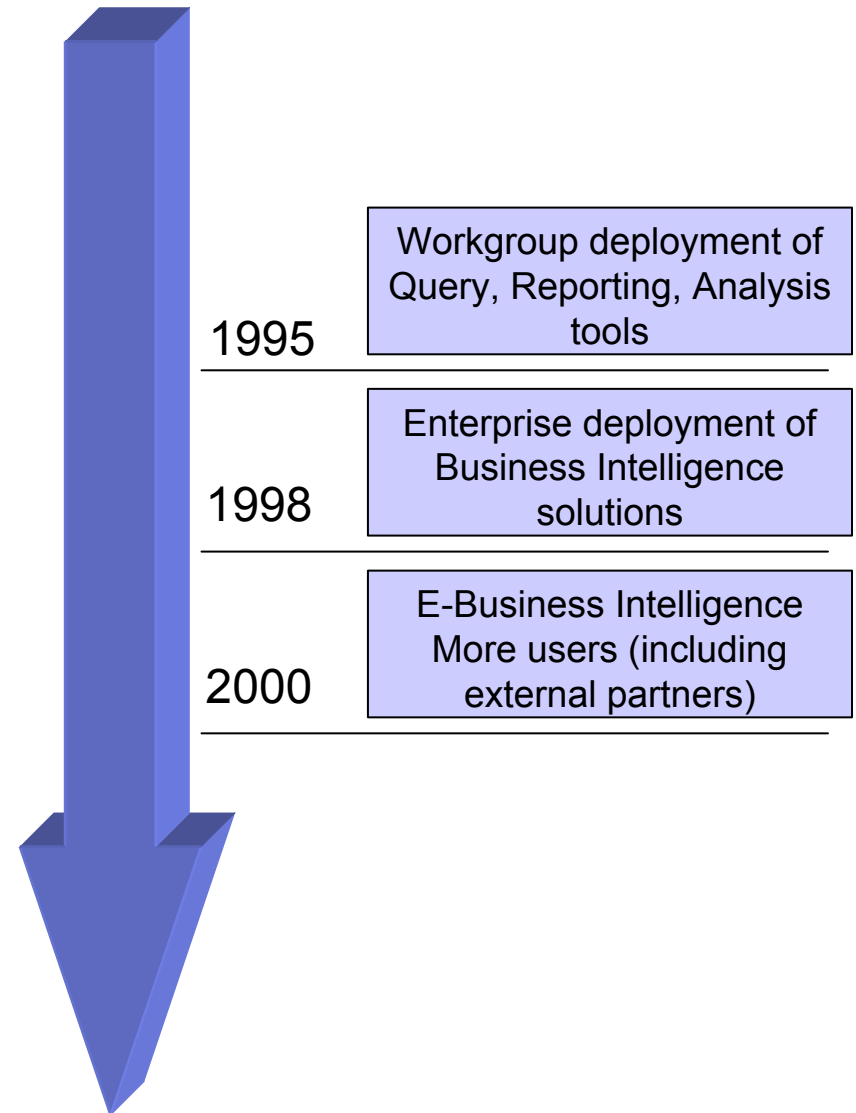

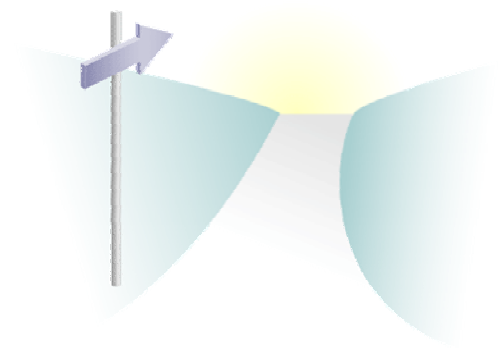(This methodology is still the most popular today)

IBM

Historical Evolution

- 70's – 80's : Infocenters – EIS – DSS

- 1989 : Howard Dresner (Gartner) – first to use 'Business Intelligence'

- 90's : Query and Reporting
  - Datawarehouse and datamarts – Analysis - Data Mining

- 2000 : e-Business Intelligence
  - Balance scorecard

- 2002 : EPM – CPM – BPM - BAM
  - ▸ designed to provide real-time business performance monitoring and analysis based on roles within an enterprise

1995	Workgroup deployment of Query, Reporting, Analysis tools
1998	Enterprise deployment of Business Intelligence solutions
2000	E-Business Intelligence More users (including external partners)

Trends and Drivers

- Huge expansion into Operational BI
  - ▸ Pricing analytics leads retail, Basel2 for banks

- Mandated Regulatory compliance creates privacy, accountability, & security concerns
  - ▸ HIPAA, SOX, Basel II, IAA are having WW effect
  - ▸ But data quality issues will still cause project failures

- Shrinking tolerance for decision making latency
  - ▸ BAM Boom, Real time BI, Active Warehouse, BPM
  - ▸ KPIs & scorecards for the masses = middle managers pressure

- Data mart consolidation will resume

- Metadata emerging from dark ages
  - ▸ Impact analysis & profiling tools lead to repositories!

Trends & Drivers

- Real-time BI

  ▸ A lot of the potential issues around business risks or less-than-expected performance have sources at the operational supply chain," said John Hagerty, an analyst at AMR Research in Boston . "By shining light on that area, it might allow people to cut some of these operational issues at the source."

  ▸ BPM / BAM

IBM

Skills, Not Technology, Inhibit Enterprise BI

By 2006, large enterprises will need three times as many BI personnel as they did in 2001 (0.7 probability).
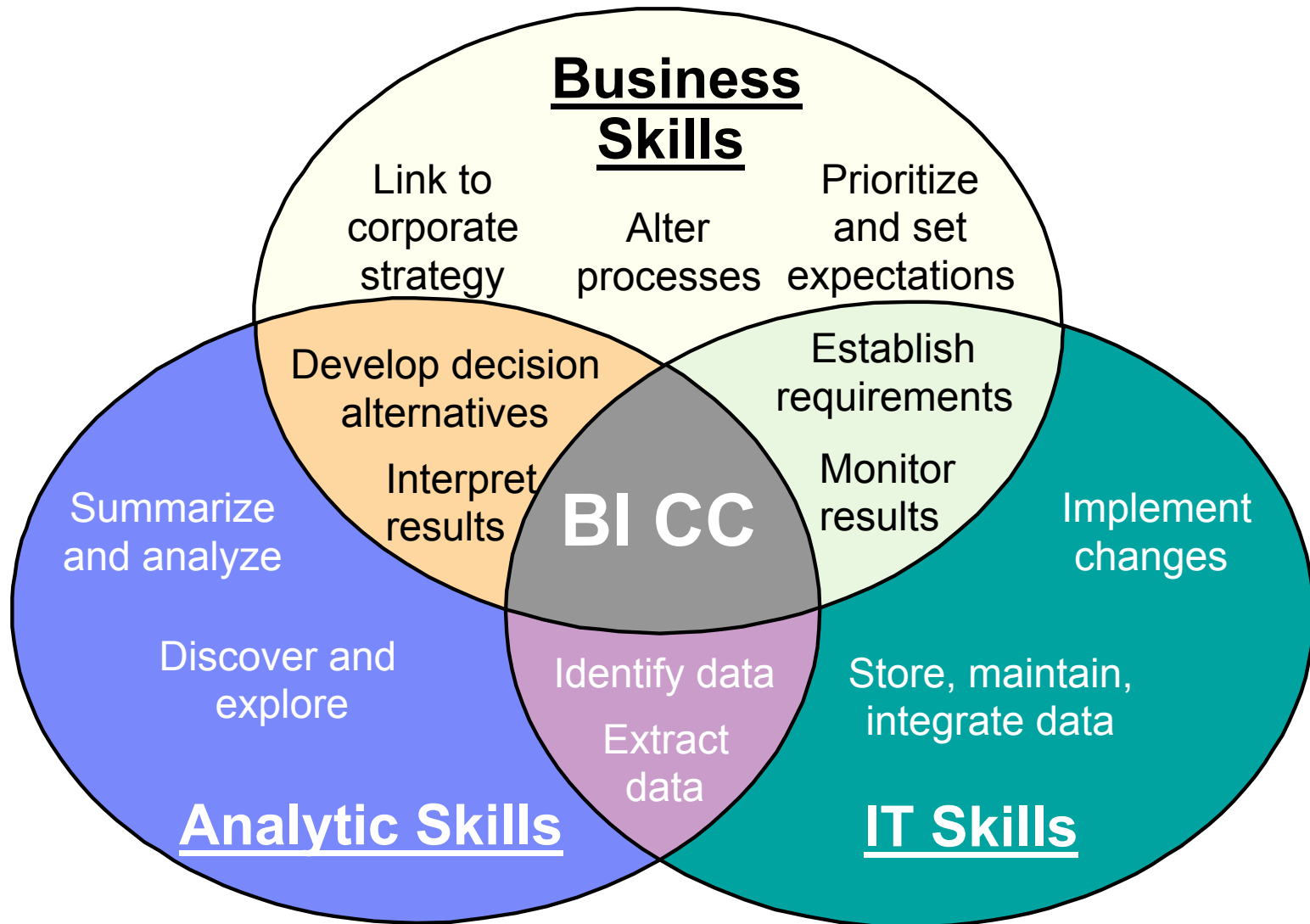
By 2005, demand for BI staff will outweigh supply by 2-to-1 (0.8 probability).

By 2004, enterprises that do not recognize and leverage analytic skills and staff, and do not invest in them, will be unable to meet strategic objectives (0.8 probability).

By 2004, enterprises that do not organize their BI activities (e.g., by starting a BI competency center), will be unable to deliver BI on a strategic level and will face at least 50 percent higher cost of implementation (0.8 probability).

Gartner

Organize Analytic Skills in a BI Competency Center

<u>Business Skills</u>

Link to corporate strategy

Alter processes

Prioritize and set expectations

Develop decision alternatives

Establish requirements

Interpret results

Monitor results

Summarize and analyze

BI CC

Implement changes

Discover and explore

Identify data

Extract data

Store, maintain, integrate data

<u>Analytic Skills</u>

<u>IT Skills</u>

Gartner

Business Intelligence and Datawarehousing on iSeries

DB2 Information Management Software

Laurent ZELLER

Information Management Software Technical Sales Specialist

laurent_zeller@be.ibm.com

@business on demand software

iSeries Parallelism – 2 types

- **Symmetric Multiprocessing (SMP)**

  ▶ assigns both the CPU and I/O processing to tasks that run the query in parallel.

  ▶ This form of SMP allows multiple database operations to take place simultaneously on multiple processors

  ▶ can only be used if the system feature, DB2 Symmetric Multiprocessing for OS/400, is installed.

- **I/O Parallelism**

  ▶ Storage management automatically places data on DASD units evenly across the system.

  ▶ I/O Parallelism describes the ability to load data into main storage simultaneously from different storage devices.

    • This method of access to data may be implemented for a particular job by changing the query attributes using the CHGQRYA command, or by changing the system value QQRYDEGREE.

iSeries - Indexing

- Indexes are used to provide efficient access to data in various environments (both OLTP and OLAP)
  - ▶ Scan of index generally more efficient than table scan as index entries are smaller than record stored
  - ▶ DB2 Optimizer is aware of existing indexes and uses the most appropriate for selecting data

- Indexes types
  - ▶ *Binary Radix Tree:* a root node with entries pointing to the other level of the tree's nodes. At the lowest level, the so-called leaves contain the relative addresses of the records.
    - • improve the performance of queries that select a small percentage of rows from the table
  - ▶ *Dynamic Bitmap*: a dynamically built (at query time) list of bits (bit vector) is created by the optimizer
    - • Optimum when column contains few distinct values
    - • Improve the performance of queries and reduces index storage overhead
  - ▶ *Encoded Vector (EVI):* the index keeps track of the distinct values that can be found in the key column. Instead of having a separate bit array for each distinct value, it assigns a code to each distinct key value (using a Symbol Table).

- Parallel Index Building and Maintenance included in O.S.
  - ▶ Can be used for permanent indexes as well as for temp index building while running ad-hoc queries

iSeries – Parallel Data Load

- The data loader command has the ability to load data from non-IBM rdbms (Oracle, SQL Server) or flat files into DB2/400 files in parallel when the DB2 SMP feature installed

- It breaks the import file into smaller blocks. The blocks are submitted in parallel, so the entire file is processed at the same time.
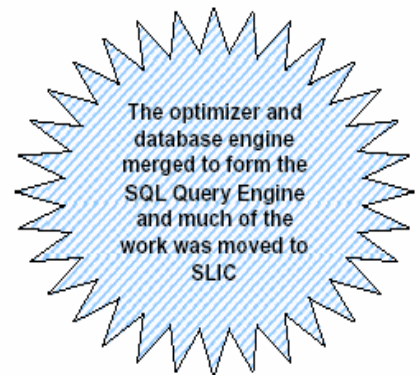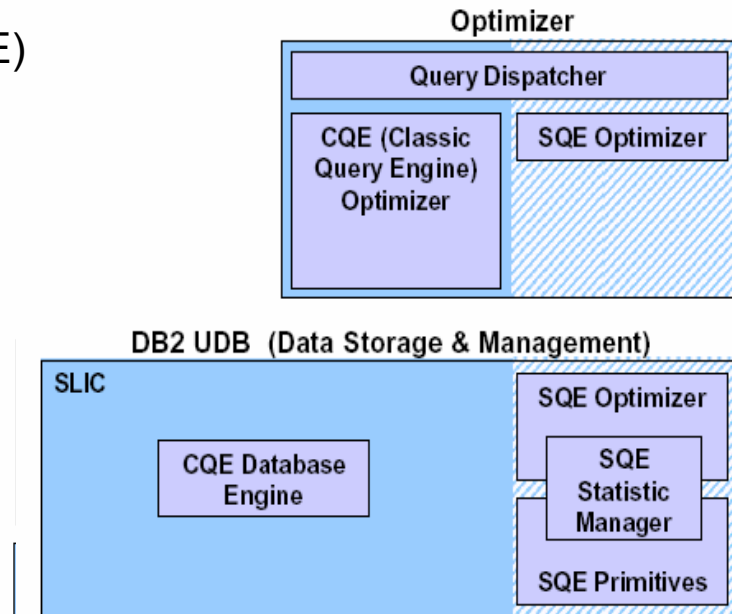
iSeries – Star Schema Join Support

- What is a Star Schema Join ?
  - A typical BI type request based on several dimensions
    - Ex : Sales figures for a given customer, in a given time period, for a given product
  - Otherwise said, the properties of such a join are :
    - That multiple tables take place in the query
    - That the selection predicates are based on the dim. tables rather than on the fact table
    - That an equi-join is used between any dim. table and the fact table to locate and select the relevant fact rows
    - That the equi-join between any dim. Table and the fact table can return many rows, while the intersection of the equi-join between all dim. Tables and the fact may result in a relatively small number of fact rows being returned.

- Without this support, BI type of queries can be difficult to optimize and execute.
  - The optimizer and db engine have specific techniques for recognizing, optimizing and running these star schema joins.

- Must be enabled using QAQQINI query options file
  - Can be scoped to the system, job or individual query

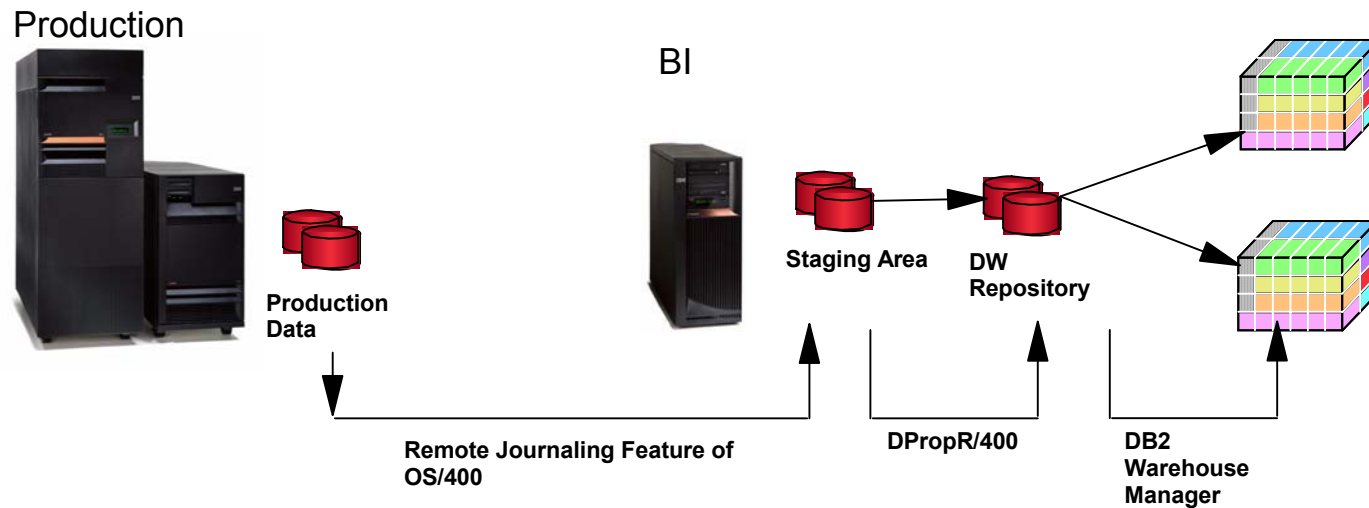# iSeries – SQL Query Engine (SQE) enhancements

- As of V5R2, the optimizer and the database engine have merged to form the SQE.
  - ▶ Much of the workload is now managed by SLIC (System Licensed Internal Code)

- Optimizer now determines what engine will handle the query request through the Query Dispatcher
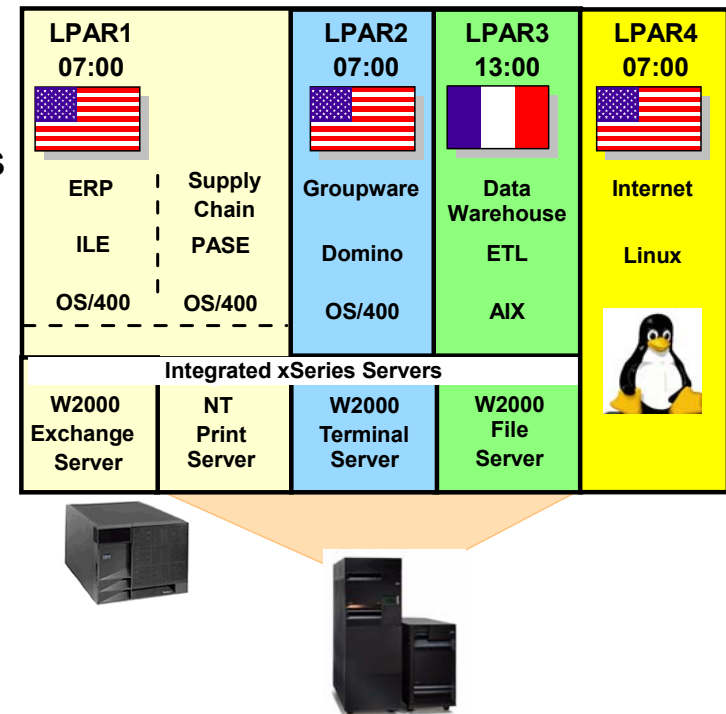 - ▶ Classic Query Engine (CQE)
  - ▶ SQL Query Engine (SQE)

Optimizer

Query Dispatcher	
CQE (Classic Query Engine) Optimizer	SQE Optimizer

DB2 UDB (Data Storage & Management)

SLIC

CQE Database Engine

SQE Optimizer

SQE Statistic Manager

SQE Primitives

The optimizer and database engine merged to form the SQL Query Engine and much of the work was moved to SLIC

iSeries remote journaling

- Advantages:
  - ▸ Remote Journaling handles Data Transport
  - ▸ Separate BI workloads from Production
  - ▸ Tune BI Box for SQL processing (SMP, EVIs, other)

Production

BI

Production Data

Staging Area

DW Repository

Remote Journaling Feature of OS/400

DPropR/400

DB2 Warehouse Manager

iSeries multiple OS Support

- **Logical Partitioning**
  - Software controlled Dynamic Resource Allocation
  - Extremely Granular
  - Linux (and DB2 UDB in Linux)
  - AIX

- **High Speed xSeries Attach**
  - Run Windows server under the covers of iSeries
  - Leverage iSeries reliability, disk management, resources (tape)

- **Server Consolidation**

LPAR1 07:00		LPAR2 07:00	LPAR3 13:00	LPAR4 07:00
ERP	Supply Chain	Groupware	Data Warehouse	Internet
ILE	PASE	Domino	ETL	Linux
OS/400	OS/400	OS/400	AIX	

Integrated xSeries Servers

W2000 Exchange Server	NT Print Server	W2000 Terminal Server	W2000 File Server

IBM value proposition for Business Intelligence on iSeries

DB2 Information Management Software

Laurent ZELLER

Information Management Software Technical Sales Specialist

laurent_zeller@be.ibm.com

@business on demand software

DB2/400

You know it

...

You use it everyday

...

**Why not base
your datawarehouse on it ?**

For Datamarts - DB2 OLAP Server for iSeries

- DB2 OLAP Server is a multidimensional database on which you can create, build, load and maintain high performance reporting datamarts/applications providing

  ▶ "speed of thought" Query Response Times

  ▶ No complex query language or database knowledge required

- DB2 OLAP is based on OPEN Standards

  ▶ Accommodate just many End-User Interfaces

    • Scorecards, Spreadsheets, Portals

- Includes OLAP Mining for detection of deviations

- Includes features for

  ▶ Hybrid Analysis using Integration Server

  ▶ OLAP Builder access to heterogeneous data sources

IBM DB2 OLAP Server

- **OLAP Engine**
  - Navigation
  - Calculations
  - Aggregations
  - Slice/Dice, Drill Up/Down

- **Allow both native multidimensional storage & hybrid analysis**

- **Includes**
  - an Application Manager for Administration
  - Cell-Level Data Security
  - Aliases and Attribute dimensions
  - Multi-user write back

- **Multi-platform support (UWO)**
  - AIX   OS/400
  - OS/390   Windows
  - Sun Solaris Linux
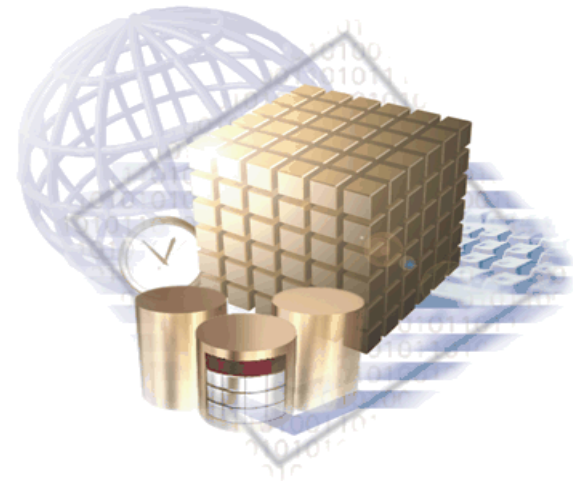
IBM DB2 OLAP Server

- Provides powerful analytical capabilities
  - ▶ Includes a library of financial, statistical and mathematical functions
  - ▶ Includes business trends analysis
  - ▶ Enables Business Performance Management and Scorecarding

- Enables authorized users for write-back
  - ▶ Improve existing business strategies through the implementation of planning, forecasting and other « what-if » applications.

IBM DB2 OLAP Server

- **Hybrid Analysis** – lets you define logical cubes as having a lower detailed portion residing in the relational database
  - ▶ Benefit:
    - combines efficiency of mass data scalability with the sophisticated and performant data analysis of DB2 OLAP Server
    - eliminates the need to load and store detail data in physical cube
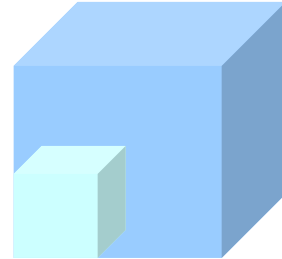
- **Parallel Load/Calc** - Leverages parallel systems to shorten batch indows
  - ➢ takes advantage of up to 4 engines in a symmetric multiprocessors
  - ➢ automatic subdivision of tasks for parallelization for both load and calculation processes

- **Integration Server** - Captures and maps source data for OLAP cubes, including Hybrid Analysis

Software evaluated
OLAP Integration Server

a tool that allows you to transform your relational data source into DB2 OLAP Server databases that are the basis of your OLAP system.

Relational Database → OLAP Integration Server ↔ DB2 OLAP Server databases
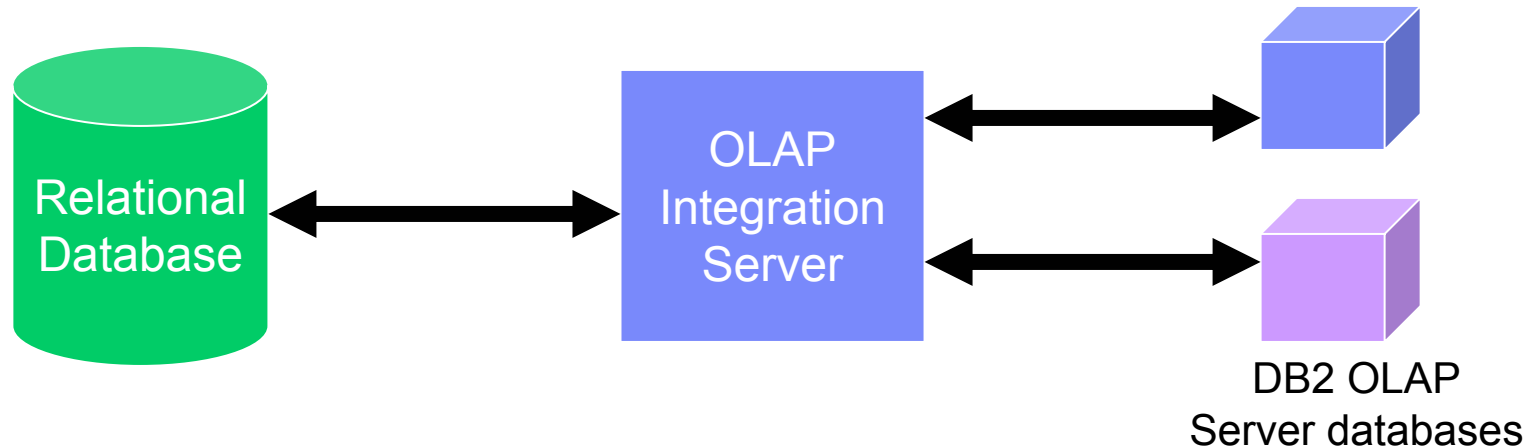
What does OIS do ?

- Maps data from relational data to multidimensional databases
- Stores OLAP definitions in a catalog
- Builds OLAP models from relational metadata
- Creates multiple metaoutlines from a single OLAP model
- Defines multiple DB2 OLAP Server databases
- Offers SQL drill-through to detailed data

OIS – OLAP Model and Metaoutline

- **OLAP Model**:
  - ▸ A graphical tool for building reusable OLAP models that are based on relational tables/views.

- An OLAP model presents a dimensional view of a relational schema
  - ▸ data from a fact table becomes data within an accounts dimension
  - ▸ data from a dimension table becomes metadata describing DB2 OLAP Server dimensions
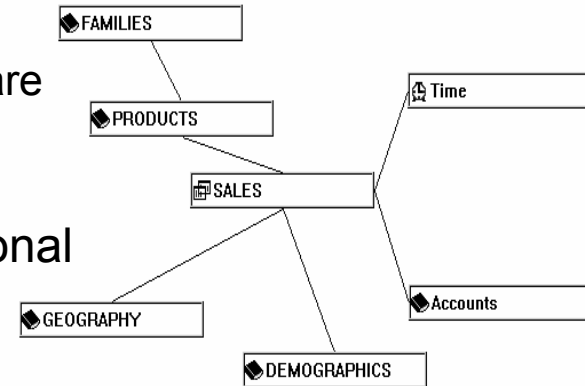
- **OLAP Metaoutline**:
  - ▸ A graphical outline editor
  - ▸ An OLAP metaoutline is designed from an OLAP Model. The OLAP metaoutline is then used to build DB2 OLAP server outlines
  - ▸ defines the dimensions and the semantic levels - members - within the dimensions of a DB2 OLAP Server outline.
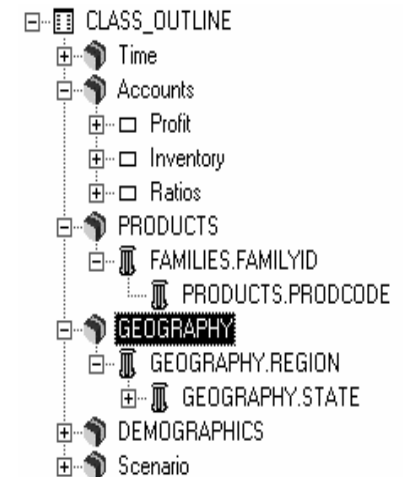
IBM Software Group

Demo :
DB2 OLAP Server
DB2 OLAP Integration Server
DB2 OLAP Analyzer

DB2 Information Management Software

Laurent ZELLER

*Information Management Software Technical Sales Specialist*

laurent_zeller@be.ibm.com

@business on demand software

DB2 OLAP Server for iSeries
Optional component – Warehouse Builder (= ETL component)

- Graphical ETML (Extract-Transformation-Move-Load) tool for
  - Building and maintaining your datawarehouse
    - Definition of ETML jobs from identified source
    - Population of defined target datawarehouse
    - Ease transformations and cleansing

- Allows automation of the ETML processes
  - By use of built-in scheduler
  - Allows transaction history tracking

- Leverage iSeries investment
  - By allowing reuse of Query/400 objects
  - By supporting full DB2/400 SQL
  - By supporting pre/post exec. of iSeries programs
  - By allowing load/calc of DB2 OLAP server cubes

ERP/SCM
Customer
Operations
Sales
Finance
Other
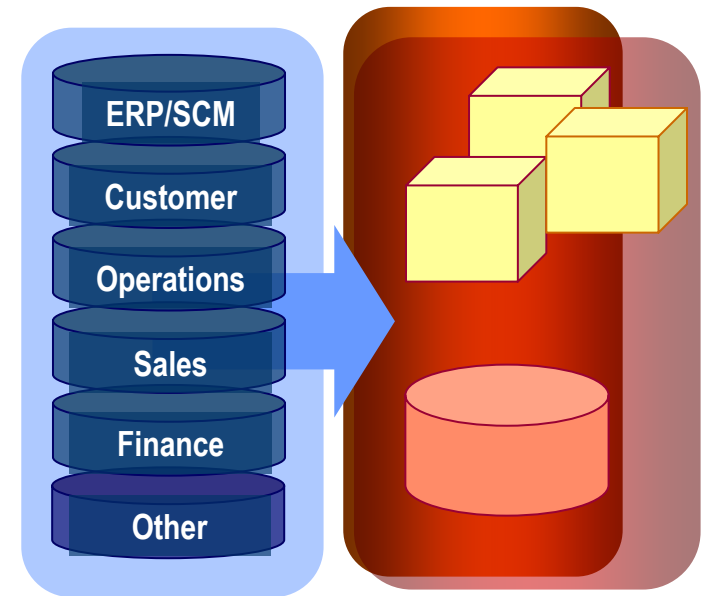
Software evaluated
DB2 OLAP Analyzer

- Interactive, intuitive analysis
- easy to use, highly graphical displays, robust analytics
- leverages the power of DB2 OLAP Server
- enterprise deployable
- broad range of applications types
- extendable using open, mainstream development tools

robust analytics, including 'traffic lighting' & ranking

direct manipulation

intuitive cube navigation

charts, spreadsheets, pinboards, forms
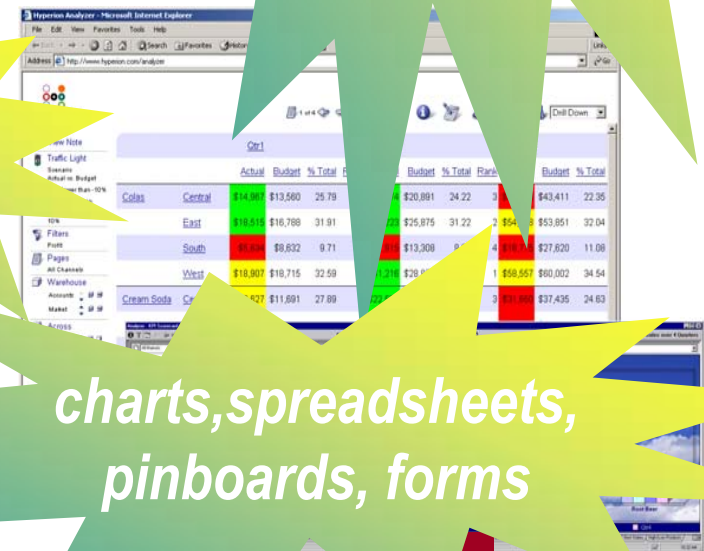
ShowCase®

In partnership with SPSS/Showcase for full i-Series solution

- **SPSS Warehouse Manager**
  - ▶ Enable your company to easily, efficiently, and securely manage datawarehouses.
    - manage and monitor warehouse security and access
    - ensure data simplification, resource allocation
    - for efficient use of hardware resources
  - ▶ Protect the integrity of your data and information systems

- **SPSS OLAP Analyzer (runs natively on OS/400)**
  - ▶ Interactive, intuitive analysis
  - ▶ easy to use, highly graphical displays, robust analytics
  - ▶ leverages the power of DB2 OLAP Server
 - ▶ enterprise deployable

IBM Software Group | DB2 Information Management Software
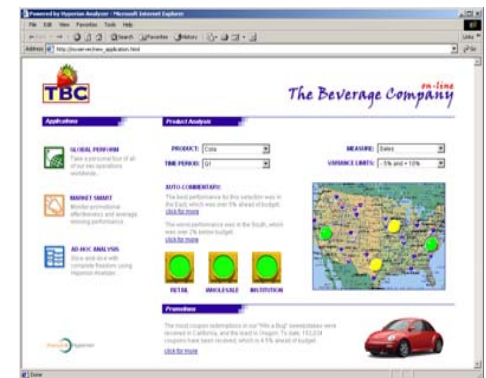
In partnership with SPSS/Showcase for full i-Series solution

- **SPSS Query**
  - ▶ Build your own queries from a full scale GUI
    - • empowers end users to access key business information with far less need for IT involvement.
    - • lets both business and technical users securely access the company's relational or multidimensional databases

- **SPSS Report Writer**
  - ▶ Create powerful, highly-formatted reports for ad-hoc data analysis.
  - ▶ Professional looking reports are only a click away with Report Writer's built-in templates.
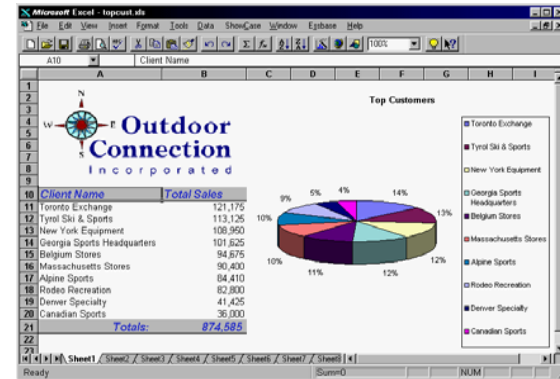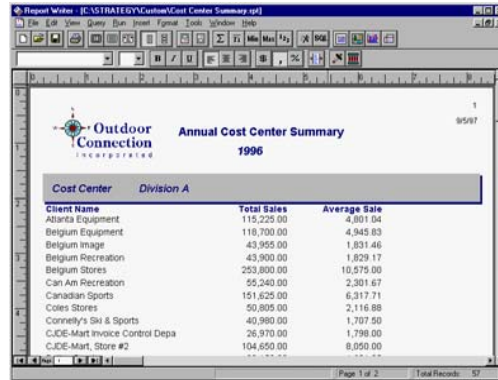
ShowCase Warehouse Manager is a prerequisite to ShowCase Report Writer and Query

In partnership with SPSS/Showcase for full i-Series solution

- **SPSS Enterprise Reporting**
  - ▶ Establish a central, searchable repository for reports and other critical information
  - ▶ Notify individuals or groups by e-mail automatically when certain types of reports or other content are updated
  - ▶ Enable report developers to link several reports on a single page, without requiring them to do programming
  - ▶ Provide for the creation of personal home pages containing links to frequently needed reports, documents, and even other applications
  - ▶ Share query and report definitions for both scheduled and dynamic execution
 - ▶ Automate the scheduling and distribution of routine reports
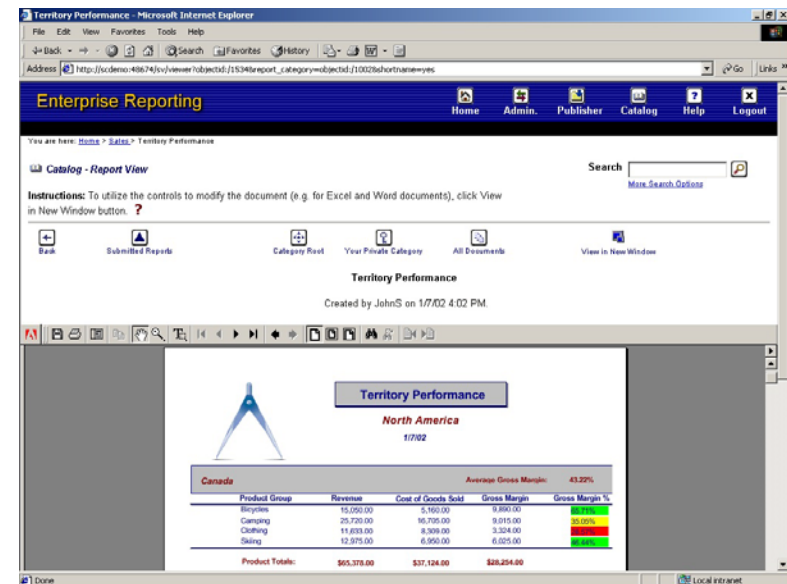
For ETL - DB2 Data Warehouse Center and Warehouse Manager iSeries agent

▪DB2 Data Warehouse Center

- ►DBA administration console
- ►Database schema & user maintenance
- ►Access to most RDBMS's & flat files
- ►Schedules and monitors database tasks
- ►150+ data transformations
- ►Loads data into DB2 data warehouses
- ►CWM Standards adherence

▪DB2 Warehouse Manager adds:

- ►Extracts & transformations on remote servers via agents
- ►Information Catalog end user metadata
- ►Connectors to MQ, Tivoli Site Analyzer

DB2 Warehouse Center
extract, transform, load, schedule, administrate

DB2

DB2 Warehouse Manager

ETL agents

Information Catalog

Metadata
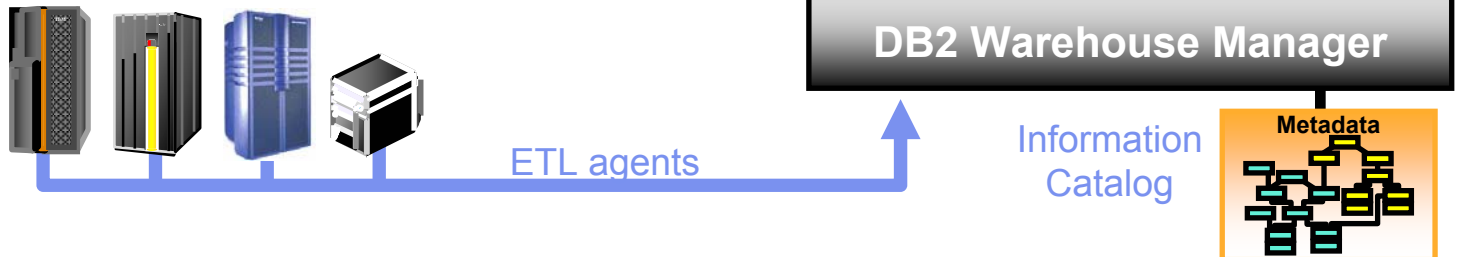
Demo : DB2 Warehouse Center and Manager

DB2 Information Management Software

Laurent ZELLER

Information Management Software Technical Sales Specialist

laurent_zeller@be.ibm.com

@business on demand software

IBM value proposition for Business Intelligence on UWL (unix, windows, linux)

DB2 Information Management Software

Laurent ZELLER

Information Management Software Technical Sales Specialist

laurent_zeller@be.ibm.com

@business on demand software

DB2 Data Warehouse Manager and Center

- **DB2 Data Warehouse Center**
  - ► DBA administration console
  - ► Database schema & user maintenance
  - ► Access to most RDBMS's & flat files
  - ► Schedules and monitors database tasks
  - ► 150+ data transformations
  - ► Loads data into DB2 data warehouses
 - ► CWM Standards adherence
- **DB2 Warehouse Manager adds:**
  - ► Extracts & transformations on remote servers via agents
  - ► Information Catalog end user metadata
  - ► Connectors to MQ, Tivoli Site Analyzer

DB2 Warehouse Center
extract, transform, load, schedule, administrate

DB2

ETL agents

DB2 Warehouse Manager

Information Catalog

Metadata

# DB2 Warehouse Center & Warehouse Manager

- **Register and access data sources**
  - ▸ DB2, Oracle, Microsoft, Sybase, Informix, flat file sources, and others

- **Define extraction and transformation steps**
  - ▸ Over 100 built-in transformations leveraging SQL
  - ▸ SQL, Warehouse programs, UDF, SP, other exec

- **Define data movement & warehouse population**
  - ▸ Full refresh, history, and incremental data movement
  - ▸ Use Agent technology for point-to-point movement

- **Model, automate, and monitor processes**
  - ▸ Schedules, triggers, dependencies, retries, notifications

- **Manage and interchange metadata**
  - ▸ Standards-based, OMG CWMI
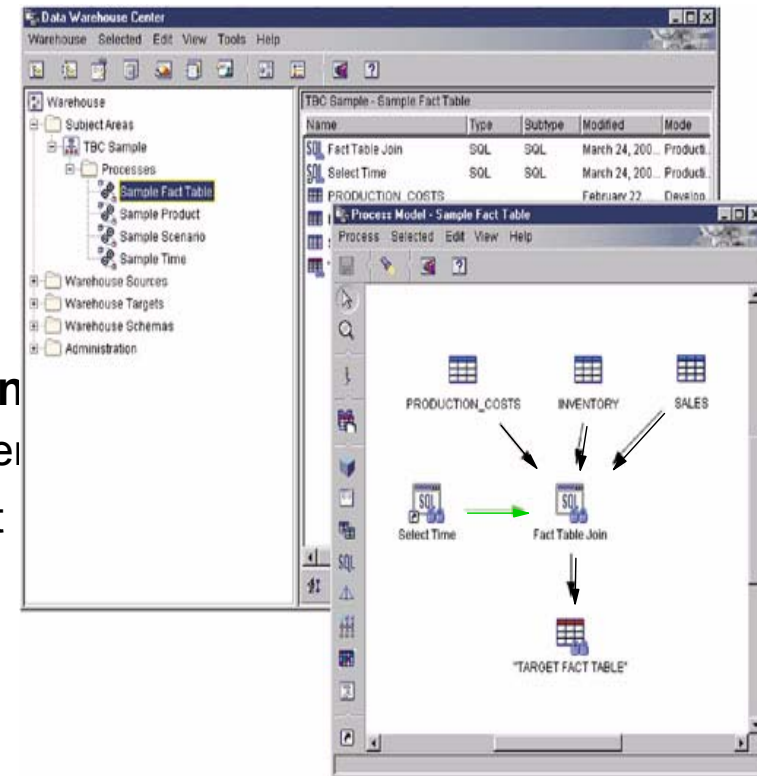
DB2 UDB - Multidimensional Clustering Block Indexes

- A CREATE TABLE option
  - ▸ Co-locates rows with same key values in same data block
  - ▸ Transparent to applications and end user tools

Region

| East 1997 | East 1998 | North 1999 | South 1999 | West 2000 |

Year
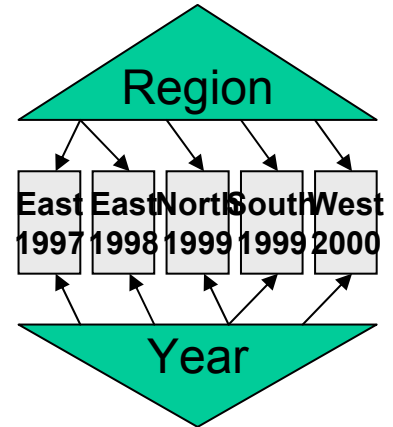
Data blocks without MDC

SKU	Store	Date	Qty	Amt
101	**21**	**04/02**	**1**	**1.50**
101	**21**	**04/02**	**1**	**2.38**
101	7	04/02	2	3.00

SKU	Store	Date	Qty	Amt
101	7	04/02	1	1.50
101	**21**	**04/02**	**3**	**4.10**

Data blocks with MDC

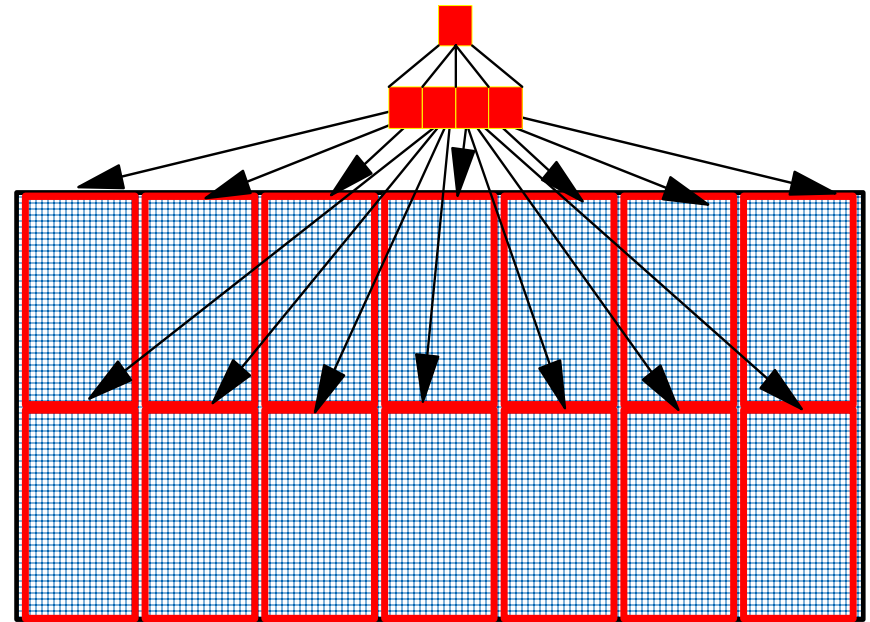| SKU	Store	Date	Qty	Amt
101	**21**	**04/02**	**1**	**1.50**
101	**21**	**04/02**	**1**	**2.38**
101	**21**	**04/02**	**3**	**4.10**

SKU	Store	Date	Qty	Amt
101	7	04/02	2	3.00
101	7	04/02	1	1.50

Row Indexes, Block Indexes (Rids & Bids)

**Row Indexes - 1 index
entry per row**
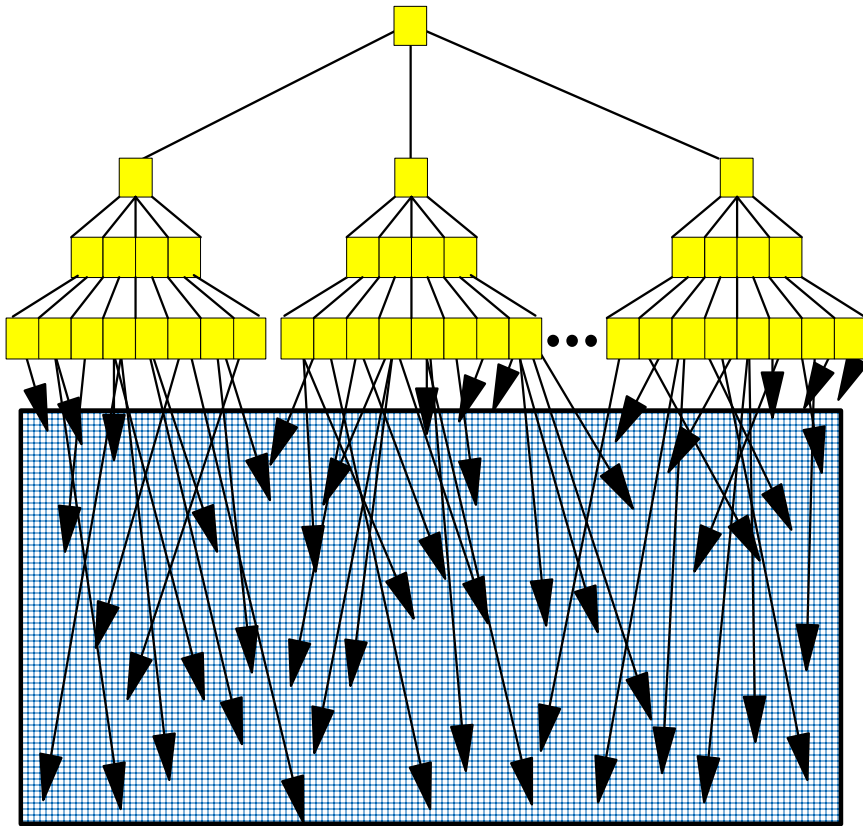
**Block Indexes - 1 index entry
per block**

Multi-Dimensional Clustering Benefits

- Faster Query speed
  - ▶ Access only the data pages necessary - no searching
  - ▶ Get all the data in a few accesses versus 100s of page reads
  - ▶ Reduces CPU & disk I/O use for other users to leverage

- Simple & Intuitive --Multidimensional keys and Star Schema friendly
  - ▶ Perfect for OLAP style hierarchical analysis

- One index entry covers entire data page, not one per row
  - ▶ Reduces index size --Saves disk, faster queries

- Reduces DBA reorg's --row clustering managed by DB2

- Faster Deletes -- just drop a few data pages

- Faster Inserts -- store the record, but rarely insert an index entry

DB2 Cube Views

- Multi-dimensional metadata inside DB2
  - It is not a physical cube being stored in DB2 ! – only a description of it (metadata)
  - Import/export for metadata exchange bridges
  - Includes an XML based API for creation of bridges to other sources

- A GUI for its management – the OLAP Center
  - Wizard to design "cube views" metadata
  - And to define efficient multidimensional MQTs
  - DB2 optimizer query rewrite to MQTs for performance

- Client interfaces
  - Office Connect
  - QMF for Windows
  - Bridges to other tools

Cube Views
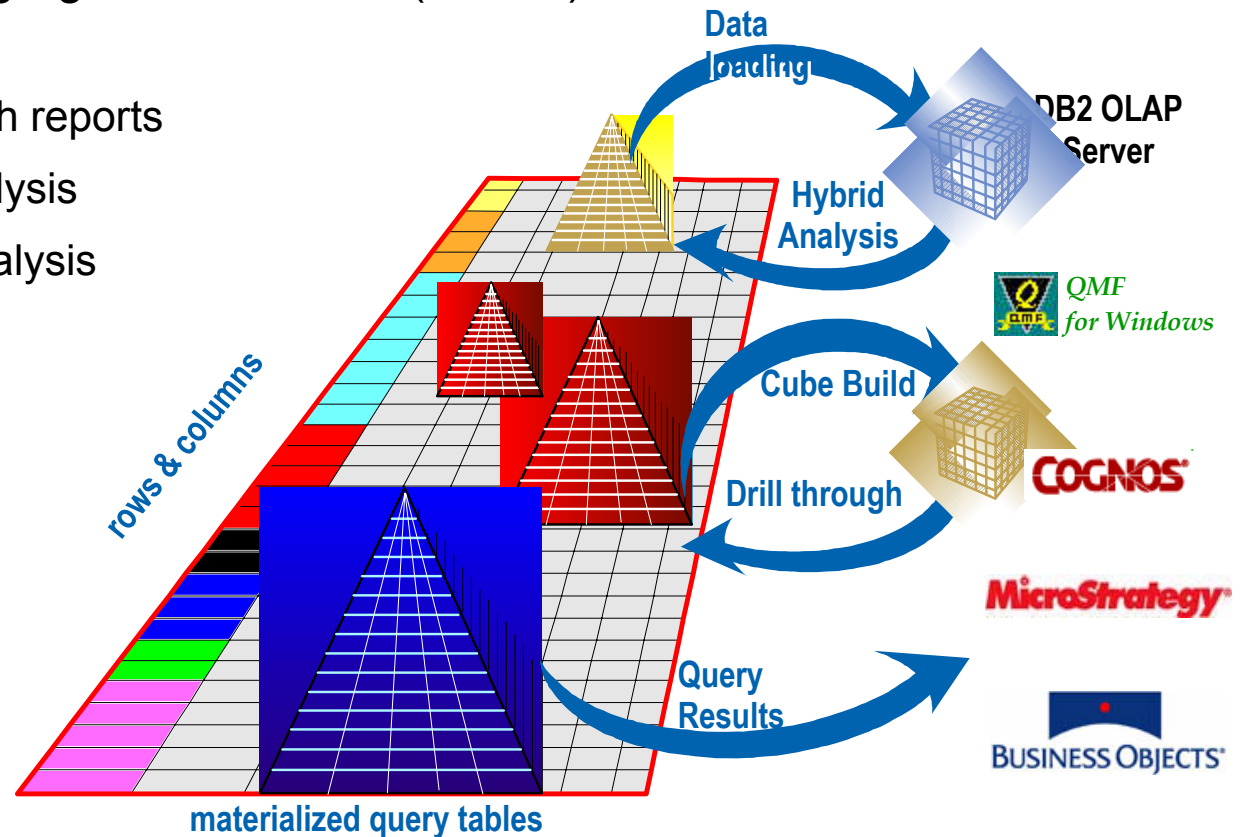
aka advanced MQTs
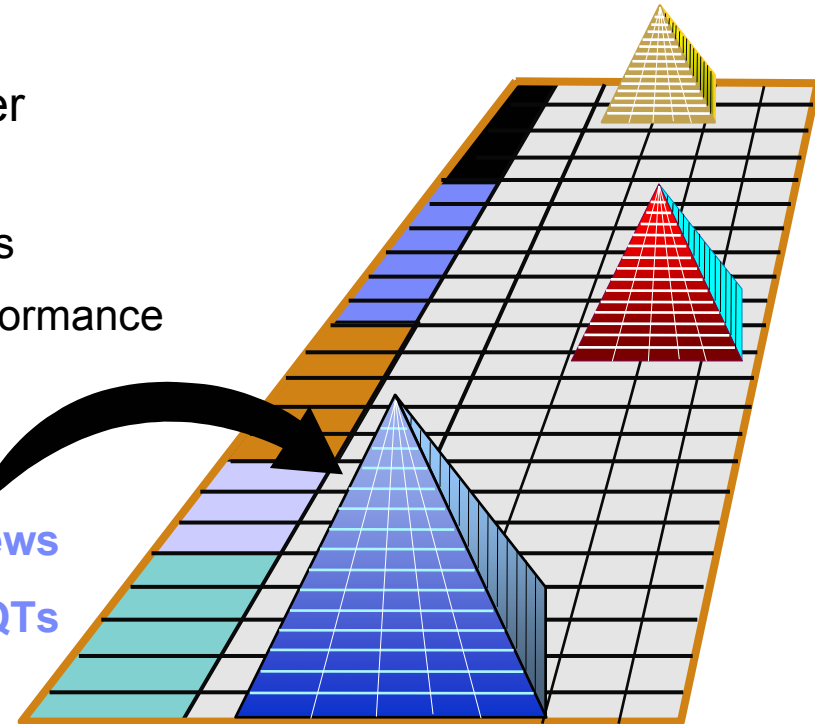
Metadata Bridge With DB2 Cube Views
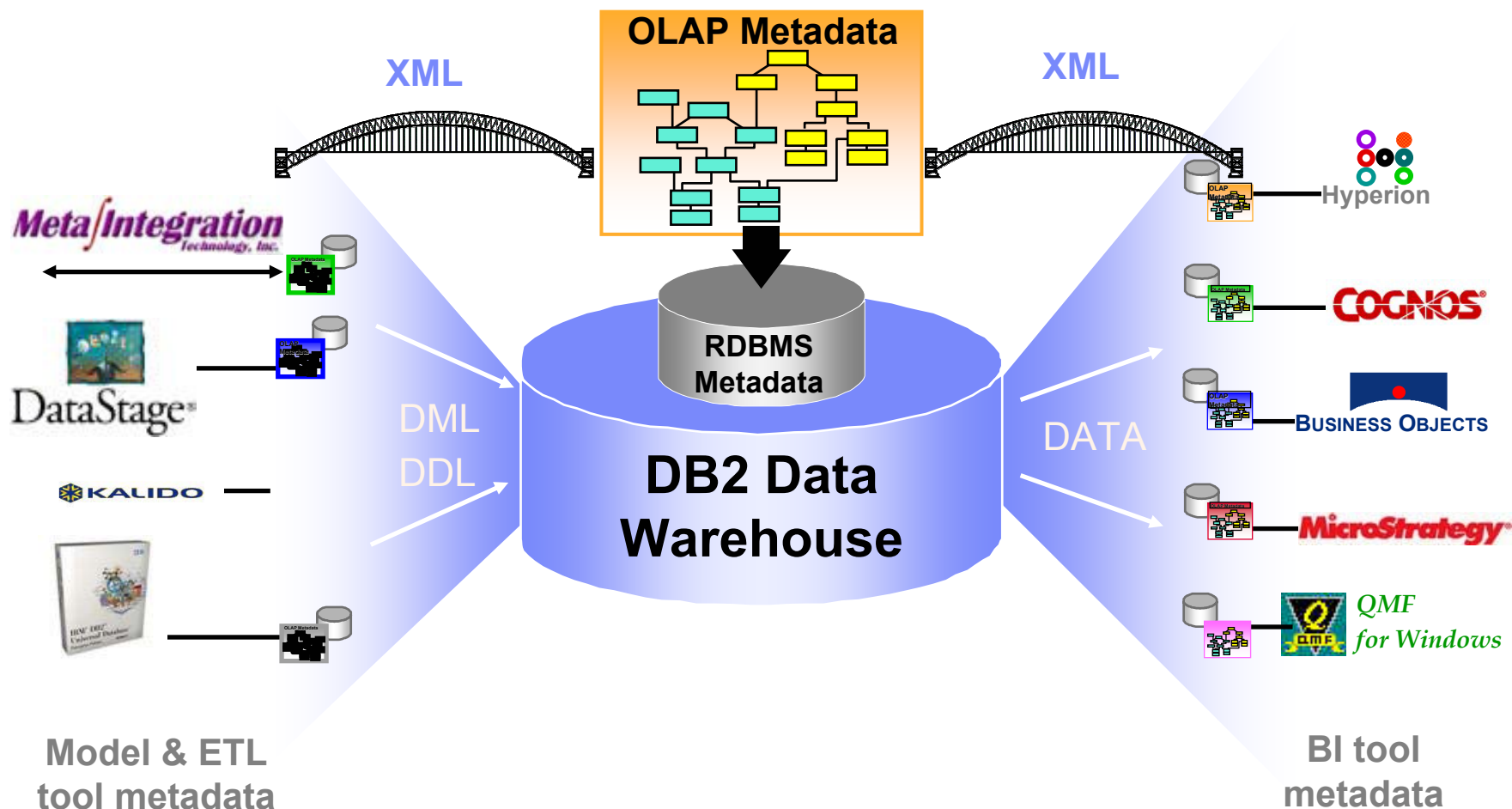
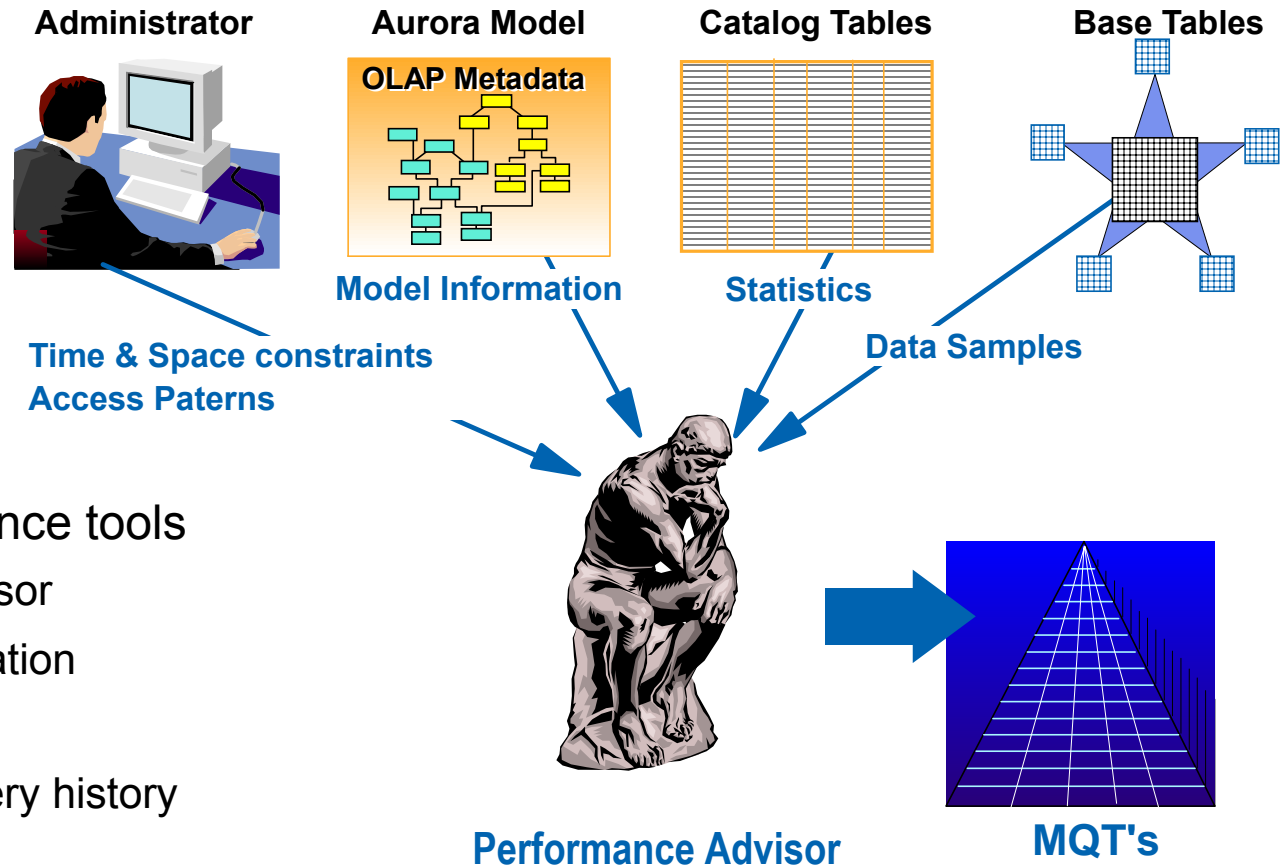# Ensure efficient and performant Multidimensional data storage

- Use of multi-level aggregate « caches » (MQT's)
  - Faster cube load
  - Faster drill through reports
  - Faster hybrid analysis
 - Faster Ad-hoc analysis
  - Faster Reporting

Ensure efficient and performant Multidimensional data storage

- Use of DB2 Optimizer + Multidimensional clustering (MDC) for ensuring performance
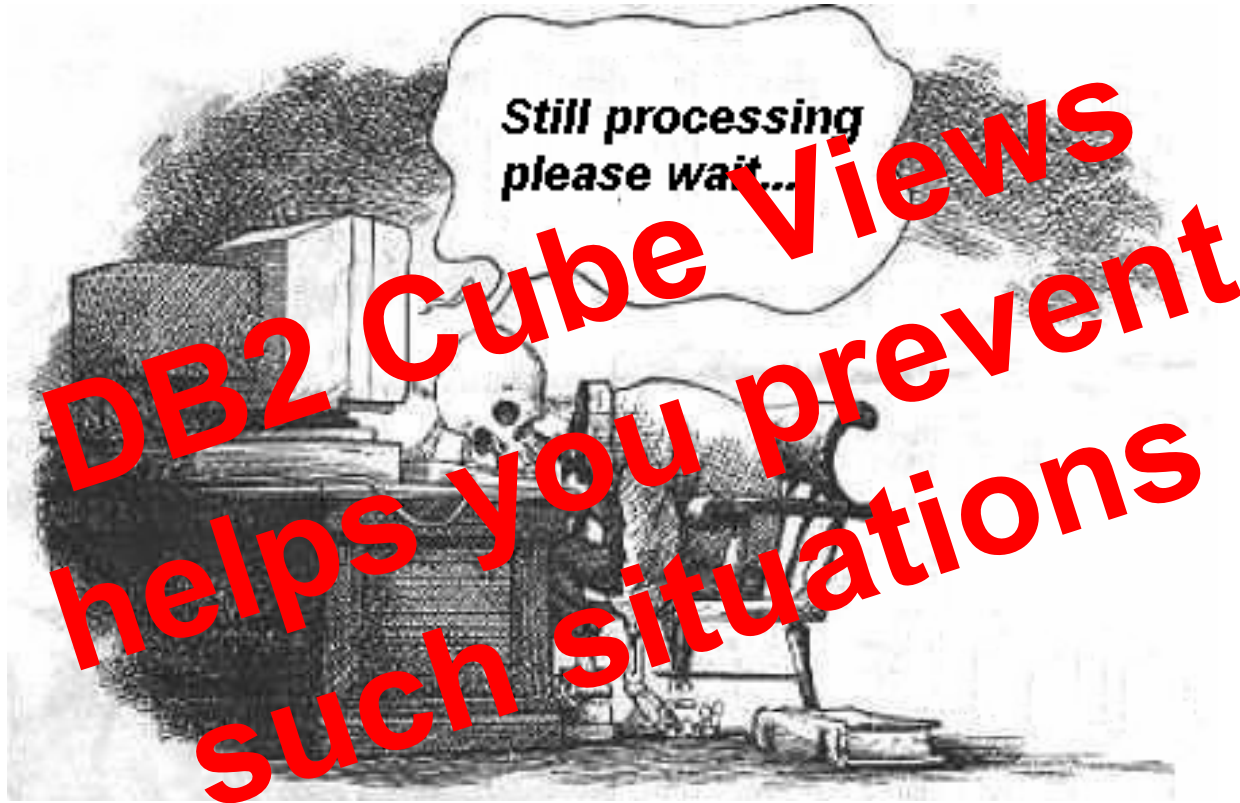
Administrator **Aurora Model** **Catalog Tables** **Base Tables**

OLAP Metadata

Model Information **Statistics**

Data Samples

Time & Space constraints
Access Paterns

- Use of new performance tools
  - ▸ Model based advisor
    - • For MQT creation
  - ▸ MQT advisor
    - • Based on query history

Performance Advisor **MQT's**

Cube Views Performance – "5-month Database"
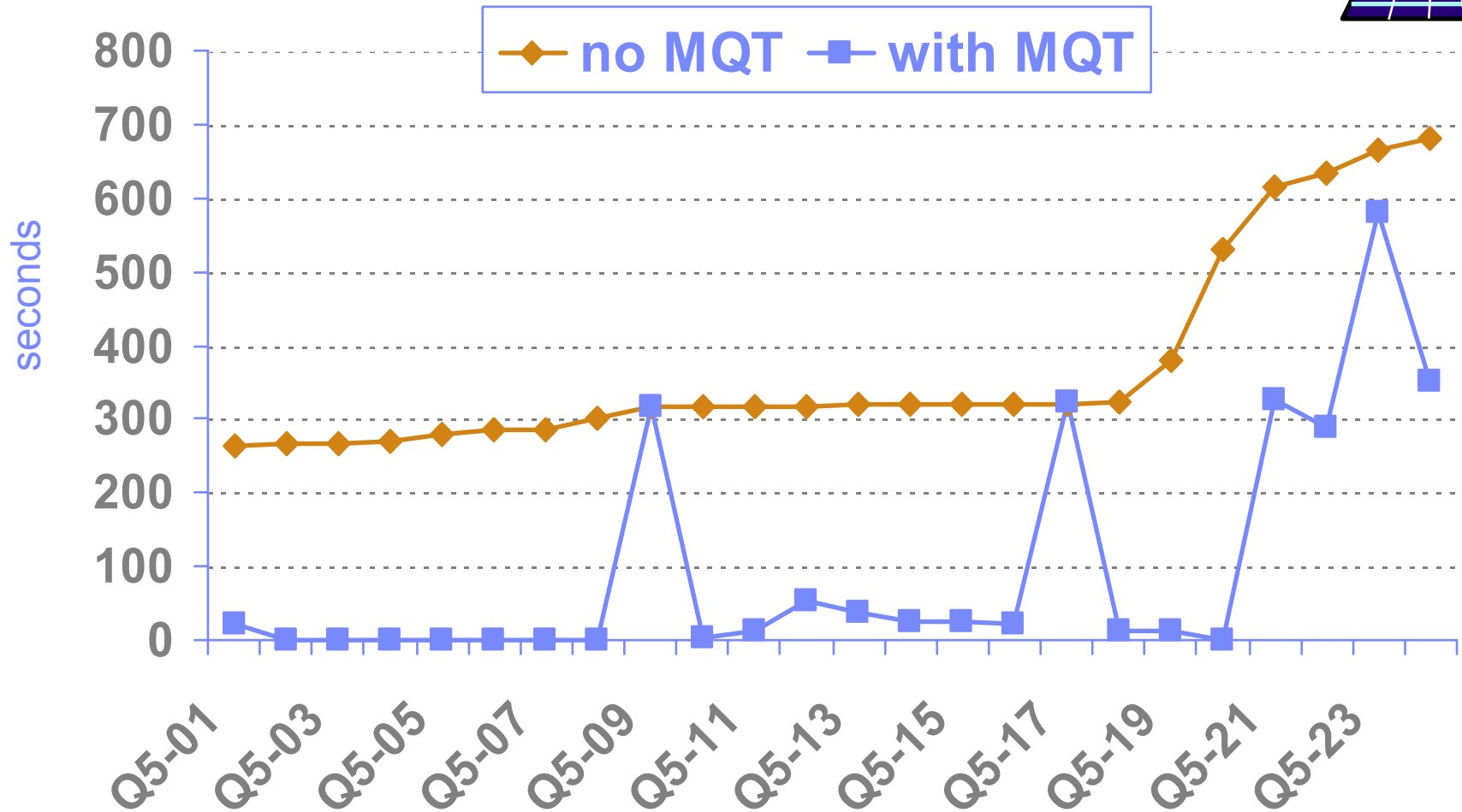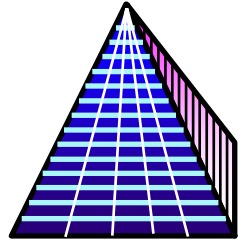
Some fun …. The killer query

Query Patroller V8

- Intercept query from the client
- Flexible result set options
  - ▶ Queue handling of reports returned
  - ▶ Take less disk space
- Allow QP to manage more than one database
- Improved interoperability with end user tools
- Simplify install

Query Patroller

queue

SQL 14

SQL 15

SQL 21

SQL 23

explains

answers

answer sets queued

DB2 Optimizer
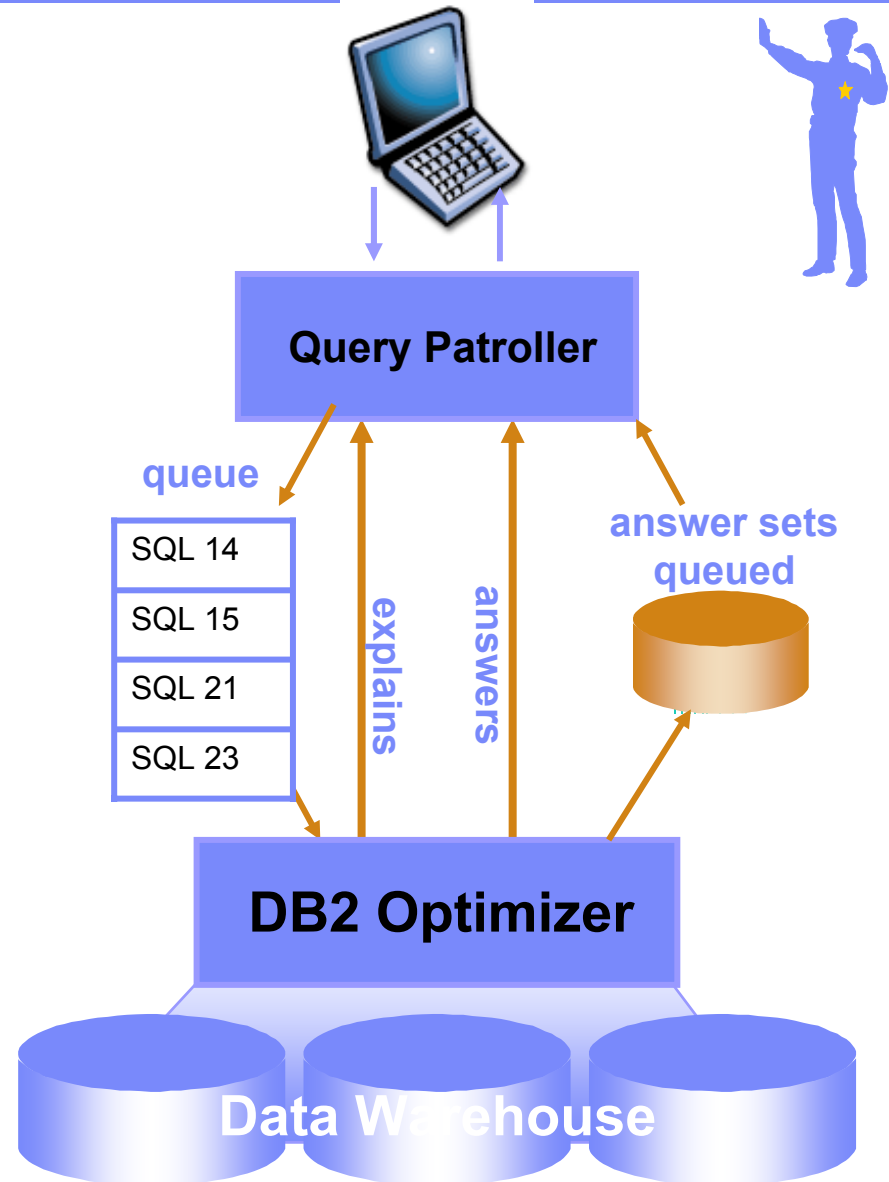
Data Warehouse

Query Patroller

- DB2 Query Patroller helps data warehouse DBAs:

  ▶ Set and enforce proactive limits on resource usage

  ▶ Collect and analyze information on data warehouse usage

  ▶ Monitor and take action on current query activity

  ▶ Put more information and capability in the hands of selected power users

- **Query governing:** before execution by DB2 engine

- **Reporting:** detail and summary information on past queries

- **Monitoring:** enables DBAs to view and take action on queries that have been recently submitted

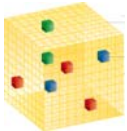- **Performance** improvements: query throughput, database hot/cold spots

DB2 QMF V8.1

- **What's New in QMF for Windows**
  - ▸ Drag & Drop Query Building
  - ▸ Diagram View for Query Building
  - ▸ OLAP Client w/ DB2 Cube Views
  - ▸ Slice, Dice and Drill
  - ▸ Executive Dashboards w/ QMF Visionary
  - ▸ Database Explorer
  - ▸ Create Visual Reports
  - ▸ XML Import / Export

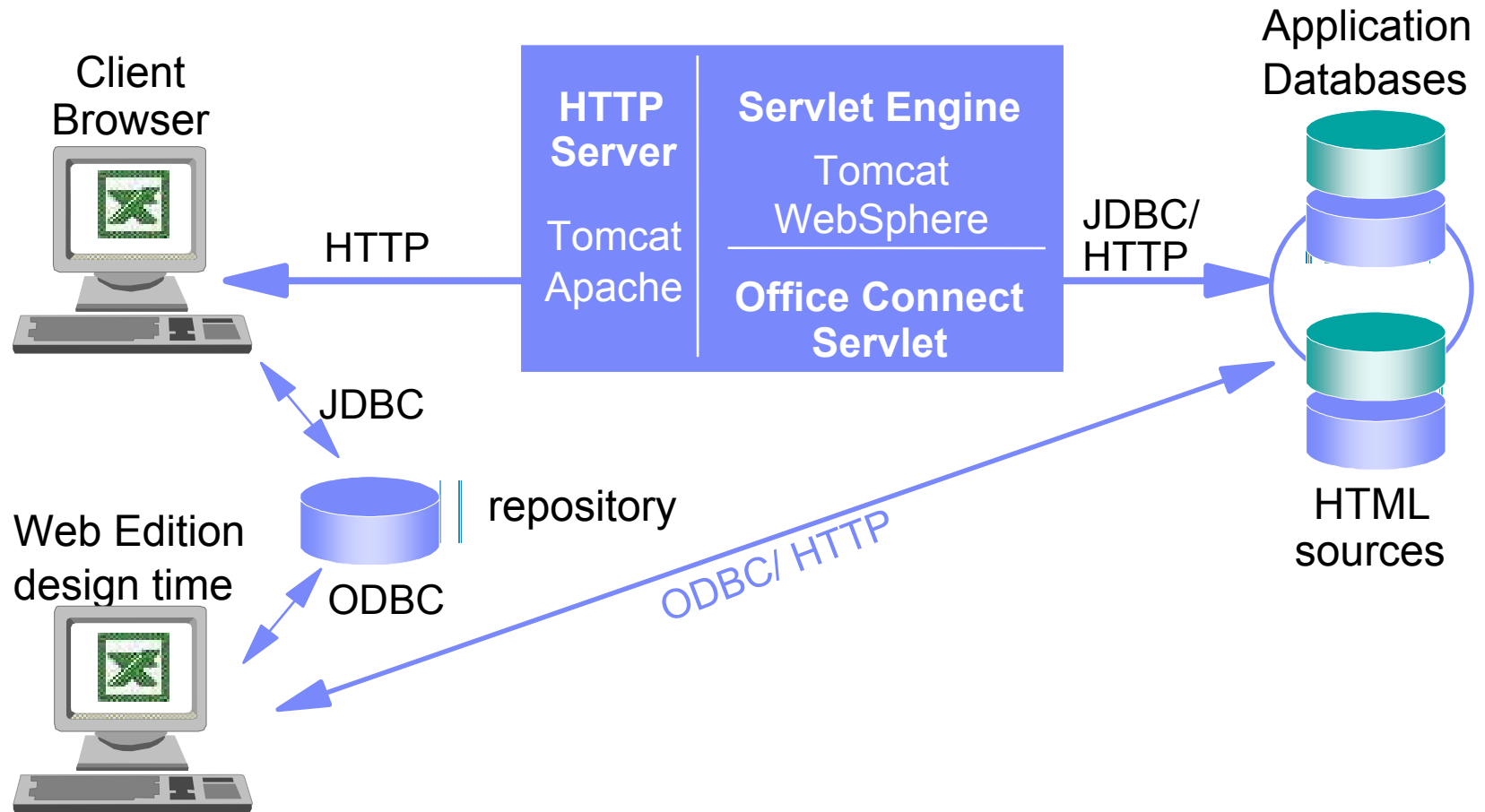|	1	2	3	4	5	6	7
1	⊟	FAMILY_FAMILY		⊟100			
2		PRODUCT_SKU		100-10		100-20	
3	⊞REG...	⊟TIME_Y...	⊞TIME_QUAR...	Profit	Margin	Profit	Margin
4			⊞Qtr1	820.34	1900.95	915.00	1775.00
5			⊞Qtr2	936.00	2117.00	1051.00	1939.00
6	⊞Central	⊟2000	⊞Qtr3	884.13	2046.11	1097.00	1974.00
7			⊞Qtr4	799.00	1917.00	897.00	1738.00
8			All values for 2000	3439.47	7981.06	3960.00	7426.00
9		All values for Central		3439.47	7981.06	3960.00	7426.00
10			⊞Qtr1	2455.40	3562.28	212.25	359.75
11			⊞Qtr2	2949.00	4121.00	304.00	478.00
12	⊞East	⊟2000	⊞Qtr3	3305.00	4504.00	314.00	490.00
13			⊞Qtr4	2436.00	3549.00	287.00	456.00
14			All values for 2000	11145.40	15736.28	1117.25	1783.75
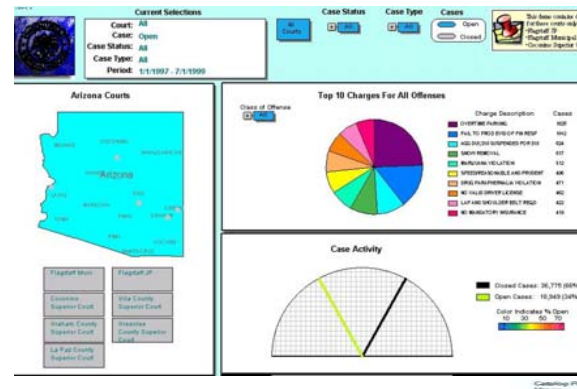
DB2 QMF V8.1

New QMF for WebSphere

- ▶ Runs under any WebSphere Application Server
- ▶ Browser Access – no plugins!
- ▶ Access existing QMF queries and reports
- ▶ Create SQL or Prompted queries
- ▶ OLAP Client w/ DB2 Cube Views
- ▶ Slice, Dice and Drill
- ▶ Integrated Charting
- ▶ Java Class API
- ▶ Web Services API
- ▶ XML import / export

WebSphere. software

QMF for WebSphere v8.1

Server: db2 v81   Set Server
User: <none>   Change User

Actions
Create
  List
  SQL Query
  Prompted Query
  Procedure
Open
  From Server
  From File
Manage
  Documents
  Connections
  Global Variables
Customize
  Appearance
Help

Command   Run

IBM QMF for WebSphere v8.1

Copyright (C) 1999-2003 IBM Corp. All Rights Reserved. IBM is a registered trademark of International Business Machines Corporation.

Copyright (C) 1999-2003 Rocket Software, Inc. All Rights Reserved. Rocket is a registered trademark of Rocket Software, Inc.

Office Connect

Client Browser

Application Databases

HTTP Server

Tomcat
Apache

Servlet Engine

Tomcat
WebSphere

Office Connect Servlet

HTTP

JDBC/ HTTP

JDBC

repository

Web Edition design time

ODBC

ODBC/ HTTP

HTML sources

Intelligent Miner Modeling, Scoring, Visualization

SQL Modeling

PMML Model

SQL Scoring

DB2 Data Warehouse

Intelligent Miner

- Neural Nets
- Associations
- Time Series
- Radial Basis
- Decision Trees
- Regression

Answers

clusters

scores

predictions

Outliers

Decisions

visualize
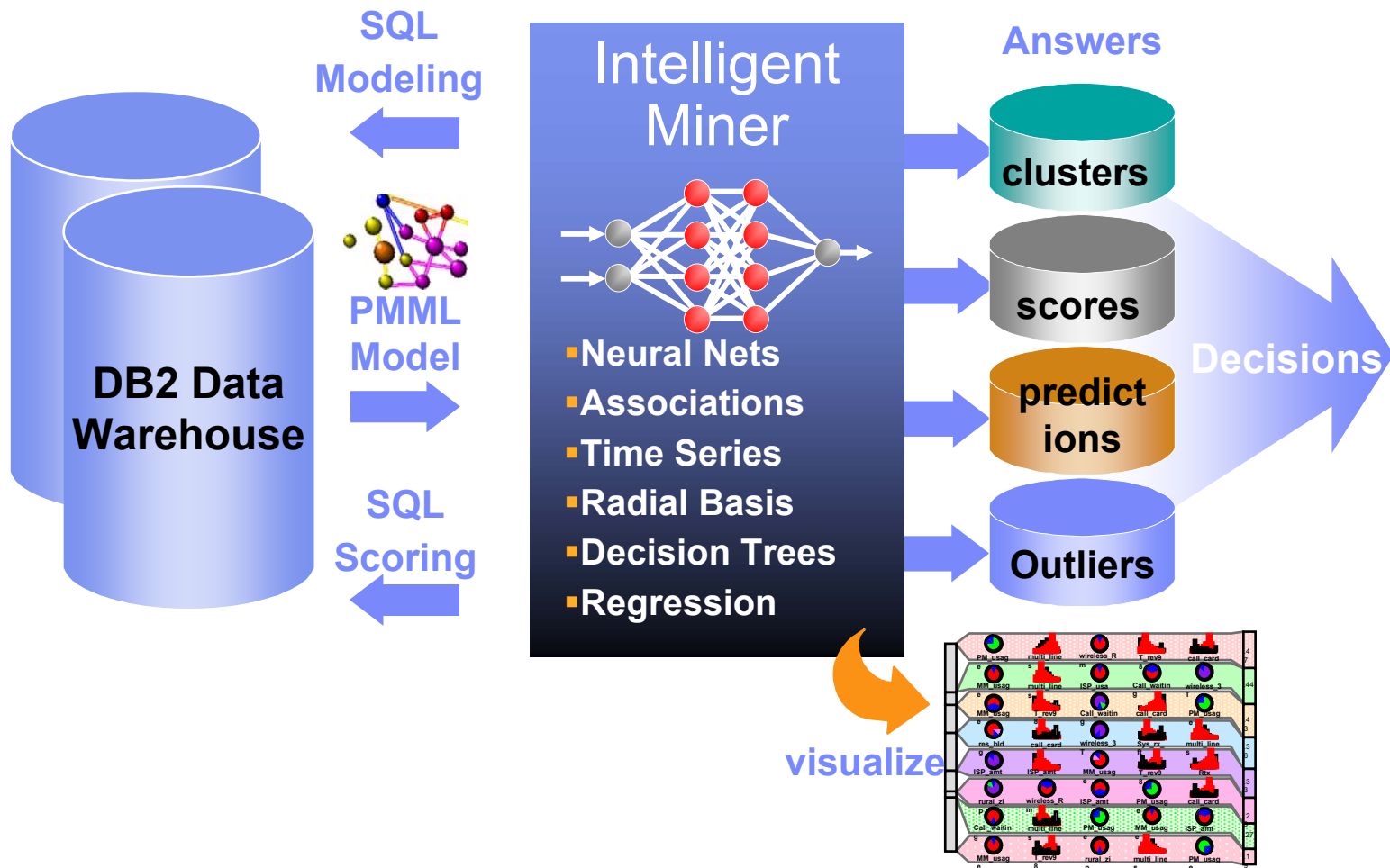
DB2 Information Integrator - Federation

- Transparency
    - ▶ Appears to be one source
- Heterogeneity
    - ▶ Integrates diverse data sources
    - ▶ Structured, XML, unstructured, messages, Web, …
- Distributed optimization
- Uses
    - ▶ Monthly, Quarterly reports
    - ▶ *On Demand* requests
    - ▶ Limit data going into EDW

DB2

Information Integrator (federated server)

wrappers

XML   **Informix**   **DB2**   **Oracle**   **SQL Server**   **Flat Files**

DB2 Information Integrator

- Heterogeneous federation technology
  - ▶ for structured data & content
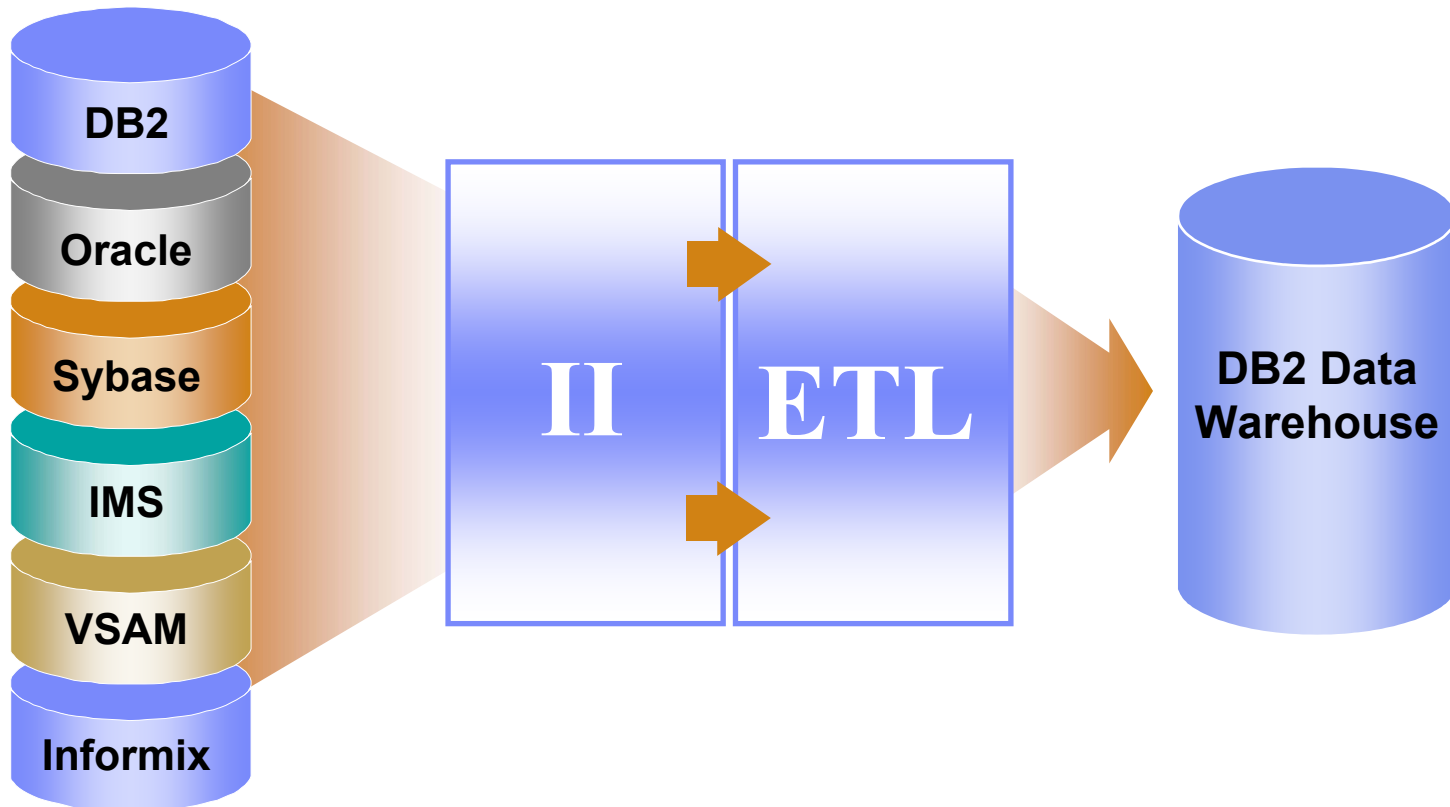
Conclusions

DB2 Information Management Software

Laurent ZELLER

Information Management Software Technical Sales Specialist

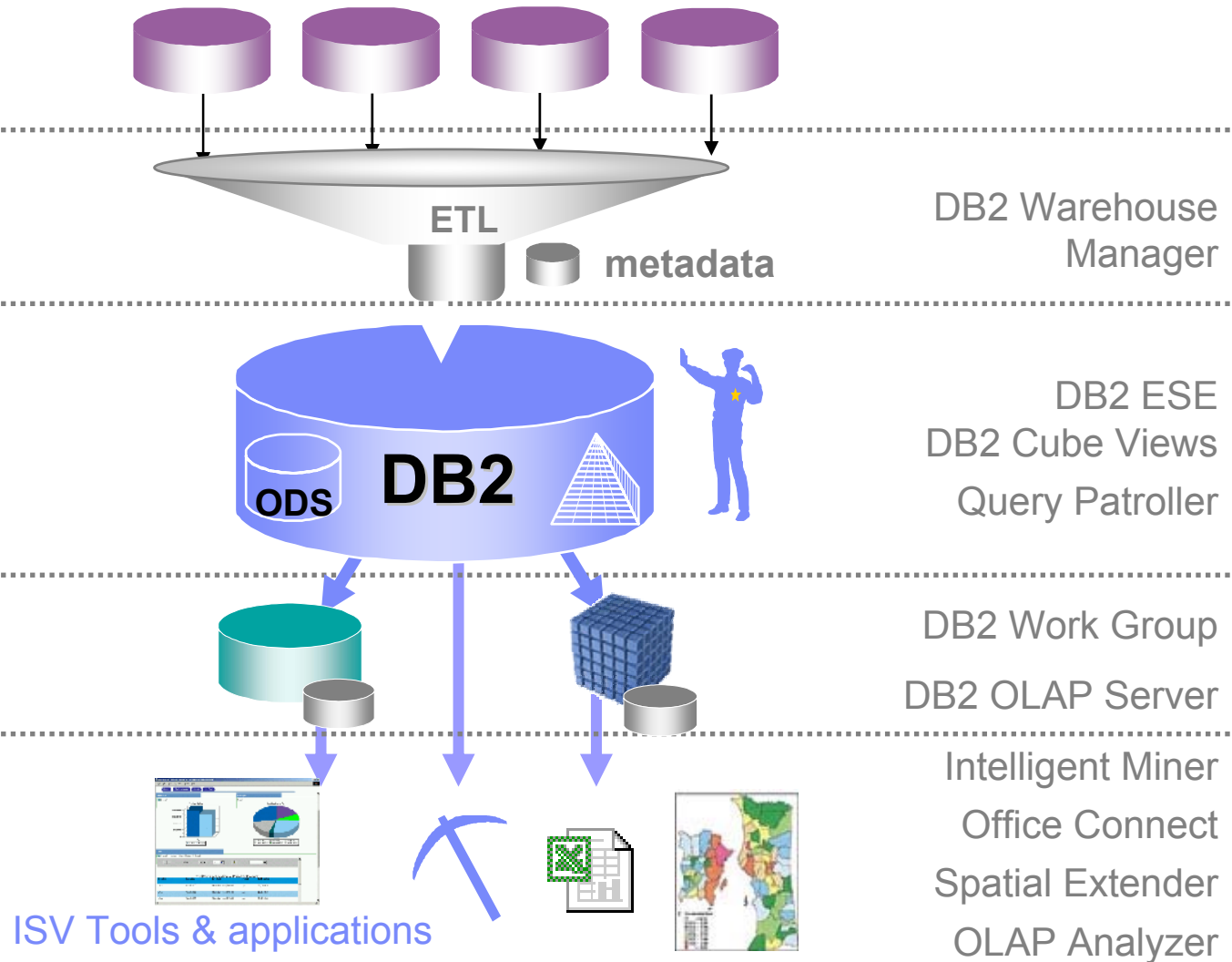*laurent_zeller@be.ibm.com*

@business on demand software

DB2 Framework for Business Intelligence
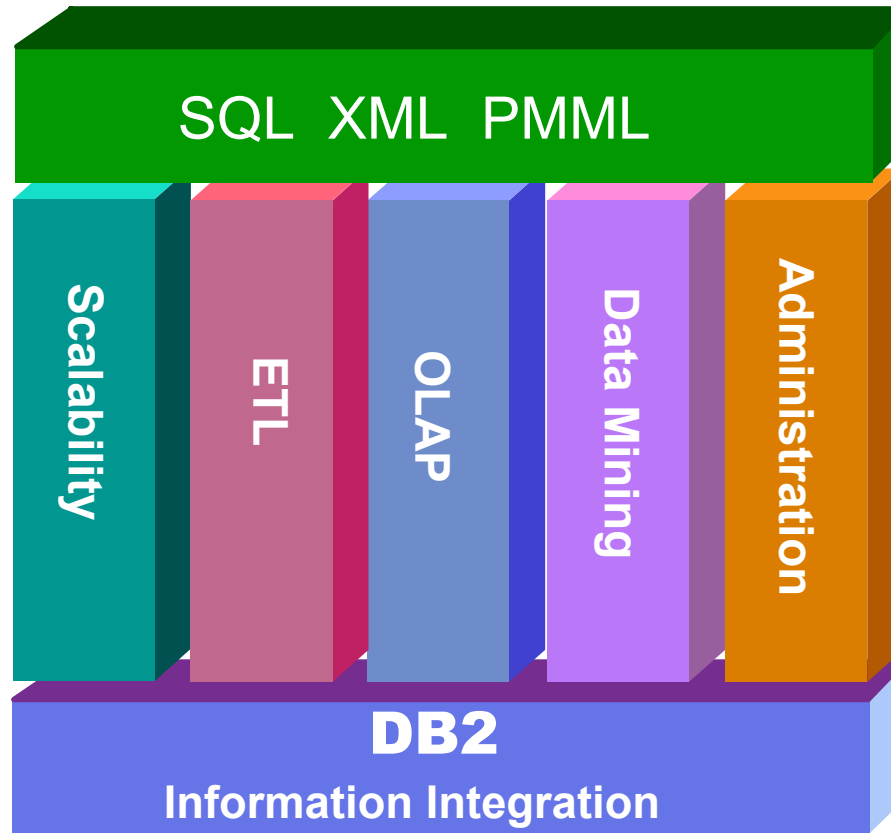
Build BI function into the database, as part of an integrated BI platform, exposed solely through open, standard interfaces, and collaborating with partners for other layers of the BI architecture

Where to look for more information ?

- iSeries Business Intelligence
  - ▶ http://www.ibm.com/servers/eserver/iseries/db2/prod_bi.html

- iSeries DB2
  - ▶ http://www.ibm.com/servers/eserver/iseries/db2

- IBM eServer Business Intelligence solution finder
  - ▶ http://www.ibm.com/servers/solutions/finder/CSFServlet.wss?mvcid=campaign&packageid=1000&campid=c124sp

- DB2 Business Intelligence
  - ▶ http://www.ibm.com/software/data/db2bi/

- DB2 Warehouse Manager
  - ▶ http://www.ibm.com/software/data/db2/datawarehouse/index.html

- DB2 OLAP Server/Analyzer
  - ▶ http://www.ibm.com/software/data/db2/db2olap/

Where to look for more information ?

- **IBM DB2 Datawarehouse Edition**
  - ▶ http://www.ibm.com/software/data/db2/udb/dwe

- **DB2 Cube Views**
  - ▶ http://www.ibm.com/software/data/db2/db2md/

- **Intelligent Miner**
  - ▶ http://www.ibm.com/software/data/iminer/

- **DB2 Information Integrator**
  - ▶ http://www.ibm.com/software/data/integration/
  - ▶ http://www.ibm.com/software/data/integration/db2ii/
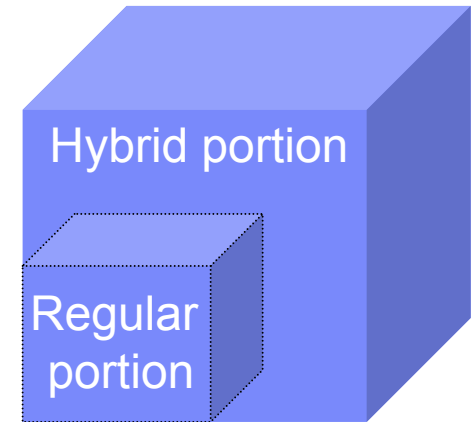
backup

OIS enabling Hybrid Analysis (HA)

■ **What is Hybrid Analysis ?**

▶ A new DB2 OLAP Server V8.1 feature that allows part of a multidimensional cube to reside in relational data source and still looks like a single, seamless cube to end users.

Hybrid portion

Regular portion

■ **How does Hybrid Analysis works ?**

▶ The HA portion of the cube (member names and cell values) are computed dynamically by SQL queries against the source database, instead of being pre-loaded and pre-calculated in OLAP Server

  ▶ Supported relational databases: IBM DB2, Oracle (8.1.6 or later), MS SQL Server

▶ Requires OLAP Integration Services (OIS) to build cubes.

▶ At Query time, the OIS metadata catalog is accessed by the Hybrid Manager to resolve the mapping to the relational database

OIS for Hybrid Analysis

- Advantages

  - ▸ The best of both world: *Combine the vast <u>scalability</u> of relational databases with the <u>powerful analytical functionality</u> of OLAP server*

  - ▸ More convenient way to «drill-through»

- Trade-offs

  - ▸ User queries that span the hybrid analysis portion of the cube usually run slower than access to the multi-dimensional data.

  - ▸ Data must be kept well-organized. Indexes are required.

IBM

OIS enabling drill-through reporting

- In general, "drill-through" is the ability to access additional information for a cell in a multidimensional cube.

  ▶ A Drill-Through report is based on an *intersection level* (member combination) that spreadsheet users can double-click to start the drill-through process.

  ▶ Spreadsheet users view or customize pre-defined Drill-Through reports that retrieve the relevant detail columns from the relational data source.

- As a drill-through mechanism, Hybrid Analysis differs from Integration Services Drill-Through Reports.

  ▶ OIS Drill-Through reports are static reports.

  ▶ Hybrid Analysis is a "live" analytical environment.

  ▶ *Hybrid Analysis does NOT replace OIS Drill-Through Reports. A Hybrid Analysis cube can still have Drill-Through Reports for its "regular" portion.*

Advantages of Drill-Through Reports

- Flexibility of information shown: Can have different reports for different intersections in the cube.

- Can see other table columns.

Advantages of Hybrid Analysis

- Easy to administer compared to managing several Drill-Through reports.

- Available from all existing client applications

- User interface is more seamless.