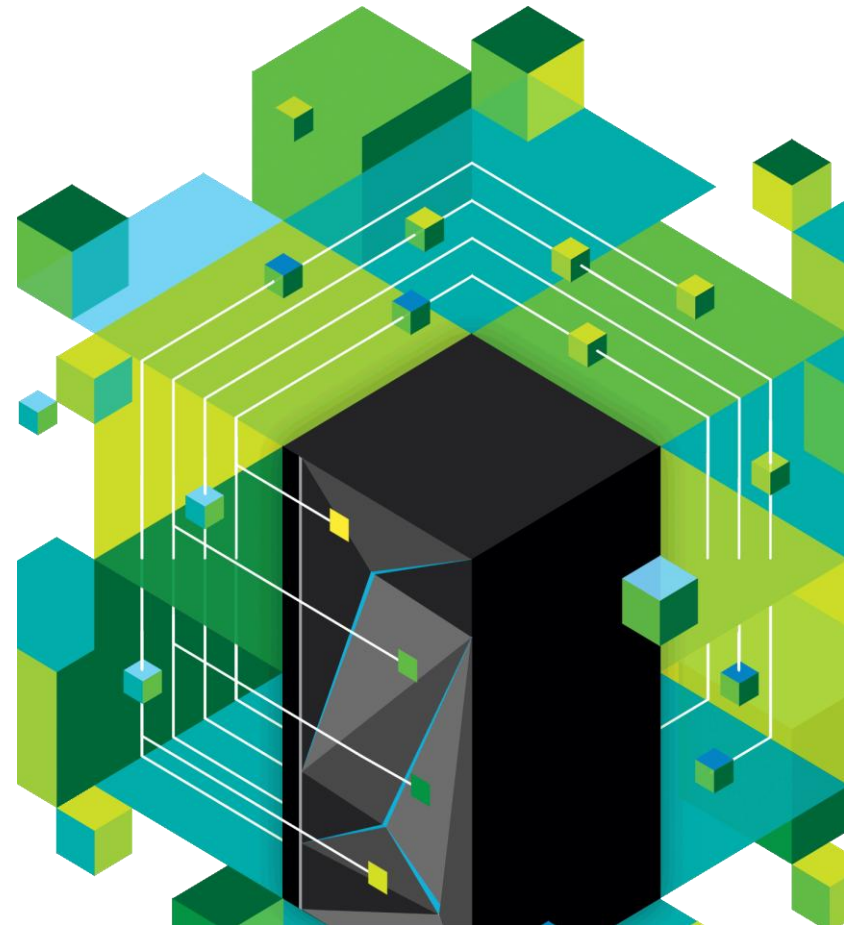# Intelligent Business Using DB2 for i

Doug Mack

Mackd@us.ibm.com

Follow us @IBMpowersystems
Learn more at www.ibm.com/power

The Agenda

- **What the heck is Big Data?**

- **Now let's talk about Analytics for the i client**

    – DB2 for i Enablers for BI/Analytics

    – Architectural Considerations/Solutions

    – The BI Tooling

Power is performance redefined

Analytics is expanding from enterprise data to big data

Volume

12 terabytes

of Tweets create daily

Velocity

5 million

trade events per second

Variety

100's video feeds

from surveillance cameras

350 billion

meter readings per annum

500 million

call detail records per day

80% data growth

are images, video, documents…

Suppose you had to scan 100Terabytes of data
With 1 node of a system with 50MB per sec, it would take 23 Days.
Imagine if you had 1000 nodes – It would be 30 minutes!

Big Data For Crime Fighting

▪**Memphis Police Department - Blue CRUSH (Criminal Reduction Utilizing Statistical History) crime-fighting methodology**

▪**Enables rich picture of the various circumstances surrounding a particular crime:**

- *Where did it occur?*

- *What was the weather?*

- *What was the weather the day or week before?*

- *Did the crime take place over a three-day weekend?*

- *Was there a payday for government workers?*

- *Was there a gun show at the local convention center?*

- *Was there a recruitment day at a local university?*

Memphis PD:
Using analytics to fight crime before it happens.

By following crime trends, Memphis Police can address, and even prevent, new ones.

How Memphis PD Got Smarter

Larry Godwin, Director of Police Services for Memphis, pioneered a way to focus existing patrol resources more intelligently—which has reduced overall crime volume by 30%.

Power is performance redefined

New Health Care Applications

Healthcare is "dying of thirst in an ocean of data"

"Medicine has become too complex. Only about 20% of the knowledge clinicians use today is evidence-based."

Steven Shapiro
Chief Medical & Scientific Officer
University Pittsburgh Medical Center

Medical info is doubling every 5 years

81% of physicians spend < 5 hrs / month reading medical journals

1.5M errors in the way medications are prescribed, delivered and taken

$750B, or 30 cents of every dollar, is wasted in US alone

Power is performance redefined

Scalable: Different types of analytics require a scalable IT infrastructure

Different types of analytics (OLTP, Data warehouse, Streaming Data, OLAP, Operational Analytics, Ad-hoc reporting, time series, deep analytics)…
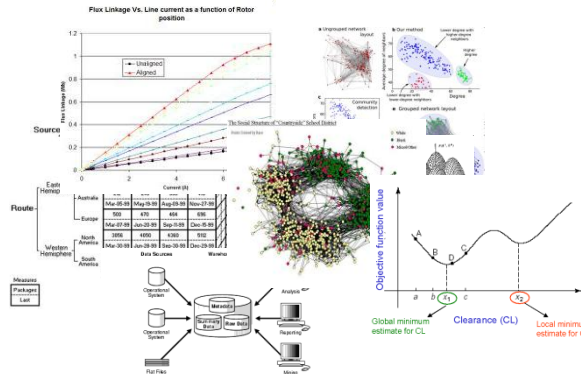
…need to access data differently and require compute *and* storage resources that are distinctly different and often highly scalable.

Vestas
No. 1 in Modern Energy

Analytics challenge: Vestas is addressing the challenge of wind turbine placement by analyzing petabytes of varied data such as weather reports, tidal phases, geospatial and sensor data, satellite images, deforestation maps, and weather modeling to pinpoint installation. The output drives the business model for their client's wind turbine farms.

IT Architectural approach: Vestas realized early on that a new IT architecture was crucial to achieving this objective. To that end, they deployed an IBM system capable of 150 trillion calculations per second, combined with a distributed parallel file system with near storage compute capabilities. All within a highly energy efficient footprint.

Benefit: The analysis, which used to take weeks, can now be done in less than one hour.

"Using IBM software and systems, we can now answer these questions quickly to identify new markets for wind energy and help our clients meet aggressive renewable energy goals." – Lars Christian Christensen, Vice President of Plant Siting and Forecasting, Vestas Technology R&D

Vestas IBM Video

6   **Power is performance redefined**

So What is IBM Doing about Big Data?

- **Creating offerings and services around key technologies**
  - Large Scale Data warehousing
  - Statistical Analysis
  - Stream Analysis
  - Hadoop
  - Massively parallel processing clustered configurations
  - Query Acceleration
- **It addresses the EXPONENTIAL growth of data !**
  - 90% of data has been created in the last 2 years. WOW
    - Sensors gather climate information
    - Posts to social web sites
    - Cell phone GPS signals
    - Web page interaction

Power is performance redefined

Example: IBM Systems for Smarter Analytics

IBM Systems is the market leader in support of analytical workloads

Dell 9%
Other 32%
HP 23%
Sun 9%
IBM STG 27%
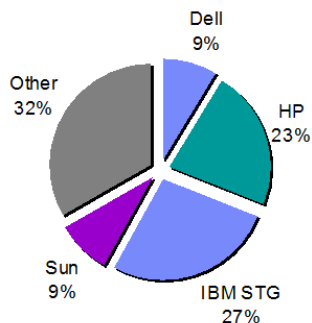
Workloads:
Data Analysis/Data Mining
Data Warehousing/Data Mart
Scientific/Engineering/Industrial R&D

IDC Server Workload study 2011

What the Analysts are saying

"When we commenced this exercise we expected to find that there were some areas in which IBM excelled and others in which Oracle did so. We have been surprised to find that that is not the case and that the ***IBM Smart Analytics System out-competes Oracle Exadata in almost every area we have examined"*** *Philip Howard, Bloor Research*

Three-year costs for Smart Analytics System 7700 are ***43 and 40 percent less than those for Oracle and Teradata systems*** *respectively.* Source: Cost/Benefit Case for Enterprise Warehouse Solutions, International Technology Group

## IBM and Analytics at a glance:

- More than **$14B** in Acquisitions Since 2005
- More than **10,000** Technical Professionals
- More than **7,700** Dedicated Consultants
- **Largest** Math Department in Private Industry
- More than **27,000** Business Partner Certifications

Power is performance redefined

Scaling out to support more data or more users

Shared Nothing, Massively Parallel Design
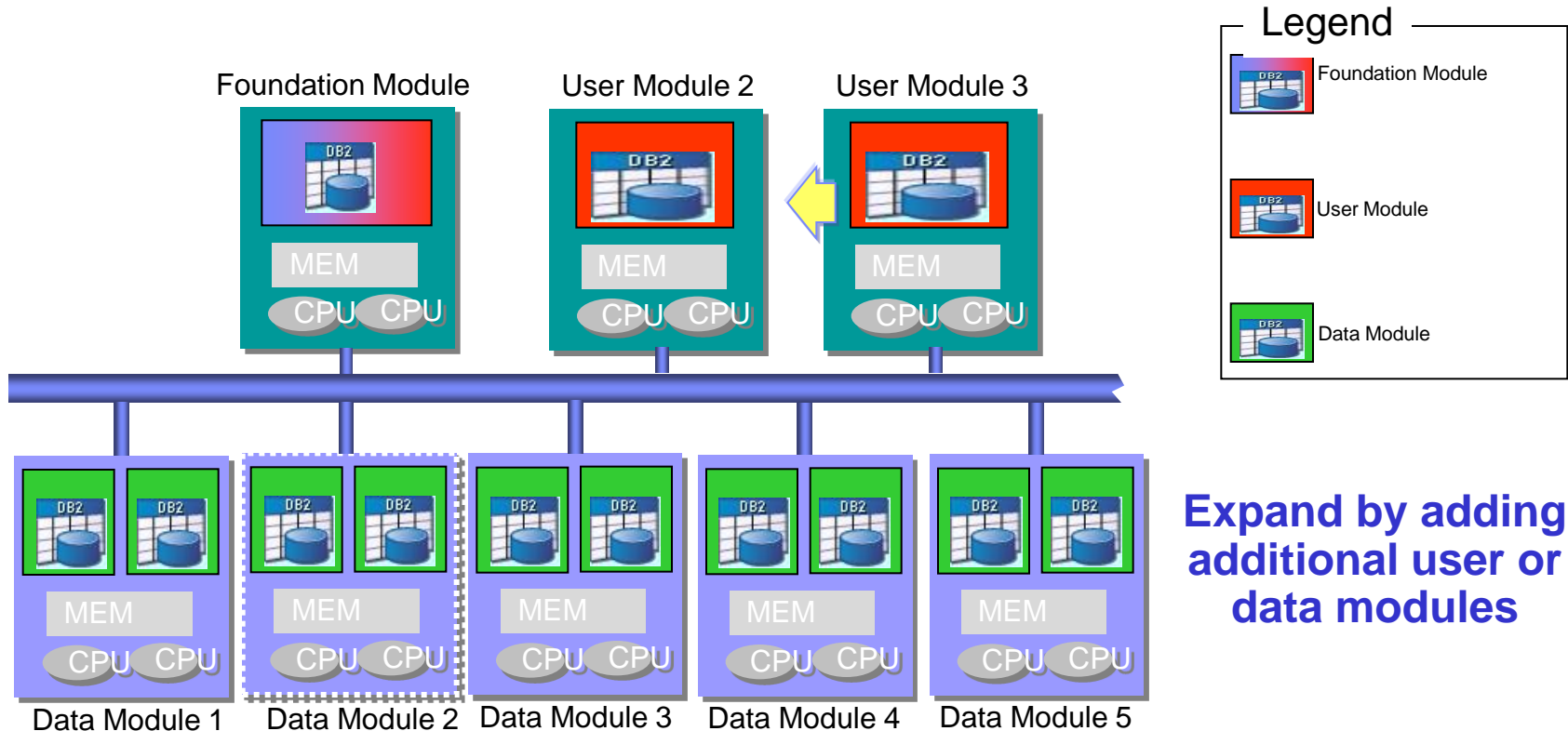
Expand by adding additional user or data modules

- **Balanced system design**
  - **System modules with optimal processor, memory, and I/O specifications**
- **Scale-out by adding additional system modules**
  - **Which always include balanced I/O**
- **Proven "best practice" for large scale data warehousing**

Power is performance redefined

Big Data Technologies

- **Info Streams**

  - IBM Research developed language to analyze web data in real time

  - Info Streams is a specific proprietary language using commodity hardware

  - Examples of applying Infostreams

    - USC project to monitor the sentiment of voters in the recent presidential election campaigns – tweet by tweet by tweet !!

    - Challenge…how do you deal with 60,000 tweets per minute ? WOW !!!

- **Big Insights puts an enterprise solution together around Hadoop**

  - Hadoop is a framework from Apache for running applications (aka jobs) on large clusters built on commodity hardware capable of processing petabytes of data

    - Hadoop implements a computational paradigm named **Map/Reduce**, where the application is divided into self contained units of work, each of which may be executed or re-executed on any node in the cluster

    - The goal is to move the processing closer to the data – not the reverse !!

  - Big Insights simplifies the environment with installation, visualizations, enterprise level security and availability, etc.

    - Big Sheets Visualizations; Adapters for DB2, Netezza, etc.

    - PIG, or PIGLatin is the progamming language; HIVE is the SQL like query language
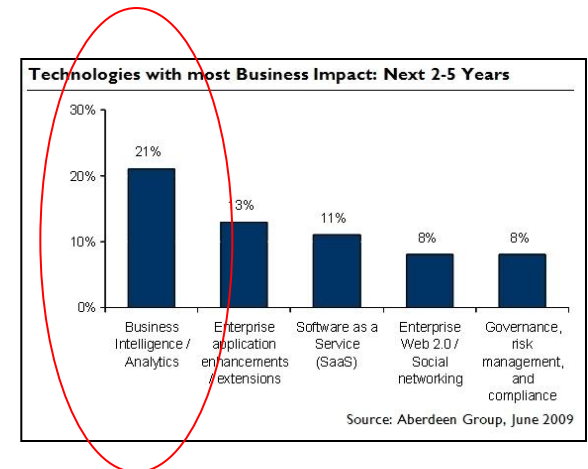
- **Social Media Analytics**

  - Also known as Cognos Insights

    - Analyzes web data (what it can….e.g., facebook pages are blocked) to understand "pulse" of the brand or company

Power is performance redefined

The Agenda

- What the heck is Big Data?

- **Now let's talk about Analytics for the i client**

  - DB2 for i Enablers for BI/Analytics

  - Architectural Considerations/Solutions

  - The BI Tooling

Power is performance redefined

What about the i Client?

- **SMB companies have similar requirements for analytics, but more constraints:**
  - Budget is smaller
  - Staff is limited– Admin, DBAs
  - "*The full capabilities of most BI tools **are overkill** for the typical SMB (Small and Medium Business).*" – Aberdeen Group

- **The i market is 85% SMB**
  - Heavily relying on old technology (Query/400) for reporting

- **Requirements for the IBM i SMB BI Client**
  - Low cost
  - Simplified Infrastructure
  - Leverages existing environment (DB2 for i)
  - Reduced dependency on I/T
  - Without sacrificing the benefits of enterprise BI

Technologies with most Business Impact: Next 2-5 Years

Business Intelligence / Analytics	Enterprise application enhancements / extensions	Software as a Service (SaaS)	Enterprise Web 2.0 / Social networking	Governance, risk management, and compliance
21%	13%	11%	8%	8%

Source: Aberdeen Group, June 2009

"The IT departments at midsize companies are usually fairly small. They're often very clever at getting things done more efficiently and with less expense than their larger brethren, but they have their limits. For example, if an SMB's IT department has 10 people and three or more of them would be needed on a BI team, that's a substantial percentage of the total IT head count"- Claudia Imhoff

http://searchbusinessanalytics.techtarget.com/news/2240113703/Coping-with-midmarket-BI-project-management-challenges-Tips-for-SMBs?asrc=EM_NLN_16106429&track=NL-544&ad=859763)

Power is performance redefined

Analytics and Business Intelligence

Business Intelligence

Predictive Analytics

REPORTING
WHAT HAPPENED?

Query/
Reporting

ANALYSIS
WHY DID IT HAPPEN?

OLAP

PREDICT
WHAT WILL HAPPEN?

Data Mining

MONITOR
*WHAT JUST
HAPPENED?*

Dashboards/
Scorecards

**Historical Data
Data Warehouses and Data Marts)**

**Real-Time Data
ODS = Operational Data Store**

Source Systems
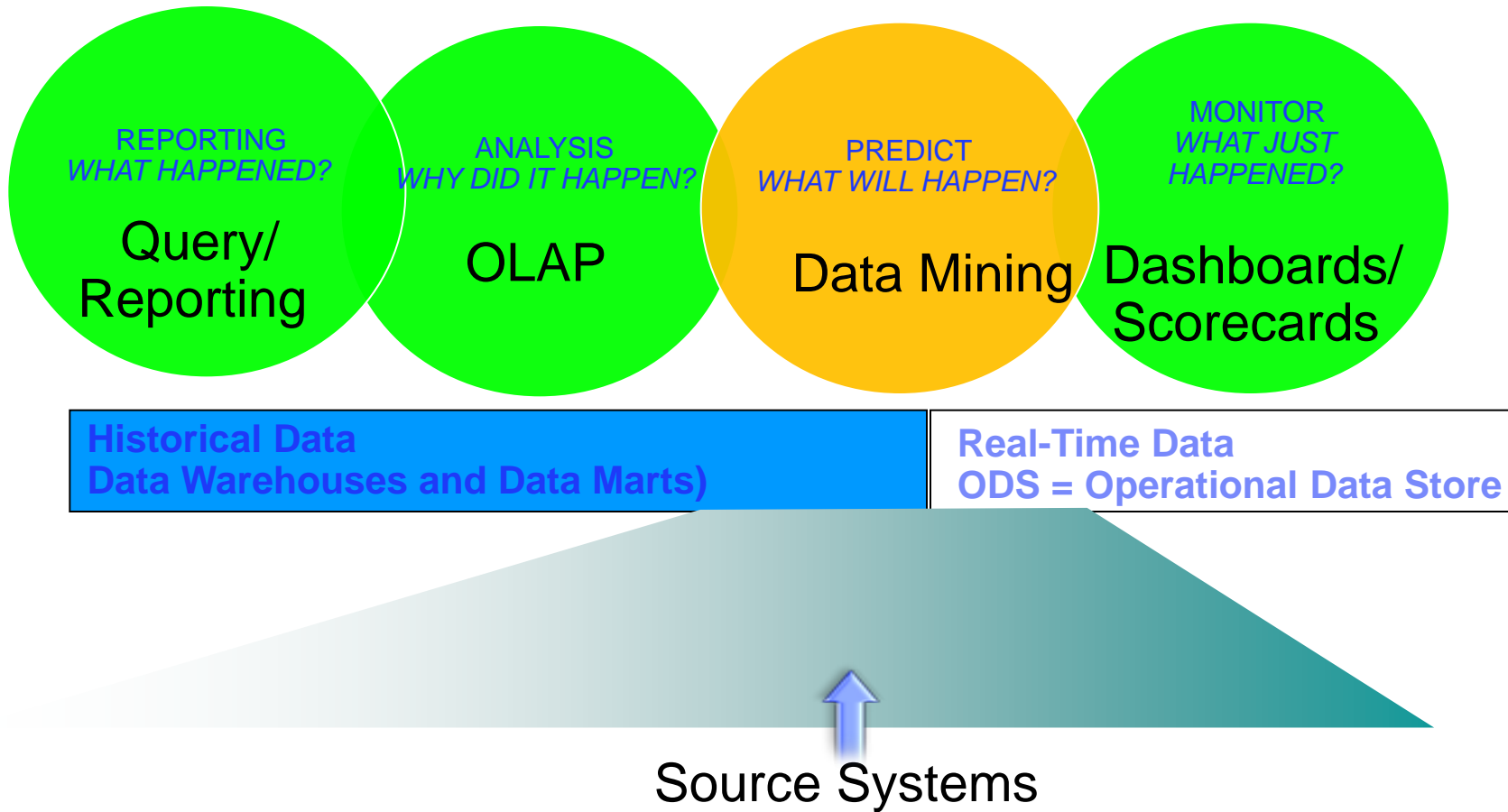
Source: The Data Warehousing Institute, Smart Companies in the 21st Century, July 2003

Power is performance redefined

DB2 Web Query for i – Intuitive, Insightful, Extensible
Fast, easy access to business information assets for query, reporting and analysis

- Easily spot trends or exceptions in data with real time reports

- Give Executives the means to track how the business is performing through intuitive Key Performance Indicator dashboards

- Create self-service reporting environment that eliminates dependency on I/T

- Provide data to spreadsheet aficionados painlessly and in real time

- Execute and distribute reports in many different formats - on demand or scheduled – via e-mail or saved for later view

- Integrate reports into existing applications for seamless access to data

Power is performance redefined
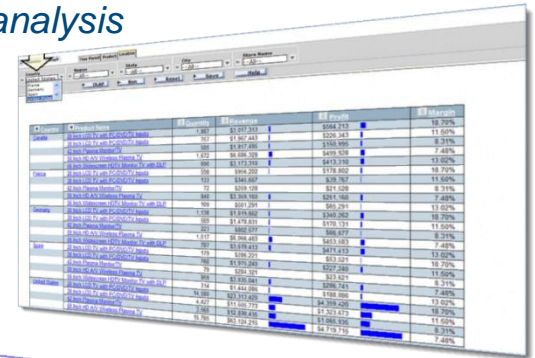
Importance of Metadata

- ***Impact* of it's absence:**

  – the 1999 NASA Mars Climate Observer mission failed because of a metadata problem…

  – thrust calculation data was provided in the US measurement scale of pounds/square foot, but was interpreted as metric numbers representing Newtons/second.

  – the *data* was correct. The *understanding* and *usage* of it was not.

  – Result: the orbiter went missing

  (it probably crashed on Mars)

  and a $300M mission failed!

The Difference between a GUI based query tool and Business Intelligence is a META DATA Layer which DB2 Web Query deploys

The Meta Data Layer allows you to:

- **Document your database (which is probably not documented now)**

- **Standardize on the meaning of data in one place (single version of truth)**

- **Expand your community of report authors because they don't have to know the intricacies of the data or SQL**

- **Integrate DB2 functions and stored procedures into your reports (very powerful and flexible)**

- **Define relationships between tables/files, and between data elements to hide complexities of the database and provide advanced analytics like auto drill down capabilities (with no programming required)**

Power is performance redefined

Leveraging a Meta Data Layer to Shield Complexities of Database

- **Standardize Field/Column Formats**
  - Ex, use commas, set currency symbol, suppress leading zeros

- **Standardize/Decompose Date Fields**
  - Ex: Integer defined as MMDDYYYY

- **Create Filters**
  - Ex: Define a set of countries as "Europe"

- **Define JOINS and Dimensional Reporting**

- **Create Business Views**
  - Organize Columns/fields for easier report development

- **Define Stored Procedures, Views, UDFs to Meta Data Layer for some REALLY COOL function**
  - Row/Column Level Security
  - Text Search Engine
  - Bring in weather report, local news and more

Power is performance redefined

The Power of Meta Data: Example

- We have a FIELD defined that has multiple sub-fields embedded

- Our report wants to show CREDIT RATING – based on the value of that field, we want CREDIT RATING to be set to "GOOD", "BAD", or "AVERAGE"

- What we need to do is pull out the 10th digit of that CUST_CODE field and apply a rule.

- BUT WE DON'T WANT TO HAVE TO DO THIS FOR EVERY REPORT THAT INCLUDES CREDIT_RATING

10 character Cust_Code

751246	935	0
191300	371	6
747538	619	8
771482	503	6

Credit Rating

Orders placed in last year

Customer Number
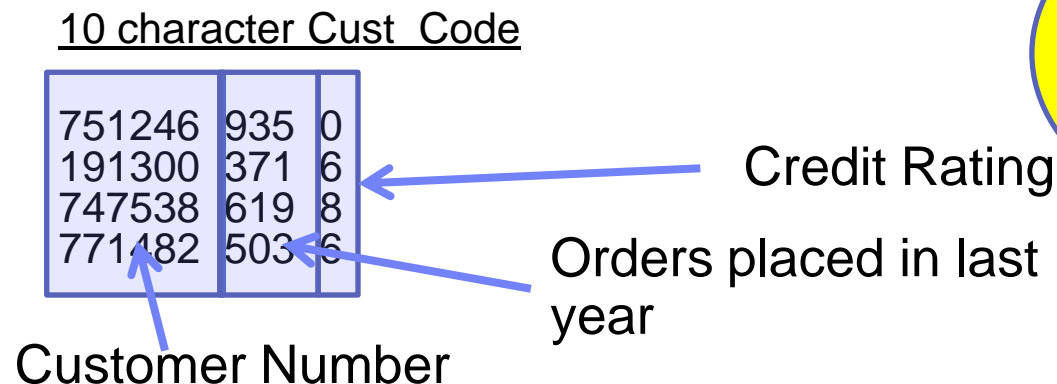
For credit rating >7 is "Good", 4-6 is "Average", <4 is "Bad"

The Power of Meta Data: Example

- **Create a new field in meta data called CREDIT_RATING**

- **Define the rule, and TEST**

IBM i client issue: Query/400 reporting nightmares!

Many i clients have hundreds if not thousands of Query/400 definitions that have proliferated over time

- Redundancy of reports is commonplace

- Maintenance of reports is a nightmare

- Dependency on report "authors" is problematic

- Query/400 CANNOT use the latest DB2 for i Query processing technologies

- Extraneous steps taken for simple problems (getting data to a spreadsheet, or doing advanced data calculations)

Power is performance redefined

Why Modernize

- Turn DATA into INTELLIGENCE

- Delivered at the right time, in the right formats

- Take advantage of SQL Capabilities

- Document the database, and standardize on data meaning (single version of the truth)

- Get I/T out of the Report Writing Business
    - With self service reporting solutions

- Eliminate redundancy

- Adapt more quickly to changes in the operational systems/databases

- Deploy reports in a variety of highly intuitive ways
    - Mobile, Excel, Dashboards, OLAP

- Improve the perception of the "AS/400"

Power is performance redefined

The Greatest BI Tool in the world is useless if it doesn't perform !

"The most widespread technical problem reported by practitioners was slow query performance."

- **Survey of over 2000 companies that have implemented Business Intelligence Applications**
  - The BI Survey 8 – Nigel Pendse

Infrastructure Matters….

- **Scalability**
  - The average data warehouse doubles in size every 18 months

- **Optimized for Workload**
  - Complex Querying and Aggregation is very different from OLTP workloads and requires different tuning

- **Resilient**
  - Analytics are as important in terms of availability as line of business applications

- **Minimize Risk**
  - Leverage what you do best to minimize risk
  - Security policies; HA/DR infrastructure; Operational policies and procedures

Power is performance redefined

DB2 for i Enablers for Analytics – Built into DB2 for i

- **SQL Query Engine (SQE)**

  – Real Time Statistics

  – Automated Index Creation and Query Adjustment

- **Database Parallelism**

  – Leverage Multiple Cores

- **Materialized Query Tables**

  – Aggregate Processing

- **Encoded Vector Indexing**

  – IBM Patented Query Acceleration Technology

- **Management Tools – System i Navigator**

  – SQL Plan Cache, Monitor Analysis of Queries

  – Index Advice
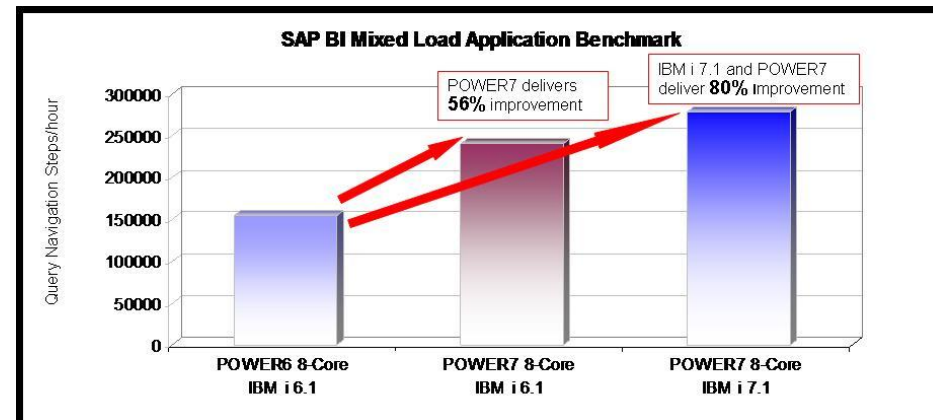
*See detailed certified benchmark results at
http://www.sap.com/solutions/benchmark/bimxl.epx

Power is performance redefined

DB2 Symmetric Multiprocessing (feature of IBM i)

- **SELECTING**
  - Index scan or probe
  - Table scan or probe via bitmap or RRN list
  - Table scan
- **JOINING**
  - Index scan or probe
  - Hash
- **GROUPING**
  - Index scan or probe
  - Hash
- **ORDERING**
  - Index scan or probe
  - Sort

SMP enabled

- Creating temporary indexes for joining, grouping or ordering is SMP enabled

Power is performance redefined
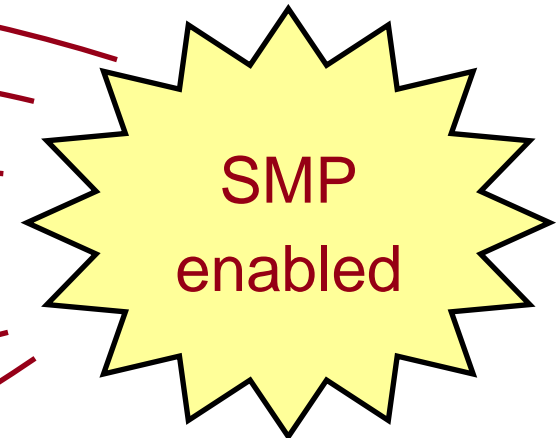
SQL Query Engine

- **Advanced query optimization, query execution engine, and management tools**
  - Part of DB2 since V5R3
    - Enhanced with each subsequent release
  - Leverages more DB2 performance and management facilities
    - Encoded Vector Indexes, Materialized Query Tables
    - SQL Plan Cache (part of IBM i Navigator)

- **7.1 Enhancements**
  - Adaptive Query Processing (AQP)
  - Support for Logical Files

- **NOTE: CQE, or "Classic" Query Engine, is also part of DB2 to support non SQL Standard Interfaces for accessing DB2**
  - Query/400
  - Some ISV Applications
  - Limited ability to leverage

Power is performance redefined
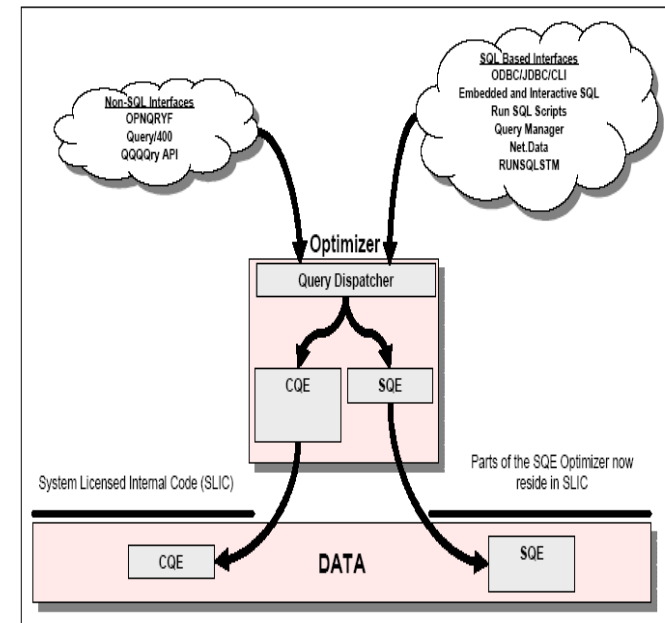
Materialized Query Tables (MQT)

Before...

Final Select

36

Hash Scan

Temporary Distinct Hash Table

6001215

Table Scan
STAR1G.ITEM_FACT

```
SELECT year, quarter, month,
        SUM(revenue_w_tax) AS srevenue_w_tax,
        SUM(revenue_wo_tax) AS srevenue_wo_tax,
        SUM(profit_w_tax) AS sprofit_w_tax,
        SUM(profit_wo_tax) AS sprofit_wo_tax,
        SUM(quantity) AS squantity,
        COUNT(*) as number_items_per_group
FROM    ITEM_FACT
GROUP BY year, quarter, month;
```

After...

**Highlight
MQT
In V5R4**

Final Select

36

Table Scan
STAR1G.TEST_MQT

Base table(s) replaced by MQT

Without MQT...
Scan and aggregate
6,000,000 rows
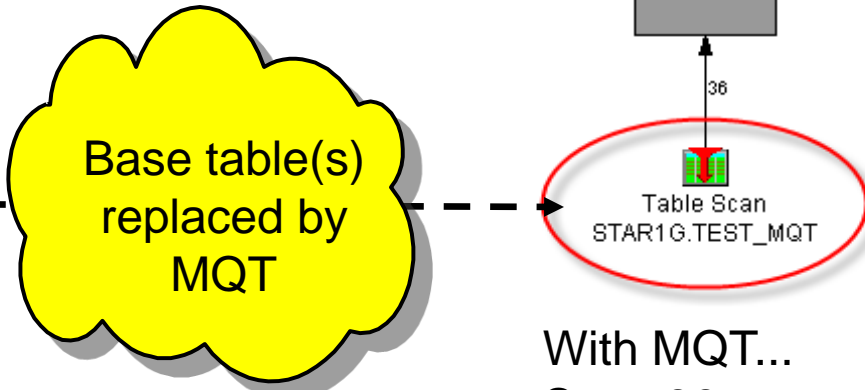
With MQT...
Scan 36 rows

V6R1: Improved Refresh Performance!

Power is performance redefined

Encoded Vector Index (EVI) Aggregates (7.1)

Symbol Table				
Key Value	**Code**	**Count**	**Include Sum()**	**Include Sum()**
Arizona	1	5000	1500	2005
Arkansas	2	7300	3200	450
…				
Wisconsin	49	340	575	1200
Wyoming	50	2760	210	0

Optional (7.1)

Vector	RRN
1	1
17	2
5	3
9	4
2	5
7	6
50	7
49	8
5	9
…	…

- Symbol table contains information for each distinct key value
    - Each key value is assigned a unique code (key compression)
    - Code is 1, 2, or 4 bytes depending on number of distinct key values

- Rather then a bit array for each distinct key value, use one array of codes

Power is performance redefined

Establish Best Practices for Managing Query Performance

- **DB2 for i Administration**
  - Part of i Navigator
  - Performance Analysis Tools
    - Database Monitors
    - SQL Plan Cache
    - Visual Explain
    - Index Advice
    - Index Evaluation
    - Reporting Functions
      - Save to spreadsheet
      - Filters
  - Documentation Tools
    - Database Navigator
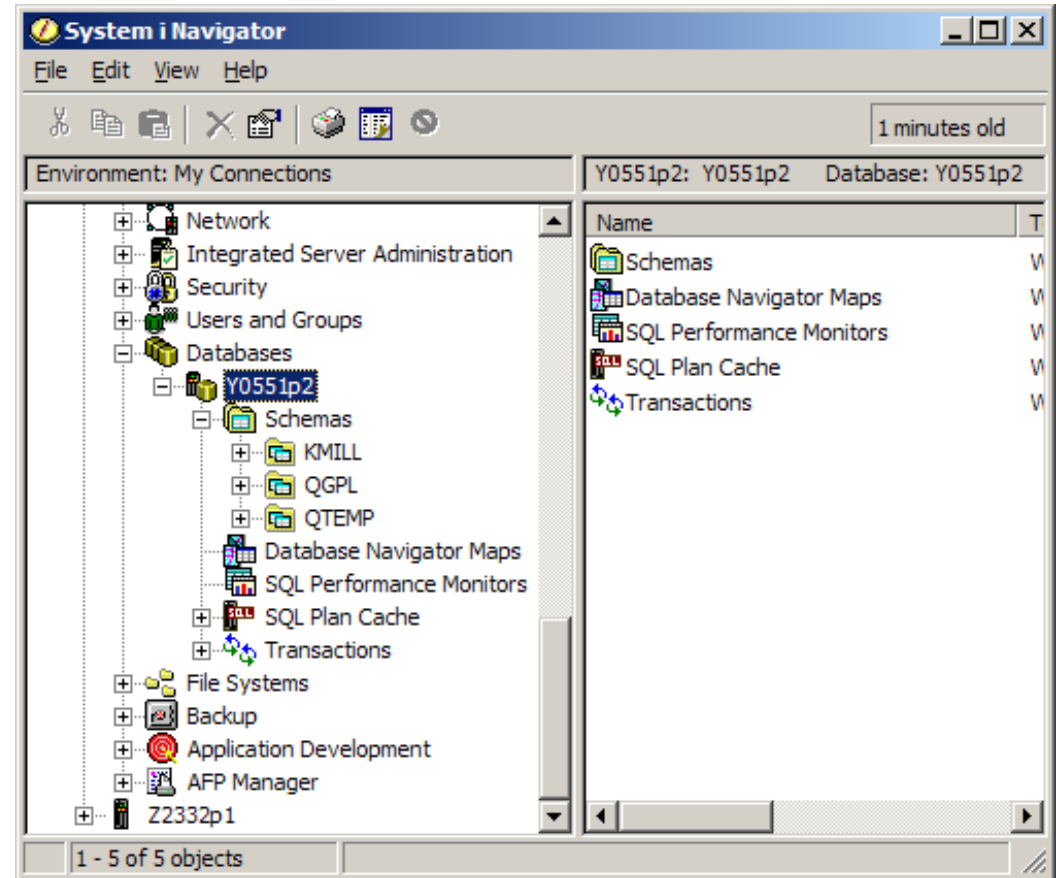
What is the right INFRASTRUCTURE to support IBM i BI?

Operational Reporting	**Operational Data Store**	**Data Warehouse**
•Deployed On Existing i server •Simple reporting and BI •No data replication or transformations •Mixed workloads (OLTP and heavy query)	•Optimized environment for operational reporting •Separate BI from OLTP workload •Multi-purpose 2nd Sustem •Simple, low cost data replication •<u>Foundation</u> for data warehouse	• Extend value of IBM i for Business Intelligence to Data Warehouse • Add an ETL tool for data transformation • CDC (transport) and an ETL (transform) can provide near real time analytics • Fully leverage advanced DB2 i technology
DB2 Web Query	**IBM i for Business Intelligence**	**IBM i for Business Intelligence + ETL and Services**

Mixed Workload *Diminished Efficiencies*	**Workload Optimized** *Simplify - Accelerate Value - Reduce Cost*	
DB2 Web Query Standard Edition Getting Started Services Memory and CPW !	**IBM i for BI Solution Additional IBM i Licenses for other uses**	**IBM i for BI Solution Extract/Transform/Load (ETL) Tool Additional Services**

Lower ◄————— **Complexity/Cost** —————► Higher

Option 1: Querying Production Databases

- **Isolate workloads as best you can**
  - Create separate subsystem for query jobs
    - DB2 Web Query installs into its own subsystem
  - Keep memory pool LARGE with reduced number of active jobs

- **Indexing and aggregate strategy plays a major role in query performance**
  - MQTs, EVIs, and traditional indexes
    - Read white papers on MQTs and Index and Statistics Advice
      - http://www-03.ibm.com/servers/enable/site/bi/strategy/index.html

- **Leverage Query Governors**
  - QQRYTIMLMT

- **Consider Report Delivery**
  - Do you need "real time" vs. batch?
  - Control queries with parameterized reports
  - Use DB2 Web Query "Active" technologies
    - Once the report is run you are no longer impacting system

- **Let DB2 do as much work as possible**
  - DB2 Web Query tips and techniques (e.g., joins)

Power is performance redefined
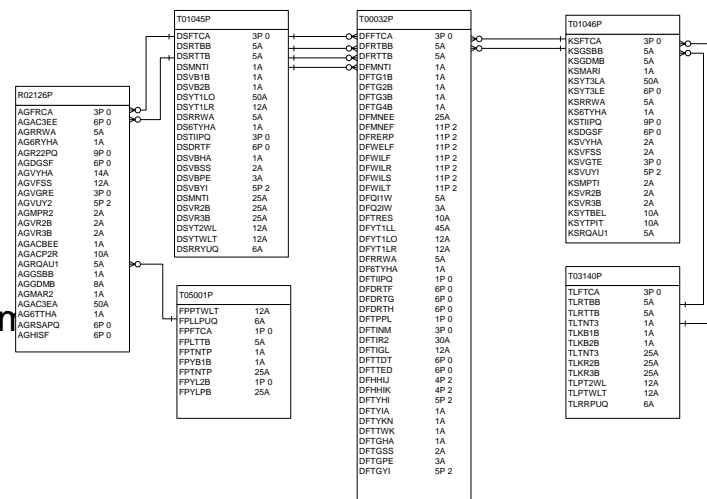
Option 2: Isolation through an *Operational Data Store*

- **What is an Operational Data Store (ODS)?**
  - A COPY of the operational (transaction) data base most often used for reporting purposes
  - ISOLATED from production workloads
  - Kept up to date based on requirements
    - Near real time?
  - The data model stays pretty much the same as production, with possibly some minor changes
    - Could have subset of fields/ columns
    - Could contain more historical data than production systems
    - Could contain minimal data transformations
    - Could encrypt/mask certain data elements
  - Enhanced for performance
    - Aggregations of the data
    - Indexing
  - Security model might be different
  - Platform is TUNED for reporting purposes
  - Get to IBM i 7.1 and POWER7 for this workload

Power is performance redefined

Example: Populating the ODS with IBM's Infosphere CDC

- **Changed Data Capture (CDC) Replication Software**

  - Based on journaling

    - Remote or Local

  - Requires both before and after images to be store in journal receiver

  - Techniques to do INITIAL loads with sync points

  - Java GUI to set up and monitor (no programming required)

  - Minimal transformations

Monitoring and Configuration

DB2 for i   Journal Receivers   Source Engine   TCP/IP   Target Engine   DB2 for i ODS

Power is performance redefined
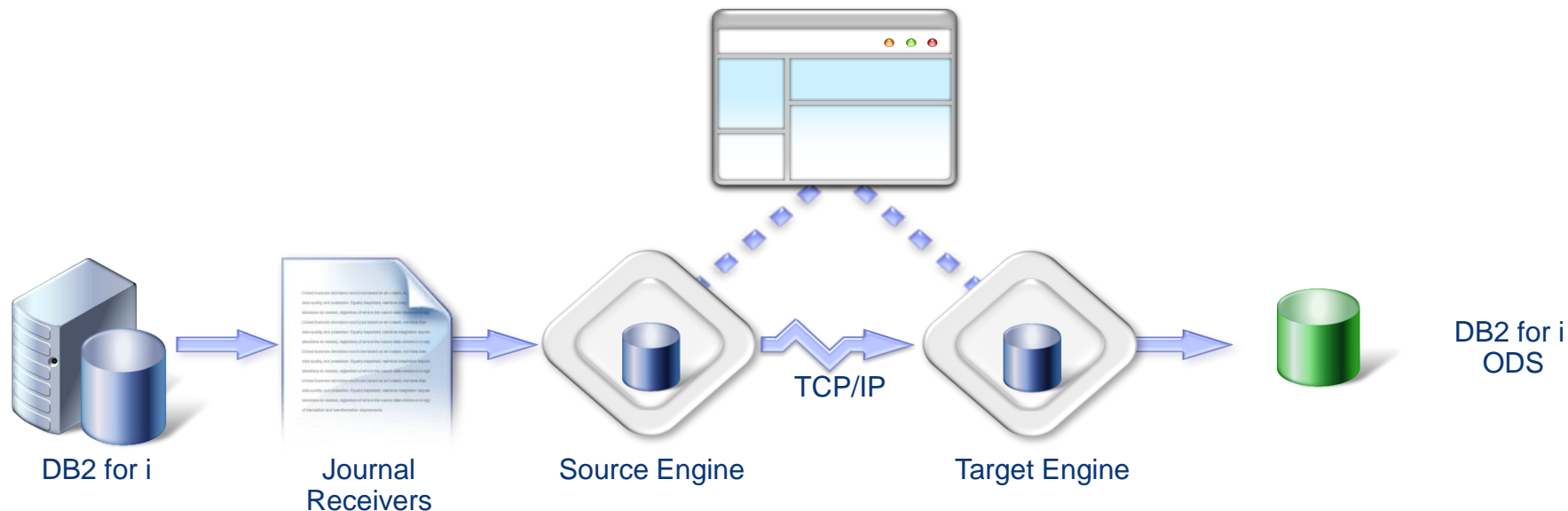
New IBM i for Business Intelligence Packaged Solution

Production System
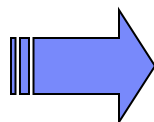
IBM i for BI

Extract, Transport and Load Data using **InfoSphere Software**

Operational Reporting in **3 Days**
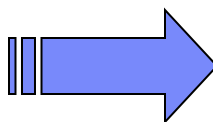
IBM i for Business Intelligence

- Installation Services Included in Price
- Includes DB2 Web Query Standard Edition
- Replicate production database to 2nd system
- Isolate query workloads (tune, optimize)
- LOW Entry price point that you can build on
- Offered in 3 Sizes: Small/Medium/Large

Power is performance redefined

Option 3: The Data Warehouse

- **Adds Extract, Transform, and Load**

Data Warehouse

ETL

Aggregations

BI

Operational Source Systems
(Production Databases)

The Data Warehouse Model

PRODUCTS

PRODUCT_NUMBER	5P 0
PRODUCT_DESCRIPTION	42A
BRAND_CODE	5A
BRAND_DESCRIPTION	20A
ORIGIN_CODE	5A
ORIGIN_DESCRIPTION	20A
FAMILY_CODE	5A
FAMILY_DESCRIPTION	20A
COST	9P 2
BASE_PRICE	9P 2
PRODUCT_WEIGHT	9P 4
PRODUCT_VOLUME	9P 4
LOAD_DATE	DATE
LAST_CHANGE_TIME	TSTP
STATUS_FLAG	1A

INVOICE_LINES

INVOICE_NUMBER	7P
INVOICE_LINE_NUMBER	3P
PRODUCT_NUMBER	5P
CUSTOMER_NUMBER	10
SELLING_COMPANY	5A
SUPPLY_WAREHOUSE	5A
QUANTITY_ORDERED	11P 0
QUANTITY_SHIPPED	
TOTAL_DISCOUNT	
NET_PRICE	
BASE_PRICE	
UNIT_COST	
EXTENDED_COST	
EXTENDED_PRICE	
MARGIN	
SALES_REP	
COMMISSION_VALUE	
INVOICE_DATE	
SHIP_DATE	DATE
DELIVERY_DATE	DATE
INVOICE_TIME	TIME
MONTH_NUMBER	2P 0
WEEK_NUMBER	2P 0
LOAD_DATE	(DATE)

...ERS

...ER_NUMBER	10A
...ER_NAME	35A
ADDRESS_LINE_1	35A
ADDRESS_LINE_2	35A
	35A
..._CODE	2A
...DE	10A
...ACT_NAME	35A
...HONE	15A
..._REP_DEFAULT	5A
...OMER_CATEGORY	5A
...OMER_CLASS	5A
...N_CODE	5A
...DATE	DATE
...CHANGE_TIME	TMSTP
STATUS_FLAG	1A

> Meaningful table and column names

> Complex calculations already done

> De-normalized design reduced to only 3 tables

> Only includes the columns we care about

> Dates are true date columns

Data Warehouse Schema

Power is performance redefined

Transformation Example: Surrogate Keys

Customer File - US	
CUSTNO	CUSTNAME
1001	John Smith
1002	Mary Jones
1003	Chris Anderson
1004	David Perry

Customer File - Canada	
CUSTNO	CUSTNAME
1001	Harry Potter
1002	Jeremy Carr
1003	Penny Hayes
1004	Debbie Thornton

Surrogate key is a sequential number with no correlation to replaced value(s)

Customer File - Data Warehouse			
CUSTNUMBER	CUSTNAME	REGION	OLDNUM
1	John Smith	US	1001
2	Mary Jones	US	1002
3	Chris Anderson	US	1003
4	David Perry	US	1004
5	Harry Potter	CANADA	1001
6	Jeremy Carr	CANADA	1002
7	Penny Hayes	CANADA	1003
8	Debbie Thornton	CANADA	1004

PK

Secondary Index

Power is performance redefined

Star / Snowflake Schema Processing in DB2 for i

Final bitmap or RRN list represents local selection

Item_Dim keylist

Itemkey

Store_Dim keylist

Storekey

Date_Dim keylist

Datekey

Keylist used as local selection (LPG)

Fact_Table

| Itemkey |
| Storekey |
| Datekey |
| Sales |
| Quantity |

Final Bitmap or RRN list

3 Bitmaps or RRN lists *AND*ed together

EVI 1
Itemkey

EVI 2
Storekey

EVI 3
Datekey

Dynamic Bitmap or RRN list

Dynamic Bitmap or RRN list

Dynamic Bitmap or RRN list

Select store, item, date, weather, sum(sales), sum(quantity)
from item_dim, store_dim, date_dim, fact_table
where **itemkey in (...keylist...) and storekey in (...keylist...)
and datekey in (...keylist...)**
and itemkey=itemkey, storekey=storekey, datekey=datekey
group by store, item, date, weather

Power is performance redefined

ETL Alternatives

- **Do it yourself**
  - Custom coding
  - Consider use of SQL vs. RPG record level processing

- **IBM i based (DB2 Web Query Meta Data Integration)**
  - Information Builder's Data Migrator
    - www.ibi.com
  - Coglin Mill's Rodin DB2 Web Query Edition
    - www.thinkrodin.com

- **High End (AIX LPAR)**
  - IBM InfoSphere Data Stage
    - Strong source and target support
    - Parallelism built into the load processes
    - Many data transformations built in

Power is performance redefined
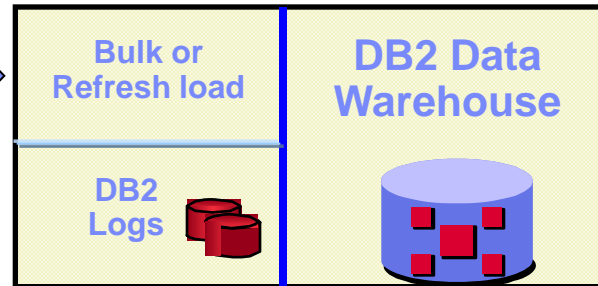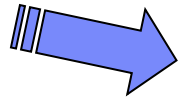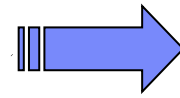
Expanding the IBM i for BI into Data Warehousing

Production System

Purchase <u>additional</u> ~~ETL~~ Tools/Services

IBM i for BI
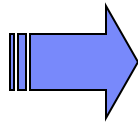
Extract and Transport Data **using InfoSphere Software**

Bulk or Refresh load

DB2 Data Warehouse

DB2 Logs

Transform <u>and</u> Load Data into DB2 Data Warehouse

IBM i for Business Intelligence

- ETL process TRANSFORMS and cleanses data
  - Bulk load or refresh
  - Scheduled or continuous updates
- Restructuring of the data improves analytics

  (for example: create a customer profiling database)
- Isolate query workloads (tune, optimize)

Power is performance redefined

Lastly, a word about Cognos as it relates to IBM i

- **All of the previously mentioned DB2 for i query optimization issues apply !**

- **Cognos does not run in IBM i, however, can access DB2 for i**

- **Cognos CAN run in an Linux on Power or AIX Partition**

  – For BEST PRACTICES for running Cognos in an AIX partition, refer to:
    • http://www-304.ibm.com/partnerworld/wps/servlet/ContentHandler/whitepaper/aix/v6r1_cognos/methods

  – Exploiting PowerVM in a Cognos environment
    • http://www.redbooks.ibm.com/Redbooks.nsf/RedpieceAbstracts/sg247842.html?Open

IBM

Questions
&Answers

Power is performance redefined