

B01

Introduction to IMS in the Business Intelligence Community

pete_sadler@uk.ibm.com



Miami Beach, FL

October 22-25, 2001

Trademarks



The world depends on it

■ The following terms are trademarks or registered trademarks of the IBM corporation in the United

States and/or other countries:

- ▶ AIX
- ▶ DataHub
- ▶ DataPropagator
- ▶ DB2
- ▶ DRDA
- ▶ ESA
- ▶ IMS/ESA
- ▶ IBM
- ▶ MVS, MVS/ESA
- ▶ OS/2
- ▶ OS/400
- ▶ Parallel Edition
- ▶ RS/6000
- ▶ SQL/DS
- ▶ Visualizer
- ▶ VM
- ▶ VSE

■ The following terms are trademarks of other companies as follows:

- | | | | |
|-------------------------------|-------------------------|---------------|---|
| ▶ HP-UX | Hewlett Packard | ▶ Solaris | SUN Microsystems, Inc. |
| ▶ Company | | ▶ SQL * Net | Oracle Corporation |
| ▶ INFORMIX | Informix Software, Inc. | ▶ Sybase | Sybase, Inc. |
| ▶ INGRES | Ingres Corporation | ▶ UNIX | AT&T Corporation |
| ▶ Microsoft | Microsoft Corporation | ▶ VisualBasic | Microsoft Corporation |
| ▶ ODBC | Microsoft Corporation | ▶ Windows | Microsoft Corporation |
| ▶ Open Client/
Open Server | Sybase, Inc. | ▶ ETI | Evolutionary Technologies
Incorporated |
| ▶ Oracle | Oracle Corporation | | |

Agenda



The world depends on it

Business Intelligence from DL/I Data
Direct access to IMS data
Data Propagation and Replication
Building and Maintaining DataMarts
Using OLAP and Data Mining

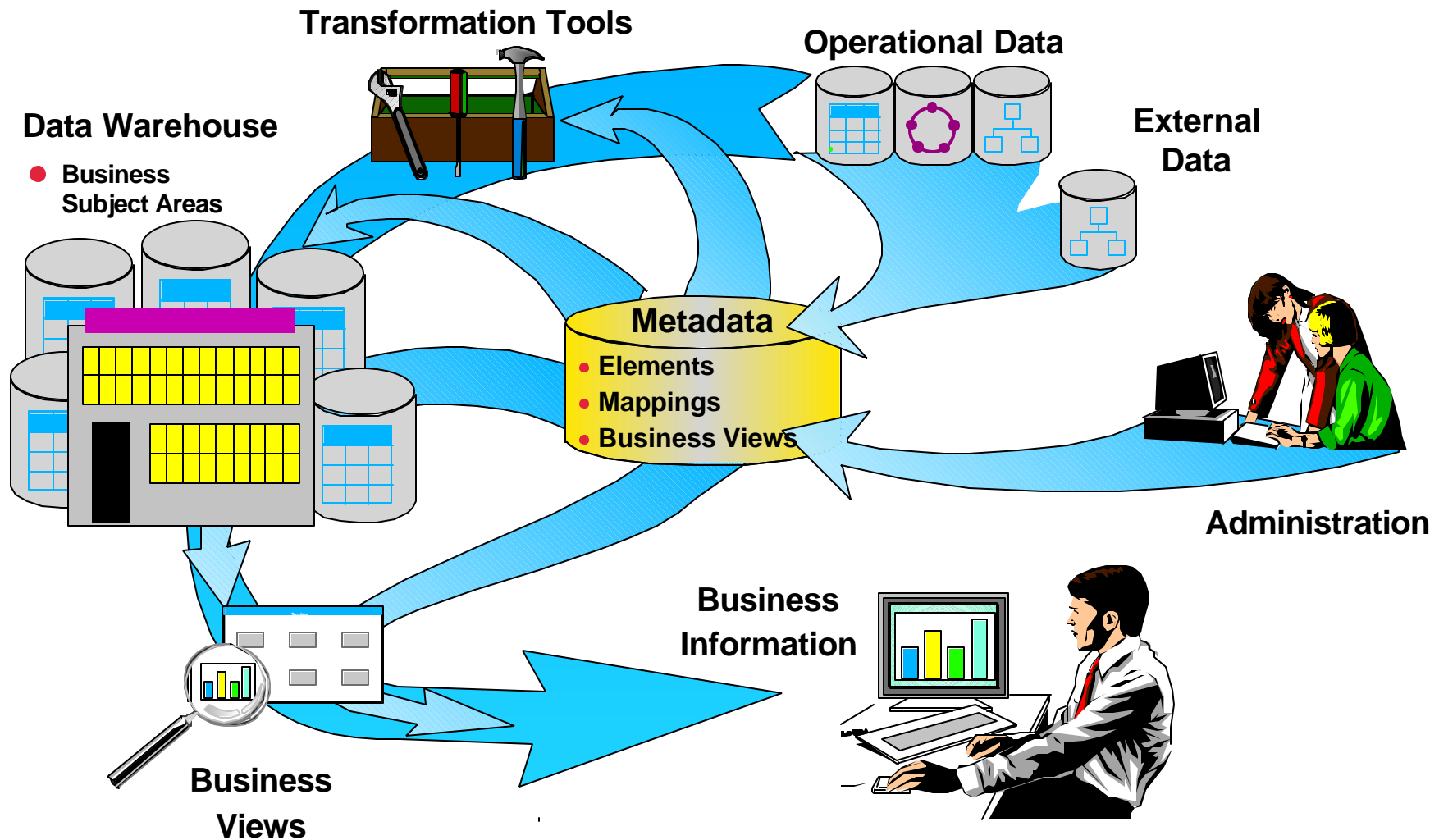
Objective

Show new ways of exploiting IMS data

Data - to - Information Process



The world depends on it

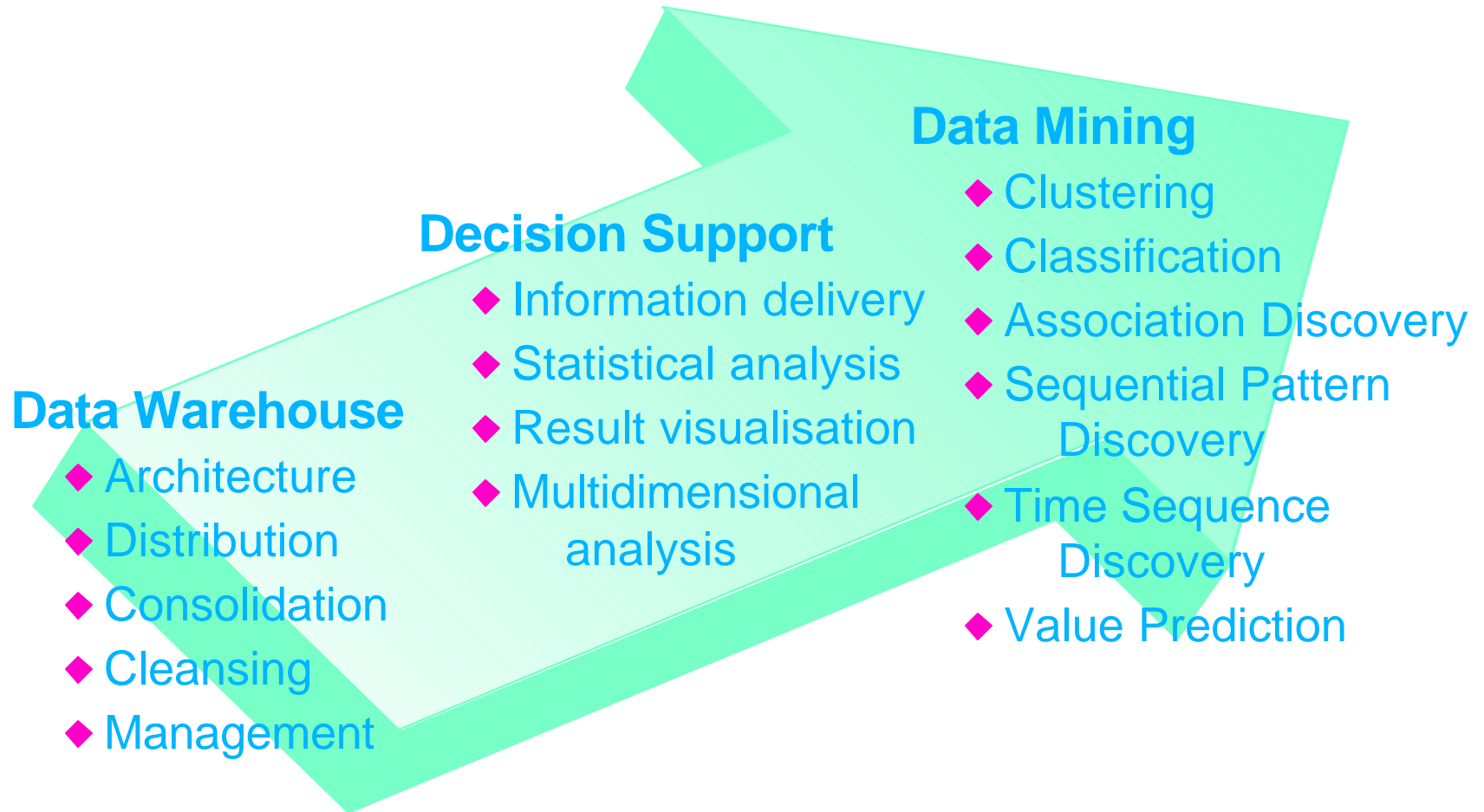


Torture the data until it confesses!

Business Intelligence Components



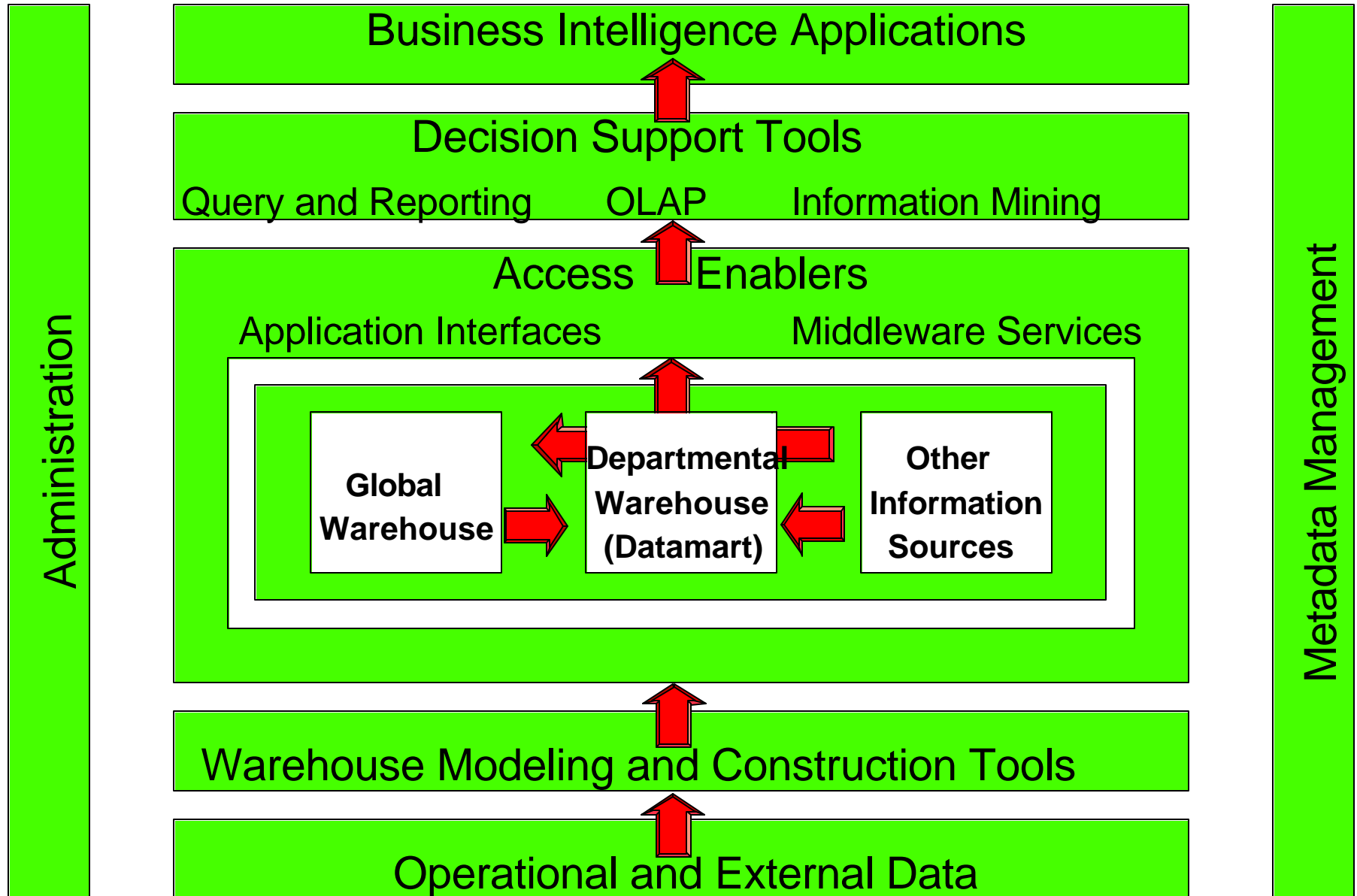
The world depends on it



Business Intelligence Architecture



The world depends on it



OLTP compared with OLAP



The world depends on it

■ OnLine Transaction Processing

- OLTP workloads are characterised by many users creating, updating and retrieving individual DB record
- Focus is high volumes: hundreds or thousands of transactions per section.
- Speed is everything.
- Time element is not stored with data
- Runs the enterprise.
- Examples: Bank Transactions, Check out lines, etc.

■ OnLine Analytical Processing

- OLAP workloads are characterised by few users running large/complex aggregation queries.
- Low volumes.
- Long processing time: hours or days.
- Time element stored with data - needed for historical perspective.
- Answers analytical questions: Why, How Much, When, etc.
- Examples: Market analysis, Trend detection, etc.

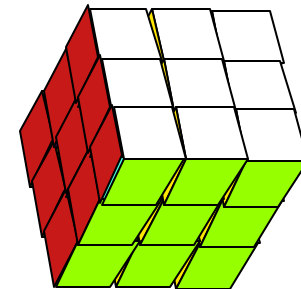
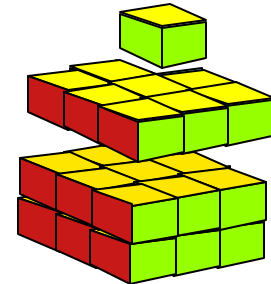
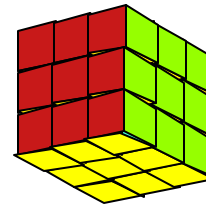
OLAP Characteristics



The world depends on it

- **Defined by Dr. E. F. Codd & widely adopted throughout the industry**
- **Software Optimised for Planning and Analysis**
 - Multidimensional view
 - Drill-down
 - “Slice & Dice”

Examples follow.....
- **Driven by End Users Need**
- **Complements RDBMS and Data Warehouses**
- **Leverages existing investments**
 - Back-end transaction systems
 - Front-end reporting systems

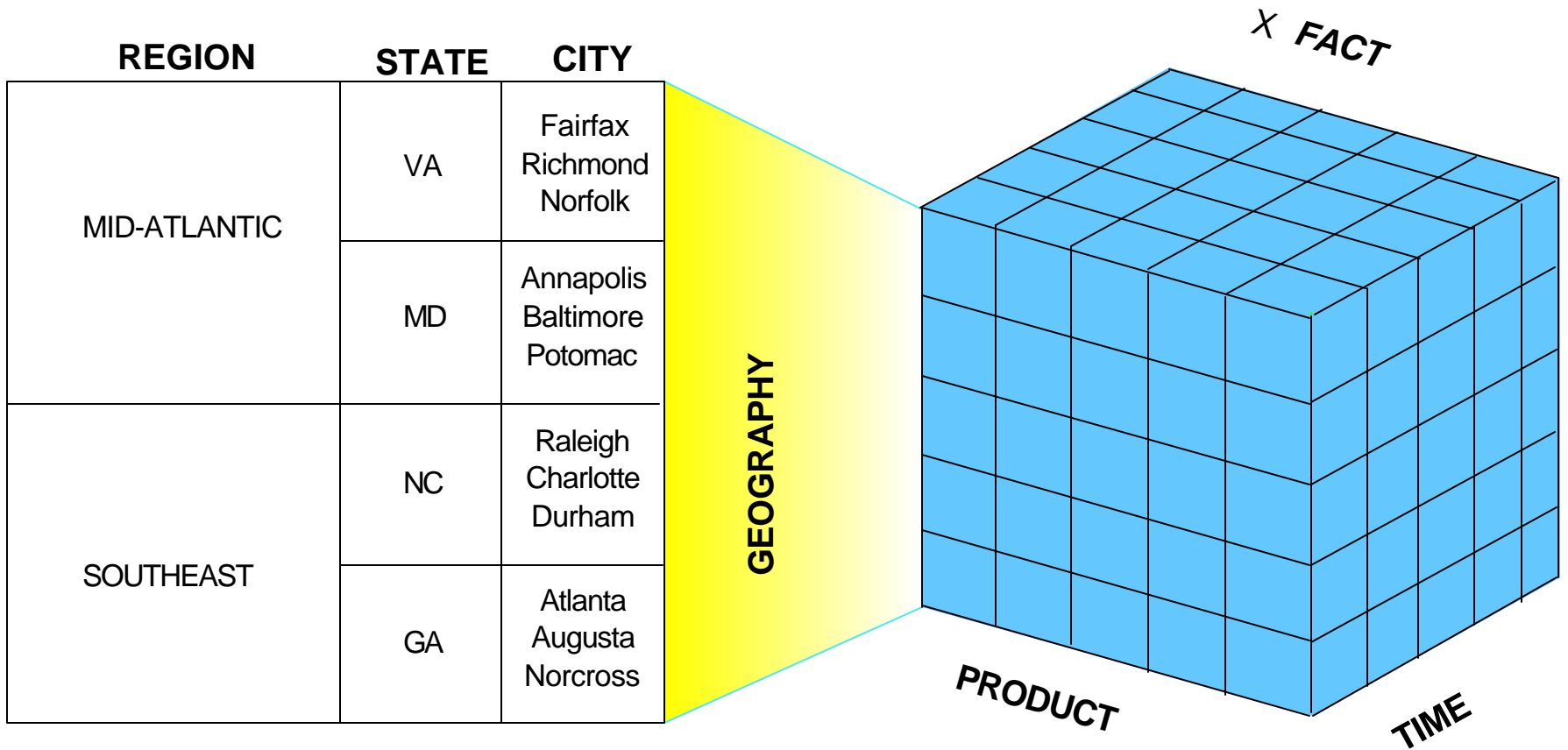


3-Dimensional Example



The world depends on it

GEOGRAPHY



Multi-dimensional Example



The world depends on it

Sales Table

Sales Units	Period Key	Product Key	Geography Key
20	1	1	1
25	1	2	2
25	2	2	3
100	3	3	3
100	4	4	4

Geography Table

Key	Desc	Region	Level
1	NW R	E	48
2	SE R	N	46
3	NW E	S	48
4	NW S	S	56

Product Table

Key	Desc	Type	Category
1	Food	Grocer	A
2	Sports	Ball	A
3	Hardwa	Tool	B
4	Rec	Pool	C

Period Table

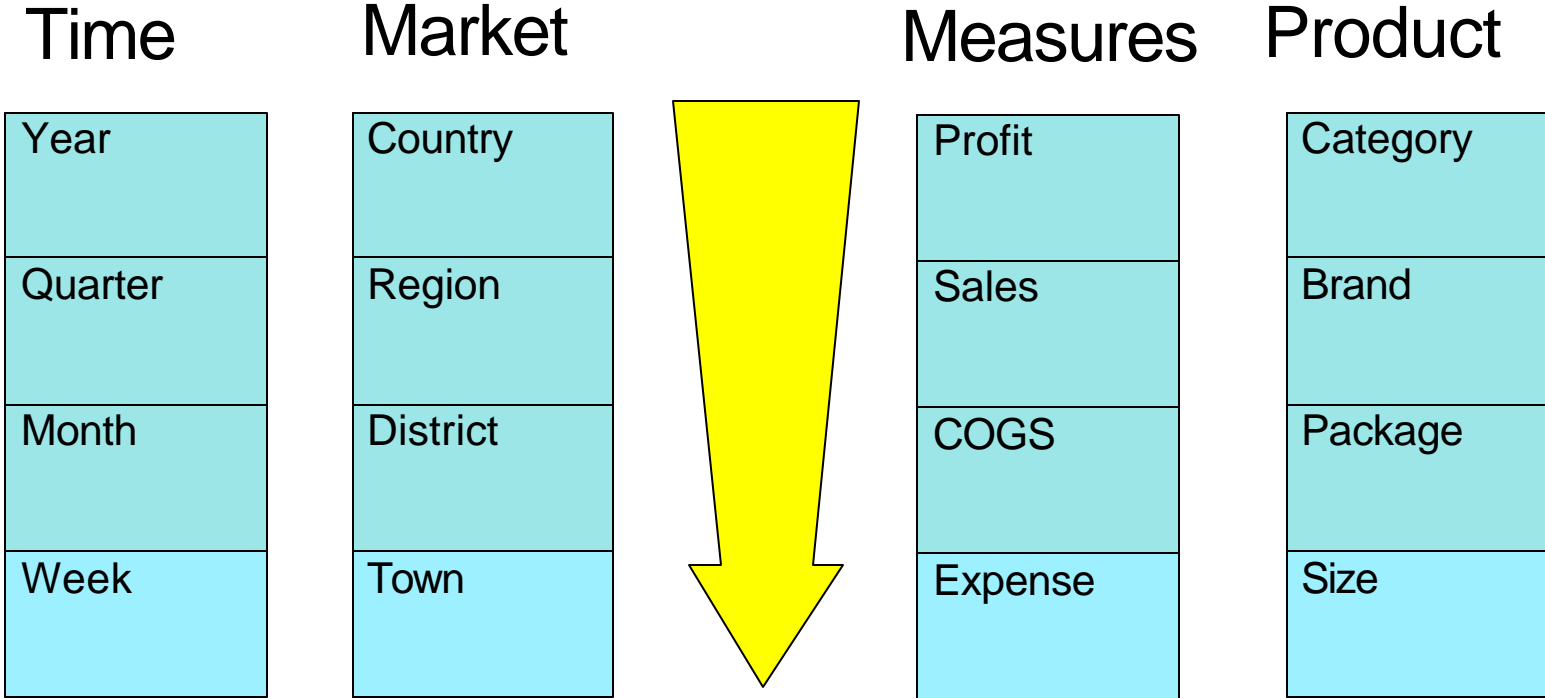
Key	Year	Month	Quarter
1	1996	Jan	1
2	1996	Apr	2
3	1997	Jul	3
	1997	Dec	4

Drill Down.....



The world depends on it

★ Looks at components in greater detail down same dimension



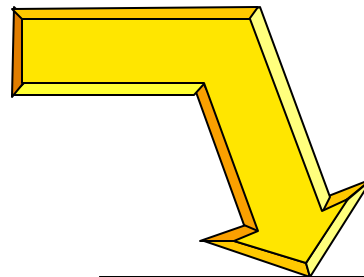
Rotate the Cube.....



The world depends on it

★ Changed dimensions which are breaking down values

		Bud	Act	Bud	Act
2000	East				
	West				
2001	East				
	West				



		1994	1995	1996	1997
East	Food				
	Drink				
West	Food				
	Drink				

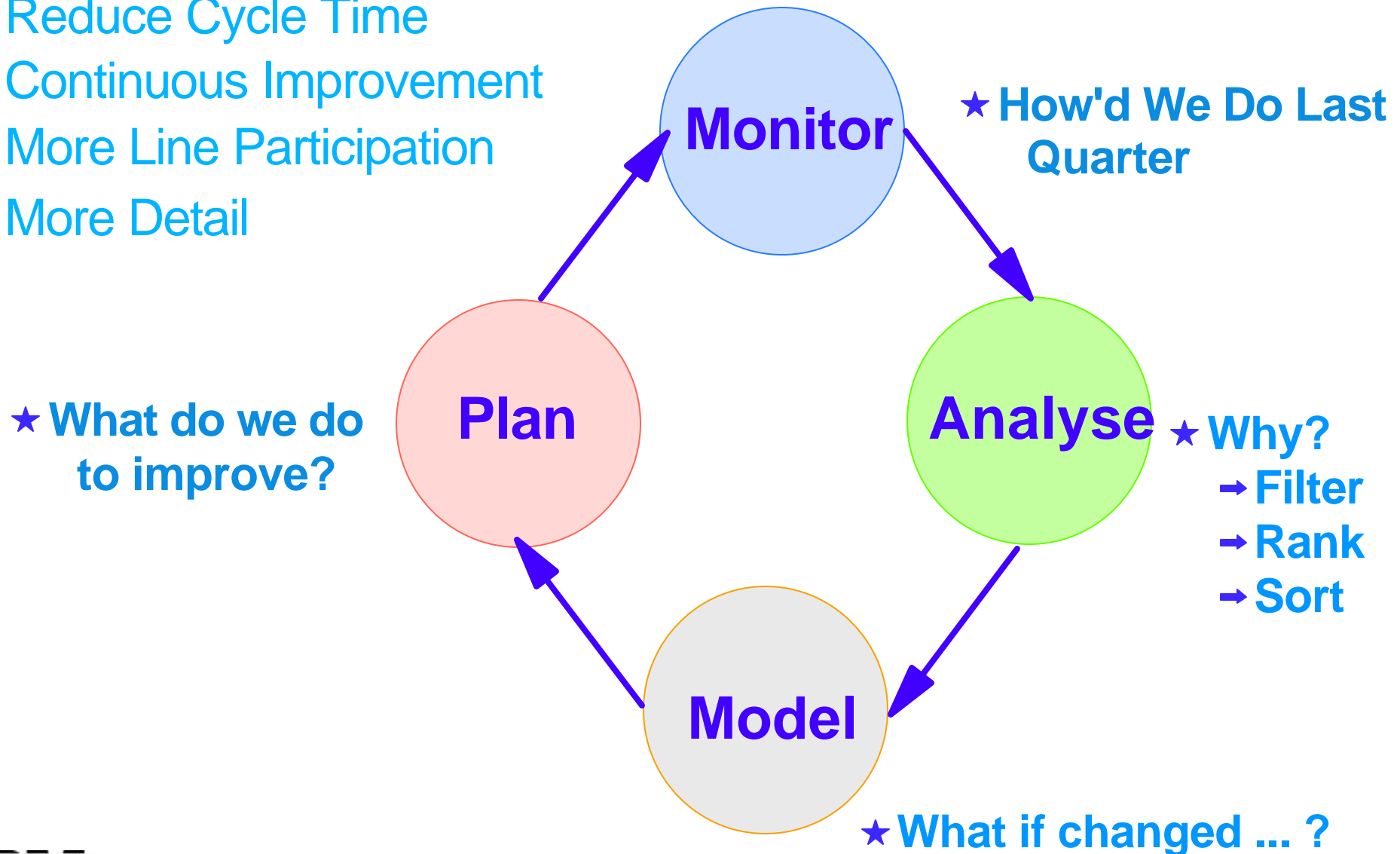
Role of OLAP



The world depends on it

Goal: Manage Business Better

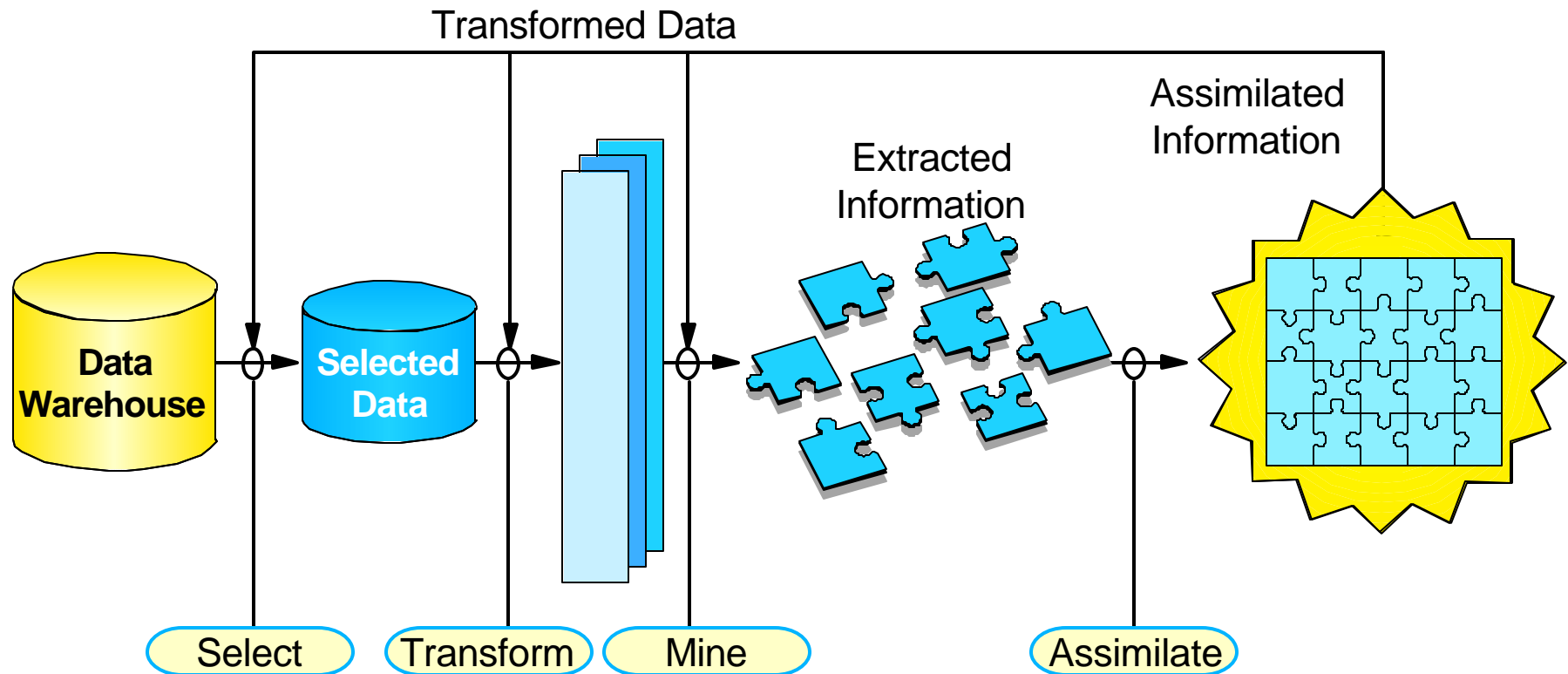
- ✓ Reduce Cycle Time
- ✓ Continuous Improvement
- ✓ More Line Participation
- ✓ More Detail



The Data Mining Process



The world depends on it



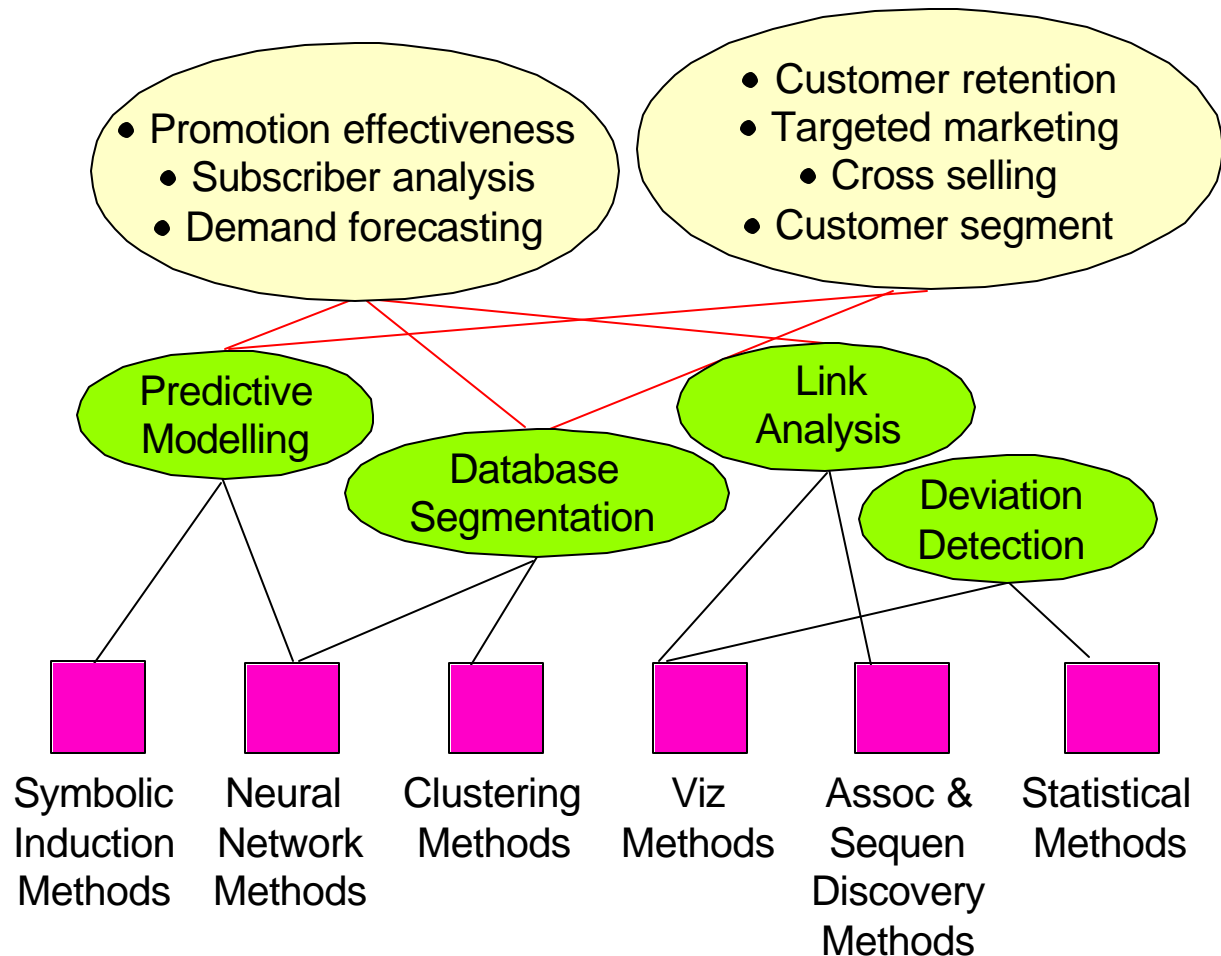
... the process of extracting ***previously unknown***, ***comprehensible***, and ***actionable*** information from large databases and using it to make crucial business decisions.

Mining Techniques



The world depends on it

Intelligent Miner



Applications

Data Mining Operations

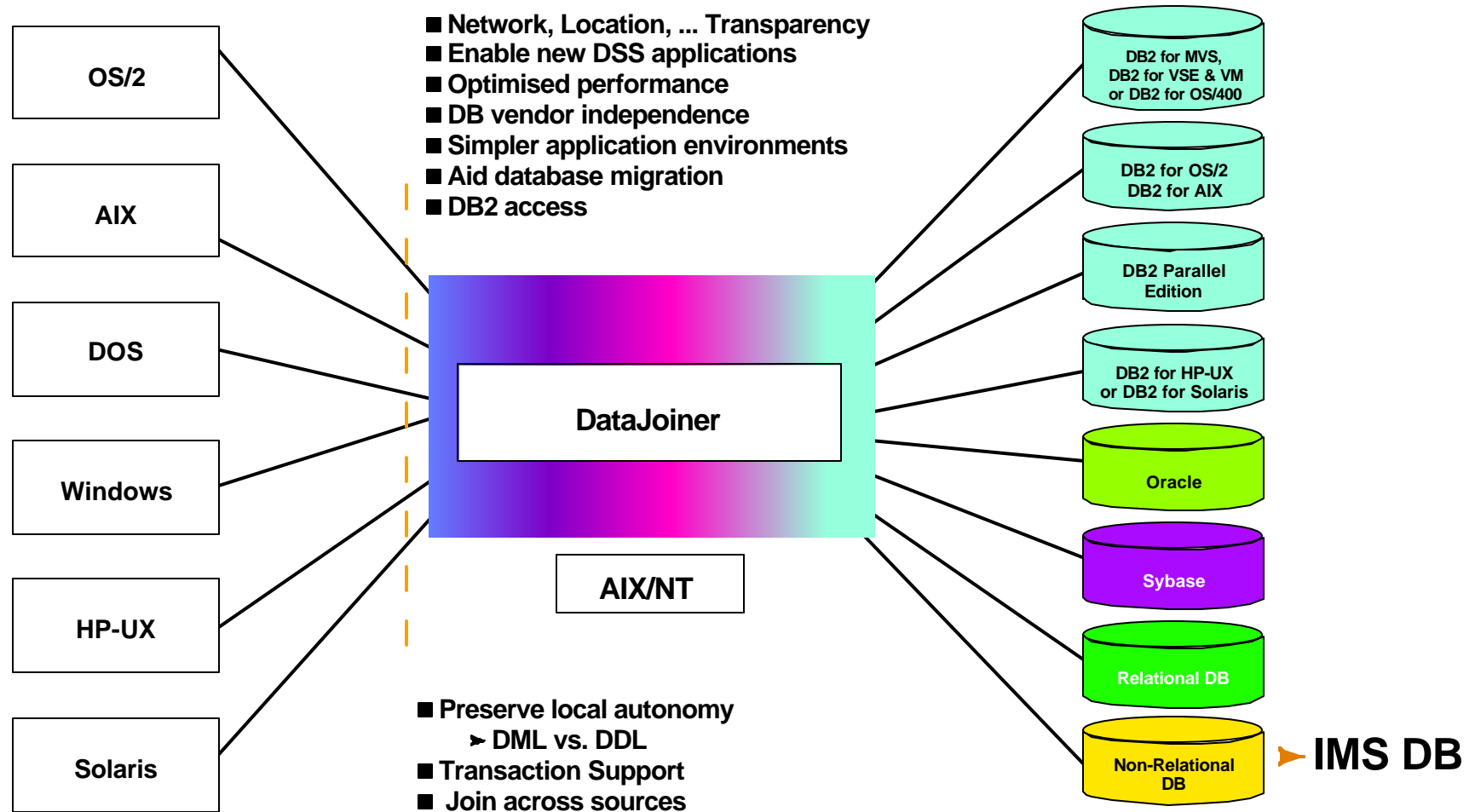
Data Mining Techniques

Direct SQL Access to Data



The world depends on it

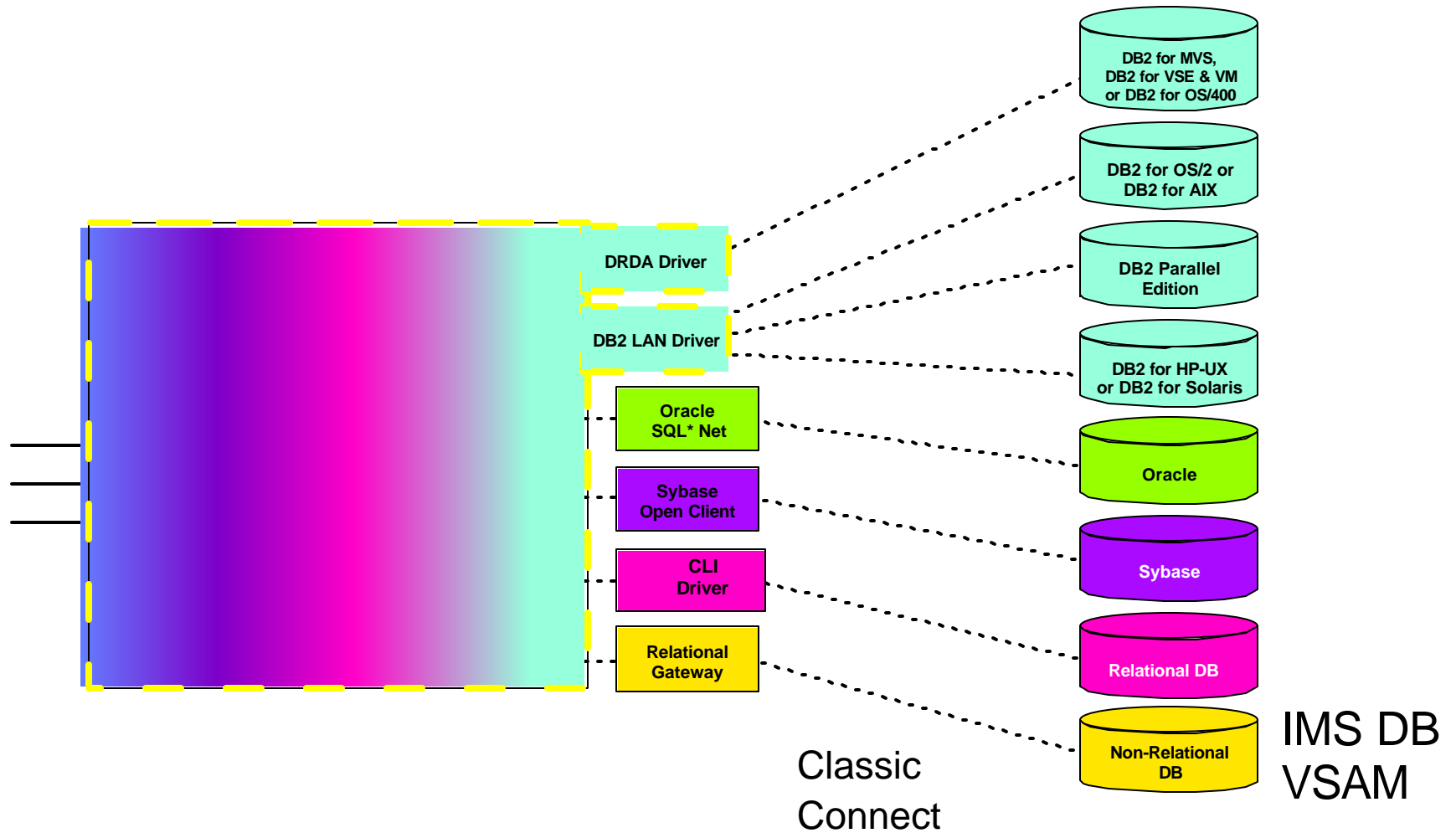
DataJoiner - Single DBMS Image



DataJoiner Server Support



The world depends on it



Direct Data Access Characteristics



The world depends on it

Operational Data



- Current data values
Single value only
- Single copy of data
Simpler management



- Current data values
Erratic repeatability of query
- Encoded data
e.g. M/F, 53 (colour), W29 (location)
- Impact on Operational system
Unpredictable workload
- Network volumes
Multiple segments, not summarised result

Data Characteristics



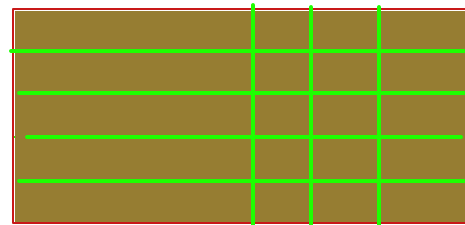
The world depends on it

Operational Data



Application Oriented
Limited Integration
Constantly Updated
Current Values Only
Supports Daily Opns

Informational Data



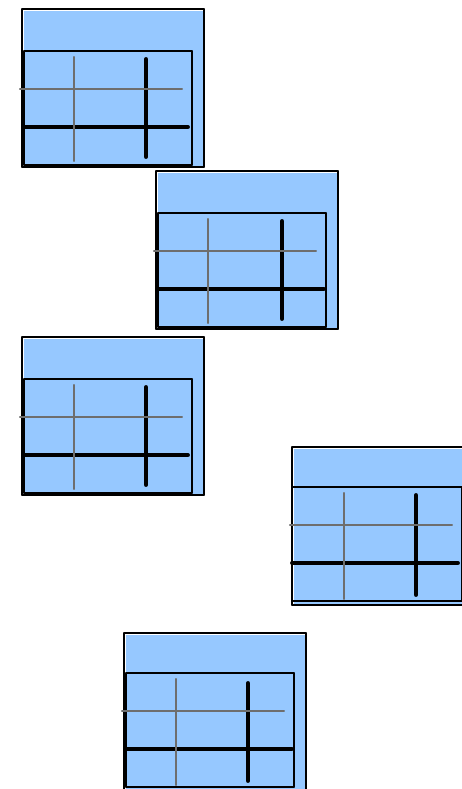
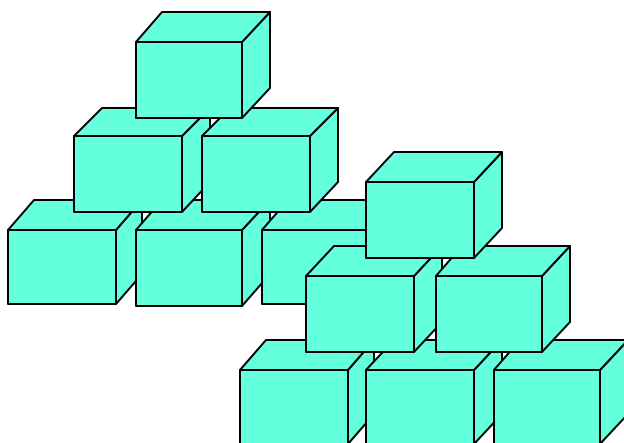
Subject Oriented
Integrated
Non-volatile
Values Over Time
Supports Decision Making

Operational and Informational Data are Fundamentally Different

Building the Warehouse



The world depends on it



Copy to the Warehouse

Decode, cleanse, summarise, enhance data
Time series values
Isolate from Operational System
Cyclic Update - reproducible results

How?

DataRefresher or ETI*Extract - mass extract
IMS DataPropagator - for changes

What, and Where, is the Data?



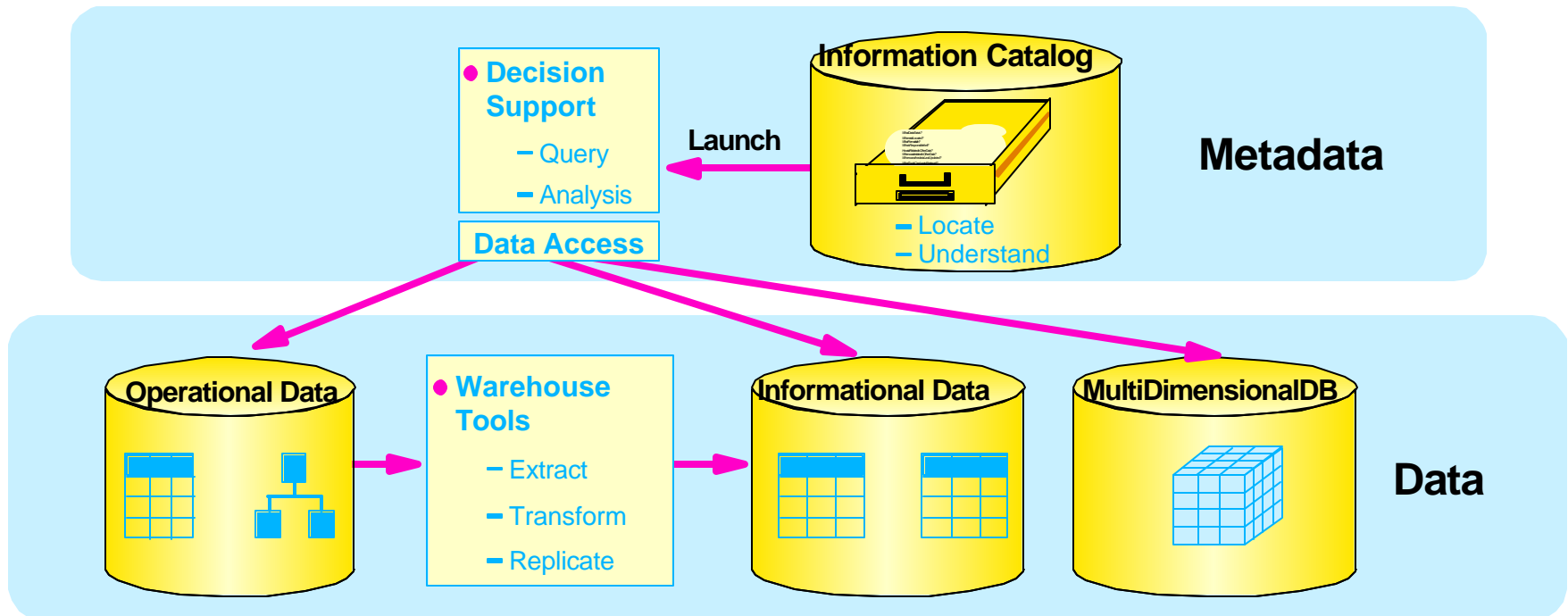
The world depends on it

Information Catalog

Business End User



- Metadata store for business end users
- Contains technical and business metadata



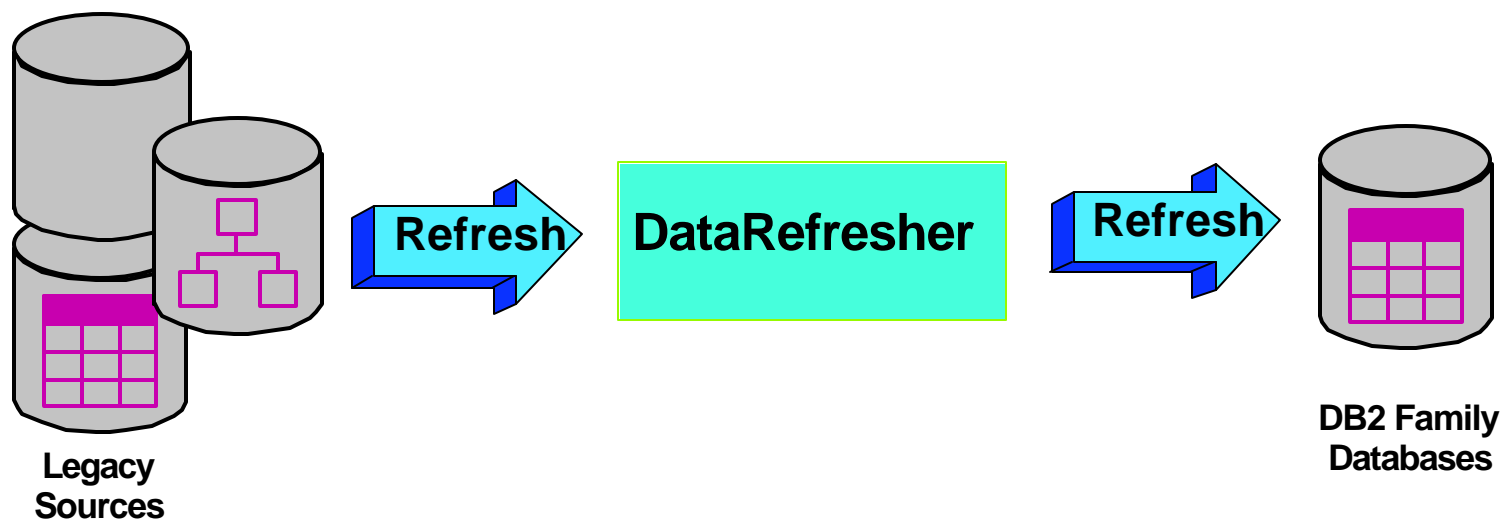
Mass Extract - DataRefresher



The world depends on it

- **Build informational databases on the DB2 family**

- From any data source on MVS and from DB2/VM
- Heterogeneous join across sources
- Eliminate extract application development and maintenance
- Data enhancement
- Client/Server implementations



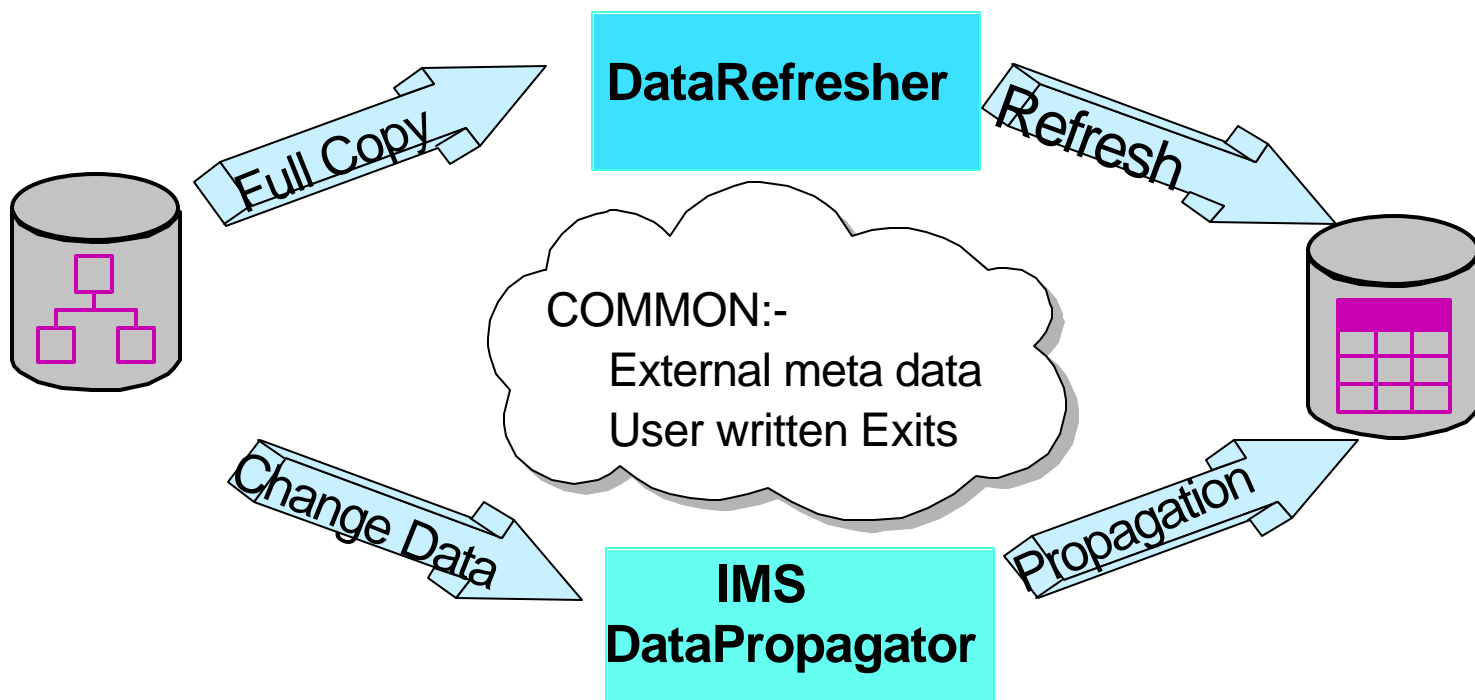
Capturing Changes - DR and IMS DProp



The world depends on it

- **Build informational databases on the DB2 family**

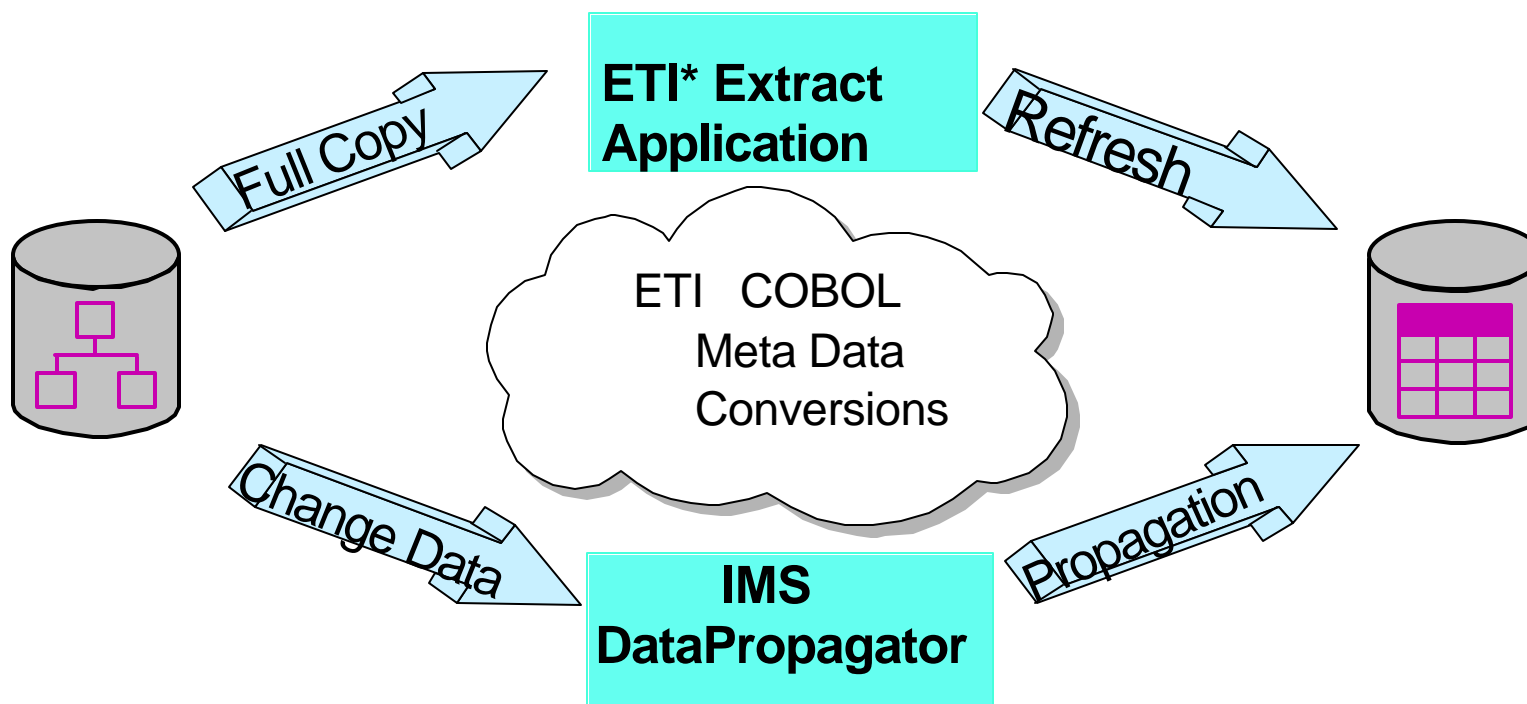
- From any data source on MVS
- Heterogeneous join across sources
- Eliminate extract application development and maintenance
- Data enhancement
- Client/Server implementations



Capturing Changes - ETI EXTRACT and IMS DProp

- **Build informational databases on the DB2 family**

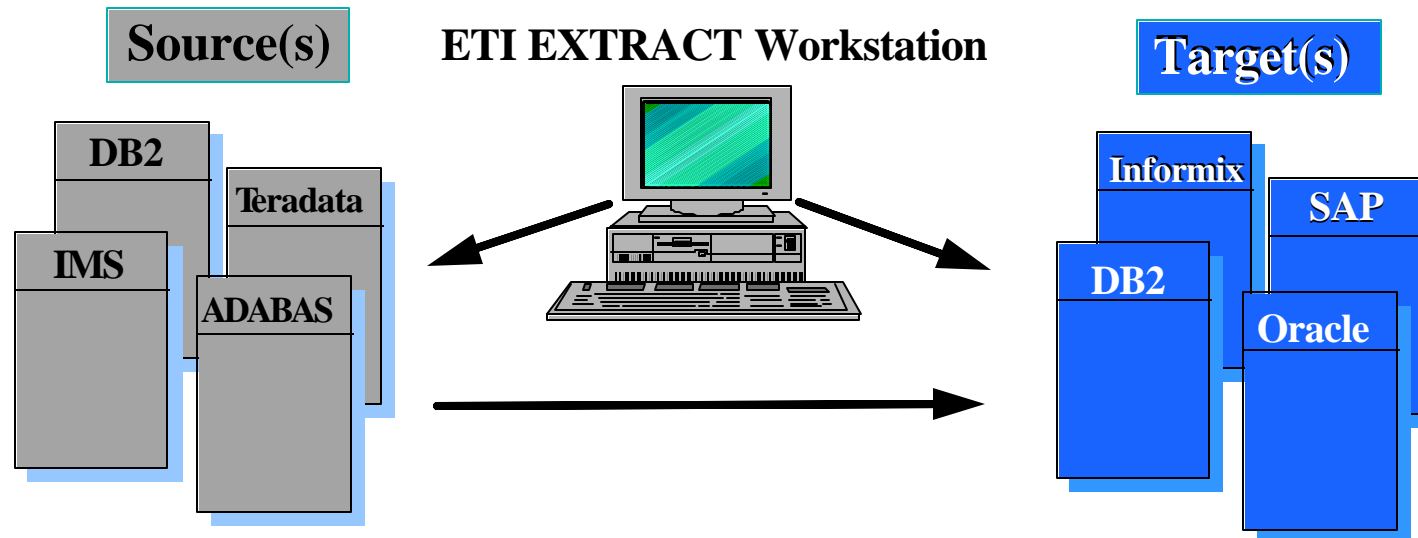
- Prepare Meta Data
- Generate extract application
- Enable Data Enhancement
- Also Invoked via DPNR Propagation Exit



ETI EXTRACT Unique Process



The world depends on it



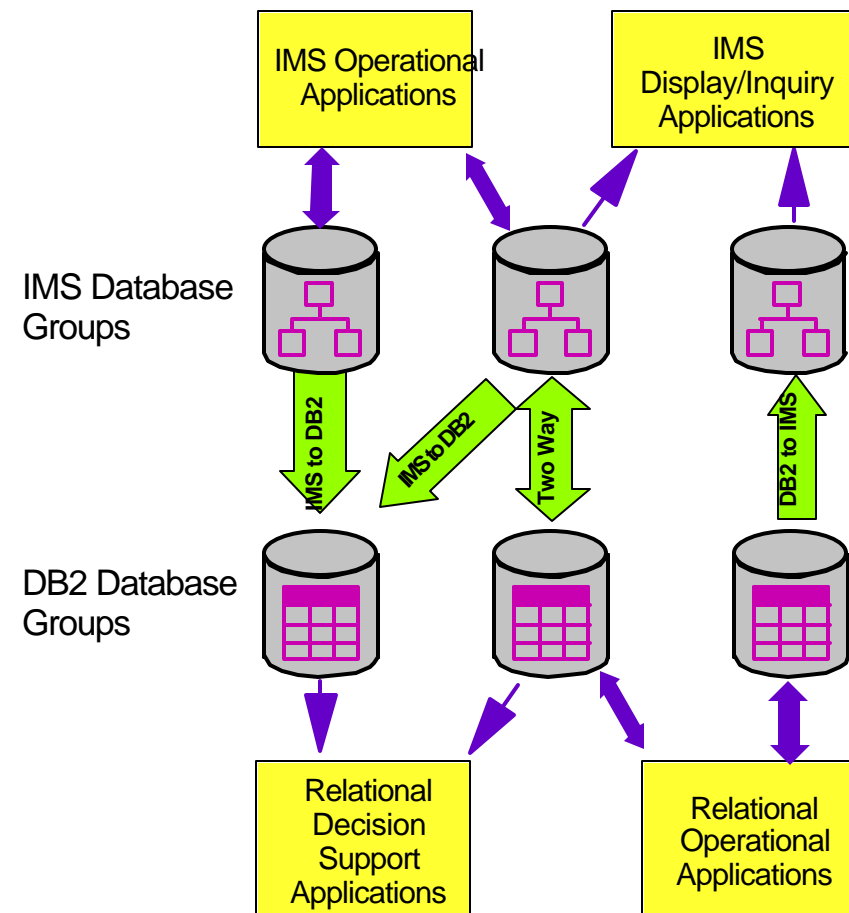
- ❖ Programs generate on ETI EXTRACT workstation
- ❖ Executive transfers programs to source and target platforms
- ❖ Programs run on source and target systems
- ❖ Data moves directly between source and target systems

IMS DataPropagator



The world depends on it

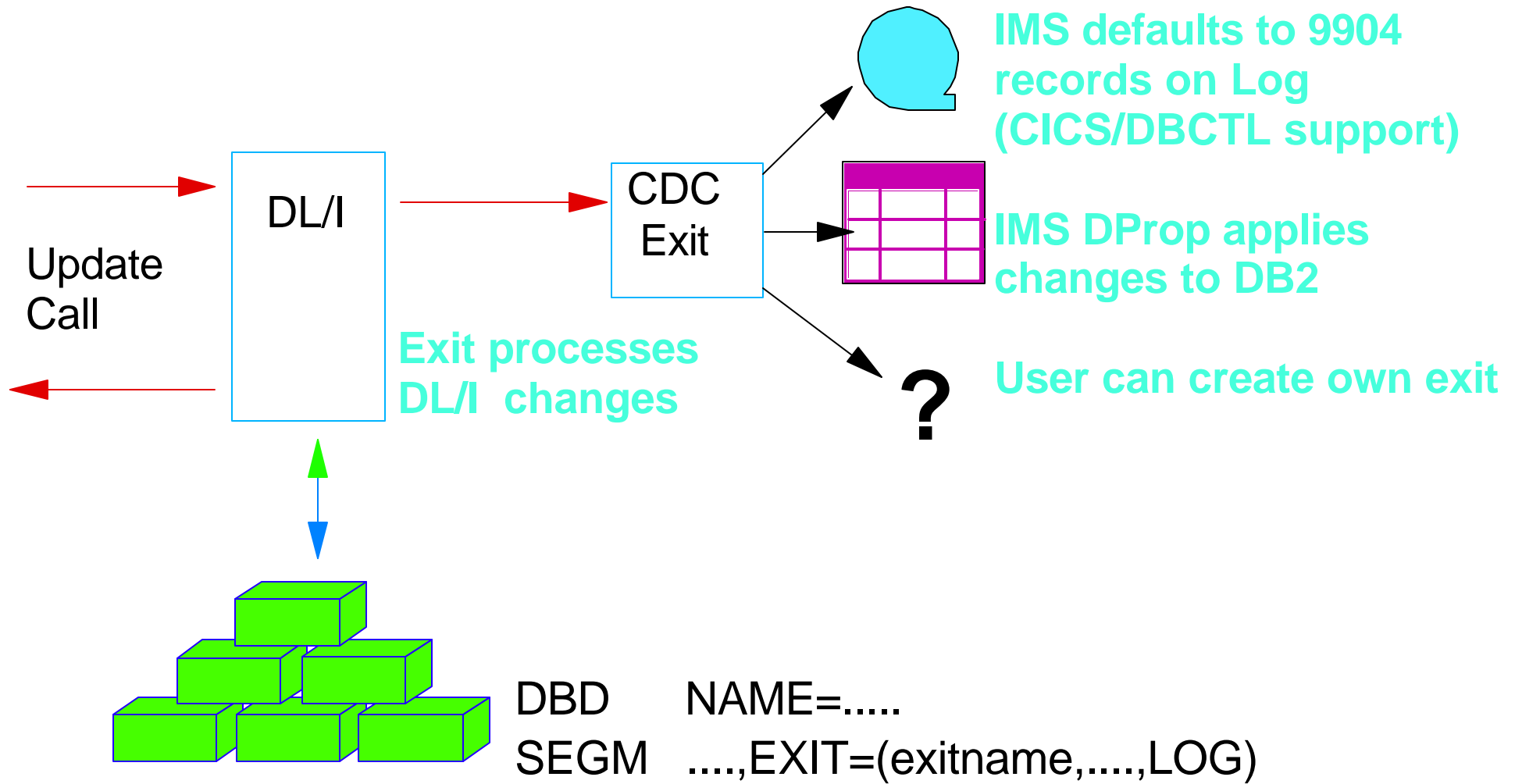
- **Decision Support**
 - User access to stable data
 - Propagate only the data of interest
 - Exploit relational technology for query
- **Application Migration**
 - Gradual, orderly migration
 - Minimal risk
- **Application co-existence**
 - Two masters
 - Data in both systems synchronised
- **No Change to Existing Applications**



IMS Change Data Capture



The world depends on it

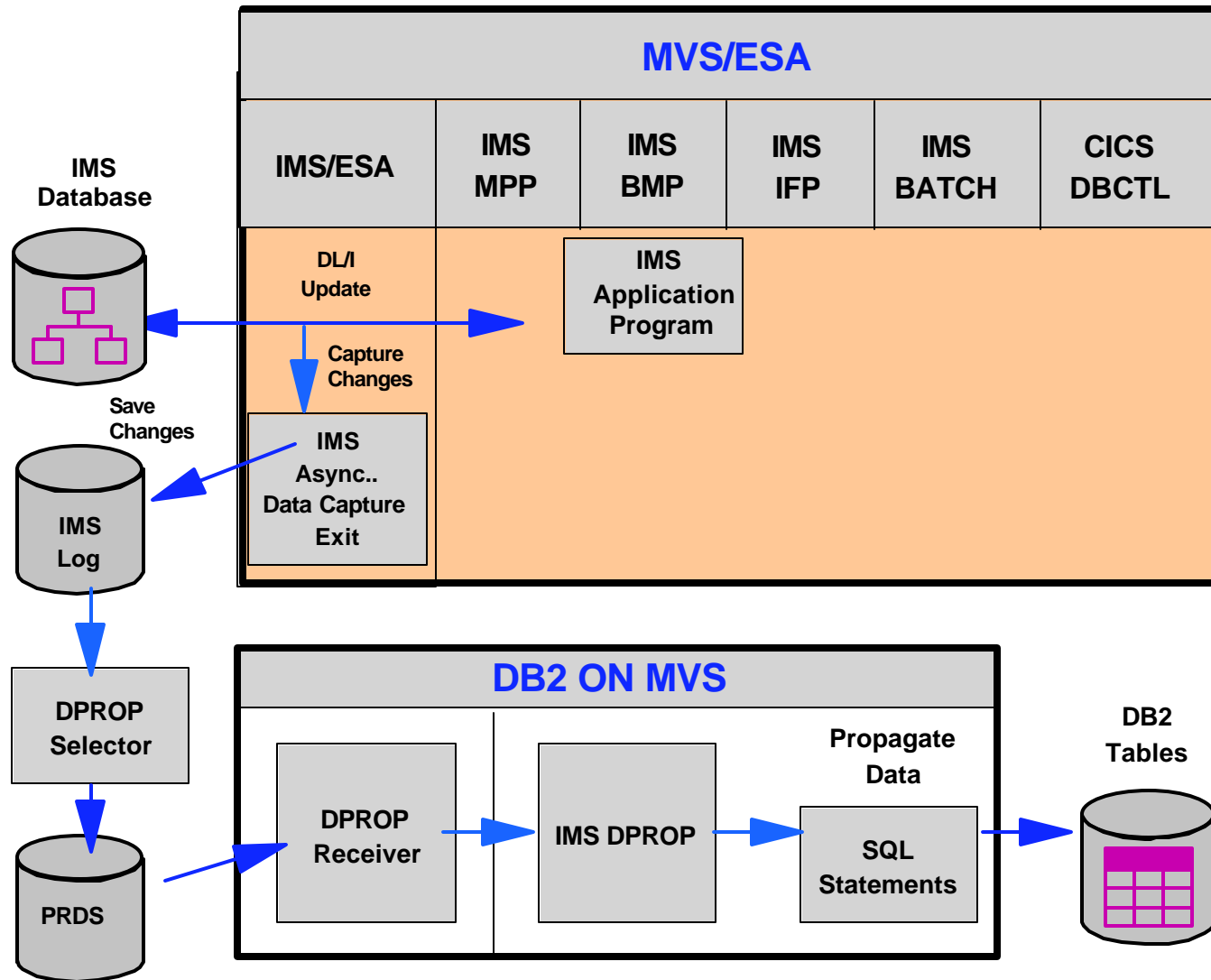


Asynchronous Propagation IMS to DB2



The world depends on it

- Updates applied at customer determined intervals



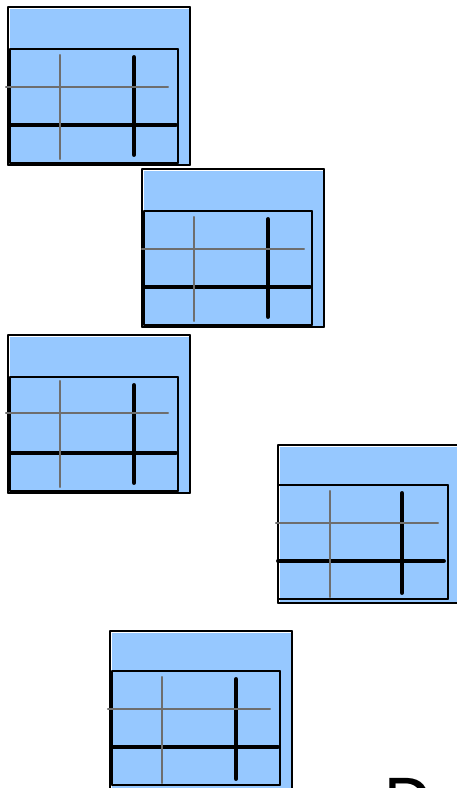
Supplying Users



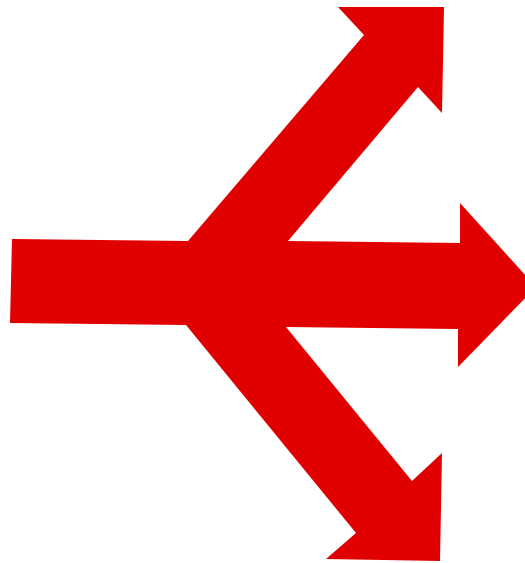
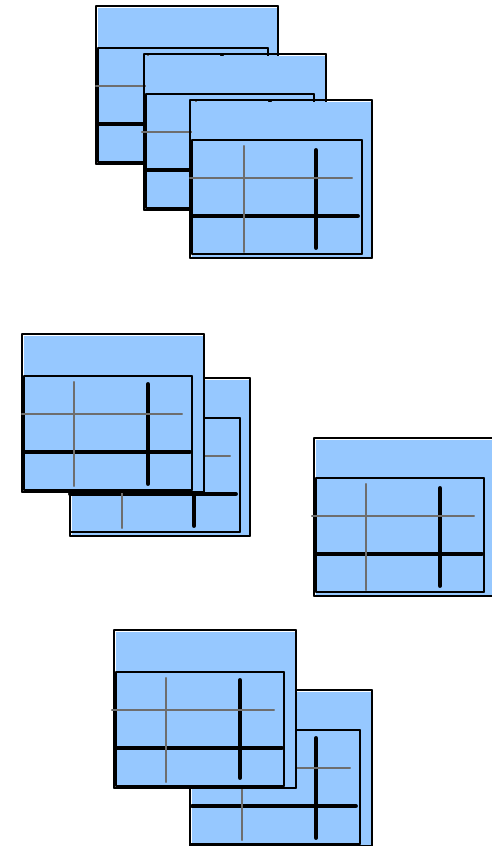
The world depends on it

Single Coherent Source, Multiple Distinct Users

Warehouse



Datamarts

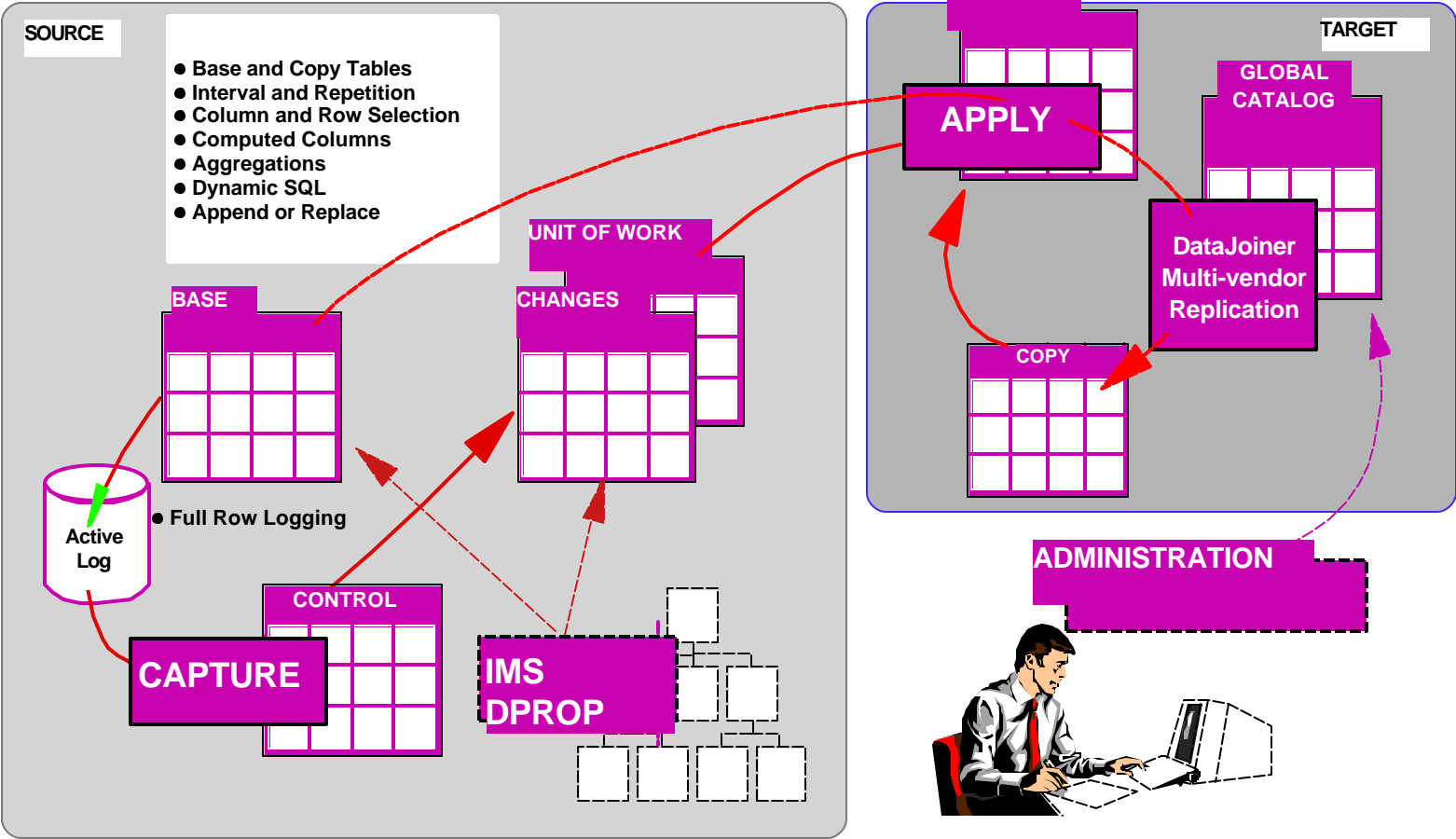


DataPropagator Relational
DataHub for control

Capture, Staging and Apply



The world depends on it



- Examples
- Apply/MVS
 - Apply/OS2
 - Apply/AIX
 - Apply/HP-UX

Data Propagator Relational

DataPropagator Relational : Features



The world depends on it

● Automated Data Copy

- Automated operation
- User specified or event driven synchronisation

● Refresh and Update

- System resource management
- Historical data

● Enhancement

- Full SQL support
 - ▶ Subset
 - ▶ Aggregate
 - ▶ Derive
- Leverage existing SQL skills

● Data Staging

- Copy consistency
- Reduced table contention
- System autonomy

● Administration

- DataHub control point
- GUI
- Initialisation tasks automated

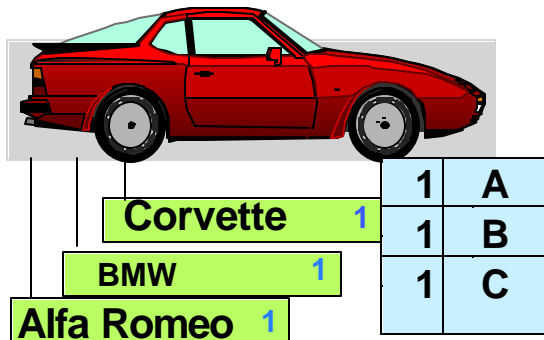
● Open Architecture

- DRDA
- SQL
- Architected interface

Application Staging Tables



The world depends on it



3rd Nov Sells the Corvette
 23rd Nov Buys a second BMW
 1st Dec Buys a Ferrari
 25th Dec Buys another Ferrari
 26th Dec Buys a Jaguar

- Delete row
- Update # of cars
- Insert new rows
- Update # of cars
- Insert new row

BASE	CONDENSED	NON-CONDENSED
CAR #	DATE Action CAR #	DATE Action CAR #
A 1	1/1 I A 1	1/1 I A 1
B 2	23/11 U B 2	1/1 I B 1
F 3	3/11 D C 1	1/1 I C 1
J 4	25/12 U F 2	3/11 D C 1
	26/12 I J 1	23/11 U B 2
		1/12 I F 1
		25/12 U F 2
		26/12 I J 1

Current view

Good for further propagation

- Remove Hotspots

Useful for current view
(Must disregard deleted rows)

Good for further propagation

Audit trail

Historical analysis

Time series

Legacy Data - the Problems....



The world depends on it

Five Legacy Contaminants Encountered in Migrations

1. **Lack of legacy standards**
2. **Data surprises in individual fields**
3. **Legacy information buried in free-form fields**
4. **Legacy myopia**
 - Multiple account numbers block consolidated view
5. **The anomalies nightmare**
 - Complex matching and consolidation

Examples follow.....

1. Lack of Legacy Standards






The world depends on it

■ Unlimited:

- formats
- structures
- attributes
- code sets

■ all within fields with the same meta labels!

	Name Field	Location
File 1	MARK DI LORENZO 	MA93
	DENIS E. MARIO	CT15
	TOM & MARY ROBERTS	IL21
File 2	DILORENZO, MARK 	6793
	MARIO, DENISE	0215
	ROBERTS, TOM & MARY	8731
File 3	MARC DILORENZO ESQ 	BOSTON
	MRS DENNIS MARIO	HARTFORD
	MR & MRS THOMAS ROBERTS	CHICAGO

2. Data Surprises in Individual Fields



The world depends on it

Metadata--"What you wish you had in your data values"

--Ken Orr

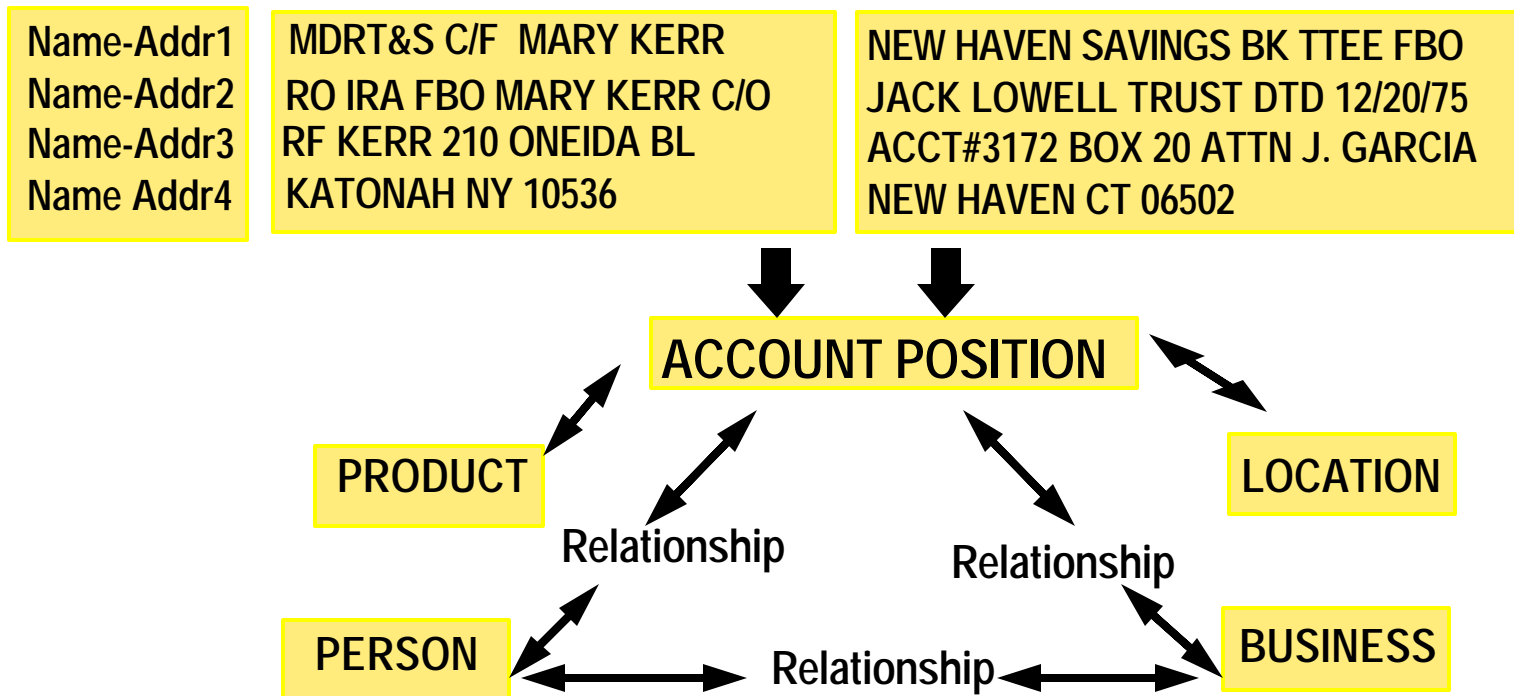
NAME	SOC. SEC. #	TELEPHONE
Denise Mario DBA	228-02-1975	6173380300
Marc Di Lorenzo ETAL	999999999	3380321
Tom & Mary Roberts	025-37-1888	
First Natl Provident	34-2671434	415-392-2000
Digital 15 State St.	101010101	508-466-1200
Astorial Fedrl Savings	LN#12-756	212-235-1000
Kevin Cooke, Receiver	18-7534216	FAX 528-9825
John Doe Trustee for K	111111111	5436

3. Legacy Information Buried - in Free-Form Fields



The world depends on it

How will you determine and extract entity relationships?



4. Legacy Myopia - No Consolidated View



The world depends on it

How will your business identify the customer when account numbers block your view?

Legacy data refers to 3 accounts under different numbers.

IBM A2R-132-867
IBM C8P-435-916
IBM Q9O-211-803



5. The Anomalies Nightmare



The world depends on it

How are you going to correctly identify and consolidate anomalies from millions of records?

CUSNUM	NAME	ADDRESS	SALES \$
90328574	IBM Corporation	187 N.Pk. St. Salem NH01456	8,494.00
90328575	IBM	187 N.Pk. St. Sarem NH 01456	3,432.00
90238495	Int. Bus Mach	187 N. Park StSalem NH 04156	2,243.00
90233479	IBM Corp	187 Park Ave Salem NH04156	5,900.00
90233489	IBM Consulting	15 Main St. Andover MA 02341	6,800.00
90234889	IBM ISSC	PO Box 9 Boston MA 02210	10,243.00
90345672	IBM Integration	Park Blvd. Boston MA 04106	15,999.00

No unique key

Anomalies

No standardising

Spelling

5. No Consolidated View



The world depends on it

How are you going to consolidate records across multiple files?

Customer File

CUSNUM	NAME	ADDRESS
9035769	IBM Corporation	187 N.Park St. Salem NH 01456

Order File

90342	ORDER #	NAME	STREET	CITY	ST	ZIP	SALES \$
90211	XA-6701	IBM Corporation	187 N.Park St.	Salem	NH	01456	8,494.00
90237	TA-8574	IBM	187 N.Pk. St.	Sarem	NH	01456	3,432.00
9021	PO-8495	Inter. Bus. Mach.	187 N. Park St	Salem	NH	04156	2,243.00
9028	SD-2363	IBM Corp	187 Park Ave	Salem	NH	04156	5,900.00
9034	YU-4889	Consulting	15 Main St.	Andover	MA	02341	6,800.00
	BN-3489	IBM ISSC	PO Box 9	Boston	MA	02210	10,243.00
	CV-5672	Integration	Park Blvd.	Boston	MA	04106	15,999.00

Data Re-engineering with Integrity



The world depends on it

VALITY -

A proven technology and methodology for data cleansing

- **Offers architecture and toolkit specifically designed to investigate, re-engineer and consolidate data**
- **Provides data investigative and matching functions that extend well beyond the capabilities of data scrubbing**
- **Enables fast and effective migration of high-quality information to new Data Warehouses, CISs & Client/Server applications**
- **Recovers buried but essential metadata**
- **Transforms large volumes of legacy data into consolidated business views**

Manage the Warehouse and Datamarts



The world depends on it

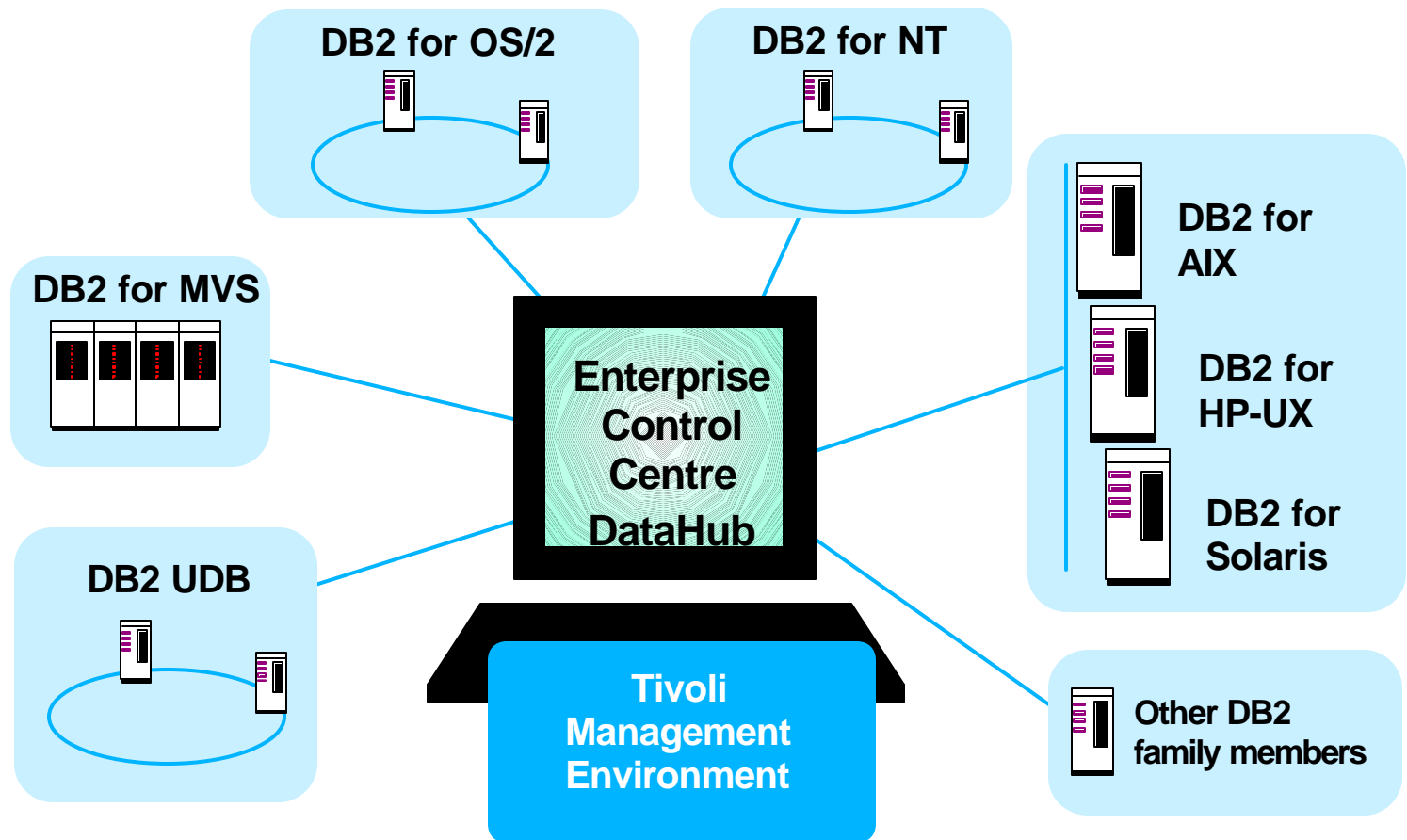
SystemView Information Warehouse DataHub - Tivoli Environment

Central point of control

Suite of tools for database administration

Automated operations

Launch platform for tools

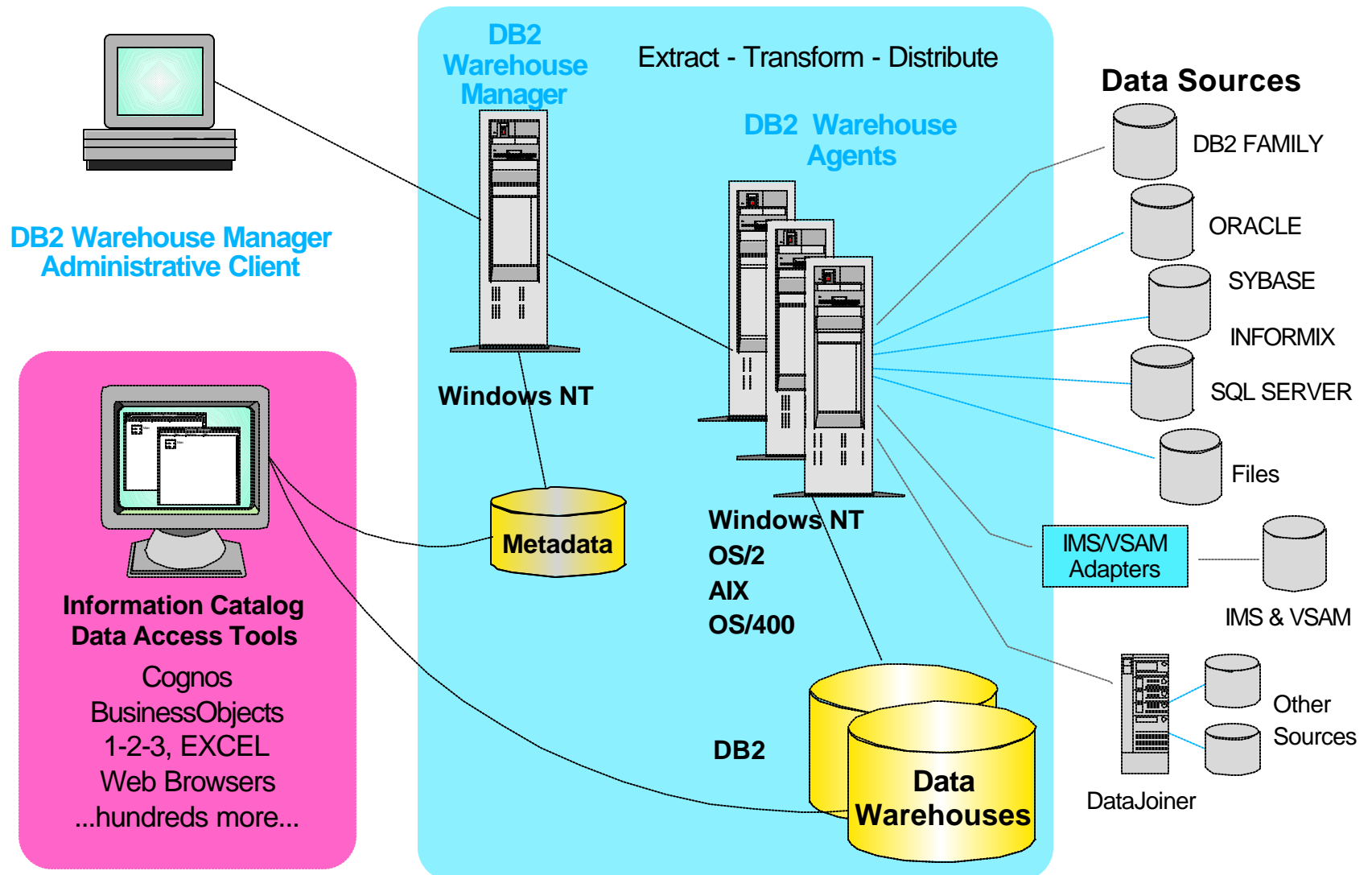


DB2 Warehouse Manager Overview



The world depends on it

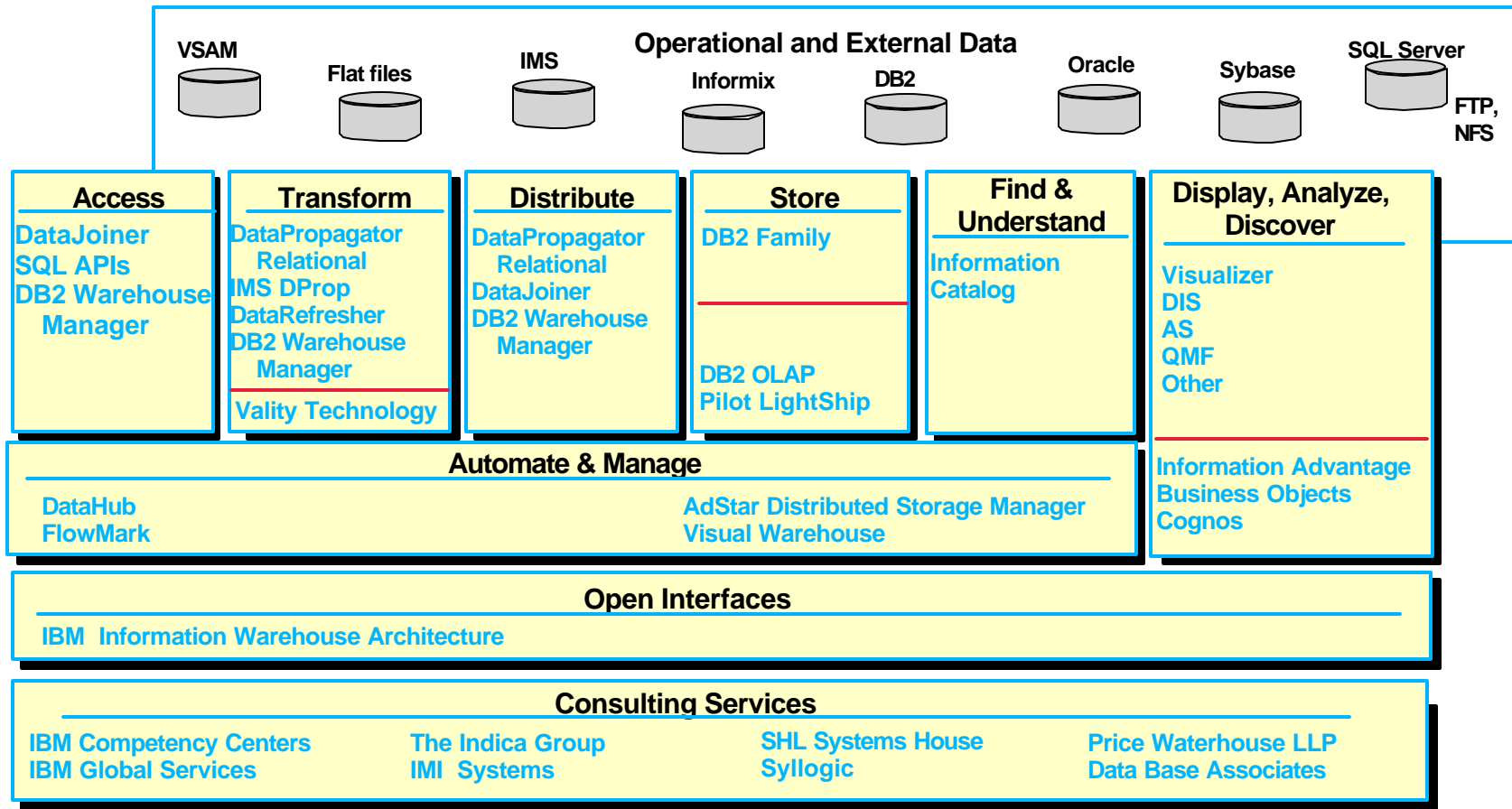
DB2 Warehouse Manager aka Visual Warehouse



Coverage of DB2 Warehouse Manage



The world depends on it



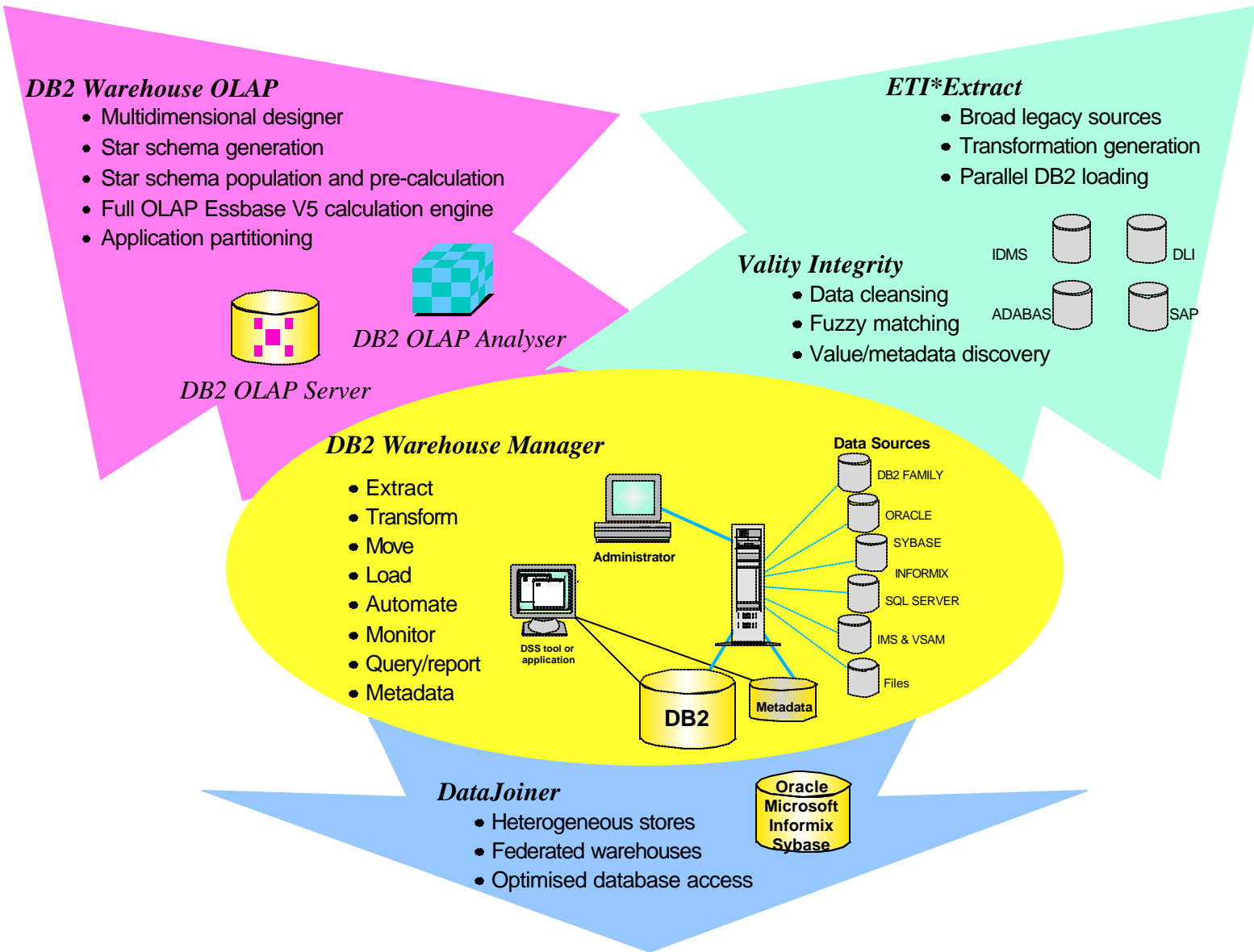
- Comprehensive Products and Services
- Flexible Implementation Alternatives
- An Open Solution Environment



DB2 Warehouse Manager Family



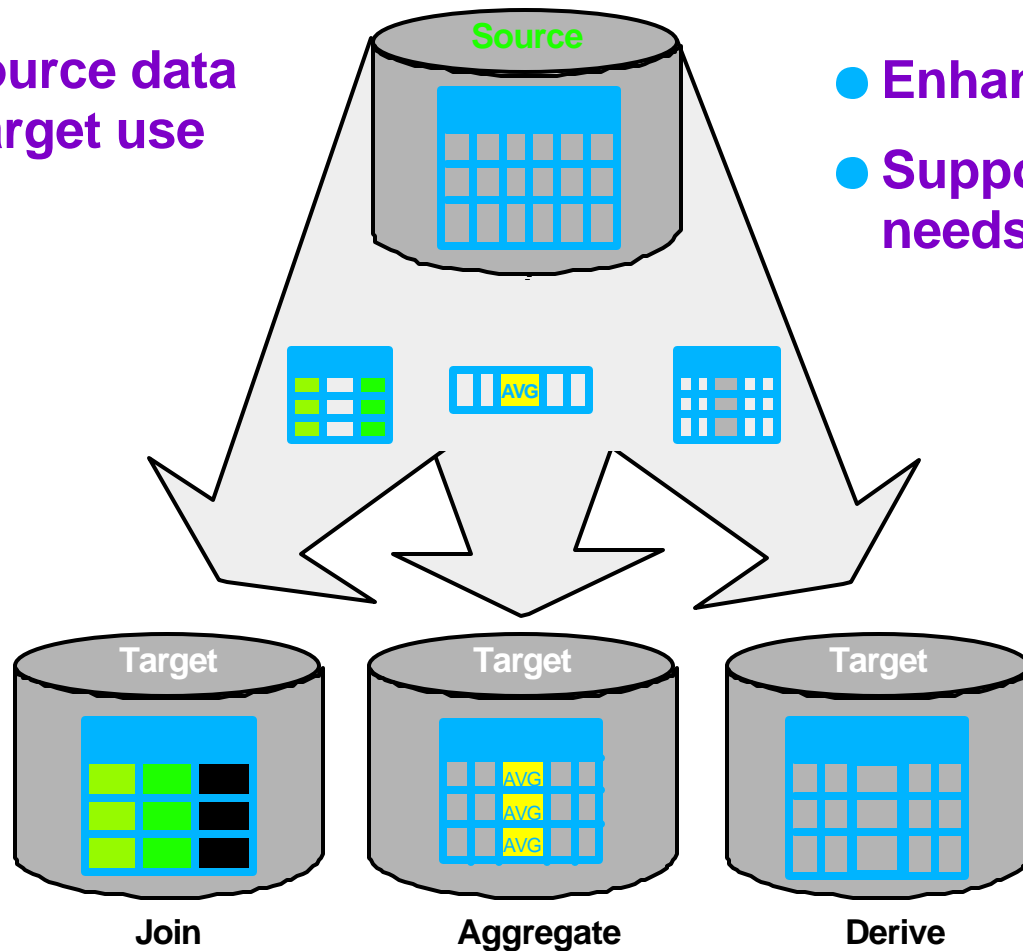
The world depends on it



Data Enhancement with DB2 Warehouse Manager

Customise source data for specific target use

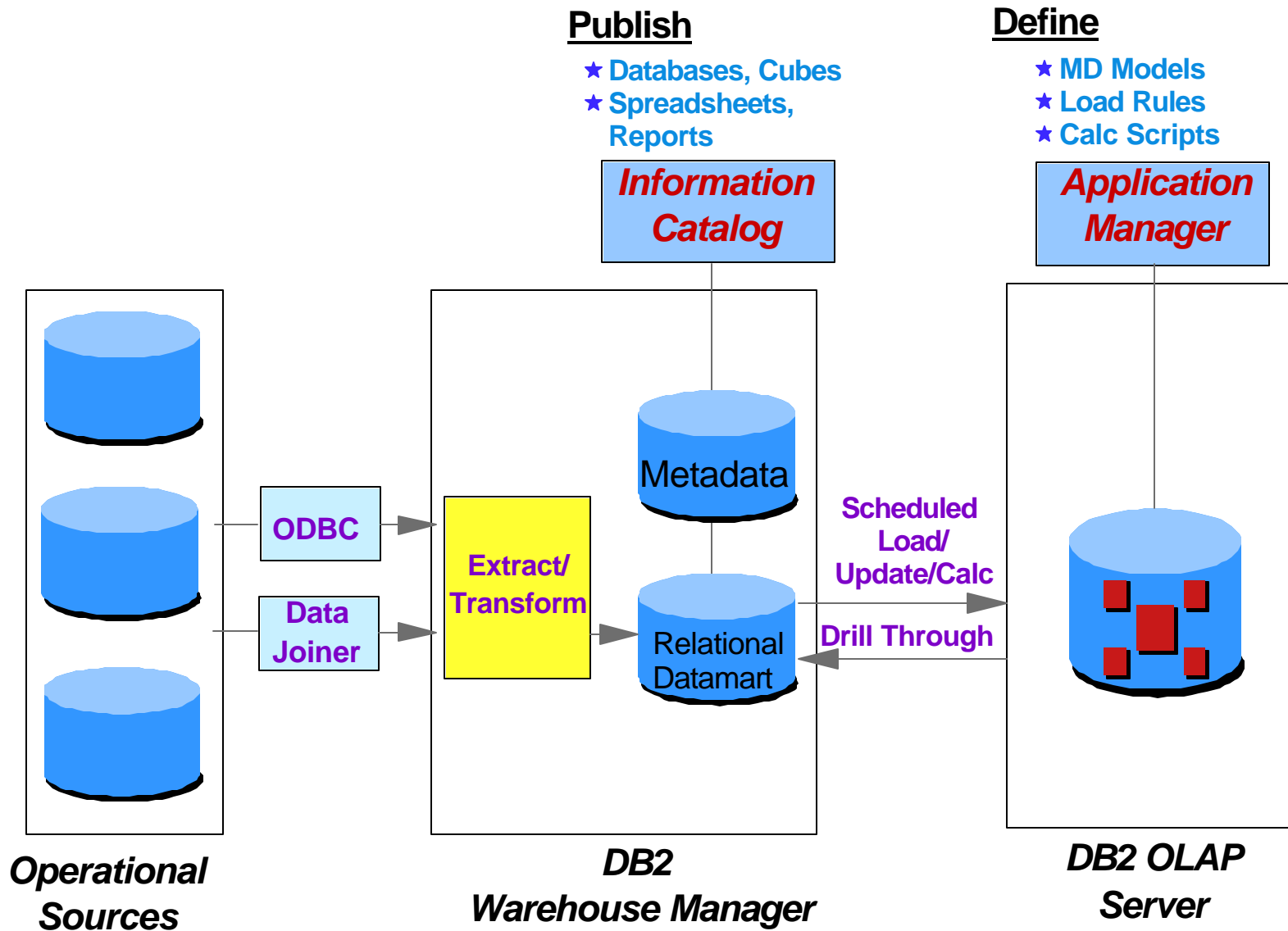
- Enhances data usability
- Supports unique application needs



DB2 Warehouse Manager OLAP Processes



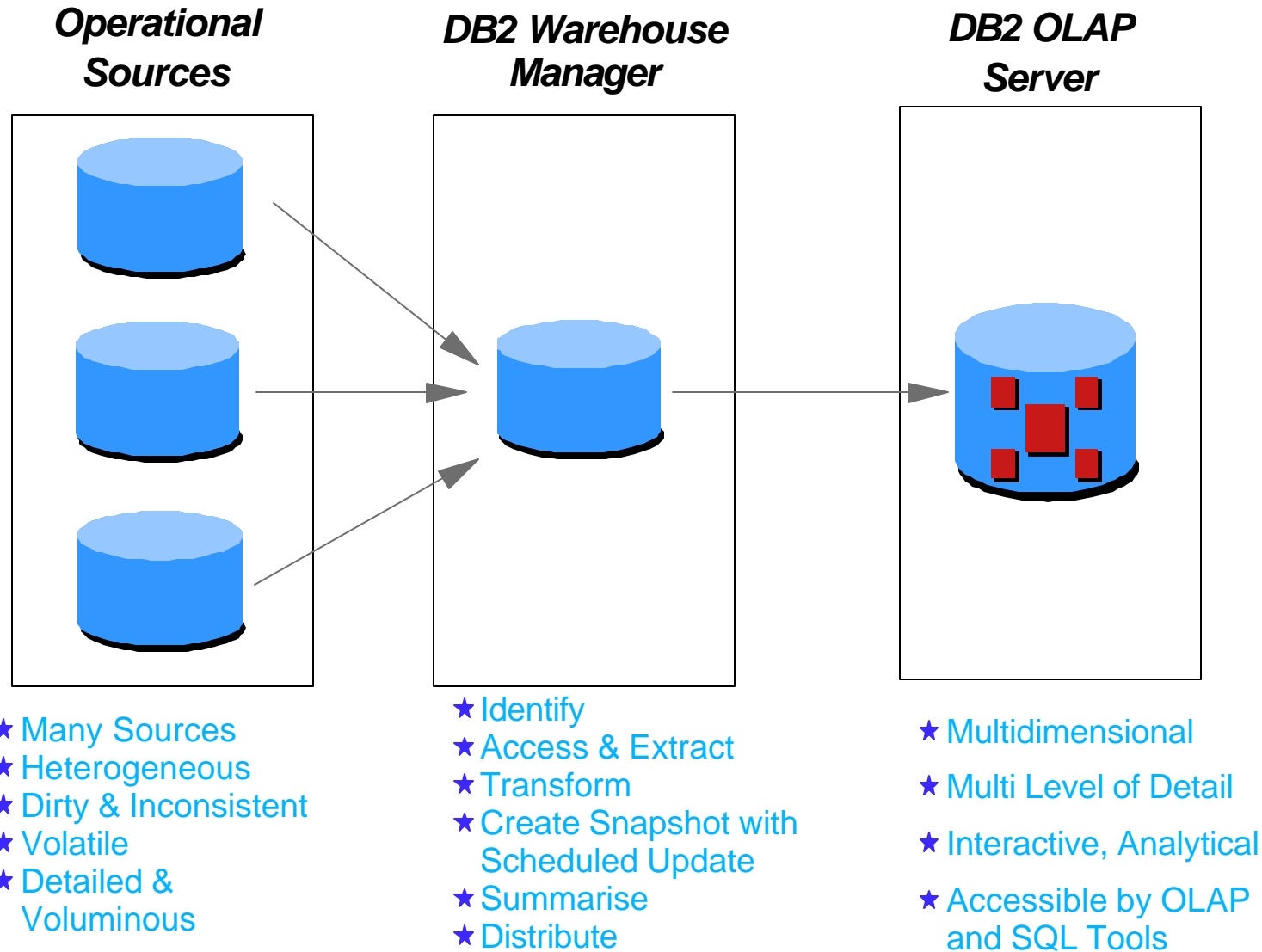
The world depends on it



DB2 Warehouse Manager Characteristics



The world depends on it



Summary



The world depends on it

Business Intelligence from DL/I Data

