



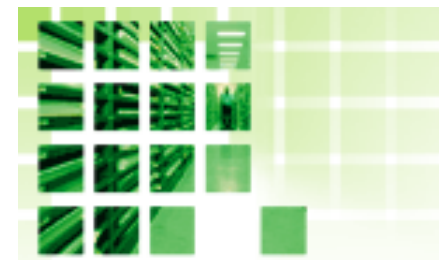
IBM Software Group

# DB2 Cube Views 8.2 Gold Consultant Briefing

*December 03, 2004*

*Pat Bates, Development Manager for DB2 Cube Views*

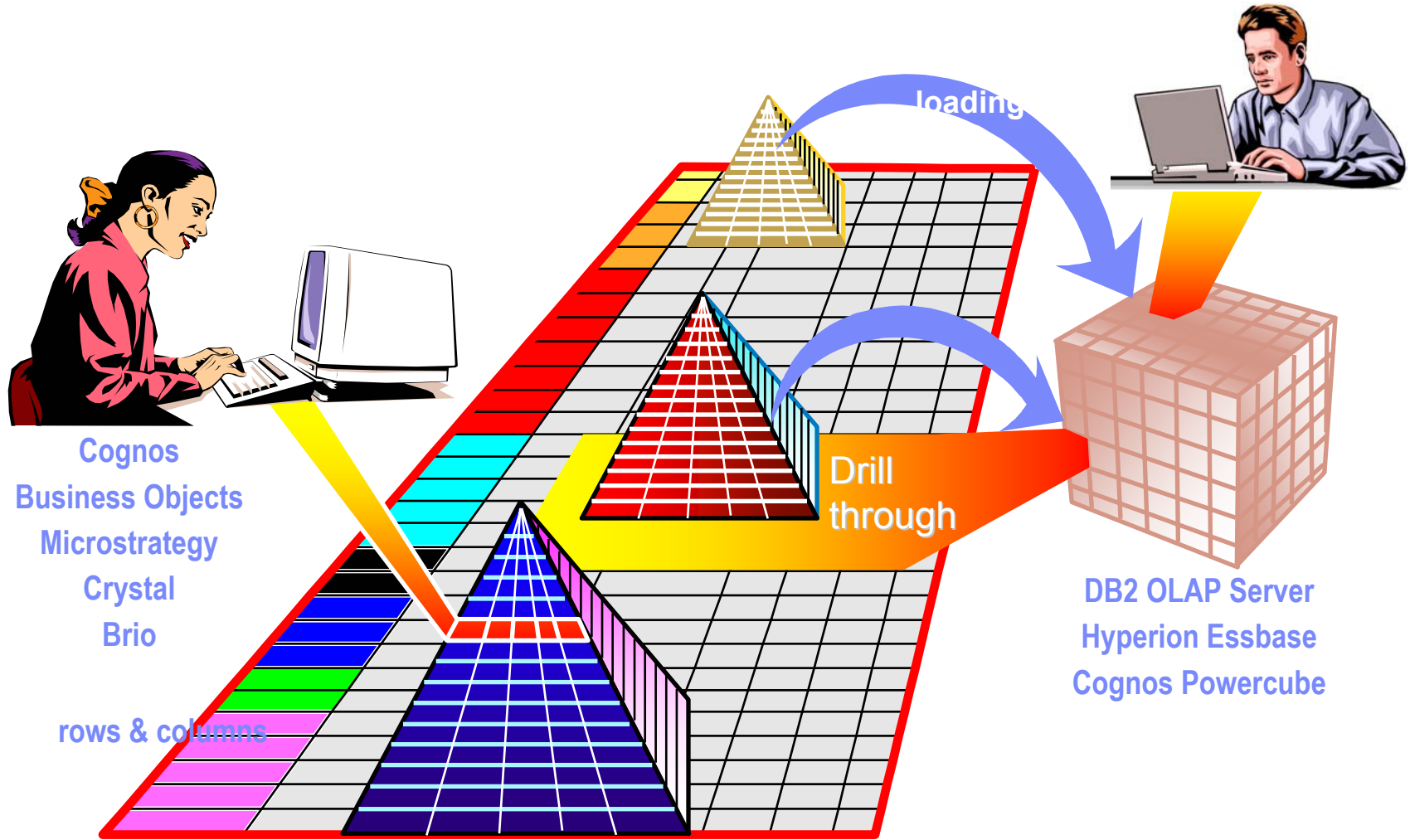
**DB2** Information Management Software



DB2 Cube Views  
Model Once, Use  
Everywhere OLAP  
Built into DB2

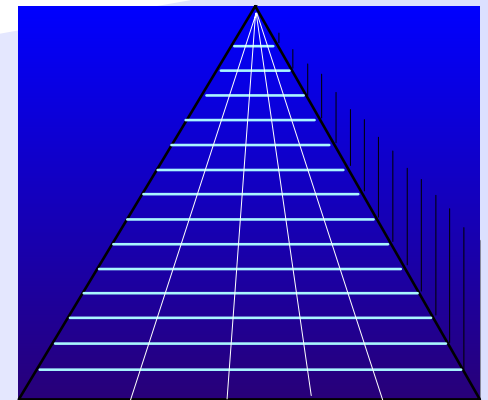
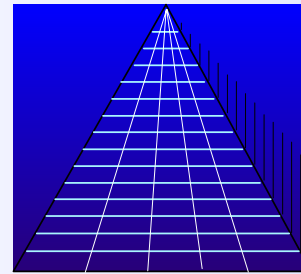
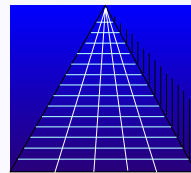
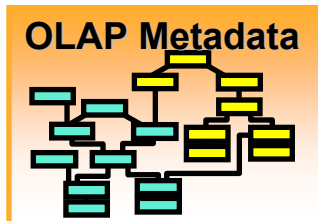


# DB2 Cube Views – Metadata, Performance, Access

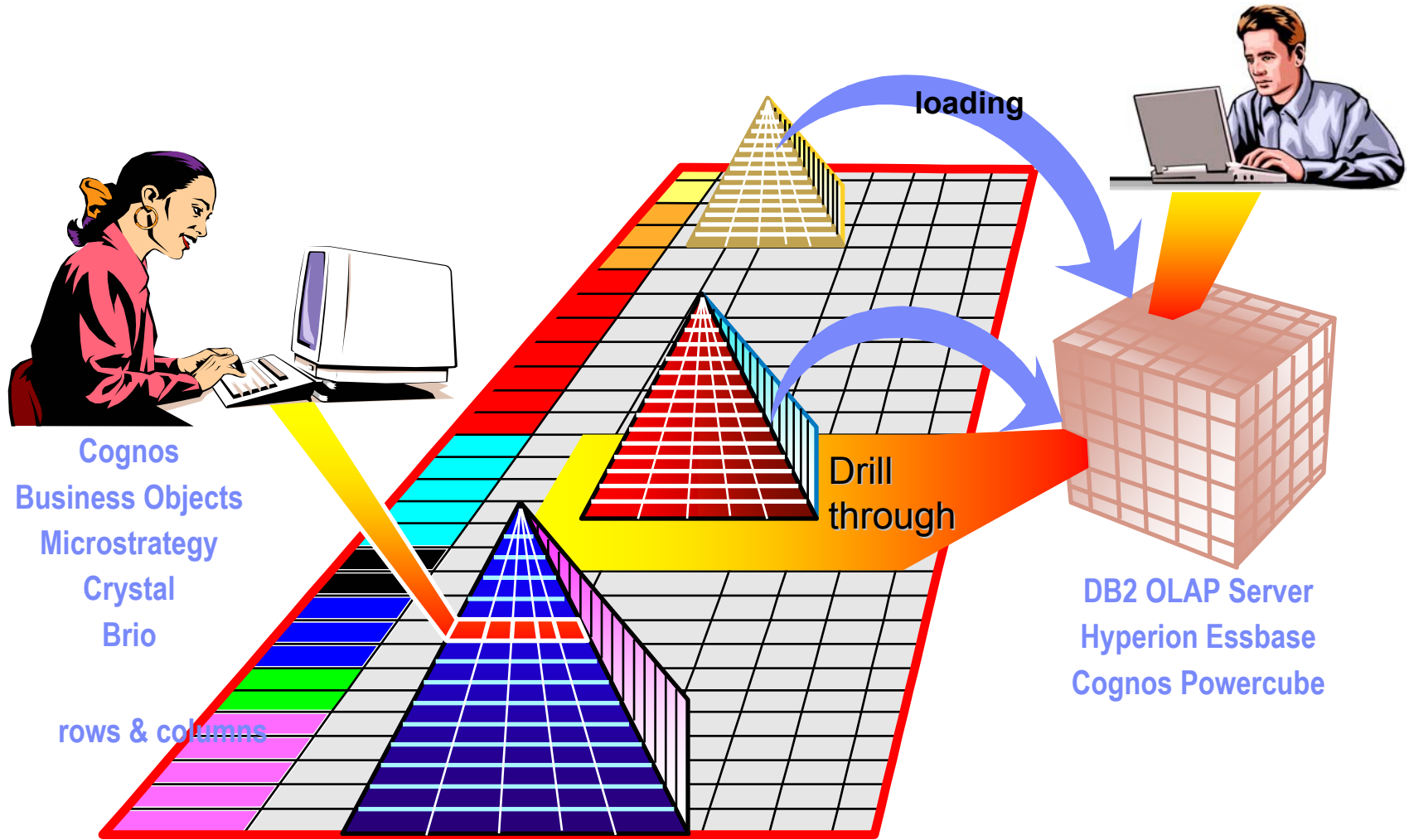


# DB2 Cube Views 8.2 Overview

- Review of DB2 Cube Views 8.1
- DB2 8.2 (Stinger) and Information Integrator Enhancements
- What's New in DB2 Cube Views
- Performance Test Results
- DB2 and Cube Views Working Together
- Brief Partner Update
- Q & A

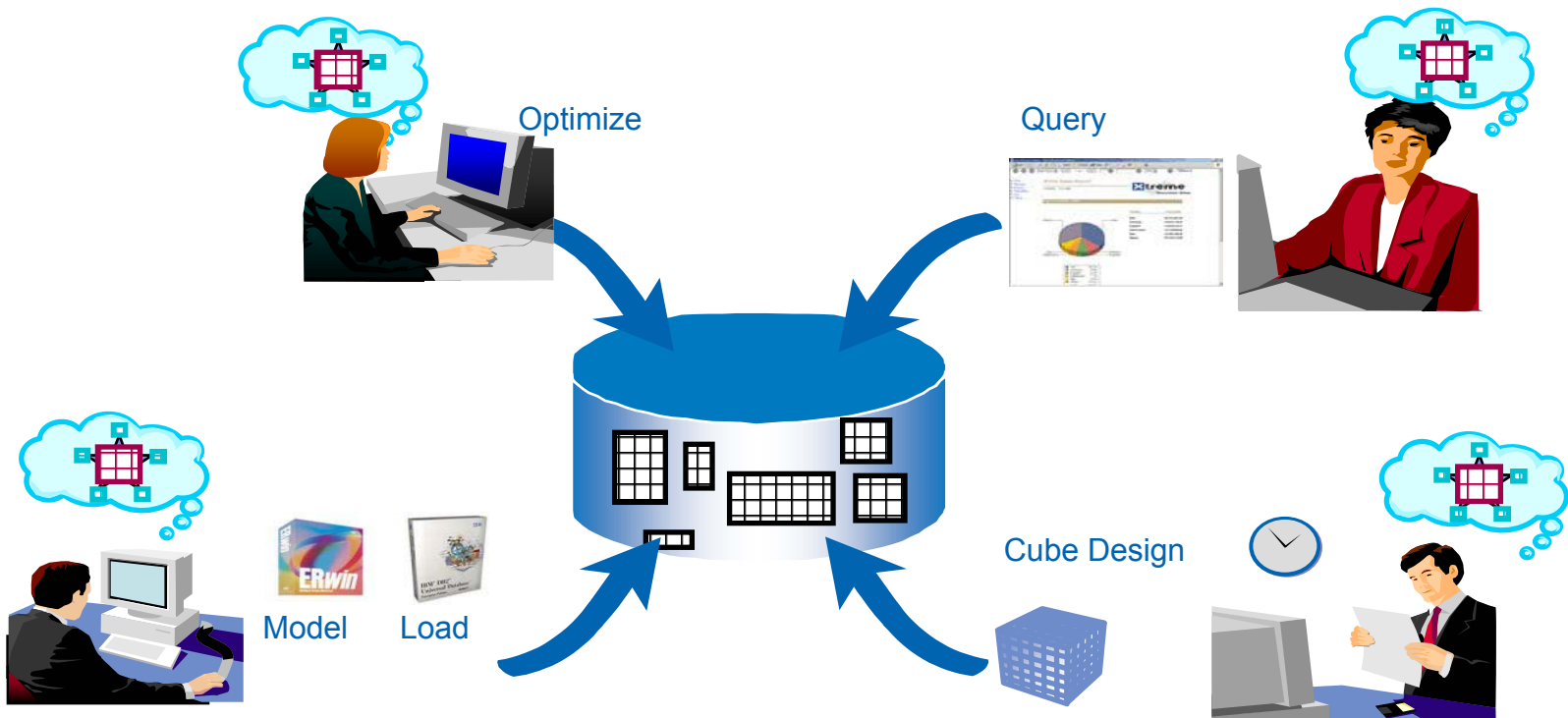


# DB2 Cube Views – OLAP Accelerator for DB2

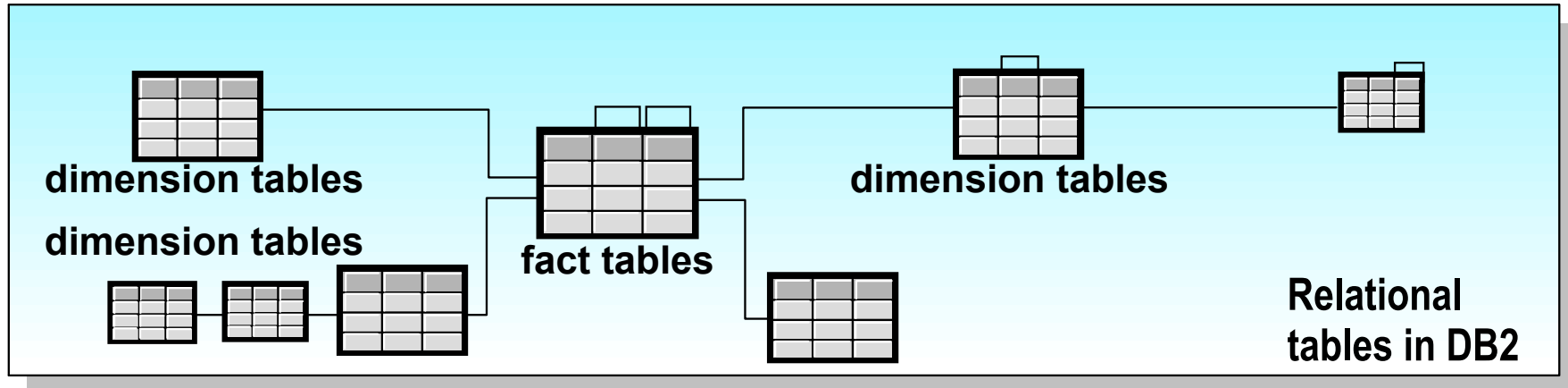


# Challenges in the World of OLAP

- Dimensional models are common
- Dimensional information is frequently lost and rediscovered



## Star Schema – The Starting Point in the Warehouse



- Key Challenges:

- ▶ Usage:

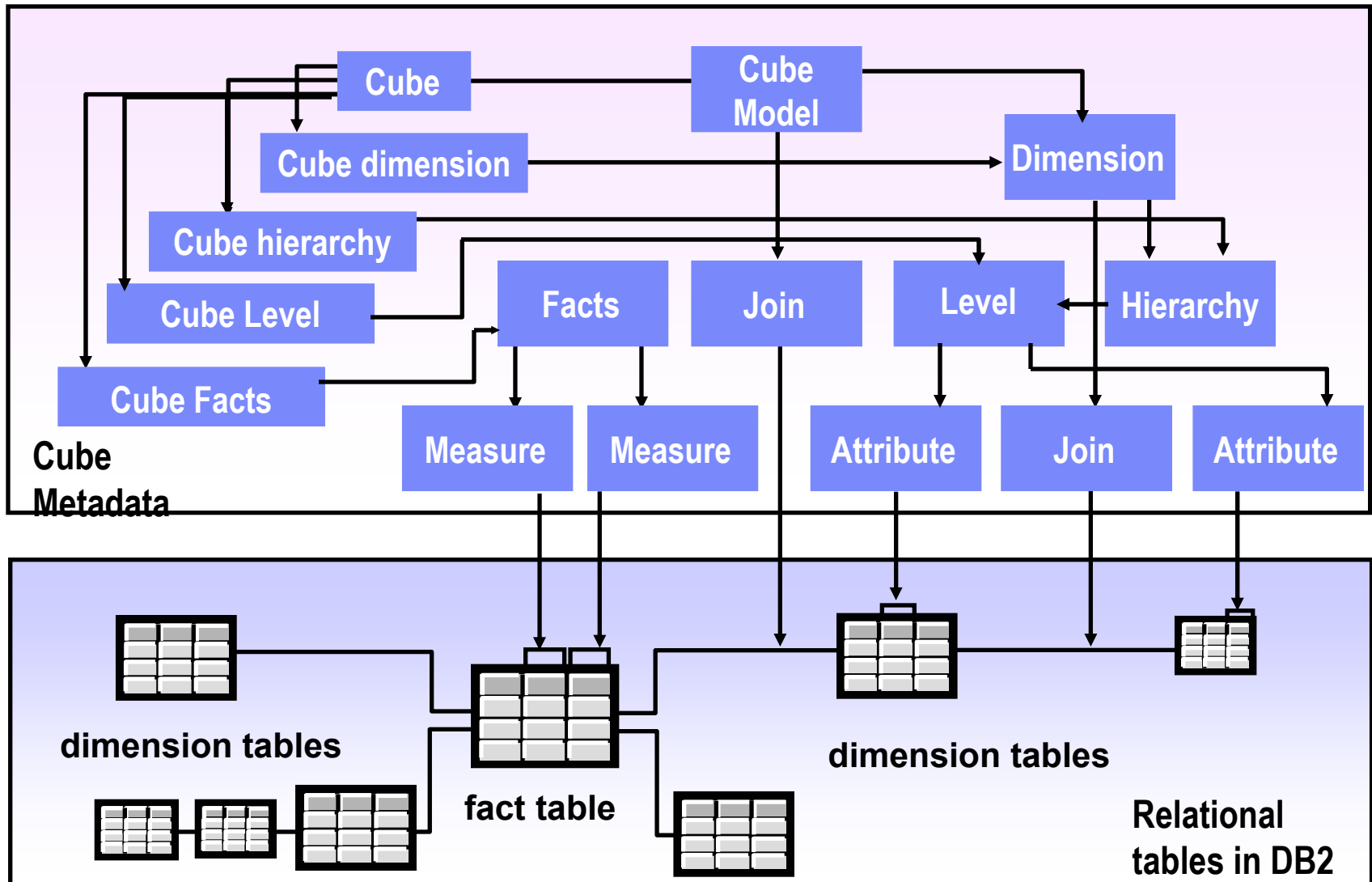
- Need to "understand" facts/formulas, dimensions, hierarchies
    - Need to share the understanding with Analytic tools

- ▶ Performance:

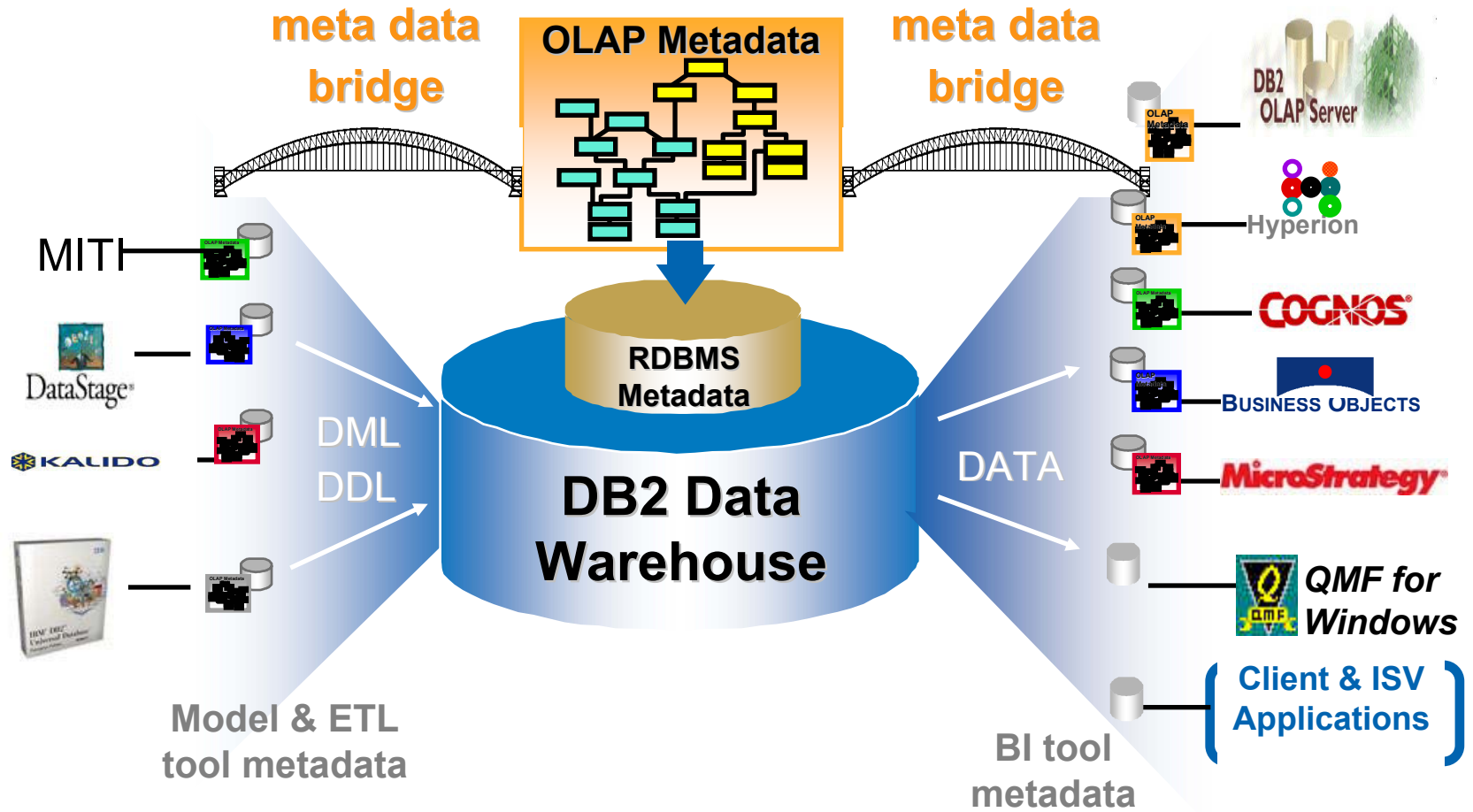
- Needs summarization
    - Needs optimization



# OLAP Model Objects - Multidimensional Metadata Catalog

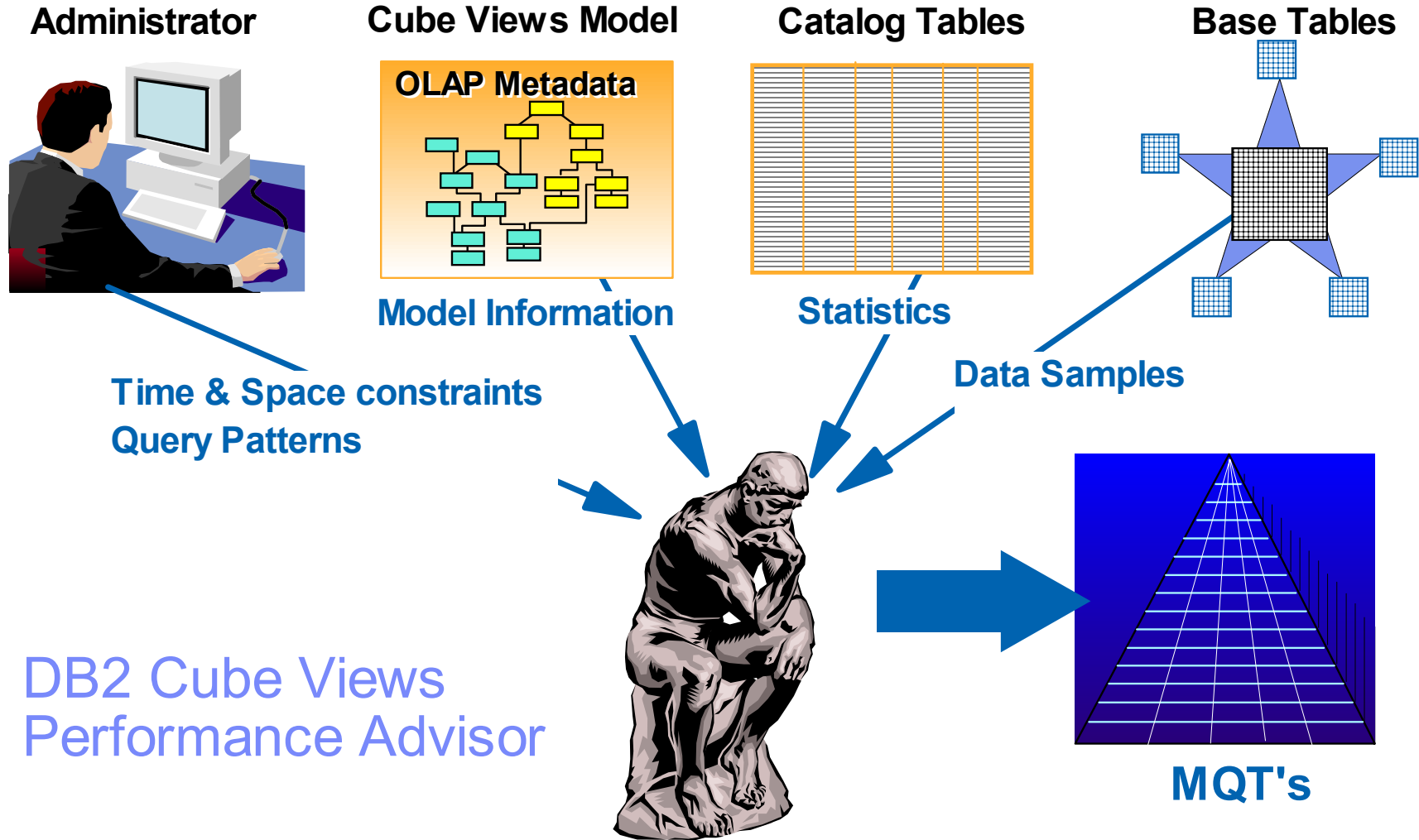


# Sharing the OLAP Metadata in DB2 Cube Views



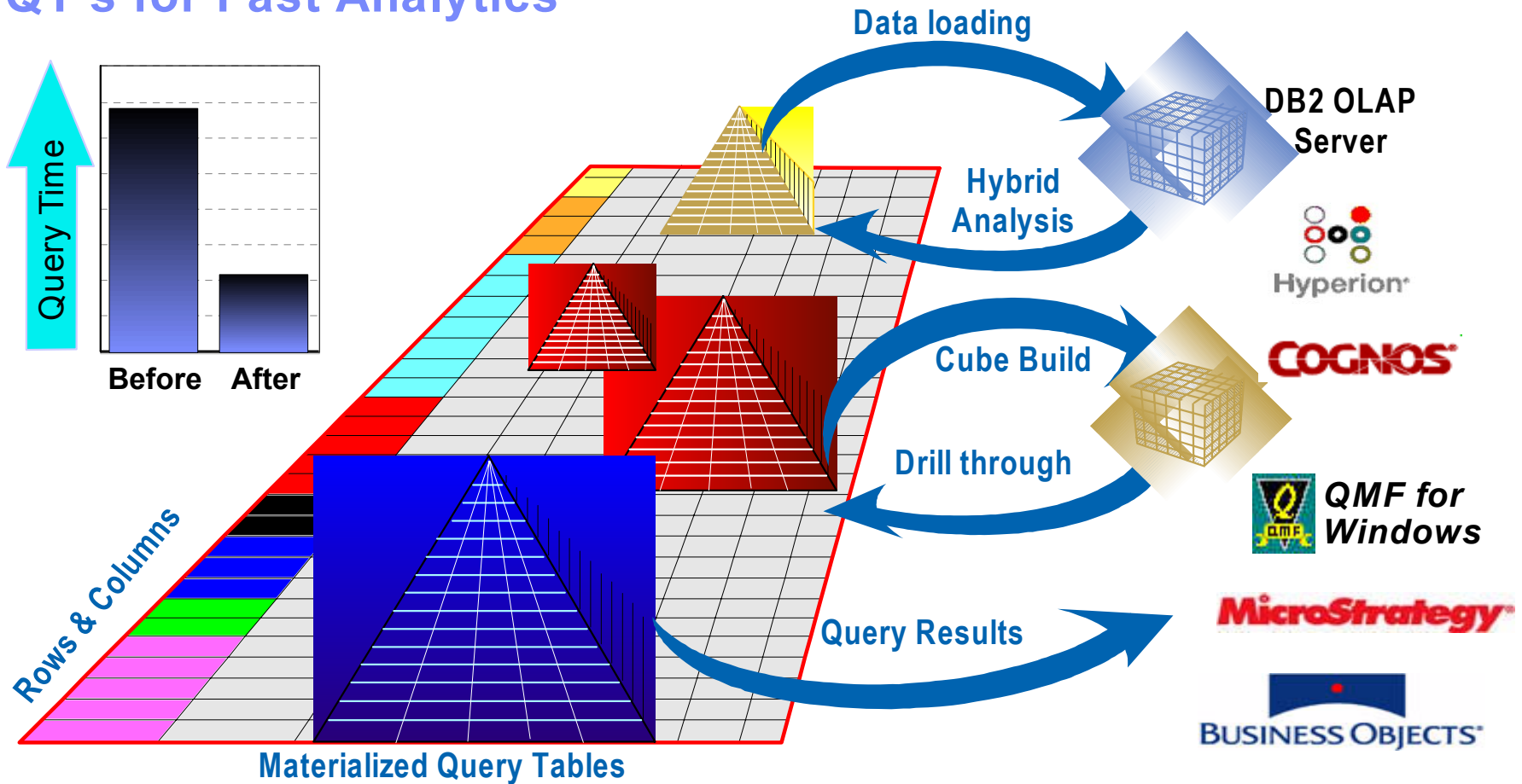


# DB2 Cube Views – Model-Based Optimization



# DB2 MQT's: Multilevel Aggregate "Caches"

## Cube Views Recommends MQT's for Fast Analytics



# DB2 and II Improvements for Cube Views

- Enhanced Functional Dependency Support - FD between any two columns enables:
  - ▶ More queries can now reroute to MQT
  - ▶ “Normalized” MQTs are more efficient/effective: skinnier and deeper
- Unique Constraints for Nick Names (Information Integrator)
- Query Rewrite / Reroute Defects Fixed



# DB2 Cube Views 8.2 – What's New Summary

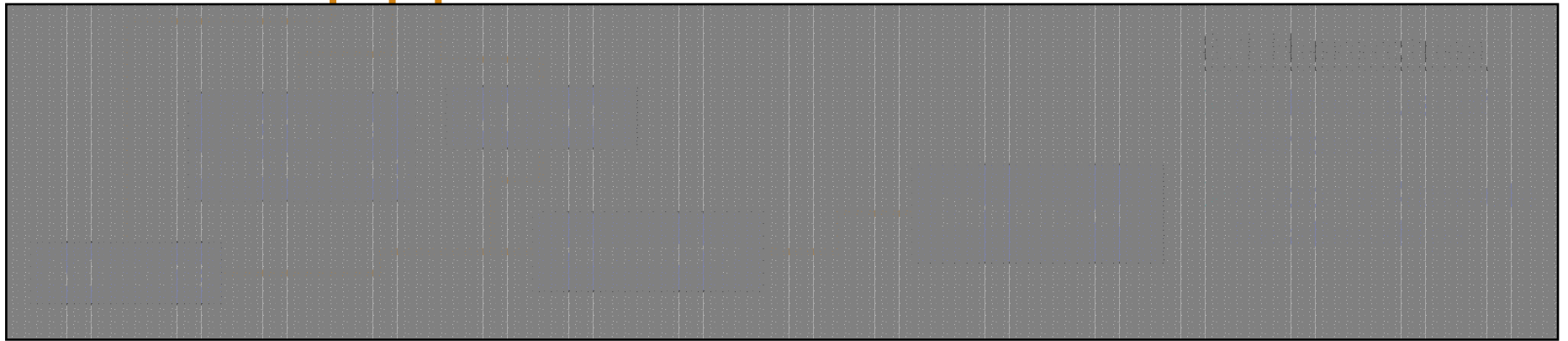
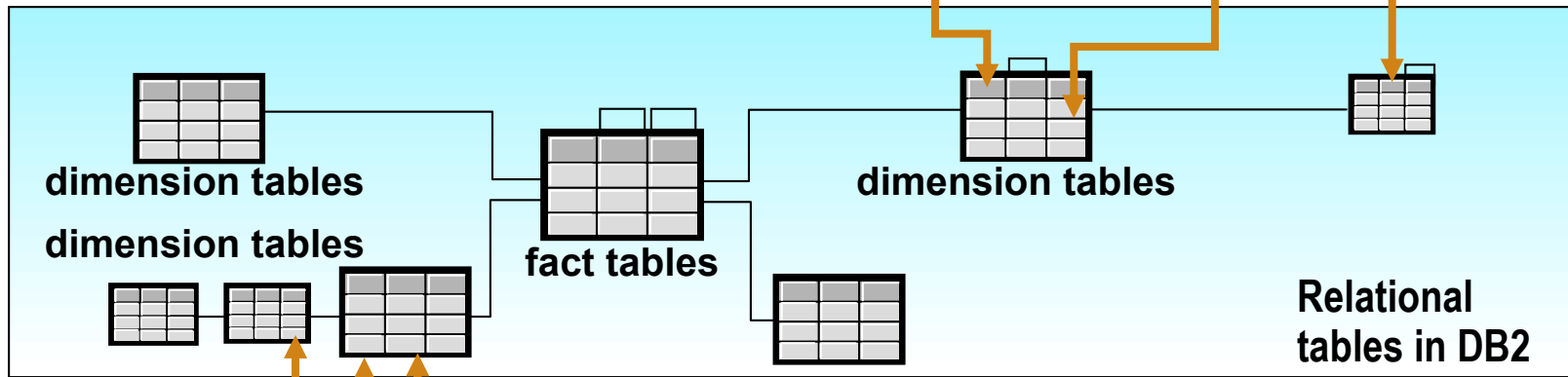
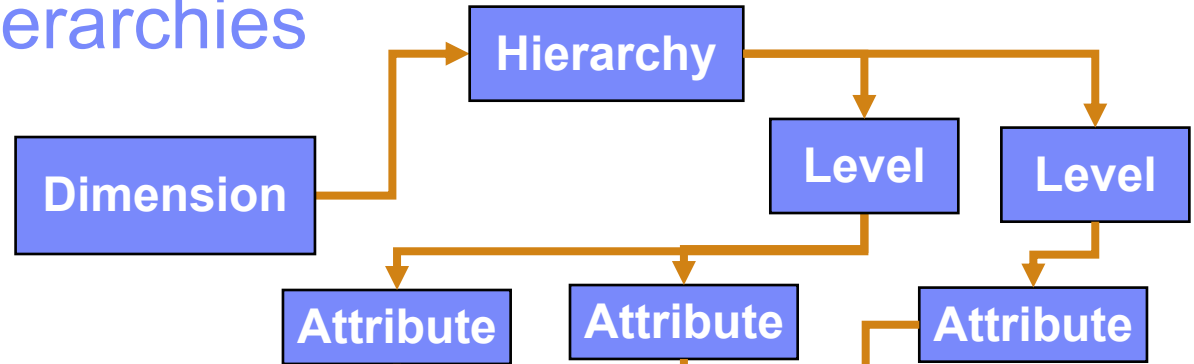
- Metadata Enhancements
  - ▶ New Level and Cube Level objects, Attribute Nullability, etc.
- Functional Dependency Exploitation
  - ▶ Normalized MQTs (Functional Dependency Exploitation)
- Federation Support
- New DBA Controls
  - ▶ Optimization Slices, Advisor Status / Cancel
- API Enhancements
  - ▶ Migration and backwards compatibility
  - ▶ Automatic creation of FD's
  - ▶ Automated Advisor Operation
- New Platform Support
- Improved Samples – CVSample Replaces MDSample



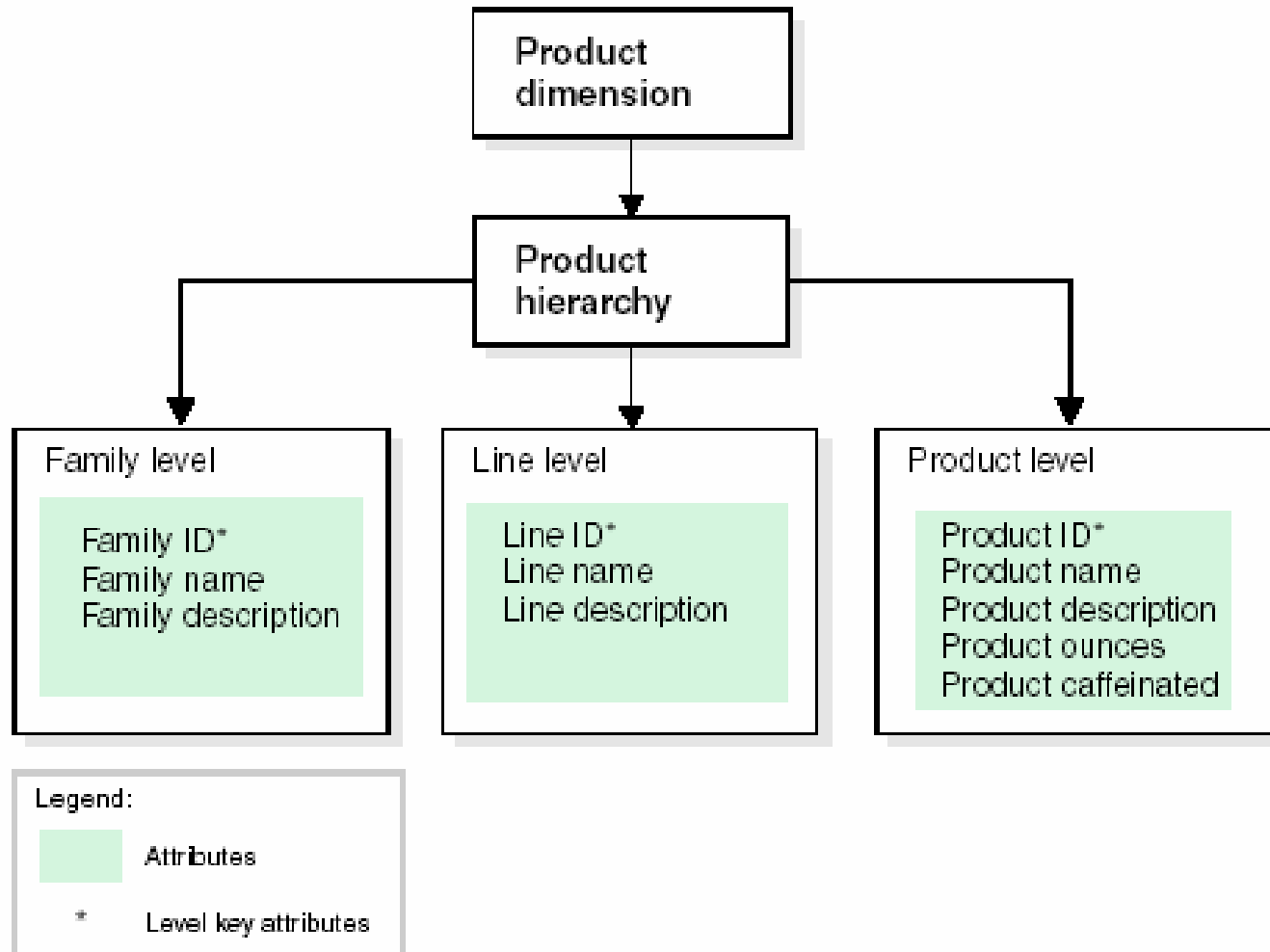
# New Levels for Hierarchies

## 8.2 Hierarchies

- Hierarchy: Ordered List of Levels
- Level: Key and Related Attributes



# Cube Views Level Example from CVSsample



# Level Object Details

- Hierarchies consist of an Ordered list of Levels
- Levels consist of
  - ▶ Level key – uniquely identifies data, multi-column keys supported
  - ▶ Default Attribute
  - ▶ (optional) Related Attributes
- Cube Hierarchies and Cube Levels follow suit
- Cube Views CREATE operation defines DB2 FDs automatically
  - ▶ User can specify that FDs not be created
  - ▶ FDs only created if DB2 will consider valid
    - Attributes must not be expressions
- Optimization Advisor can Exploit New Level Features:
  - ▶ Functional Dependencies enhance reroute rates / performance
  - ▶ New Feature: Optimization Slices



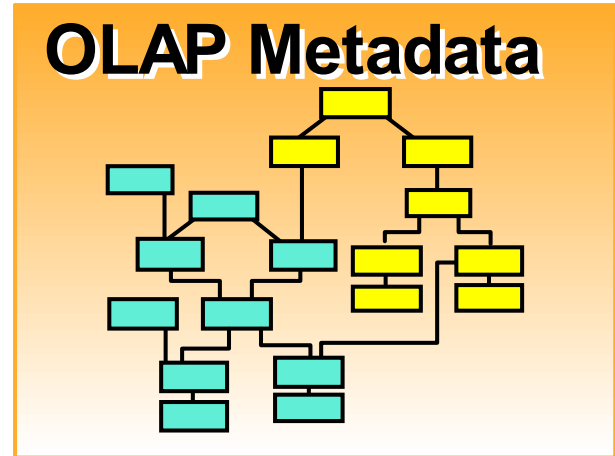
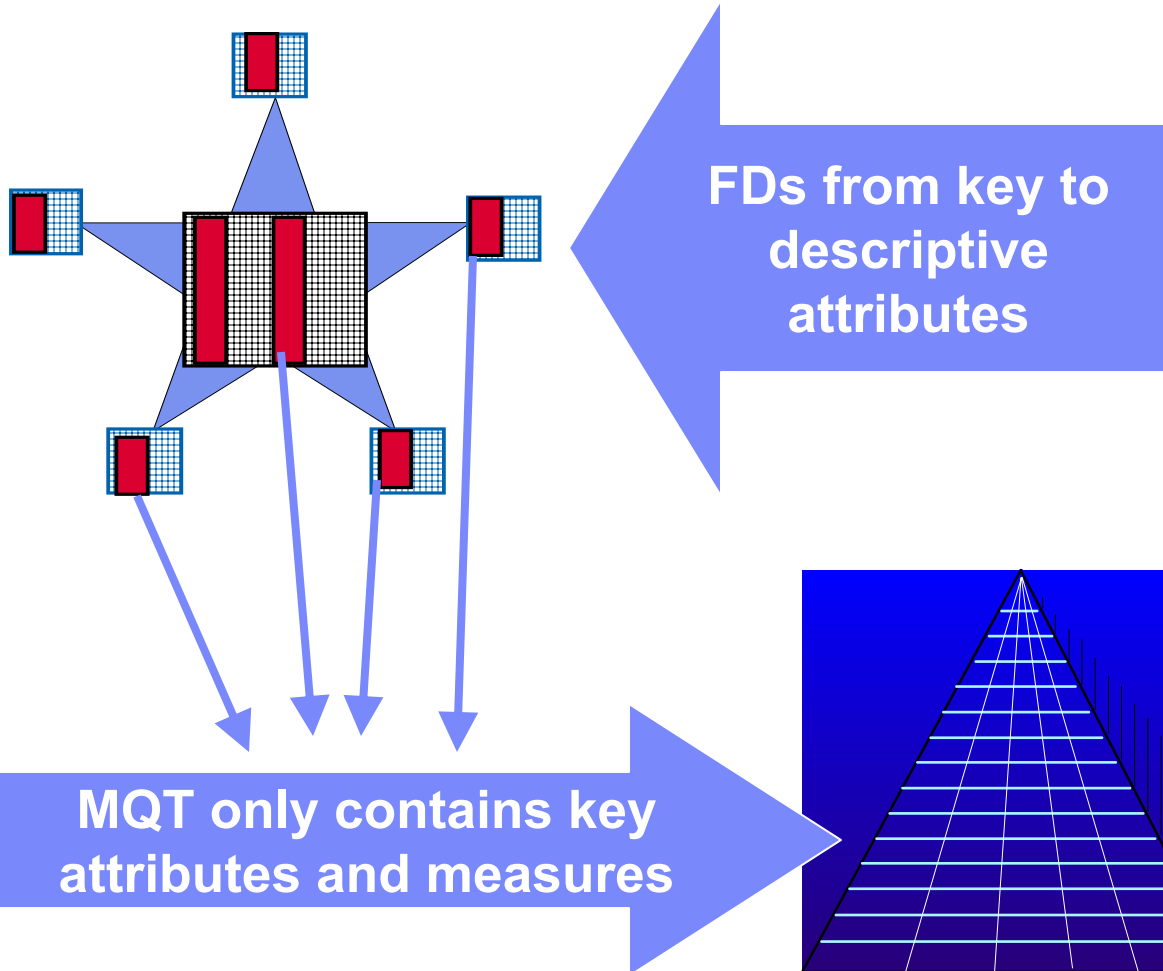
# DB2 Exploitation – Constraints and Functional Dependencies

- Functional Dependency: One (or more) column(s) uniquely determines (identifies) another column or columns.
  - ▶ Example: City\_Key => City\_Name
  
- DB2 contains semantic information about data relationships
  - ▶ Primary key
  - ▶ Unique constraints
  - ▶ Functional dependencies (new in DB2 V8.2)
  
- Keys that determine other data
  - ▶ Primary key determines all other columns in a table
  - ▶ Unique constraint determines all other columns in a table
  
- Exploit constraints and functional dependencies to make better MQTs





# Better MQTs thru functional dependencies



- Compact MQT
- Faster refresh
- Better coverage

# Benefit of Using Functional Dependency - Example

```
SELECT
SUM(T1."SALES") AS "SALES (SALESFACT)",
SUM(T1."SALES" - (T1."COGS" + T1."ADVERTISING")) AS "Profit",
T5."FAMILYID" AS "FAMILYID (FAMILY)",
T6."REGION_DIRECTOR" AS "REGION_DIRECTOR (LOCATION)",
T4."FISCAL_YEAR" AS "FISCAL_YEAR (TIME)"
```

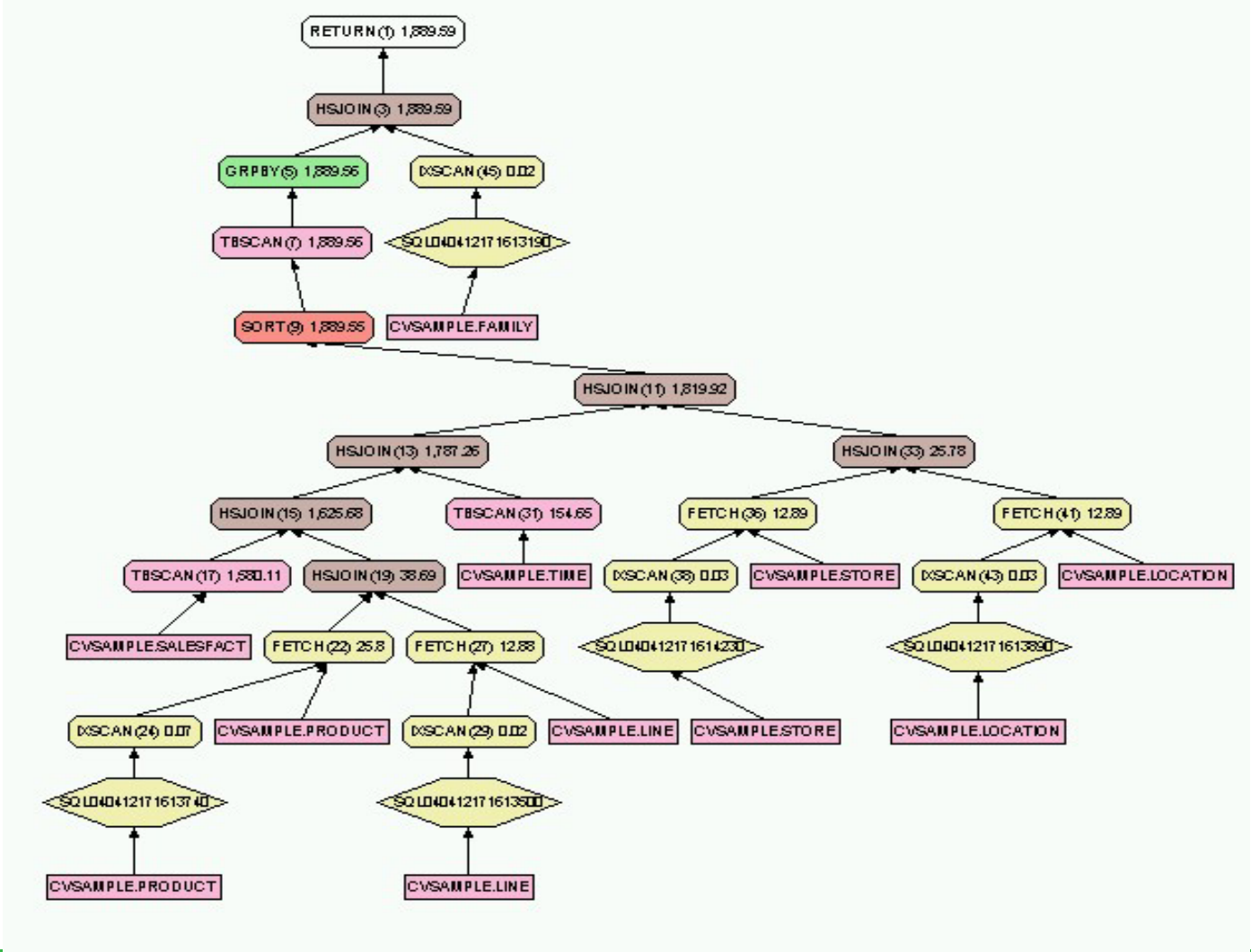
```
FROM
"CVSAMPLE"."SALESFACT" AS T1,
"CVSAMPLE"."STORE" AS T2,
"CVSAMPLE"."PRODUCT" AS T3,
"CVSAMPLE"."TIME" AS T4,
"CVSAMPLE"."FAMILY" AS T5,
"CVSAMPLE"."LOCATION" AS T6,
"CVSAMPLE"."LINE" AS T7
```

```
WHERE
T1."STOREID"=T2."STOREID" AND
T1."PRODUCTID"=T3."PRODUCTID" AND
T1."TIMEID"=T4."TIMEID" AND
T3."LINEID"=T7."LINEID" AND
T7."FAMILYID"=T5."FAMILYID" AND
T2."POSTALCODEID"=T6."POSTALCODEID"
```

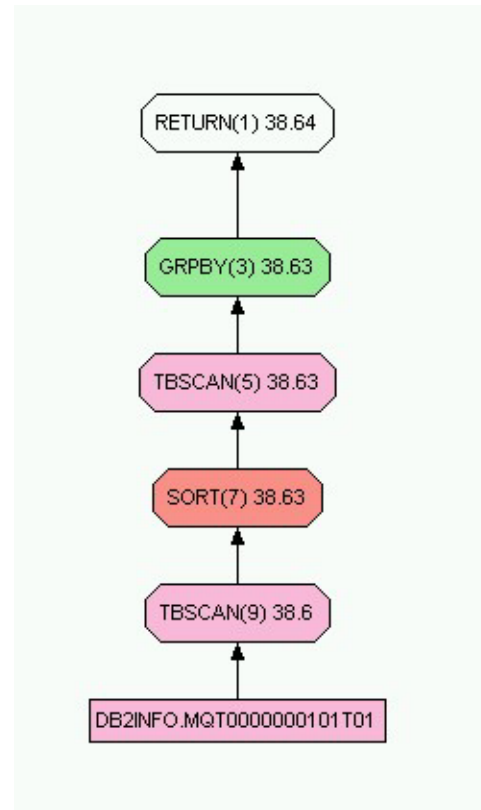
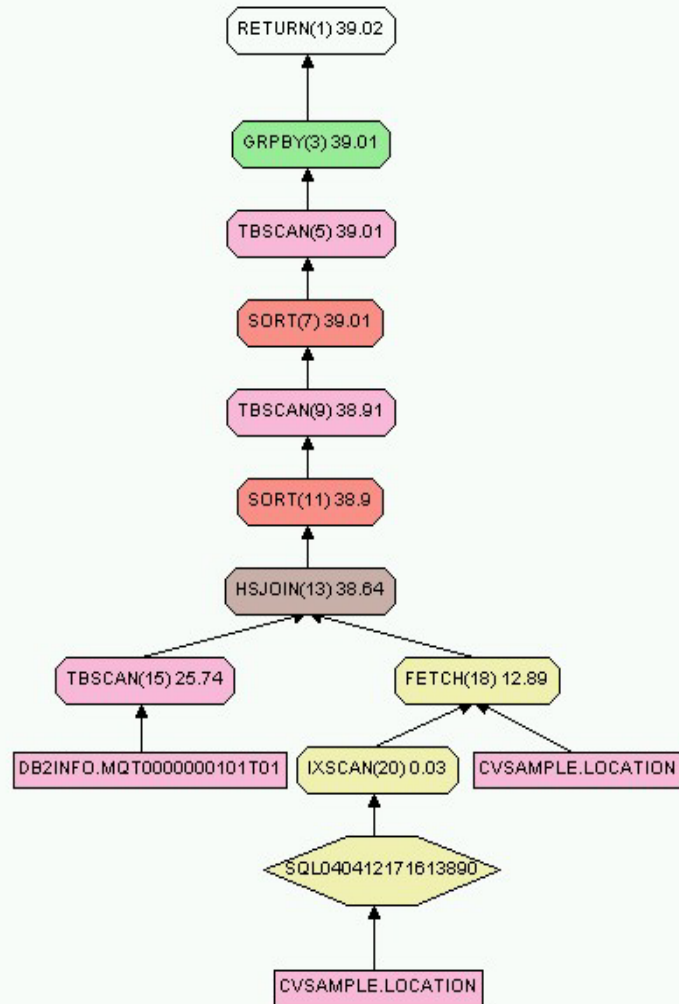
```
GROUP BY
T5."FAMILYID",
T6."REGION_DIRECTOR",
T4."FISCAL_YEAR";
```

Category	Query Cost
Using FD (8.2)	39.0196
Without FD (8.1)	38.6414
Without Any MQT	1889.5900

# Access Plan without MQT



# Access plans: Normalized vs Denormalized MQT



# Benefits of normalized MQT

- Fewer columns
- Shorter rows
  - ▶ Reduce maximum row width
  - ▶ Reduce average row width
- Less disk space
- Less temp space
- Faster refresh
- Faster reorg
- Faster runstats
- Query performance comparable
  - ▶ May vary from slower to faster



# FD Exploitation Performance Benefits Summary

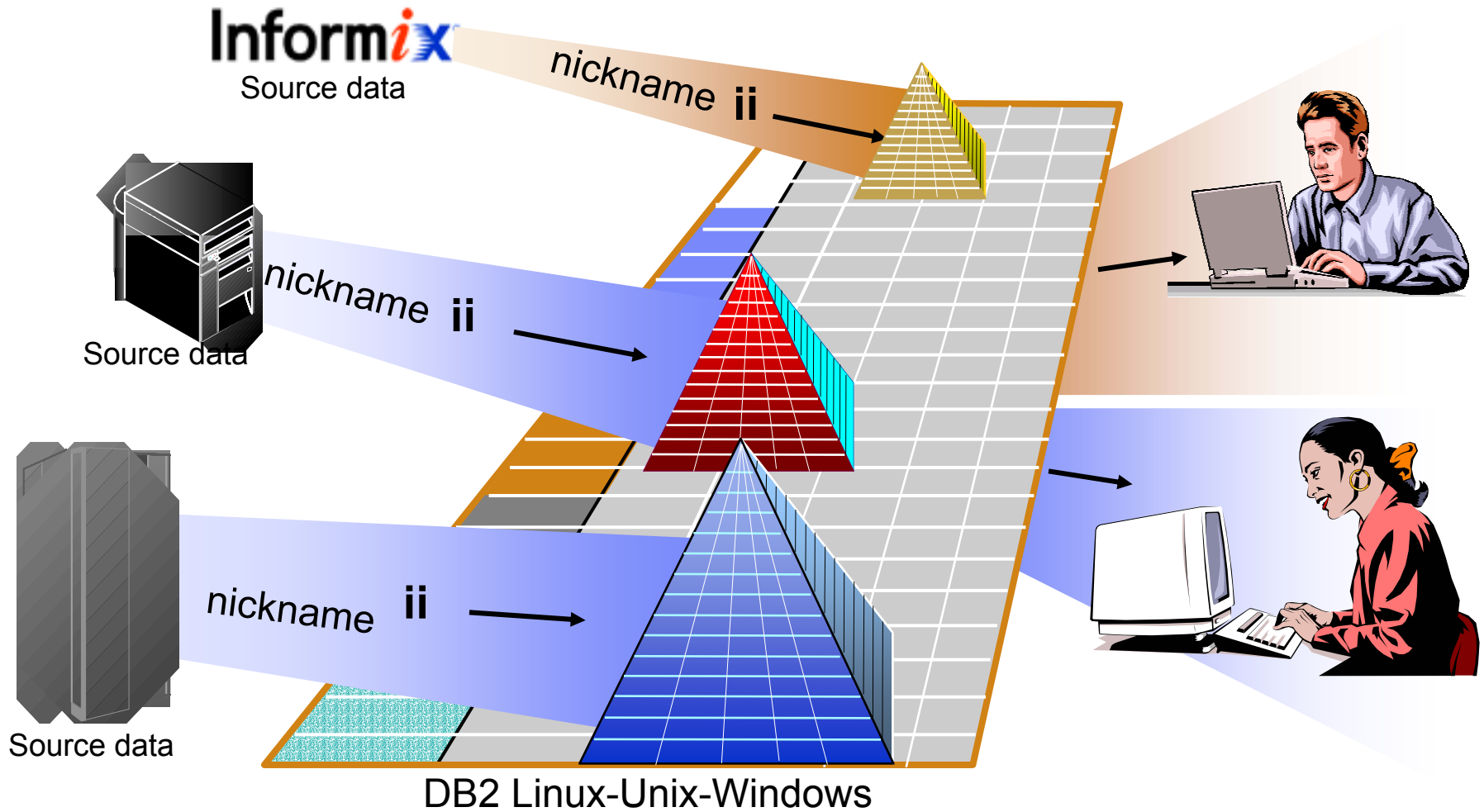
- Normalize MQTs by exploiting DB2 functional dependencies/constraints
  - ▶ Reduce MQT table size from 0% to 50%
  - ▶ Reduce MQT refresh temporary space from 0% to 80%
  - ▶ Reduce MQT refresh time from 0% to 30%
  - ▶ Reduce MQT reorg time from 0% to 60%
  - ▶ Comparable query time



## Normalization - Effect on 2 Customer Models

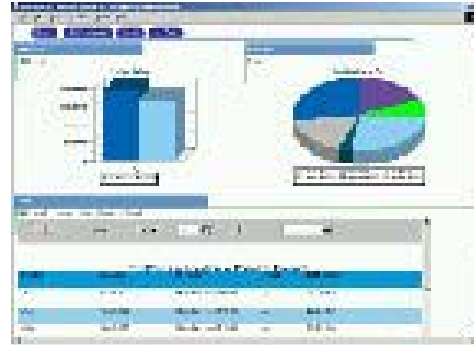
Category	Clickstream Data	E-Commerce Data
<b>MQT coverage</b>	Identical	Identical
<b>MQT rows</b>	Identical	Identical
<b>MQT columns</b>	40% Fewer	30% Fewer
<b>MQT row width (Bytes)</b>	42% Shorter	42% Shorter
<b>MQT table size</b>	37% smaller	42% smaller
<b>Refresh temporary space</b>	No change	81% less
<b>Refresh time</b>	7% faster	32% faster
<b>Reorg time</b>	60% faster	60% faster
<b>Runstats time</b>	36% faster	50% faster
<b>Query time</b>	16% faster	8% slower
<b>Query routing</b>	5% less	No change

# Cube Views via Federation

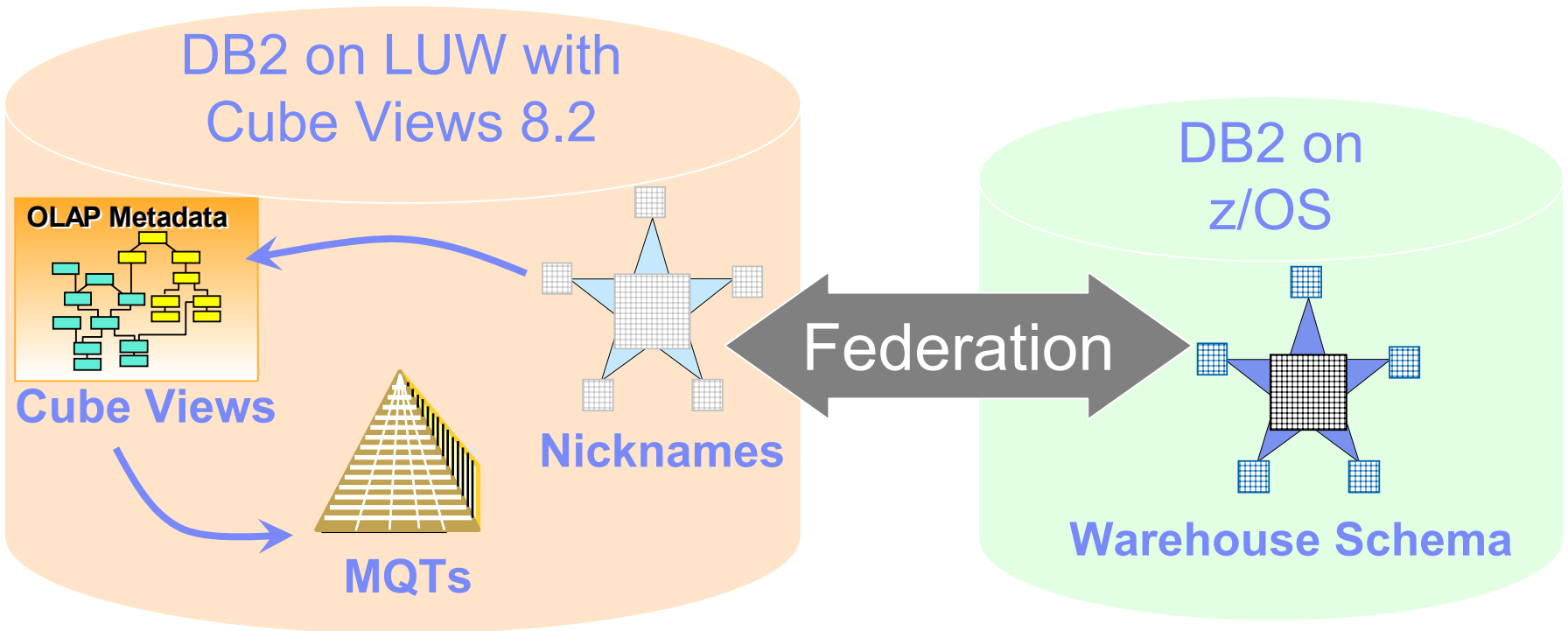
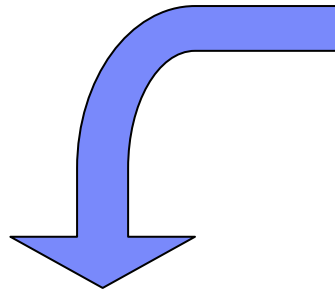




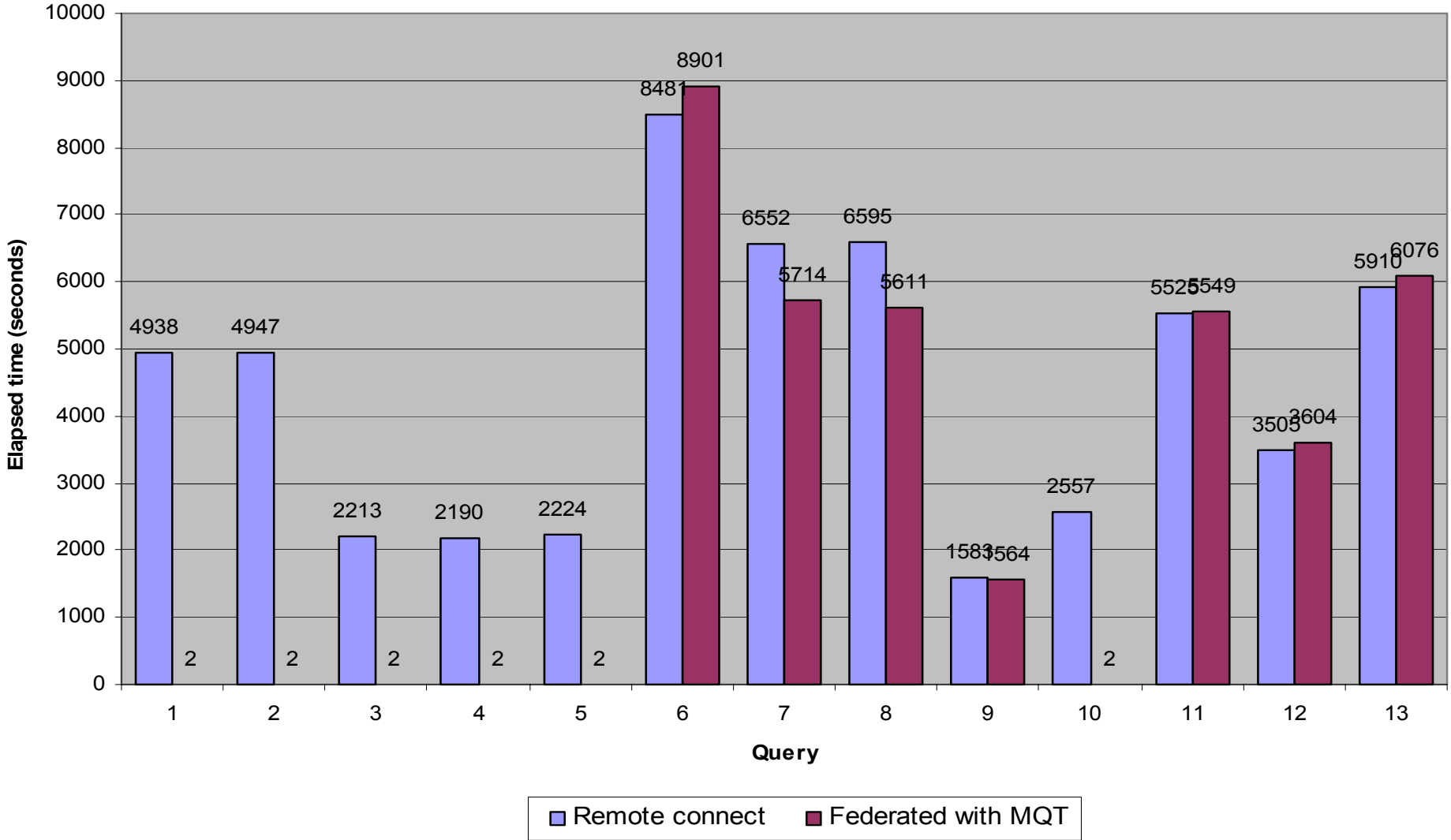
# Federated Support



BI Application



### Query time - Federated to z/OS (SurfAid model)



## Federated - Summary

- Reroute rate ranges from 15% to almost 80%
- Factors that influence reroute:
  - ▶ Specific workload – your mileage may vary
  - ▶ Advisor runtime and disk: longer sampling yields better results, bigger/more MQTs yield better reroute. Trade-offs with resources have to be made.
  - ▶ Optimization Slices will generally yield better reroute if the opt. slices match the query workload.
- Information we still don't have:
  - ▶ Dimension table co-location: queries may reroute to MQT but still join back to federated source. Need to test with co-located dim tables to determine how much of the workload can be fully satisfied by the distributed system.



# Better MQTs - Optimization Slices

**Legend**

- <Select>
- Drill-down
- Report**
- MOLAP extract
- Hybrid extract
- Drill through

**Market (Daily Sales)**

- All
- Region (Daily Sales)
- State (Daily Sales)
- City (Daily Sales)
- Postal code (Daily Sales)
- Store (Daily Sales)
- Any

**Product (Daily Sales)**

- All
- Family (Daily Sales)
- Line (Daily Sales)
- Product (Daily Sales)
- Any

**Time (Daily Sales)**

- All
- Year (Daily Sales)
- Quarter (Daily Sales)
- Month (Daily Sales)
- Day (Daily Sales)
- Any

**List of optimization slices:**

Type	Market (Daily Sales)	Product (Daily Sales)	Time (Daily Sales)
Report	City (Daily Sales)	Family (Daily Sales)	Month (Daily Sales)
<b>Report</b>	<b>Any</b>	<b>Line (Daily Sales)</b>	<b>Quarter (Daily Sales)</b>

➤ **Important Slices are Prioritized for Optimization**

# Optimization Slices

- Tightly Focused MQTs with Optimization Slices
  - ▶ Identify important Cube Dimensions and Levels
- Added to Cubes
- Tell the advisor what regions to optimize for
- A “slice” has a type and a level per dimension
  - ▶ Types:
    - Drill-down
    - Report
    - Drill-through
    - Molap-extract
    - Hybrid-extract
  - ▶ Levels: Any, All, or a specific level



## Optimization Slices Performance Results

<b>Category</b>	<b>Ariba Model</b>
<b>MQT coverage</b>	More targeted coverage
<b>MQT size</b>	3% of the original size
<b>Query routing</b>	33% increase in routing
<b>Query time</b>	15% faster
<b>Advise creation time</b>	3% of the original time



## DB2 Cube Views 8.2: Advisor: Progress Monitor and Suspend

The current progress is shown below. Time to completion is estimated and might change as the Optimization Advisor proceeds. You can click Stop at any time to have the Optimization Advisor return the current set of recommendations. Allowing the Optimization Advisor to complete more sampling will yield better recommendations.

Phase: Recommendations are complete.

Specified time limit: 10:22

Time estimate (hh:mm): 0:00

Progress: 10:22

Percentage of sampling: 0.0

Buttons: Stop, Close

### Progress:

- Time Remaining
- Rate of Sampling

### Stop the Advisor:

- Recommendation is "Good Enough"

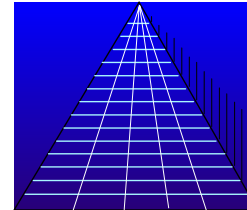
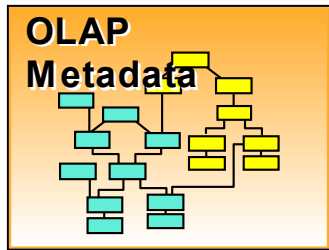
## Cube Views Advisor / The Design Advisor (V8.2)

- The Cube View Performance Advisor is **model** based
  - ▶ MQTs are recommended based on the metadata and rules
  
- The Design Advisor will be **query workload** based
  - ▶ MQTs recommended based on multi-query optimization (MQO) looking for common subexpressions (CSEs)
  - ▶ Integrated Index and MQT recommendations (MDC and Partitioning is a separate phase)
  - ▶ A “Candidate” phase and a “Recommendation” phase with an optional “Sampling” phase



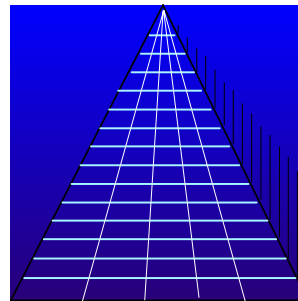
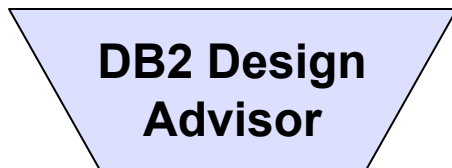
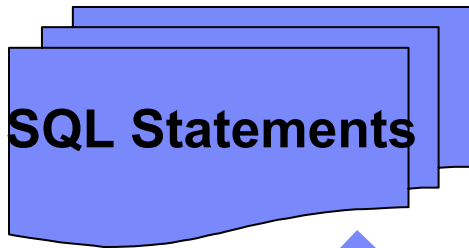


# Model-Based and Workload-Based Advisors



- MQTs Optimized for any Query Against the Cube Model

- ❖ Some queries may not reroute

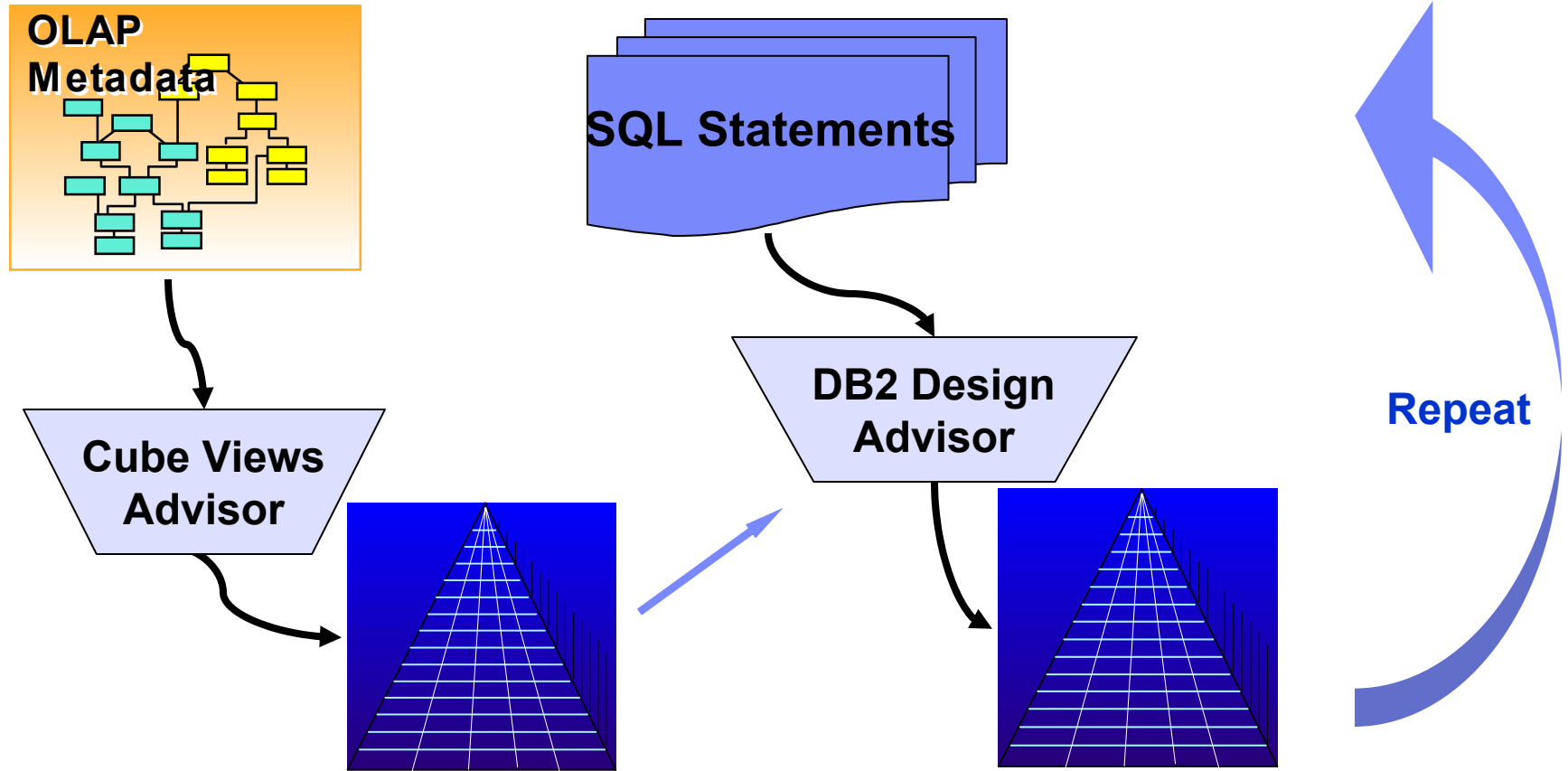


- MQTs Optimized for any Query in the Workload

- Recommends MDCs, too

- ❖ Queries outside the workload but still within the Warehouse Model may not route.

# How Do We Use the Advisors Together?



**Best**

**Performance = Model-Based MQT + Workload-Based MQT**



# Data Points – Performed via Real Customer Study

- Given 3 Major Ingredients:
  - ▶ A Cube Views Model against a real customer's schema
  - ▶ Data from the customer
  - ▶ A query workload of 57 queries representing ad-hoc queries
    - Count, Sum, Division, Avg, Min, Max, Distinct, Count Distinct
    - Group By, Group By Cube, Group By Rollup, Having, Select using Temporary Table
  
- Facts:
  - ▶ Fact table has 1,036,628 rows
  - ▶ Cube Views Query Type: Reporting
  - ▶ Cube Views Measures: Count, Sum
  
- Run the 2 Advisors in different orders and combinations to determine how they best complement each other



## Both Advisors: Findings

	# Queries Rerouted	# MQTS	MQT Size KB	MQT # Rows
<b>Cube Views Only</b>	22	1	22,656	136,509
<b>DB2 Only</b>	16	10	1856	7957
<b>Cube Views first then DB2</b>	30	6	25,888	156,430
<b>DB2 first then Cube Views</b>	25	11	24,224	144,466



## Enabled Partner Summary

Partner	8.1	8.2	Direction	Notes
Ascential	Y		2-way	
Business Objects	Y		2-way	
Cognos	Y	Y	2-way	Via MITI OEM
Embarcadero	Y	?	1-way	
FEnet	Y		1-way	Using web services APIs
Hyperion (DB2 OLAP)	Y		1-way	Two bridges (Brio + EIS)
Meta Integration	Y	Y	2-way	OEM-ed by Cognos, Embarcadero
MicroStrategy	Y		1-way	
QlikTech	Y	?	1-way	Bridge output is a load script
Rocket (QMF)	Y		1-way	Cube reporting front-end
Viador	Y		1-way	



## Other partners with bridges in progress

- Alphablox
- Arcplan
- Beacon IT (Japan)
- Cubus (Germany)
- Informatica
- Information Builders



## Cube Views 8.2 - Platforms

Hardware	OS	Server	Client
IBM pSeries 32-bit	AIX 4.3.3, 5L	✓	
IBM pSeries 64-bit	AIX 5L	✓	
Intel 32-bit (x86)	NT Workstation	✓	✓
	NT Server	✓	✓
	Windows 2000	✓	✓
	Windows XP	✓	✓
	Win 2003 Server	✓	✓
	RH Linux Pro	✓	
	<b>RHEL (Linux)</b>	✓	
	Suse Linux Pro	✓	
	<b>SLES (Linux)</b>	✓	
Itanium 64-bit (IA64)	Windows XP		✓
	Win 2003 Server	✓	✓
Sun 32-bit	Solaris 8, 9	✓	
Sun 64-bit	Solaris 8, 9	✓	
HP IPF (Itanium) 64-bit	HP-UX 11i v2	✓	



## DB2 Cube Views and Office Connect Analytics

- Office Connect Analytics Edition is no longer shipped with Cube Views in 8.2.
- Office Connect Analytics Edition (OCA) is now available as a free download:
  - ▶ FTP site TBA
  - ▶ Free, As-Is
  - ▶ No Warranty, No Support, No Future Enhancements
- OCA still makes a good demo/test tool
- Do not recommend it for production.





# Thank You!

