OLAP IN THE DATA WAREHOUSE

Chapter 15

Chapter Objectives

- Demand for OLAP
- Features of OLAP
- Hypercubes, drill-down, roll-up, slice-and-dice
- Different OLAP models
- OLAP implementation

Demand for OLAP

- Current methods of analysis provided in a data warehouse are **not sufficient for** more complex analysis in small amount of time
- Your users must have easy methods of performing complex analysis **along several business dimensions**
- Need an environment with **multidimensional conceptual view** for data
- Must be able to analyse data along **any number of dimensions** and at **any level of aggregation**

Demand for OLAP (cont..)

- **Time must be part of the dimensions** as it is critical for any analytical system
- True analytical systems must recognize the **sequential nature of time** (November comes after October)
- Users must not be penalized for complex queries, and the reports must be at a consistent speed in terms of size and the effort to formulate the query
- Dimensions must be **easily introduced to a report or be added with ease**
- **Speed of the analysis process** is very important

Demand for OLAP (cont..)

- An effective OLAP environment must not be only fast and flexible but must support **complex and powerful calculations:**
  - roll-ups
  - drill-downs
  - computation of margins (sales minus costs)
  - moving averages and growth percentages
  - trend analysis using statistical methods

Data Warehouse vs. Operational DBMS

- OLTP (on-line transaction processing)
  - Major task of traditional relational DBMS
  - Day-to-day operations: purchasing, inventory, banking, manufacturing
- OLAP (on-line analytical processing)
  - Major task of data warehouse system
  - Data analysis and decision making
- Distinct features (OLTP vs. OLAP):
  - User and system orientation: customer vs. market
  - Data contents: current, detailed vs. historical, consolidated
  - Database design: ER + application vs. star + subject
  - View: current, local vs. evolutionary, integrated
  - Access patterns: update vs. read-only but complex queries

Demand for OLAP (cont..)

	OLTP	OLAP
users	clerk, IT professional	knowledge worker
function	day to day operations	decision support
DB design	application-oriented	subject-oriented
data	current, up-to-date detailed, flat relational isolated	historical, summarized, multidimensional integrated, consolidated
usage	repetitive	ad-hoc
access	read/write index/hash on prim. key	lots of scans
unit of work	short, simple transaction	complex query
# records accessed	tens	millions
#users	thousands	hundreds
DB size	100MB-GB	100GB-TB
metric	transaction throughput	query throughput, response

Demand for OLAP (cont..)

- OLAP is the answer
  - Meet all the above requirements.
  - Can be implemented on the Web.
  - Designed for highly interactive analysis.
  - Visual presentations (charts and graphs).
  - Complements the use of data mining.
- “Providing On-Line Analytical Processing to User Analysts”: Dr. E.F. Codd

OLAP Guidelines - Dr Codd

- **Multidimensional conceptual view**: must be analytical and easy to use
- **Transparency**: make the underlying data repository, technology and the nature of the source data totally transparent to users
- **Accessibility**: Provide access to the data that is actually needed to perform the specific analysis, presenting a single view to the users
- **Consistent reporting performance**: Ensure users do not experience any degradation as the number of dimensions or DB size increase
- **Client/Server architecture**: Conform the system to this for optimum performance, flexibility and interoperability

OLAP Guidelines (cont…)

- **Generic dimensionality**: Have one logical structure for all dimensions. The data structure must not be biased towards any single data dimension
- **Multi-user Support**: All users must be able to work concurrently and ensure data access, integrity and access security
- **Unrestricted cross-dimensional operations**: System to recognize dimensional hierarchies and automatically perform roll-up or drill-down operations within a dimension
- **User-friendly Data Manipulation**: Enable pivoting, drill-down, and roll-up to be done via drag-and-drop and not by menu-driven user interface

OLAP Guidelines (cont…)

- **Unlimited Dimensions and Aggregation Levels**: Accommodate at least 15-20 dimensions within an analytical model. Each dimension must allow unlimited number of aggregation levels
- **Flexible reporting**: Provide capabilities for the user to arrange columns, rows and cells so that its easy to analyse and manipulate information
- **Dynamic Sparse Matrix Handling**: The system must be able to deduce distribution of data and adjust the storage and access to maintain good level of performance

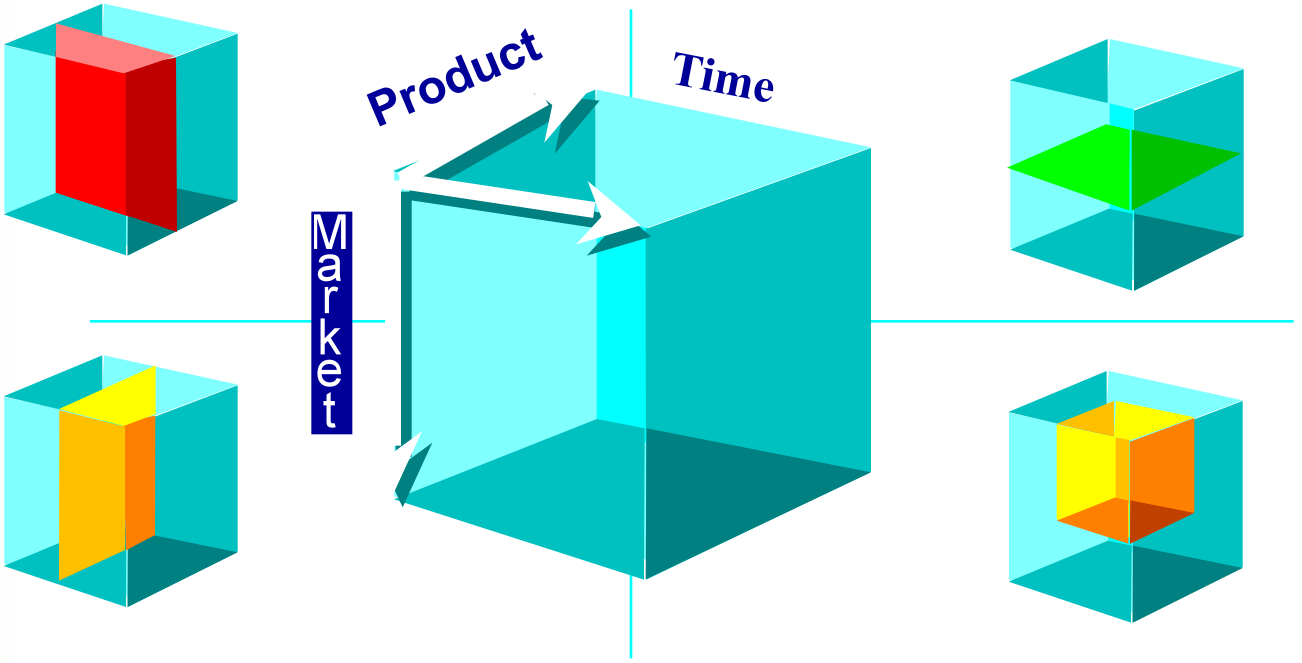
OLAP Requirements

- **Drill-through to detail level:** Smooth transition from the MDDB to the detail in the source data warehouse.
- **OLAP Analysis Model: (categorical, contemplative and formulaic).** We could describe these as parameterized static reporting, slicing and dicing with drill down, 'what if?' analysis and goal seeking models (data mining), respectively.
- **Treatment of Non-Normalized Data**. This refers to the integration between an OLAP engine and denormalized source data. Dr Codd pointed out that any data updates performed in the OLAP environment should not be allowed to alter stored denormalized data in feeder systems.
- **Storing OLAP Results:** Do not deploy write-capable OLAP tools on top of transactional systems.
- **Missing Values**: Ignore missing values
- **SQL Interface**
- **Incremental Database Refresh**: Ability to refresh OLAP data

OLAP

Product Manager's View

Regional Manager's View



Financial Manager's View

Ad Hoc View

- **Basic Features of OLAP**
  - Multidimensional analysis
  - Drill down and roll-up
  - Slice-and-dice or rotation
  - Time intelligent (year-to-date)
  - Navigation in and out of details
  - Fast response times for interactive queries
  - Easy scalability
- **Advanced Features of OLAP**
  - Powerful calculations
  - Drill-through across dimensions or details
  - Derived data values through formulas
  - Cross-dimensional calculations
  - Collaborative decision-making
  - Report generation with agent technology
  - Application of alert technology

Basic Operations for OLAP

- Drill Down
  - Navigate to higher levels of detail
  - Example: from regional analysis to specific plant analysis, further to team analysis, …

Drill Down

Volume of Product (numbers in 1000)		CellPhone		Pager	
		1001	1011	2001	2011
West	San Jose	33	12	8	12
	Boulder	45	34	20	23

Drill Down
Dimension: Location
Member: San Jose

Volume of Product (numbers in 1000)		CellPhone		Pager	
		1001	1011	2001	2011
San Jose	Team 1	20	8	6	7
	Team 2	13	4	2	5

Roll Up

- Navigate to lower levels of detail
- Example: from month analysis to a quarter analysis

Roll Up

Volume of Product (numbers in 1000)		1996			
		Qtr1	Qtr2	Qtr3	Qtr4
West	San Jose	78	45	34	56
	Boulder	90	67	87	91

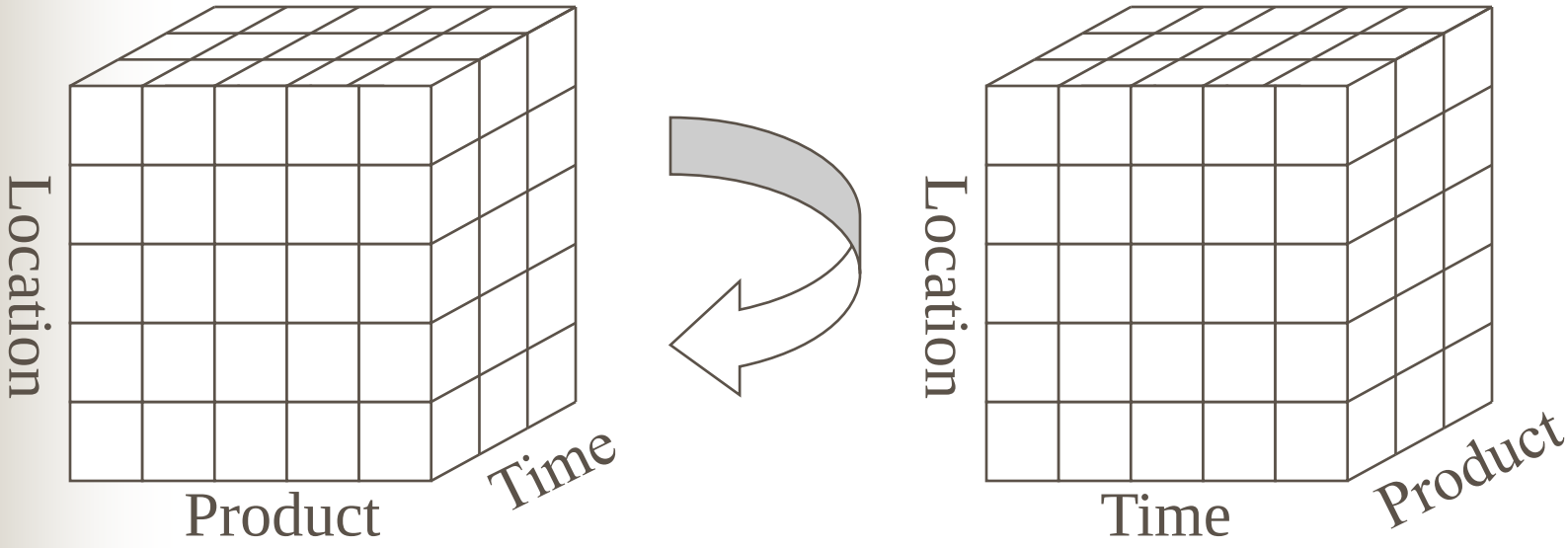
Roll Up

Dimension: time

Volume of Product (numbers in 1000)		Quarter 1		
		Jan	Feb	Mar
West	San Jose	30	26	22
	Boulder	28	30	32

- Pivot
  - Rotate the cube
  - Example: change the perspective from "Region X Product" to "Region X Time"

Pivot



Pivot (cont.)

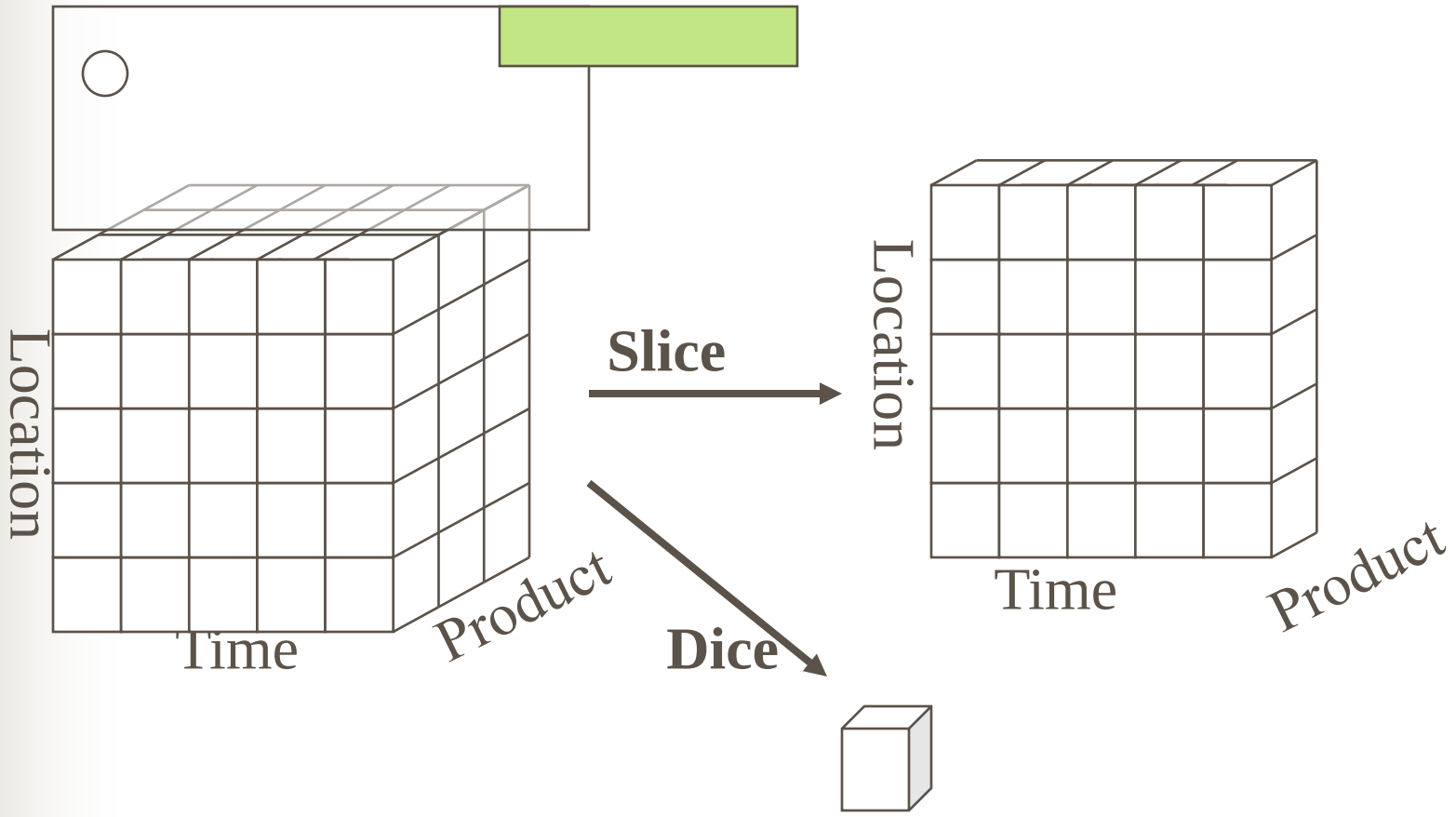
Volume of Product (numbers in 1000)		CellPhone		Pager	
		1001	1011	2001	2011
West	San Jose	33	12	8	12
	Boulder	45	34	20	23

Volume of Product (numbers in 1000)		1996 (CellPhone & Pager)			
		Qtr1	Qtr2	Qtr3	Qtr4
West	San Jose	78	45	34	56
	Boulder	90	67	87	91

Basic Operations for OLAP (cont.)

- Dice
  - Get one cell from the cube (the smallest slice)
  - Example: get the production volume of Armonk, for CellPhone 1001, in January (here, we suppose *plant*, *product model* and *month* are the smallest members in **Location**, **Product**, **Time** dimensions respectively)
- Slice
  - Cut through the cube, so that users can focus on some specific perspectives
  - Example: only analyzing on the product CellPhone

Slice & Dice



Slice & Dice (cont.)

Volume of Product (numbers in 1000)		1996 (CellPhone & Pager)			
		Qtr1	Qtr2	Qtr3	Qtr4
West	San Jose	78	45	34	56
	Boulder	90	67	87	91

Volume of Product (numbers in 1000)		1996 (CellPhone only)			
		Qtr1	Qtr2	Qtr3	Qtr4
West	San Jose	53	35	20	48
	Boulder	76	57	40	80

Dice:

Slice 1: Loc=1..2, Prod=3..4

Time=3..4

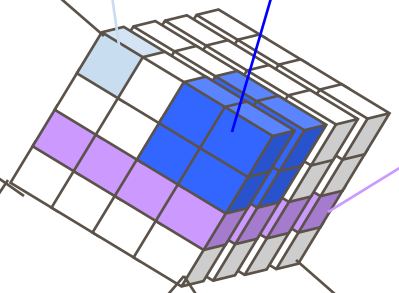
Loc=1..1

Prod=1..1

Time=4:4

Slice2: Loc=3..3, Prod=1..4

Time=1..4



Loca-tion _Key	Continent	Country
1	Europe	Germany
2	Europe	Spain
3	America	USA
4	America	Canada

Location Table

Product _Key	Vehicle _Kind	Vehicle _Size
1	Car	Small
2	Car	Big
3	Truck	Short
4	Truck	Long

Product Table

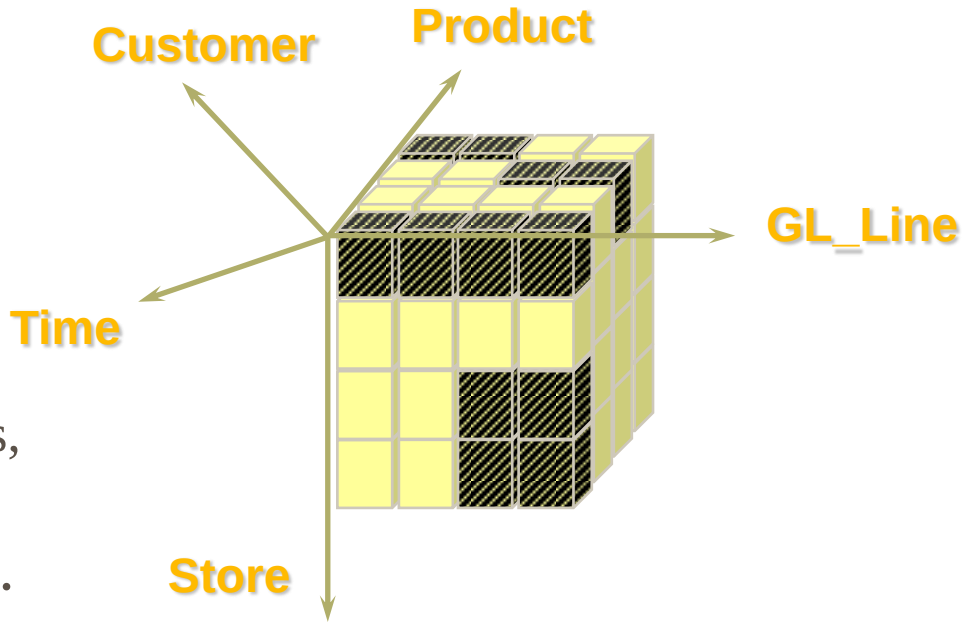
Time _Key	Year	Half _Year
1	1996	1HF
2	1996	2HF
3	1997	1HF
4	1997	2HF

Time Table

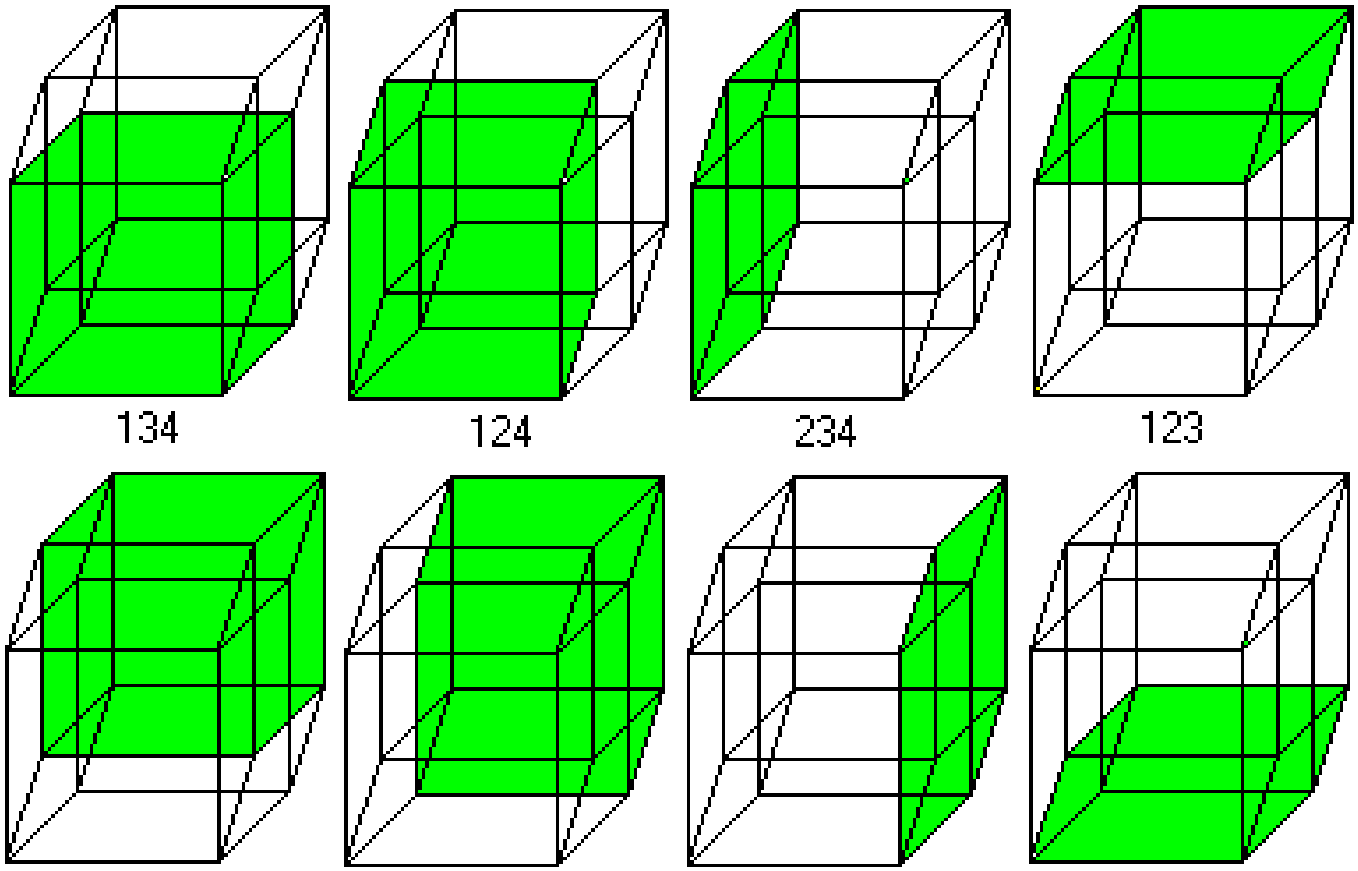
Hypercubes

- Variant of the multidimensional model
- One hypercube in the database
- Data has the same dimensions

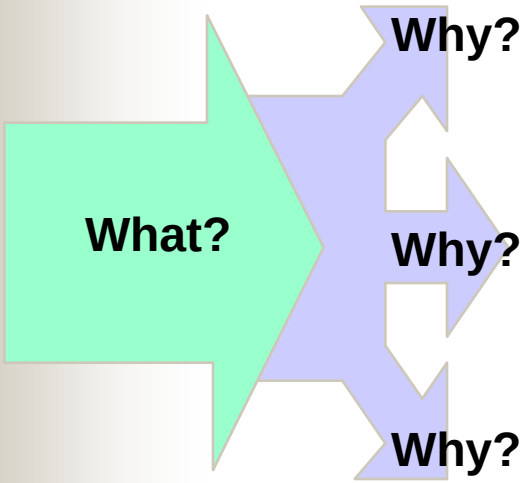
- Example: "GL" Hypercube might contain data from the general ledger and be comprised of multiple dimensions usually found in a general ledger, such as accounts, costs centers, months, and versions. The accounts dimensions might contain base.



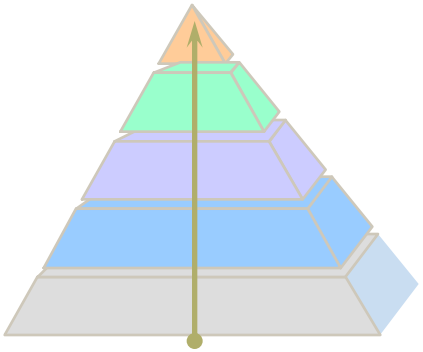
Hypercubes



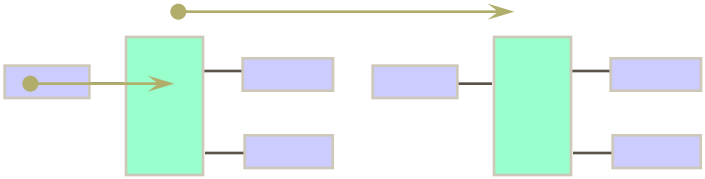
Standard Query Techniques



- Drill-up

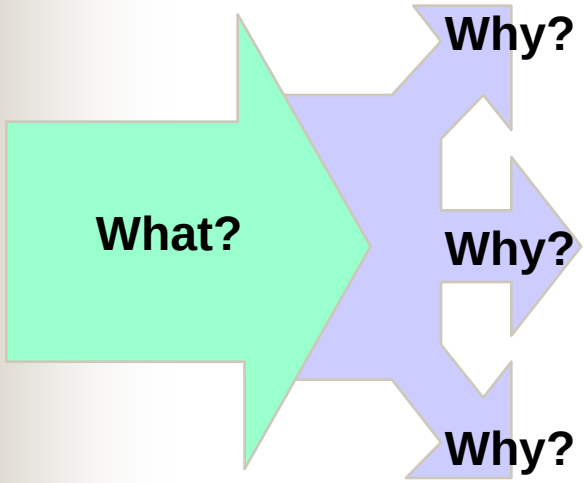


- Drill-across

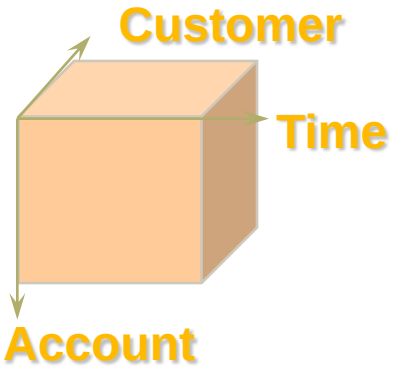


- Pivoting

Standard Query Techniques

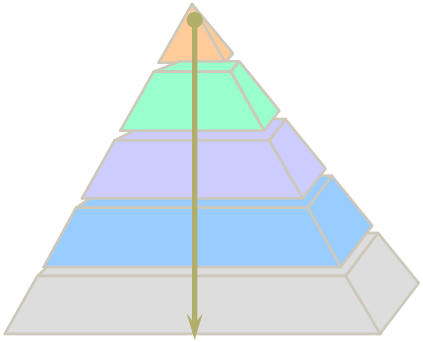


- Slice



- Dice

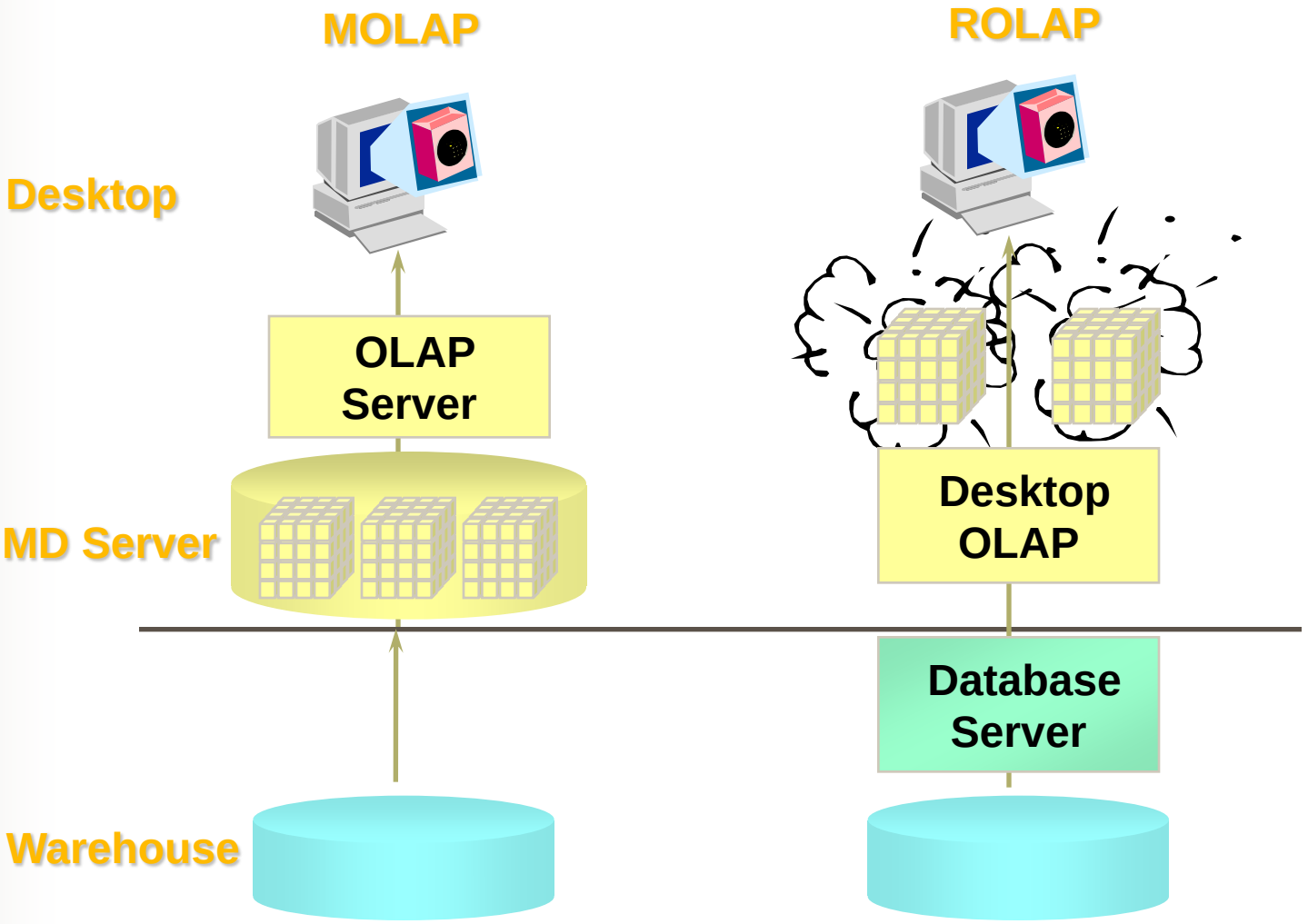
- Drill-down



OLAP Models

- Logical models
  - Cube
  - Star schema (Dimensional Model)
    - Fact Tables & Dimension Tables
  - Snowflake schema
    - A refinement of star schema, with de-normalized dimension tables.
- Physical modeling
  - Multidimensional OLAP (MOLAP)
  - Relational OLAP (ROLAP)

OLAP Physical Models

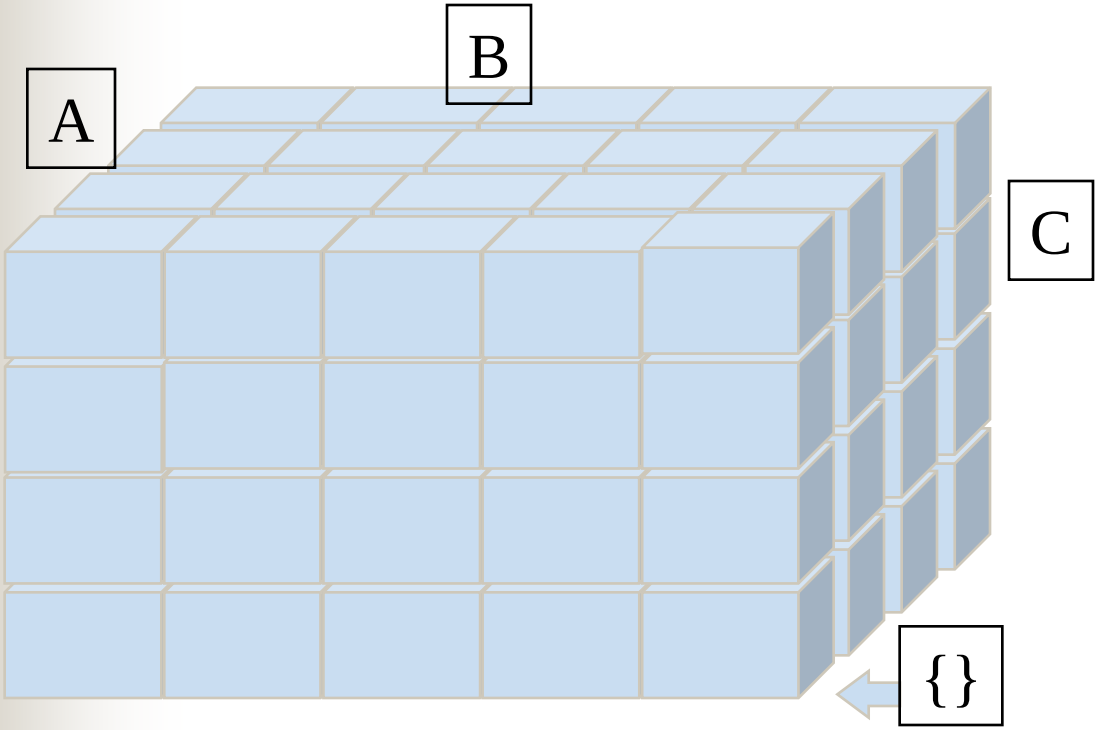


The MOLAP Model

- MDDB: a special-purpose data model
  - Multidimensional database
- Facts stored in multi-dimensional arrays
- Dimensions used to index array
- Sometimes on top of relational DB
- Products
  - Essbase

Cube Computation -- Array Based Algorithm

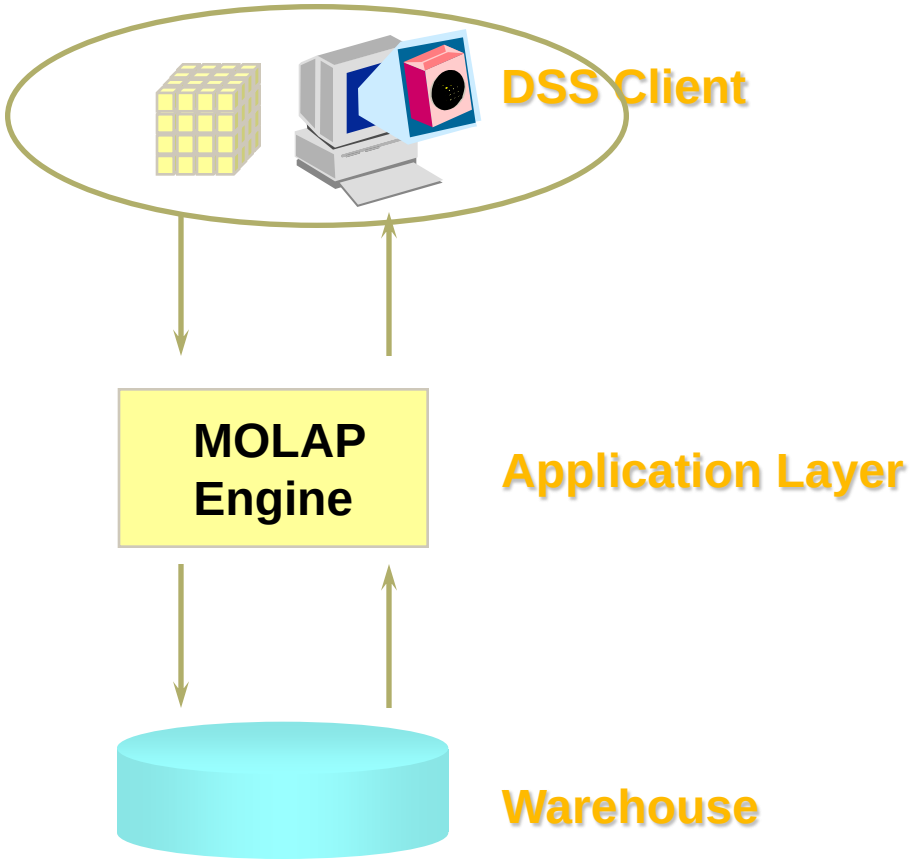
- An MOLAP approach: the base cuboid is stored as multidimensional array.
- Read in a number of cells to compute partial cuboids



{ABC}
{AB}
{AC}
{BC}
{A}
{B}
{C}
{ }

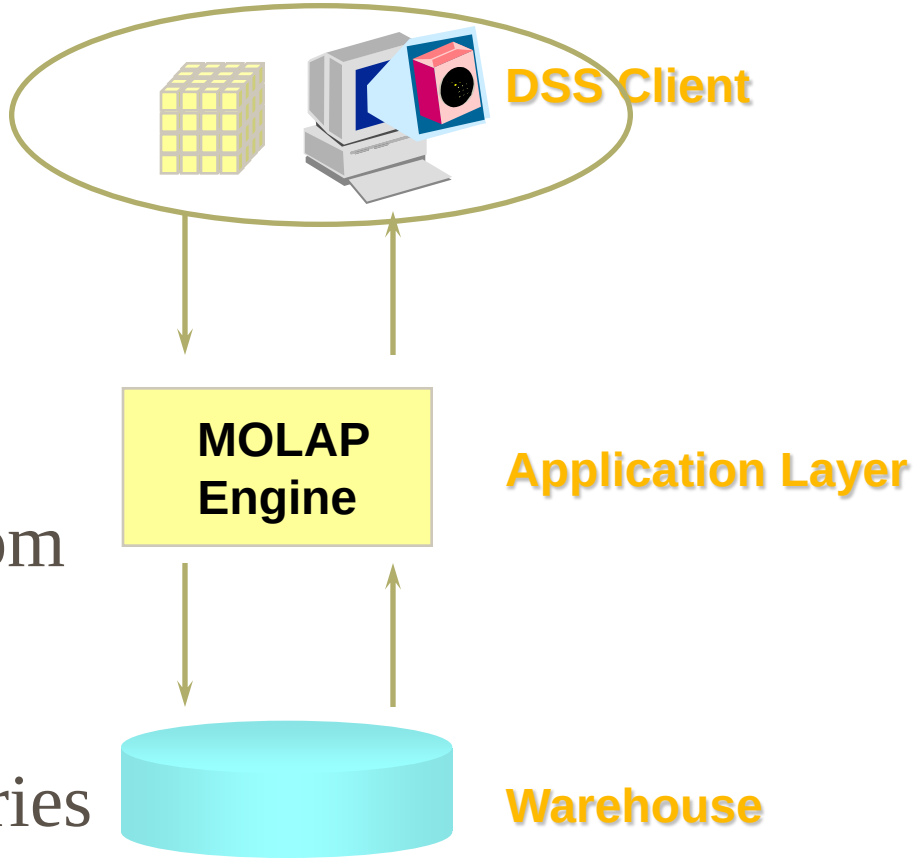
The MOLAP Model

- The application layer stores data in a multidimensional structure.
- The presentation layer provides the multidimensional view.
- The data layer contains the data warehouse tables



MOLAP Model - SUMMARY

- Data
  - Arrays
  - Cached
  - Offloaded from server
- Efficient storage and processing
- Complexity hidden from the user
- Analysis using preaggregated summaries and precalculated measures

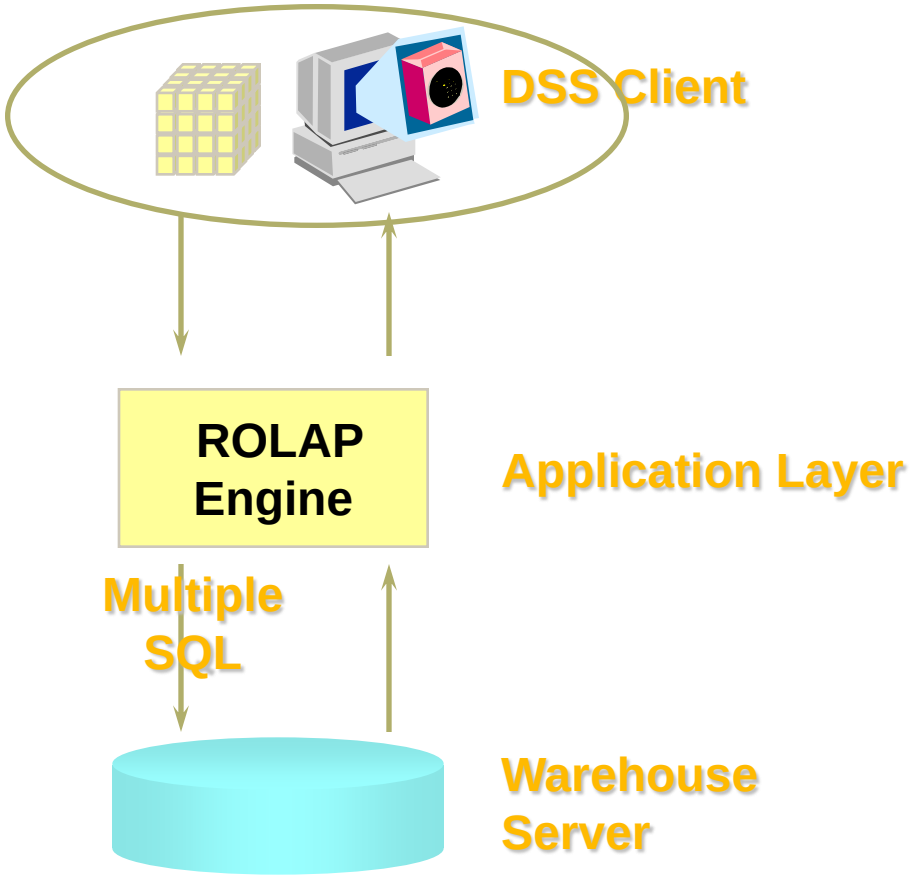


The ROLAP Model

- Special schema design: *star*, *snowflake*
- Special indexes: bitmap, multi-table join
- Dimensional Modelling using Relation DBMS
- Proven technology (relational model, DBMS), tend to outperform specialized MDDB especially on large data sets
- Products
  - IBM DB2, Oracle, Sybase IQ, RedBrick, Informix

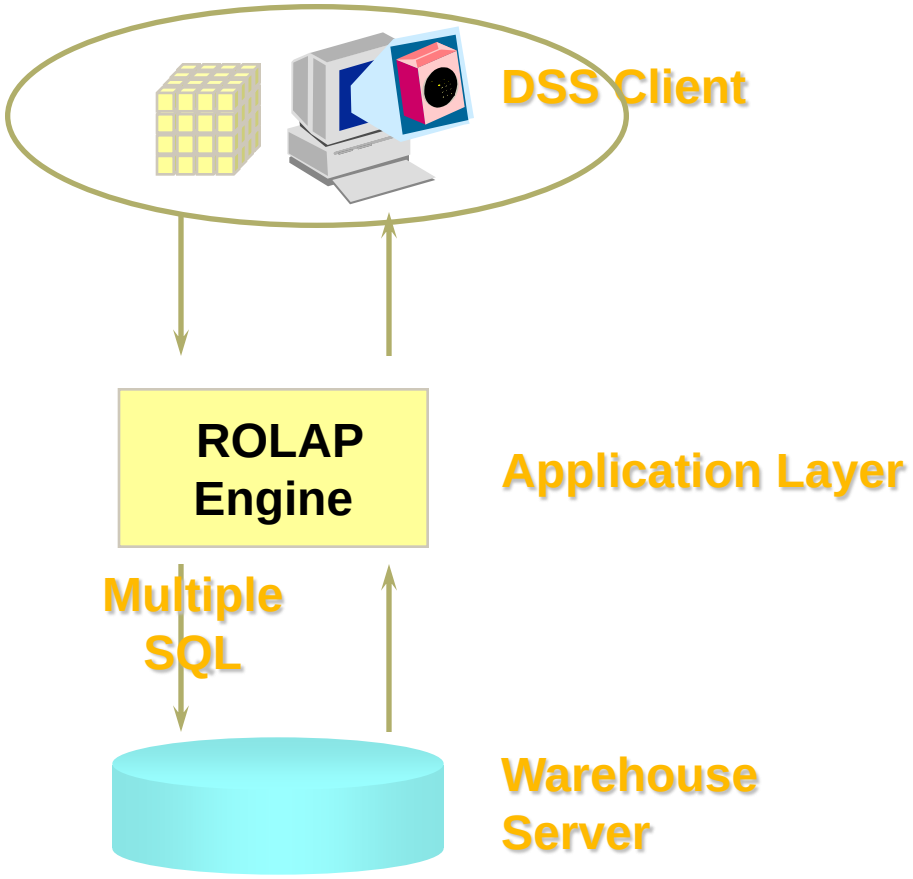
The ROLAP Model

- The warehouse stores atomic data.
- The application layer generates SQL for the three dimensional view.
- The presentation layer provides the multidimensional view.



ROLAP Model - SUMMARY

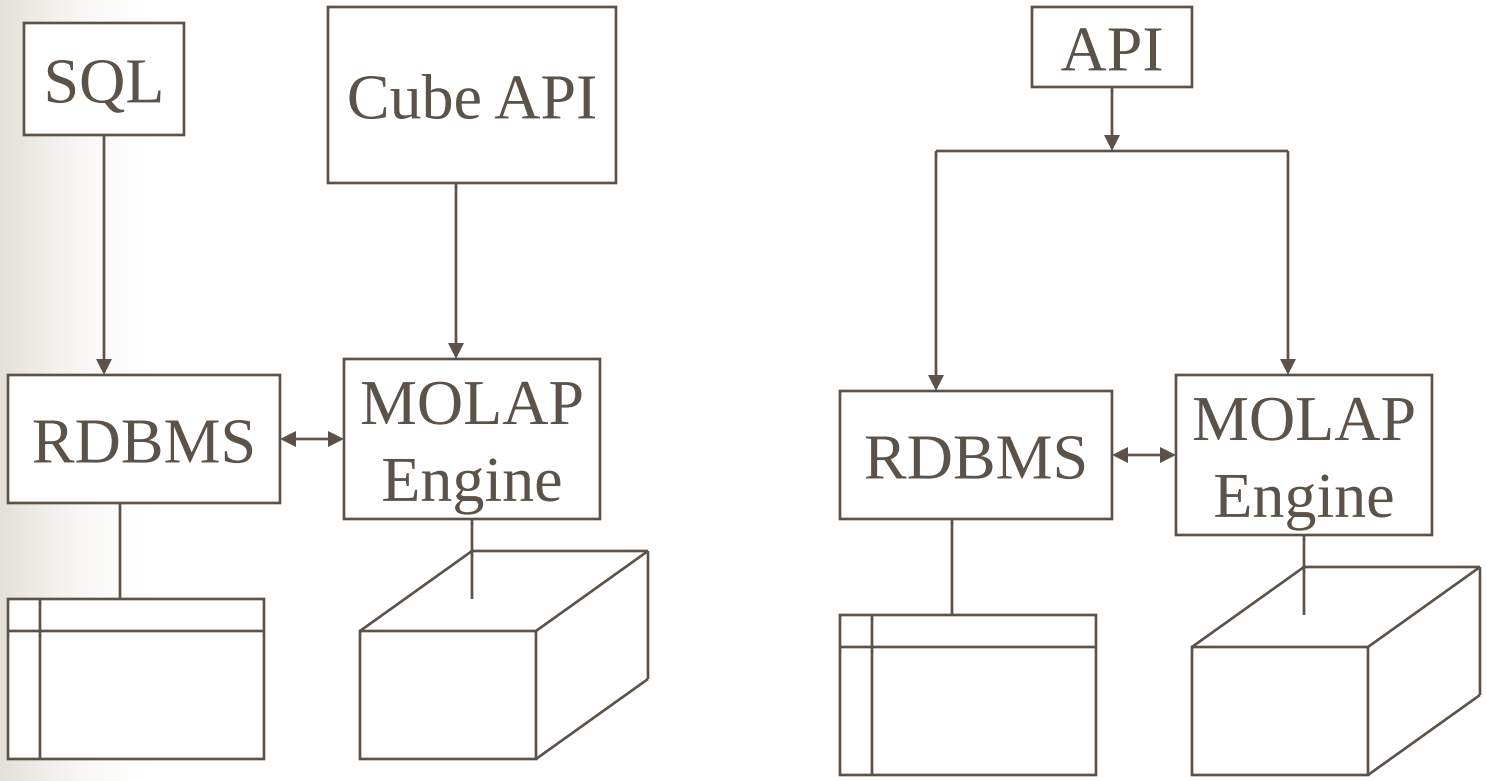
- Data and metadata in Server
- Multidimensional views of data
- High connectivity
- Unlimited
  - Database size
  - Query criteria
- Complex SQL generated by tool



HOLAP (Hybrid OLAP)

- MOLAP does not work well when the user drills down to the dimension tables.
- HOLAP = RDBMS + (MOLAP Engine)
- Dimension tables are stored inside a RDBMS, while the FT is accessible through a MOLAP engine.
- ROLAP engine still exists, but it is integrated with the HOLAP data engine.
  - Users may still use ROLAP if scalability is very important.

Hybrid APIs



ROLAP versus MOLAP

- Figure 15
- ROLAP: Relational On-Line Analytical Processing
- MOLAP: Multi-Dimensional On-Line Analytical Processing