Data Warehousing and Data Mining

Online Analytical Processing (OLAP)

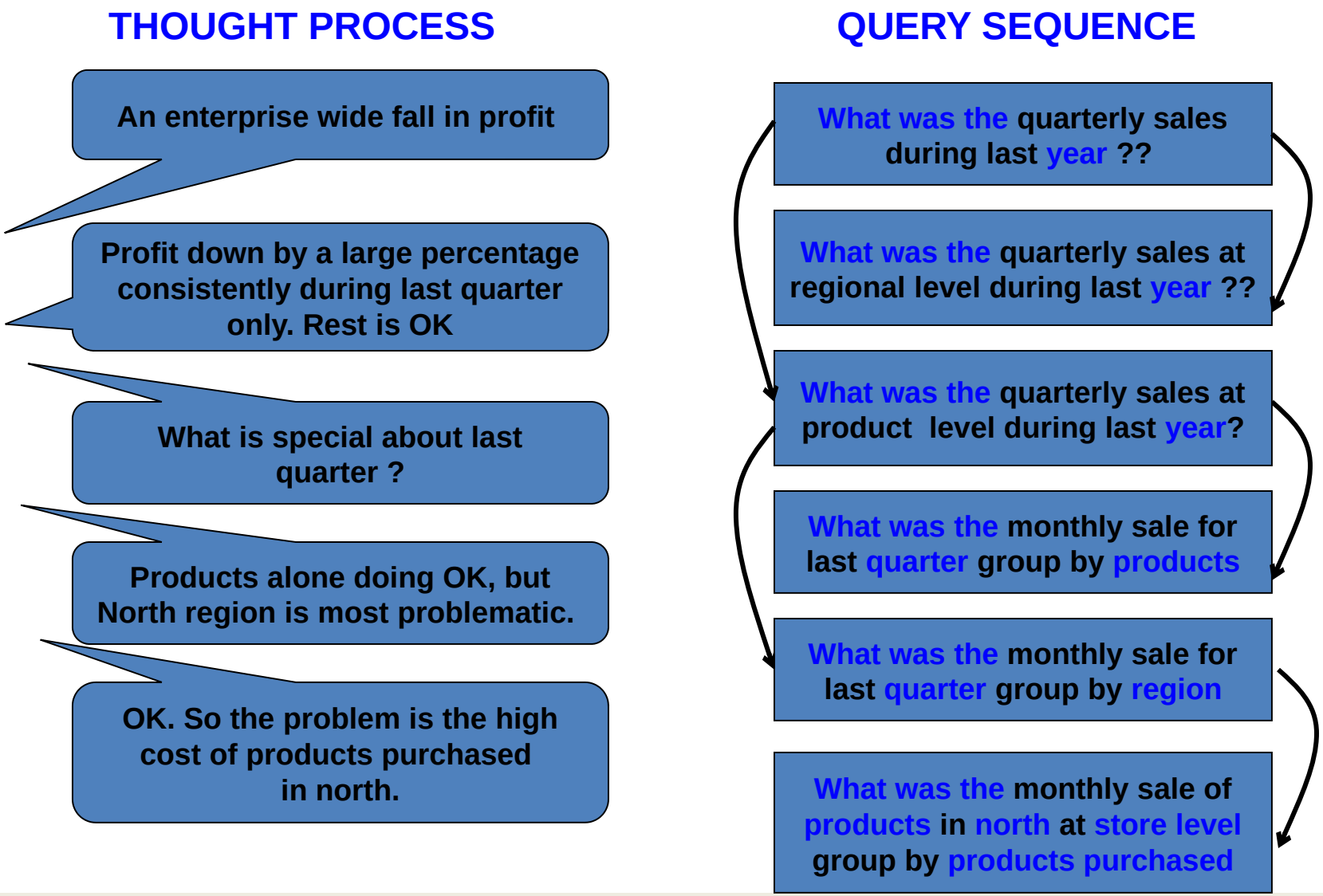
DWH & OLAP

- Relationship between DWH & OLAP
- Data Warehouse & OLAP go together.
- Analysis supported by OLAP

OLAP

- *OLAP* (Online Analytical Processing) is a category of database processing that facilitates business intelligence.
- OLAP tools are optimized for querying and reporting.
- enable users to analyze multidimensional data interactively from multiple perspectives.

Supporting the human thought process



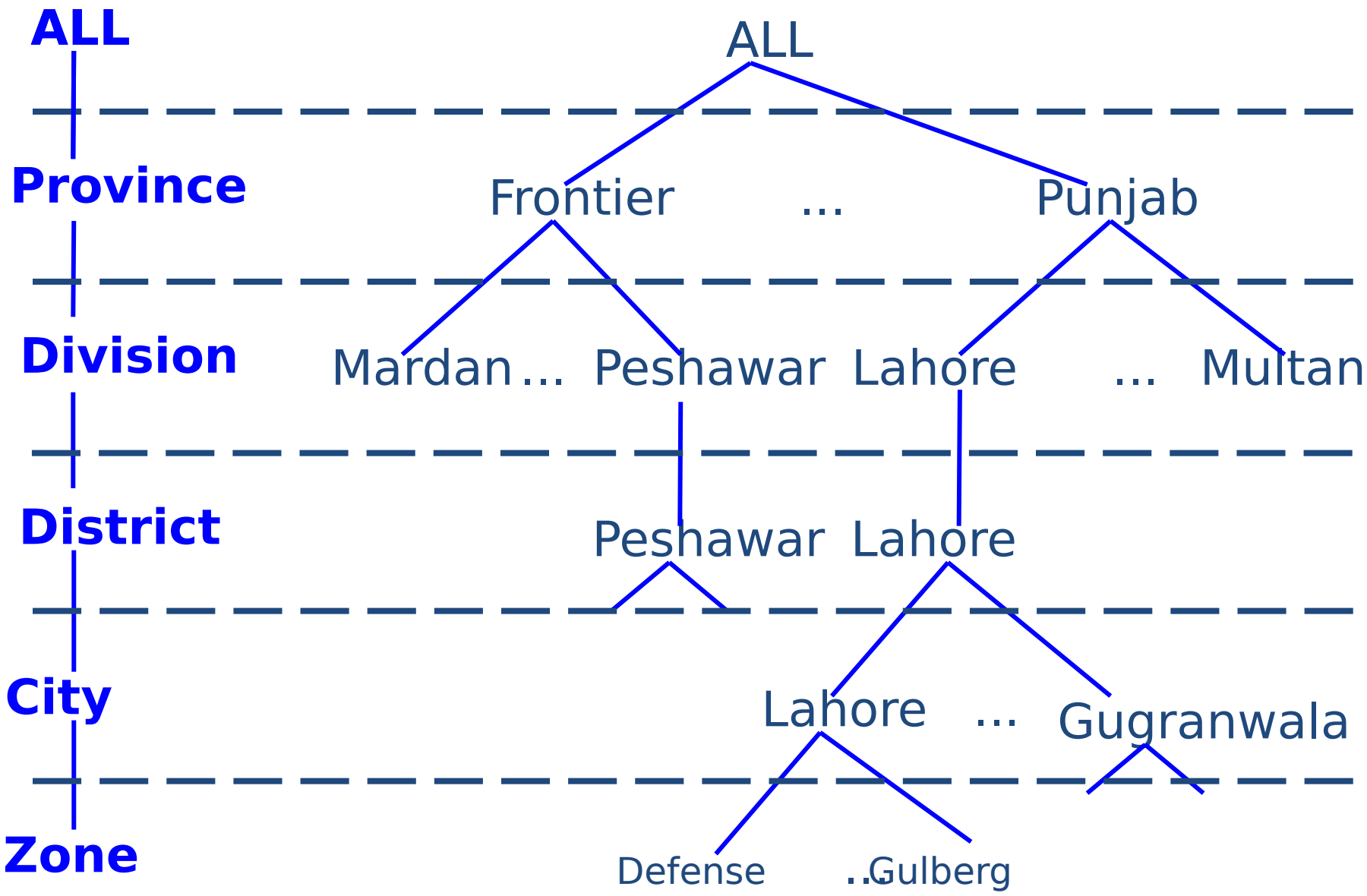
How many such query sequences can be programmed in advance?

Analysis of last example

- Analysis is Ad-hoc
- Analysis is interactive (user driven)
- Analysis is iterative
  - Answer to one question leads to a dozen more
- Analysis is directional
  - Drill Down
  - Roll Up
  - Pivot

More in subsequent slides

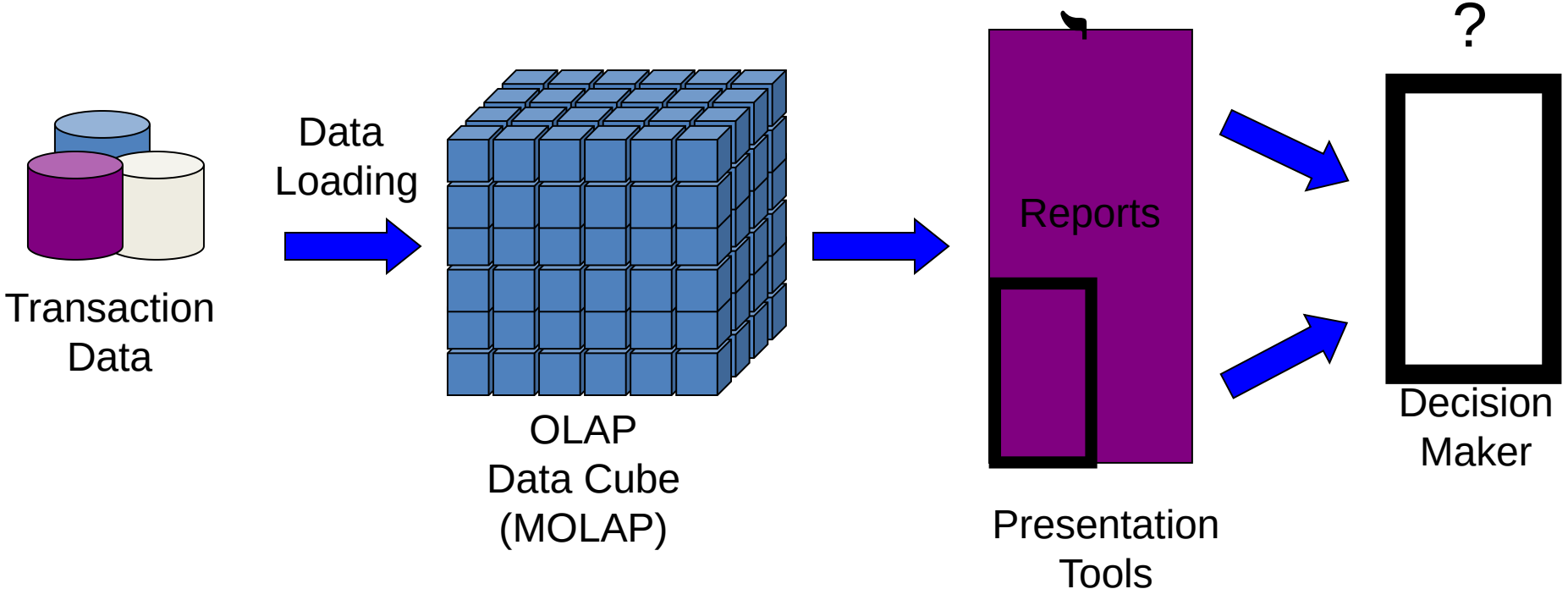
"All" possible queries (level aggregates)



Where Does OLAP Fit In?

- It is a classification of applications, NOT a database design technique.
- Analytical processing uses multi-level aggregates, instead of record level access.
- Objective is to support very
  I. fast
  II. iterative and
  III. ad-hoc decision-making.

Where does OLAP fit in?



OLTP vs. OLAP

Feature	OLTP	OLAP
Level of data	**Detailed**	**Aggregated**
Amount of data per transaction	**Small**	**Large**
Views	**Pre-defined**	**User-defined**
Typical write operation	**Update, insert, delete**	**Bulk insert**
“age” of data	**Current (60-90 days)**	**Historical 5-10 years and also current**
Number of users	**High**	**Low-Med**
Tables	**Flat tables**	**Multi-Dimensional tables**
Database size	**Med (10^9 B – 10^{12} B)**	**High (10^{12} B – 10^{15} B)**
Query Optimizing	**Requires experience**	**Already “optimized”**
Data availability	**High**	**Low-Med**

OLAP FASMI Test

- **Fast Analysis of Shared Multidimensional Information (FASMI) Test.**
- developed to evaluate the OLAP products.

how the user can decide whether a particular tool actually provides the desired OLAP functionalities and what these functionalities are.

OLAP FASMI Test

Fast: Delivers information to the user at a fairly constant rate. Most queries answered in under five seconds.

Analysis: Performs basic numerical and statistical analysis of the data, pre-defined by an application developer or defined ad-hocly by the user.

Shared: Implements the security requirements necessary for sharing potentially confidential data across a large user population.

Multi-dimensional: The essential characteristic of OLAP.

Information: Accesses all the data and information necessary and relevant for the application, wherever it may reside and not limited by volume.

...from the *OLAP Report* by Pendse and Creeth.

A DW integrated with an OLAP can be used to perform these important tasks:

– **information processing**, which is implemented through querying, providing basic statistical analysis, and reporting using tables, charts, and graphs

– **analytical processing**, in which multidimensional analysis of the data by using basic OLAP operations, such as slice and dice, drilling, pivoting, etc. is performed

- data mining, in which new and interesting hidden patterns in the data are found. Some of the DW are furnished with, or can be integrated with, data mining tools that support discovery of data associations, construction of analytical models, performance of classification and prediction, and presentation of the results using visualization tools.

Examples of OLAP Tools

- Pentaho BI
- Mondrian
- IBM Cognos
- Micro Strategy
- icCube