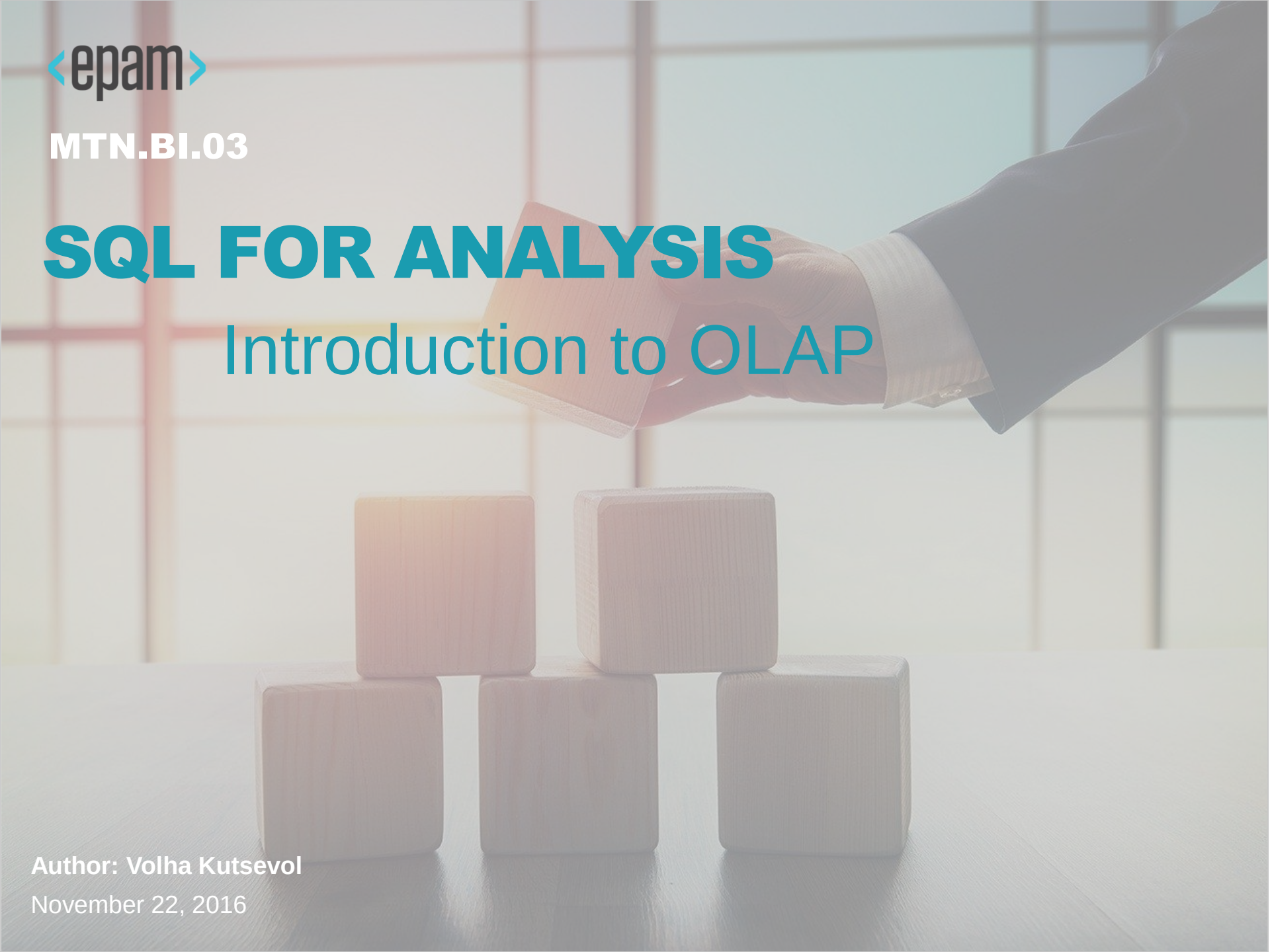


MTN.BI.03

SQL FOR ANALYSIS

Introduction to OLAP

Author: Volha Kutsevol

November 22, 2016

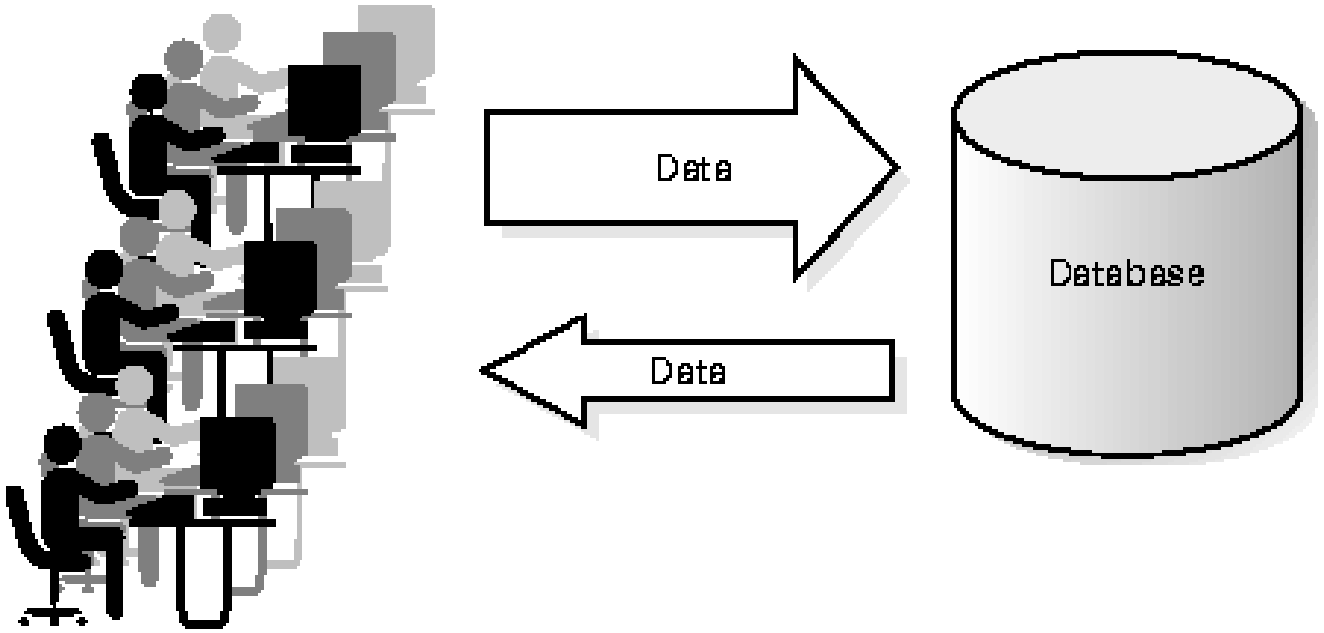
Agenda

1. Online Transaction Processing (OLTP) Systems
2. Online Analytical Processing (OLAP) Systems
3. Data Warehousing
4. Multidimensional Databases
5. OLAP Cubes
6. Relational OLAP Systems
7. Star Schemas

DIFFERENT INFORMATION WORLDS: OLAP AND OLTP

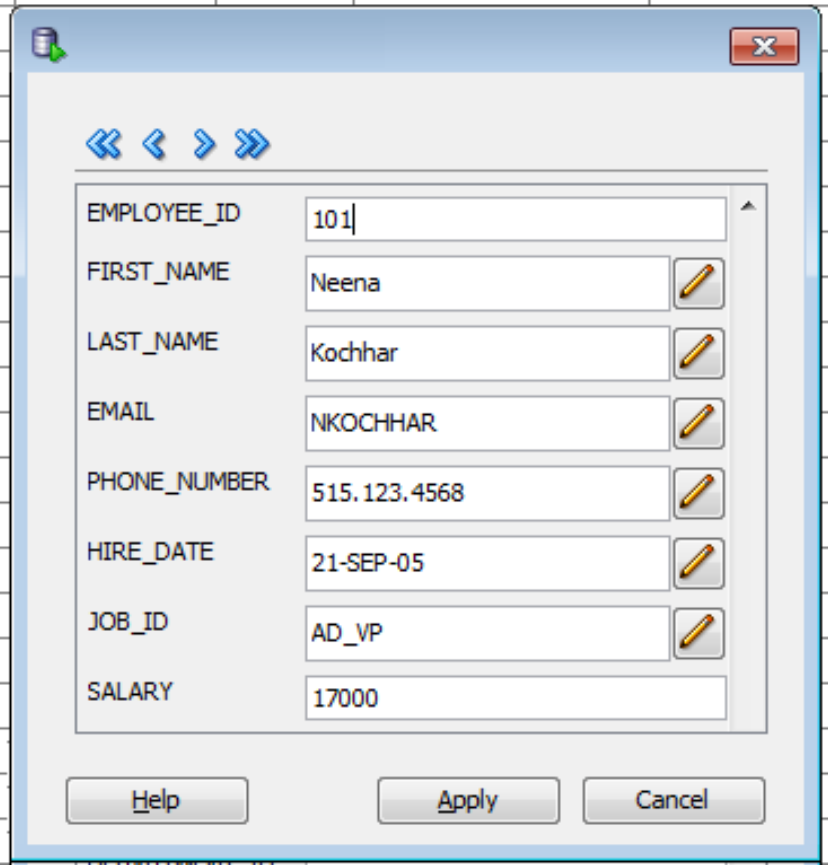
Online Transaction Processing (OLTP) System

OLTP refers to a class of systems that facilitate and manage transaction-oriented applications, typically for data entry and retrieval transaction processing.



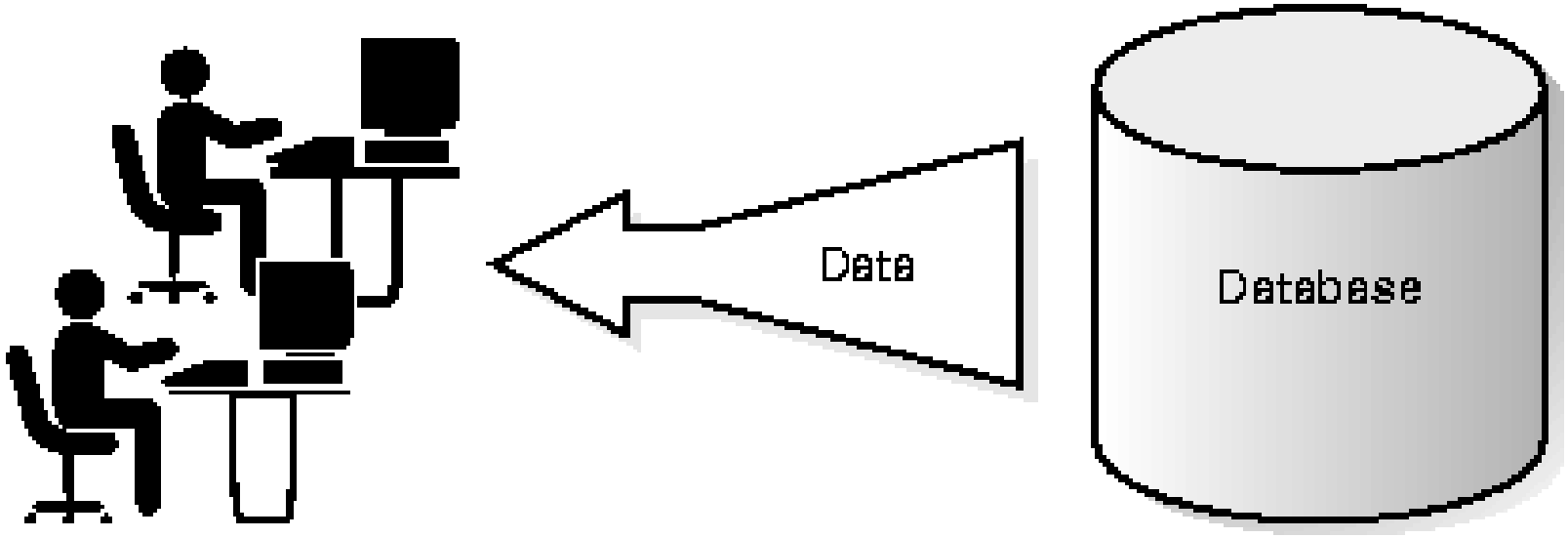
Example: HR Specialist Automated Workplace

	EMPLOYEE_ID	FIRST_NAME	LAST_NAME	EMAIL	PHONE_NUMBER	HIRE_DATE	JOB_ID	SALARY
1	100	Steven	King	SKING	515.123.4567	17-JUN-03	AD_PRES	24000
2	101	Neena	Kochhar	NKOCHHAR	515.123.4568	21-SEP-05	AD_VP	17000
3	102	Lex	De Haan	LDEHAAN	515.123.4569	13-JAN-01	AD_VP	17000
4	103	Alexander					IT_PROG	9000
5	104	Bruce					IT_PROG	6000
6	105	David					IT_PROG	4800
7	106	Valli					IT_PROG	4800
8	107	Diana					IT_PROG	4200
9	108	Nancy					FI_MGR	12008
10	109	Daniel					FI_ACCOUNT	9000
11	110	John					FI_ACCOUNT	8200
12	111	Ismael					FI_ACCOUNT	7700
13	112	Jose Manuel					FI_ACCOUNT	7800
14	113	Luis					FI_ACCOUNT	6900
15	114	Den					PU_MAN	11000
16	115	Alexander					PU_CLERK	3100
17	116	Shelli					PU_CLERK	2900
18	117	Sigal					PU_CLERK	2800
19	118	Guy					PU_CLERK	2600
20	119	Karen					PU_CLERK	2500
21	120	Matthew					ST_MAN	8000
22	121	Adam					ST_MAN	8200

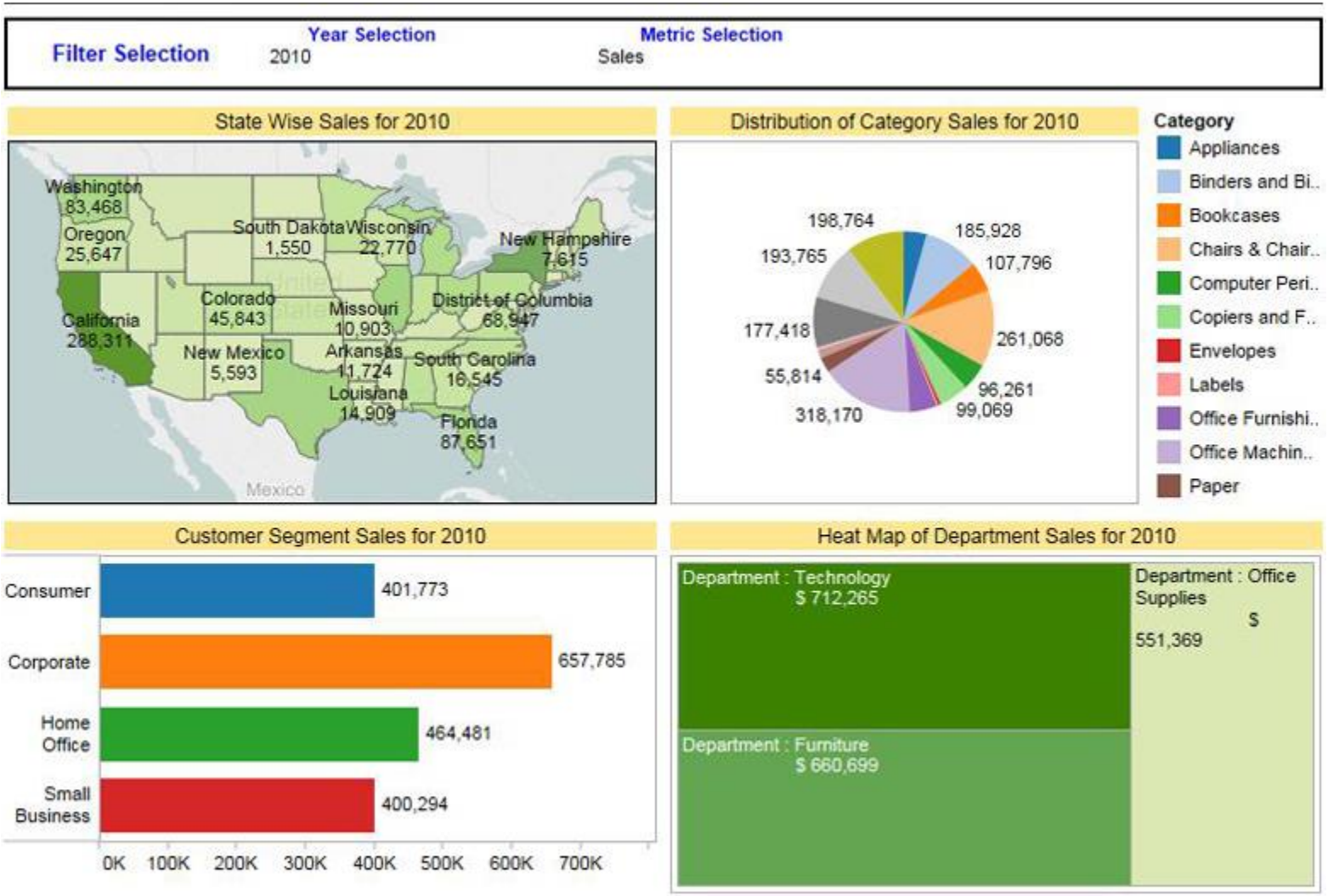


Online Analytical Processing (OLAP) System

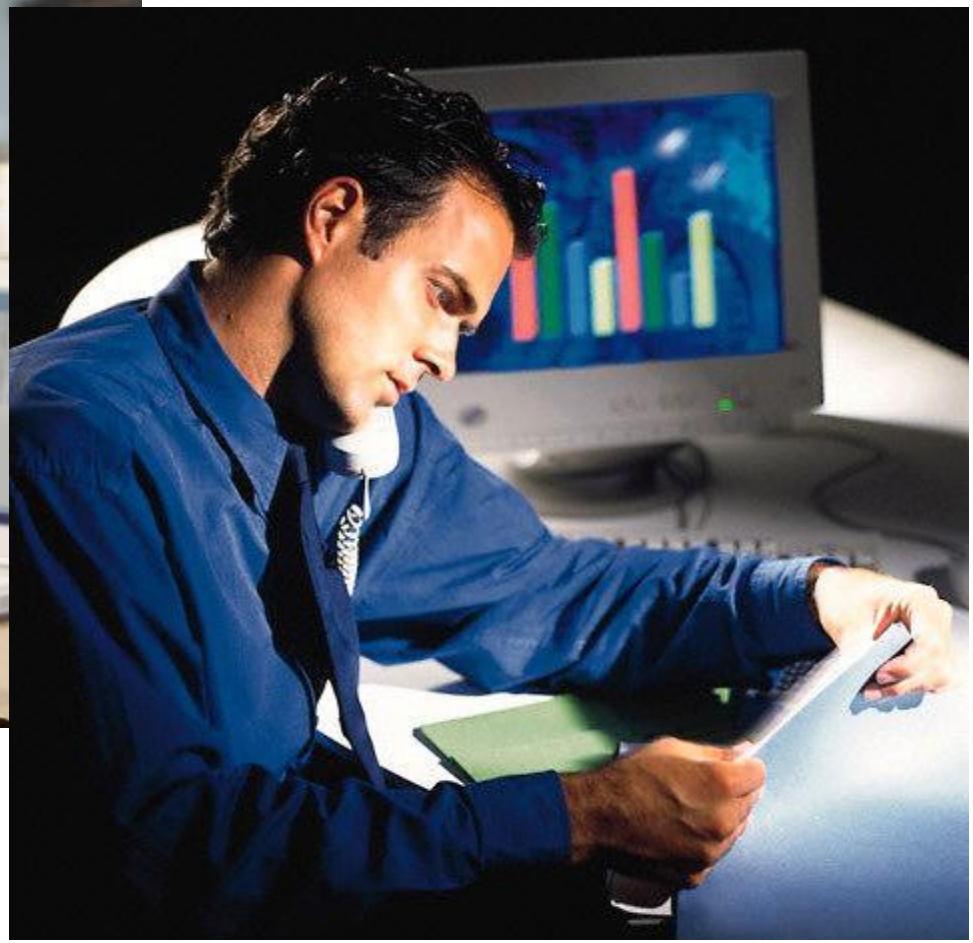
On-Line Analytical Processing (OLAP) is a category of software technology that enables analysts, managers and executives to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the user.



Example: Sales Manager Automated Workplace



OLTP vs. OLAP



OLAP and OLTP Comparison

OLTP

- Insert new object in database
- Update object's attributes
- Delete object
- View list of objects
- Print information about object
- Find particular object

OLAP

- Calculations and modeling applied across dimensions, through hierarchies and/or across members
- Trend analysis over sequential time periods
- Slicing subsets for on-screen viewing
- Drill-down to deeper levels of consolidation
- Reach-through to underlying detail data

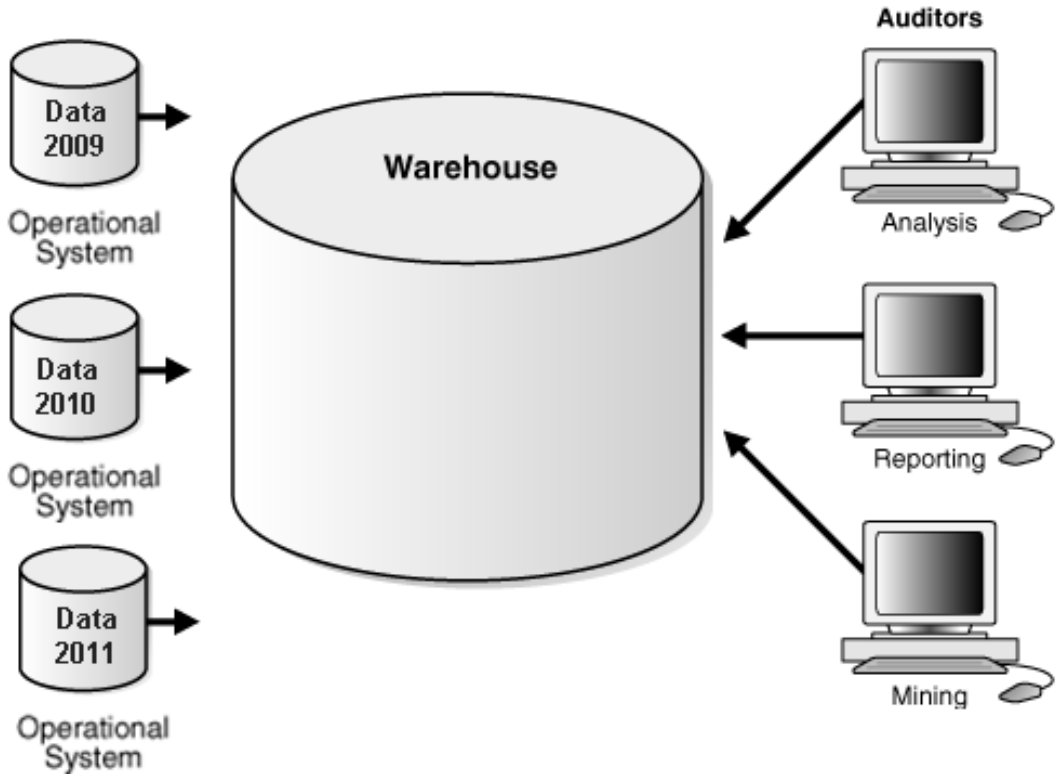
OLAP and OLTP Comparison

	OLTP System Online Transaction Processing (Operational System)	OLAP System Online Analytical Processing (Data Warehouse)
Source of data	Operational data; OLTPs are the original source of the data.	Consolidation data; OLAP data comes from the various OLTP Databases
Purpose of data	To control and run fundamental business tasks	To help with planning, problem solving, and decision support
What the data	Reveals a snapshot of ongoing business processes	Multi-dimensional views of various kinds of business activities
Inserts and Updates	Short and fast inserts and updates initiated by end users	Periodic long-running batch jobs refresh the data
Queries	Relatively standardized and simple queries Returning relatively few records	Often complex queries involving aggregations
Processing Speed	Typically very fast	Depends on the amount of data involved; batch data refreshes and complex queries may take many hours; query speed can be improved by creating indexes
Space Requirements	Can be relatively small if historical data is archived	Larger due to the existence of aggregation structures and history data; requires more indexes than OLTP
Database Design	Highly normalized with many tables	Typically de-normalized with fewer tables; use of star and/or snowflake schemas
Backup and Recovery	Backup religiously; operational data is critical to run the business, data loss is likely to entail significant monetary loss and legal liability	Instead of regular backups, some environments may consider simply reloading the OLTP data as a recovery method

DATA WAREHOUSE

Data Warehouse

A **data warehouse** is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process.



DWH Definition

- **Subject-Oriented:** A data warehouse can be used to analyze a particular subject area.
- **Integrated:** A data warehouse integrates data from multiple data sources.
- **Time-Variant:** Historical data is kept in a data warehouse. This contrasts with a transactions system, where often only the most recent data is kept. For example, a transaction system may hold the most recent address of a customer, where a data warehouse can hold all addresses associated with a customer.
- **Non-volatile:** Once data is in the data warehouse, it will not change. So, historical data in a data warehouse should never be altered.

Goals of a Data Warehouse

- Present the organization’s information consistently
- Make an organization’s information easily accessible
- Serve as the foundation for improved decision making
- Be acceptable by business community (Restructure the data so that it makes sense to the business users)
- Be adaptive and resilient to change
- Restructure the data so that it delivers excellent query performance, even for complex analytic queries, without impacting the operational systems.
- Keep information securely

Different Information Worlds

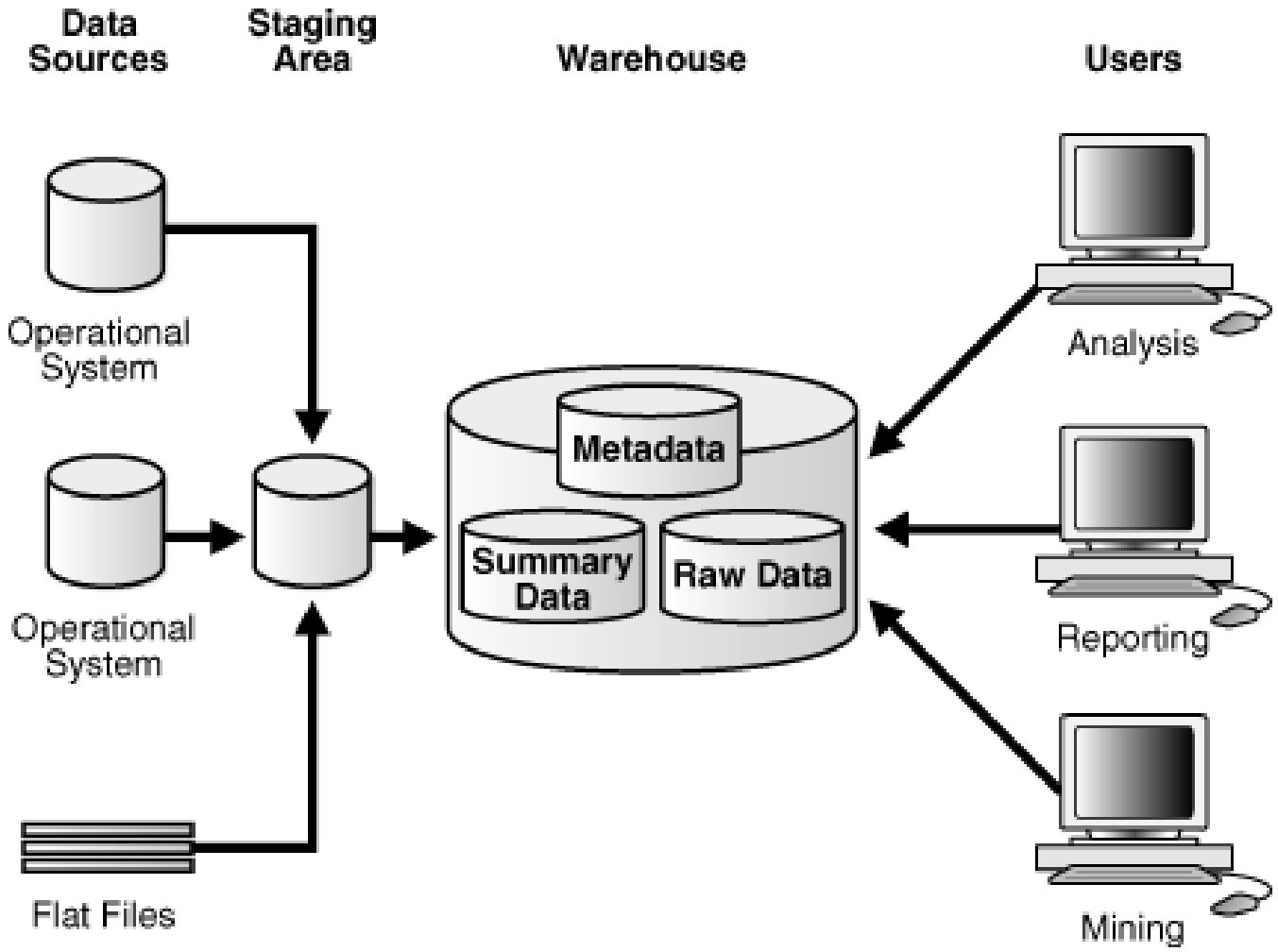
"One of the most important assets of any organization is its information. This asset is kept in two forms: the operational systems of record and the data warehouse.

The users of an operational system turn the wheels of the organization. They take orders, sign up new customers and log complaints. OS users deal with one record at a time. They repeatedly perform the same operational tasks over and over.

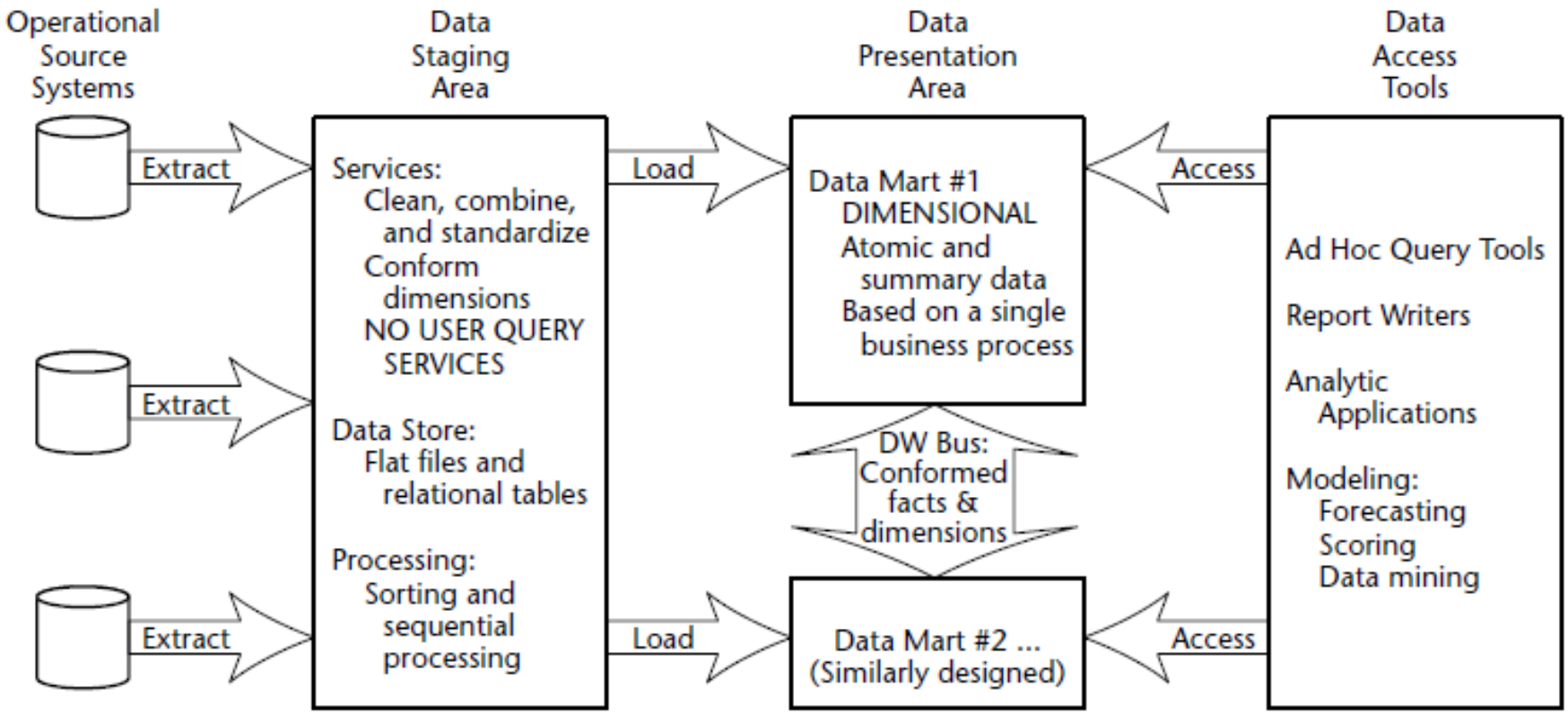
The users of a data warehouse watch the wheels of the organization turn. They count the new orders and compare them with last week's orders and ask why the new customers signed up and what the customers complained about. DWH users never deal with one row at a time. Their questions require thousands of rows be searched and compressed into an answer set…"

Ralph Kimball

Typical Data Warehouse

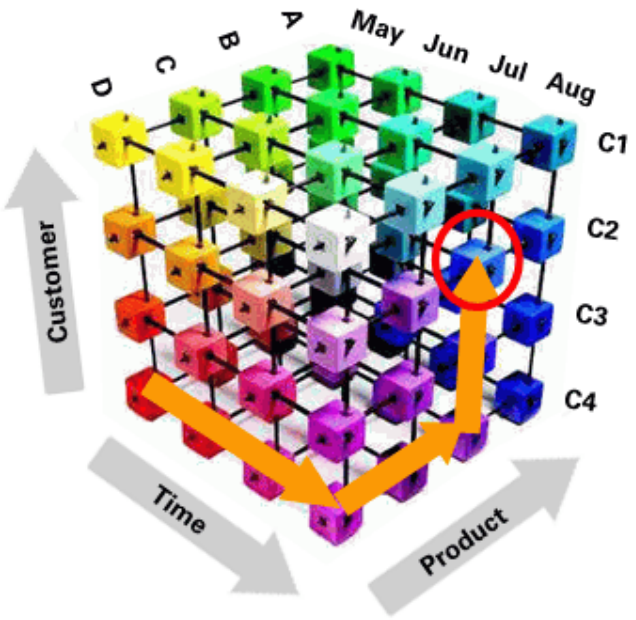


Data Warehouse Components

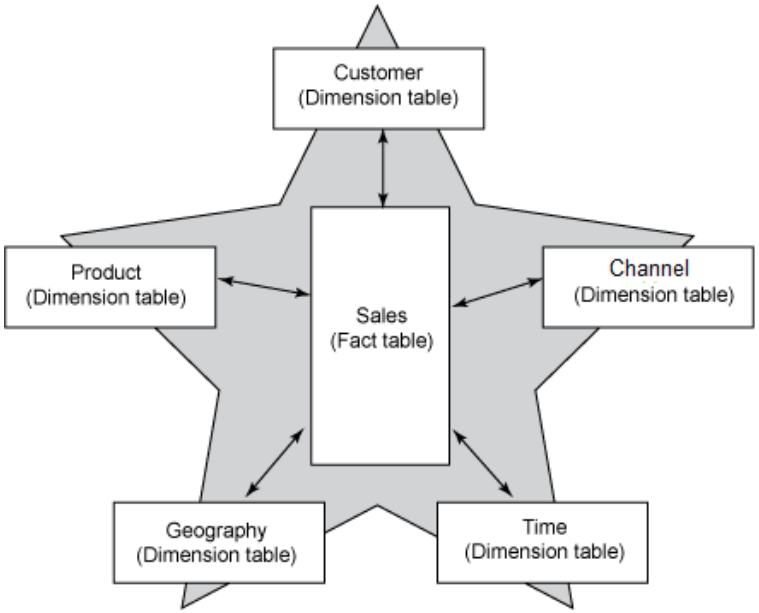


Data Warehouse Organization

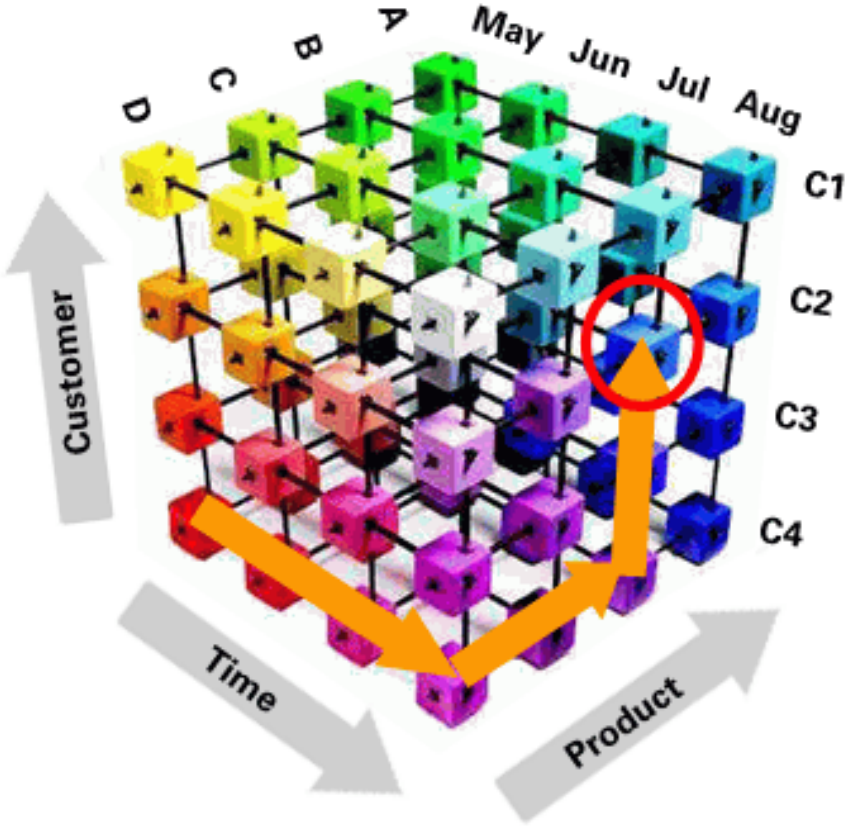
Multidimensional Database



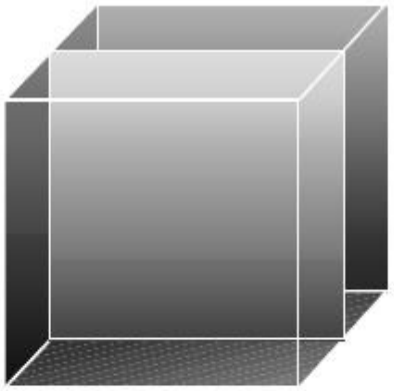
Relational Database



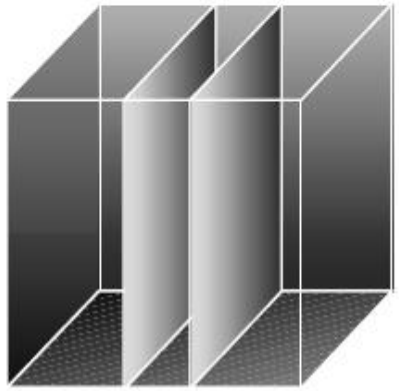
OLAP Cube



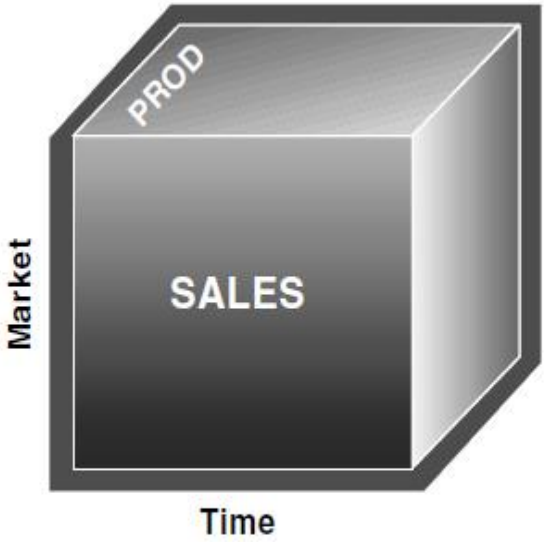
OLAP Cube



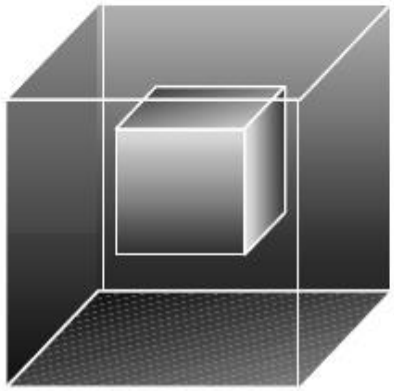
Product Mgr. View



Financial Mgr. View

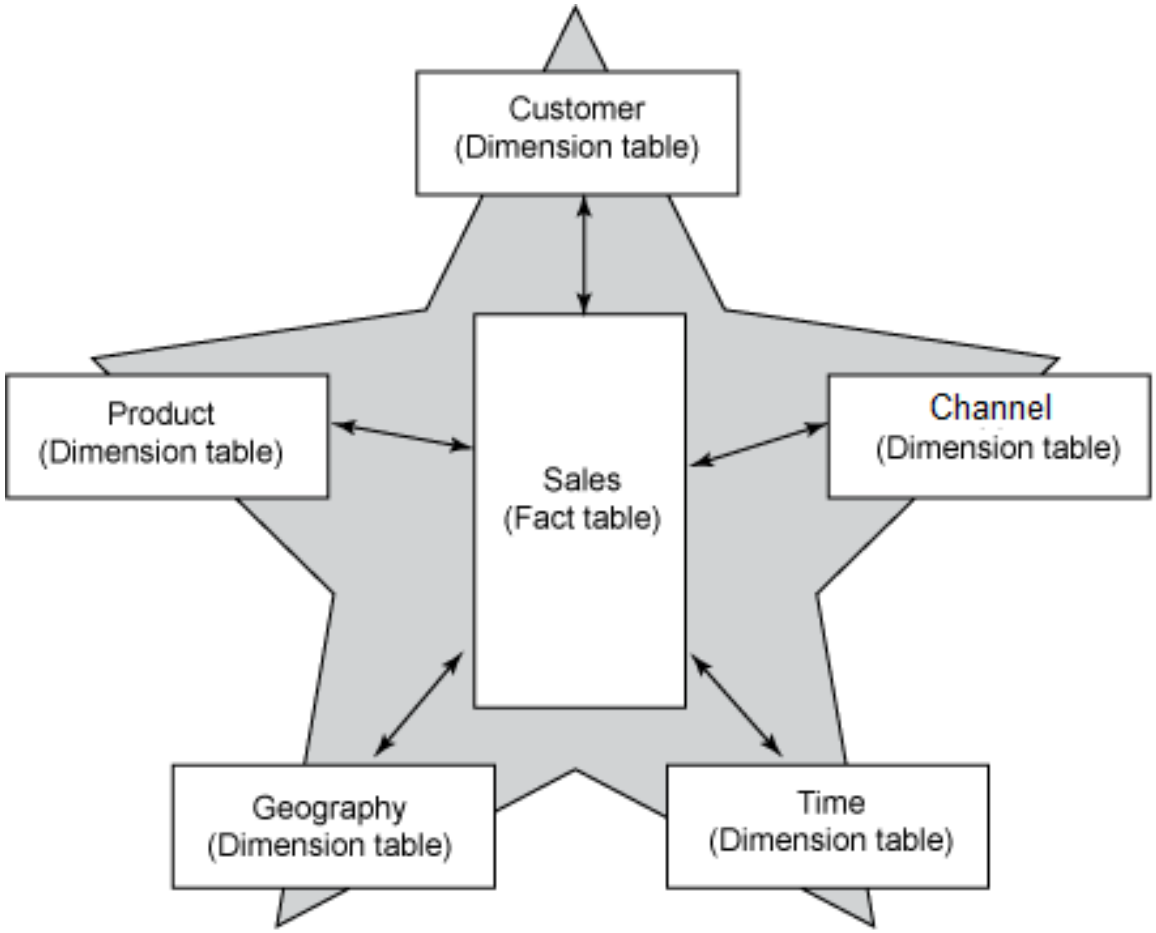


Regional Mgr. View



Ad Hoc View

Star Schema



Facts

In a star schema, the central table which contains the individual facts being stored in the database.

There are two types of fields in a fact table:

- The fields storing the foreign keys which connect each particular fact to the appropriate value in each dimension.
- The fields storing the individual facts (or measures) - such as number, amount, or price.

Dimensions

In a star schema, a table which contains the data for one of the cube's dimensions.

Dimension Table:

- has a primary key which is used to connect it to the fact table
- has one field for each level of each hierarchy contained in the dimension
- has as many attribute fields as possible
- dimension tables in a star schema are intentionally de-normalized

REPORTING

Reporting

Reporting means collecting and presenting data so that it can be analyzed.

- Strictly defined

  Strictly defined (or managed) reporting is reporting prepared by technical personnel such as developers.

- Ad-hoc

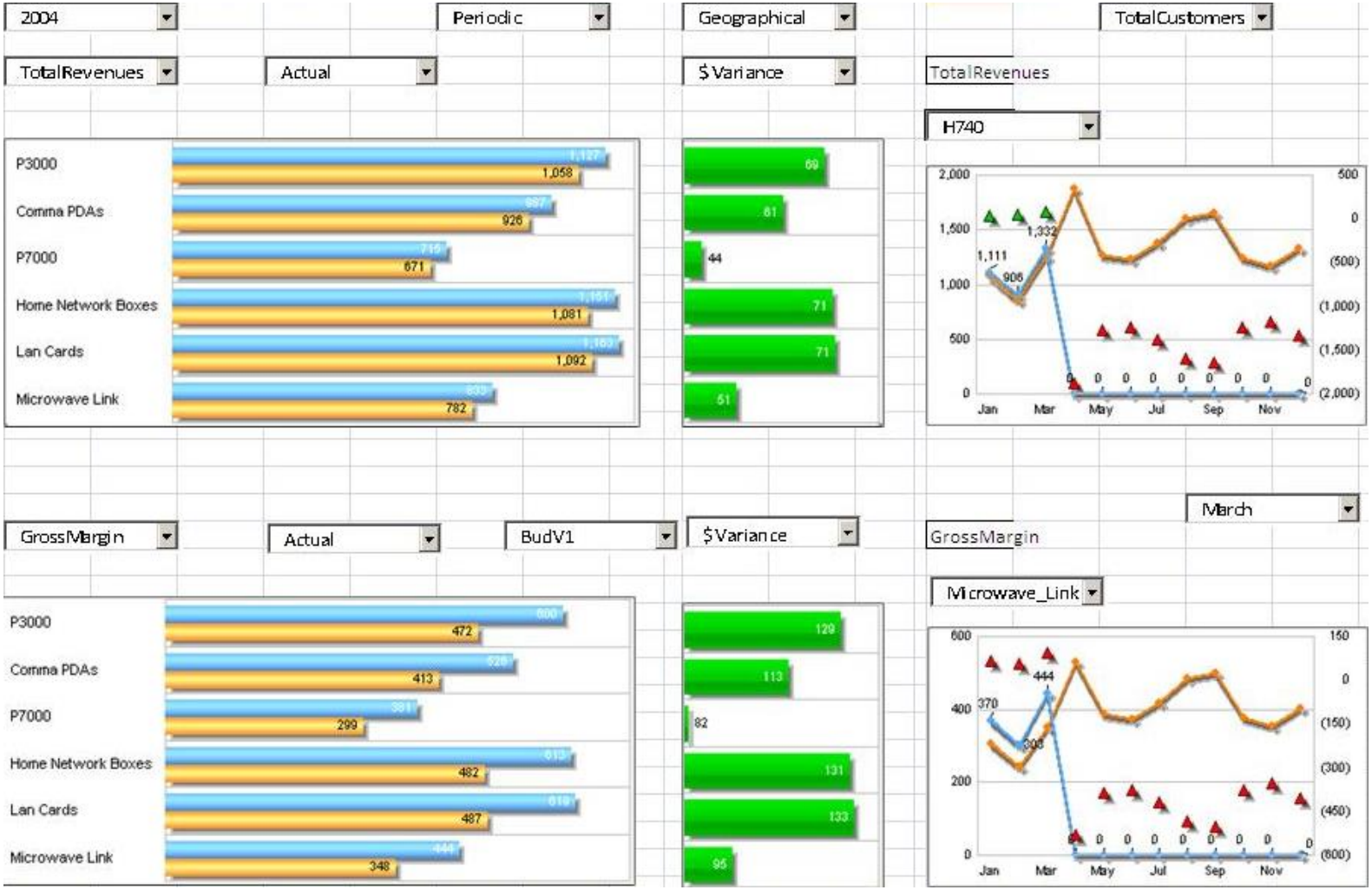
  Ad-hoc reporting is a model of business intelligence in which reports are built and distributed by nontechnical end-users. With ad-hoc reporting, all the technical user does is set up the BI solution, connect it to the data-sources, establish security parameters and determine which objects end-users can see. From that point on, the actual reports are created by business end-users.

Print-ready Report Example

Out Job #1

Job Name	Job Out #1	**Date of Claim**	05/01/2010	**Company**	Sea Pros Yachts
Quotation Number	55	**Approval**	Given	**Warranty**	2 Years
Client Name	John Doe	**Client Telephone**	03001122	**Boat Name**	CHARAB
Boat Model	Mochi #34	**Boat Hull#**	001231113455	**Boat Location**	A.T.C.L.
Engine Hours (Port)	200	**Engine Hours (Stdby)**	400		
Client Claim	There is a leak.				
Technician Report	There is no leak.				
Comment	No comment.				
Job Status	Pending	**Operator**	N/A	**Currency**	Euro (1 Dollar = 0.71 Euro)

Ad-hoc Report Example



ORACLE SALES HISTORY SAMPLE SCHEMA

Sales History Sample Schema

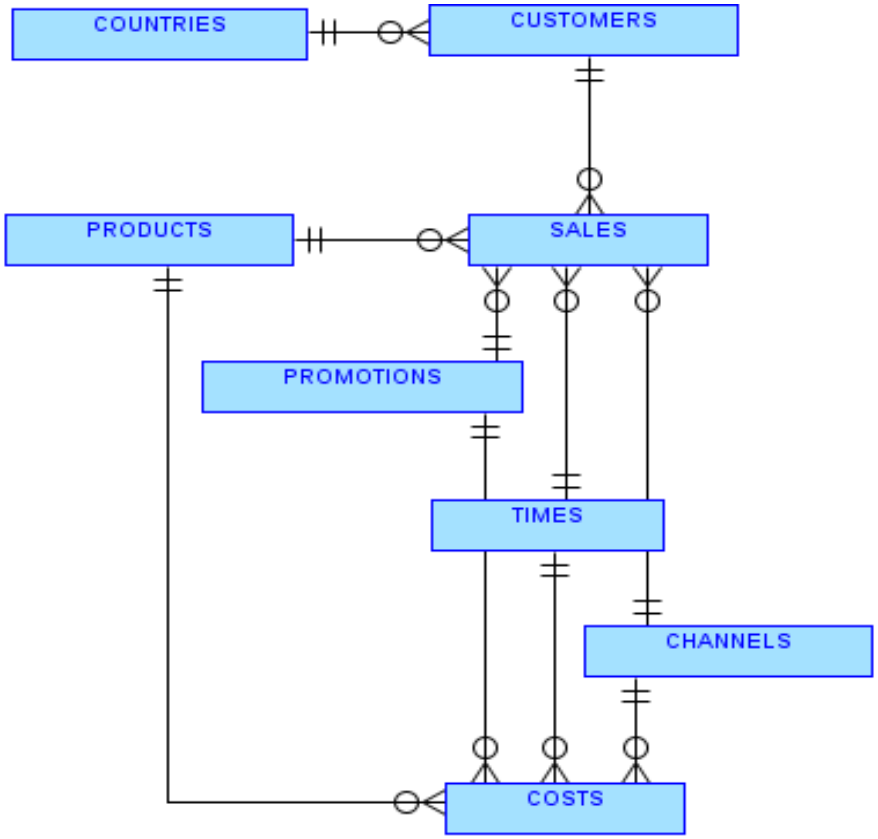
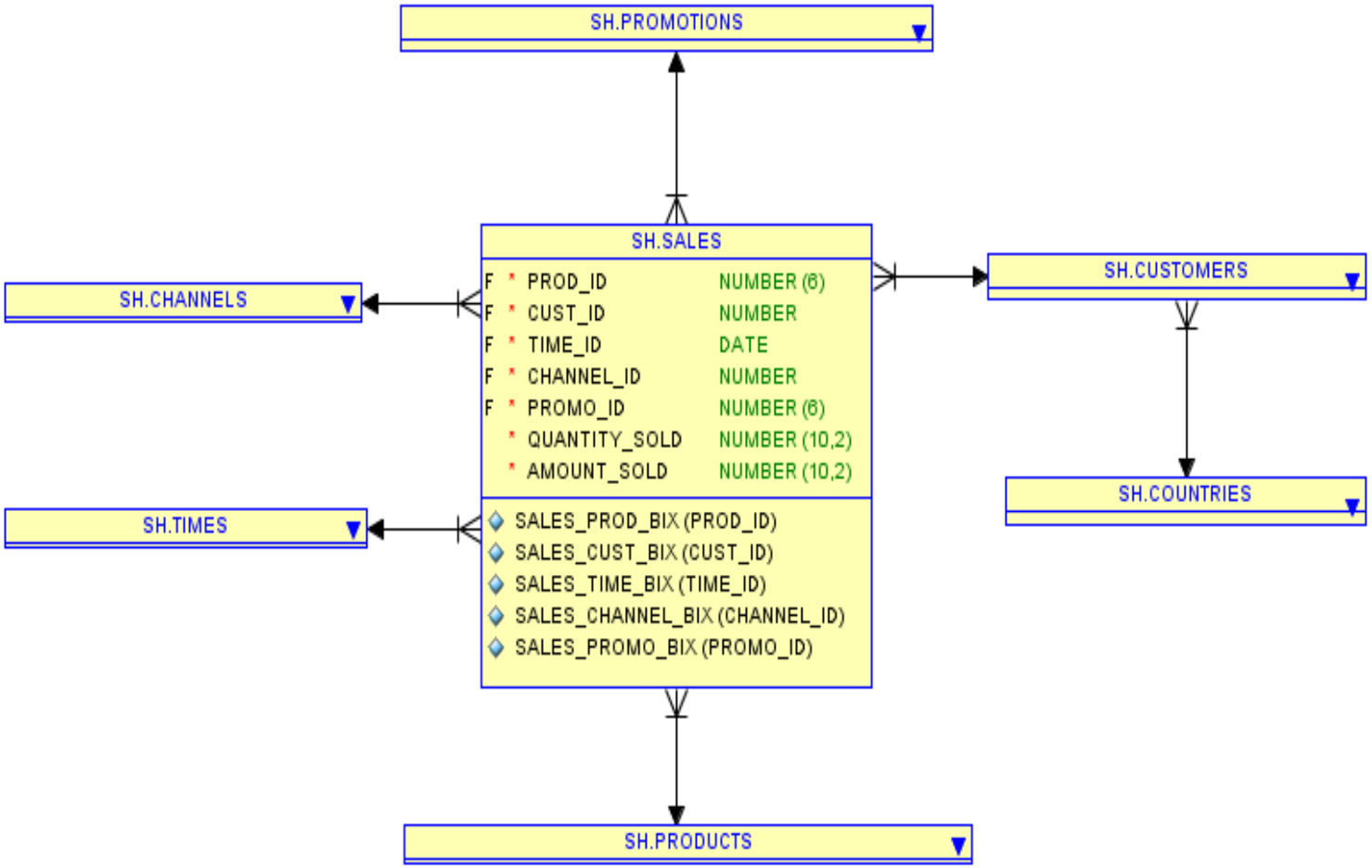
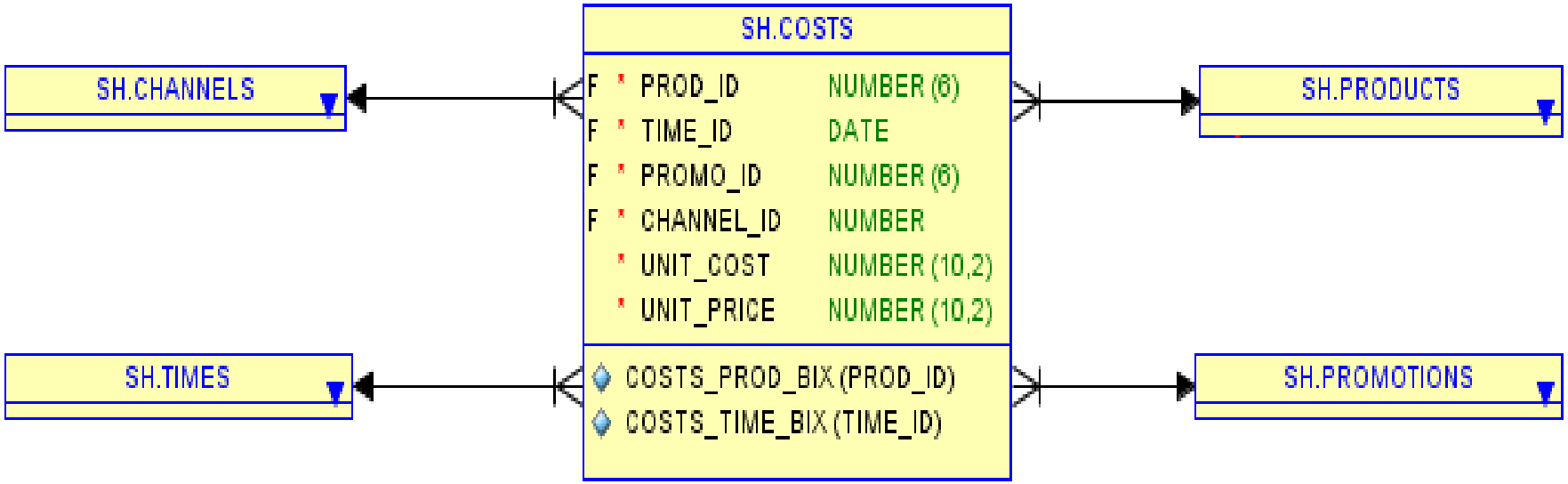


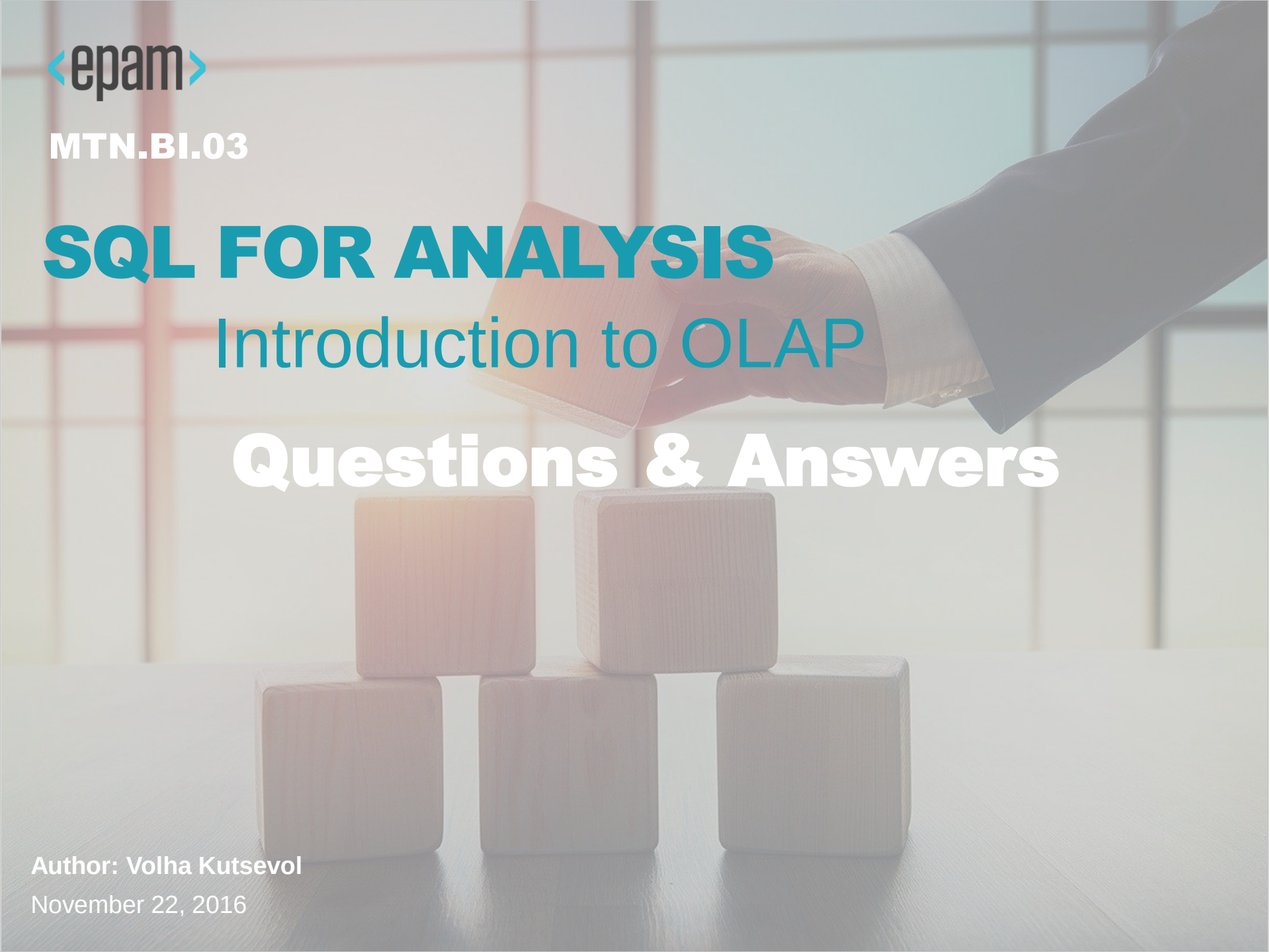
Table	Records count
COUNTRIES	23
CUSTOMERS	55500
CHANNELS	5
TIMES	1826
PRODUCTS	72
PROMOTIONS	503
COSTS	**82112**
SALES	**918843**

Sales Star



Costs Star





MTN.BI.03

SQL FOR ANALYSIS

Introduction to OLAP

Questions & Answers

Author: Volha Kutsevol

November 22, 2016