Knowledge and Intelligent Technology for Business Processes Optimization (SAP)

JASS 2008

Student: Constantine Sonkin

Professor: Dr. Viacheslav P. Shkodirev
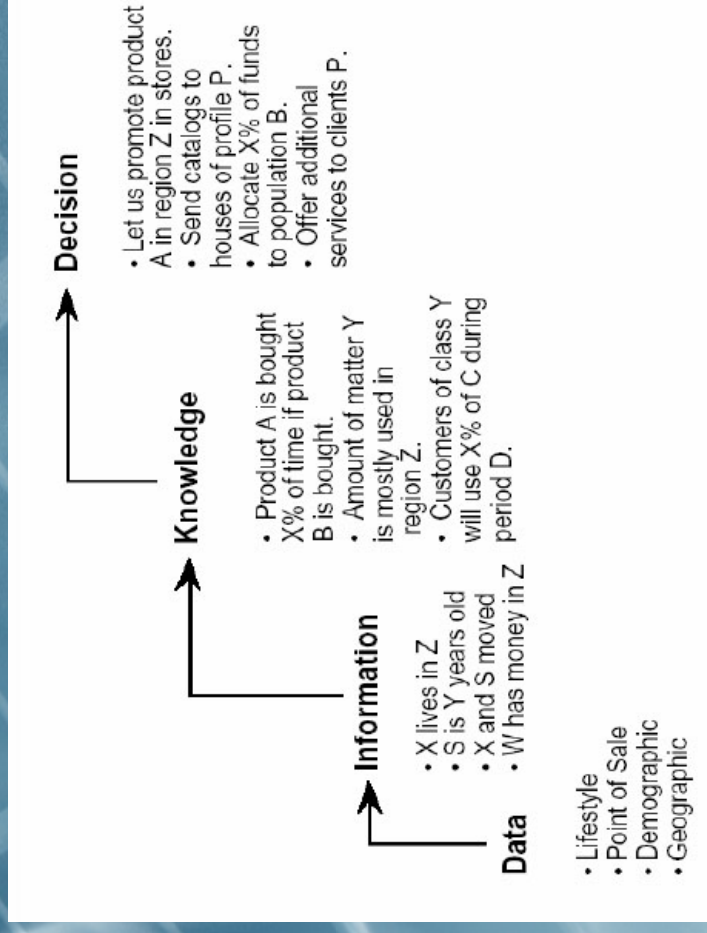
Saint-Petersburg State Polytechnical University

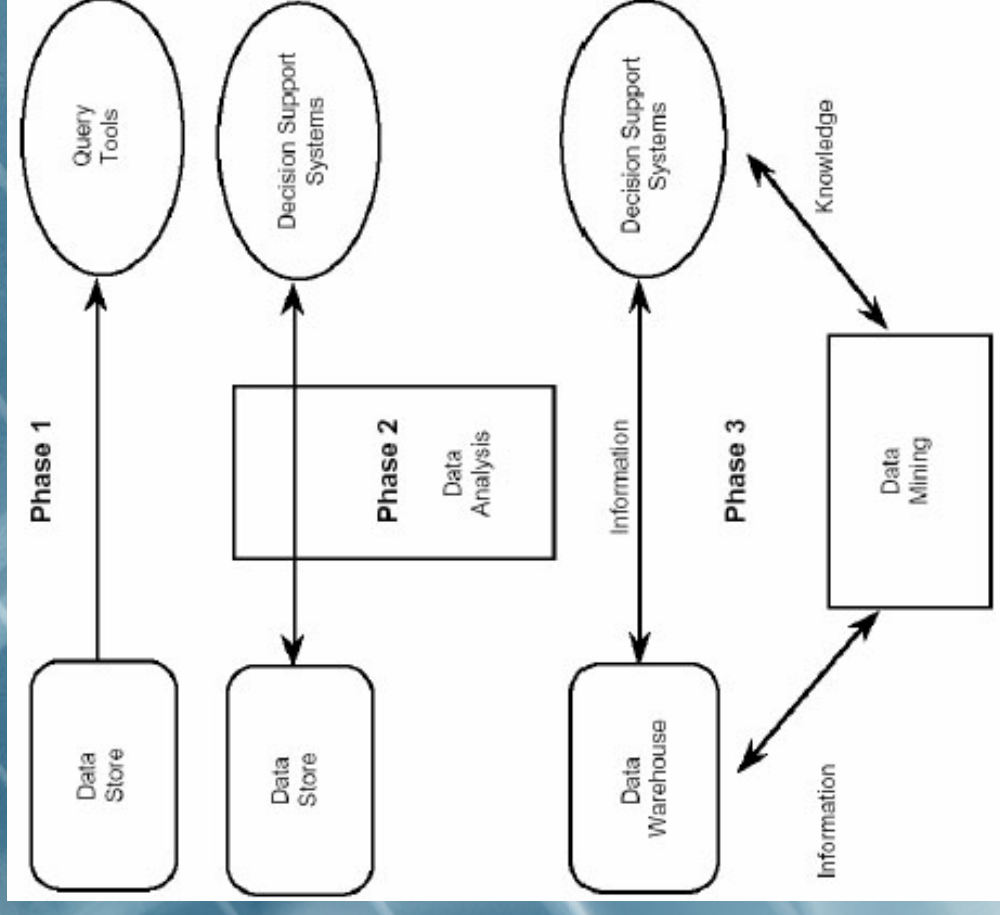# Contents

Introduction

- Main goal of <u>advanced data analysis</u> is to make <u>decisions</u> that lead directly <u>to benefits</u>

- Process from data to decisions involve:
  - database technology
  - data mining
  - business intelligence

Decision

- Let us promote product A in region Z in stores.
- Send catalogs to houses of profile P.
- Allocate X% of funds to population B.
- Offer additional services to clients P.

Knowledge

- Product A is bought X% of time if product B is bought.
- Amount of matter Y is mostly used in region Z.
- Customers of class Y will use X% of C during period D.

Information

- X lives in Z
- S is Y years old
- X and S moved
- W has money in Z

Data

- Lifestyle
- Point of Sale
- Demographic
- Geographic

Introduction.
From data to decisions

- **Phase 1**
  - structured data stores filing relevant data;
  - decision support tasks were performed centrally

- **Phase 2**
  - decision support implemented offline
  - data analysis techniques

- **Phase 3**
  - diversity of the data stores;
  - warehouse frameworks;
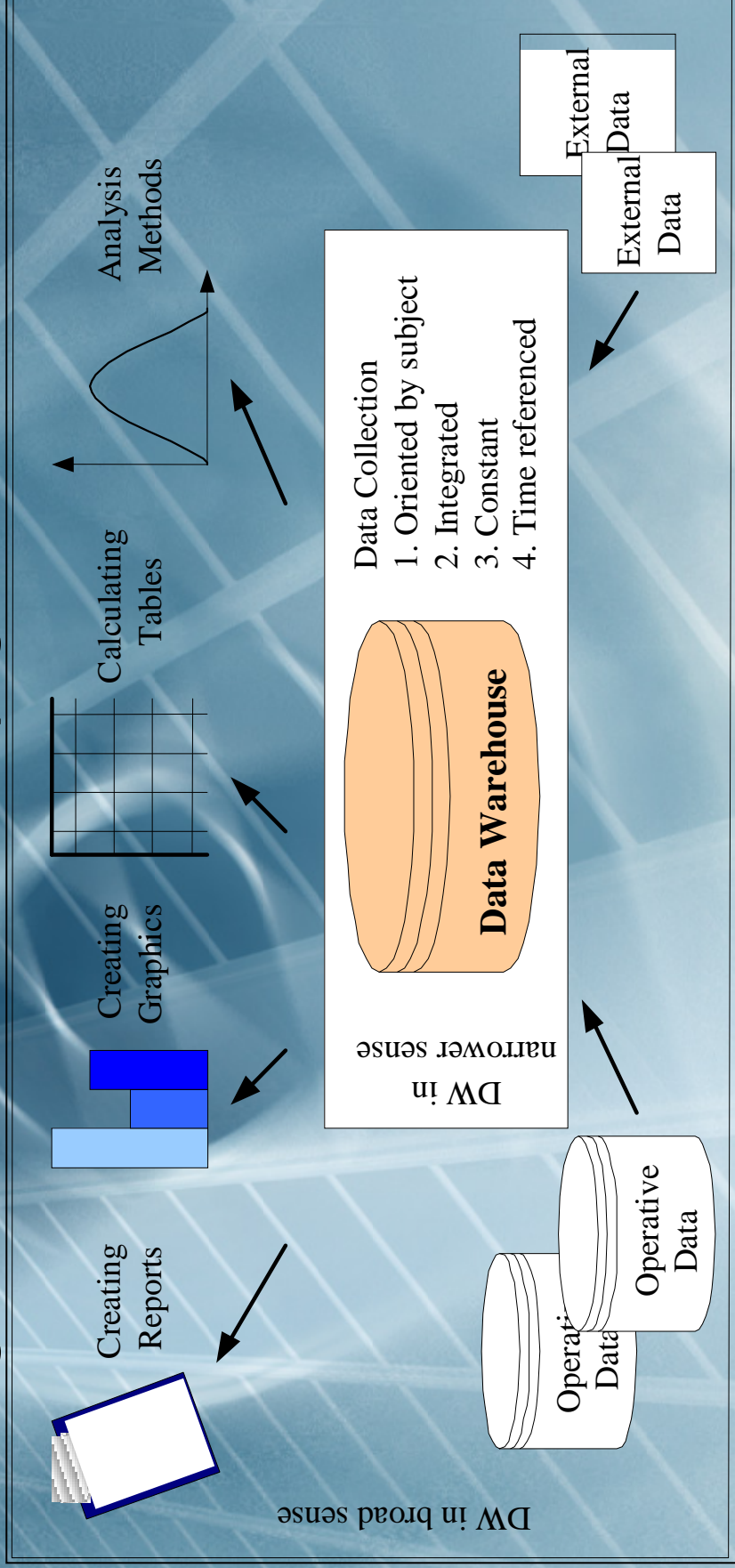 - advanced data analysis techniques

Introduction. Definitions

- **Data Warehouse:** is a process and architecture that requires robust planning to implement a platform for decision-making processes and BI.

  Includes: selection, conversion, transformation, consolidation, integration from multiple operational data sources to a target DBMS

- **Data Mining:** is the process of discovering meaningful new correlations and trends by sifting through large amounts of data stored in repositories, using pattern recognition as well as statistical, machine learning, and mathematical techniques

- **Business Intelligence:** BI is the user-centered process of exploring data, data relationships and trends - thereby helping to improve overall decision making.

Data Warehousing

- A data warehouse is a physical dataset enabling an integrated view of the underlying DataSources.

External
Data

External Data
Data

Analysis
Methods

Calculating
Tables

Data Collection
1. Oriented by subject
2. Integrated
3. Constant
4. Time referenced

Creating
Graphics

Data Warehouse

DW in
narrower sense

Creating
Reports

Operative
Data

Operati
Data

DW in broad sense

Data Warehousing

- Three-Level DW Concept:

Information Analysis
Data Storage
Data Staging

- DW is an architecture for five components:
  - BI tools
  - DW administration
  - DW Database MS
  - Data coversion and extraction
  - Operational data source

Data Warehousing

Implementation difficulties:

- DW can't be taken from a shelf – has to be compiled using a variety of components
- Clear idea of goals and benefits is a necessity
- DW is an architecture, not a product
- Can't be bought – need to be built

Data Mining

Data mining – *the extraction of hidden predictive information from large databases*

Data mining tools predict future trends and behaviors, allowing to make proactive, knowledge-driven decisions.

Three main reasons for data mining:

- Correcting data – correction of incompleteness and contradictory information

- Discovering knowledge – to determine hidden relationships, patterns and correlations from data

- Visualizing data – to humanize the mass of data, display the data

Data correcting

Discovering knowledge

Visualizing data

Data Mining

Methodology:

1. Database selection and preparation
   - identification of databases and factors to be explored
   - preparation includes filling in missing values and rectifying errors

2. Cluster and feature analysis
   - database groups are divided using clustering techniques
   - more detailed feature analysis to find the main factors

......

Data Mining

3. Tool selection
   - How many examples? How much processing? What are outputs? Simplicity of update?

4. Hypothesis testing and knowledge discovery
   - hypotheses are formed and tested (top down)
   - new relationships are discovered (bottom up)
   - what-if analyses may be performed

5. Knowledge application
   - tested rules created from the discovery process can be directly added to either procedural code or into a knowledge-based system.

Data Mining

- Data mining process:
  -

Data Mining System

Data Mining *Algorithm*

Training

Test

Model

Score Model

Results
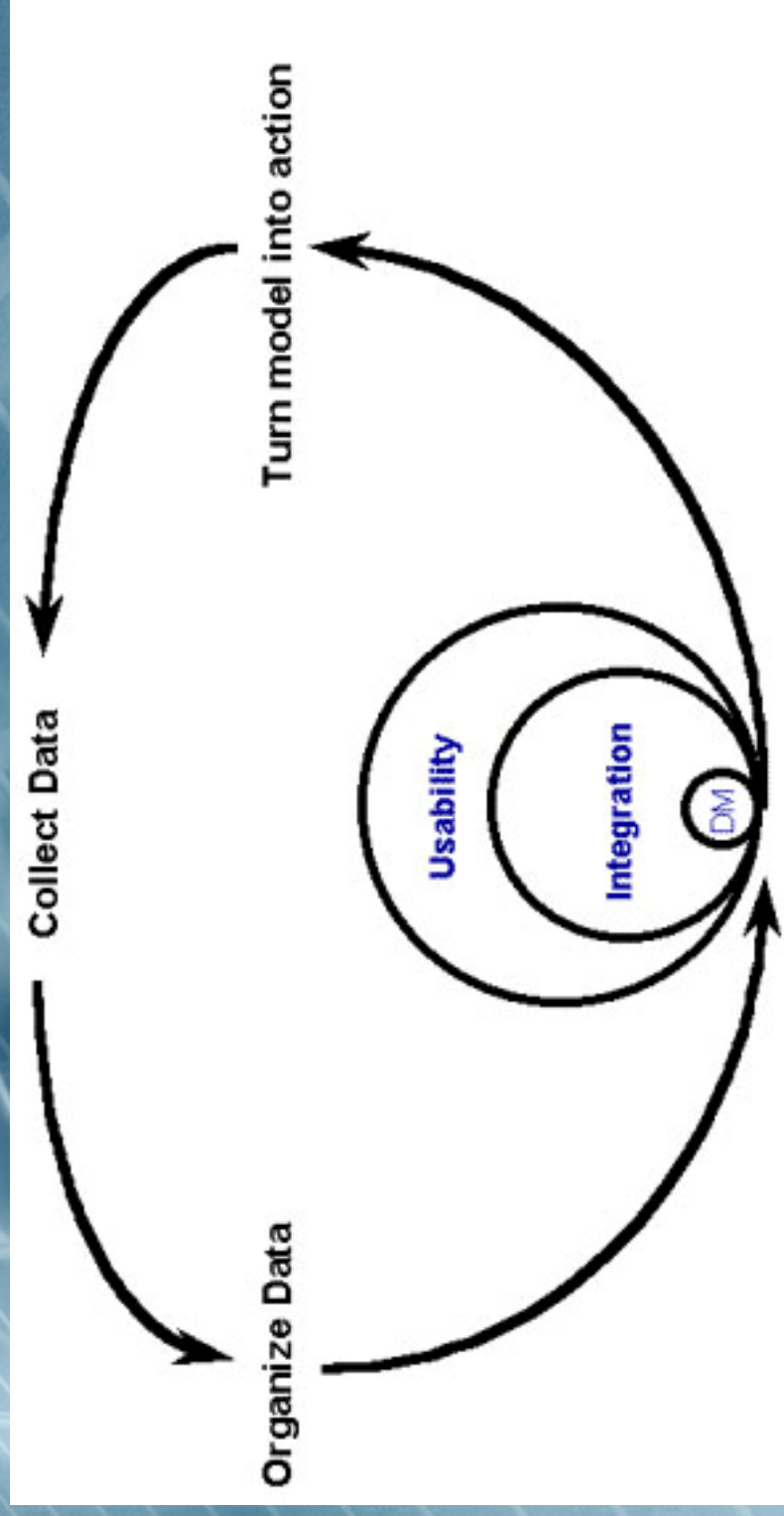
Training

Eval

Prediction

Historical Training Data

New Data

Data Mining

Techniques for DM:

- Visualization
  - colored 2-D – 4-D representation
- Statistics
  - Choice for initial analysis
  - clustering
  - factor analysis
  - prediction
  - Identification predictive factors
- Induction
  - Decision tree
  - Reasoning from specific facts to reach a hypothesis
- Neural Networks
  - unsupervised
  - supervised
  - Multilayered network architectures that learn how to solve a problem

Data Mining

- Data mining process:
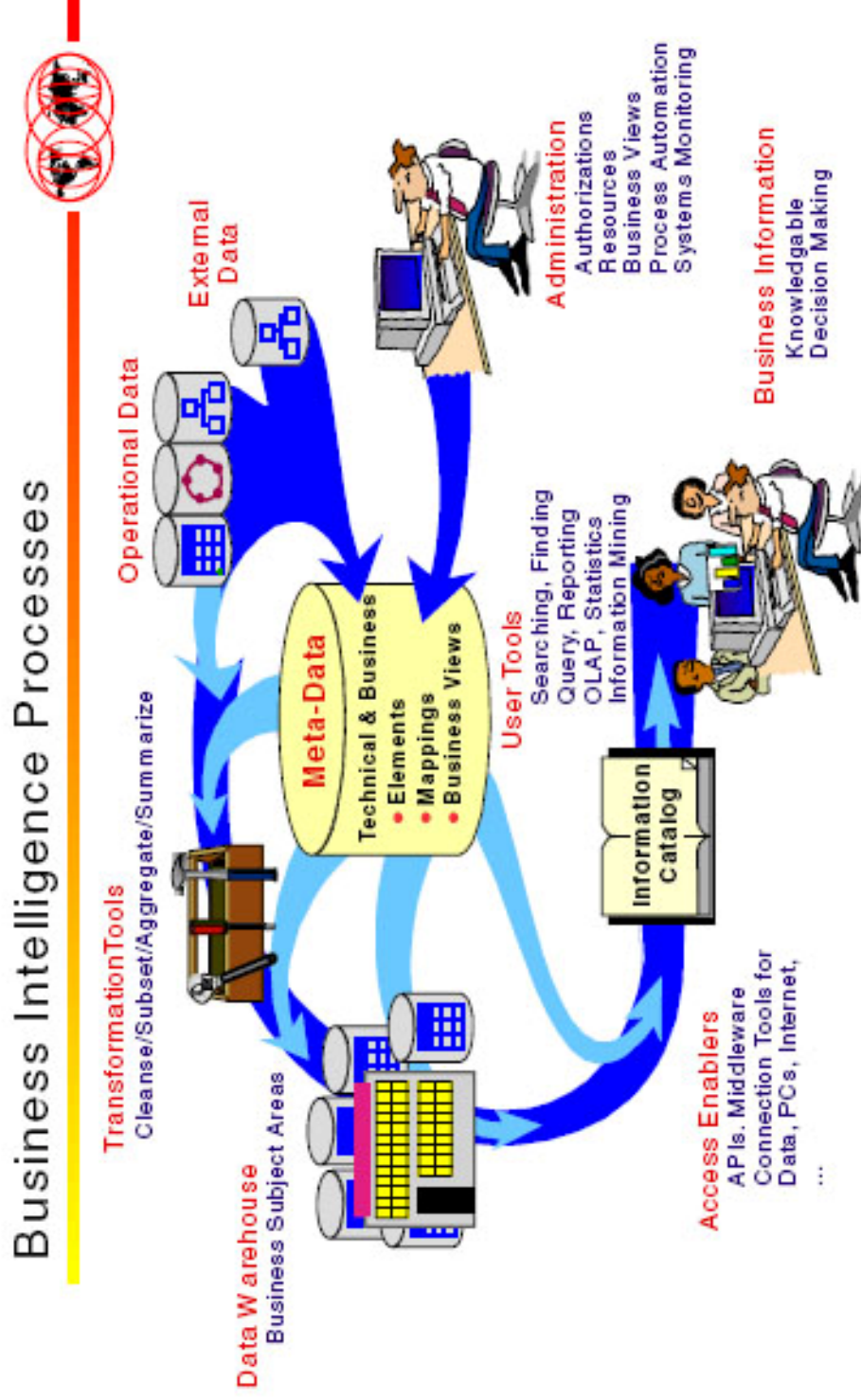
Business Intelligence

- Business intelligence (BI) – the processes and technologies used for collecting, managing, and reporting decision-oriented data.

- The business intelligence architecture is:
  - a subset of the overall IT architecture
  - an umbrella term for an enterprise-wide set of systems, applications, and governance processes
  - sophisticated analytics, by allowing data and analyses to flow to those who need them, when they need them.

Business Intelligence

■ BI architecture in six elements:

1. Data management
2. Transformation tools and processes
3. Repositories of data and metadata
4. Applications and other software tools for analysis.
5. Presentation tools and applications
6. Operational processes (security, error handling, archiving and privacy)

Business Intelligence

BI processes:

Business Intelligence Processes

TransformationTools
Cleanse/Subset/Aggregate/Summarize

Operational Data

External Data

Administration
Authorizations
Resources
Business Views
Process Automation
Systems Monitoring

Business Information
Knowledgable
Decision Making

Meta-Data
Technical & Business
- Elements
- Mappings
- Business Views

User Tools
Searching, Finding
Query, Reporting
OLAP, Statistics
Information Mining

Data Warehouse
Business Subject Areas

Information Catalog

Access Enablers
APIs. Middleware
Connection Tools for
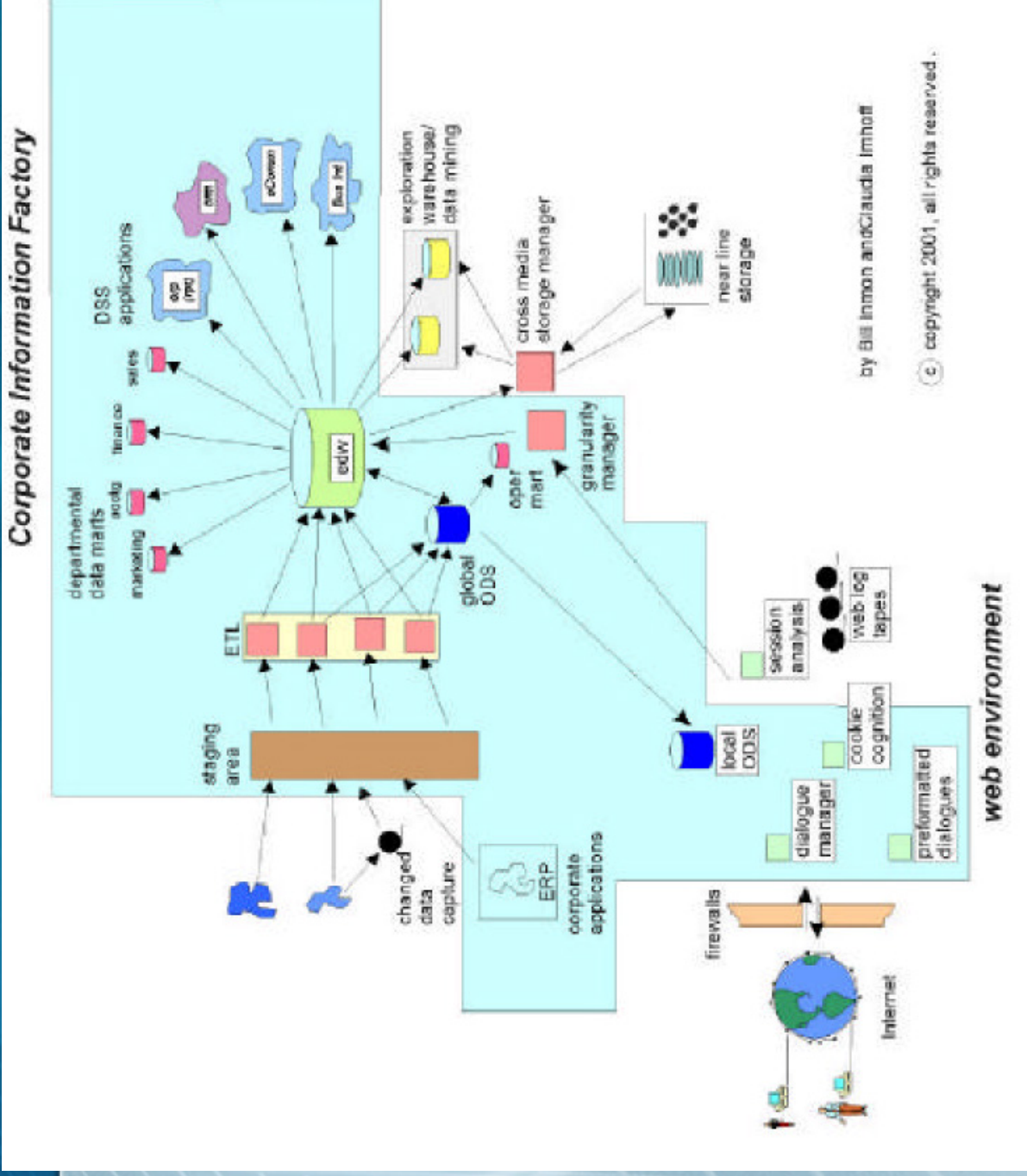Data, PCs, Internet,
...

Business Intelligence

BI processes and tasks can be summarized as:

- Understand the business problem to be addressed
- Design the warehouse
- Learn how to extract source data and transform it
- Load the warehouse
- Connect users and provide them with tools and way to find the data of interest in the warehouse
- Use the data to provide business knowledge
- Administer all these processes
- Document all this information in meta-data

SAP Development

1. SAP introduced as an ERP vendor

2. SAP and multidimensional technology
   - InfoCubes – data in form suitable to analysis

3. SAP and Data Warehouses:
   - BW 1.2b - introduction of InfoCubes and Business Content
   - BW 2.0b - introduction of ODS
   - BW 2.1c - analytical components
   - BW 3.0 – enhancement of ODS into DW; creation of analytical applications

SAP Information Factory

Corporate Information Factory

by Bill Inmon and Claudia Imhoff

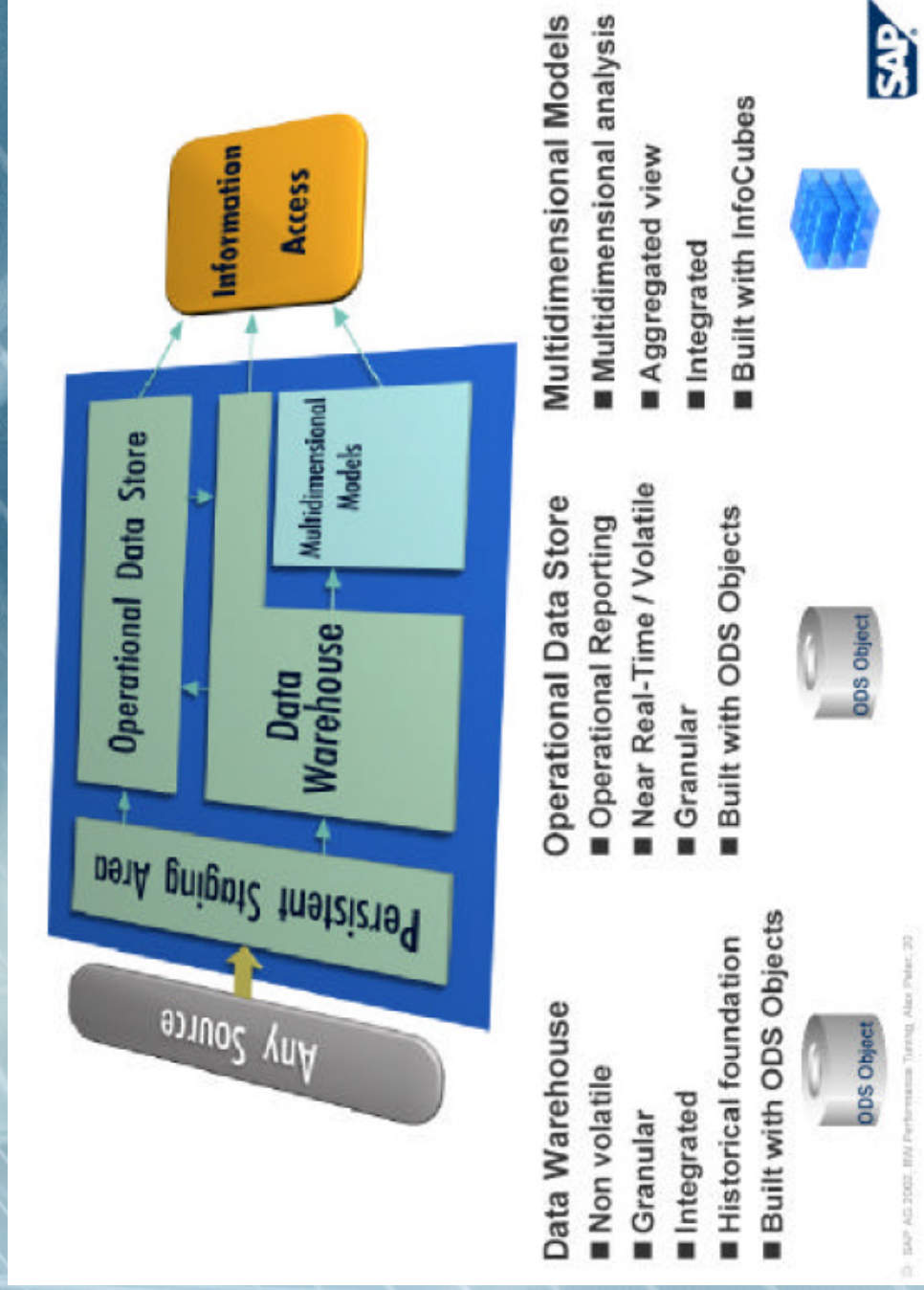(c) copyright 2001, all rights reserved.

SAP Information Factory

1. SAP support of ERP
   - SAP was the worlds pioneer in this arena
   - ERP foundation gives SAP a basis for gathering and managing data

2. SAP support of data warehouse
   - first foray was with the help of an Operational Data Store
   - ODS contains granular data ODS serves as a warehouse and operates in a mode of openness

3. SAP support for data marts
   - SAP has InfoCubes
   - InfoCubes plays a robust role

SAP Information Factory

4. SAP's support of the web environment

   ■ mySAP.com has interfaces from the web to and from the corporate infrastructure

5. SAP support for DSS applications

   ■ the concept of the "cockpit" – for management with the need for up to date and a wide variety of information.

   ■ SAP's offering of SEM – Strategic Enterprises Management includes:

      • SEM-BIC – business information collection,
      • SEM-BPS – planning and simulation,
      • SEM-BCS – business consolidation,
      • SEM-CPM – corporate performance monitoring,
      • SEM-SRM – stakeholder relationship management

SAP

Information Access

Data Warehouse
- Non volatile
- Granular
- Integrated
- Historical foundation
- Built with ODS Objects

Operational Data Store
- Operational Reporting
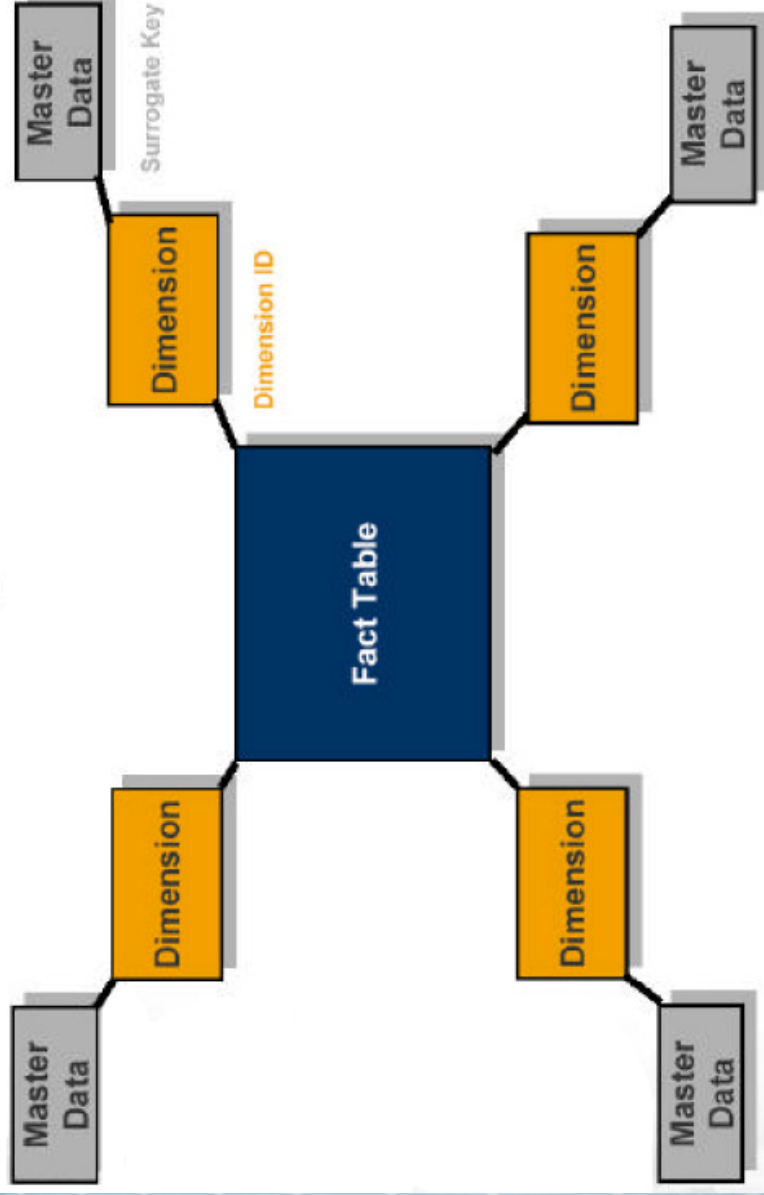- Near Real-Time / Volatile
- Granular
- Built with ODS Objects

Multidimensional Models
- Multidimensional analysis
- Aggregated view
- Integrated
- Built with InfoCubes

Operational Data Store

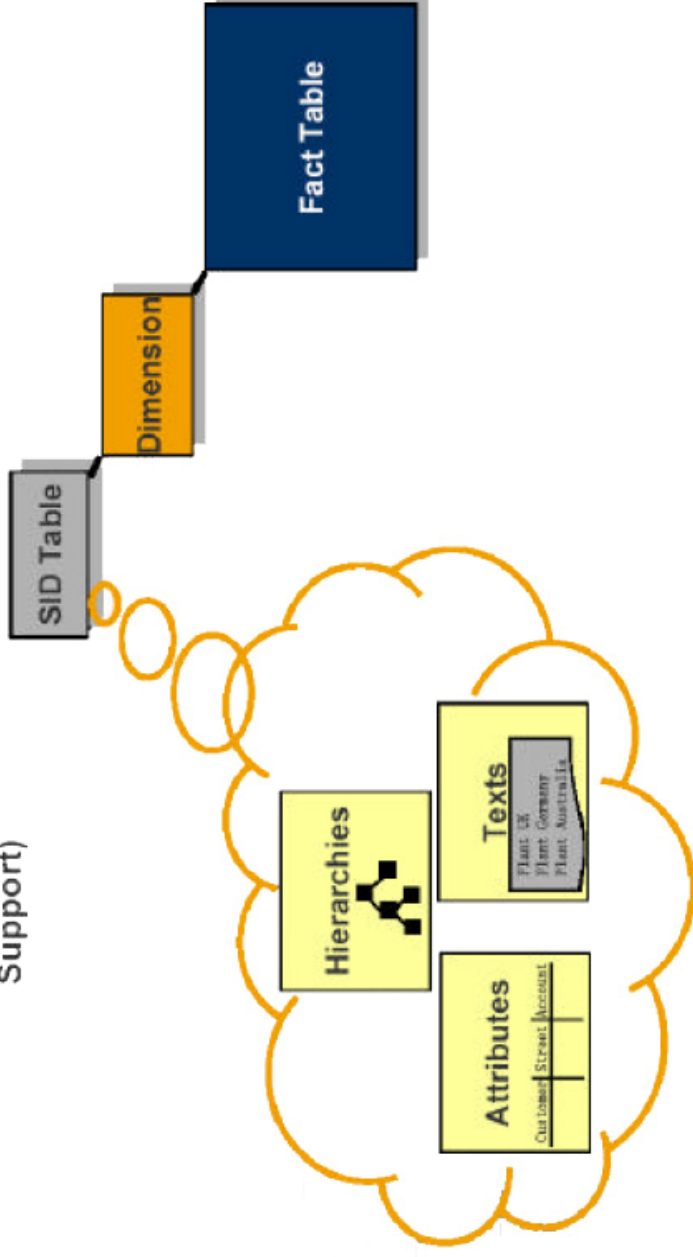Multidimensional Models

Data Warehouse

Persistent Staging Area

Any Source

ODS Object

ODS Object

SAP

© SAP AG 2002, BW Performance Tuning, Alex Peter, 20

SAP

SAP-Extended Star Schema

- ONE Normalization of Dimensions
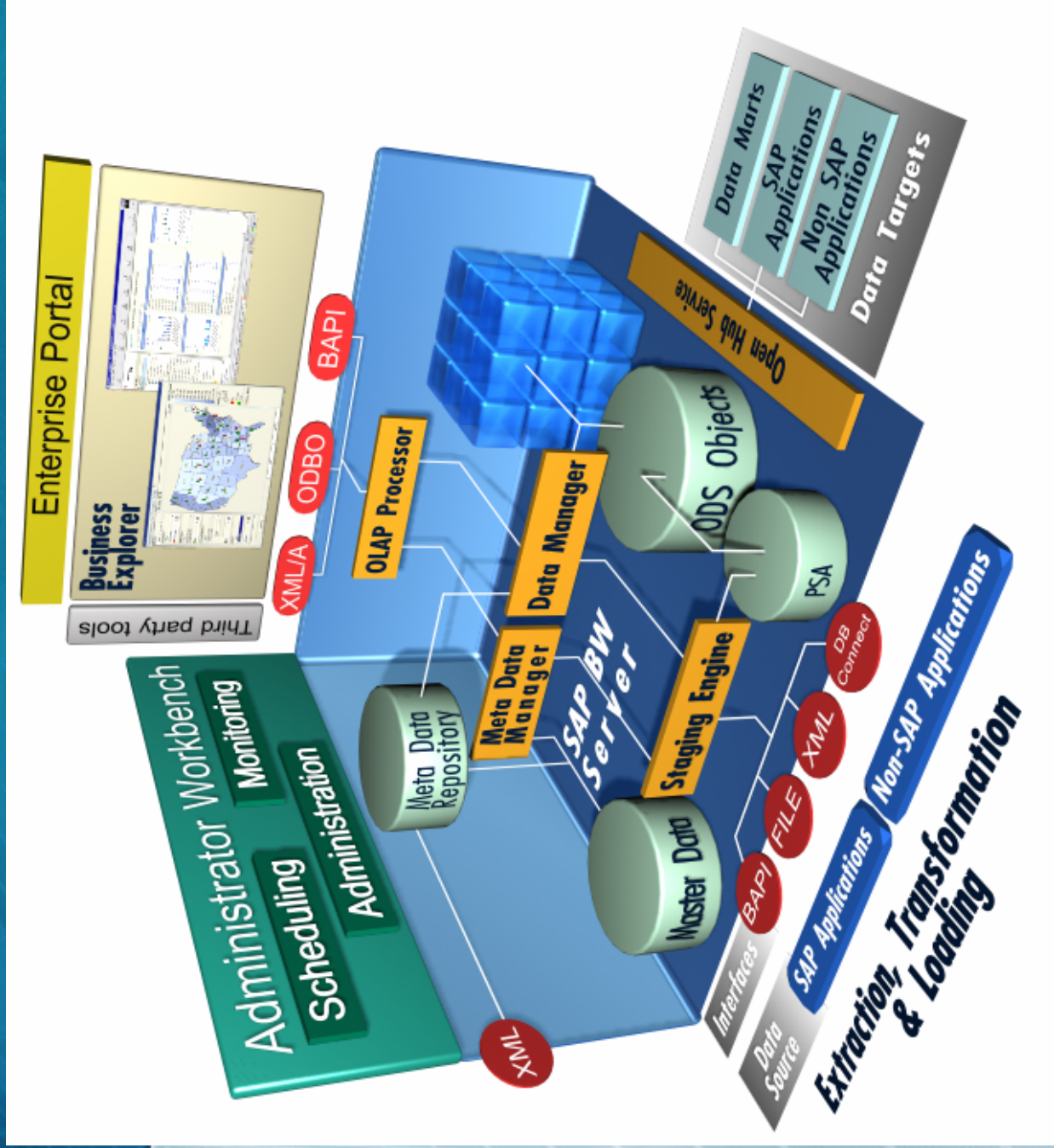- Master Data is shared, not part of the InfoCube

References

- Basics – Reporting & Analysis – Modeling & Staging – Data Mining, Helmut Krcmar SAP UCC at Technische Universität München

- Business Intelligence, Gartner Group, http://www.gartner.com

- Data warehouse, Wikipedia, http://en.wikipedia.org/wiki/Data_warehouse

- An Introduction to Data Mining, Kurt Thearling, www.thearling.com/SAP

- Data mining, Wikipedia, http://en.wikipedia.org/wiki/Data_mining

- Building Data Mining Applications for CRM, Alex Berson, Stephen Smith, Kurt Thearling

- The New Science of Winning Harvard Business School Press, March 2007

- The Architecture of Business Intelligence, Thomas H. Davenport and Jeanne G. Harris

- Business Intelligence Architecture on S/390, ibm.com/redbooks

- Multi-Dimensional Modeling with BI, SAP, May 2006

- BI Data Modeling and Frontend Design, SAP Community Network https://www.sdn.sap.com/irj/sdn/

Conclusion

Thank you for your attention

May I answer your questions?