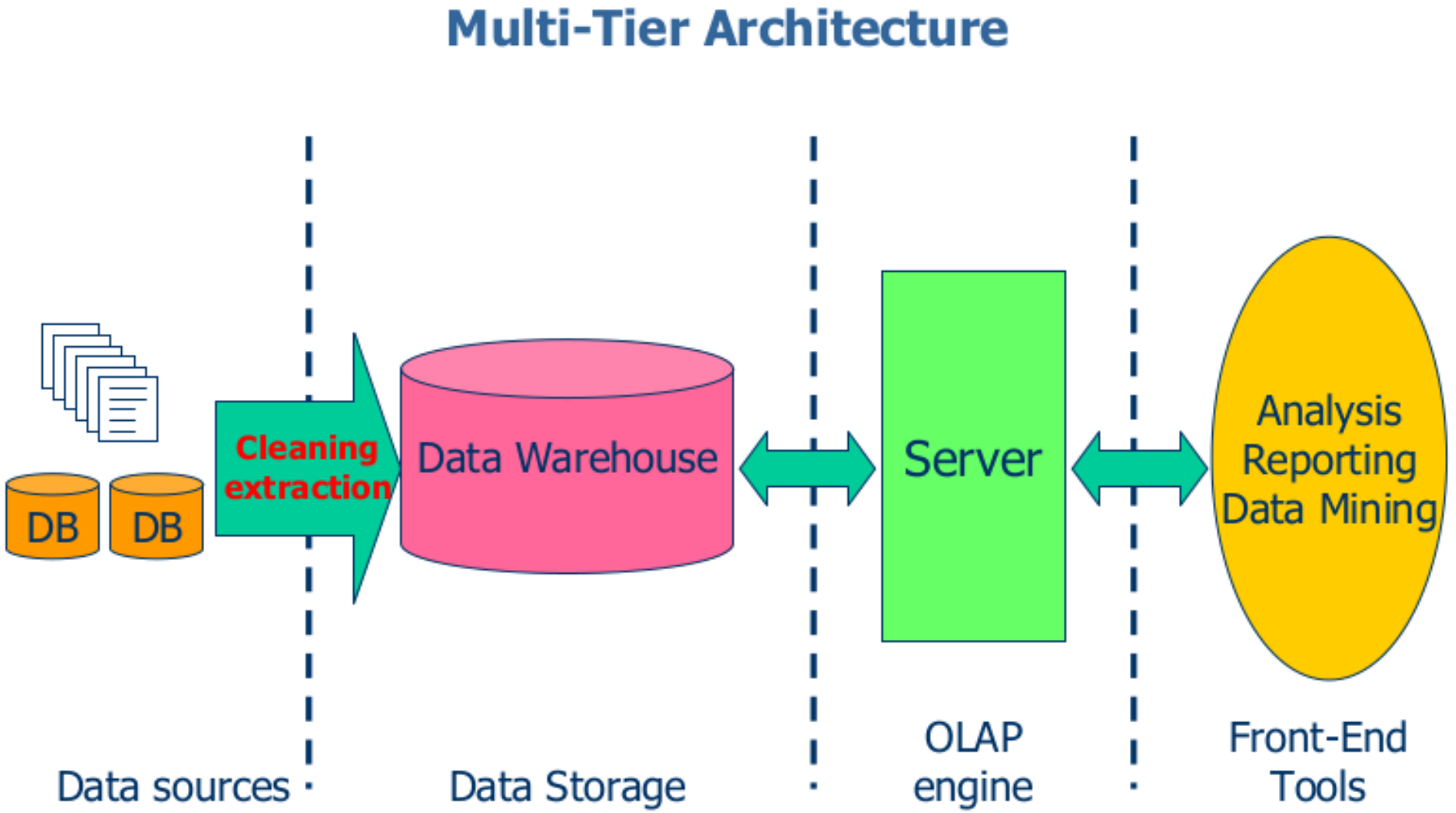
Unit I

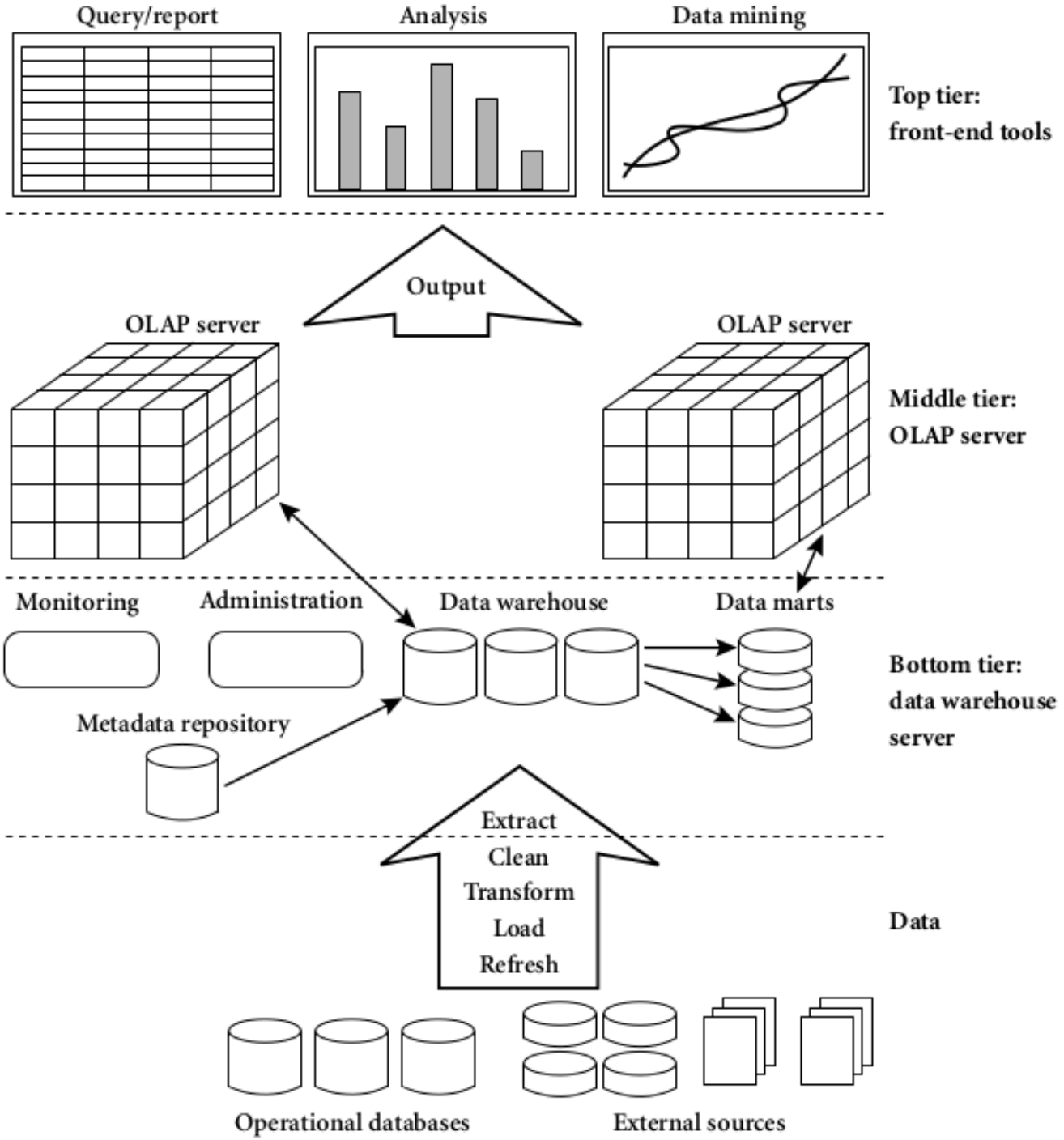
- Introduction
  - Data warehousing
  - Multidimentional data model
  - OLAP operations
  - Warehouse schema
  - DW architecture
  - Warehouse server
  - Metadata
  - OLAP Engine
  - DW backend process

FOCUS on making yourself BETTER, not on thinking that you are better

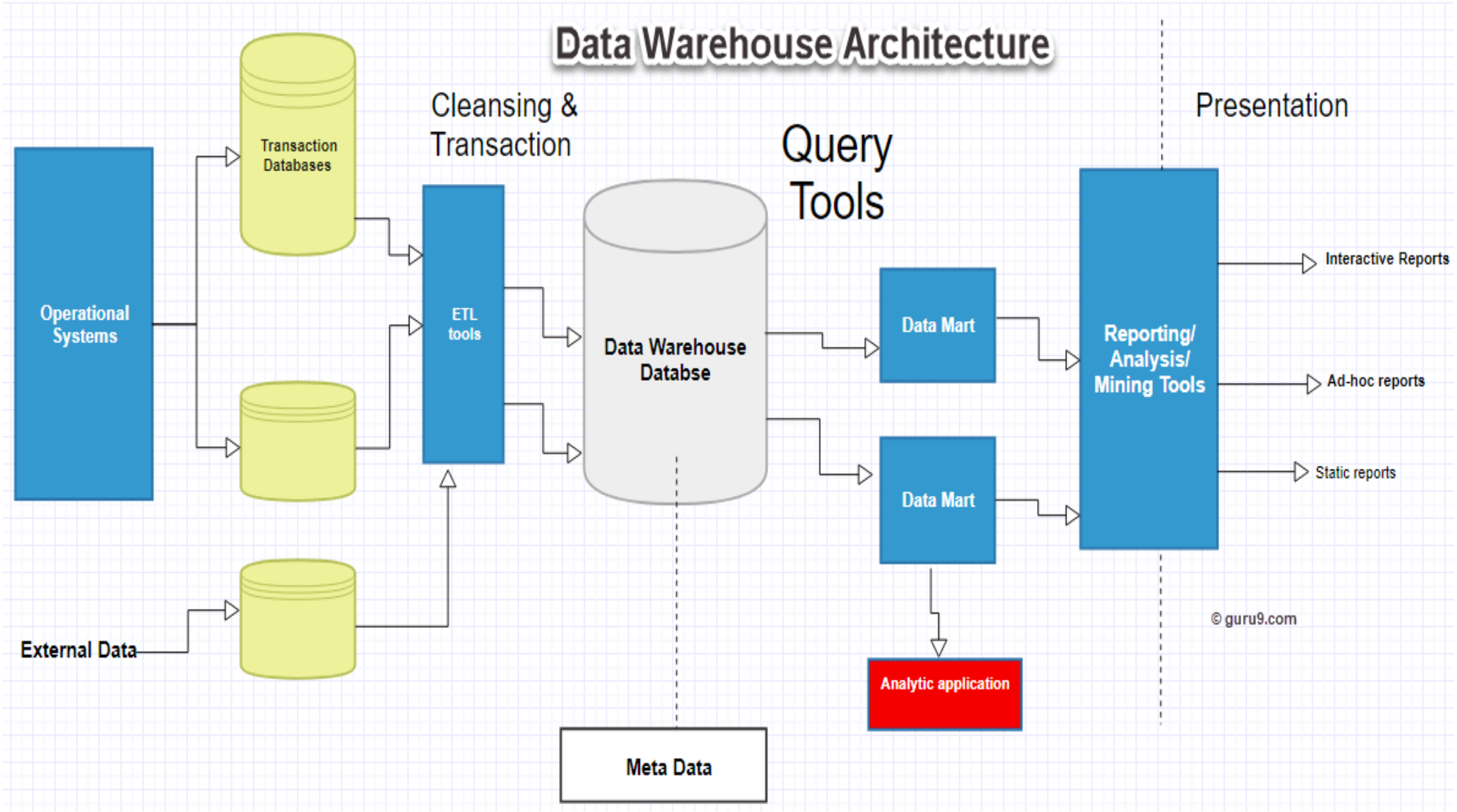
Data Warehouse Architecture



FOCUS on making yourself BETTER, not on thinking that you are better



FOCUS on making yourself BETTER, not on thinking that you are better



FOCUS on making yourself BETTER, not on thinking that you are better

- Metadata are
  - data about data, information about data stored
- In a data warehouse,
  - metadata are the data that define warehouse objects
  - Created for data names and definitions of the DW
  - Used to pinpoint access to infomation
  - Metadat is the roadmap to th contents of the DW
  - Acts as directory to locate the contents
  - Has To provide unambiguous interpretation of the contents

- It alos contains info such as
  - Data extraction/transformation history
  - Alsas names of the coulm
  - Table sizes
  - Data Modelling algo used
  - Usage statistics
  - Hierarchical relationships
  - Calcualtion performed
  - Operations during propagation,cleanng and transformation
  - Audit report

FOCUS on making yourself BETTER, not on thinking that you are better

- **What for metadata**
  - Acts as catalog
  - Mapping of data
  - To help DSS
  - Summary preparation
  - Query tools and reporting tools
  - Used in cleaning and ETL

- **Challenges**
  - Metadata **scattered** across orgn
  - May be stored in **different** formats and different types of files
  - No **standard** for defining metadata
  - No **easy and well defined** method for passing the metadata to applications

- **Metadata repository**
  - The metadata repository is responsible for physically storing and cataloging metadata.
  - Data in a metadata repository should be
    - Generic : store the metadata by generic terms instead of storing it by an applications-specific defined way
    - Integration of the metadata repository allows all business areas' metadata to be in an integrated fashion: covering all domains and subject areas of the organization.
    - The metadata repository should have accessible current and historical metadata.
  - Metadata repositories used to be referred to as a data dictionary.

- **Metadata repository**
  - Is a **Place** where metadata is stored
  - It is an **Integeral part** of the DW system that includes
    - Descriptin of the DW structure
    - Operational metadata
    - Summarizaion algorithms
    - Business metadata
    - Mapping info
    - Data related to system performance

- Metadata Repository
  - Structure of DW
  - Operational metadata
  - Algorithms used for summarization

- **Data warehouse metadata**
- are pieces of information stored in one or more special-purpose metadata repositories that include
  - (a) information on the contents of the data warehouse, their location and their structure,
  - (b) information on the processes that take place in the data warehouse back-stage, concerning the refreshment of the warehouse with clean, up-to-date, semantically and structurally reconciled data,
  - (c) information on the implicit semantics of data (with respect to a common enterprise model) , along with any other kind of data that aids the end-user exploit the information of the warehouse,
  - (d) information on the infrastructure and physical characteristics of components and the sources of the data warehouse,
  - (e) information including security, authentication, and usage statistics that aids the administrator tune the operation of the data warehouse as appropriate.

FOCUS on making yourself BETTER, not on thinking that you are better

- **Categories of metadata** (based on the contents)
  - **Business** – ownership info, business functions defn, ploicies
  - **Technical** – DB names, tables, shemas, value descriptions etc
  - **Operational** – currency of data (it's a temporal reference: "the degree to which the data is current with the world it models) and data lineage(Data lineage is defined as a data life cycle that includes the data's origins and where it moves over time. It describes what happens to data as it goes through diverse processes)

- Categoies (based on how it is used)
  - Build Time – while designing & buliding the DW, used by designers, developers, acts as the source for other metadata
  - Usage – derived from build time metata data , when DW data is ropagated, used by end users, admins, DSS applications
  - Control – used by DB ssytems, system programmers, used for historical data, security applications

FOCUS on making yourself BETTER, not on thinking that you are better