Data Warehouse
——About Database——

徐华

清华大学 计算机系 智能技术与系统国家重点实验室

xuhua@tsinghua.edu.cn

1

About Data Warehouse

- **Data Collection: Multi-source Heterogeneous Data**
  - **Multi-Source：** Business Database (Structured) 、Internet (Semi-Structured, Unstructured) 、Business Report (Structured) 、Questionnaire (Structured, Unstructured) ......
  - **Heterogeneous** ： Structured, Semi-Structured, Unstructured ，Transformed to Structured Data
- Data Collection -》Data Cleaning -》Data Integration-》Data Reduction-》Data Discretization -》Concept Hierarchy
- After Pre-processing：Depiction
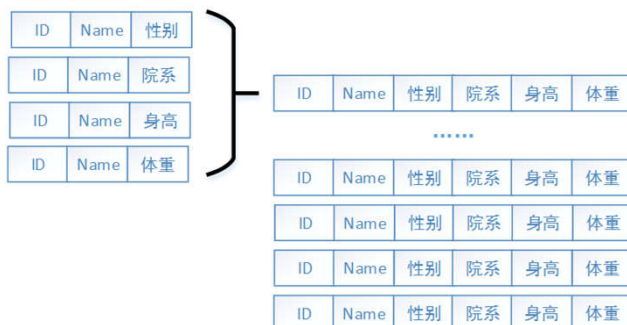
2

About Data Warehouse

⊙ **Warehouse v.s. Data Warehouse**

数据仓库/Data Warehouse   仓库/Warehouse

3

About Data Warehouse
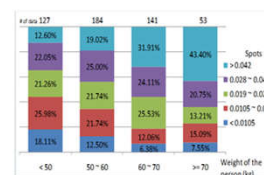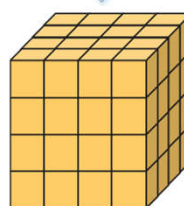
⊙ Data Warehouse

4

About Data Warehouse

◉ **Data Warehouse should be illustrated deeply.**

- ◆ **Concept**
- ◆ **Management Method**
- ◆ **Architecture**
- ◆ **How to realize**

5

Thanks !

6