Business Intelligence

Data Management and Business Performance Management

Prof. Dr. Ingo Claßen

Hochschule für Technik und Wirtschaft Berlin

Inro

Data Warehouse

Data Integration

Analytics

Cubes

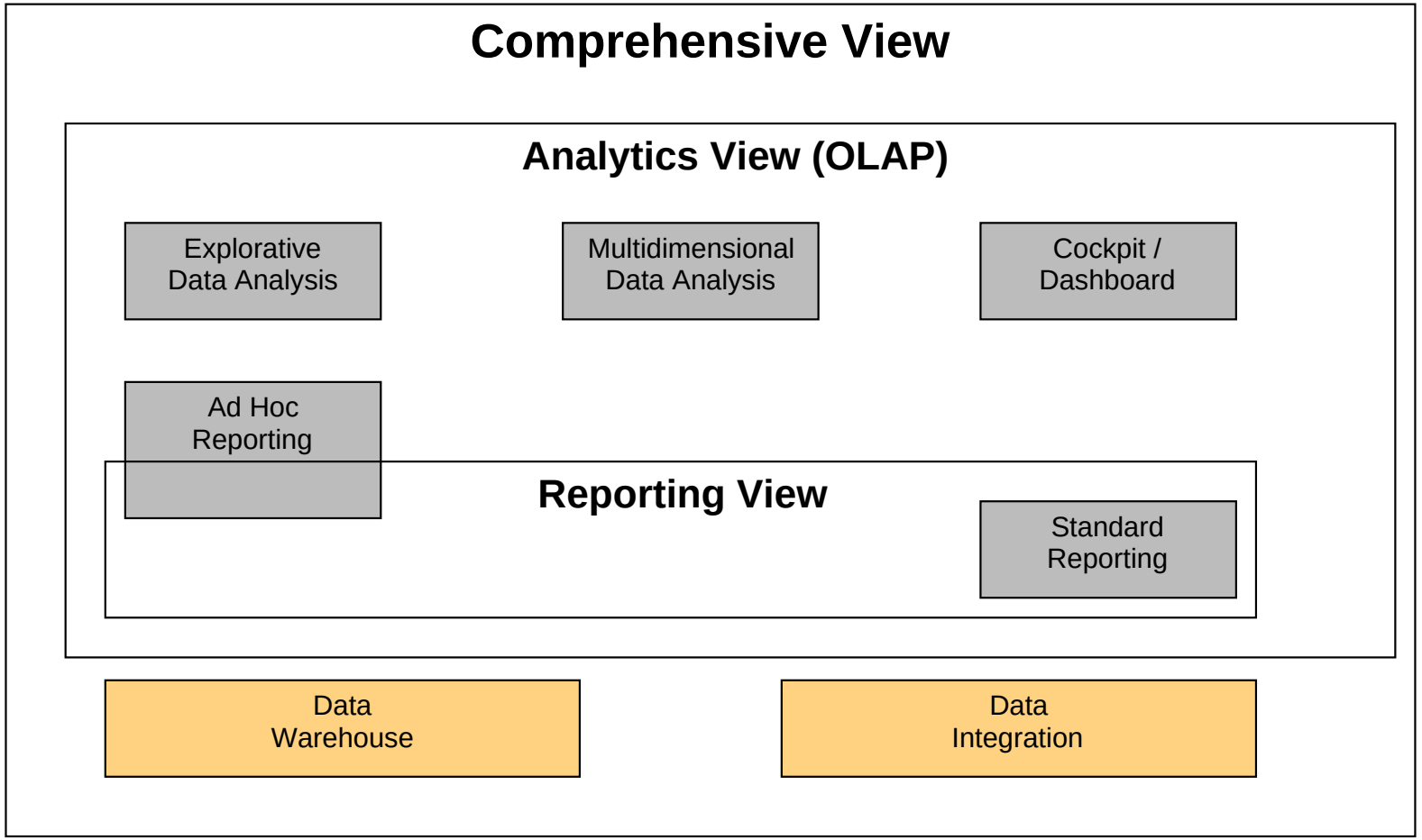
References

Business Intelligence (BI)

Umbrella term

- Applications, infrastructure and tools
- Best practices
- Access to and analysis of information
- Improvement and optimization of decisions and performance

BI Views


Comprehensive View
Analytics View (OLAP)
Explorative Data Analysis
Multidimensional Data Analysis
Cockpit / Dashboard
Ad Hoc Reporting
Reporting View
Standard Reporting
Data Warehouse
Data Integration

Data Warehouse

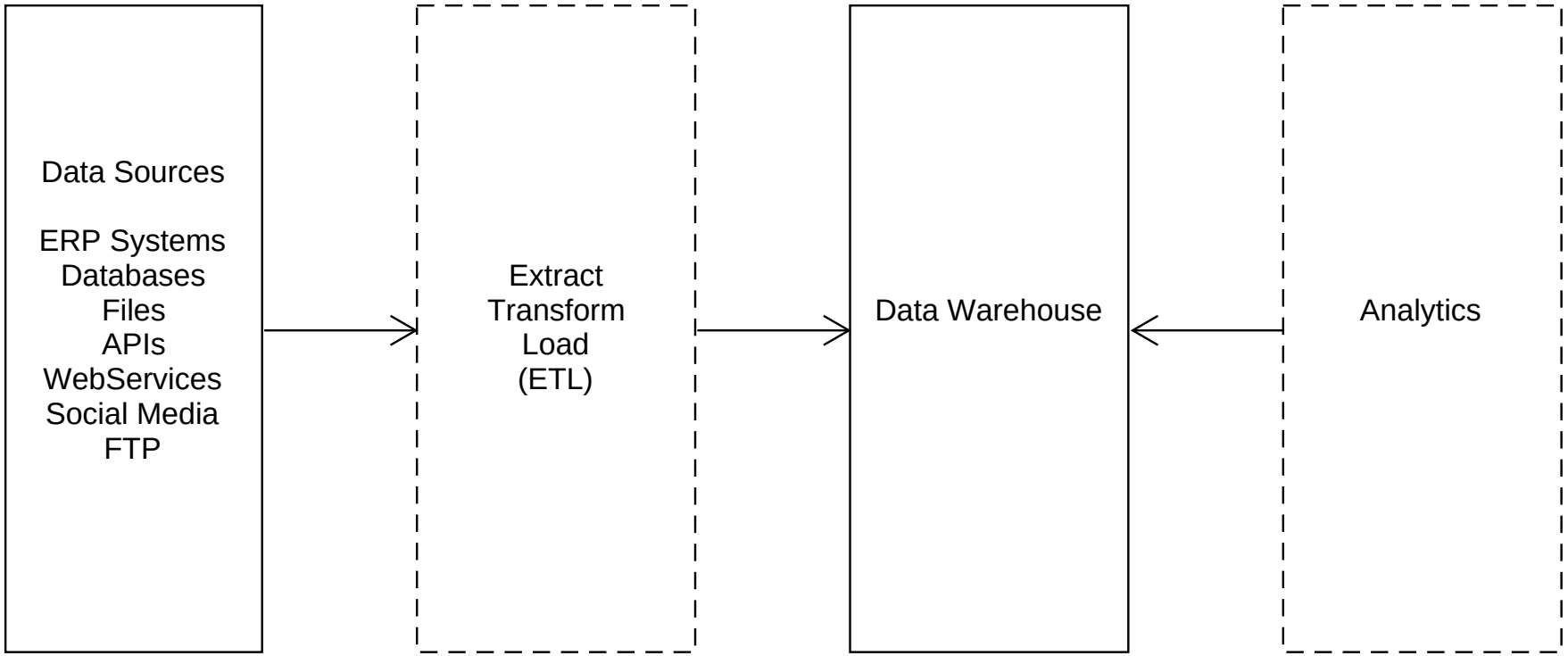
Data Warehouse is a

- Subject oriented – data are organized around sales, products, etc.
- Integrated – data are integrated to provide a comprehensive view
- Time variant – historical data are maintained
- Nonvolatile – data are not updated by users

collection of data

- usually used to improve decision making
- "single point of truth"

Architecture



Extract, Transform, Load (ETL)

Extract

- Get data from source systems

Transform

- Clean data
- Normalize data
- Transform data

Load

- Store data in Data Warehouse

ETL Operations

Input / Output

- File Access
- Database Access
- API / WebServices Access

Data Manipulation

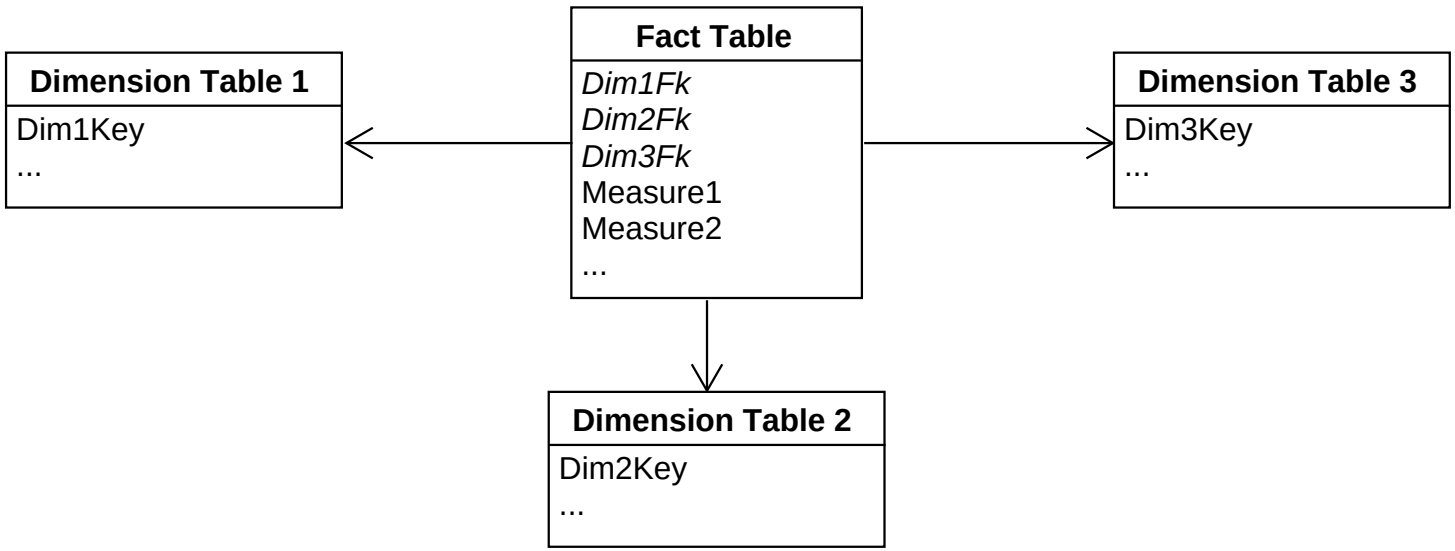
- Filter
- Split / Union
- Join
- Aggregation
- Calculations

Multidimensional Data Analysis

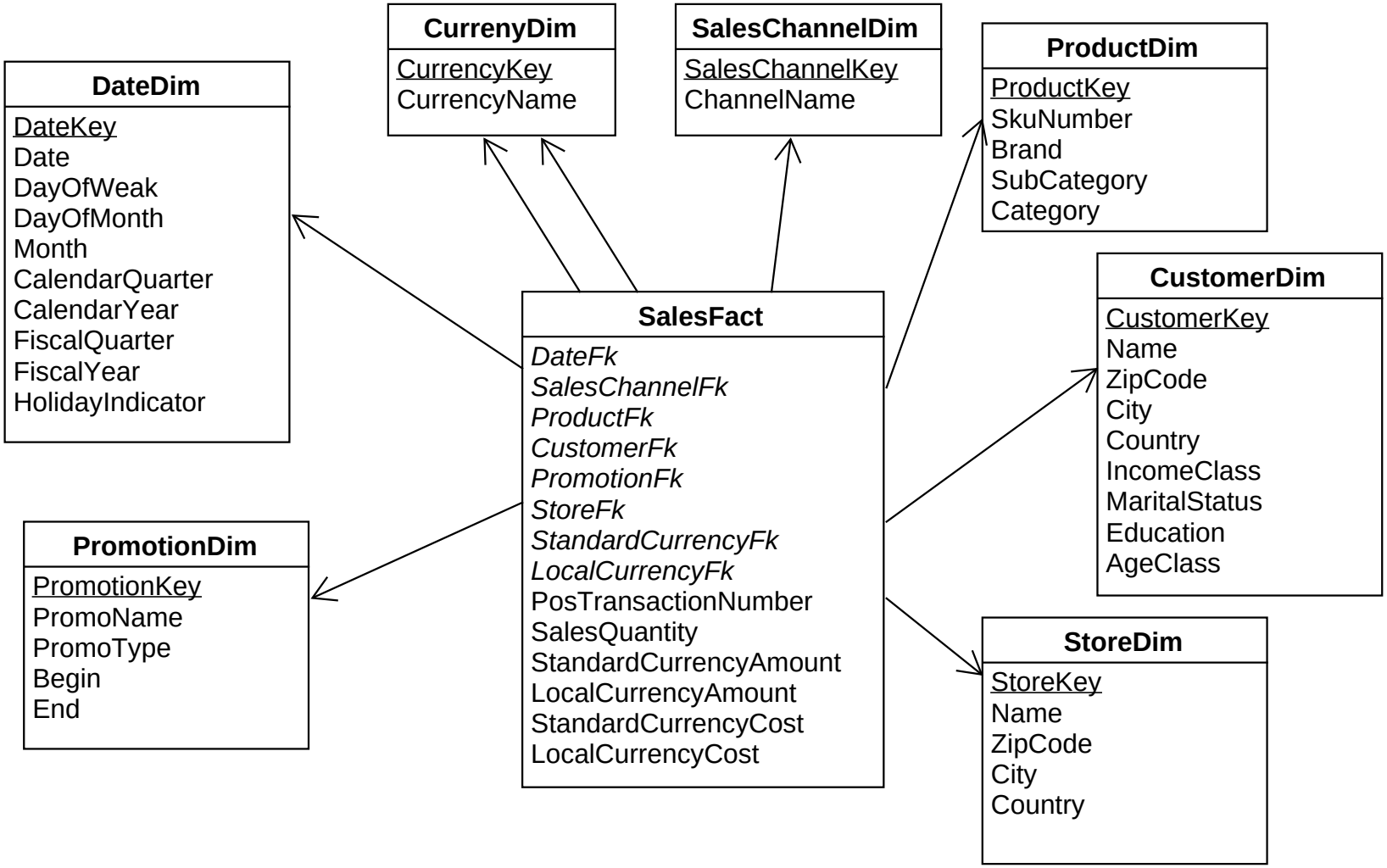
Main elements

- Measures (quantitive facts) – revenue, cost, profit
- Dimensions (descriptive information) – time, geography, product
- Hierarchies – day, month, year

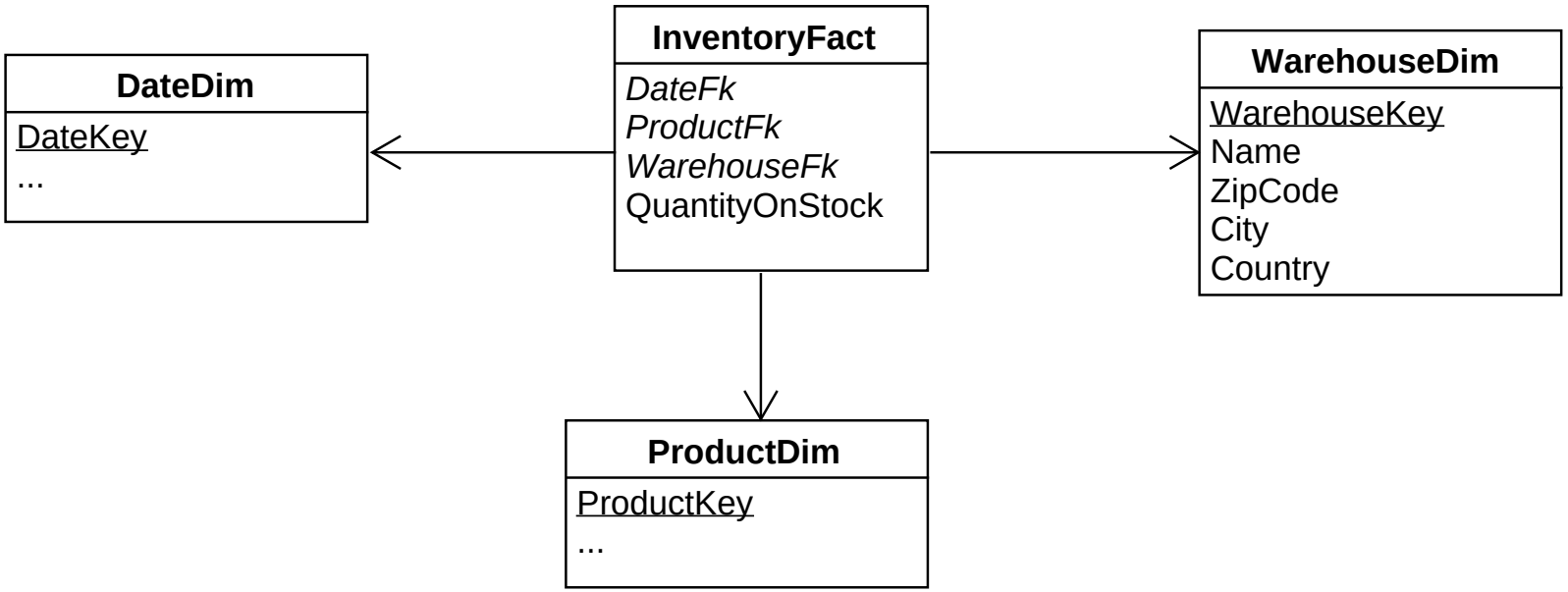
Star Schema



Sales Transaction Facts

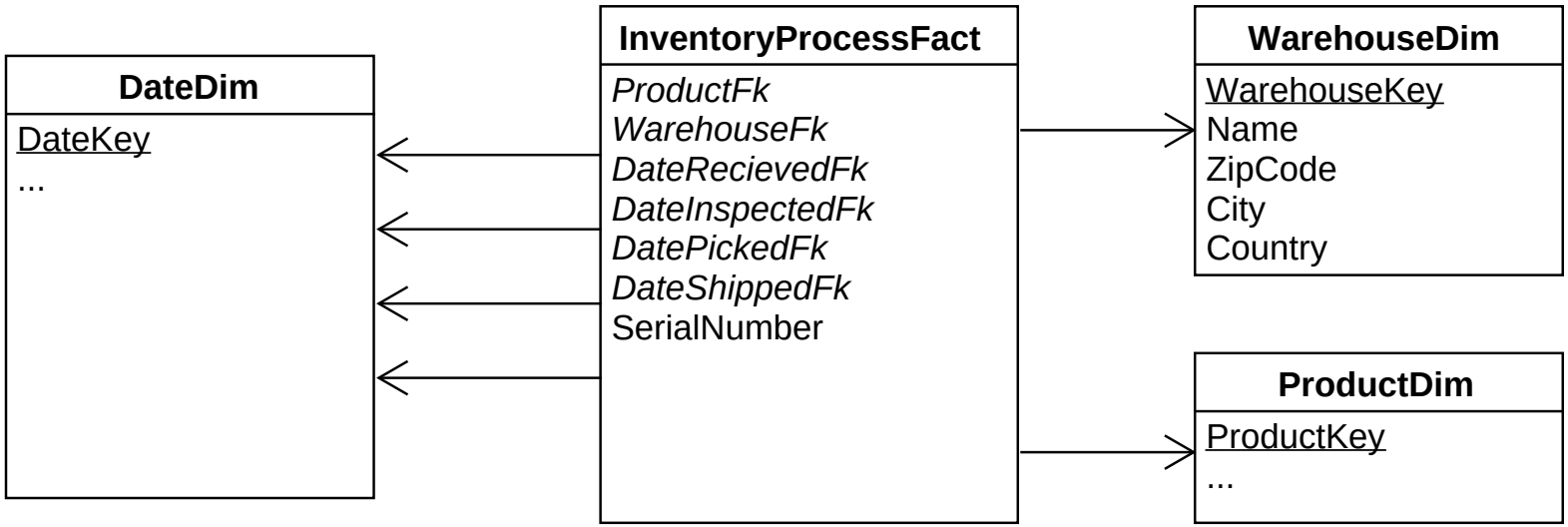


Inventory Snapshot Facts

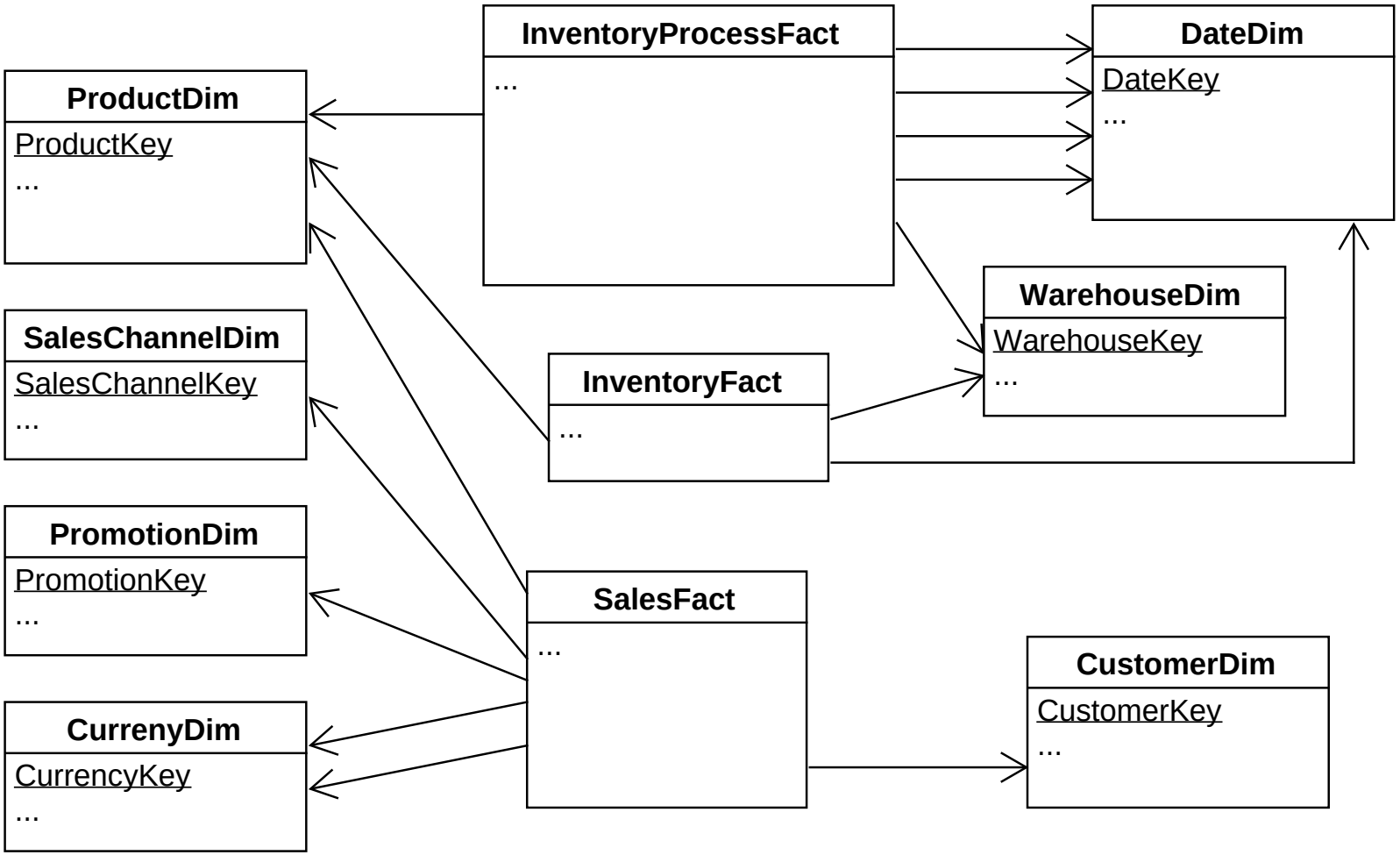


- QuantityOnStock – semiadditive

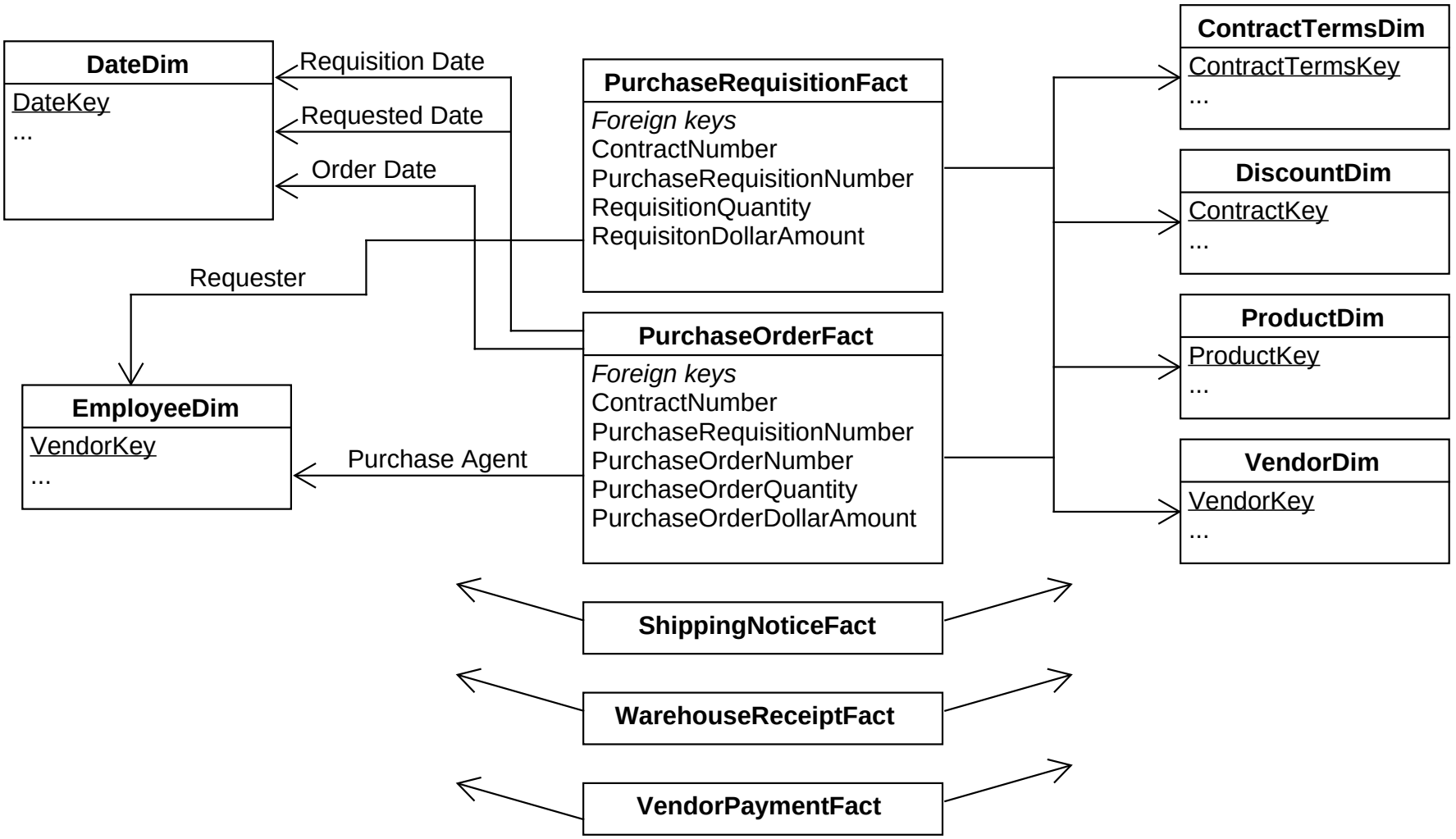
Inventory Process Facts



Conformed Dimensions



Procurement Transaction Facts



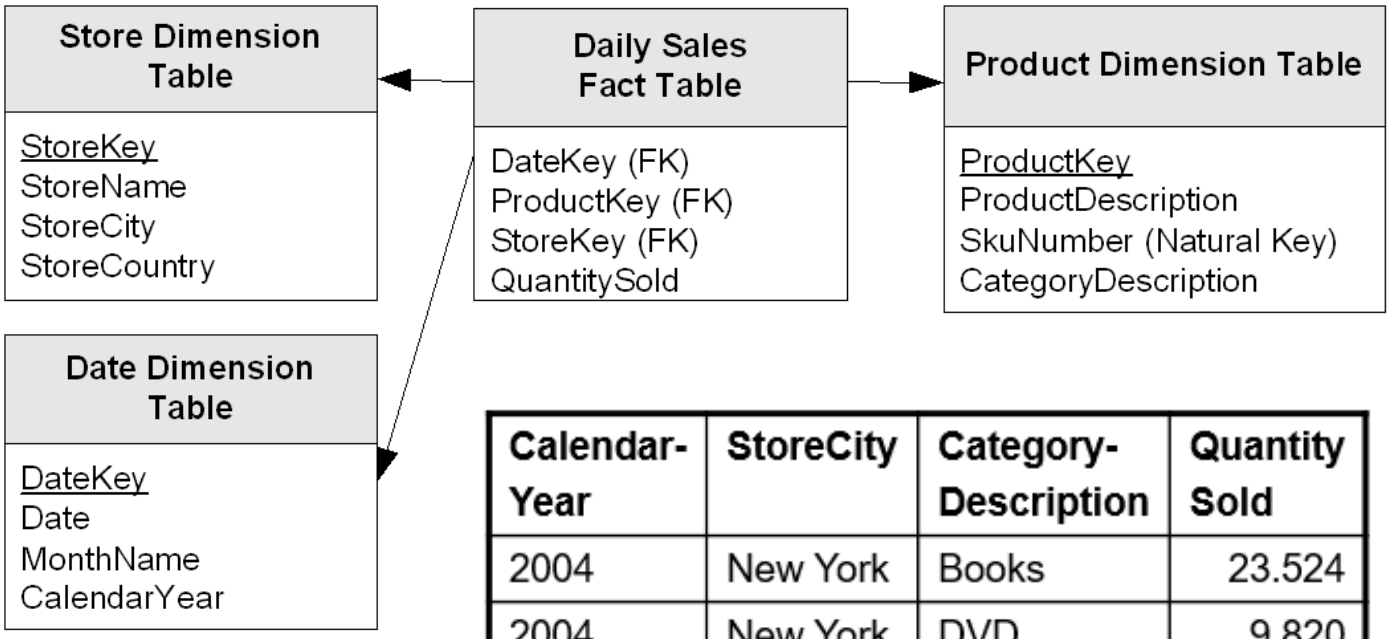
Kinds of Facts

	Fact		
Characteristic	**Transaction**	**Snapshot**	**Process**
Grain	One row per event	One row per period	One row per process
Fact Table loads	Insert	Insert	Insert and update
Fact row update	No	No	yes
Date dimension	Transaction date	End-of-period date	Multiple dates

Further Domains

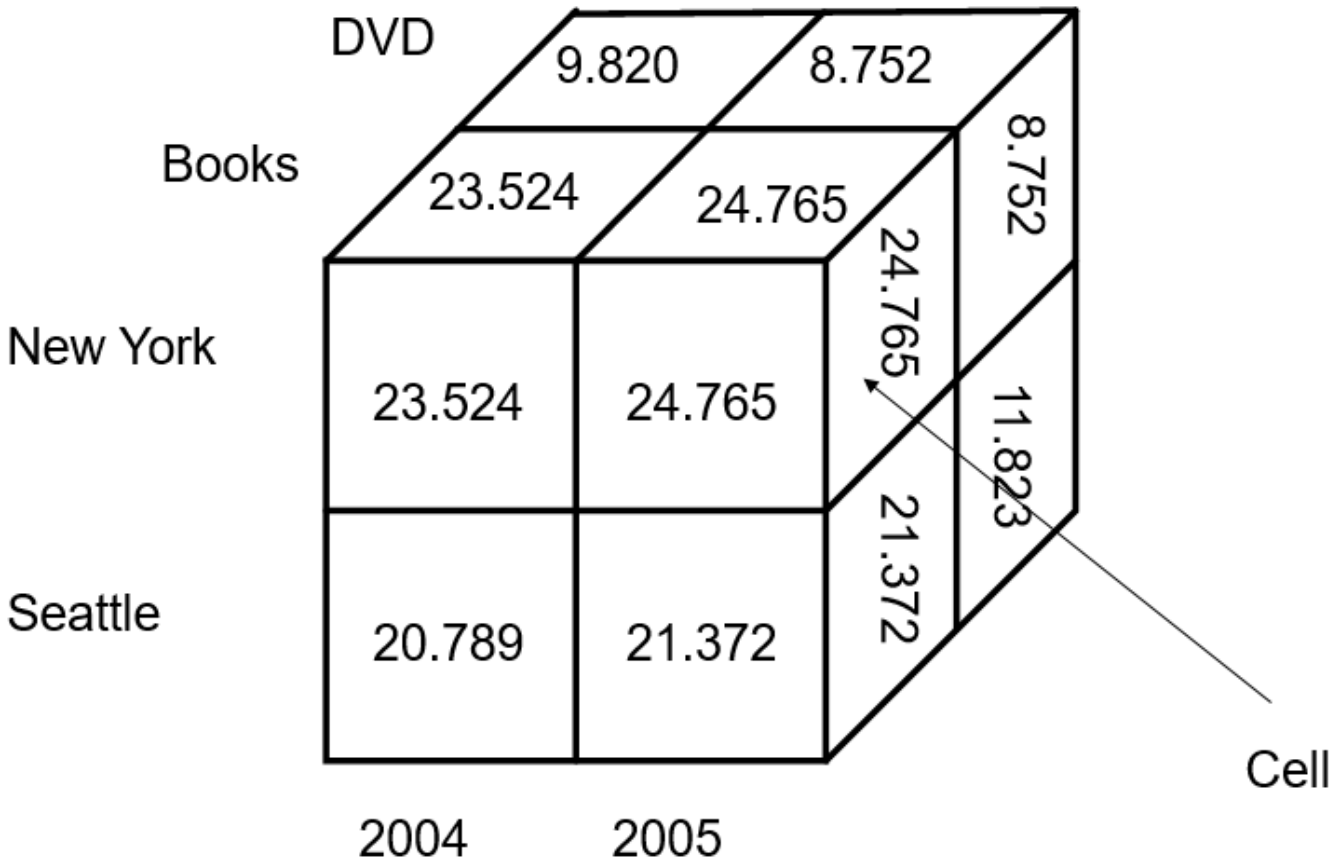
- Health care
- Telecommunications
- Banking
- Insurance
- Transportation
- Education
- Administration
- Travel

Relational View on Cubes



Calendar-Year	StoreCity	Category-Description	Quantity Sold
2004	New York	Books	23.524
2004	New York	DVD	9.820
2004	Seattle	Books	20.789
2004	Seattle	DVD	10.921
2005	New York	Books	24.765
2005	New York	DVD	8.752
2005	Seattle	Bücher	21.372
2005	Seattle	DVD	11.823

3-dimensional Interpretation



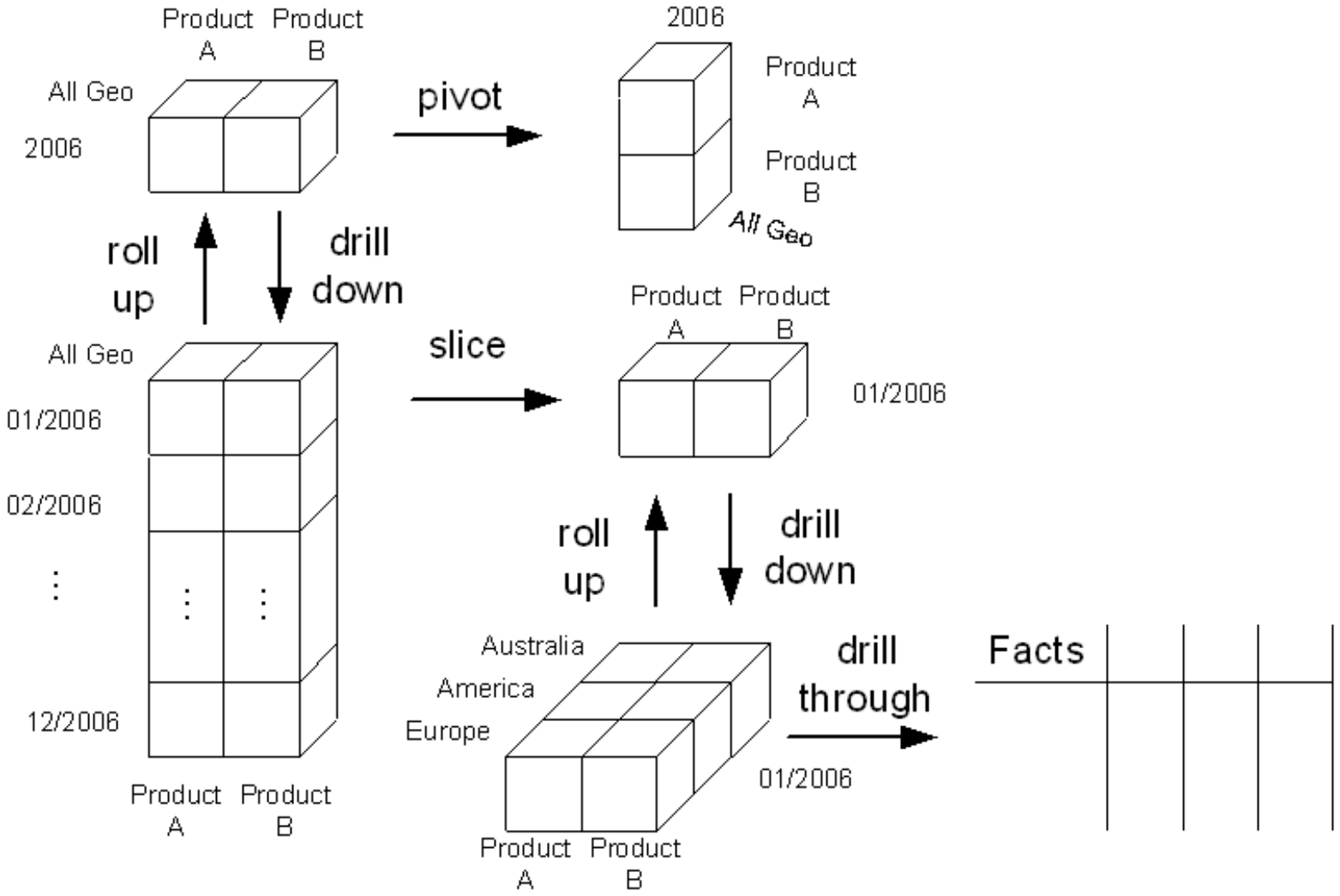
2-dimensional Representation

Common format used by tools

Cartesian dimension (**CalendarYear** x **StoreCity**)

Calendar-Year	StoreCity	Books	DVD
2004	New York	23.524	9.820
	Seattle	20.789	10.921
2005	New York	24.765	8.752
	Seattle	21.372	11.823

Cube Operations



Cube Technology

- ROLAP — OLAP based on relational database technology
- MOLAP — OLAP based on multidimensional database technology
- HOLAP — combination of ROLAP and MOLAP

References

- Ralph Kimball, Margy Ross – The Data Warehouse Toolkit
- Ralph Kimball, Joe Caserta – The Data Warehouse ETL Toolkit
- William H. Inmon – Building the Data Warehouse
- Gartner – Information Technology Glossary `https://www.gartner.com/en/information-technology/glossary`