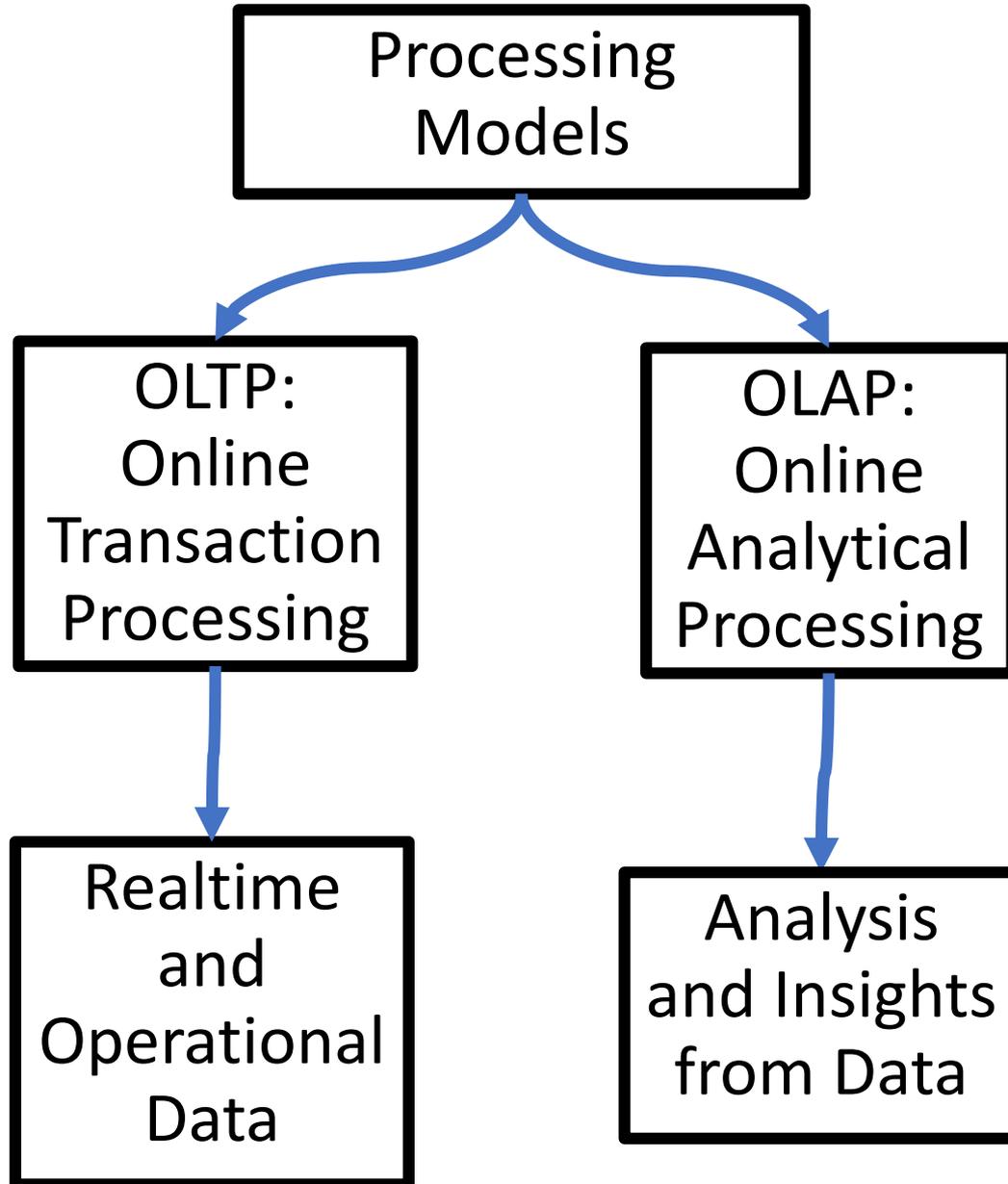
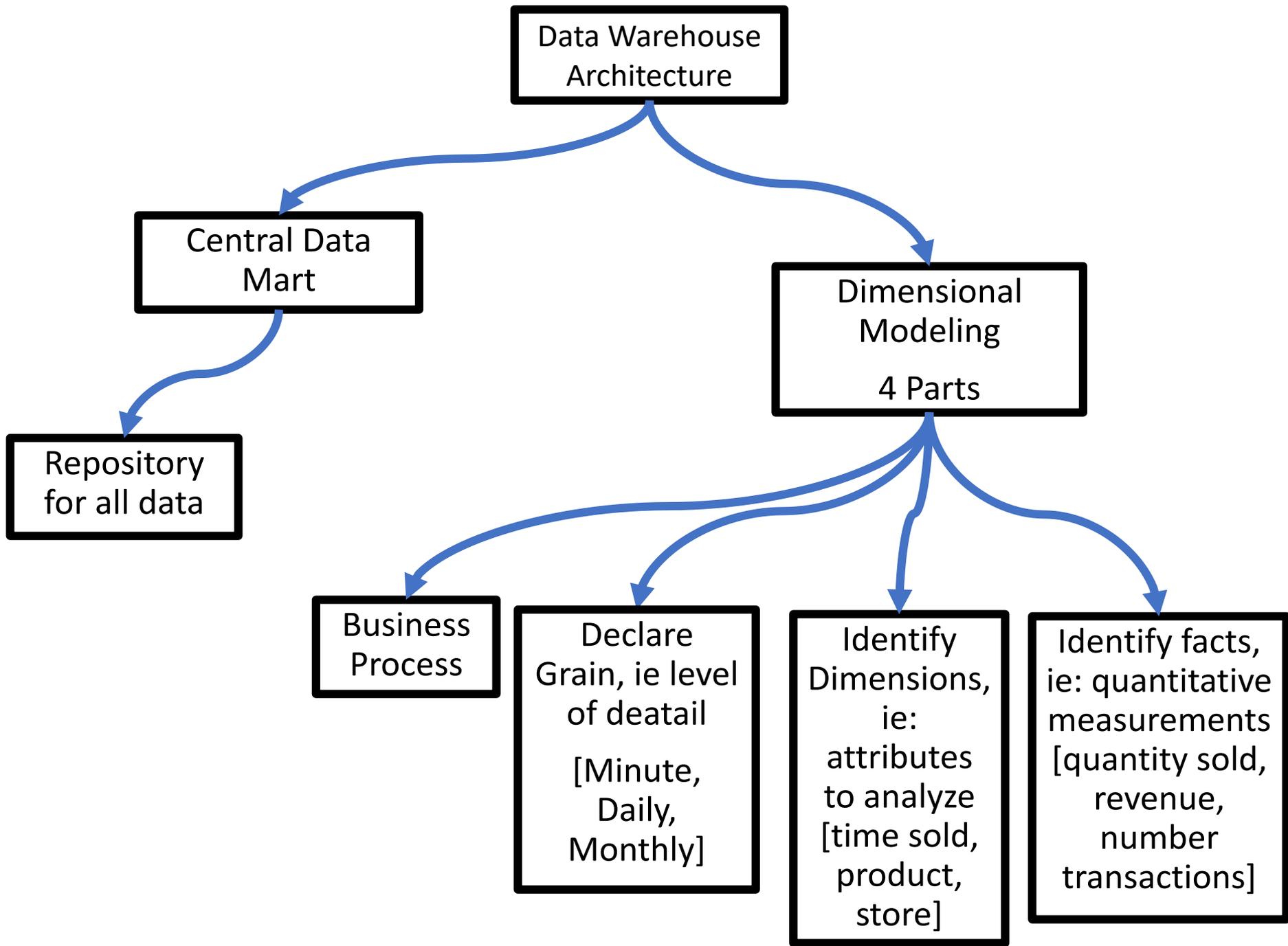
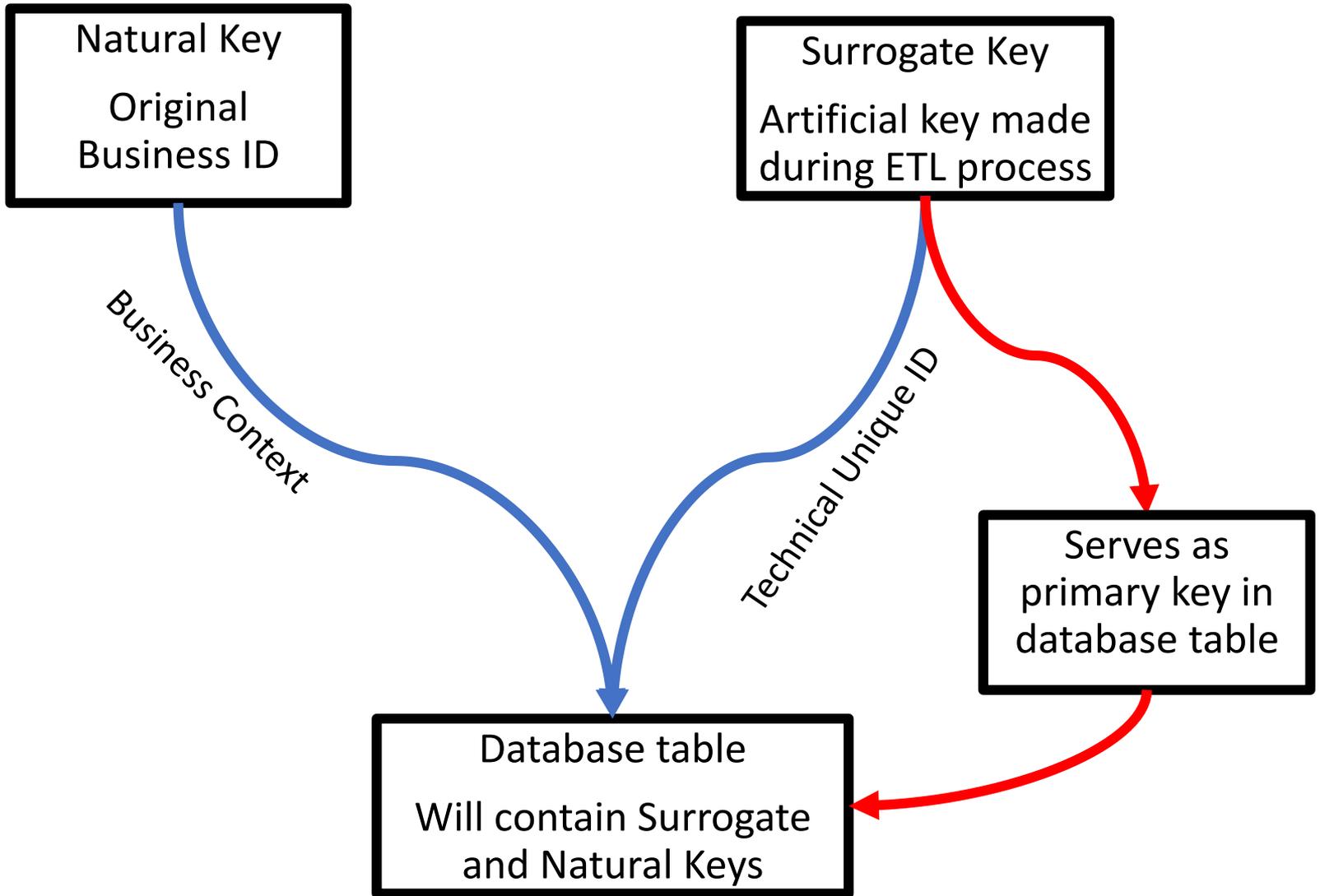


Data Warehouse Midterm Review



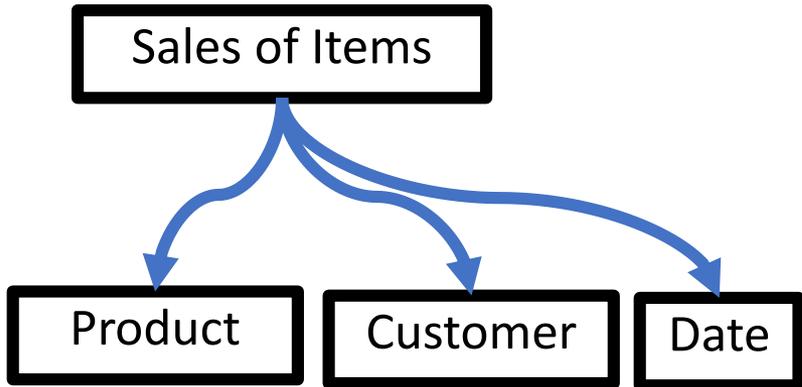


Natural Key vs Surrogate Key



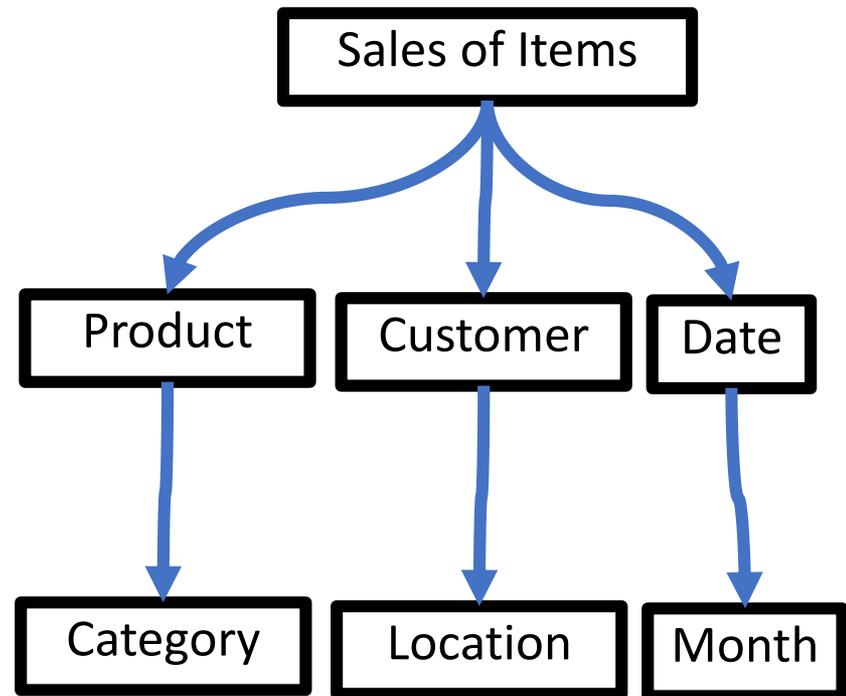
Star Schema vs Snowflake Schema

Star Schema



- Faster queries
- Fewer joins
- Simpler SQL commands
- Better for BI tools

Snowflake Schema



- Uses less storage
- Better integrity
- Easier Updates
- Better for overwrites, new rows, etc.

Business Intelligence
Framework

```
graph TD; A["Business Intelligence  
Framework"] --> B["Business  
“Why?”  
Business  
Alignment/Value"]; A --> C["Technical  
“How?”  
Capabilities/  
Architecture"]; A --> D["Implementation  
“What?”  
Data  
Sources/Integration/  
Storage"]; A --> E["Governance  
“Who/When?”  
Management and  
control"];
```

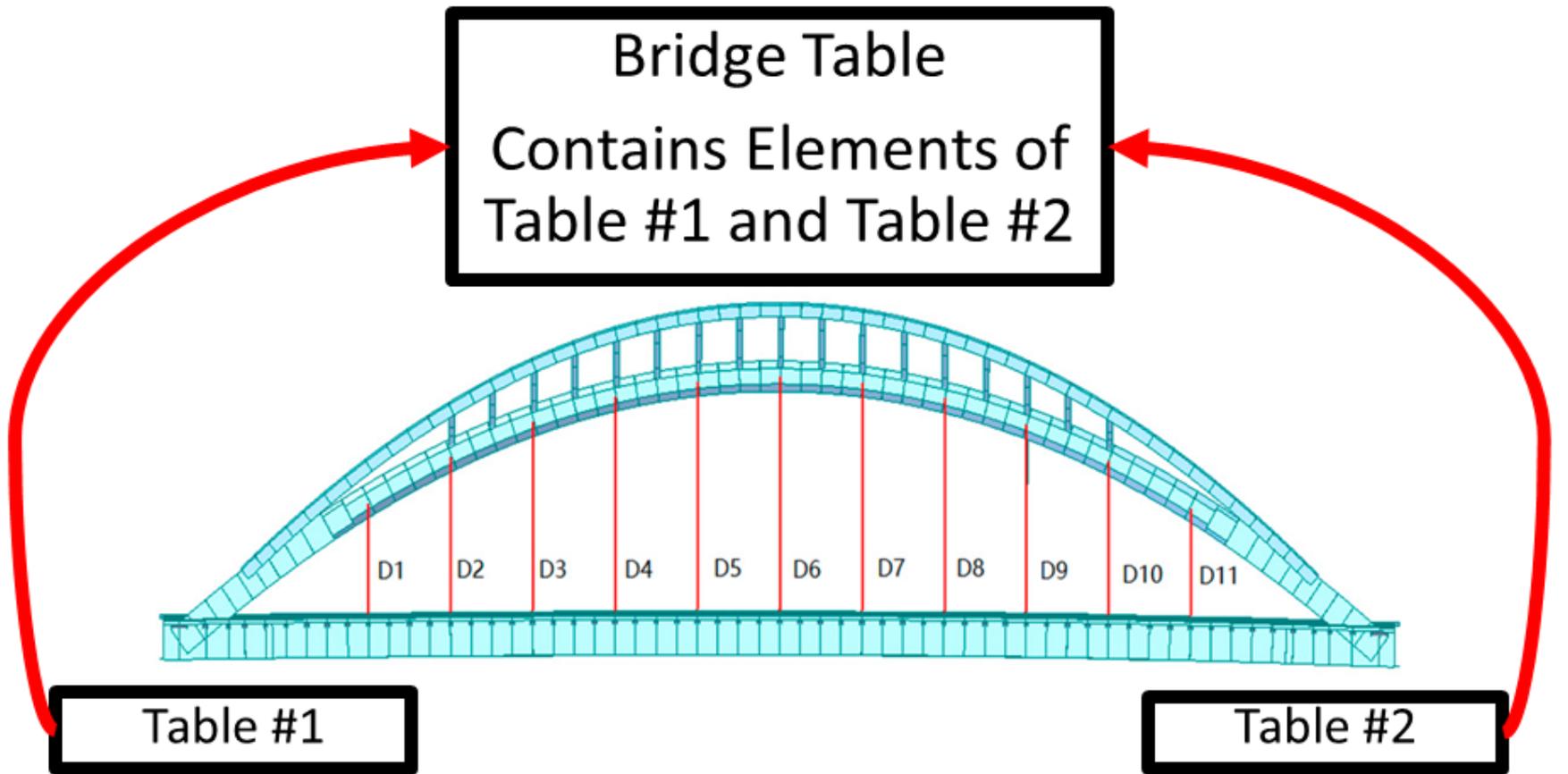
Business
“Why?”
Business
Alignment/Value

Technical
“How?”
Capabilities/
Architecture

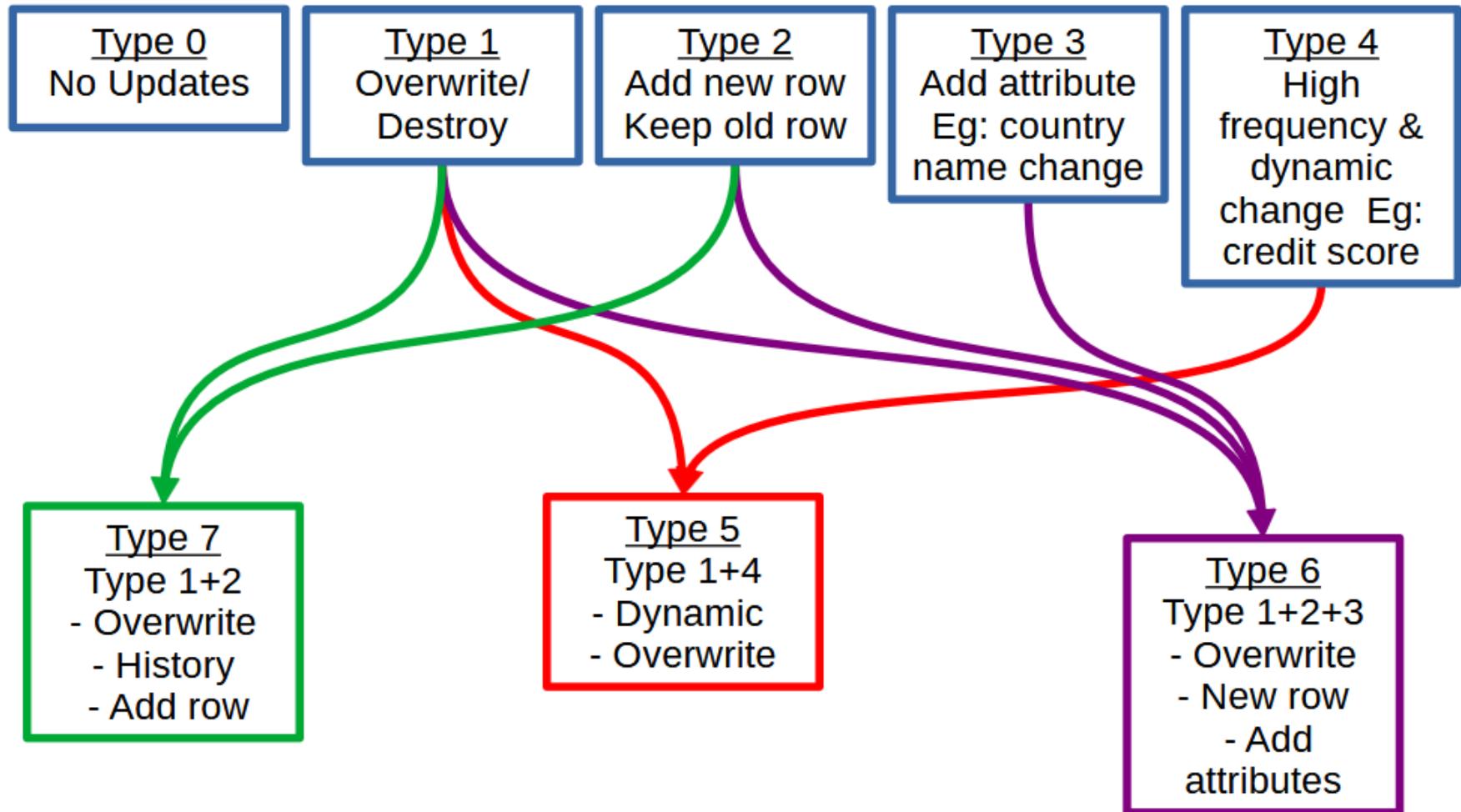
Implementation
“What?”
Data
Sources/Integration/
Storage

Governance
“Who/When?”
Management and
control

Bridge Table Graphic

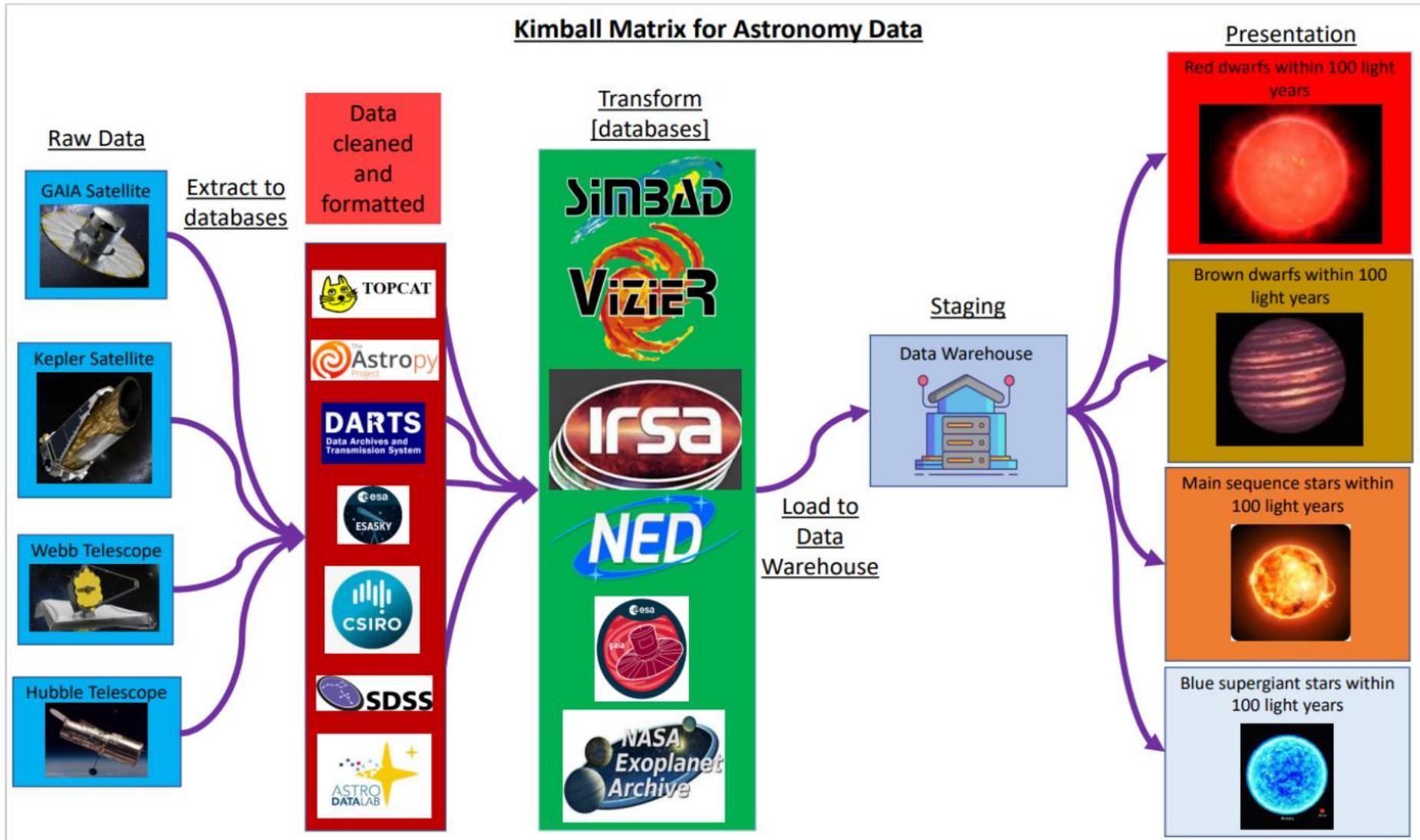


Data Warehouse History Slowly Changing Dimensions

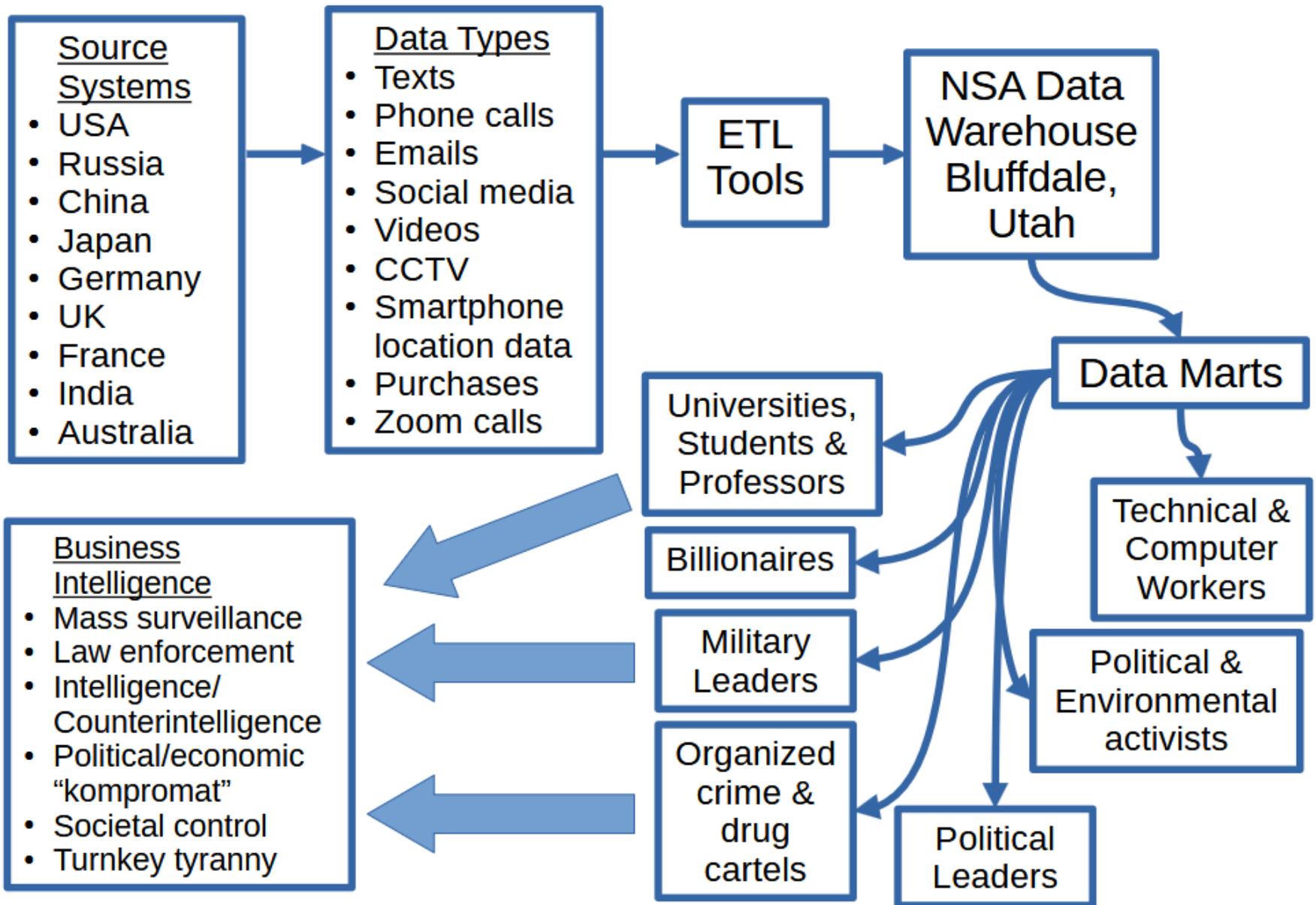


ETL Process

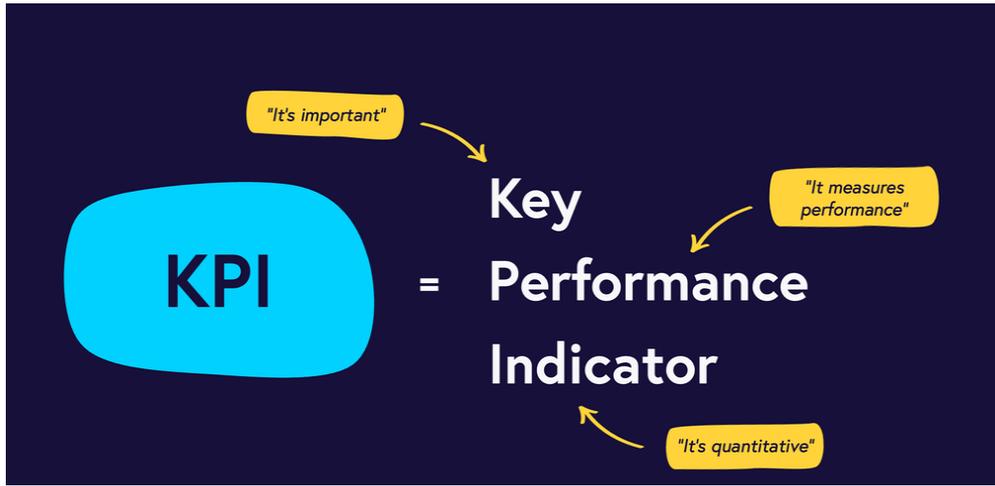
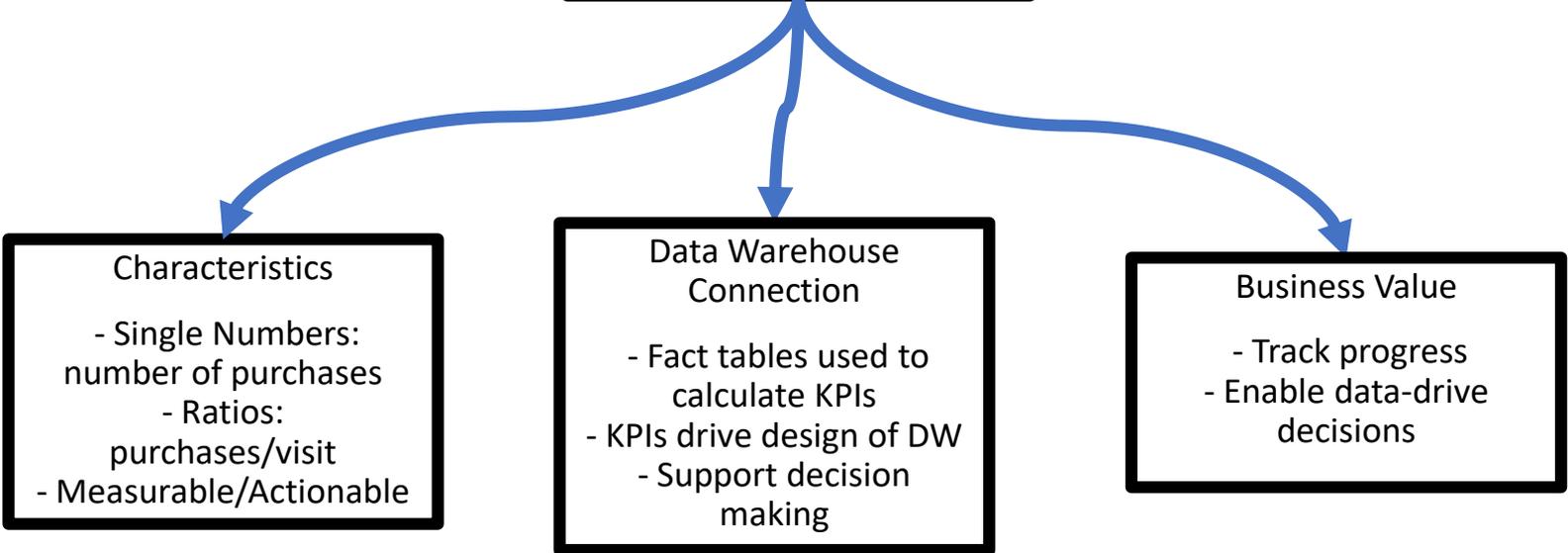
Extract, Transform and Load



National Security Agency Data Gathering Using Kimball Matrix



Key Performance Indicators (KPIs)
Quantifiable Measurements



**Conformed
Dimensions
Standardized Data**

Date Dimension:
Month, Quarter

Customer
Dimension:
ID, Name

Product Dimension:
ID, UPC #

Sales

Inventory

Marketing

Consistent across all data marts

Things to Write Out 25 Times to Encode in Memory (Memory Reinforcement)

1. OLTP (Online Transaction Processing) - real-time operational data
2. OLAP (Online Analytical Processing) - data analysis; business insights
3. Central Data Mart - Repository for all data; unified source of truth
4. Dimensional Modeling (4 Parts):
 - Business Process/activity
 - Grain/level of detail
 - Dimensions/attributes to analyze: time, product
 - Fact/measurable quantities: sales, revenue
5. Natural Key - Original business identifier
6. Surrogate Key - Artificial technical identifier created during ETL
7. Primary Key – the Surrogate Key serves as Primary Key in database table
8. Star Schema - model with fact table connected directly to dimension tables
 - Benefits: Faster queries, fewer joins
9. Snowflake Schema - Normalized dimensional model with hierarchical tables
 - Benefits: Less storage, easier updates, better for data changes
10. Business Intelligence Framework:
 - Business: Why? Value
 - Technical: How? Capabilities
 - Implementation: What? Sources/Storage
 - Governance: Who? Management
11. KPIs – Key Performance Indicators (Key: Important Performance: Measurable Indicators: Quantitative)
12. ETL Process – Extract, Transform and Load data
13. Conformed dimensions: standardized data across all data marts