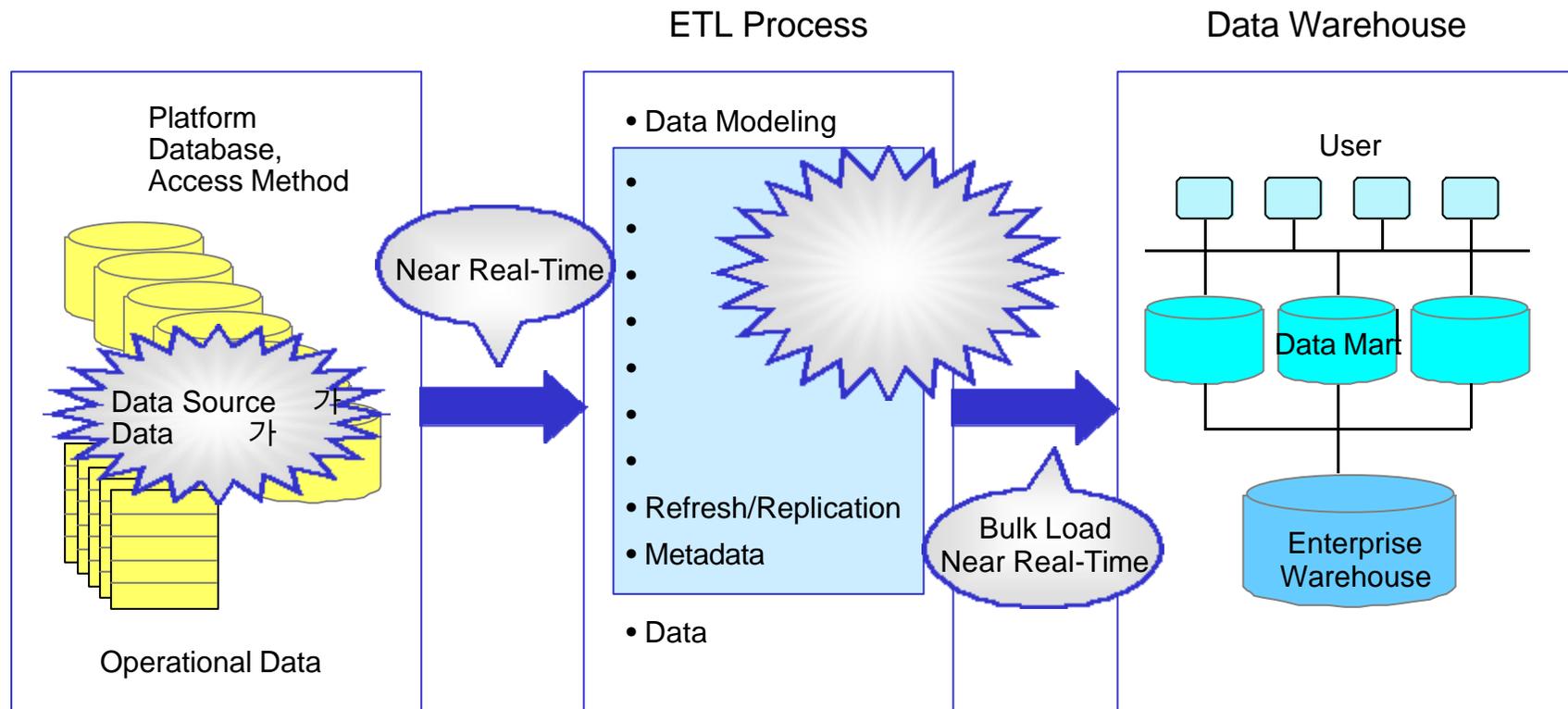


# ETL Best Practice

IK

## ETL

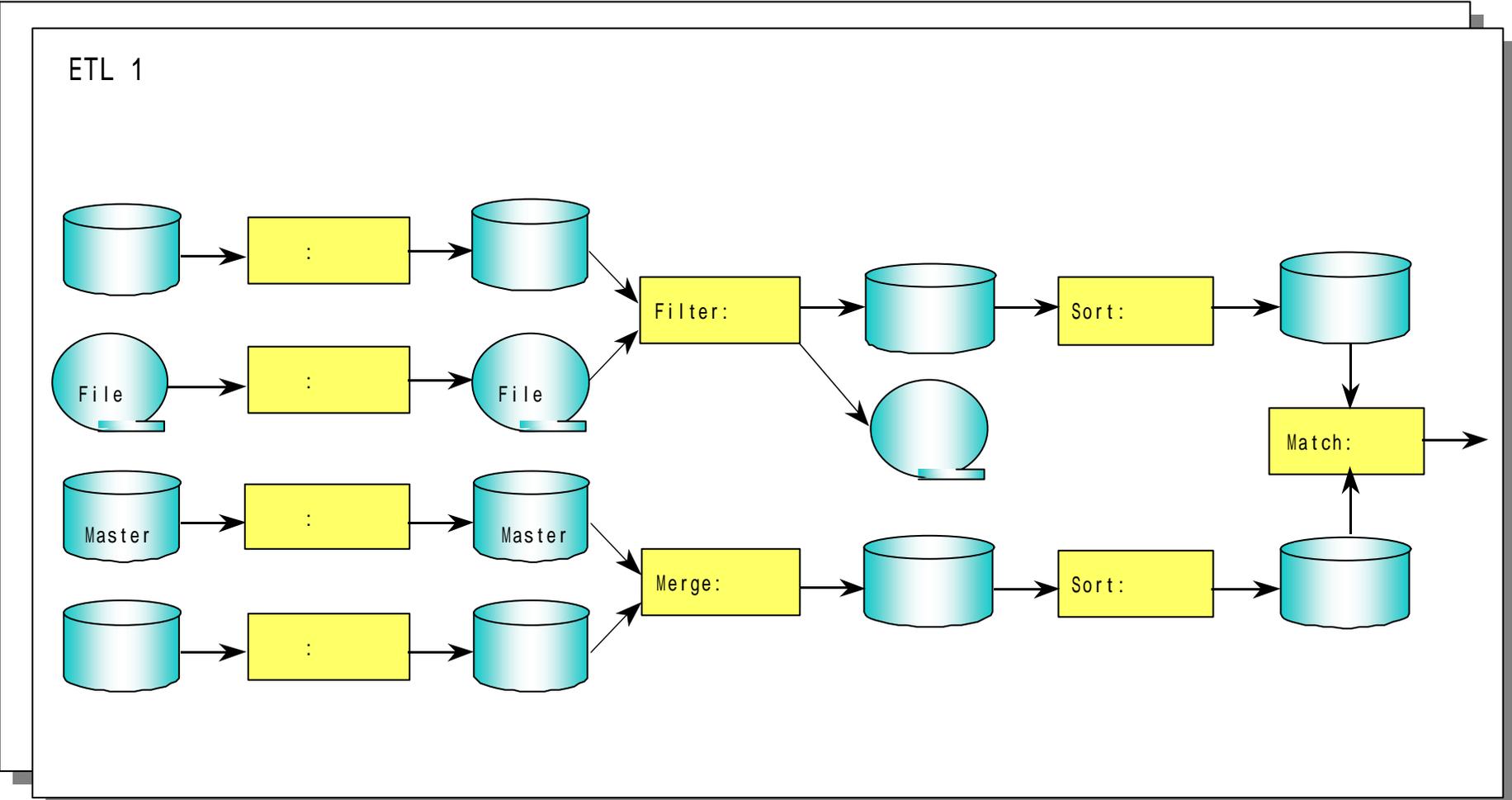
- Data Source, Data Warehouse
- Data Source, Data Warehouse
- ETL tool/system: ETL



ETL

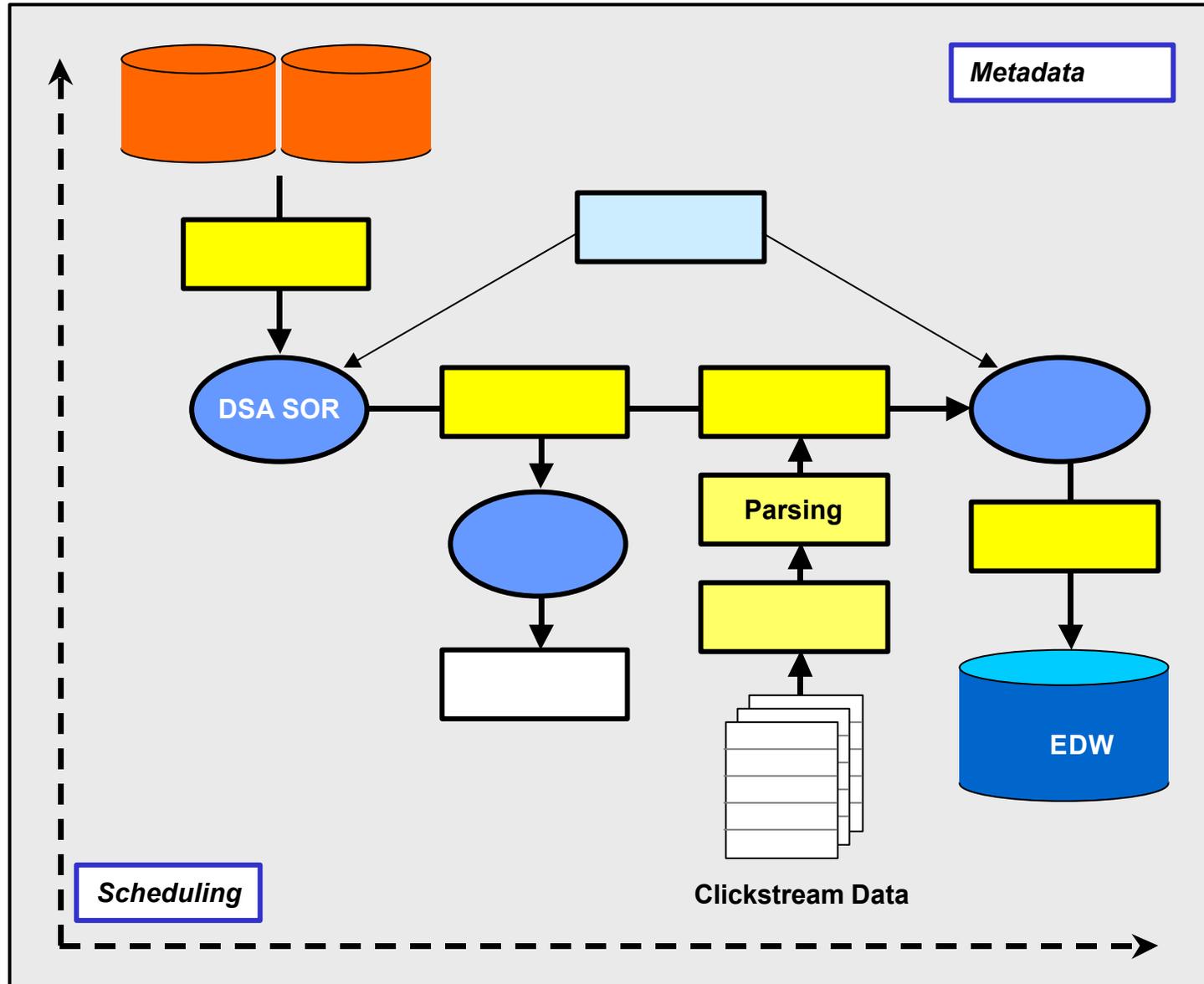
- ✓ SoR(System of Record) Table/File, Column/ 가?
- ✓ Target Column 가?
- ✓ , 가?
- ✓ Transformation 가 가?
- ✓ Process Flow Cleansing, Transformation 가?
- ✓ Metadata 가?
- ✓ Load/Refresh 가?
- ✓ 가?
- ✓ 가?
- ✓ Transformation 가?
- ✓ 가?
- ✓ 가?

## ETL Process Flow



# ETL Process

- ETL Process
- 
- 
- 
- 
- 
- Scheduling
- Metadata



- System of Record (SoR)

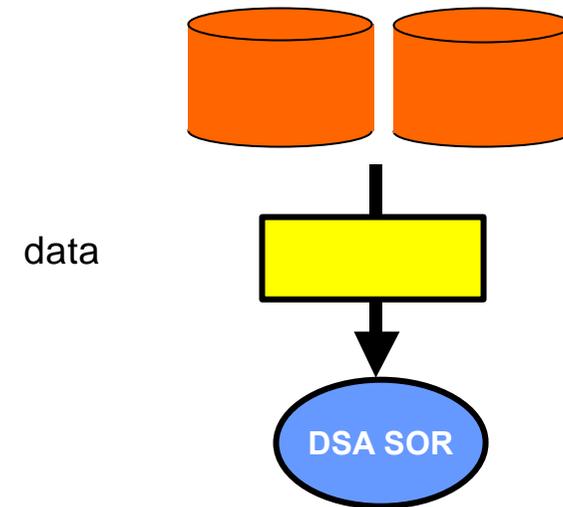
- data source
- ,
- data source

- : SoR

- data format
- data type
- concatenated field sub-field
- field
- key
- target database loading utility input file format
- utility/tool (Fast Unload )
- source system table access , 가

- 

- change data capture
- time stamp
- 



- (data cleansing)

- dirty (noise) data
- dirty data

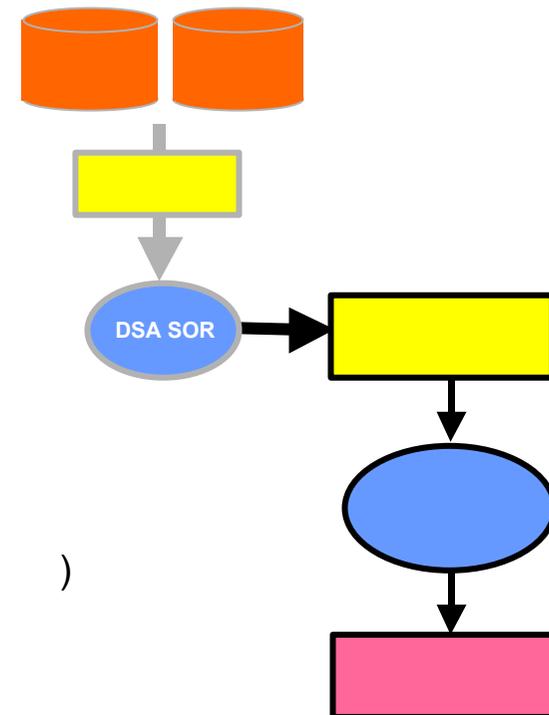
, noise . -

- 

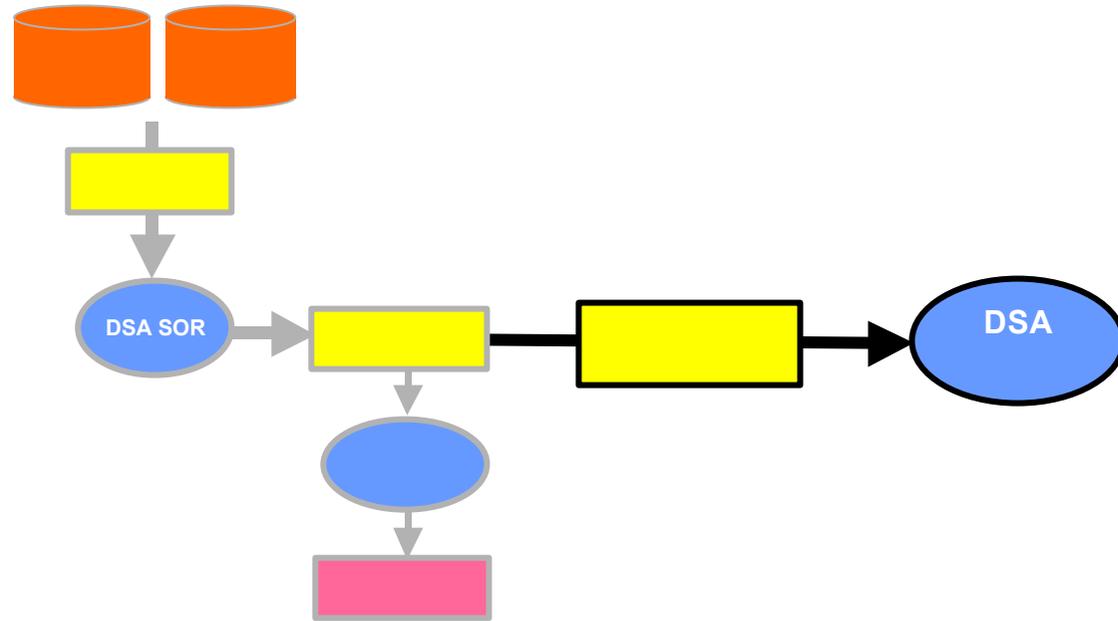
- data data
- data
- 가

- 

- |                       |                           |
|-----------------------|---------------------------|
| ▪ dummy default value | ▪ unique identifier       |
| ▪ missing value(null) | ▪ data                    |
| ▪ data                | ▪ data value ( , )        |
| ▪ field               | ▪ domain ( , data )       |
| ▪ ( ) data            | ▪ data                    |
| ▪ data value          | ▪ (referential integrity) |
| ▪ business rule       | ▪ matching                |
| ▪ primary key /       |                           |



- - source
  - source
- - match-merge/purge rule
  - row



(transformation)	code set	(selection)	data
		(enrichment)	가
	record sort, merge (grouping)		

- - ETL
  -
  
- - Data Demographics(row count, Distinct Value Count, Value Variance)
  - Domain Violation
  - Aggregation
  - Outer Join Analysis
  - Cartesian Product Analysis
  - Business Rule Validation
  - Report

## Load

- Load
  - data warehouse , data source가

	<ul style="list-style-type: none"> <li>• Incremental Update</li> <li>• Time stamp</li> <li>• CDC(Change Data Capture)</li> <li>• Frequent Load(Store &amp; Forward)</li> <li>• Bulk Load</li> <li>• Near Real-Time</li> <li>• EAI</li> <li>• Peer-to-Peer</li> <li>• Log Audit File</li> <li>• Before Image    After Image</li> <li>•                    (Refresh)</li> </ul>
Operation	<ul style="list-style-type: none"> <li>•                    : Insert, Mass Insert, Update, Delete, Mass Delete</li> <li>• DB (Utility): Check Data, Load, Reorg, Recover</li> </ul>

## Metadata

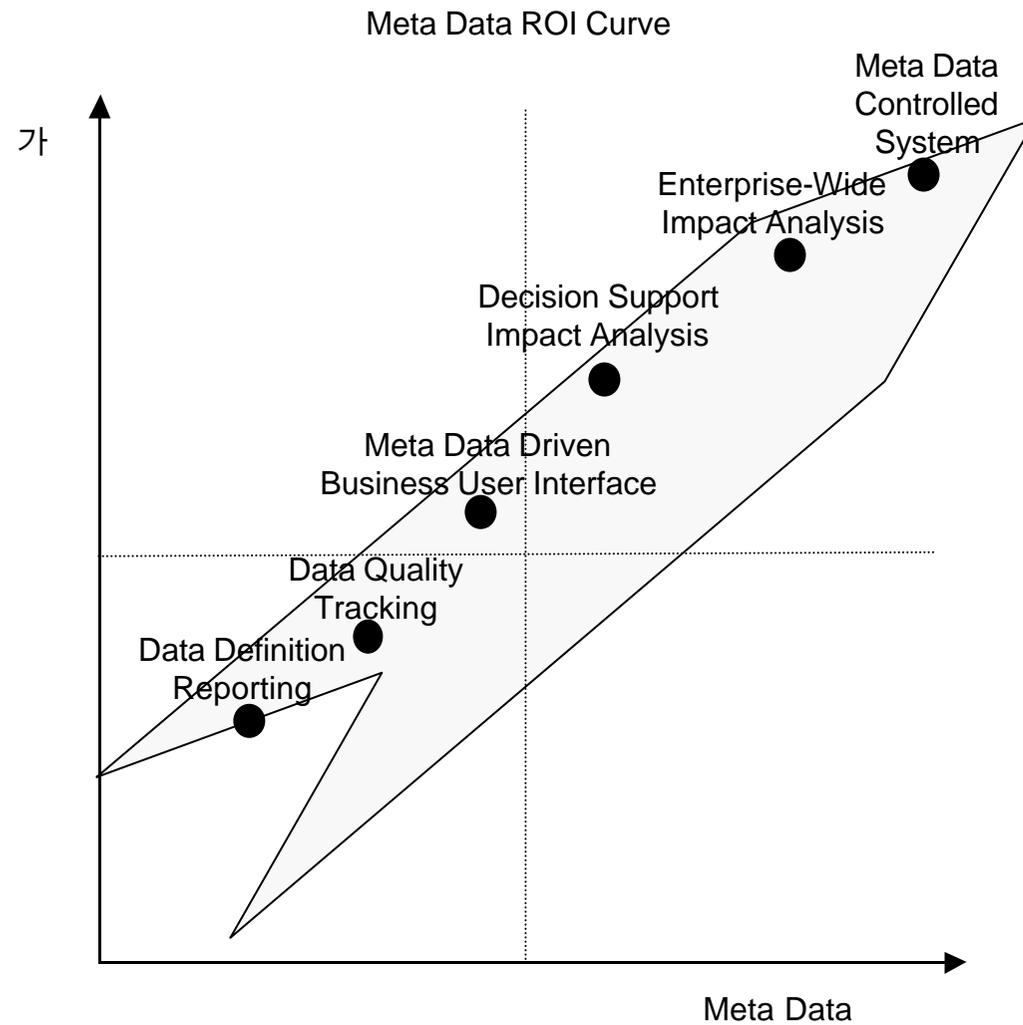
- Metadata
  - ETL metadata
  - metadata 가

Business User	
• Business	Data
• Data	
• Data	Business Rule
• Data	
• Data	, Rule
• Data	Source
• View	
• Data	
• Data	
•	Report
•	Query
• Business	

IT	
• Data Warehouse	
• Column	
• Column	Business Rule
• Column	Transformation/Cleansing Rule
• Column	, Rule
• Column	Source
• Column	Program
• Column	Default
•	
• Column	
• Data Volume	
• Table	
• Metadata	
• Security Rule	

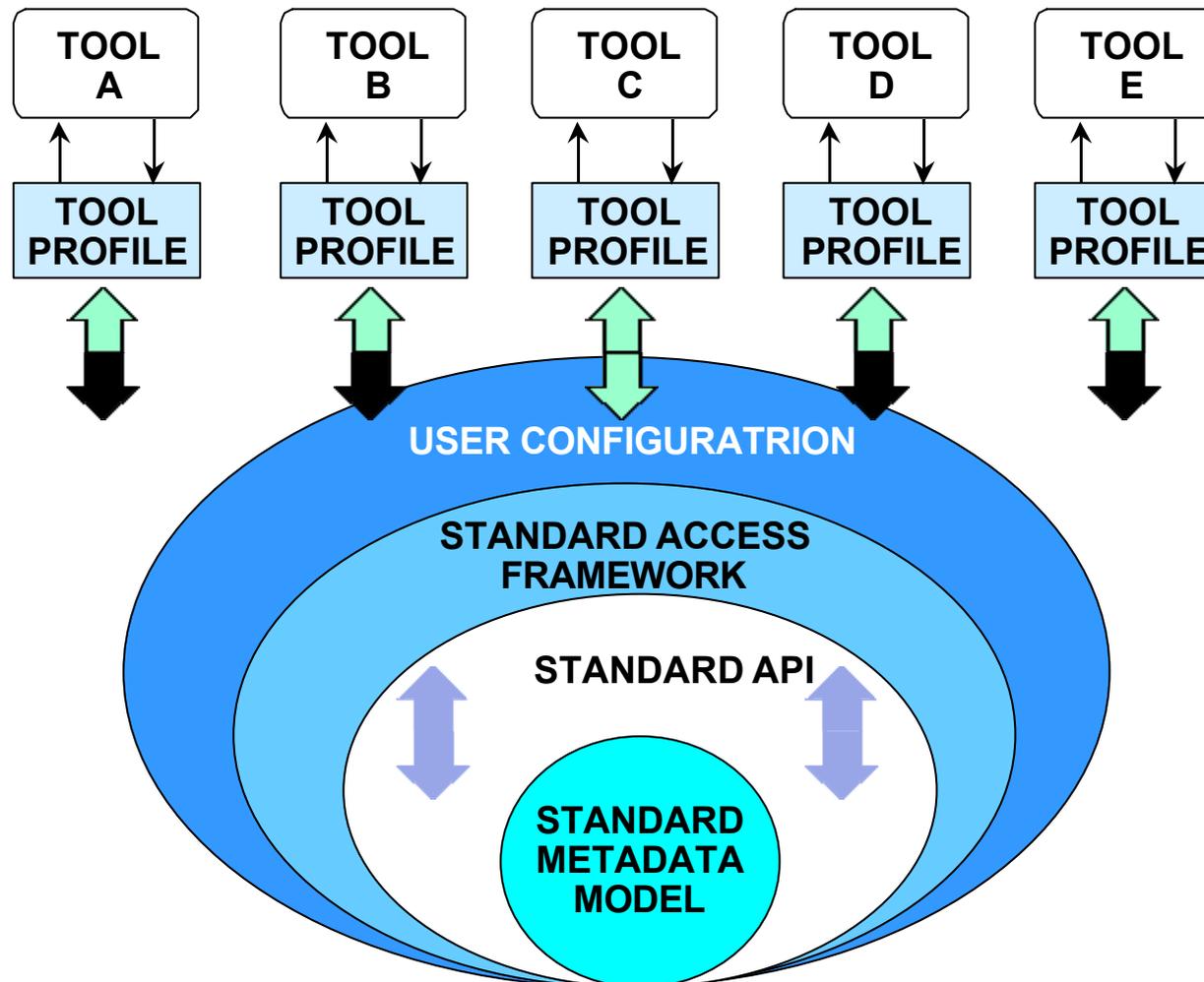
## Metadata

- Metadata 가
  - data definition report
  - data quality tracking
  - metadata driven business user interface
  - decision support impact analysis
  - enterprise wide impact analysis
  - metadata controlled system



## Metadata

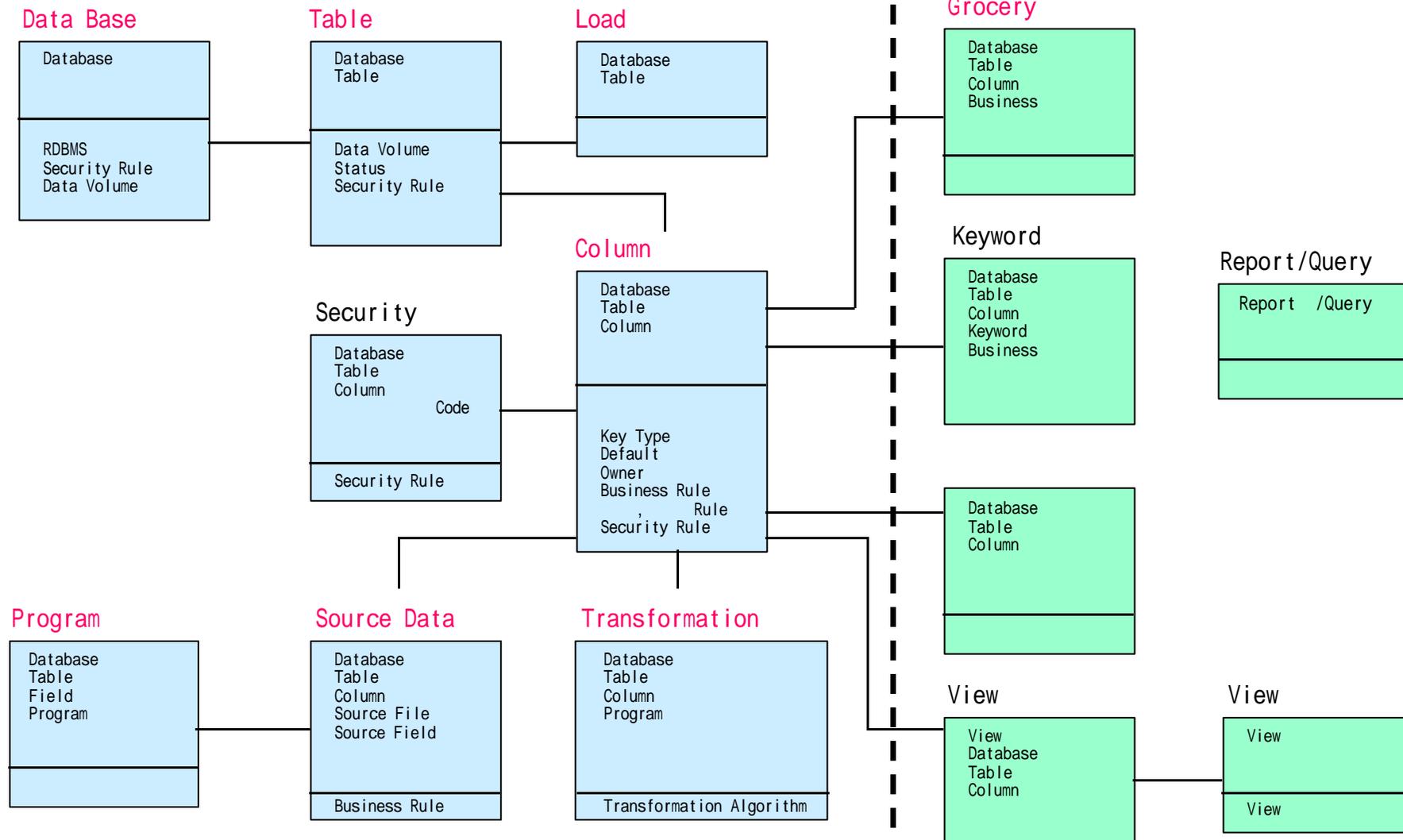
- Metadata : Metadata Exchange Standard Model(Meta Data Council)



## Metadata

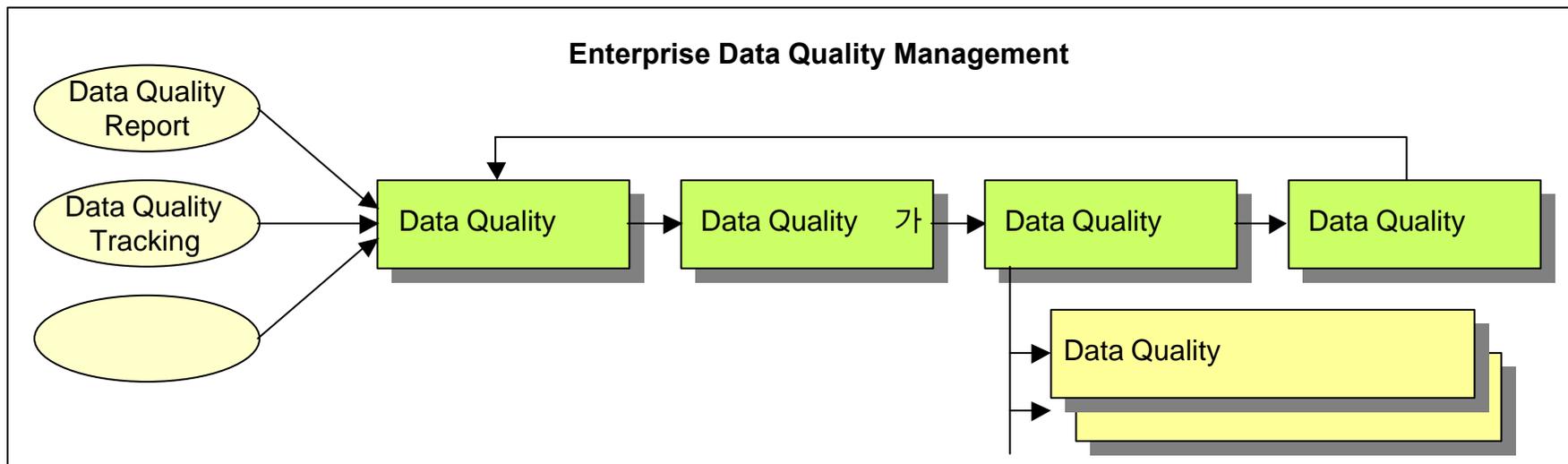
Metadata(Corporate Repository) – View

- Metadata Model

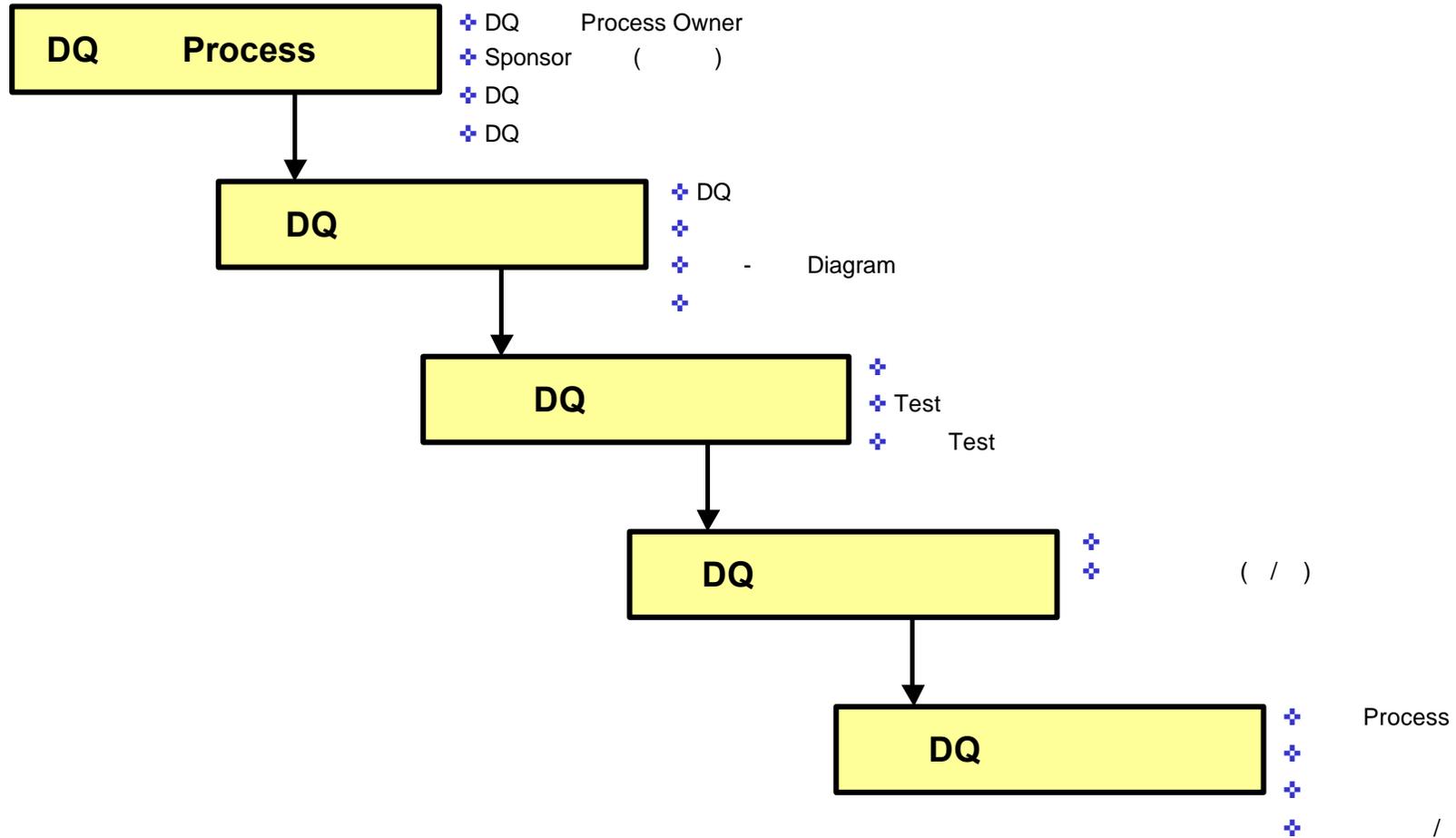


## Data Quality Management

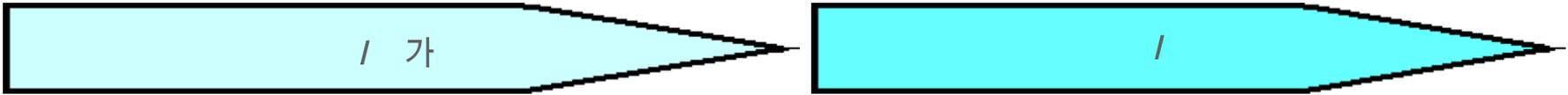
- Data Quality
  - Data Quality : DW, BI
  - , , 가
  - (correction, cleansing) (preventing)
  
- Data Quality Management – 4
  - Data Quality :
  - Data Quality 가: / 가
  - Data Quality : data cleansing, data integration, data enrichment
  - Data Quality : application



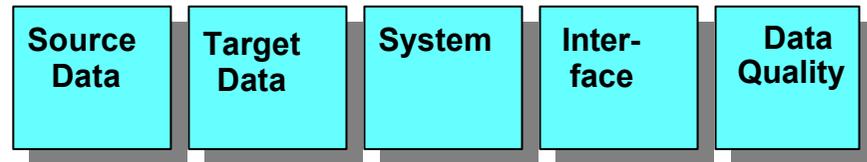
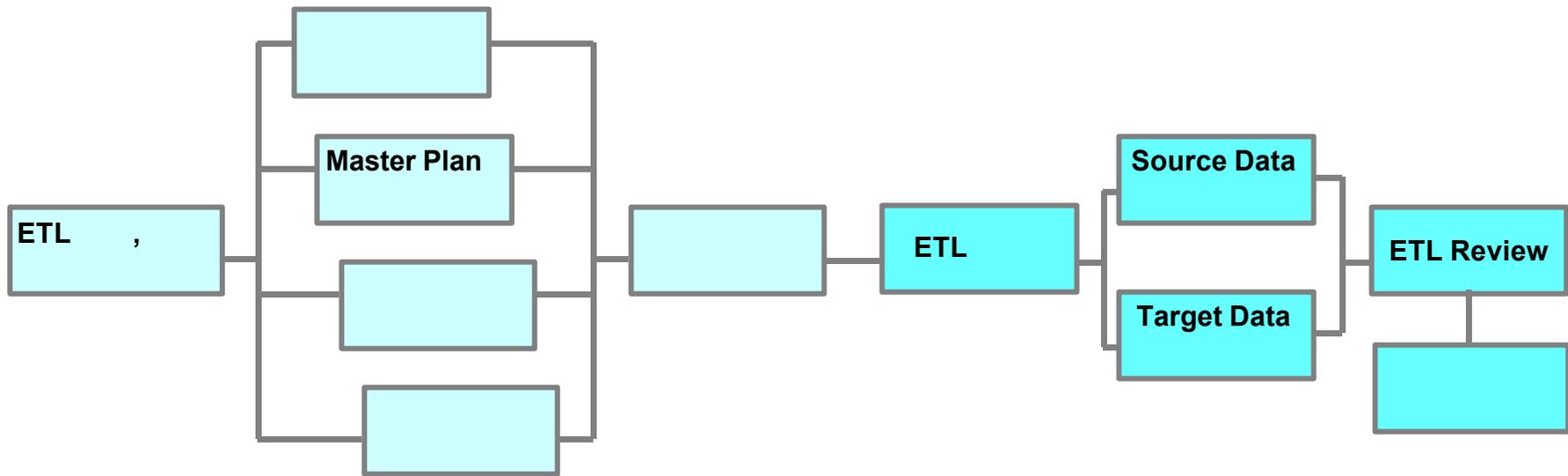
## Data Quality(DQ)



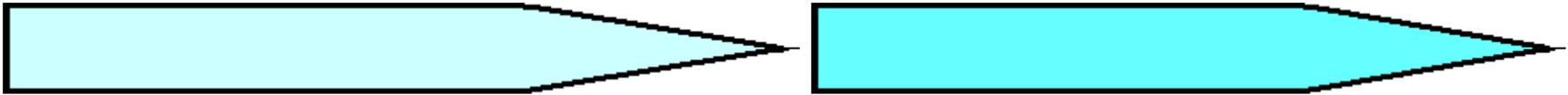
Phase



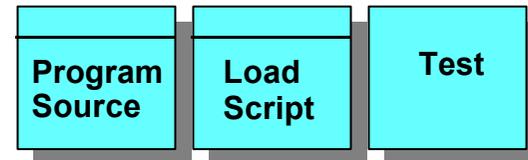
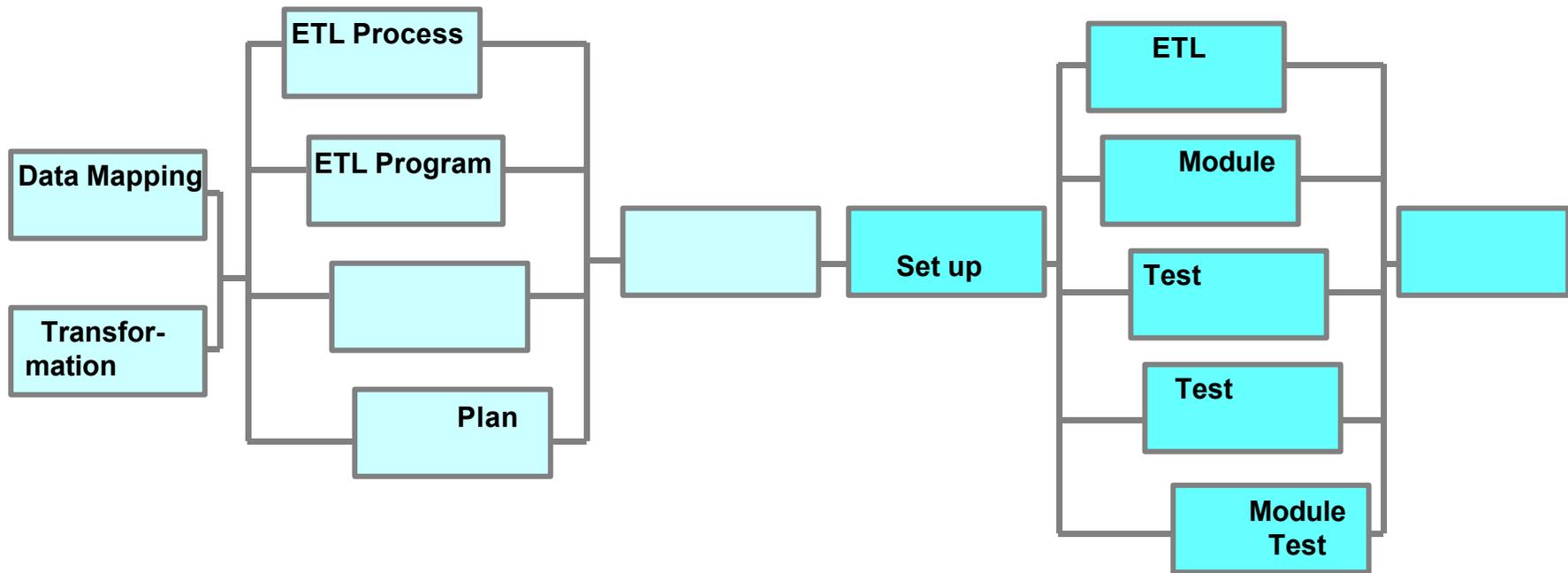
Task



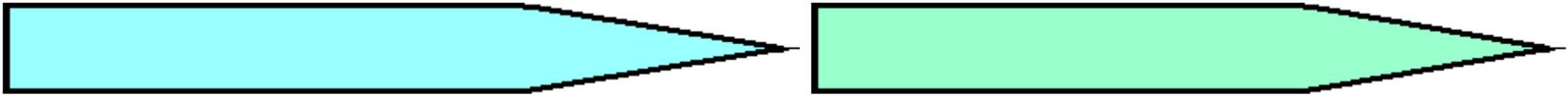
Phase



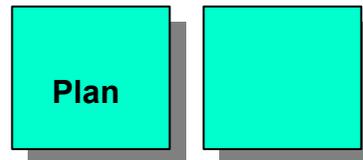
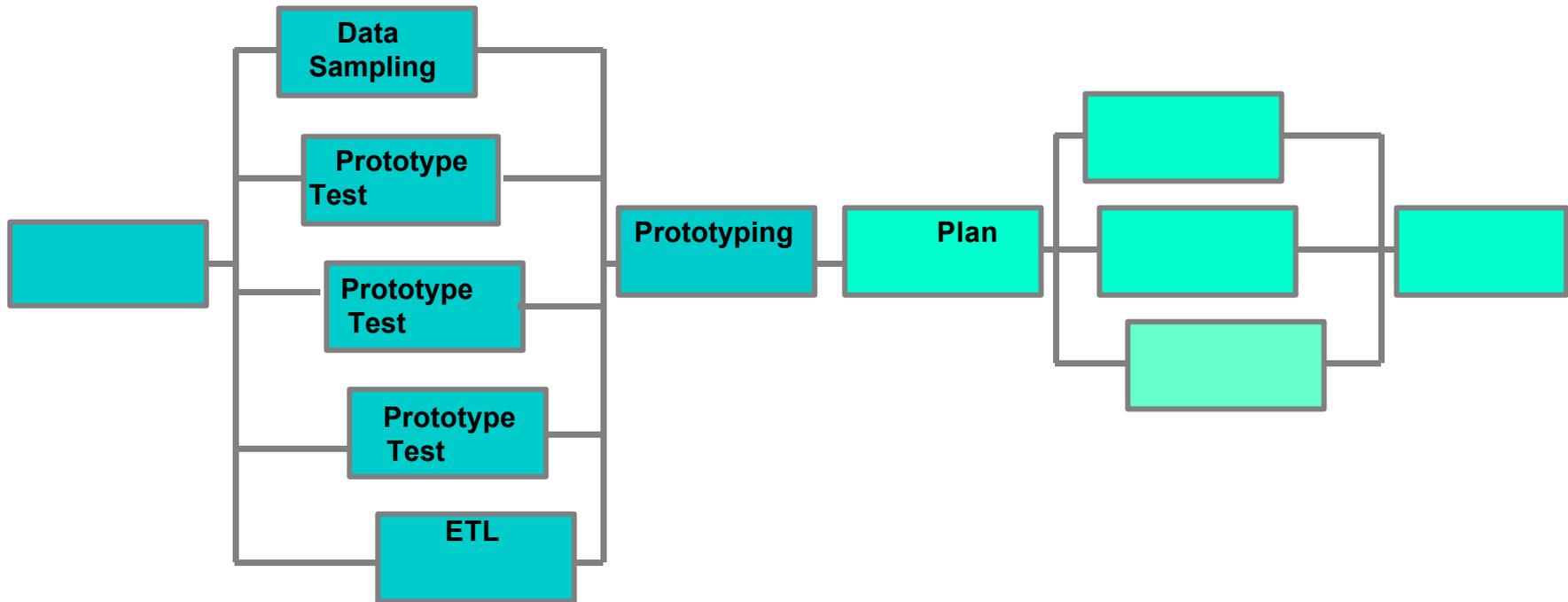
Task



Phase



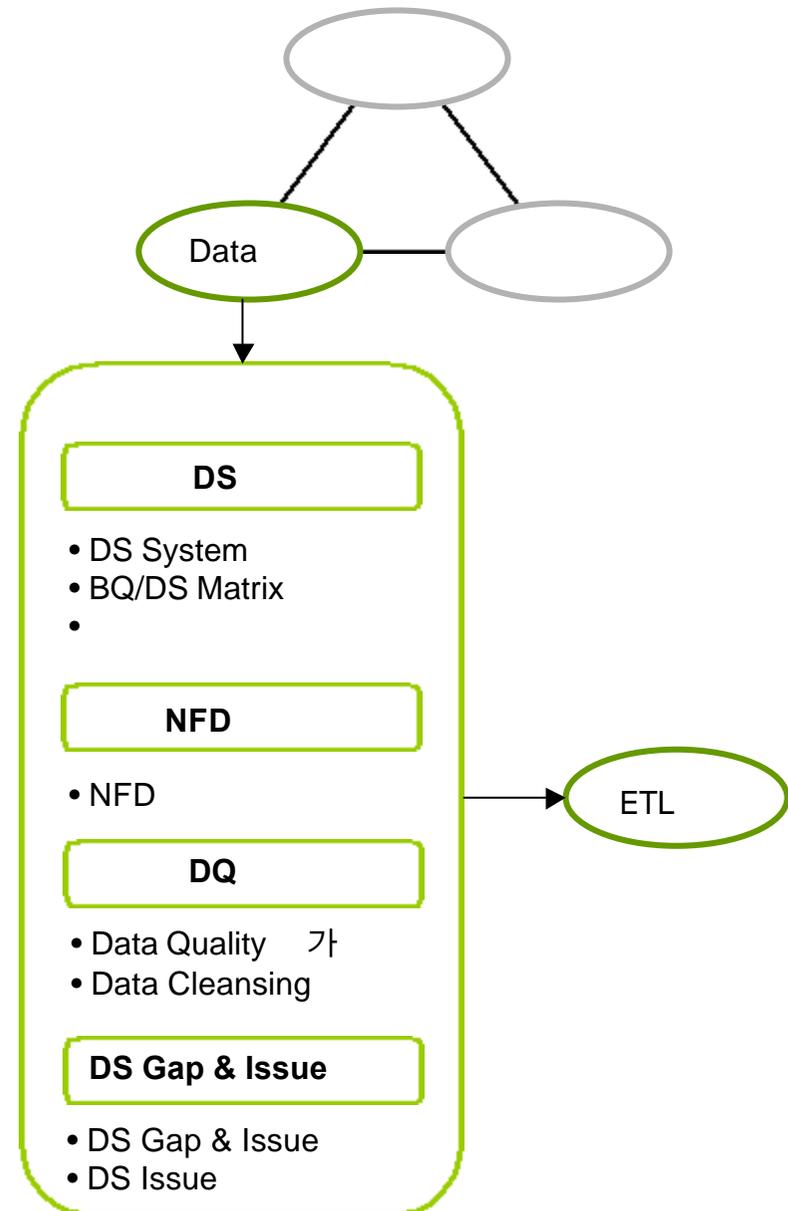
Task



	가	,	,	,	,	가
	ETL  Plan	Target, Source Data, System  Data Mapping, Transformation  Plan 가	Data  Data ETL Program	ETL Program  ETL 가 Module	Sampling Data Prototyping  가 Data Prototyping	ETL  Prototyping 가
<b>H i g h l i g h t</b>	<ul style="list-style-type: none"> <li>• Target Table</li> <li>• Target Column</li> <li>• Target Data</li> <li>• Target DBMS</li> <li>• Source Table</li> <li>• Source Column</li> <li>• Source Data</li> <li>• Source DBMS</li> <li>•</li> <li>• Key Man</li> </ul>	<ul style="list-style-type: none"> <li>• Source</li> <li>• Target</li> <li>• Target, ETL, Source 3 Interface</li> <li>• Source Data</li> <li>• Target Data</li> <li>• Target Data Source Data Mapping</li> <li>• Transformation</li> <li>• Data</li> <li>• Data Load</li> <li>• ETL Process Flow</li> </ul>	<ul style="list-style-type: none"> <li>• 가</li> <li>• Data</li> <li>•</li> <li>• Program</li> <li>• Program</li> <li>• Plan</li> </ul>	<ul style="list-style-type: none"> <li>• ETL ( ) Set Up</li> <li>•</li> <li>• ETL Program</li> <li>• Data Sampling Program</li> <li>• Test</li> <li>• Test</li> </ul>	<ul style="list-style-type: none"> <li>• Plan</li> <li>• Data Sampling</li> <li>• Prototype Test</li> <li>• Prototype Test</li> <li>• ETL</li> <li>• Data ETL</li> </ul>	<ul style="list-style-type: none"> <li>• Plan</li> <li>•</li> <li>•</li> <li>•</li> <li>•</li> </ul>

ETL Best Practice

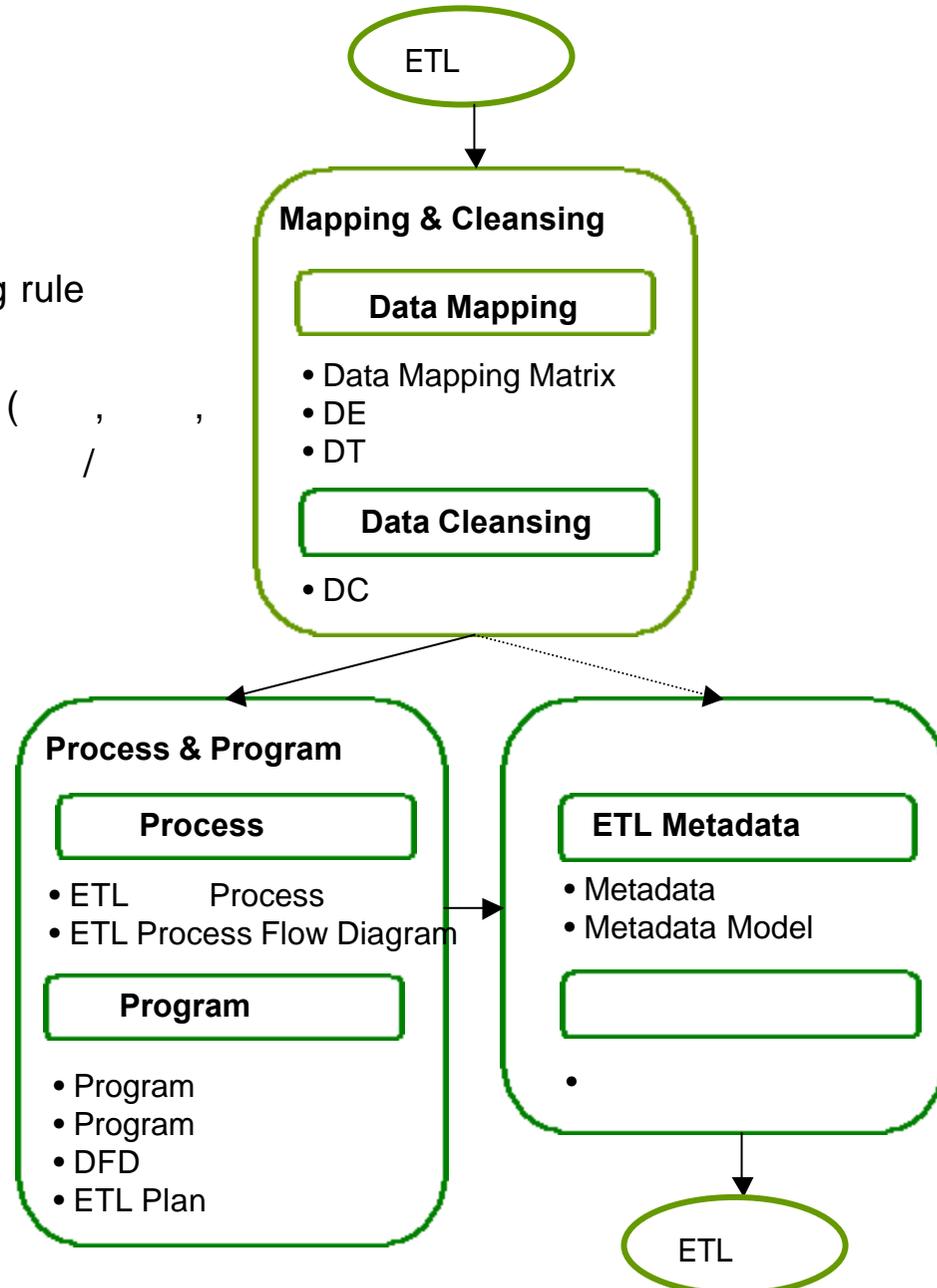
- : data
  - ETL ,
  - ETL ,
  - (business question)
  
- ( 가 )
  - SoR(system of record)
  - BQ(business question) DS(data source) matrix
  - , DT(data transformation) rule
  - NFD(not found data)
  - DQ(data quality)
    - DQ 가
    - data profiling: data(sample)
    - DQ
    - DQ 가
  - DC(data cleansing)
  - DSI(data source gap & issue)
  - ETL review 가 & BQ review



ETL

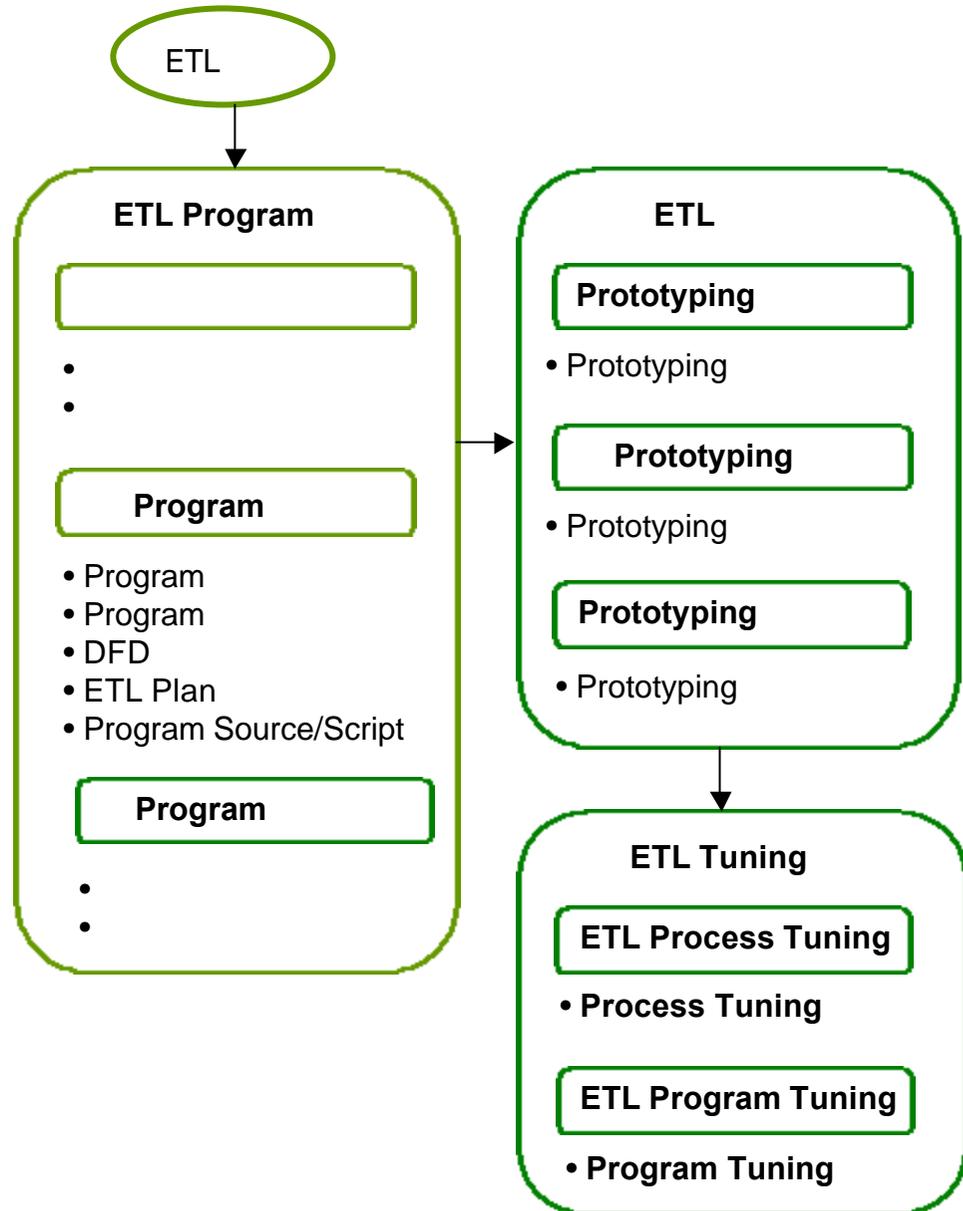
Best Practice

- ETL
  - ETL metadata ( , )
  - ETL metadata ( , )
  - ETL metadata ( , )
- ETL
  - DE(data extract) DT (data transformation) rule
  - DC(data cleansing) rule
  - ETL process
  - ETL program
    - program , DFD(data flow diagram)
    - program
  - ETL metadata , data model
  - ETL

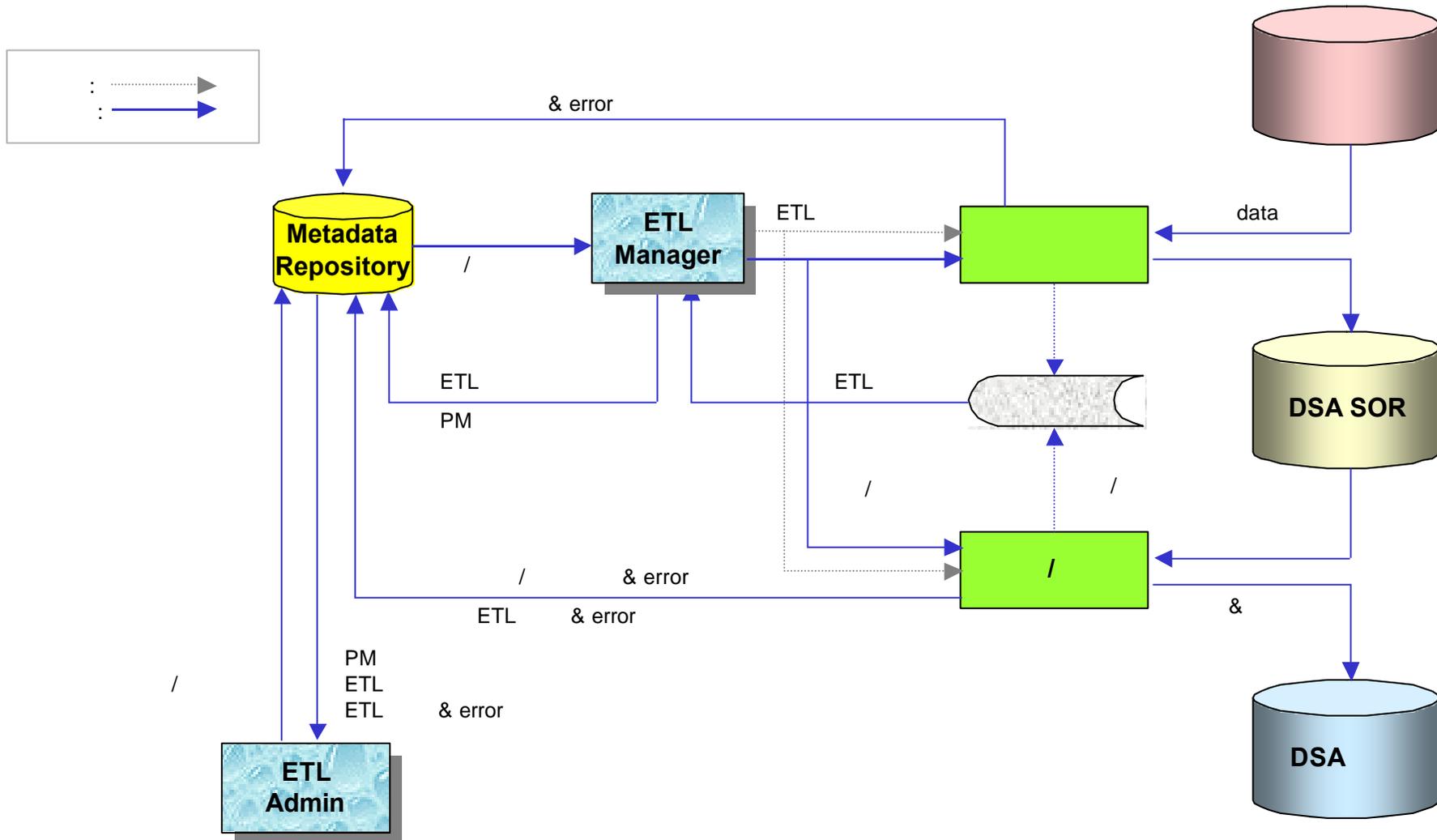


ETL Best Practice

- : process program , prototyping, tuning
  - program dependency ,
  - plan
  - scheduling, metadata
  - ,
- ( 가 )
  - ETL
  - ETL program &
  - ETL (prototyping)
    - 
    - 
    -
  - ETL tuning



## ETL



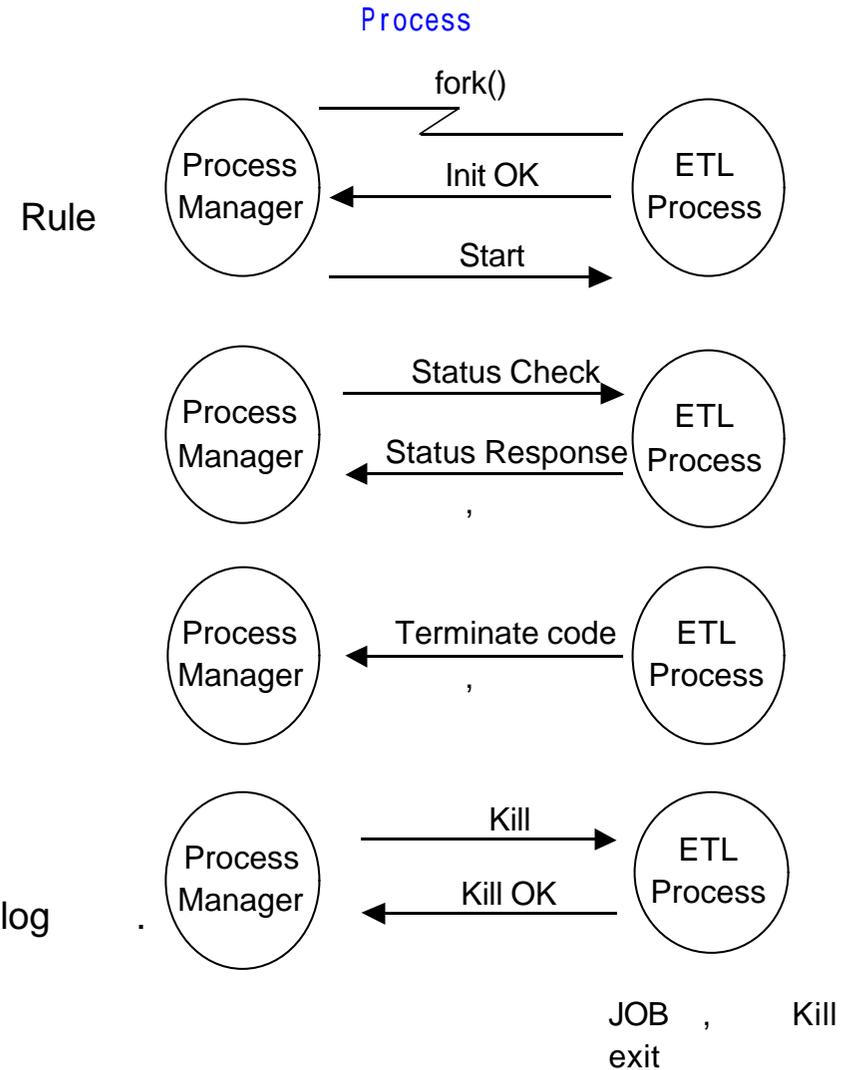
## Scheduler

### ● Process

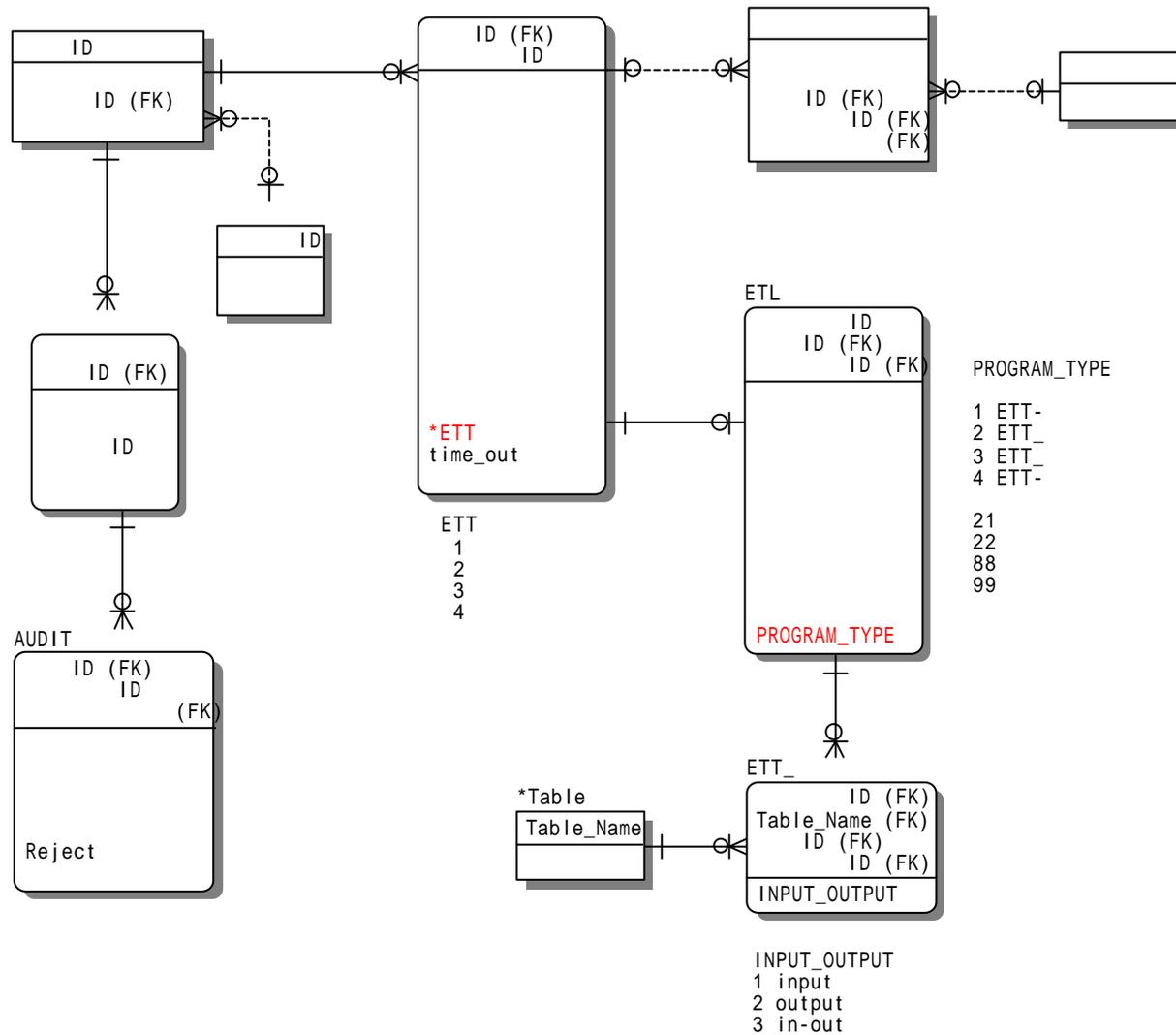
- USER :
- JOB :
- : , , loading
- 
- JOB : JOB
- Message : ETL
- LOG : ETL Logging
- :

### ● Scheduling

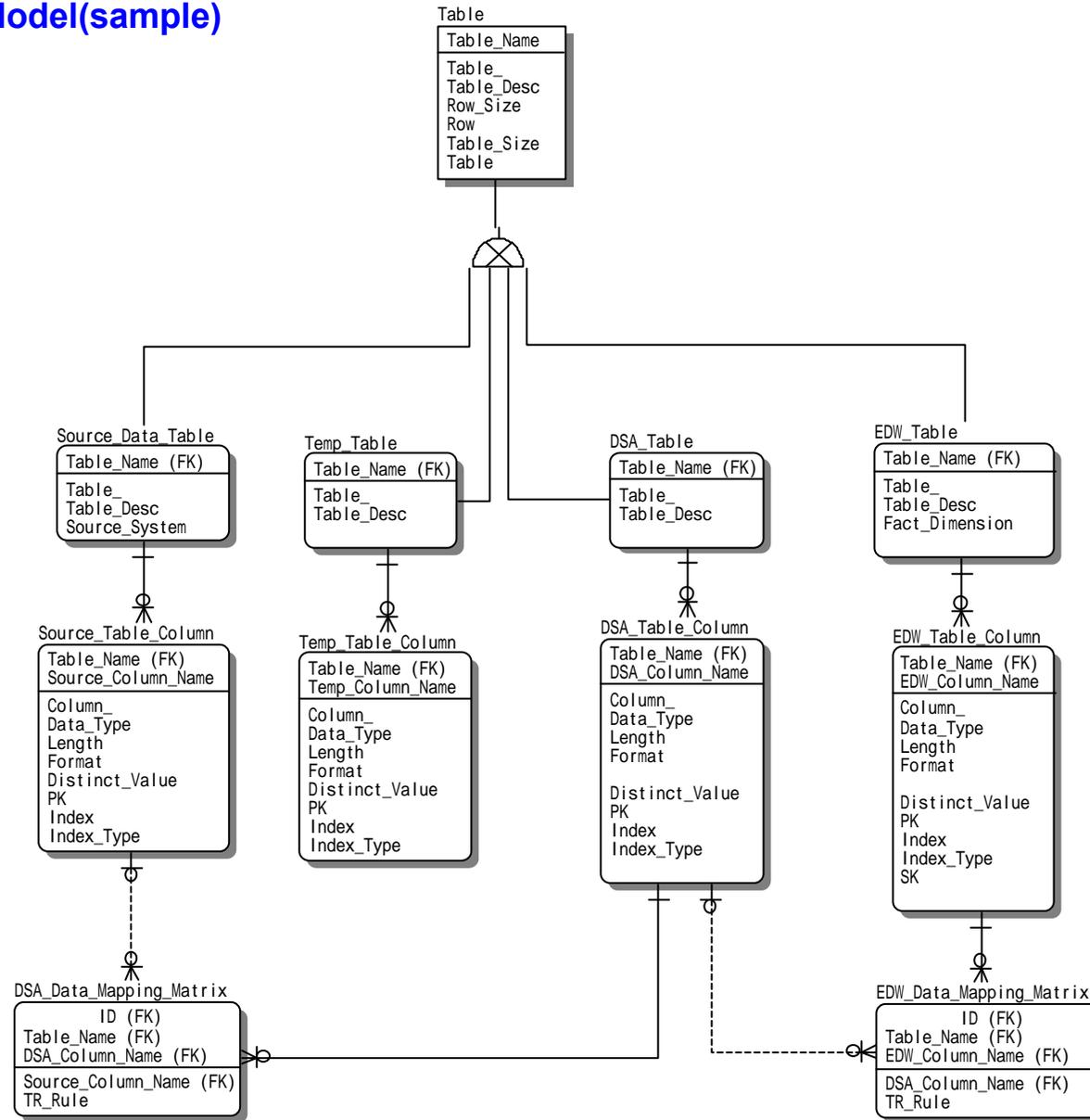
- :
- : JOB , ,
- : JOB
- : JOB



## Metadata – Data Model(sample)

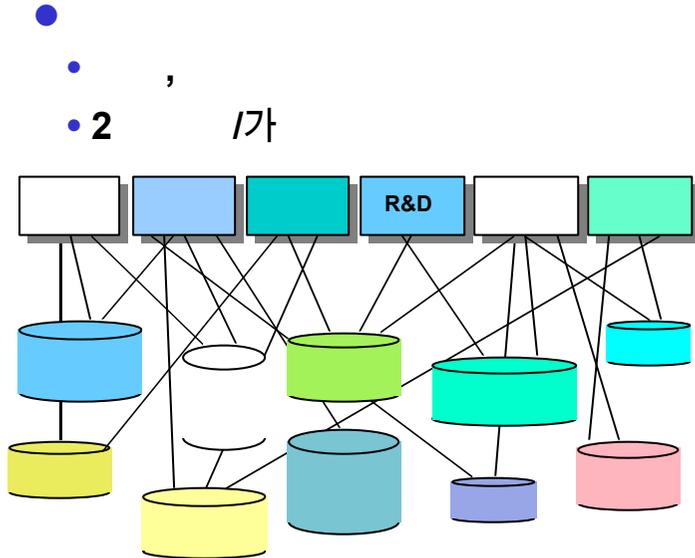


Metadata – Data Model(sample)

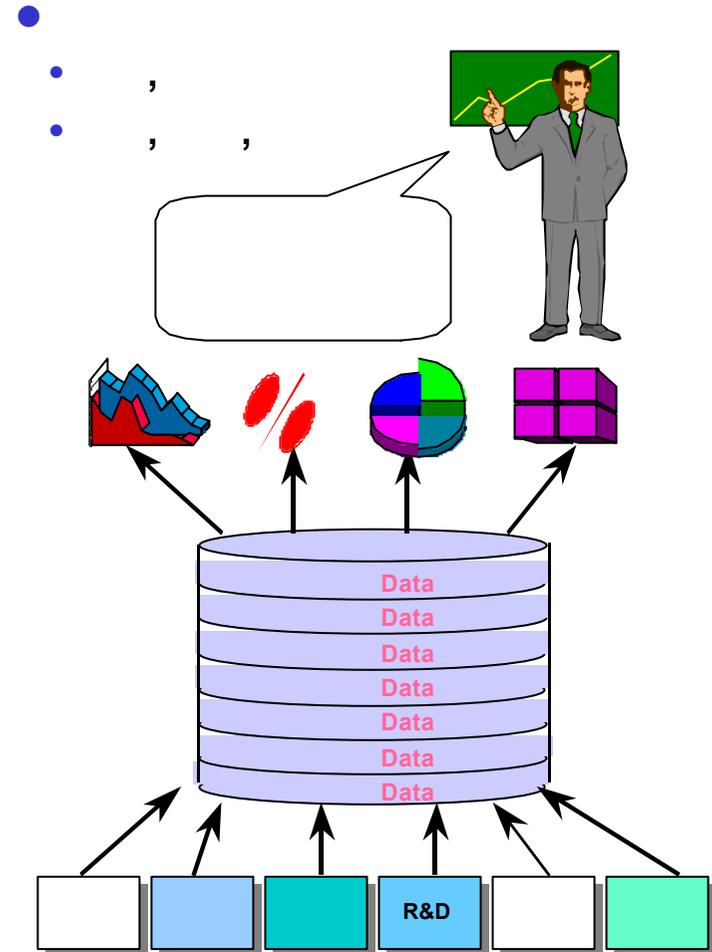
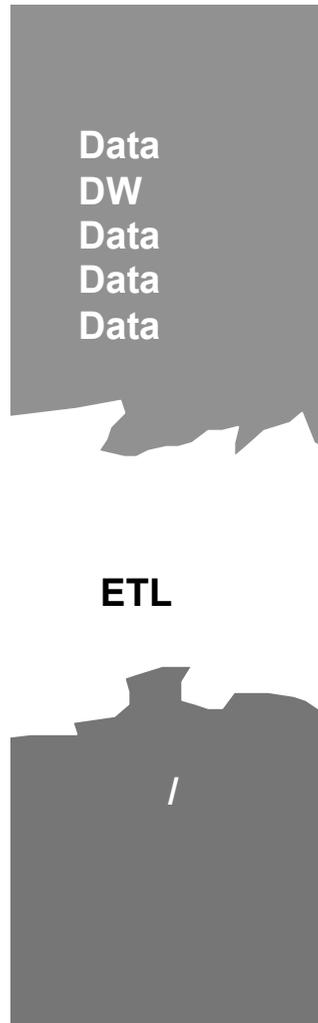


ETL :

version ( )

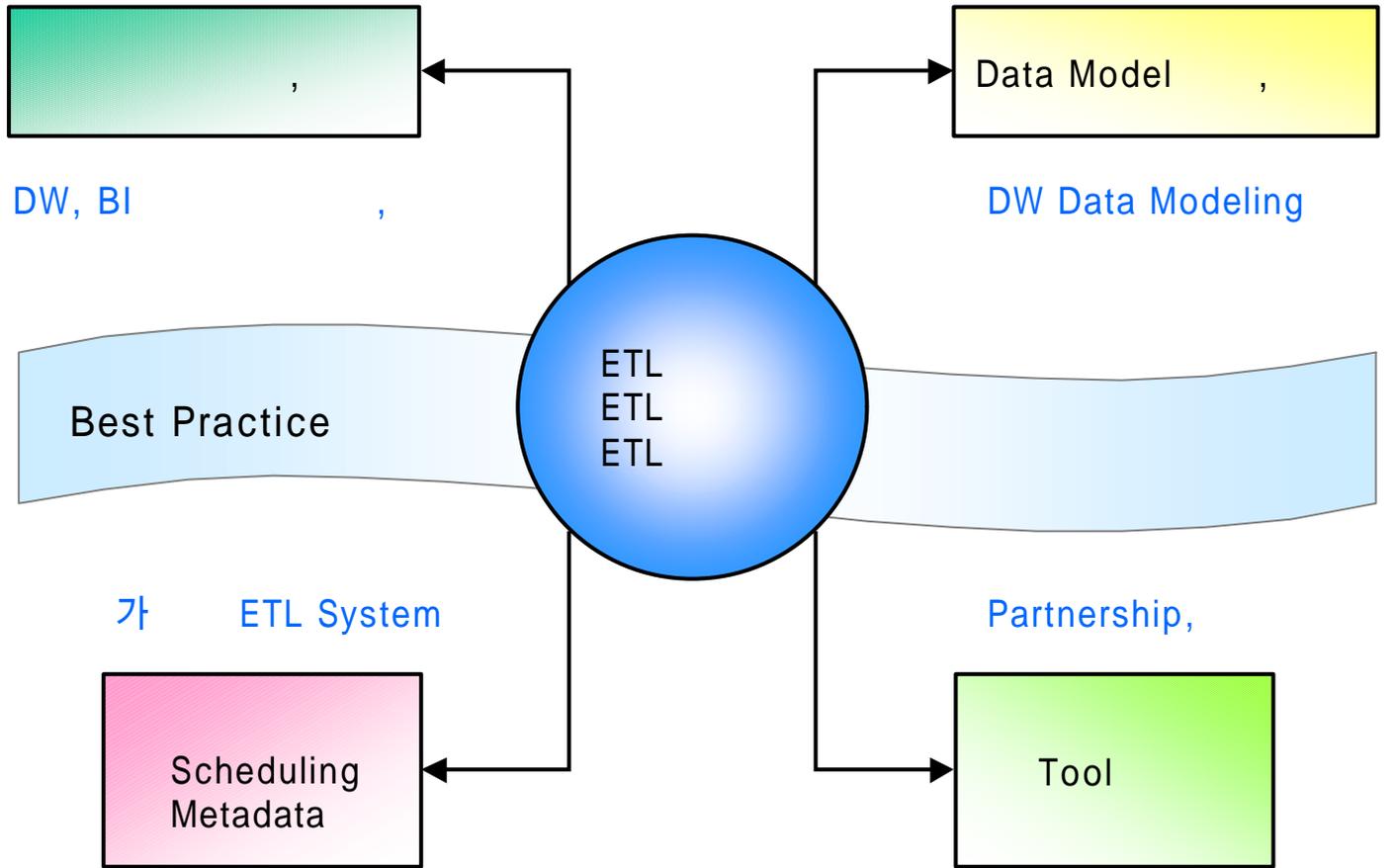


Operational System



Business Intelligence

IK



# Contact Us

Tel: 02-3775-2890

Email: [con@iksolutions.com](mailto:con@iksolutions.com)

URL: [www.iksolutions.com](http://www.iksolutions.com)

IK