

# Building a **Big Data** Warehouse

*Joris Bontje*  
*Big Data Hacker*

*@mids106*  
*jorisbontje@godatadriven.com*



GoDataDriven

PROUDLY PART OF THE XEBIA GROUP

# About Me

Big Data Hacker

Data Driven Solution Architect

Hadoop Trainer





GoDataDriven

PROUDLY PART OF THE XEBIA GROUP

HOME

WHY US

SOLUTIONS

CUSTOMERS

TEAM

CONTACT

# HELLO! WE ARE GO DATA DRIVEN

We **know** Big Data. We **create** data driven solutions.

We **offer...**



Solutions



Services



Training

# About GoDataDriven

**bol.com**<sup>eo</sup>



**ING** 

 **SanomaMedia**

 **kpn**

**TomTom** 

 **Rabobank**

 **GREENHOUSE GROUP**

 **Belastingdienst**

 **Reed Business Media**



 **essent**



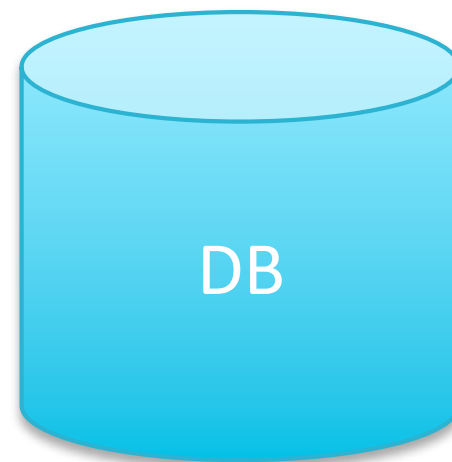
# Data Warehouse Evolution

*In computing, a data warehouse is a database used for reporting and data analysis.*

[http://en.wikipedia.org/wiki/Data\\_warehouse](http://en.wikipedia.org/wiki/Data_warehouse)

# Database Architecture (1.0)

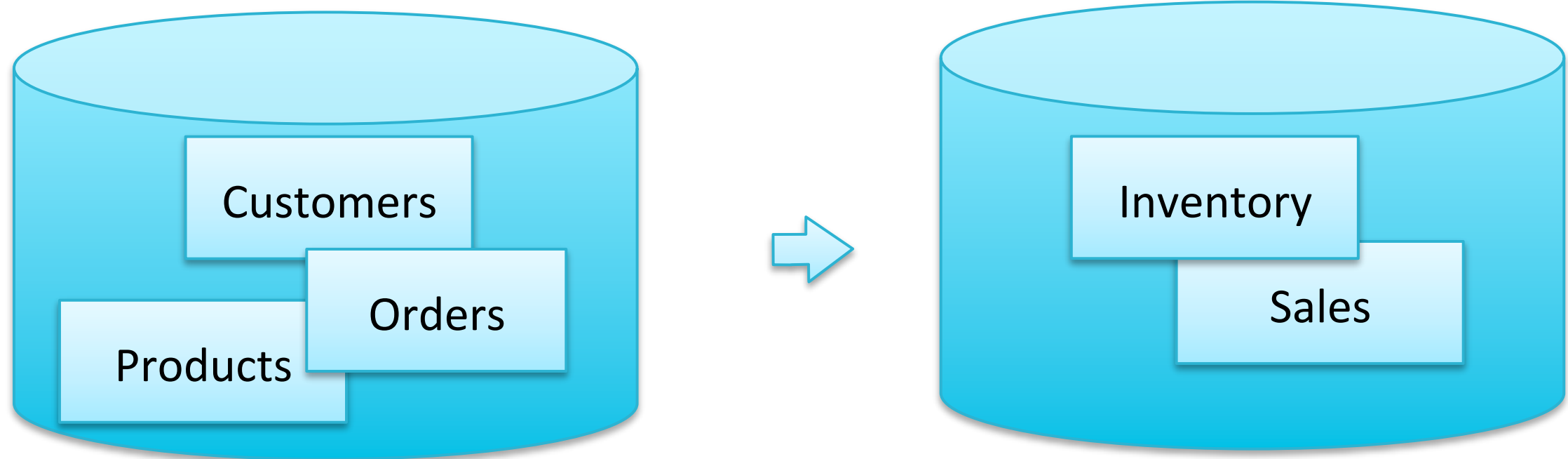
Products  
Customers  
Orders



Inventory  
Sales

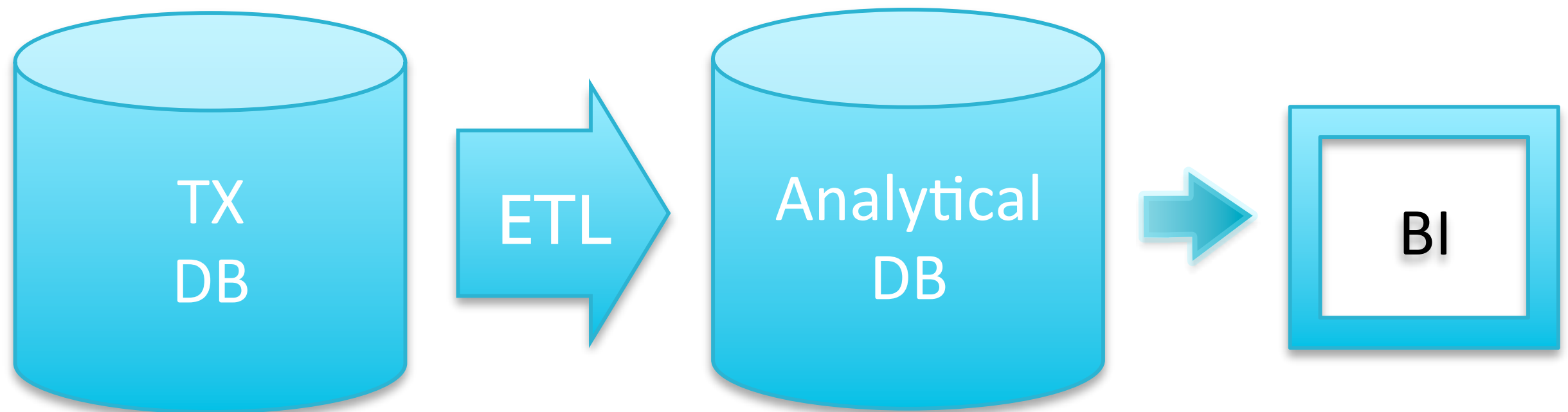


# Analytical Database (2.0)

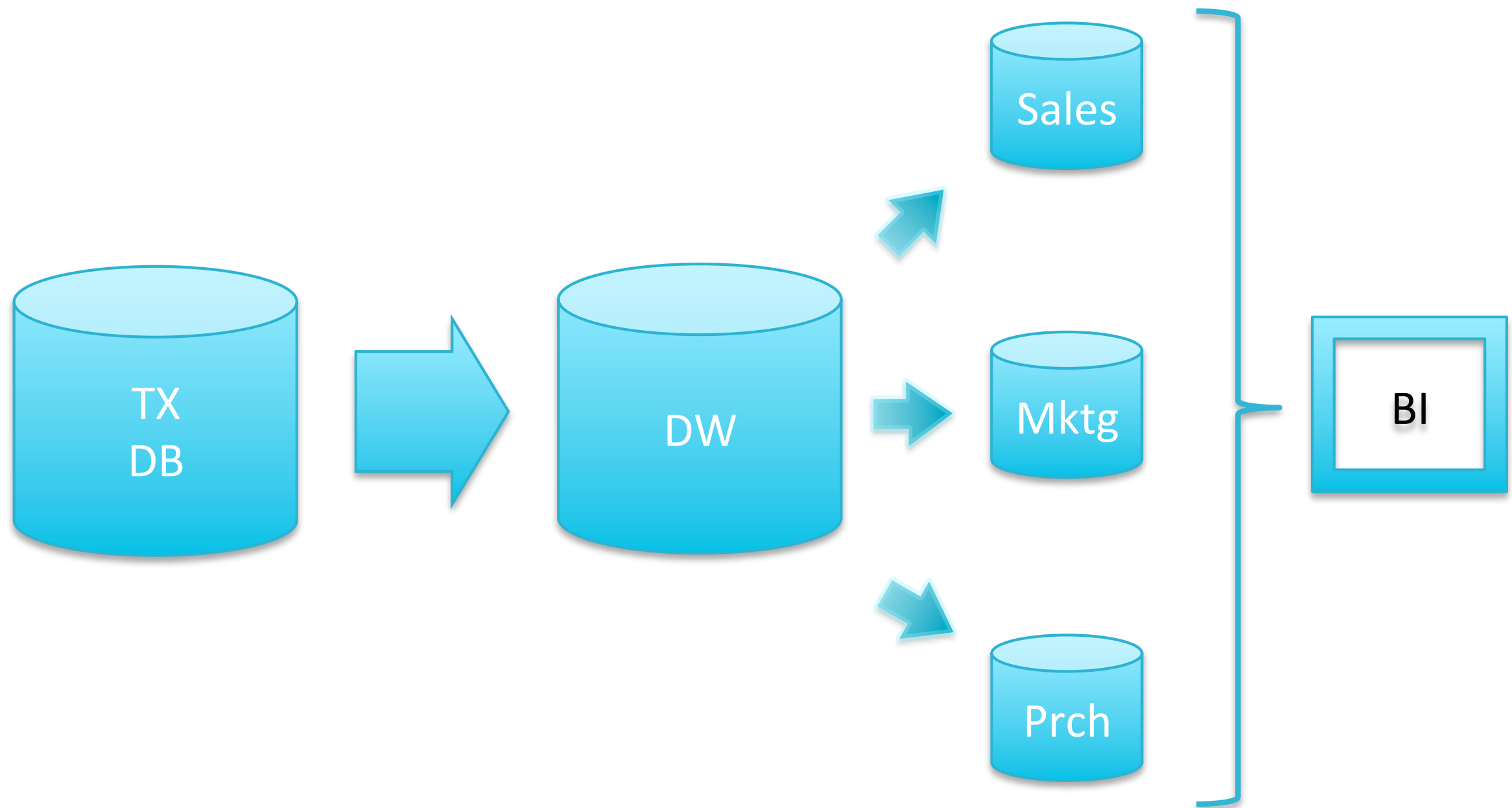




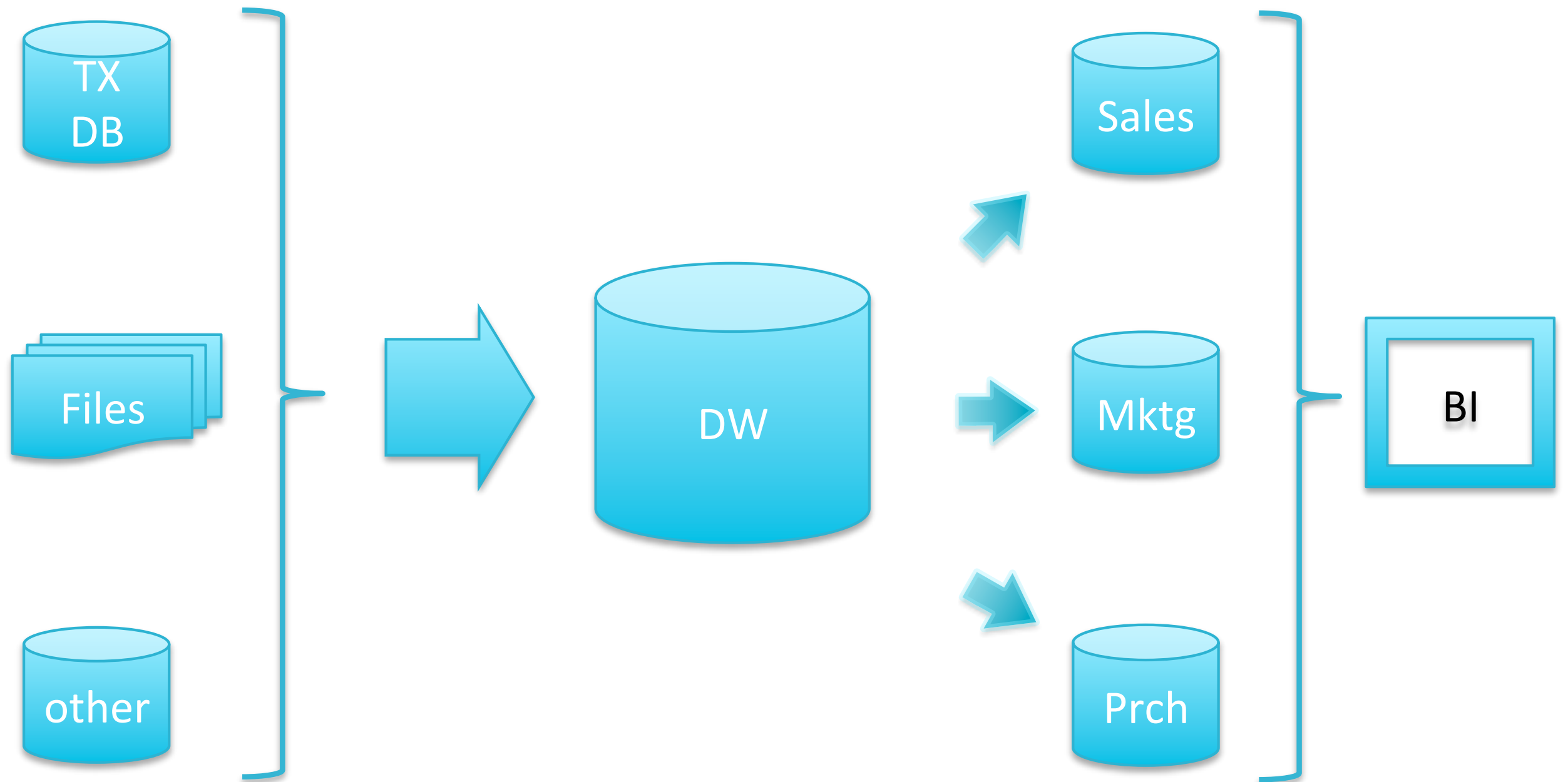
# Basic DWH Architecture



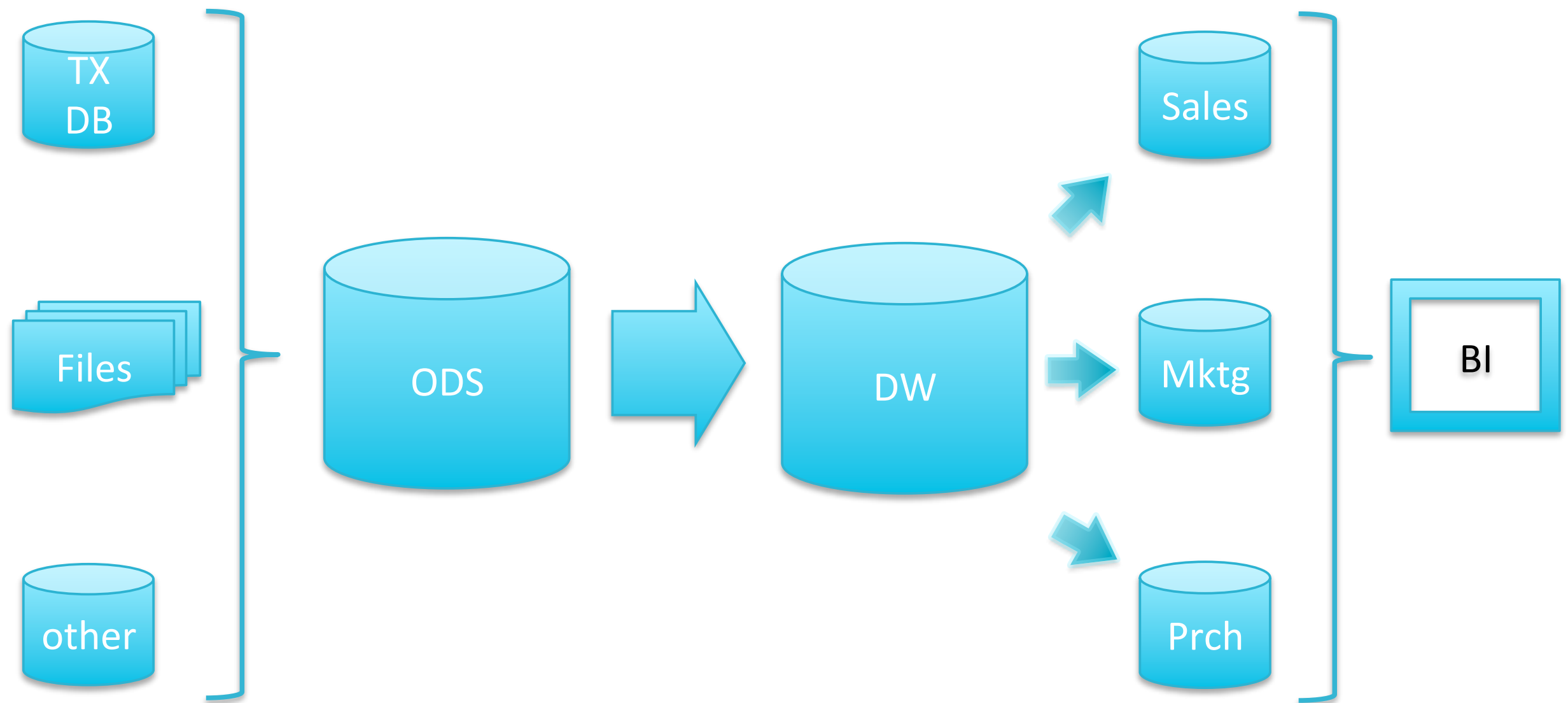
# Data Marts



# Multiple Data-Sources

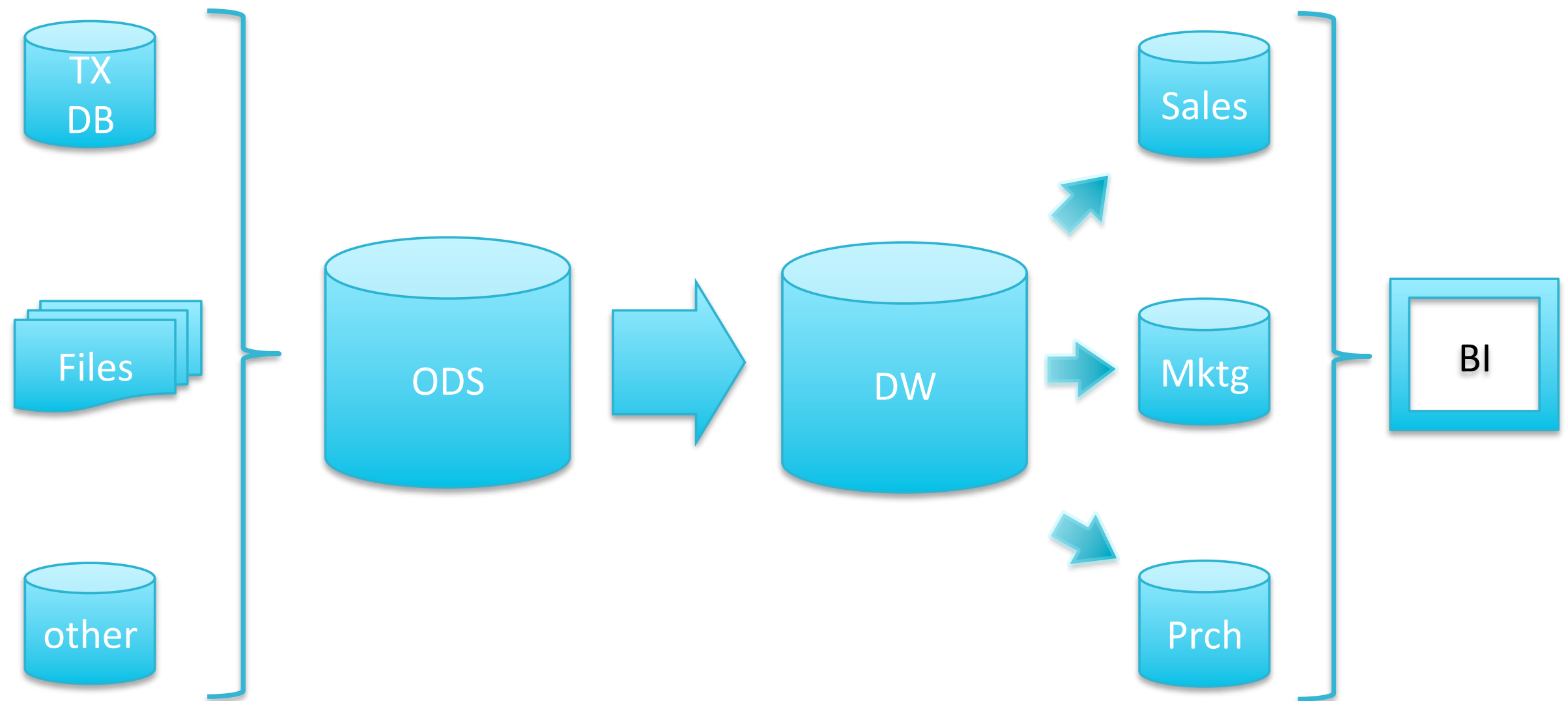


# Operational Data Store

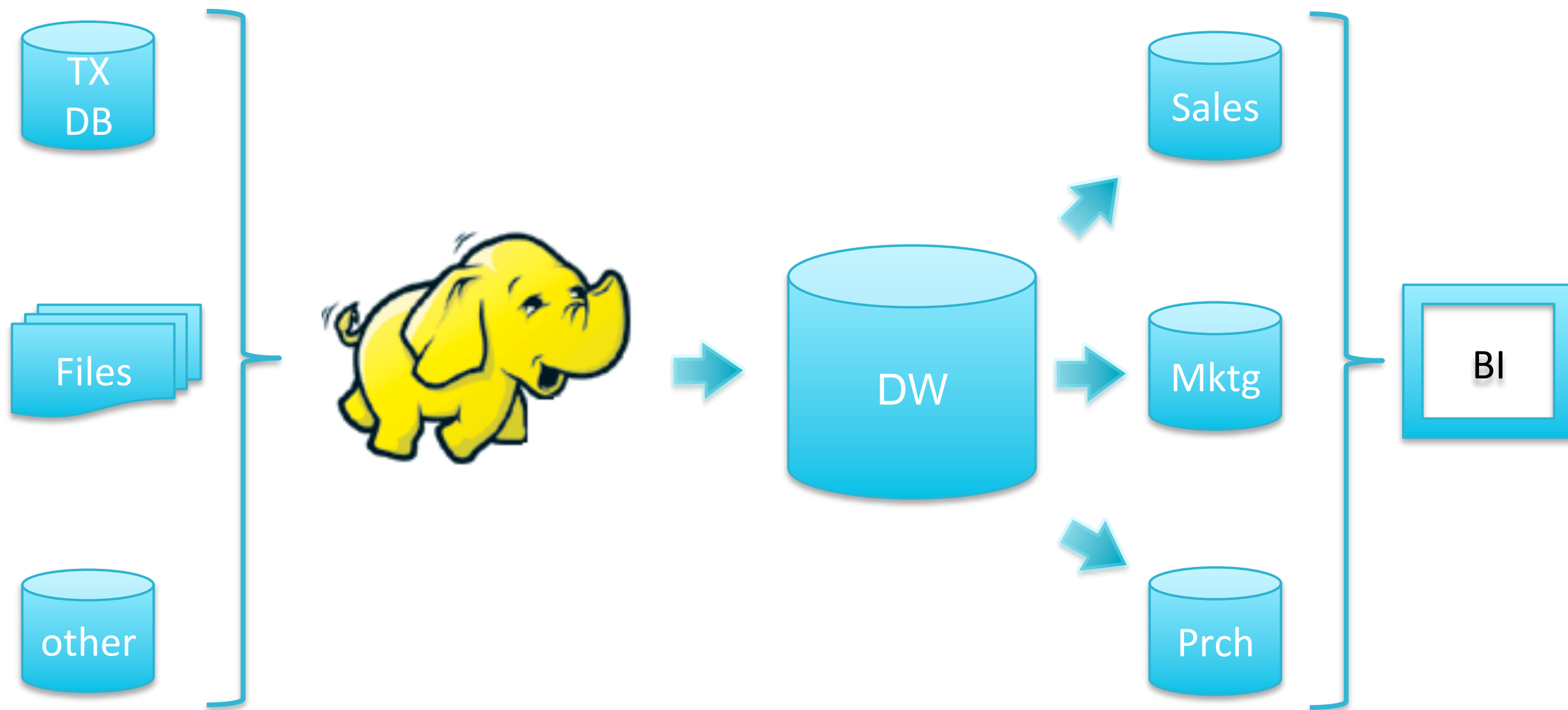


# Hadoop

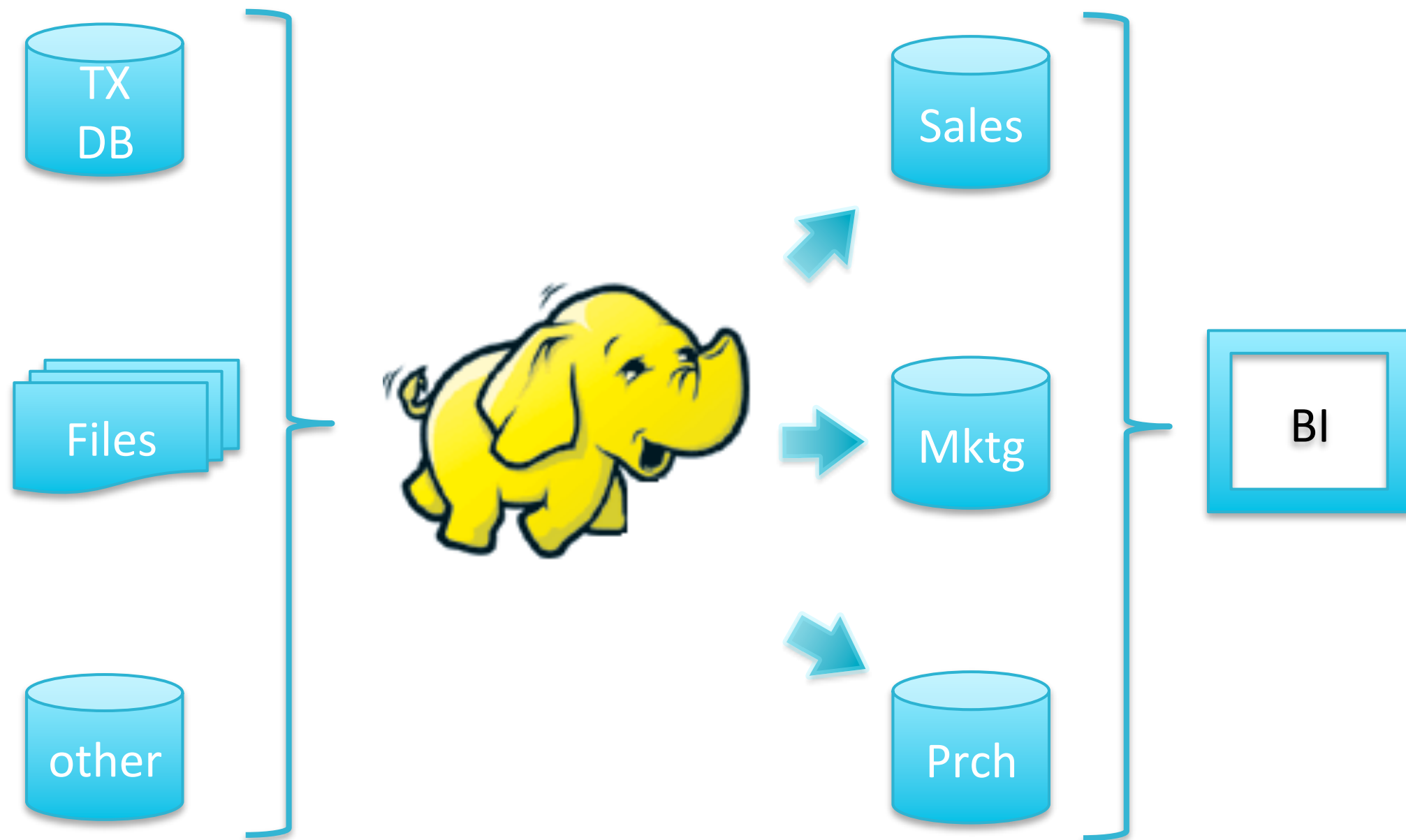
# No Hadoop



# ETL Engine

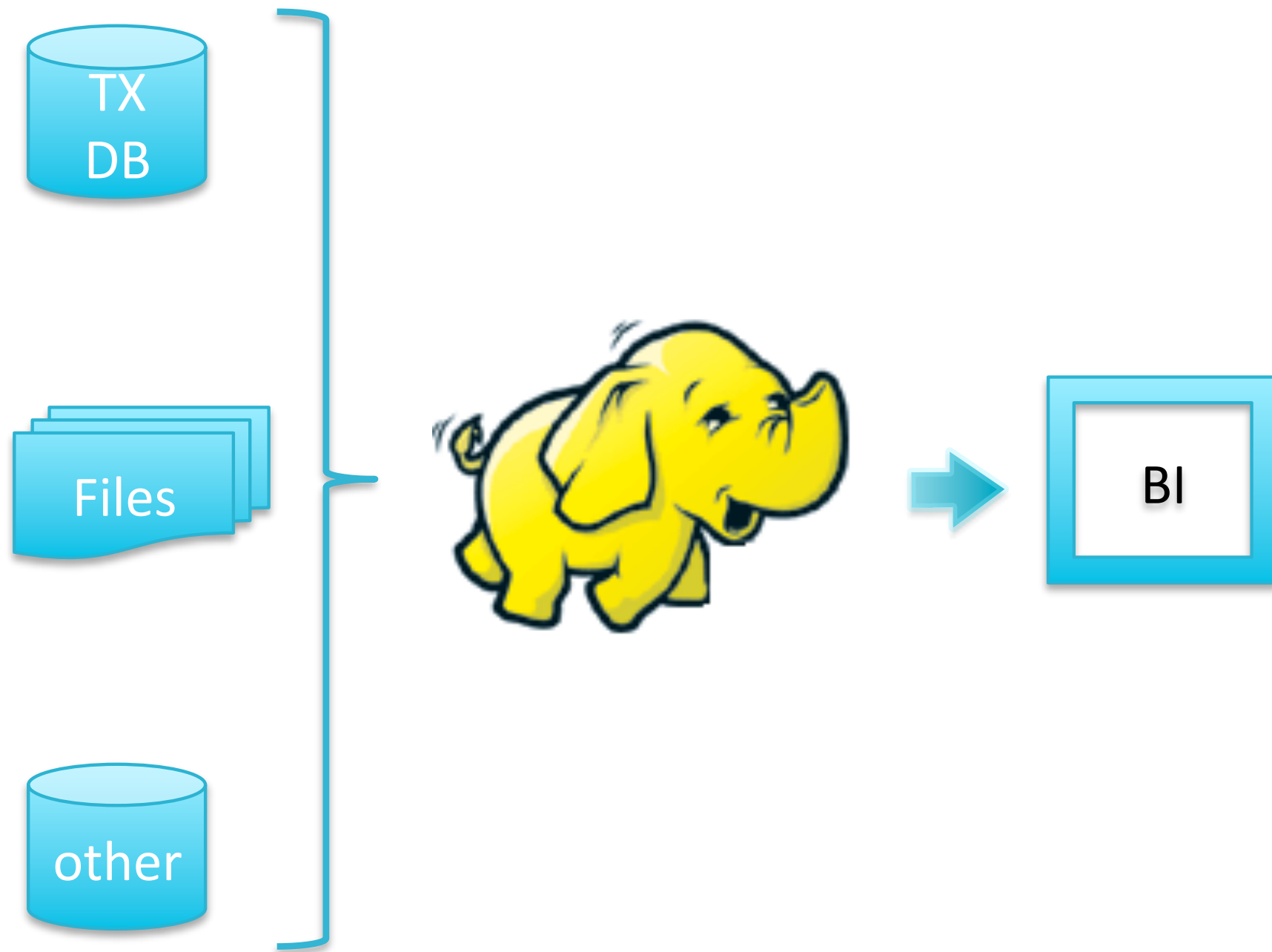


# Tiered Data Warehouse





# Analytical Query Engine



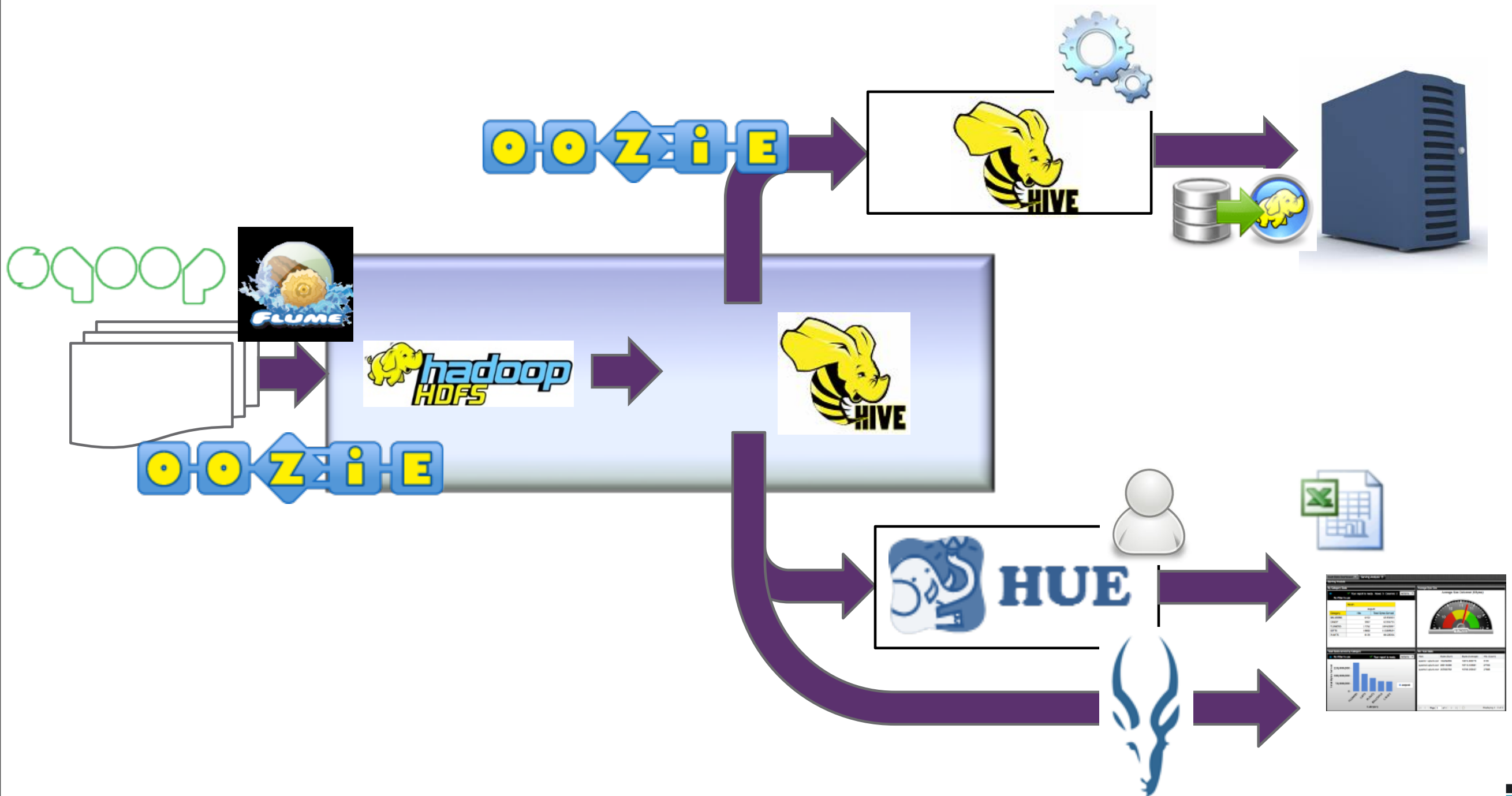
# Tools

# Tools



# Tools Applied

# Tools Applied



# Considerations

# Considerations

Big Data is dirty

Automate everything

Monitoring and QA become the same thing

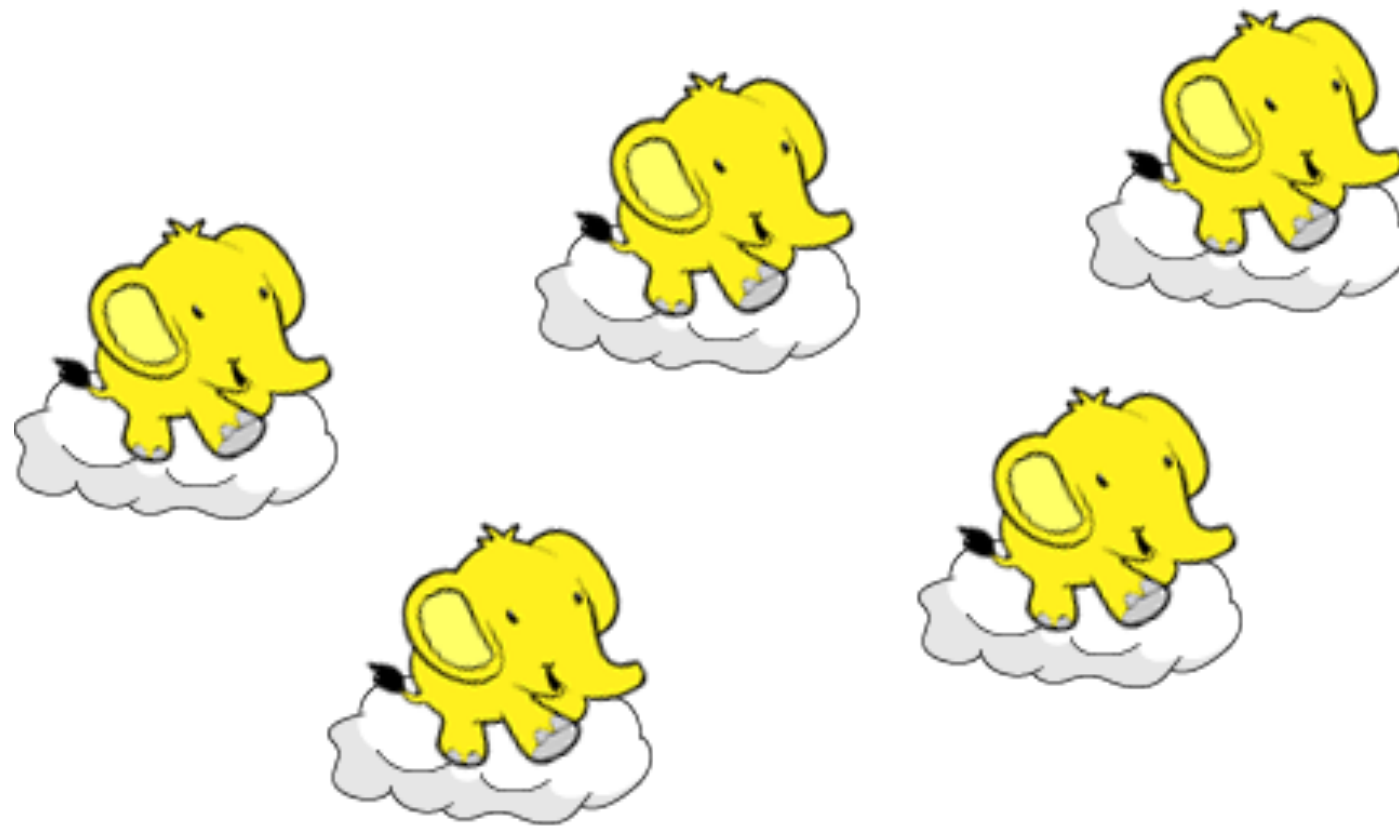


# My Past Trends

Big Data Forum 2012



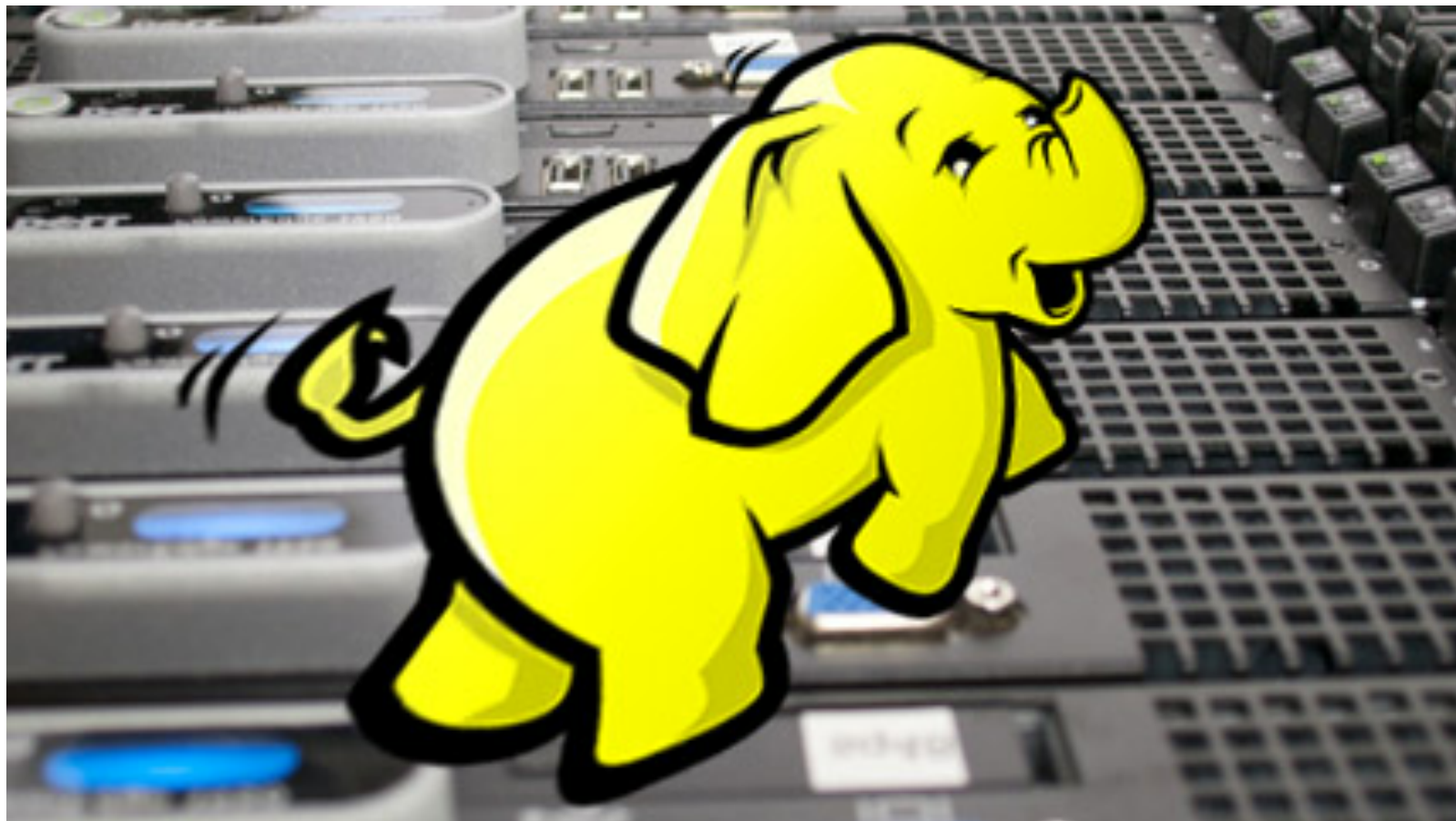
# My Past Trends



Cloud / On-demand



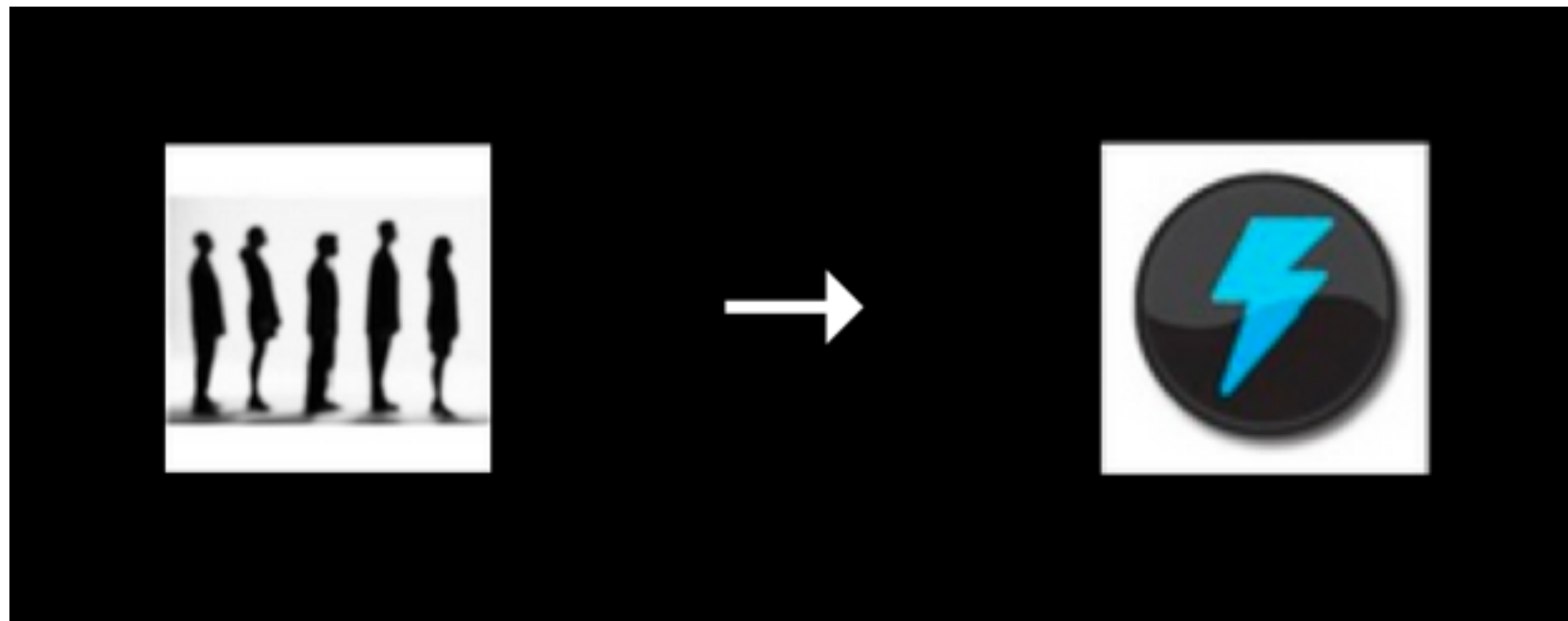
# My Past Trends



Hadoop Hardware



# My Past Trends



Batch → Real-Time



# New Trends

XebiCon 2013

# Trends

## **Impala**

Open Source, Real-time Query engine  
for Hadoop



# Trends

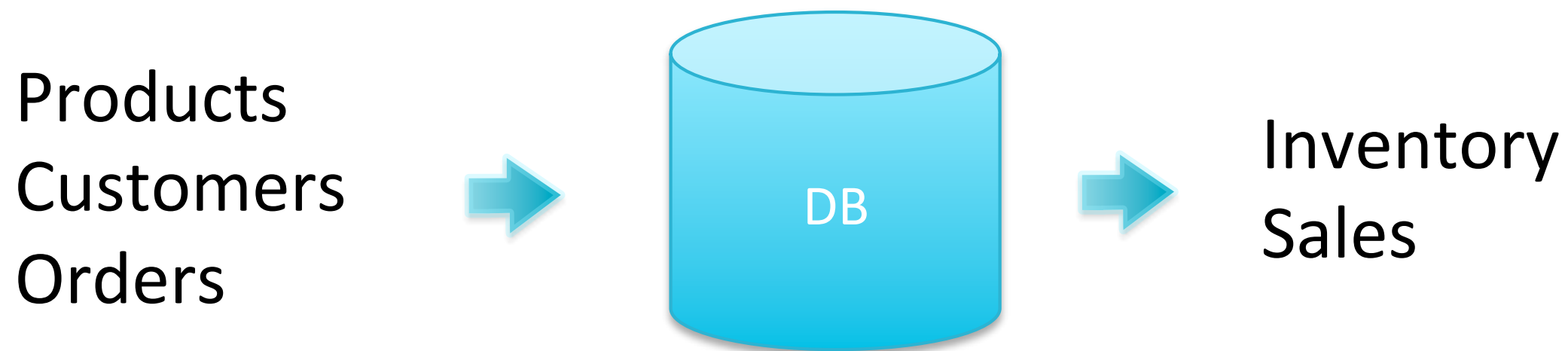
## **HCatalog**

Table Management

Defacto standard for Hadoop metadata



# Simple Database Architecture



# The future?

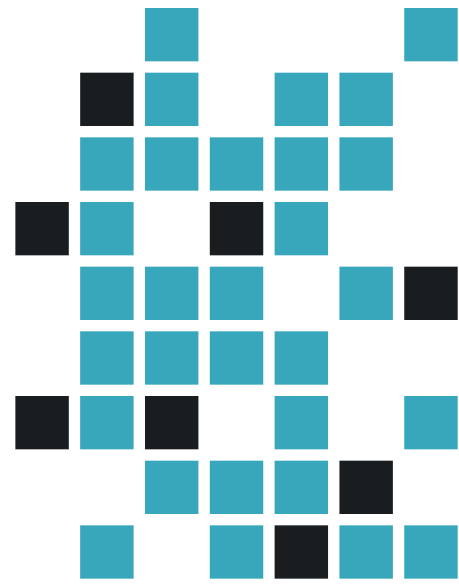
Products  
Customers  
Orders



Inventory  
Sales







# GoDataDriven

We're hiring / Questions? / Thank you!

*Joris Bontje*  
*Big Data Hacker*

*@mids106*  
*jorisbontje@godatadriven.com*