From Provider To Client - The Journey of Data In A Fintech Firm

Tomáš Košlab
Marek Bukáček

FinMason Europe

June 25, 2021

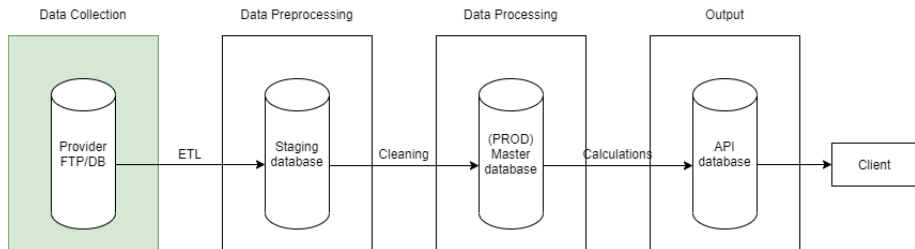Katedra matematiky FJFI ČVUT v Praze

Table of Contents

Obsah

Data Collection

Data Collection	Data Preprocessing	Data Processing	Output
Provider FTP/DB	Staging database	(PROD) Master database	API database

ETL → Cleaning → Calculations → Client

Data Collection

- Cloud solutions - AWS, Azure, Google Cloud, . . .
- API
- FTP
- from vendor's own database
- ETL - extract, transform, load
- save data into staging database
- student - load MS Excel file into Matlab, save into .mat file

Preprocessing

Data Collection | Data Preprocessing | Data Processing | Output

Provider FTP/DB → **ETL** → Staging database → **Cleaning** → (PROD) Master database → **Calculations** → API database → Client
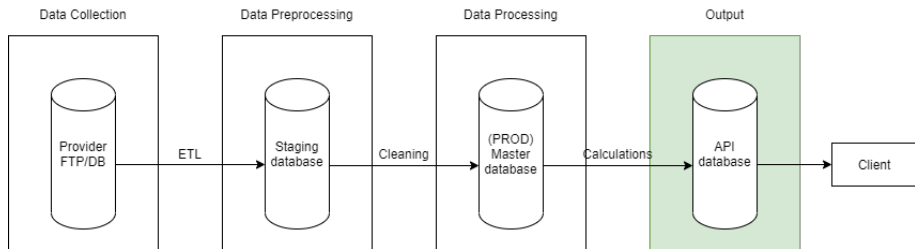
Preprocessing
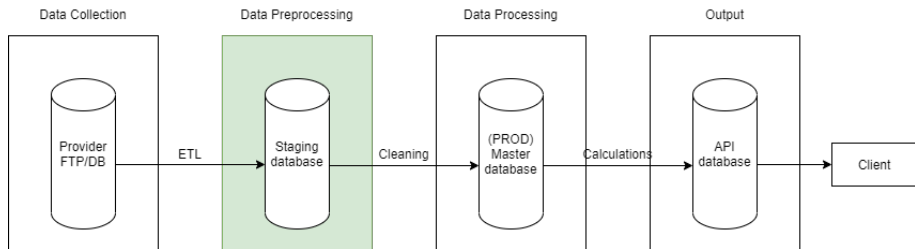
- work with data that are saved in our system
- data validation
- data cleansing - make sure the data is correct, relevant and not duplicated
- after this stage data should fulfill the requirements of used models
- student - outlier detection, reduction of dimensionality, . . .

Data Processing
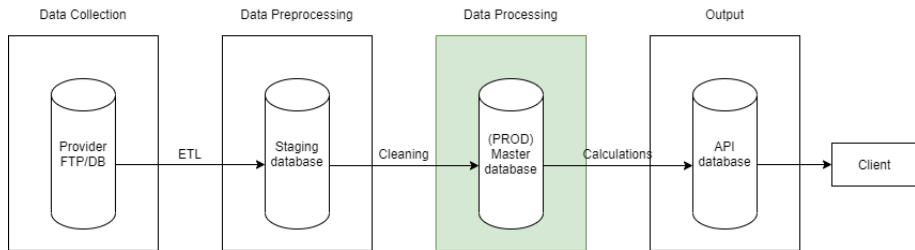
Data Processing

- actual calculations and processing of data
- machine learning, models, statistical approach and inference
- save "raw"results
- student - solve your problem :)

Output

- check raw results and make them available for clients
  - csv
  - refresh website
  - refresh database
  - API
- pre-calculated vs on-demand
- student - visualize and present your results (graphs, charts, tables)

Obsah

Continuous Integration Cycle

Ticket

- describe the problem - nothing is "absolutely trivial"!
- keep documentation of your work
- Jira, Workzone, VersionOne, . . .
- link other similar issues

Local development

- student - changes directly in production code
- local Git branch - if you collaborate on a branch, use git pull before starting your work
- if you work with databases, work in DEV environment
- prepare your feature/patch and commit it to your local branch
- Flyway - for database migration
- pull request - code review
- CI testing (unit tests)

From local to production

1. release branch
2. testing - integration tests, end-to-end test
3. user acceptance testing
4. merge to production

Thank you for your attention!