

ANASTASSIA BLYSHKO

VEIKO KAMPURA

**EESTI LOTO USE CASE ABOUT MIGRATING DATA
TRANSFORMATIONS FROM PENTAHO PDI TO DBT**



30.09.2025
Mövenpick Hotel, Tallinn



How to migrate your data transformations from Pentaho (PDI) to dbt?



Agenda

- Intro
- Legacy Architecture
- New Architecture
- Migration Process
- Conclusion

Eesti Loto

- Estonia national lottery operator for more than 50 years
- 100% state-owned, with a mission to offer responsible entertainment.
- Every ticket supports Estonian culture, sports & education.
- 200 000 players per month
- 60% of Estonian residents played Eesti Loto lotteries at least once

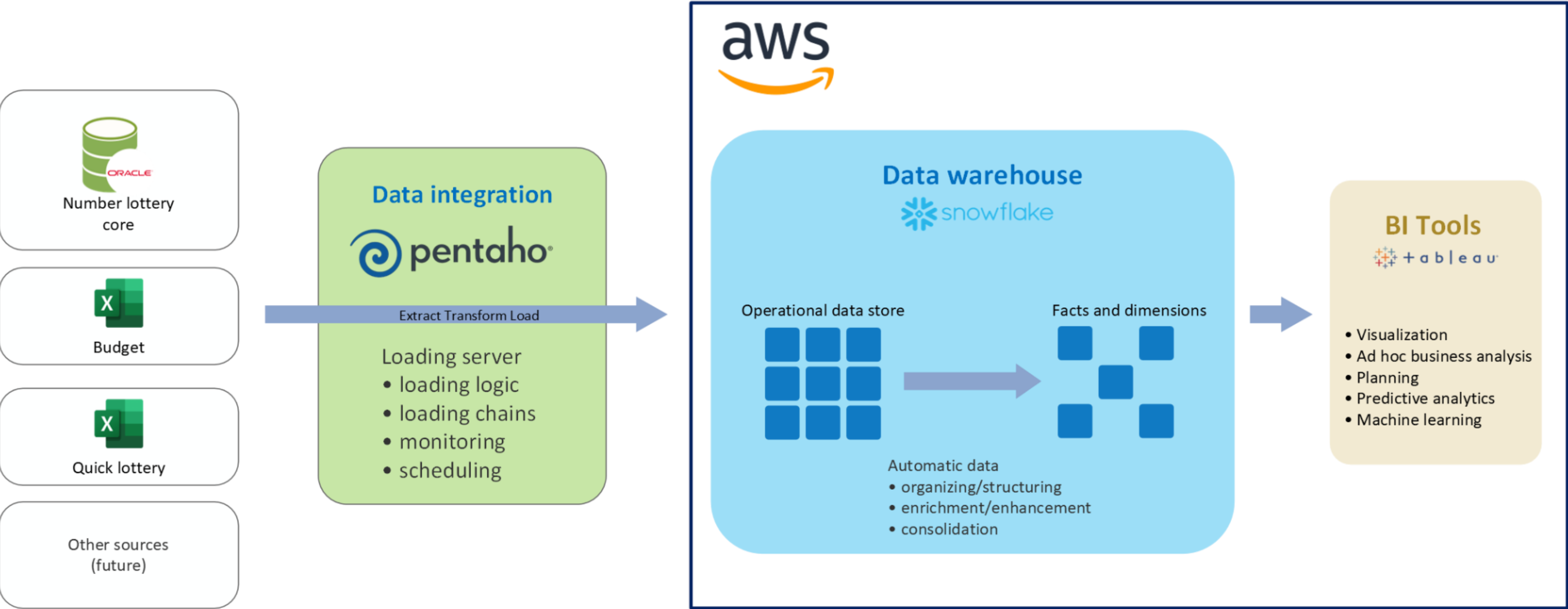
Monthly players
200 000

DWH Timeline



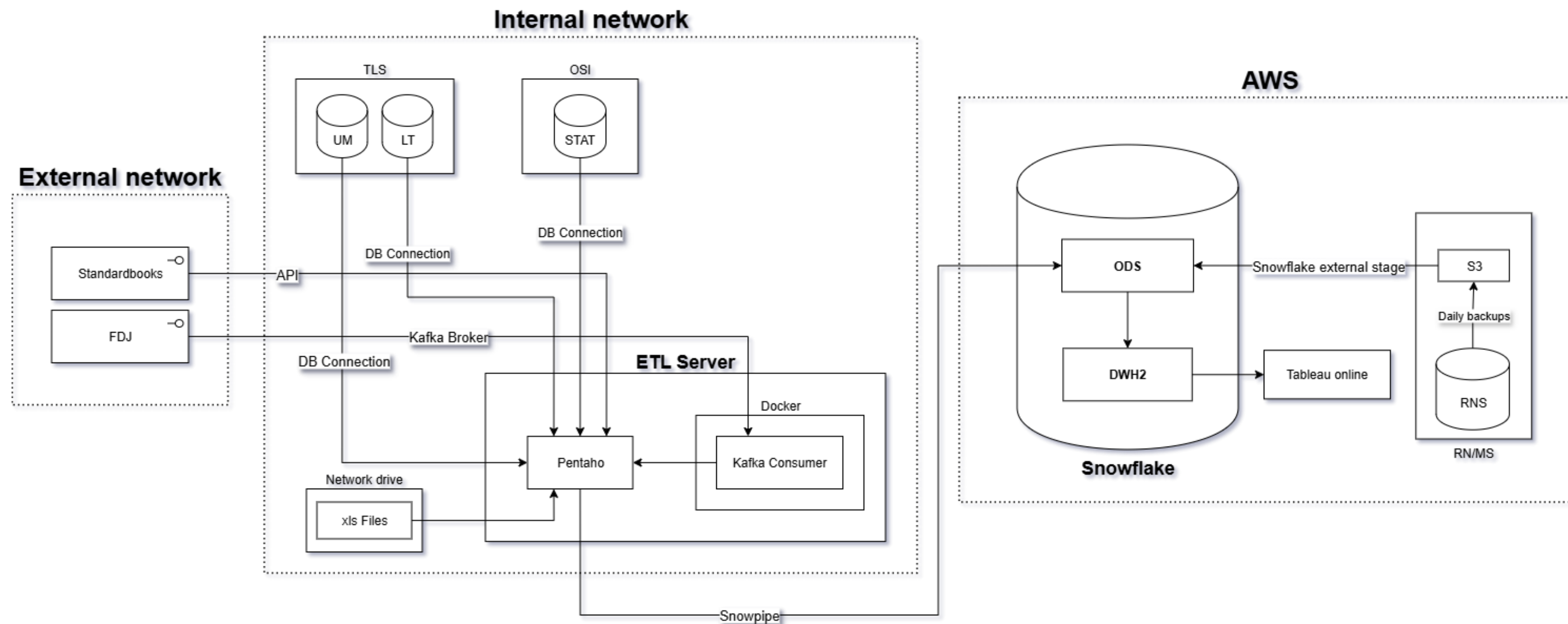
Legacy Architecture

Initial



Eesti Loto

After 4 Years



Problem Statement

Challenges

- Scalability & Maintainability
- Version Control & CI/CD
- Testability
- Transparency

Decision to move forward from Pentaho and find open source, self-hostable and easy to learn solutions for

- central orchestration
- data ingestion / synchronization
- dimensional model / business logic transformations

with focus on tools where we have existing knowledge base

Managing Dependencies in Pentaho



New Architecture

Orchestrator

Legacy Solution

- Crontab to run Pentaho jobs and separate python scripts.
- Pentaho jobs to orchestrate transformations and other steps: if/then/else, file system operations, sending emails etc

Considered Solutions

Category	Airflow	Dagster	Kestra
Open Source	3	2	2
Maturity	3	2	2
Community	3	2	1
Testing / CI-CD	2	3	2
AWS Managed Service	3	0	0
Compatibility with Sync	3	3	1
Existing Know-How	3	1	3
	20	13	11



Data Ingestion / Synchronization

Legacy Solution

- Implemented as Pentaho transformations to sync data from files and databases, custom python scripts to sync from API-s

Considered Solutions

Category	Airbyte	Python ETL Library	Custom Python
Open Source	3	3	3
Ease of Development	1	1	1
Light Resource Demand	1	2	3
Flexibility	1	2	3
Compatibility with Orchestrator	2	3	3
Existing Know-How	2	1	3
	11	13	16



Transformations

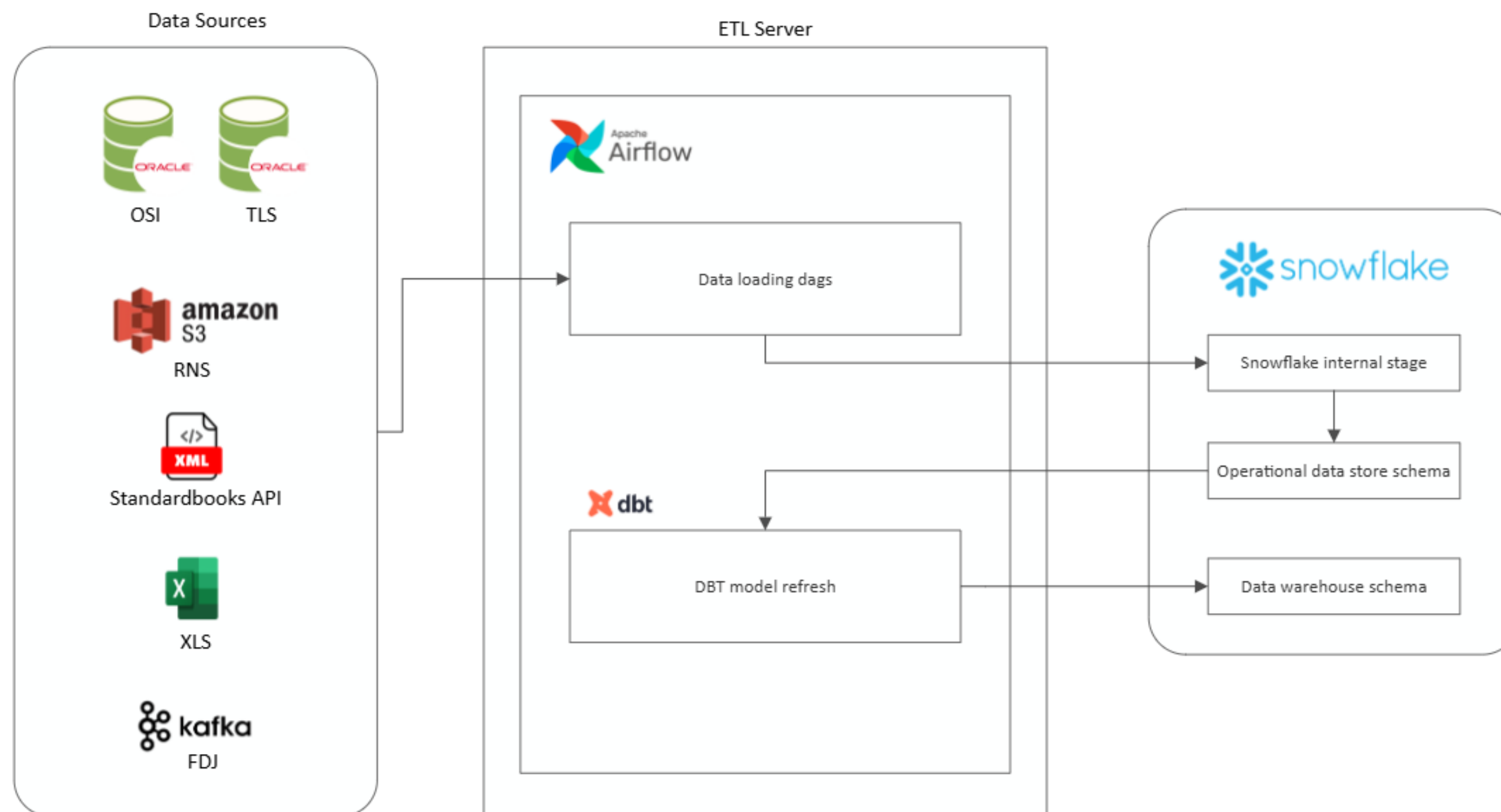
Legacy Solution

- Dimensional model / business logic transformations implemented as hand-written SQL embedded in Pentaho jobs.

Considered Solution

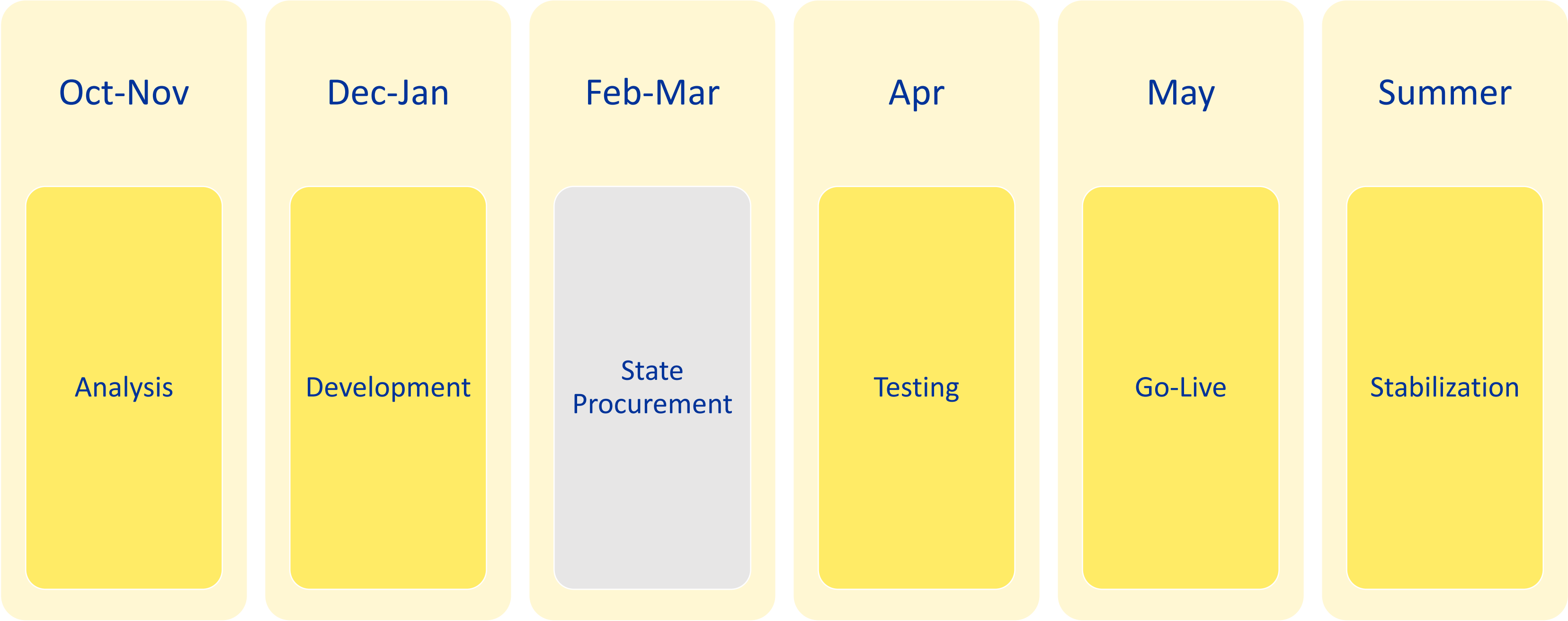
- Chose dbt as a tool moving to be the industry standard
 - Widespread with huge community
 - Built in incremental loading capability
 - Built in dependency management
 - Snapshots & Seeds
 - Automated testing
 - Automated documentation

New Architecture



Migration Process

Migration Process Timeline



Development

Data Sync Coding

- 5 source types
- 3 incremental loading scenarios

Storing Results

- Secondary database to hold artifacts of the new data flow

Orchestration

- Airflow install and conf
- Setup of loading DAGs

Transformations

- dbt install and conf
- Migration of transformations

CI-CD

- Automatic refresh of Airflow code base after commit to Github

Testing

Legacy and new data flows were run in parallel for 2 weeks

- Data validation
- Regression
- Performance & Load
- User Acceptance

Go-Live

After all tests were done and all known bugs fixed, we switched to the parallel database and stopped Pentaho loadings

- Documentation and knowledge transfer
- Monitoring and optimization
- Plan for decommissioning legacy Pentaho

Conclusion

Results & Benefits

- Developer Productivity & Velocity
- Maintainability & Reliability
- Data Quality & Trust
- Scalability & Performance
- Modern Engineering Practices
- Observability & Monitoring
- Strategic Flexibility & Future-Proofing

Lessons Learned & Next Steps

We learned and re-confirmed that

- You need to move with the times
 - Your initial perfect choice is probably not perfect after 4 years
- Migrations are rarely only for tool replacement
 - Consider them early on as data product modernization projects
- Plan (for the future), while building for the present
- Trust what you know

Next, we will

- Transferring the test environment to the new solution
- Ongoing improvements (data sync, incremental loadings, testing)...

Thank you!

GOLD SPONSORS



Google Cloud



Sparkle



Microsoft

SILVER SPONSORS



UNIVERSITY OF TARTU



SOLITA

SelectZero

veriff



VALI IT

PROEKSPERT

TERASKY

dbt Labs

SUPPORTING PARTNERS



Supersimple



SHROOMWELL
WELLNESS
— FARMING —
INNOVATION

 **Tallinn
Data
Week** ²⁵

Please give your feedback!

Help us be data-driven 😊
Scan the QR code to rate the event.

