

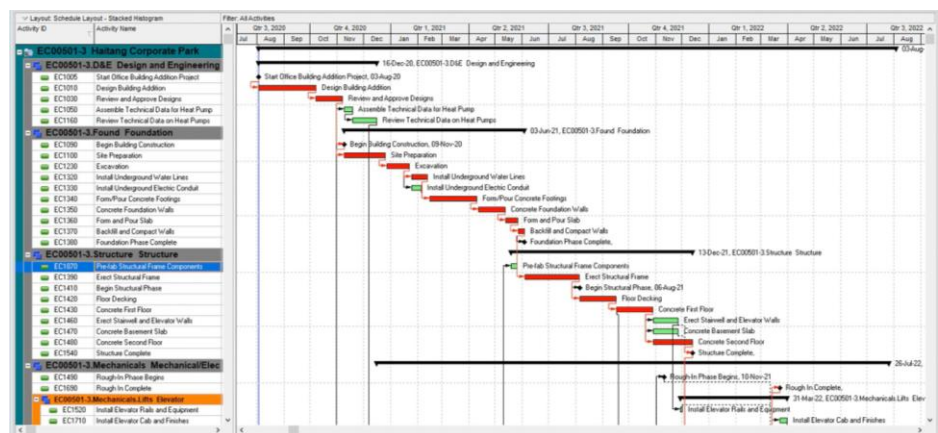
Data Pipelines Demystified

Marco Frisenda
Isaac Dyer

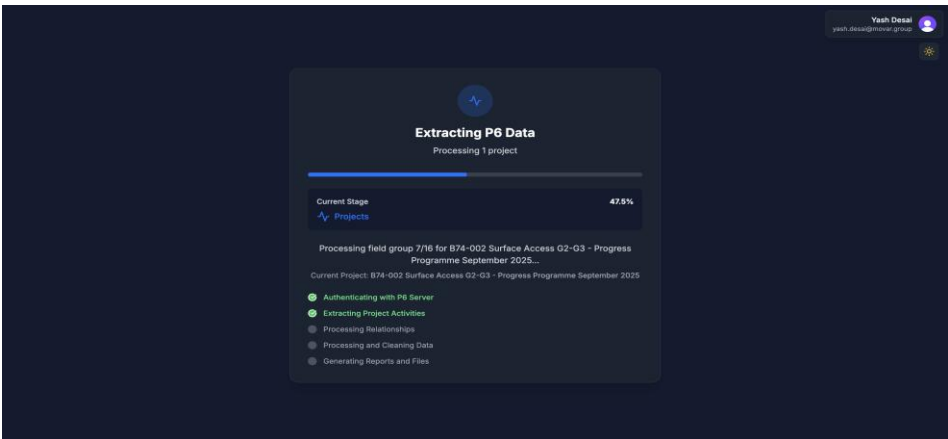


Data pipelines demystified : Extraction and Visualisation Techniques

Agenda



Extraction



Processing



Visualisation

Understanding the data – Data Extraction

Things to consider:

- Requirements
- Clean data vs messy data
- Time and effort to process

To simplify the process, you must consider the pipeline from end to end and clarify not just how to get what you want but also, what you want.

Data Extraction (ii)

Bronze



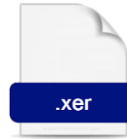
Pros

- Easy to interpret and set up
- Useful for proof-of-concepts

Cons

- Difficult to scale and track over time
- Transformations and cleaning is manual

Silver



Pros

- Conciseness of data extraction
- Full baselines (XML)

Cons

- Mid-level understanding to process effectively
- Less manual than excel but still manual

Gold



Pros

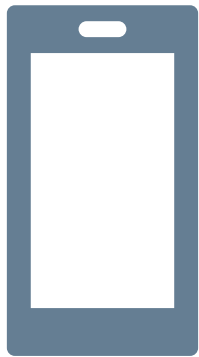
- Uniform data makes processing simple
- Once set up, requires little manual input

Cons

- Set up costs
- Changes in strategy require overhauls in backend
- Requires expertise to snapshot and orchestrate

Processing and storing

- Data models – A representation of how data is structured. Entities, Attributes, Relationships, Data types.
- Cleaning data via transformations
- Analysis of information
- Regression lines



Processing and Storing (ii)

Bronze



Pros

- Tailored data structure
- Easy to understand

Cons

- Processing and storing is inefficient
- Large amounts of data causes crashes
- Limitations on analysis and visualisations

Silver



Pros

- Is excellent for storing in a database
- Table to Table matching

Cons

- Datamodels can become confusing
- More complex visuals like gantt charts require expertise

Gold



Pros

- Can be fetched at any time
- Storage for later use is clean

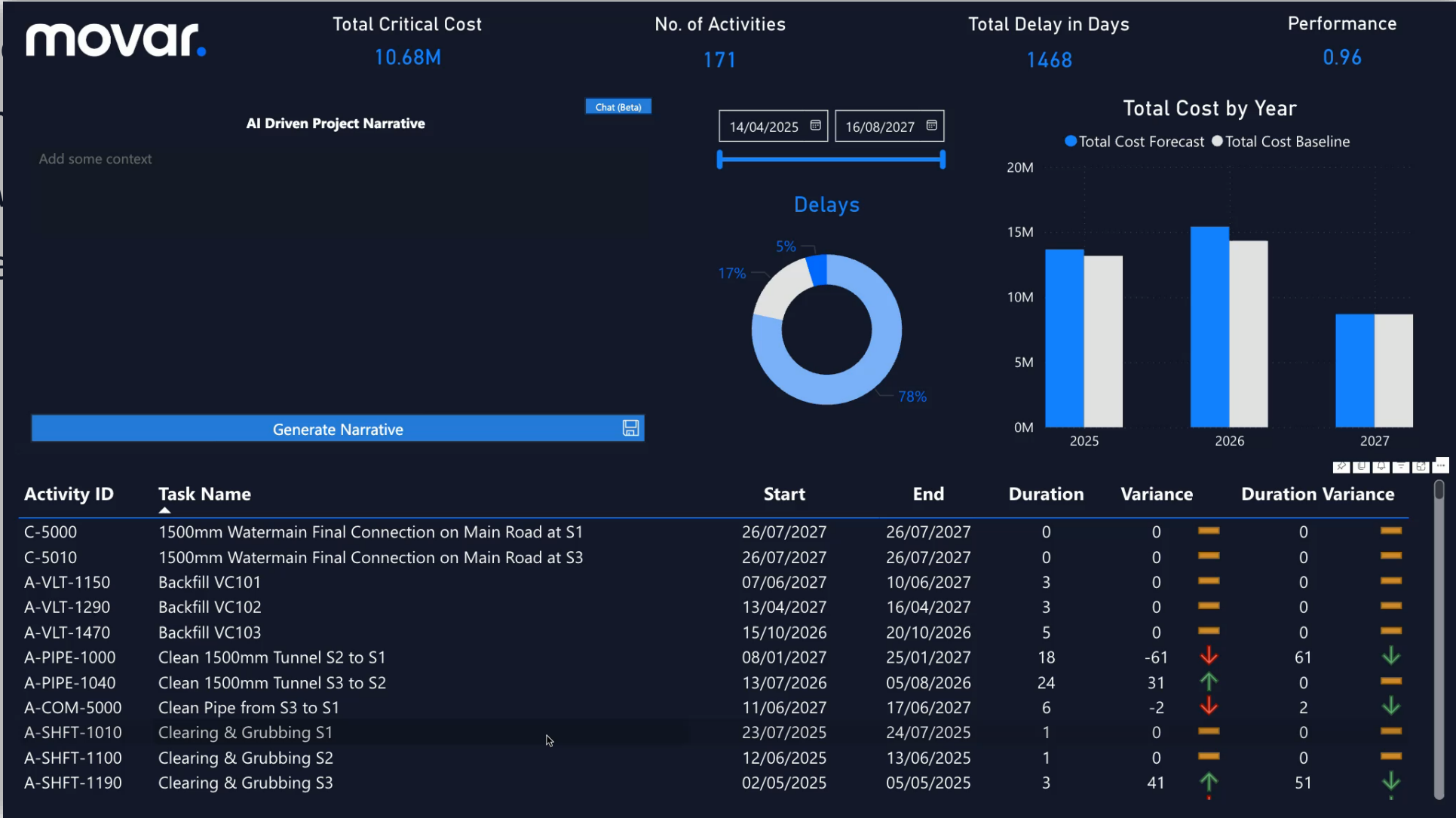
Cons

- Requires a workspace which can be costly
- Requires stricter PMO processes and orchestration

- Power
- Dem
- Shov
- Leve

mover.

Add some context



EXPECTED OUTCOMES

Enhanced Governance

With a powerful data pipeline, project information isn't limited to lengthy monthly reports. But it is instead easily accessible and expedient.

Increased Efficiency

Combining all sources of information into one place allows the PM to spend less time cross-checking and spend more time empowering decision making.

Proactive Project Management

Waiting for the end of the month to create a summary of project progress is a thing of the past. Use data to get ahead of the curve.

Cost Savings

Live data pipelines allow for more finely tuned project controls. Preventing and escalating data driven risk becomes a part of BAU.

Recap

From a data perspective, key points are:

- **Start small and keep scale in mind**
- **Efficiency at the start, peace of mind at the end**
- **Punchy, concise, action-oriented visuals**



Movar.Netlify.app

From a Project Control/Planning point of view, the information that is paramount:

- **Coding - “Golden thread”, to enable linking the correct information across different inputs**
 - **Project coding (when extracting directly from Primavera Database)**
 - **Activity Coding (to enable finding the right information and linking it to other sources)**



A big thank you to all our sponsors

ORACLE®
Construction
and Engineering



movar.



th3rdcurve®

