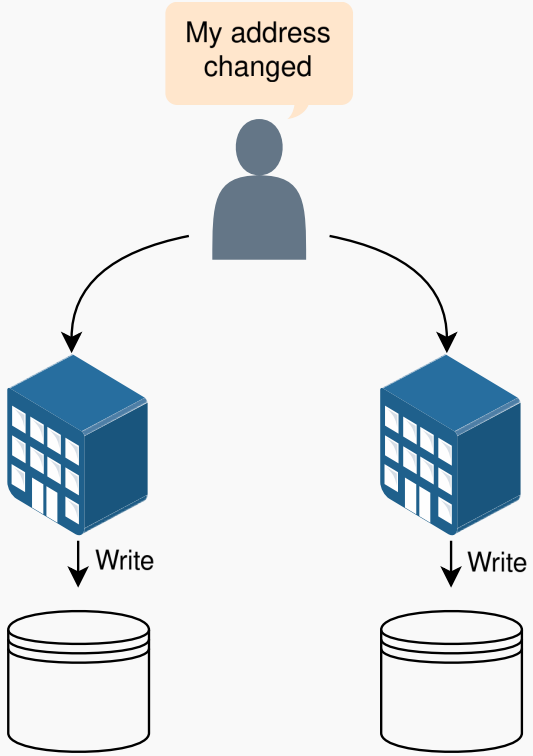
Optimizing Write Performance in Decentralized Personal Data Ecosystems

Jitse De Smet

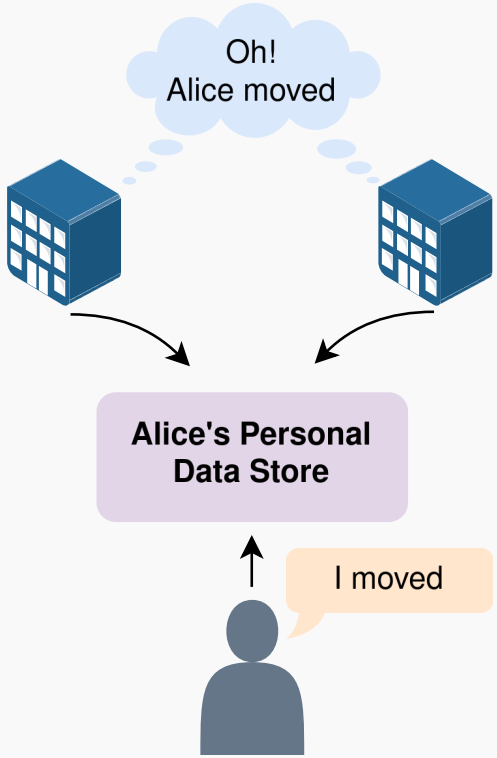
5 September 2024

Importance of decentralized personal data ecosystems

Data around processors

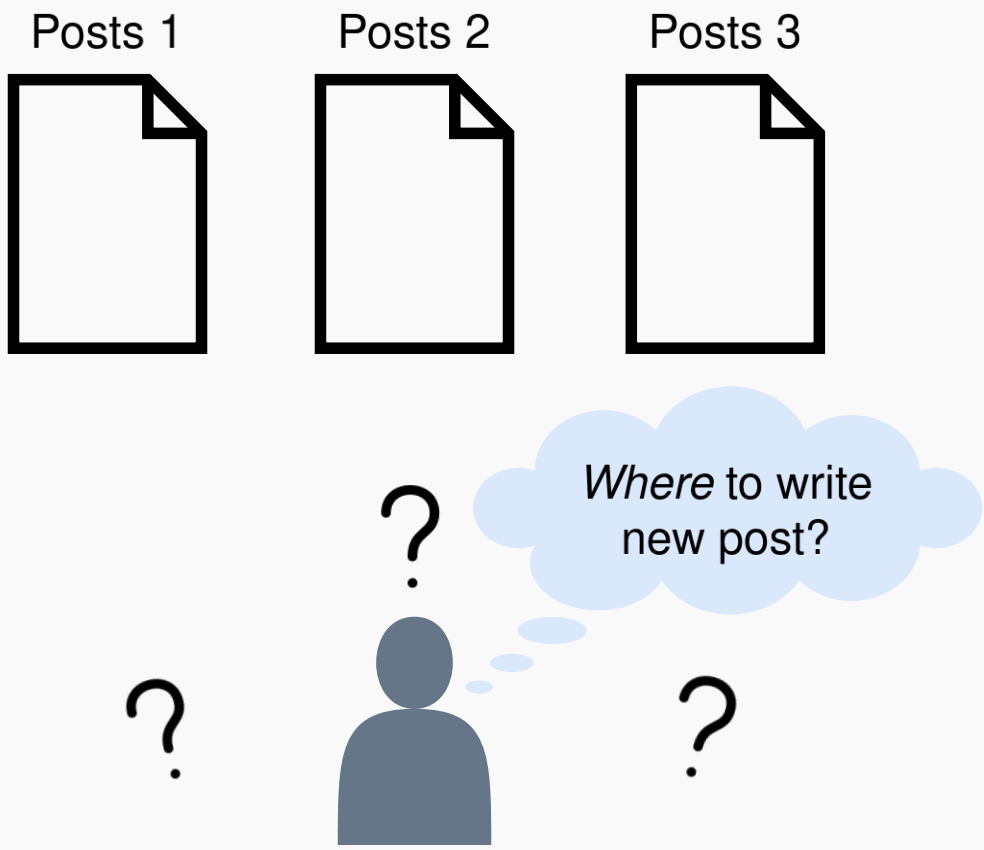


Data around producers



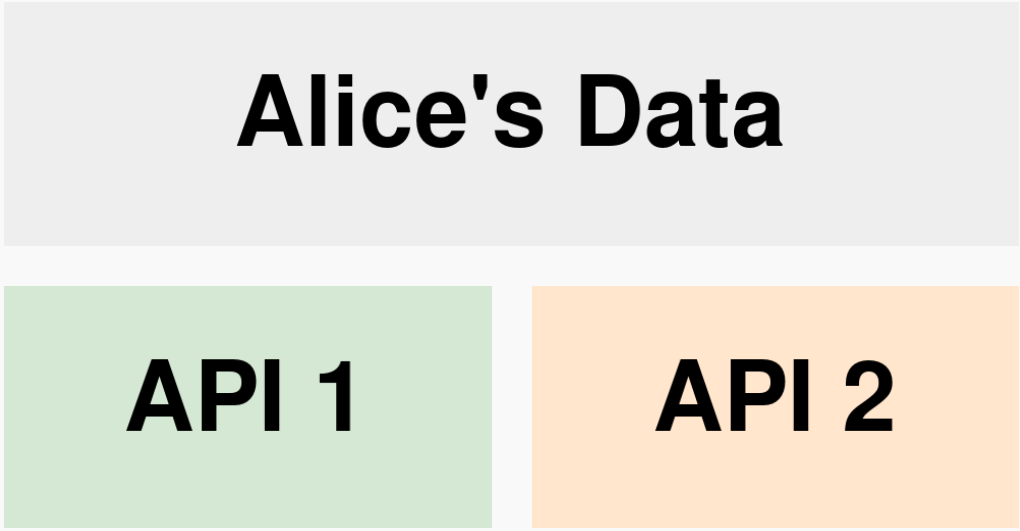
- ✓ Relieves SMEs cost
- ✓ Data quality
- ✓ Sparks innovation

More profound update issues are lurking

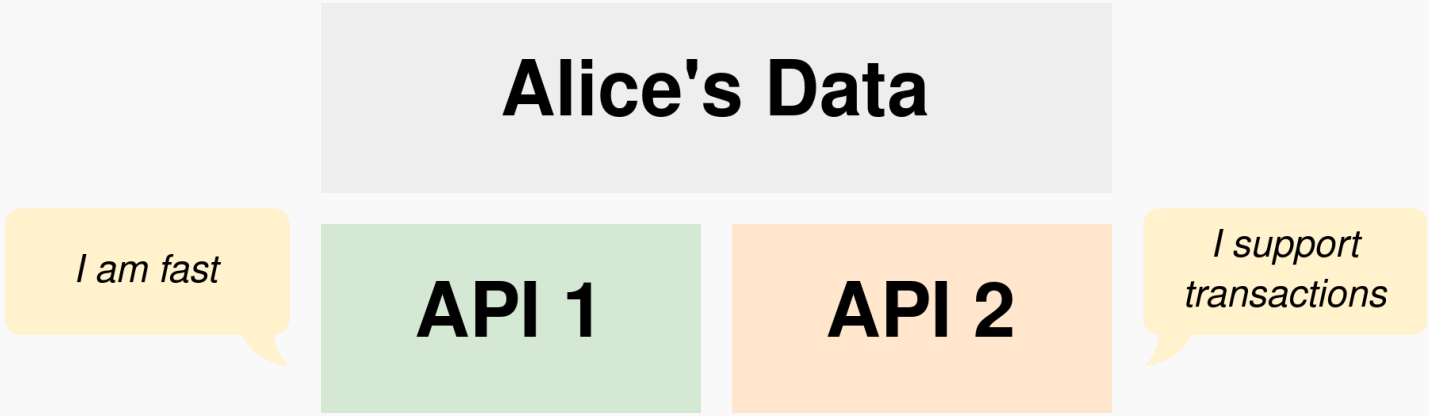


WP 1: Study of update interfaces in a decentralized environment

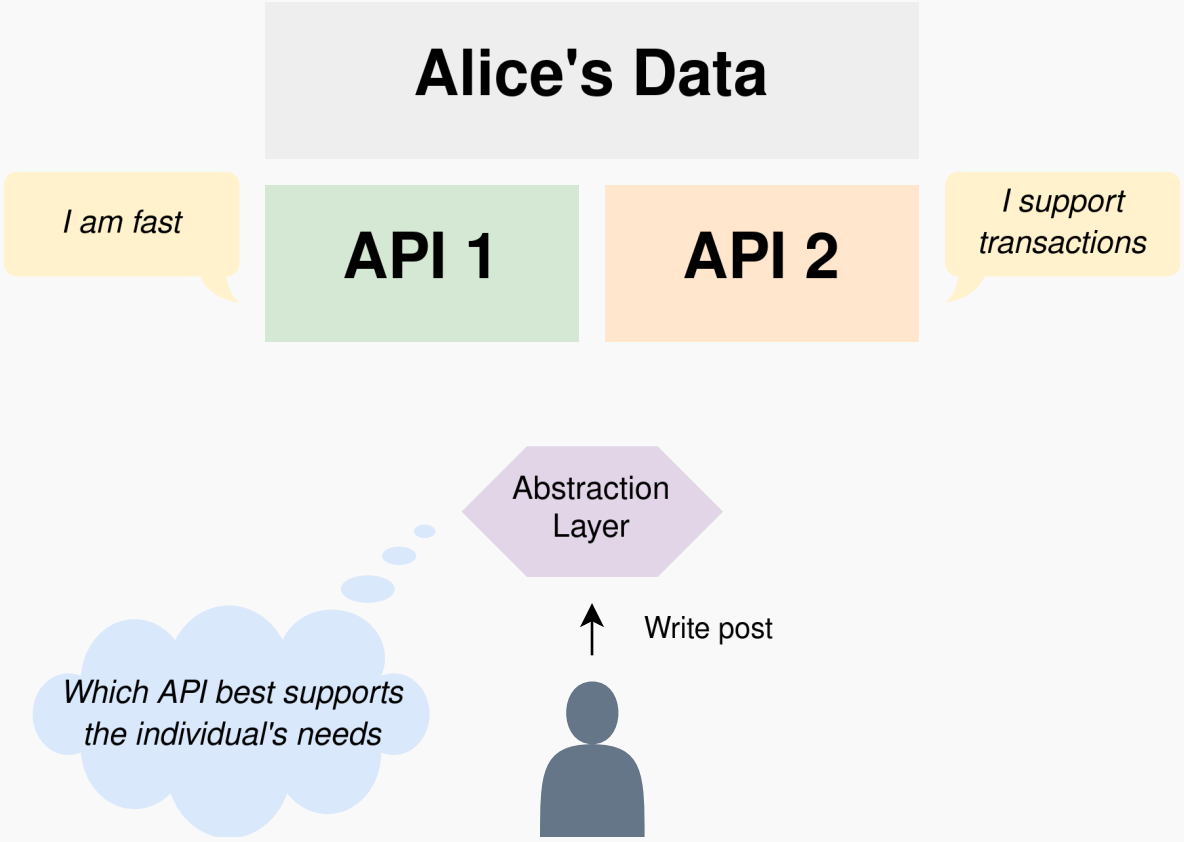
WP 1: Data store exposes multiple update interfaces



WP 2: Interfaces are self-descriptive

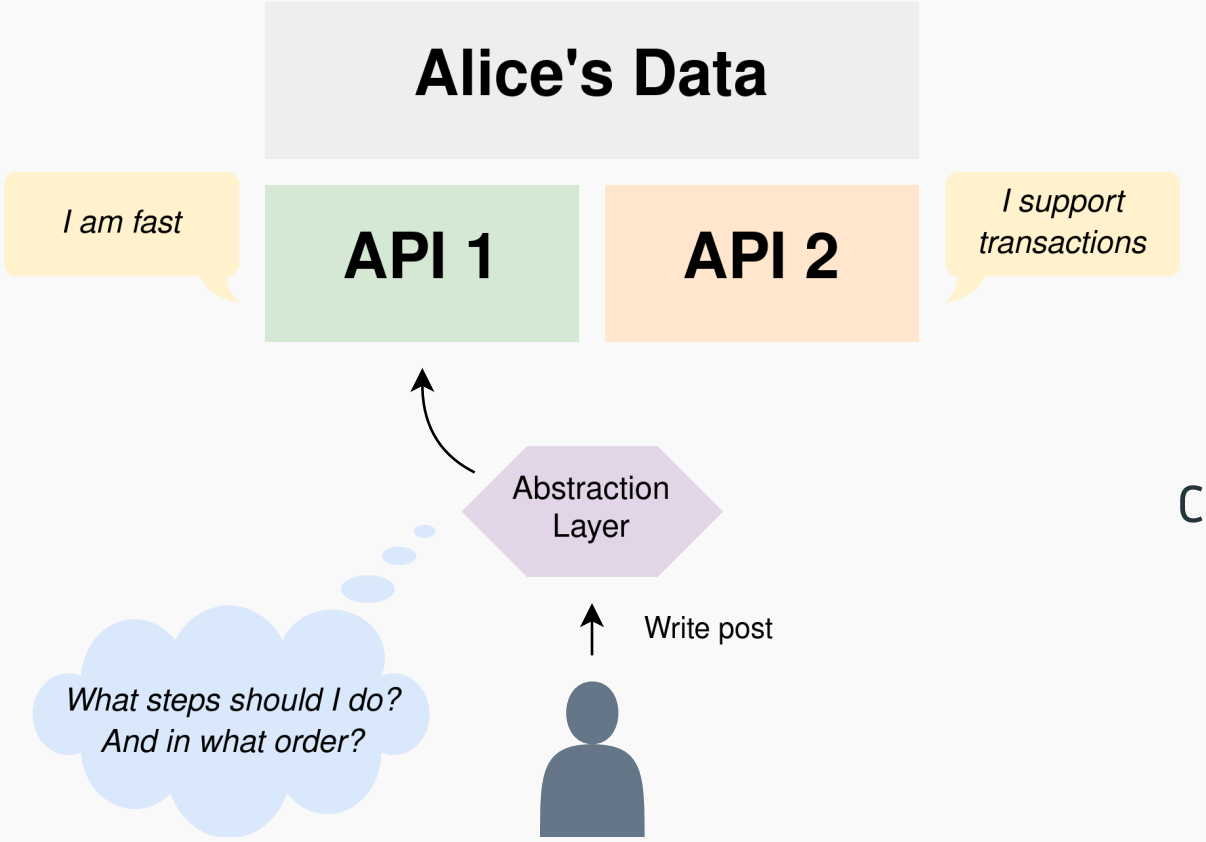


WP 2: Discover data store interfaces & select



3 years experience:

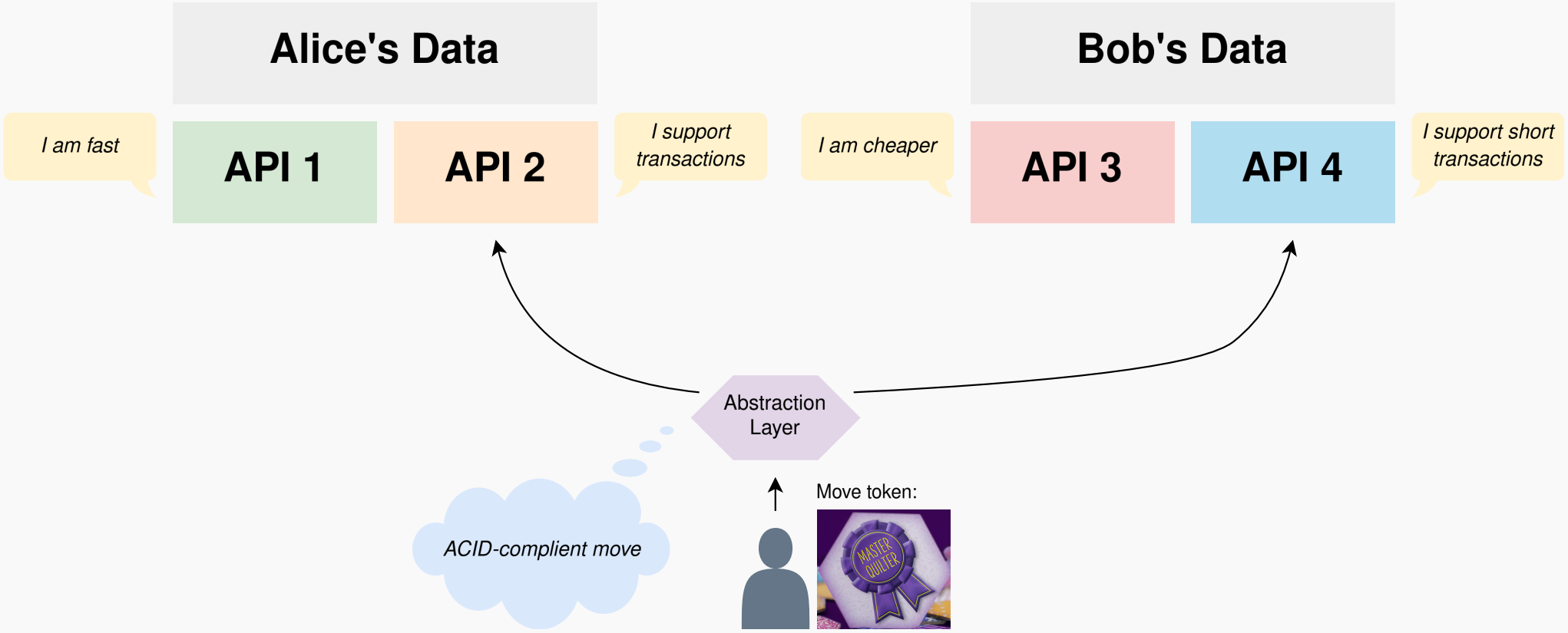
WP 2: Update query plan for a single permissioned data store



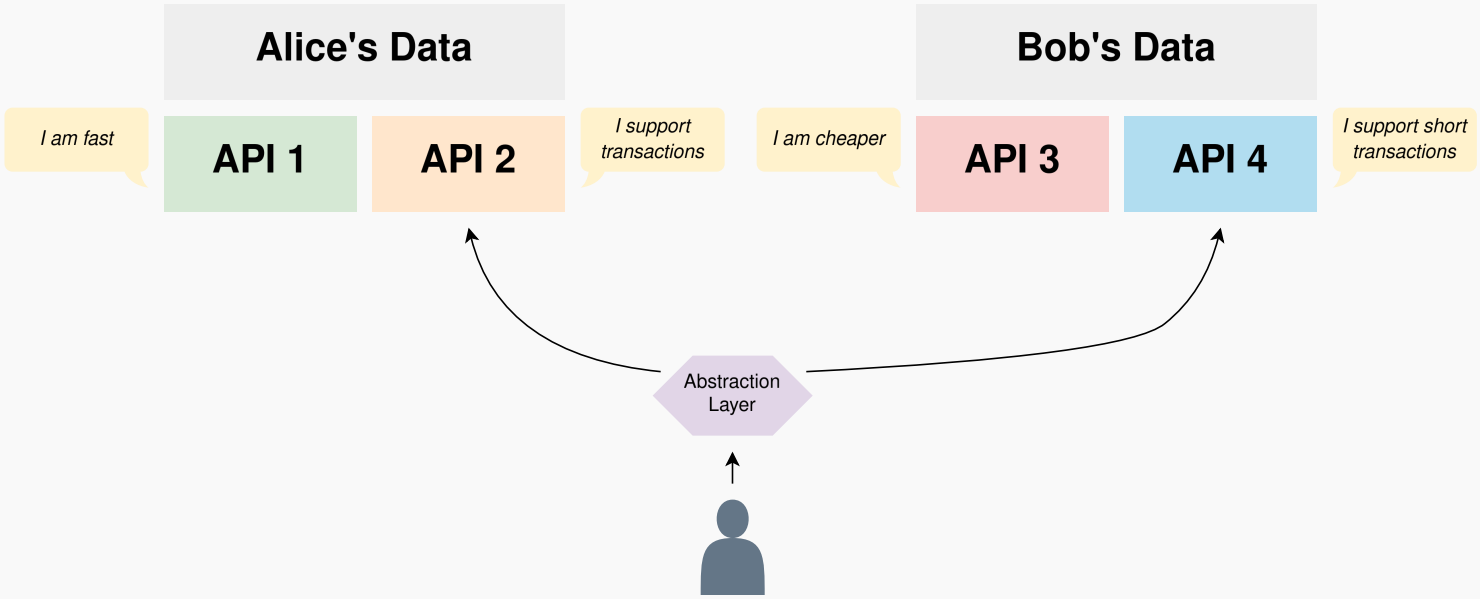
Collaboration with:

TU WIEN

WP 3: Update query processing across heterogeneous interfaces



Conclusion

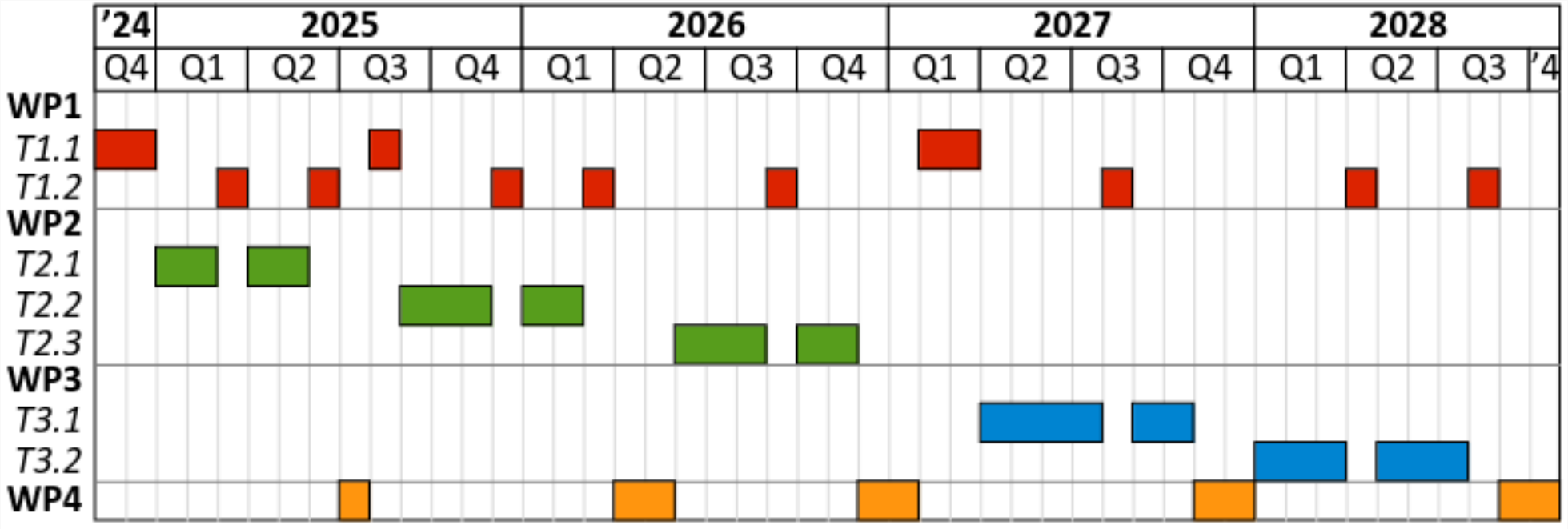


- ✓ Insight to update interfaces
- ✓ Insight to update query planning

- ✓ Relieves SMEs cost
- ✓ Data quality
- ✓ Sparks innovation

Supplemental slides

Work Plan


'24
2025
2026
2027
2028
Q4
Q1
Q2
Q3
Q4
Q1
Q2
Q3
Q4
Q1
Q2
Q3
Q4
Q1
Q2
Q3
'4
WP1
T1.1
T1.2
WP2
T2.1
T2.2
T2.3
WP3
T3.1
T3.2
WP4

Risks

- **Expressivity turn out infeasible**
  Migration: Statistical approach
  Cost: Loss of accuracy
- **Many interesting interfaces**
  Migration: Smaller selection
  Cost: Loss of heterogeneity
- **Update query planning too complex**
  Migration: Focus on single algorithm
  Cost: Loss of quality
- **Cross interface transactions cause infeasible overhead**
  Migration: Focus on native within-data store-transactions
  Cost: Loss of innovation

AI

- RAG to a single store so the AI really becomes a companion of your life
- NLP of natural language to query

Why me?

Jitse De Smet

- Fresh graduate, ready for the challenge
- Excellent study trajectory
- 3 years of development experience with research query engine
- Deep exposure to topical complexities through Master's thesis