



FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

Pavel Koupil, Sebastián Hricko, Irena Holubová

MM-infer: A Tool for Inference of Multi-Model Schemas

25th International Conference on Extending
Database Technology
Online (Edinburgh, United Kingdom)
29th March-1st April, 2022

Motivation

- ❖ *Variety* feature of Big Data
- ❖ *Level of support* of multiple data models in MMDBMS *varies greatly*
 - ❖ No unified approaches exist
 - ❖ Necessity of multiple *model-specific model and query constructs*
 - ❖ There is *no solid formal background*
- ❖ We need a representation that would allow us to
 - ❖ Capture all the existing data models, preferably in a *unified way*
 - ❖ *Query* across multiple *interconnected*, possibly *overlapping* models
 - ❖ Perform correct and complete *evolution management*
 - ❖ Enable *data migration*
 - ❖ Permit *integration* of new data models

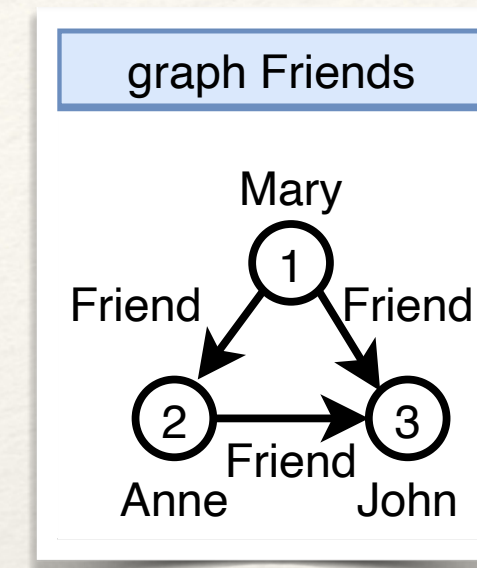


table Customer

CustomerId	FirstName	Address	Credit
1	Mary	...	3000
2	Anne	...	2000
3	John	...	5000

collection Order

```
{ OrderId : 220,
  Paid: true,
  Items: [
    { ProductId: T1, Name: toy,
      Price: 200, ItemQuantity: 2},
    { ProductId: B4, Name: book,
      Price: 150, ItemQuantity: 1 } ] }
```

column family Orders

CustomerId	Orders
1	[220, 230, 270, ...]
2	[10, 217]
3	[94, 137, 214, 370]

key / value pairs Cart

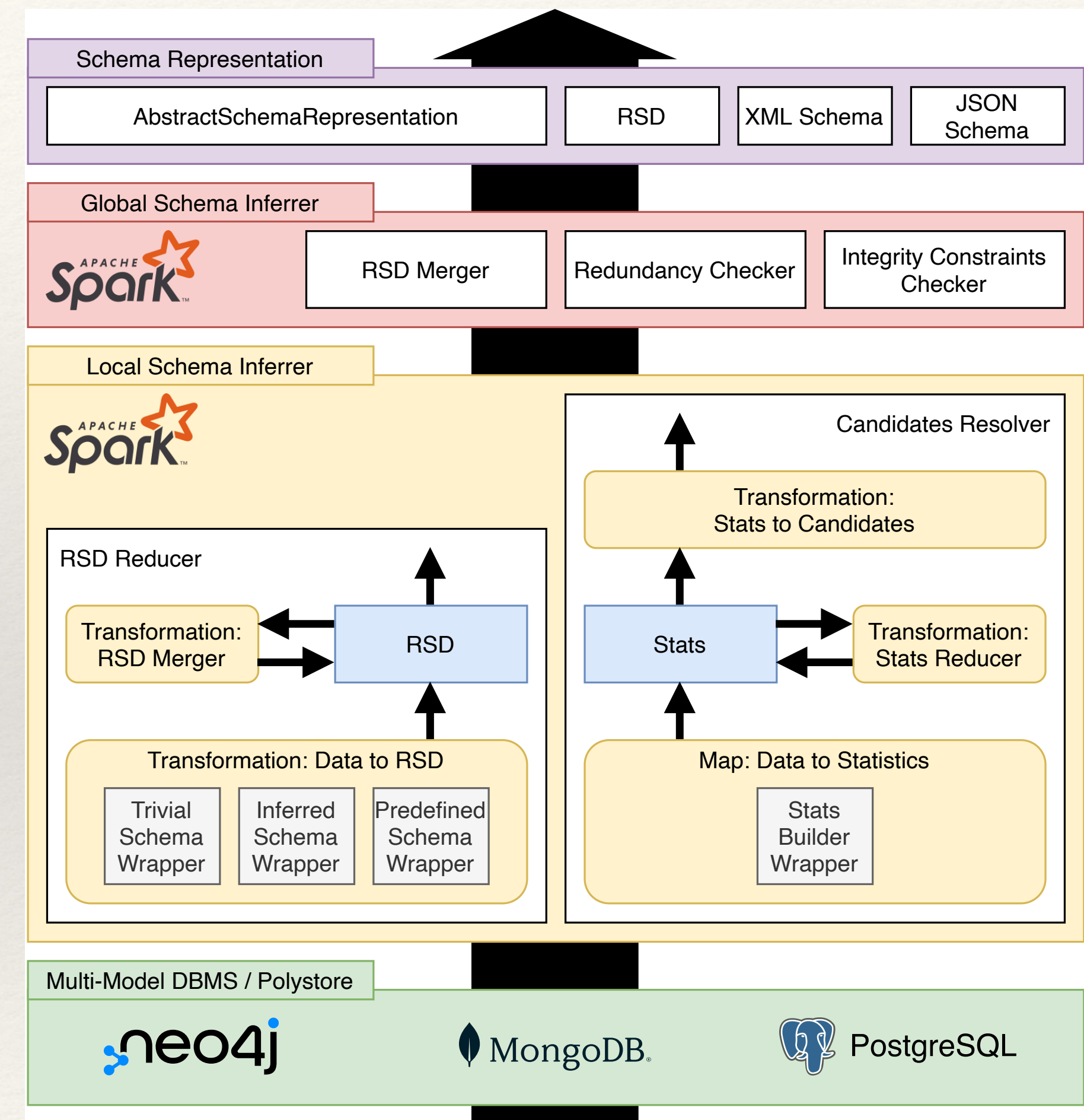
1	→ Product: T1, Quantity: 2 Product: B4, Quantity: 1
2	→ Product: H1, Quantity: 1
3	→ Product: B3, Quantity: 2

MM-infer: An Extensible Framework

- ❖ Multi-model schema inference modular framework
- ❖ Enables to *represent a multi-model schema* using a schema category
 - ❖ Can be *automatically extracted* from a conceptual model, e.g., ER model
 - ❖ Can *cover* (combinations of) *existing* logical *data models*
 - ❖ Can be *visualized using a multigraph*
- ❖ Enables to map the categorical model to any (combination of) DBMS(s)
 - ❖ *Implementation specifics* of particular DBMS(s) are *hidden* to user (demonstrated using MongoDB and PostgreSQL)
- ❖ Enables to transform the data from the underlying DBMS(s) to an instance category and vice versa, i.e., *data migration*
 - ❖ Instance category serves as a mediator for a unified representation of data instances

MM-infer: Modular Architecture

- ❖ Modular architecture
 - ❖ *Allowing simple extension* towards additional data models (DMBSs)
- ❖ *Advanced data management tasks* built on top of the categorical representation
 - ❖ Querying
 - ❖ Evolution Management
 - ❖ Data Migration
 - ❖ Schema Inference



Running Example

1. Unibench Scenario

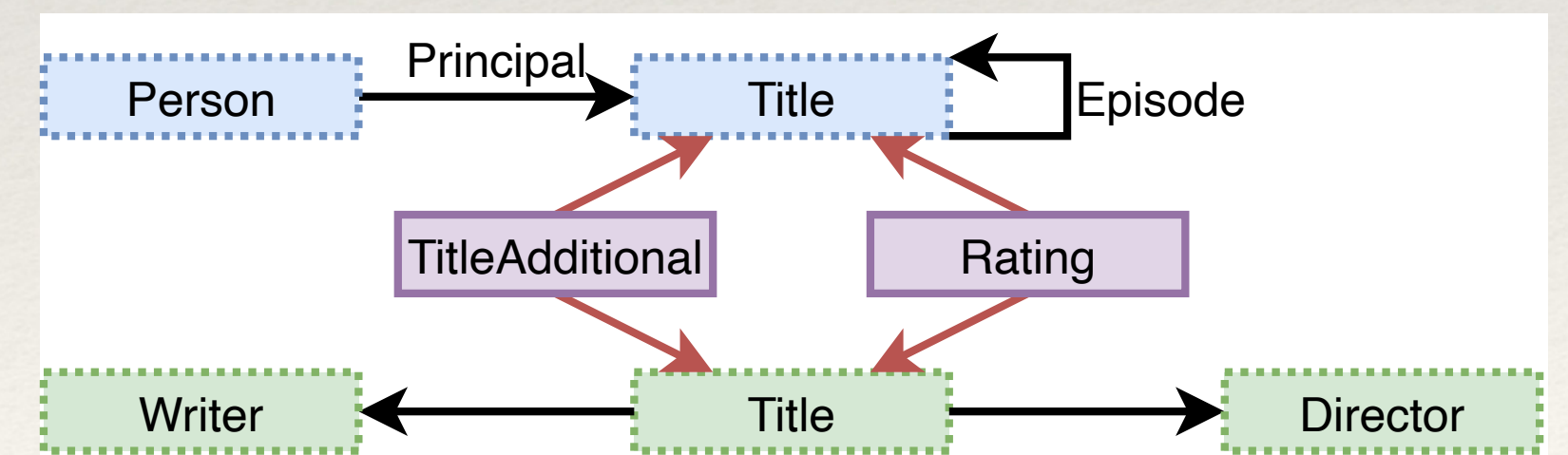
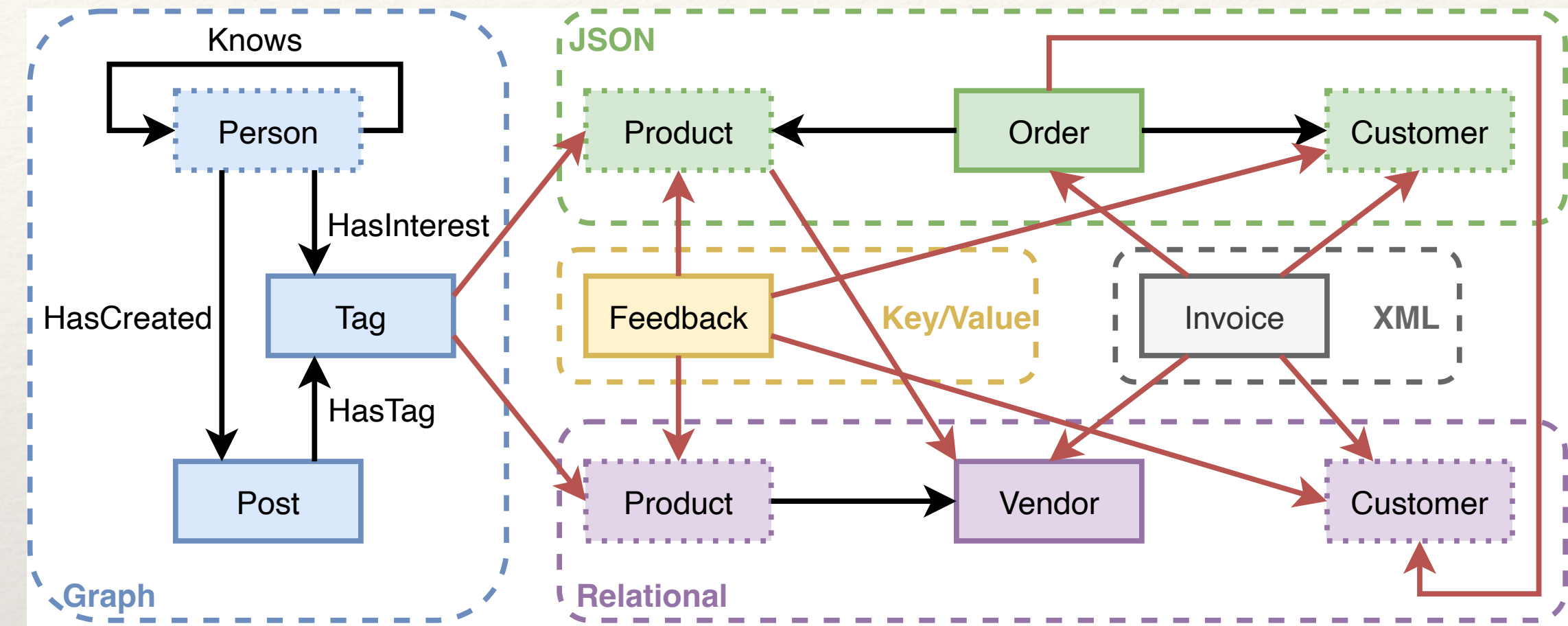
1. Selecting DBs

2. ...

2. IMDB Scenario (real data)

1. Inferred local schemas and candidates

2. Inferred multi-model schema



Summary

- ❖ Current version of MM-cat provides
 - ❖ Modular and extensible framework
 - ❖ A tool for *user-friendly schema inference* of multi-model data
 - ❖ Captures specifics of popular data models

- ❖ Extensions currently being added and future work:
 - ❖ vyjadreni schematu ruznymi formaty
 - ❖ Intra and inter model modifications of the schema, i.e., *evolution management*
 - ❖ Automatic *schema inference* over schema-less (and schema-mixed) multi-model data and its *mapping to schema category*

Thank you for your attention...