Toward A Model-Driven Design Tool for Big Data Architectures

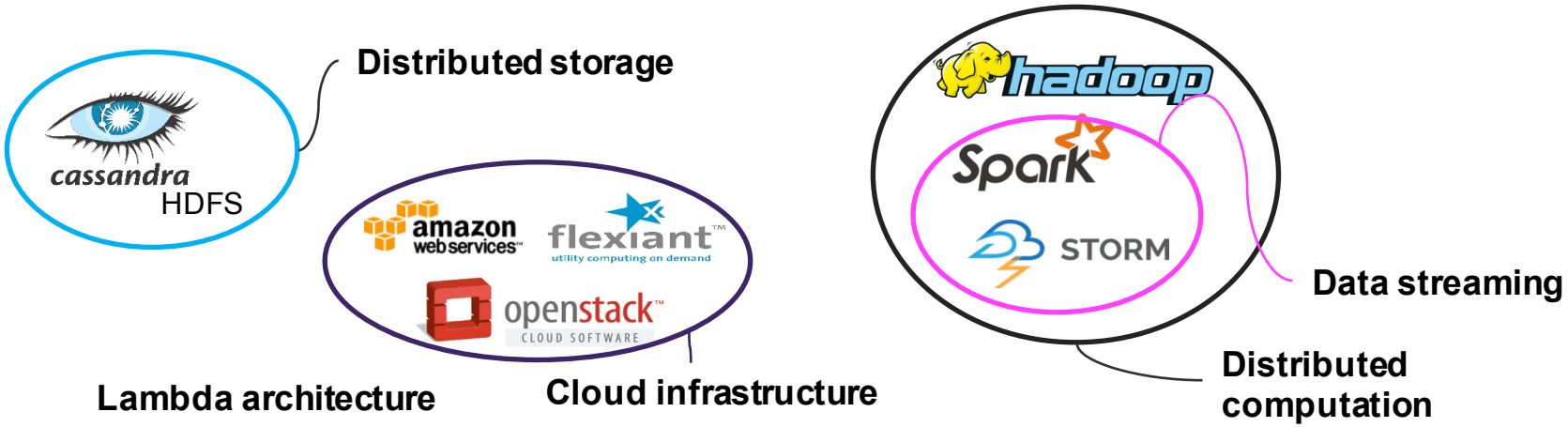
Michele Guerriero

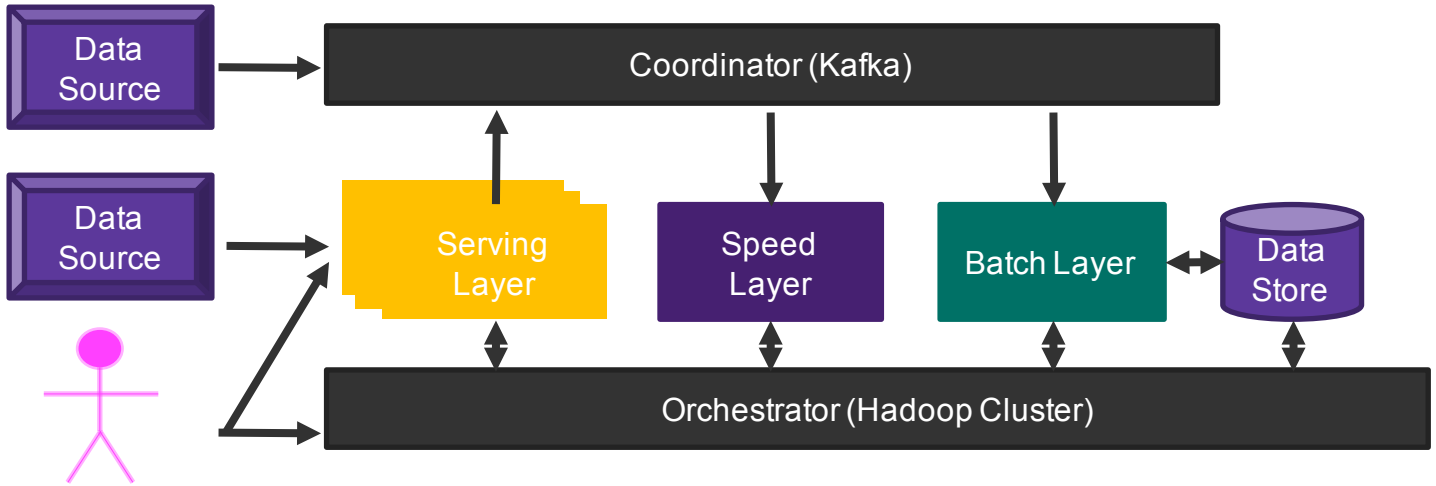
Saeed Tajfar

Damian Andrew Tamburri

Elisabetta Di Nitto

Introduction





Questions

Goals

How many Big Data technologies do I need to know and combine?

Simplify software design and reduce costs

Which resources, how many do I need and how do we configure the deployed technologies?

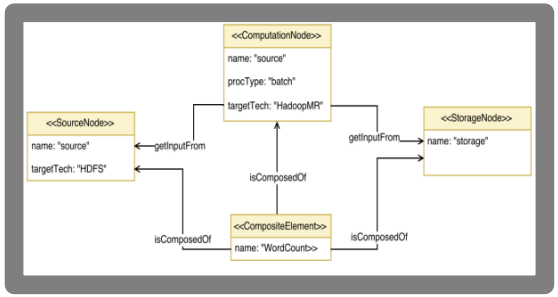
Simplify Deployment

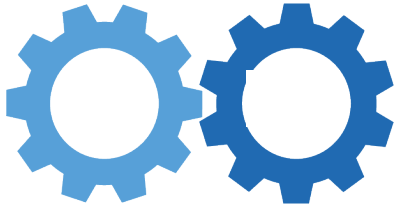
What if I want to know properties and performance of my application?

Support Analysis

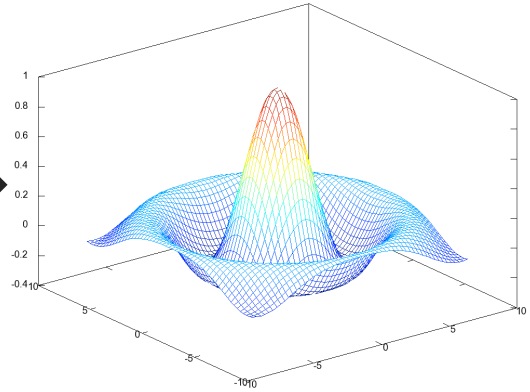
POLITECNICO
MILANO 1863

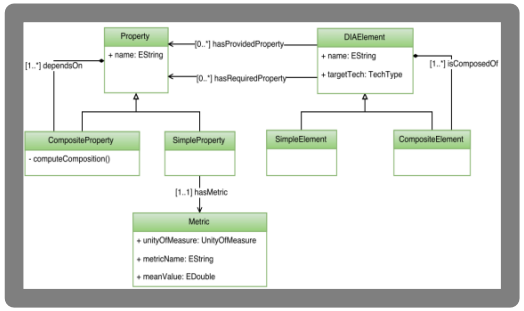
Model-Driven Engineering





Analysis





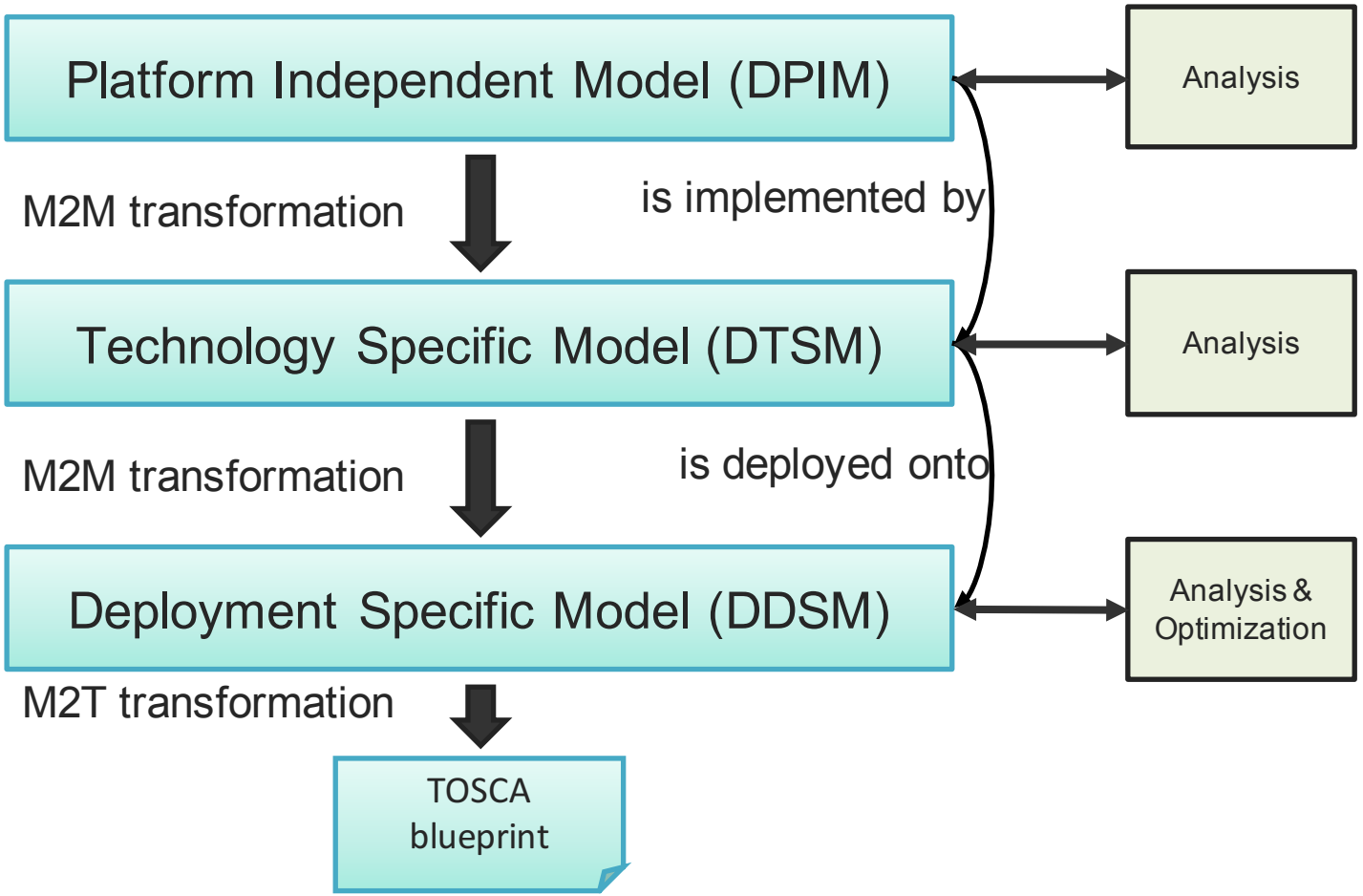
```
tosca.nodes.Database:
  derived_from: tosca.nodes.Root
  properties:
    db_user:
      type: string
      description: user account name for DB administration
    db_password:
      type: string
      description: the password for the DB user account
    db_port:
      type: integer
      description: the port the underlying database service will listen to data
    db_name:
      type: string
      description: the logical name of the database
  requirements:
    - host: tosca.nodes.DBMS
  capabilities:
    - database_endpoint: tosca.capabilities.DatabaseEndpoint
```

Deployment blueprint

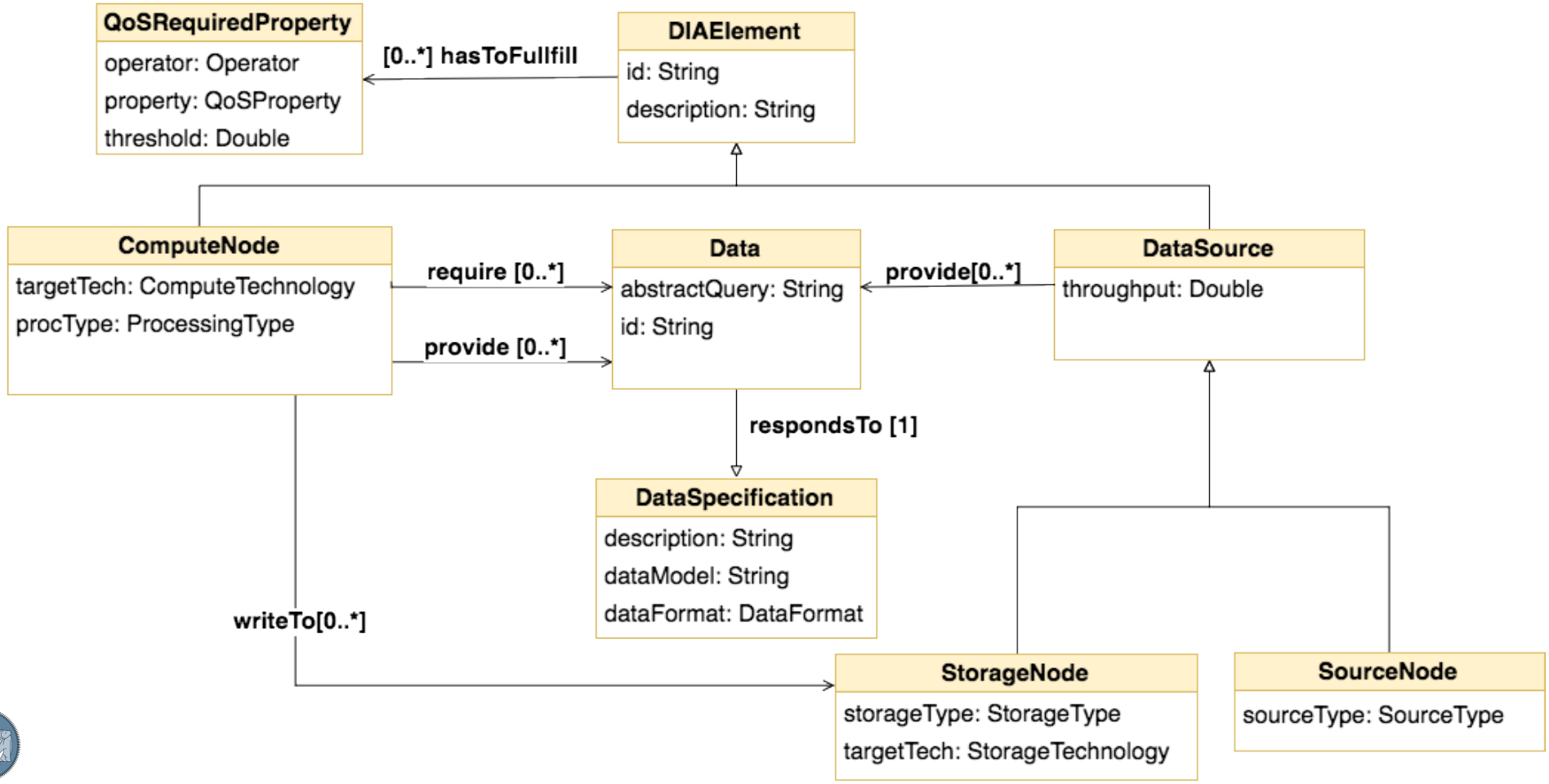
DICE

POLITECNICO
MILANO 1863

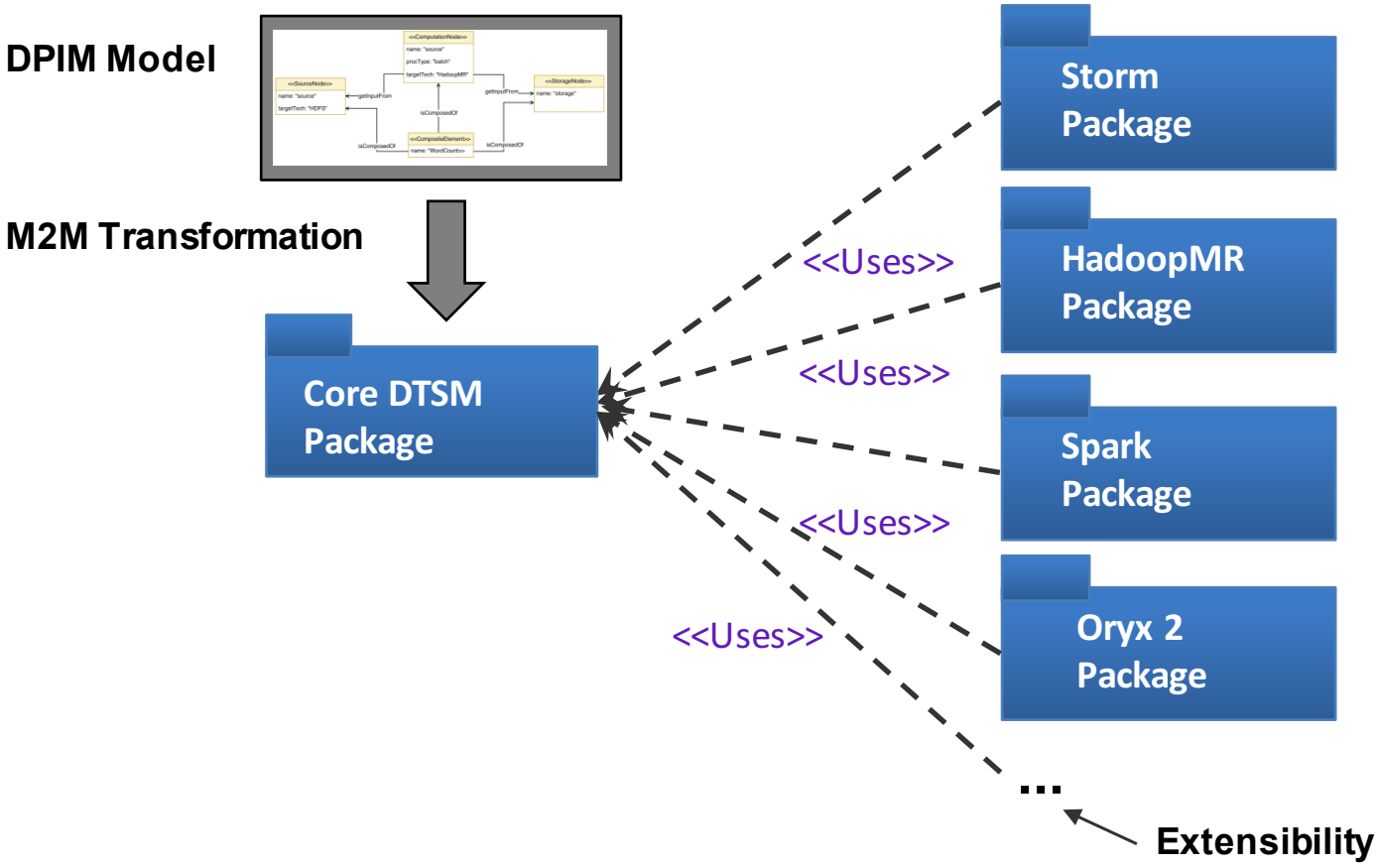
Model-Driven Big Data Design Architecture



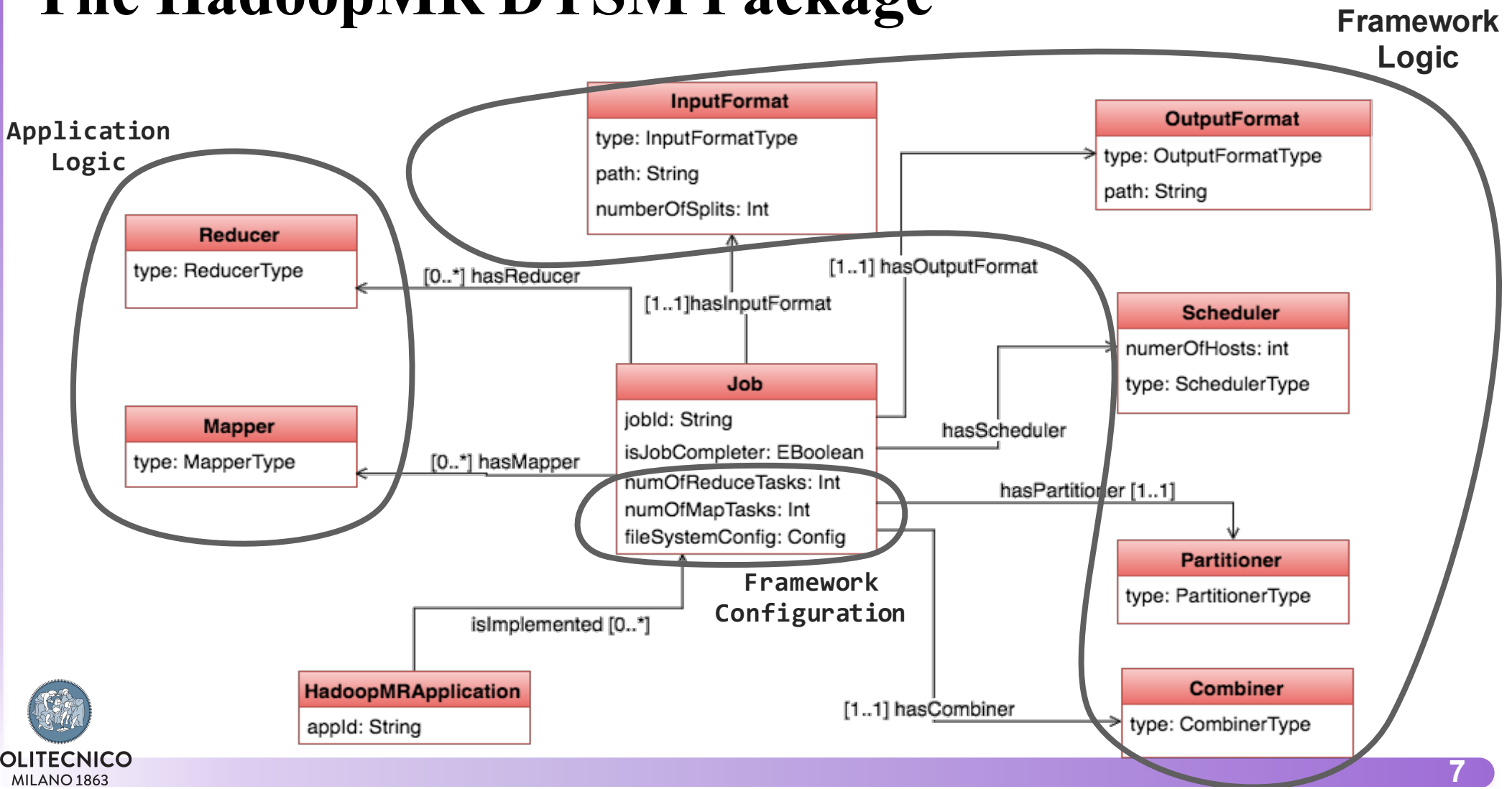
DPIM Meta-Model



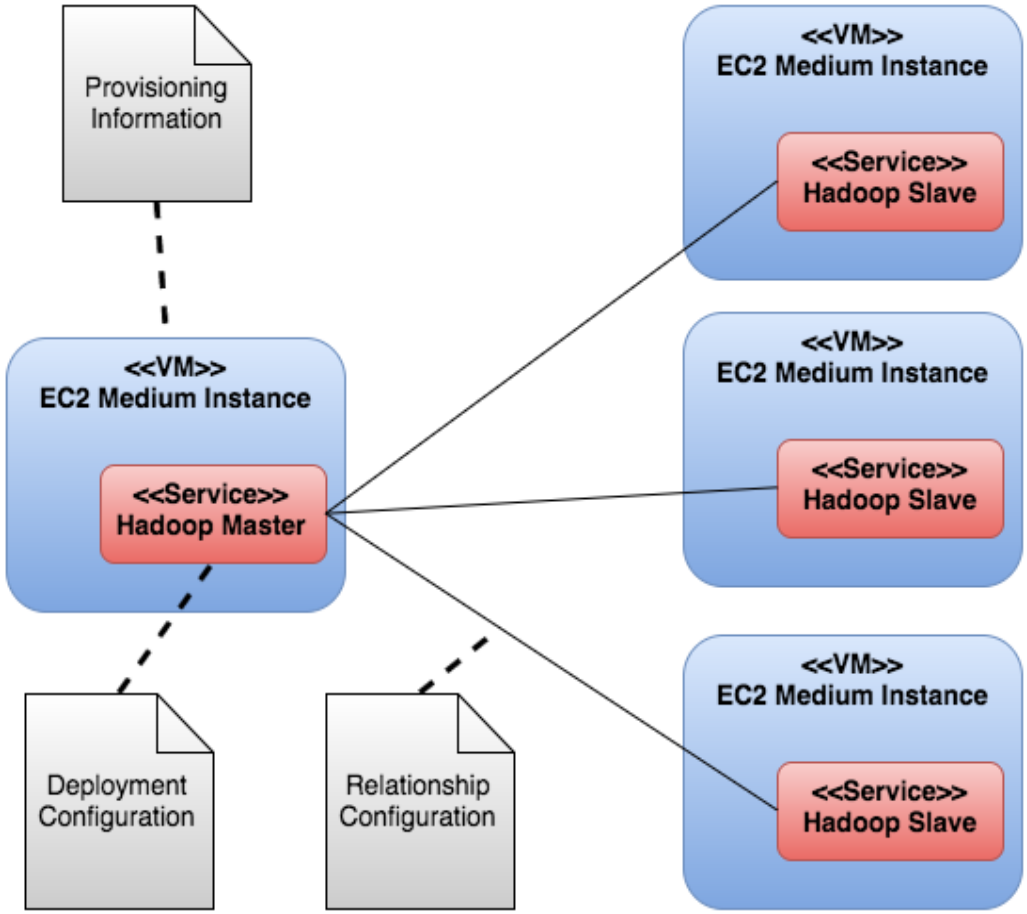
DTSM Meta-Model



The HadoopMR DTSM Package

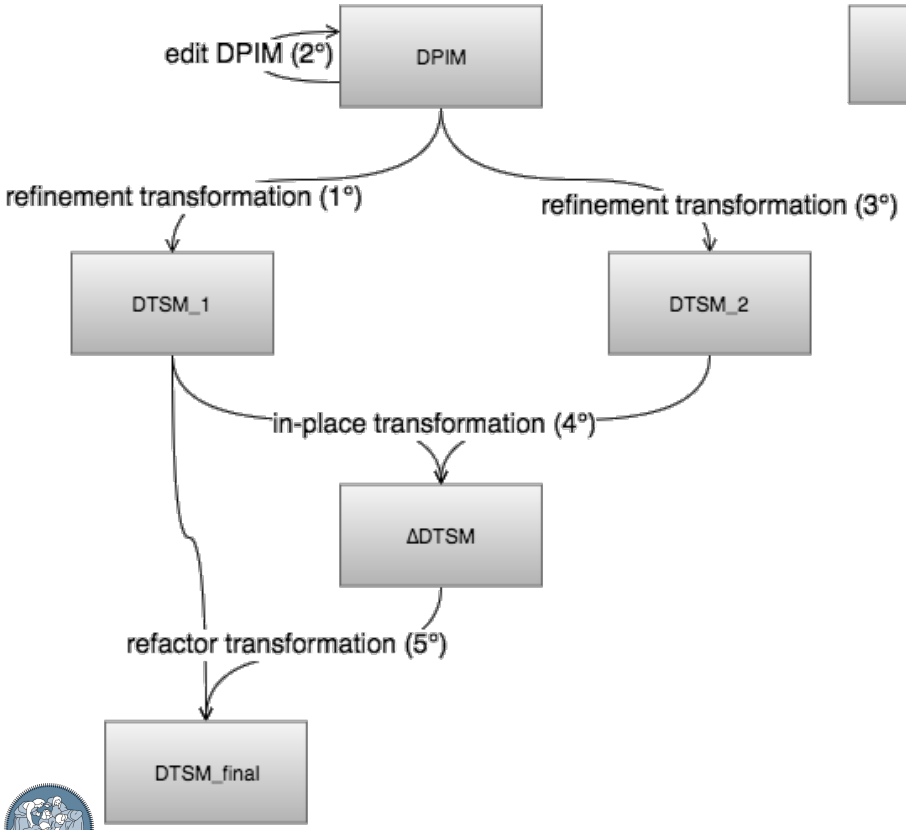


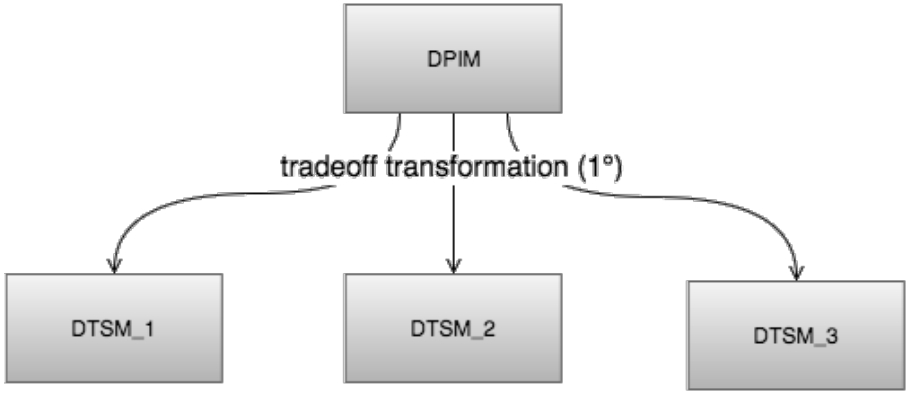
DDSM Meta-Model



- Build **Deployment Topology**
- Standards adoption (**TOSCA**)
- Deployment **Technological Packages**

Transformations Set





Conclusion and Future Works

- Preliminary steps toward the **model-driven engineering of Big Data applications**.
- **Great potential** behind MDE for data-intensive applications!
- Future steps:
  - increase **models expressiveness** (data quality, privacy concerns)
  - **validation** against industrial case studies
  - increase **automation** mechanisms
  - **technological support**

Thanks!