Toreador

TrustwOrthy model-awaRE Analytics Data platfORm

Conceptual Overview

Marco Anisetti - Università degli Studi di Milano – Dipartimento di Informatica

GARR Workshop April 19, 2016

Consortium

Project Coordinator
PI: Prof. Ernesto Damiani

BIG DATA National Lab Units:
UNIMI, UNIBO, UNIBA, SUN, UNITO, POLITO

Mission

- Many companies and organizations in Europe have become aware of the **potential competitive advantage** they could get by timely and accurate **Big Data analytics**, but lack the IT expertise and budget to fully exploit BDA.
- The TOREADOR project is aimed at overcoming some major hurdles that until now have prevented many European companies from reaping the full benefits of Big Data Analytics.

Objectives

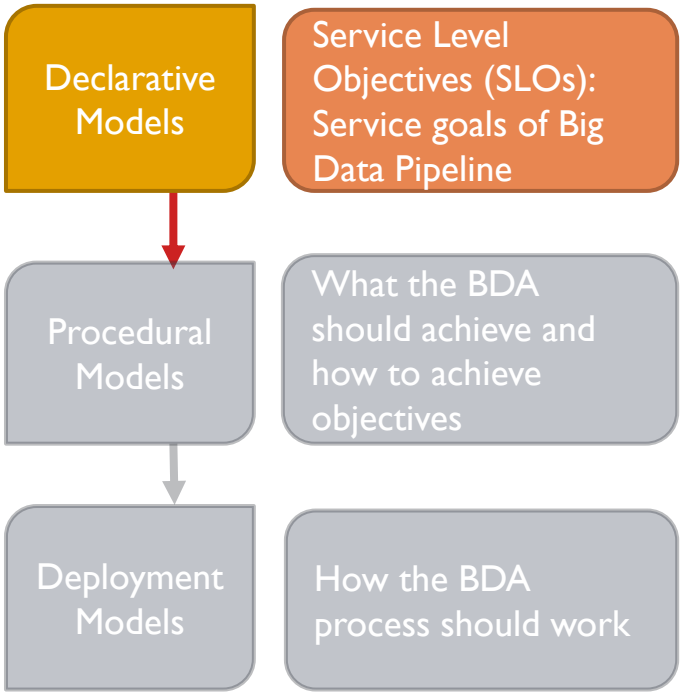
- Specification of a fully **declarative** framework and a **model set** supporting Big Data analytics.
- **MBDAaaS** model-based **BDA-as-a-service** providing models of the entire Big Data analysis process and of its artefacts.
  - **Automation** and **commoditization** of Big Data analytics.
  - Enabling it to be **easily tailored** to domain-specific customer requirements.
- **SLA** and **assurance** approaches to guarantee contractual quality, performance, and security of BDA.
- Design and development of **automatic deployment** of TOREADOR analytic solutions.

Pilots

- SAP: “The Application Log Analysis Pilot” (security log).
- LIGHTSOURCE: “The Energy Production Data Analysis Pilot” (sensor-based management of photovoltaic).
- JoT: “The Clickstream Analysis Pilot” (e.g., fraud control).
- DTA: “The Aerospace Products Manufacturing Analysis Pilot” (data related to manufacturing process).

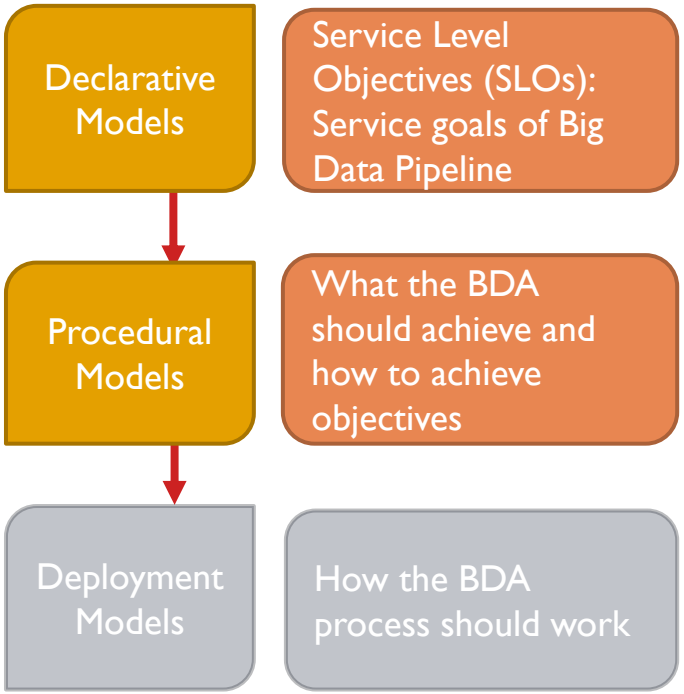
TOREADOR Overview: Declarative Model

- Specify goals as service level objectives (SLO).
- Five areas: representation, preparation, analytics, processing, display and reporting.
  - A single model because aspects of different areas may impact on the same procedural model template.
  - Based on a controlled vocabulary.
    - Controlled names (e.g., anonymization).
    - Values in an ordinal scale as strings or numbers (e.g., obfuscation, hashing, k-anonymity).
  - Insufficient to run a Big Data Analytics.
  - Incomplete.



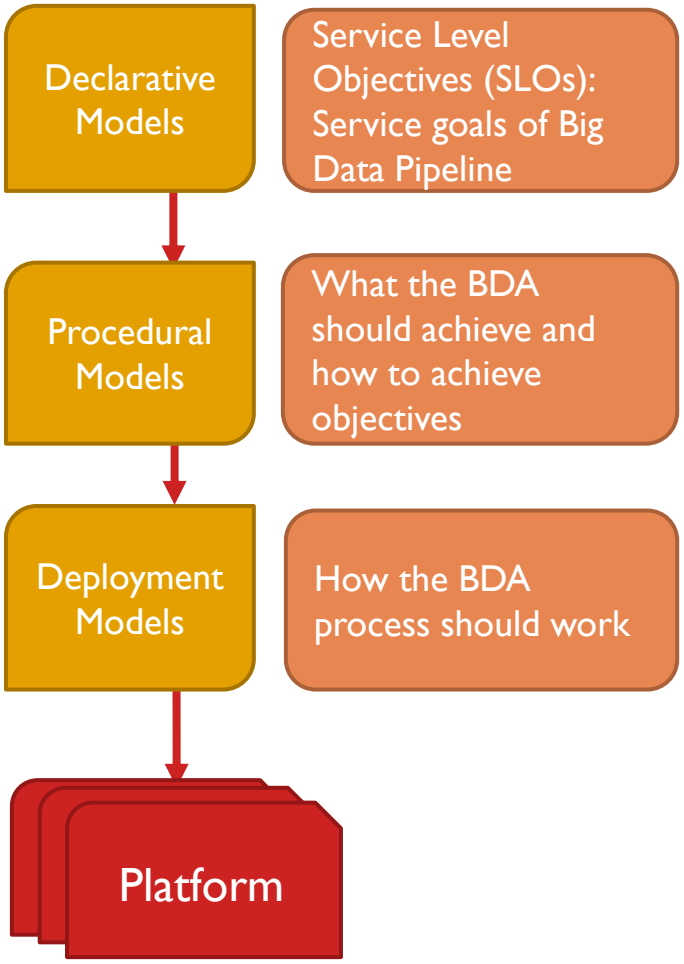
TOREADOR Overview: Procedural Model

- Complete the declarative model with all information needed for running analytics.
  - Simple to map SLOs on procedures.
  - Platform independent.
- Specified using a function that takes as input SLOs and returns as output procedural templates.
  - Procedural templates are pre-calculated based on defined SLOs.
- Templates express competences of data scientist and data technologist.
  - Declarative models used to select the (set of) proper templates.



TOREADOR Overview: Deployment Model

- Specify how procedural models are incarnated in a ready-to-be-deployed architecture.
- Drive analytics execution in real scenarios.
- To be defined for each TOREADOR application scenarios.
- Platform Dependent.



The contribution of TOREADOR

- Use of TOREADOR models in non ICT areas.
  - e.g., social sciences, human sciences, healthcare.
- Based on TOREADOR pilots' experiences.
  - ready to be deployed BIG DATA models.
- Deployment campaign on Italian Universities Consortium platforms (e.g. CINECA, GARR).

Thank you