Aitor Murguzur¹, Johannes M. Schleicher², Hong-Linh Truong², Salvador Trujillo¹, and Schahram Dustdar²

¹ IK4-Ikerlan Research Center, Spain
² Distributed Systems Group, TU-Wien, Austria

DRain: An Engine for Quality-of-Result driven Process-based Data Analytics
The URBEM Scenario

A holistic view of the city

Free from isolation and integrated into the city.

URBEM project: http://urbem.tuwien.ac.at
Motivation

Processes/services can offer the same functionality, each with the associated QoR parameters.
Process Configuration

PROCESS FAMILY

PROCESS VARIABILITY

Negative variability modeling

Base model and fragments

Positive variability modeling

Legend: √ Completed | • Enabled
The DRain Framework: Overview

**MODELING**

- BASE MODELS & FRAGMENTS
- COMPUTATIONAL & DATA MODELS
- INTENT & QUALITY OF RESULT MODELS (IQoRM)

**SELECTION, EXECUTION, PROCESSING & CONFIGURATION**

- RDMS
- NoSQL

**Domain Model**

**Variability Model**

**PROCESS SELECTOR SERVICE** (SelM)
**PROCESS INTERACT. SERVICE** (DRi, AdaWR)
**PROCESS ENGINE SERVICE** (EnM)

**Data API**
**User API**

**deploy**

**Data acquisition**
**intent request**

**DRain**: QoR-driven selection and configuration of data-aware processes
A **base model** represents the commonality shared by a process family in a particular domain. DRi activities (variation points) enable QoR-driven late binding.

A **fragment** describes a single variant realization option for each variation point within a particular base model.
DRi activity

During operation

DRi activities indicate possible points in a base model and/or fragment where data interaction and fragment binding occur (fragment selection based on context data).

- *DRi* activities are automatically determined at runtime in subsequent QoR-driven data interactions.
**Intent and Quality of Result Model (IQoRM):** It is used to construct a data analytics task and its strategy, and thus represents constraints for the desired behaviour.
Variability & Domain Models

VARIABILITY MODEL

All configuration alternatives

DOMAIN MODEL

The domain knowledge and semantics

Mappings
- feature names are mapped to **BaseModel** (→) and **Fragment** (→) individuals **hasProcessKey** data property.
- variation point features are linked to **hasServiceName** data property of **DataEndpoint** (→) individual.
- configuration model attributes are related to data model variables from each **DataEndpoint** (→).
The DRain Approach: Overview

**DRain:** QoR-driven selection and configuration of data-aware processes
Selection, Execution, Processing & Configuration

QoR-driven Process Config.
(1) Intent request
(2) QoR-based base model selection
(3) Analytics process execution
(4) QoR-based data endpoint selection
(5) Context data gathering
(6) DRI resolution (fragment binding)

**RESOLUTION STRATEGIES: How to proceed with different base model and data endpoint choices?**
- **no-solution**
  - No solution for the given QoR.
- **one-solution**
  - One solution for the given QoR.
- **n-solutions**
  - More than one solution. Not sure what to choose? In this case, we adopt a simple *get-first* solution.
Evaluation

PROVIDED MODELS

- 30 base models were created, each with 4 DRi activities and different time and cost constraints for a synthetic URBEM scenario.
  - Each DRi activity contained 2 fragments, providing each data endpoint different data quality (availability and accuracy).
  - 24 data endpoints with different QoR and 5 data values were parsed in each data interaction (DRi activity).

METRICS

- The aim was to assess the computation time of the engine.
  - Time for base model retrieval (TSelM): This metric measures the time required for intent-driven and QoR-based base model searching.
  - Time for data endpoint retrieval (TDRi): This metric defines the timespan from DRi activity initialisation to the moment when the process interaction service finds a suitable data endpoint for the given QoR and invoked the particular REST resource to collect data.
  - Time for fragment solving (TAdaWR): This metrics measures the time required to establish context values and find a suitable fragment once data is gathered from a REST resource.

- 200 intent-requests were processed.
  - On a 13-inch MacBook Air with 8GB 1600 MHz DDR3 RAM, and Core i7 running @2 GHz
  - get-first strategy was enabled as default n-solutions strategy.
Evaluation: TSelM

The difference between the minimum and maximum time required for a base model retrieval based on a user-defined QoR is about 3ms.
The average time for data endpoint selection and processing is reasonable at **184.019ms**.
Evaluation: TAdaWR

The average time required to complete the runtime configuration (for processing context values and determining a fragment choice) is about **2.986ms**.
Issues

- Error handling: How can we guarantee the correct intent request execution?
- Strong validation: Assessment in a complex URBEM scenario.
Summary

- Context-aware and data-intensive environments require abstractions to select relevant analytic processes (exposed as WFaaS) and data endpoints based on user-defined QoR, and flexibility in terms of runtime process configuration.

- The framework is capable of QoR-driven runtime process configuration, with reasonable selection, processing and configuration performance.
  • Employing a high-level IQoRM.

- Evaluation cases diverse and large scale, but not ideal.
  • Still good indication.
Future Work

- **Industrial** empirical evaluation (URBEM).
- Adapt the **IQoRM** for a more domain-specific environment.
- Adopt **ranking and selection algorithms/dimensions** using QoR.
Thank you!

Questions?

@amurguzur #bpm2014
amurguzur@ikerlan.es