DICE: Quality-Aware DevOps For Big Data Applications

Giuliano Casale
Imperial College London

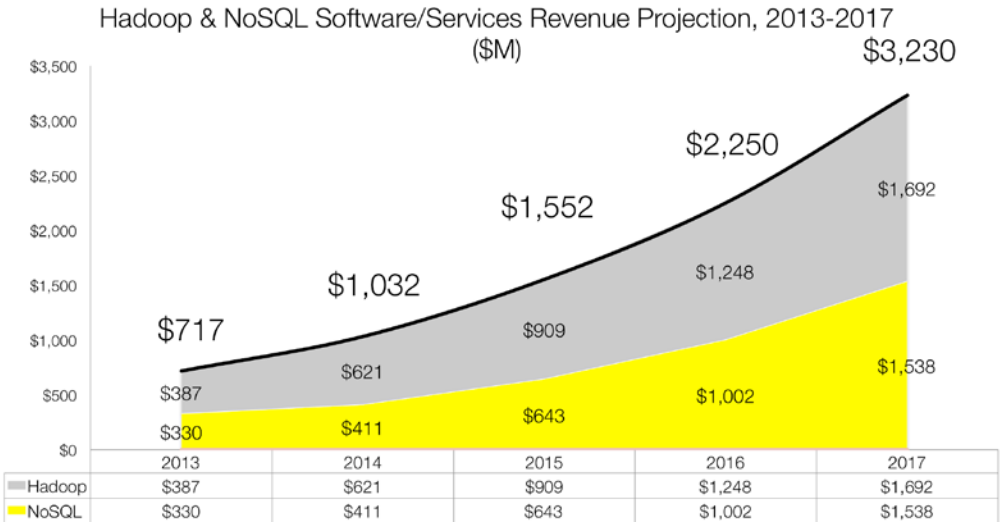
DICE Horizon 2020 Project
Grant Agreement no. 644869
http://www.dice-h2020.eu

Funded by the Horizon 2020
Framework Programme of the European Union

The Rapid Growth of Big Data

- Software market rapidly shifting to Big data
  - 27% compound annual growth rate through 2017 (IDC)
  - Popular technologies such as Spark, Hadoop, and NoSQL boost Big Data adoption and revenues from new services

Source: IDC



Source: Wikibon

Business issue: 65% of Big data projects still fail (CapGemini)

What problems EU SMEs face?

An example from our consortium

Traditional market:
Legacy software systems

- *Learning curves*
- *Initial prototype*
- *Risk of failure*
- *Fast-paced market*

Customers with legacy data now ask for Big Data technologies

(+ others…)

Growth in sight, but …

DICE: DevOps for DIAs

Mission: support SMEs in developing high-quality cloud-based data-intensive applications (DIAs)

- Horizon 2020 research project (4M€, 2015-18)
- 9 partners (Academia & SMEs), 7 EU countries

Imperial College London

XLAB NOT IDLE

@ IeAT

Universidad Zaragoza

NETFECTIVE Creative Technology

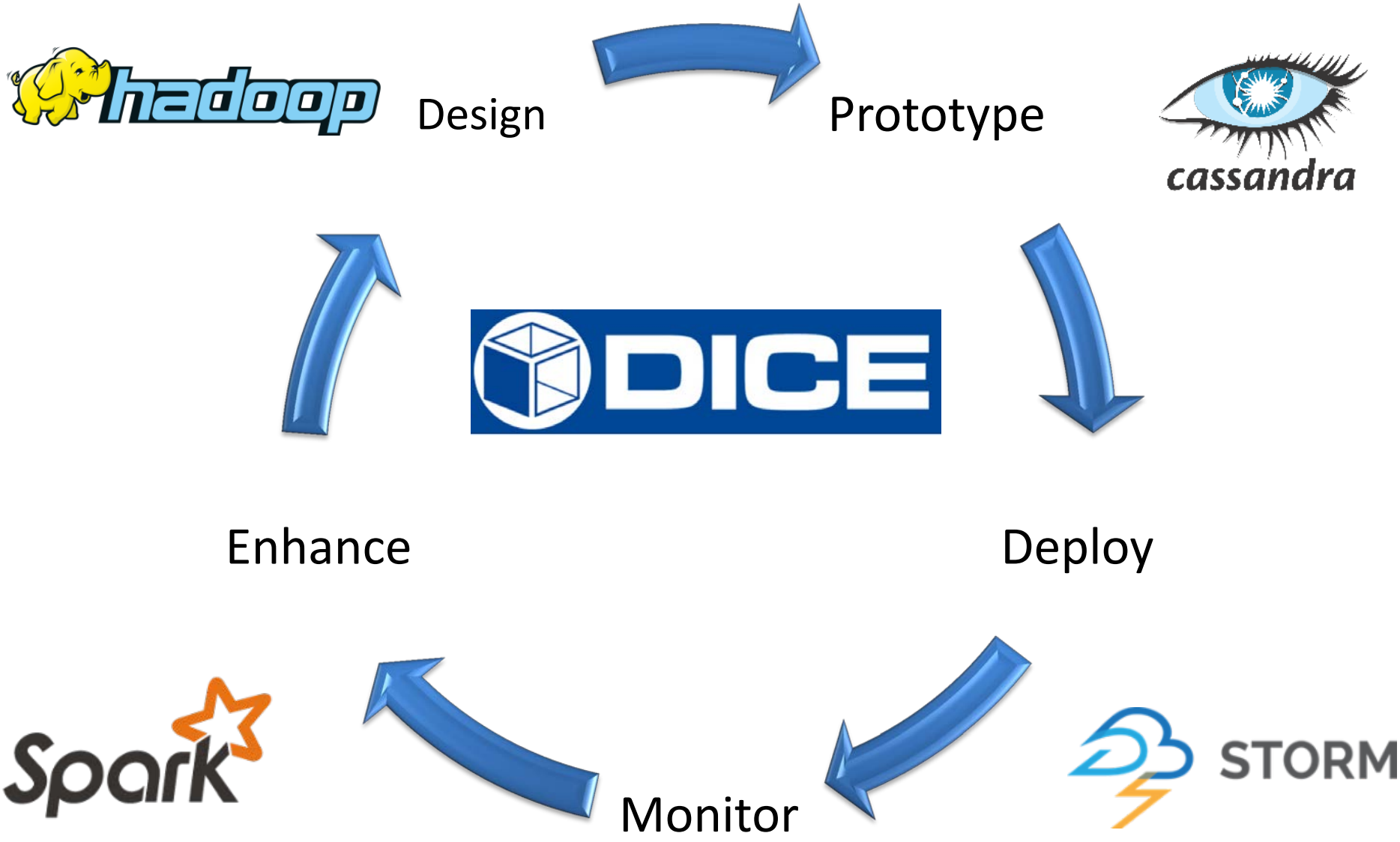
ATC ATHENS TECHNOLOGY CENTER

POLITECNICO DI MILANO

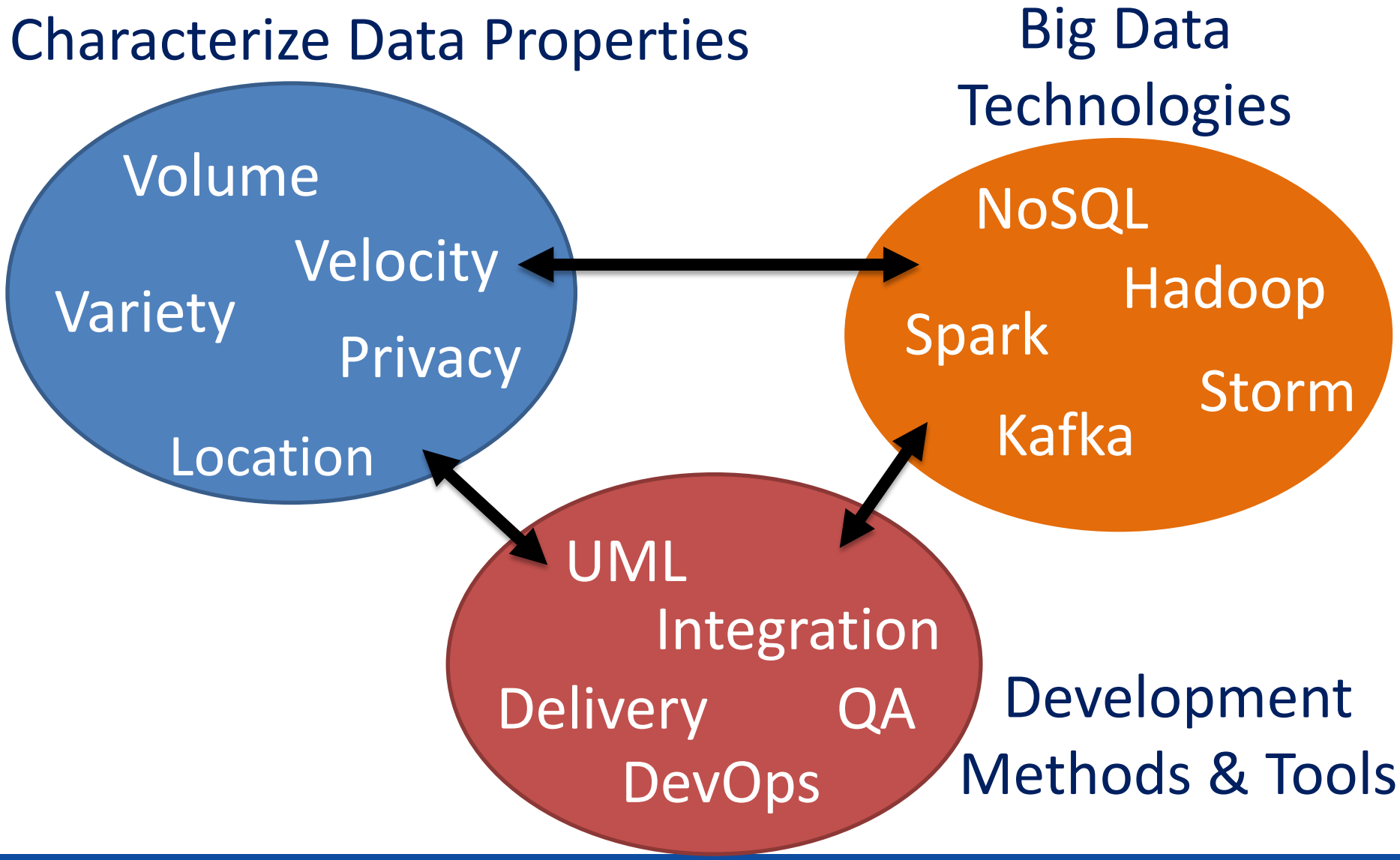
prodevelop Integración de tecnologías

flexiOPS

DICE: Quality-Aware DevOps for Big Data

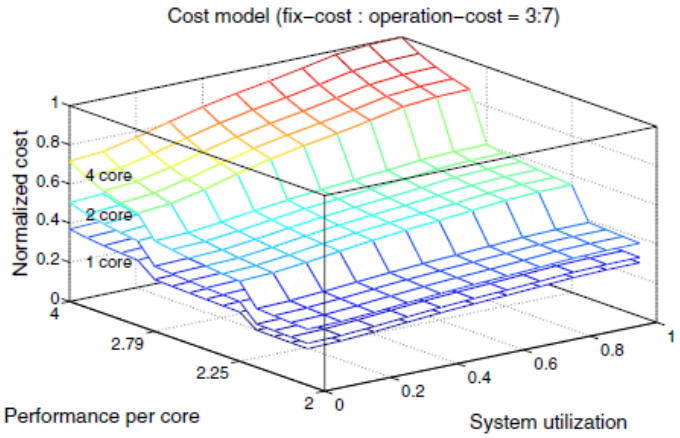


How to support DIA development?

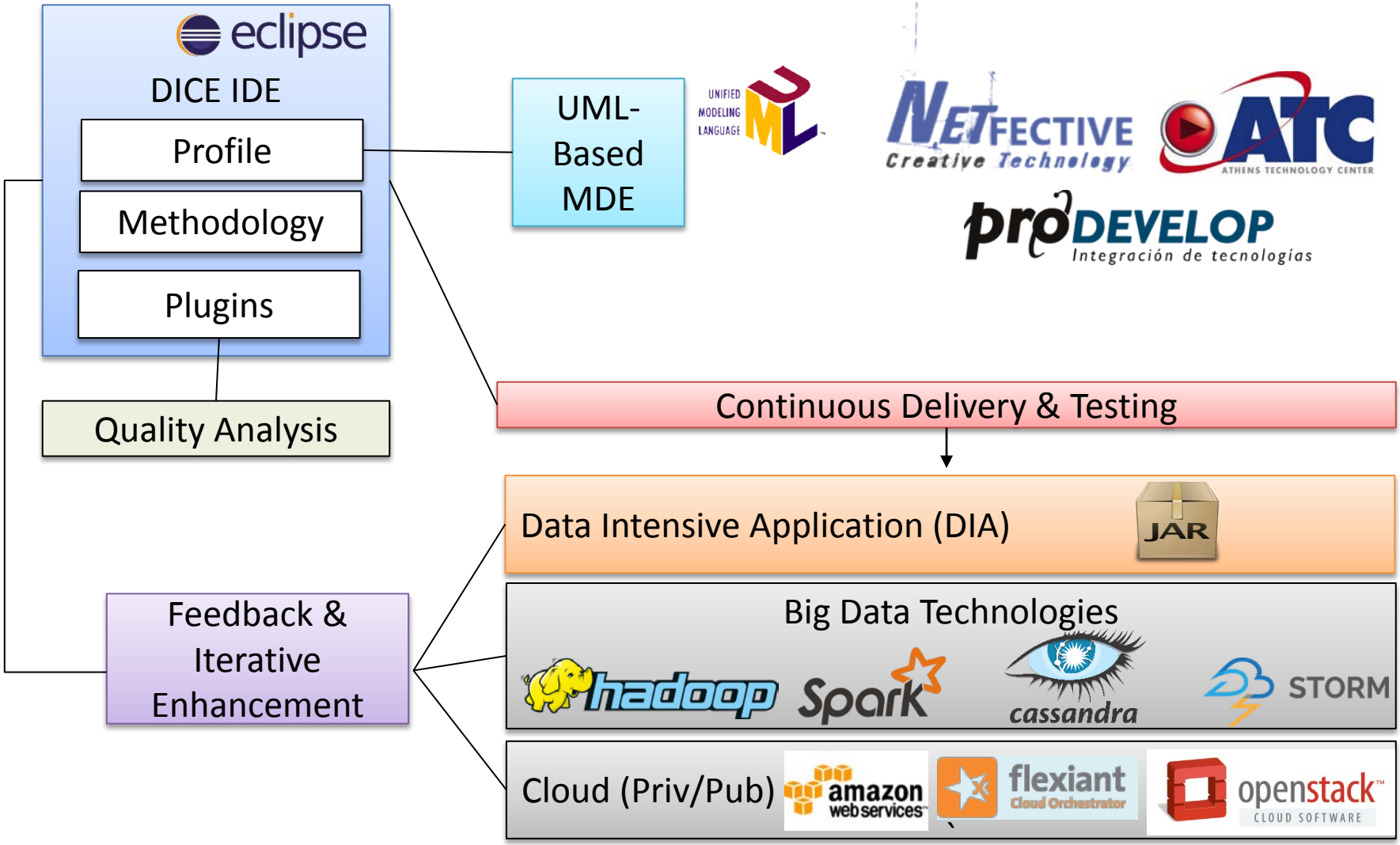


What do we mean by Quality?

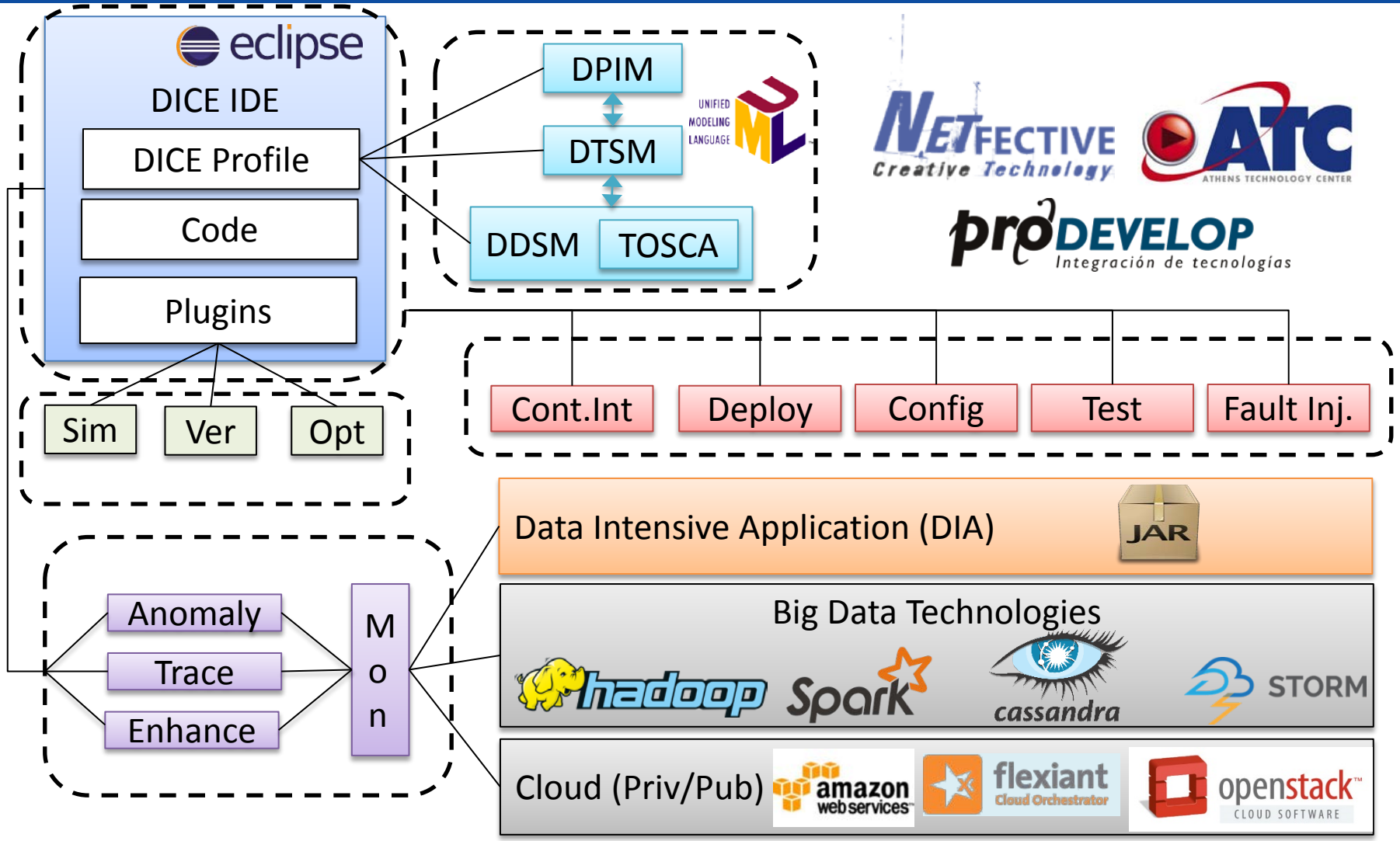
- Reliability
  - Availability
  - Fault-tolerance
- Efficiency
  - Performance
  - Costs
- Correctness
  - Privacy & security
  - Temporal metrics



DICE Framework



DICE Framework



DICE Workflow - Dev

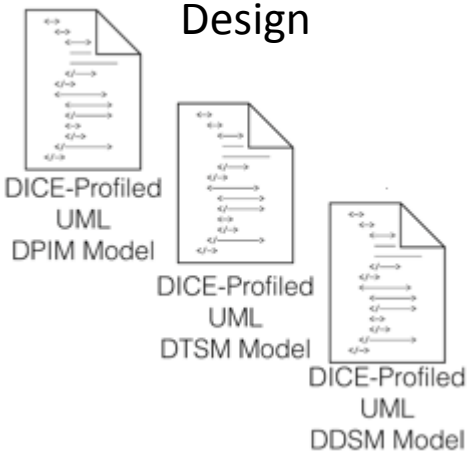
"Will the DIA meet SLAs and costs?"

DICE Eclipse IDE

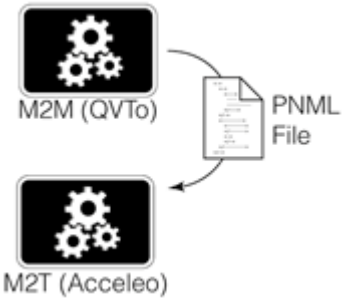
Enhance

"I want to design for Big data"

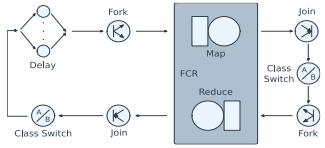
Design



Transform to Formal Models

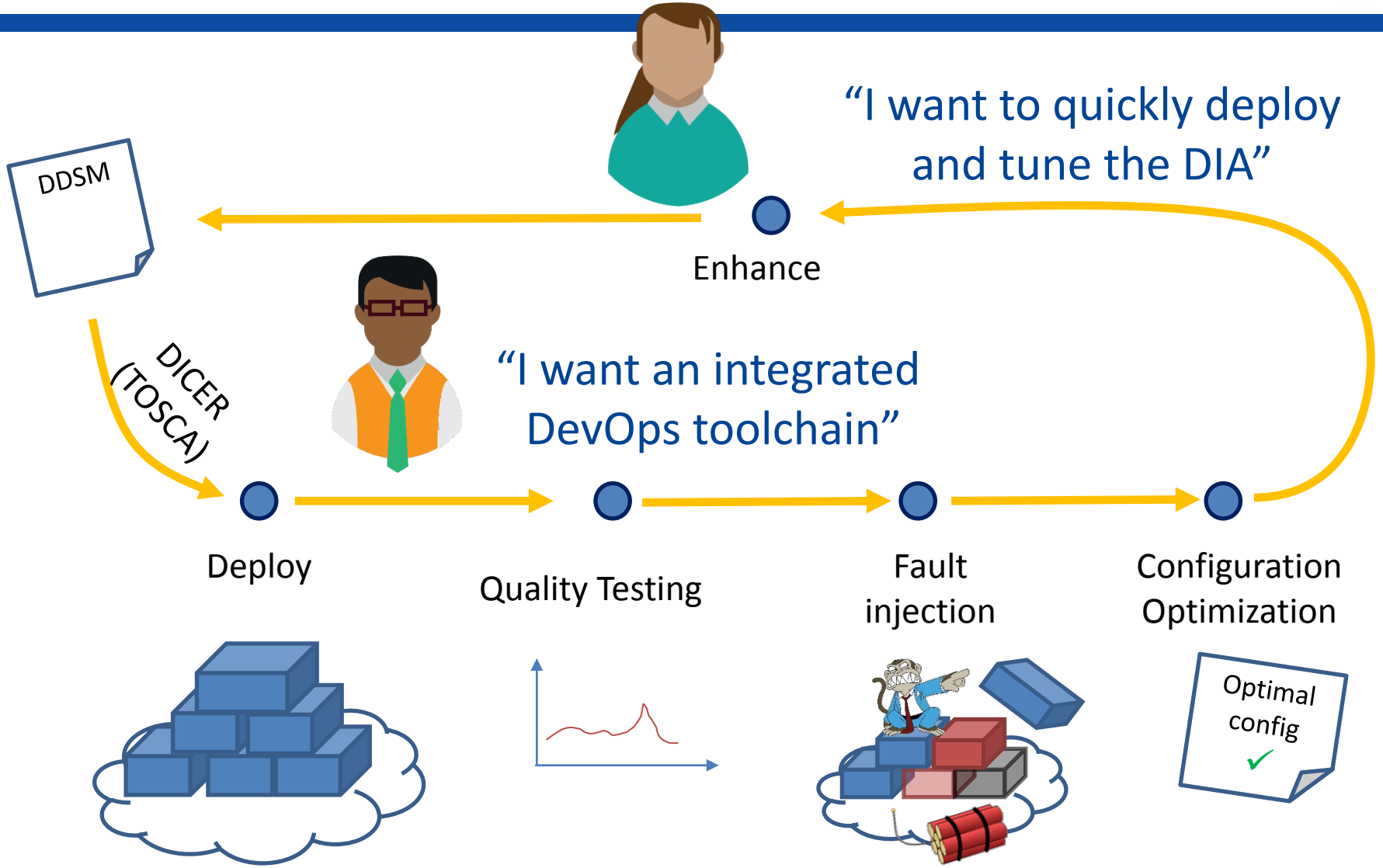


Simulate & Verify



Optimize

DICE Workflow - Ops



DICE Workflow: Unified Toolchain

Following the DevOps paradigm, DICE delivers a unified toolchain for the enterprise team

DICE Eclipse IDE

Enhance

Enhance

Deploy

Quality Testing

Fault injection

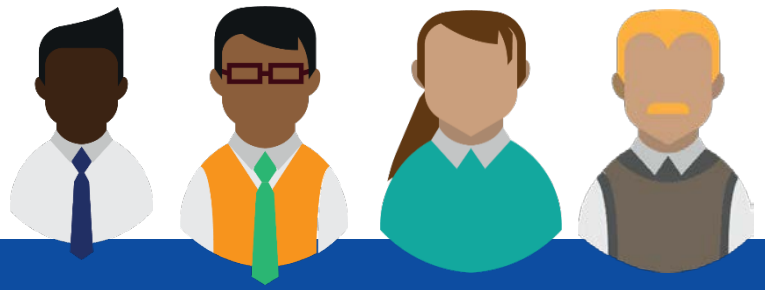
Configuration Optimization

Design

Transform to Formal Models

Simulate & Verify

Optimize



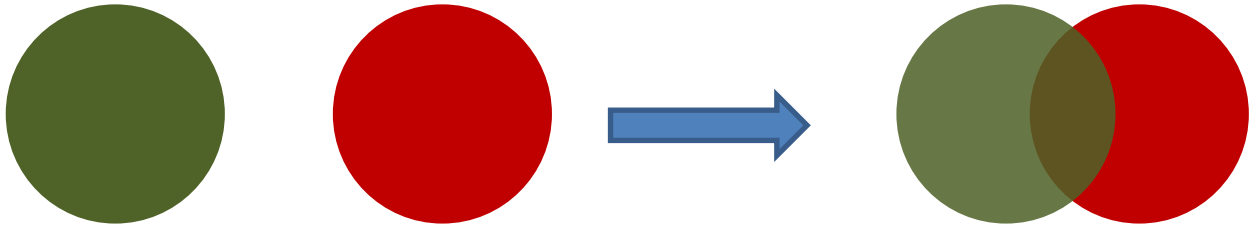
DICE Approach, Tools, Architecture and Methodology

DICE Horizon 2020 Project
Grant Agreement no. 644869
http://www.dice-h2020.eu

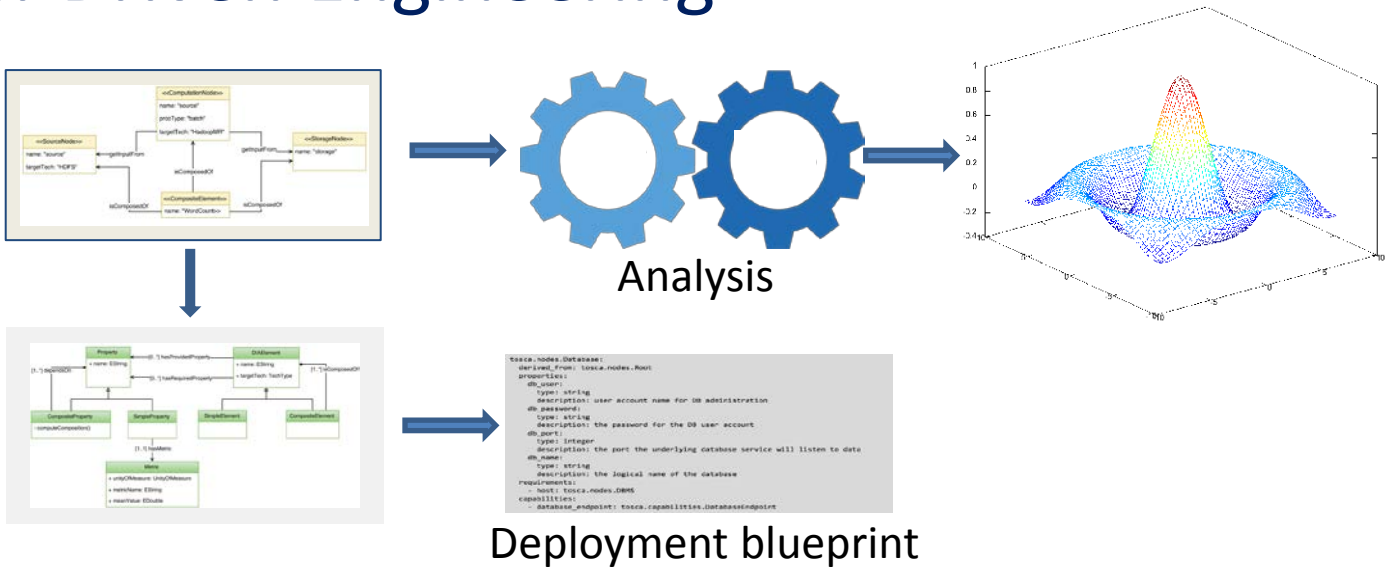
Funded by the Horizon 2020
Framework Programme of the European Union

Ingredients of technical approach

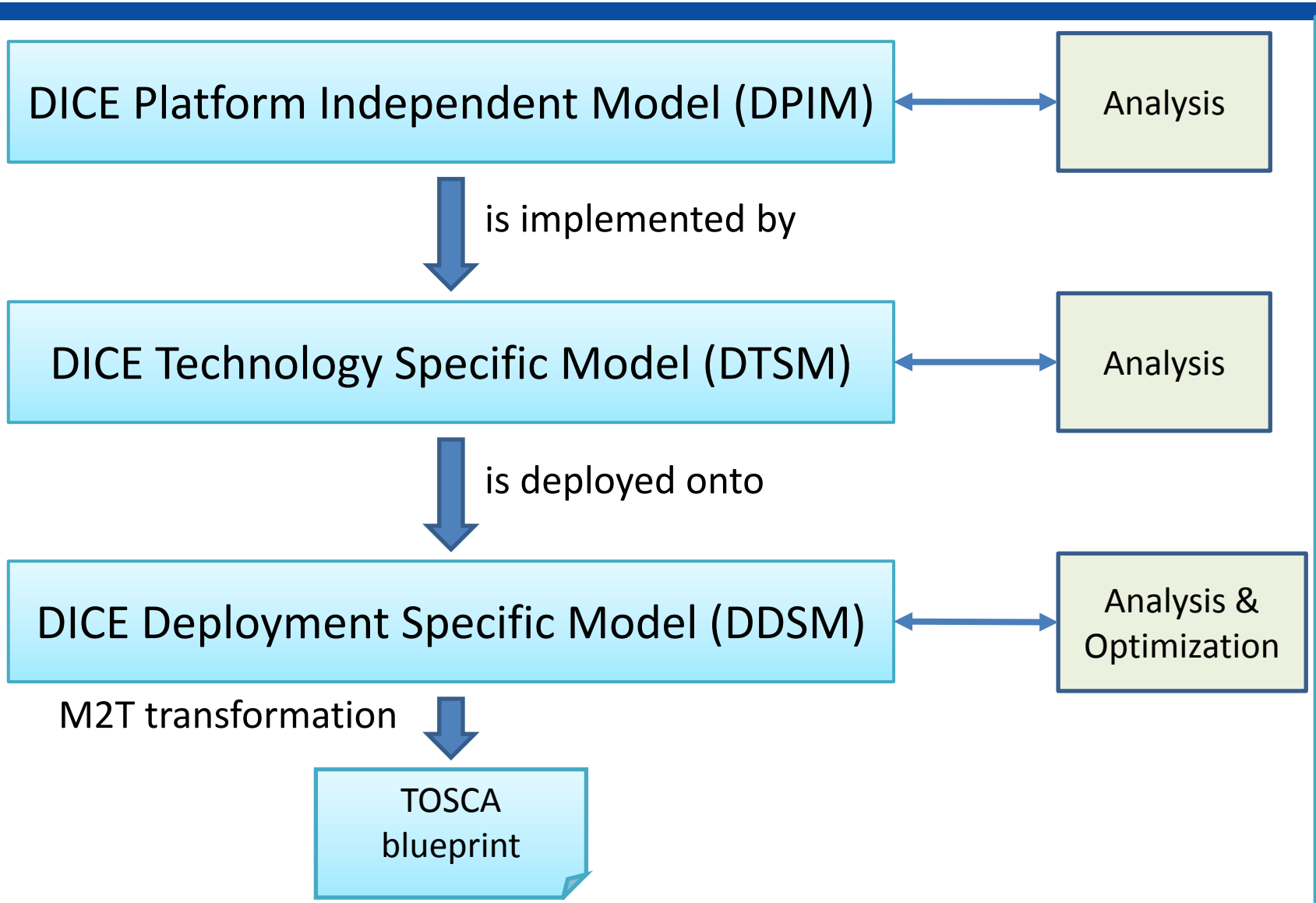
- DevOps



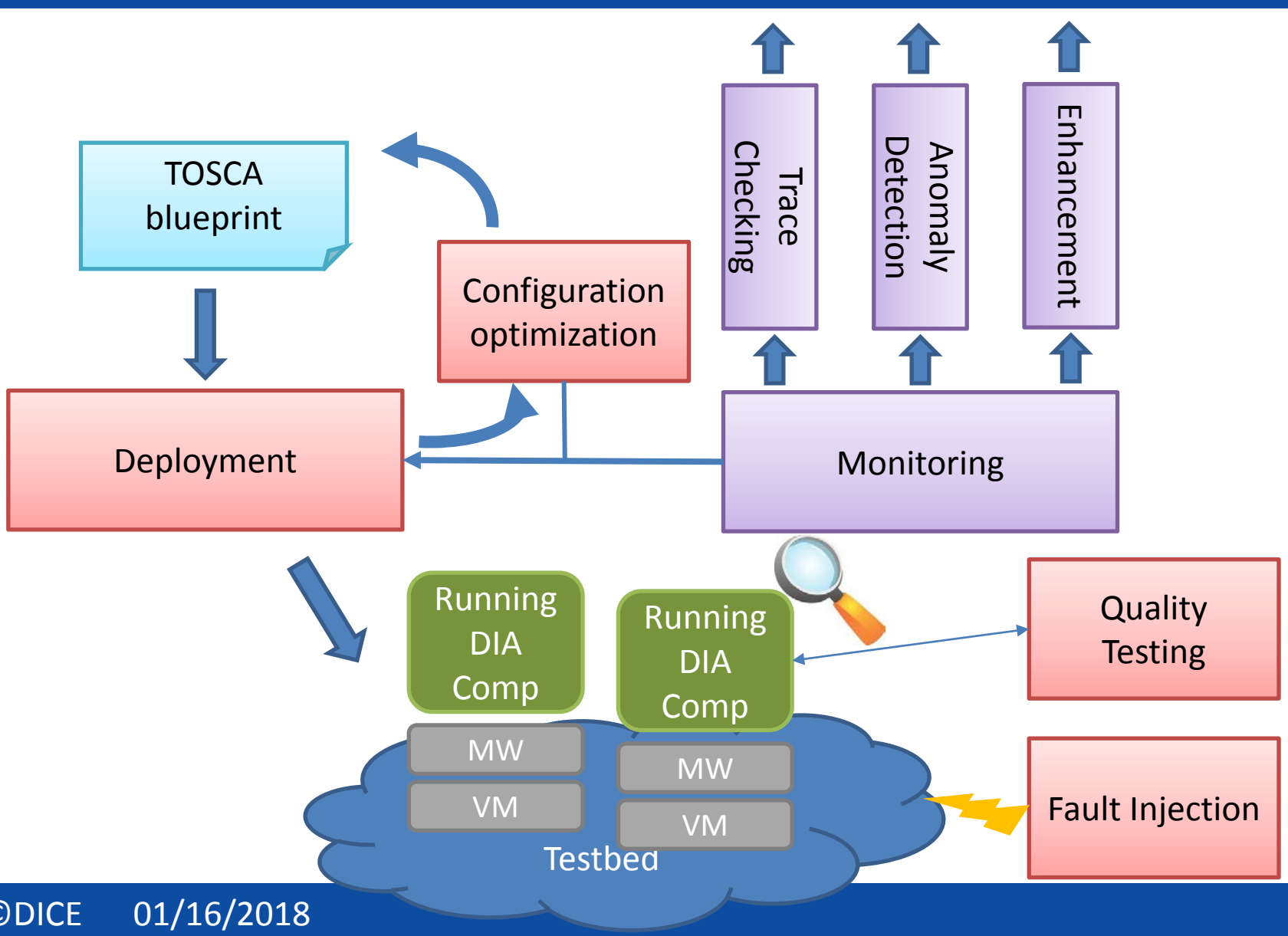
- Model-Driven Engineering



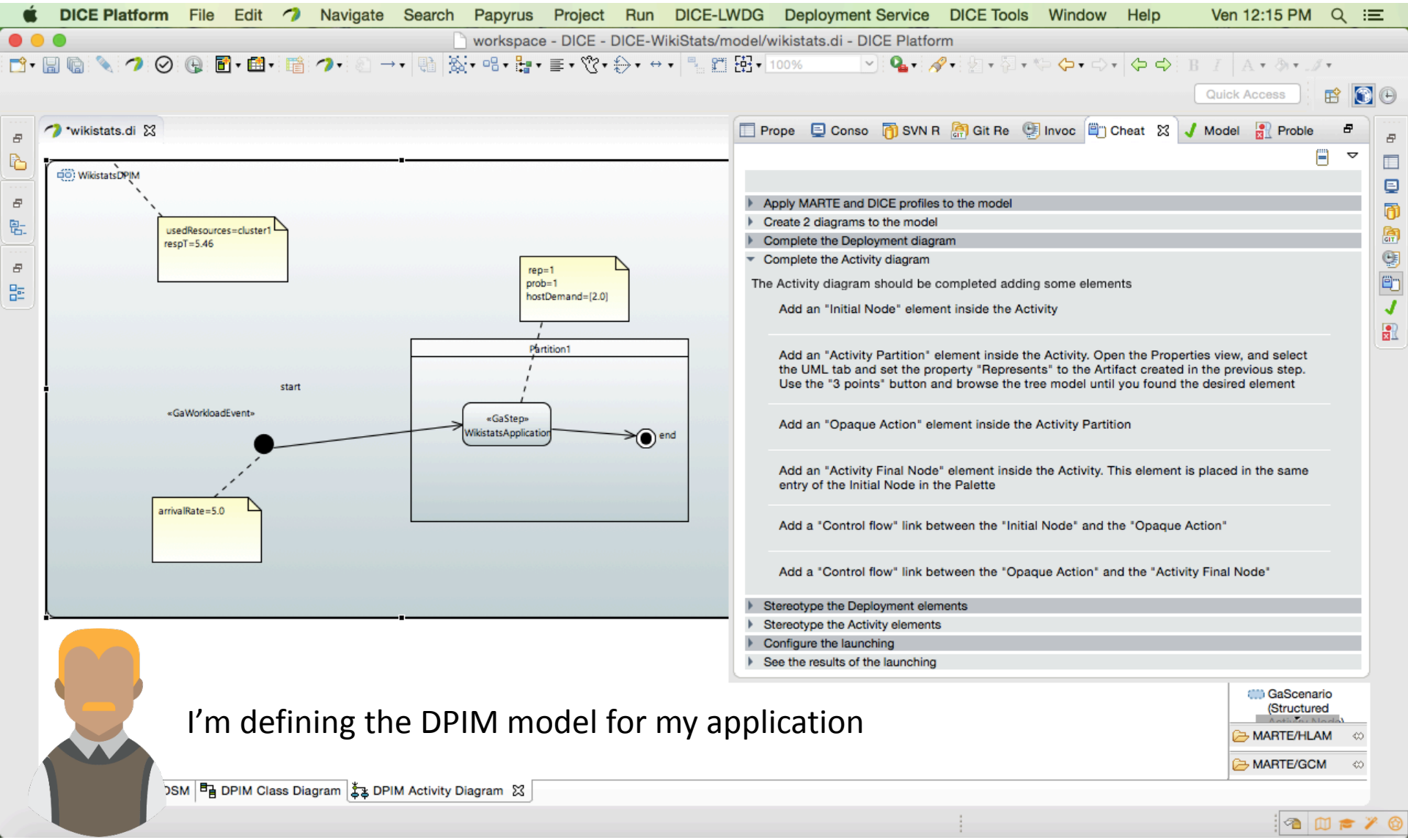
DICE incremental modeling and analysis



DICE deployment, monitoring and testing

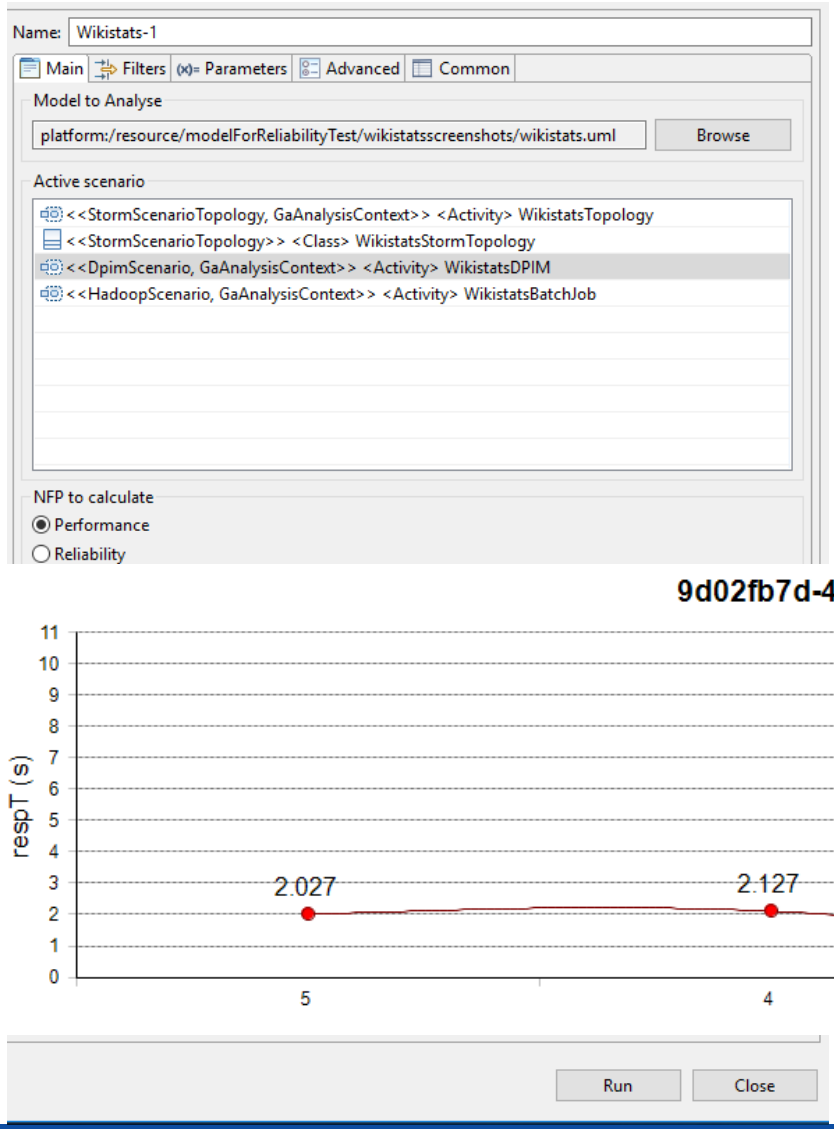


Modeling in the IDE

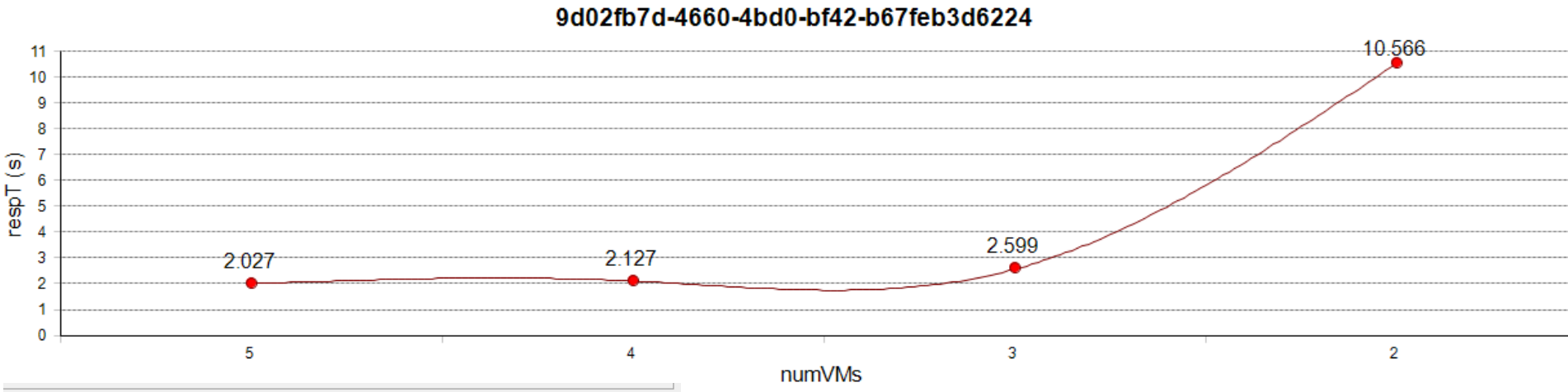


I'm defining the DPIM model for my application

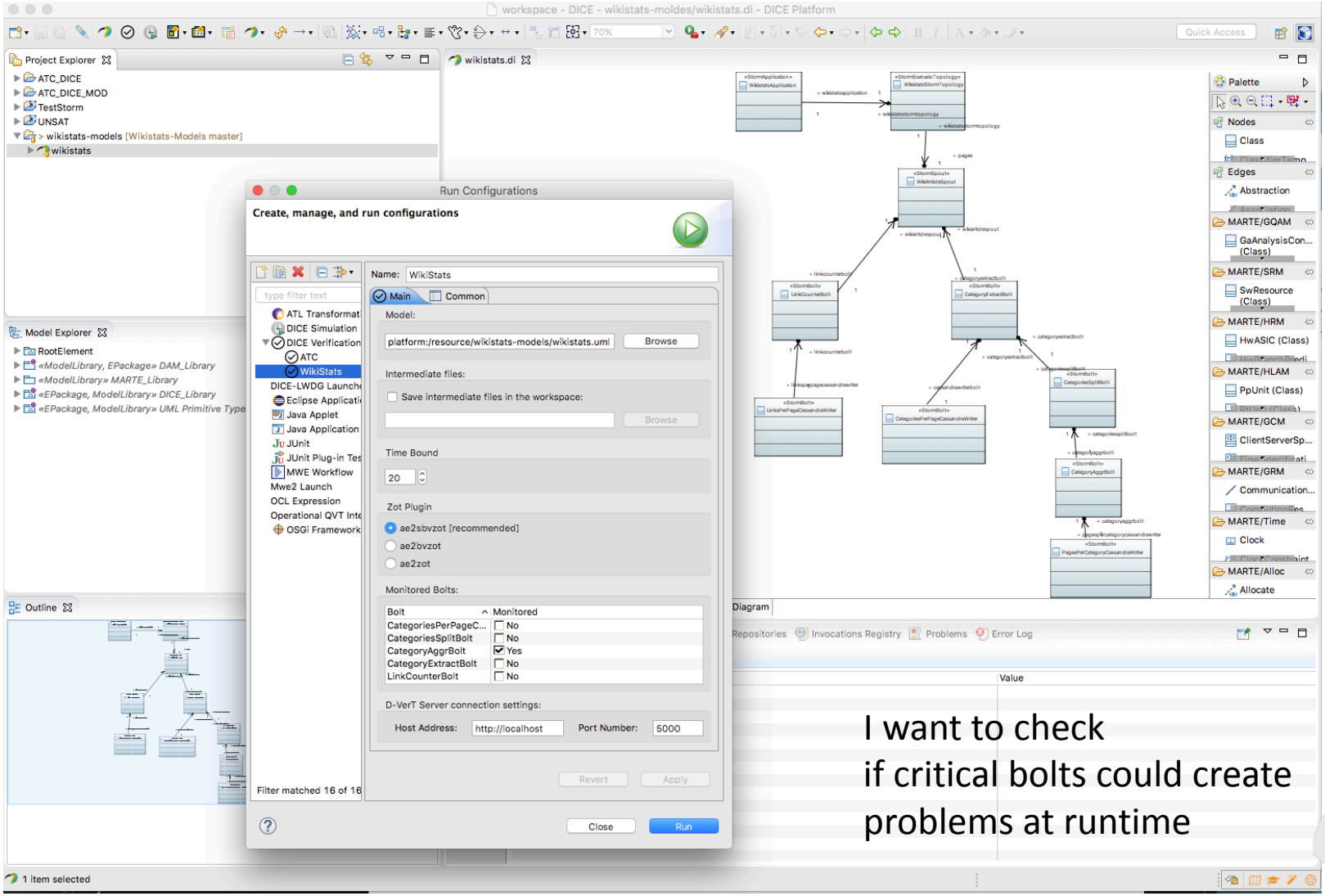
DPIM-level simulation



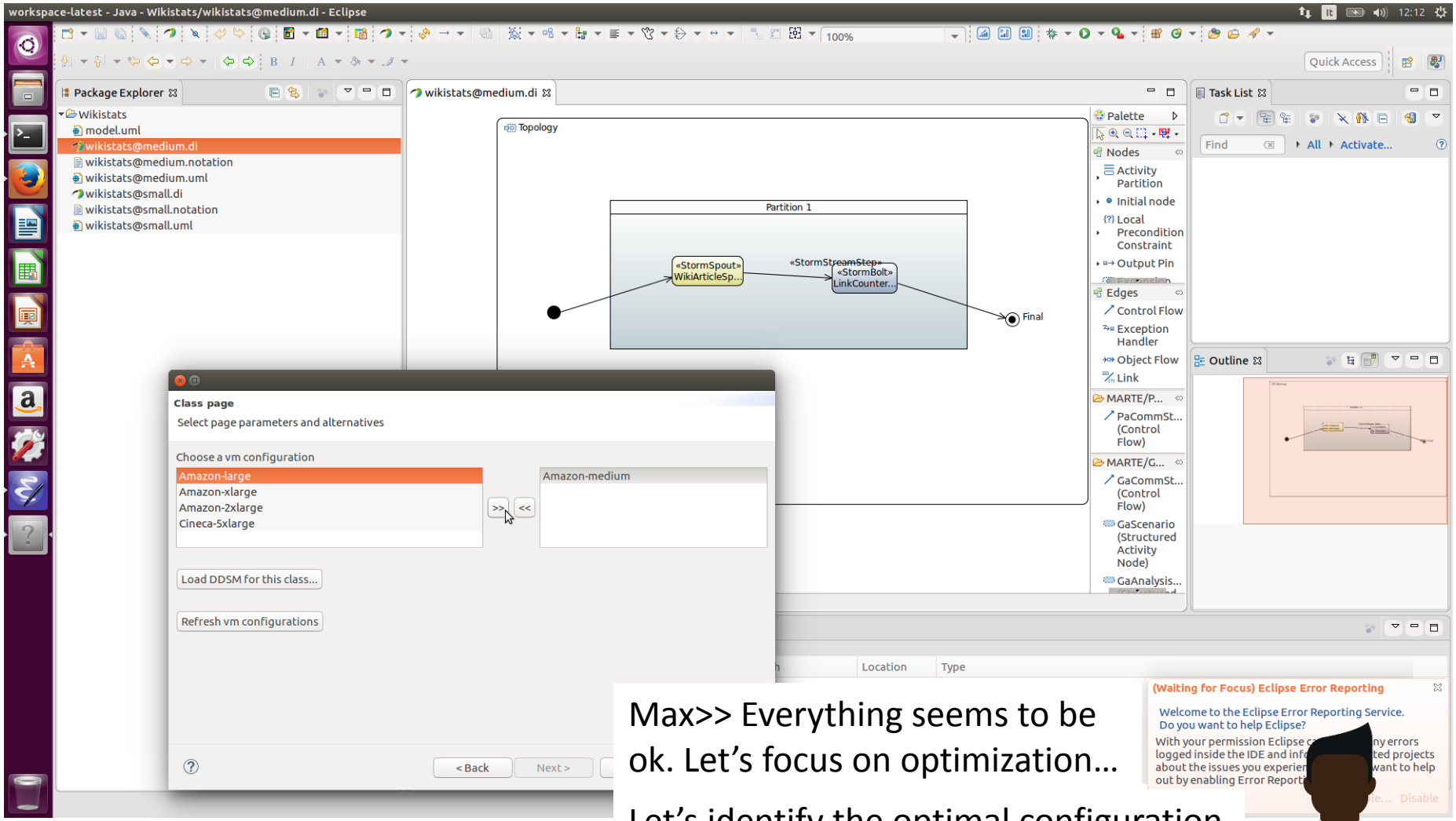
A two VMs configuration will not be acceptable for this application



The verification step



The optimization step

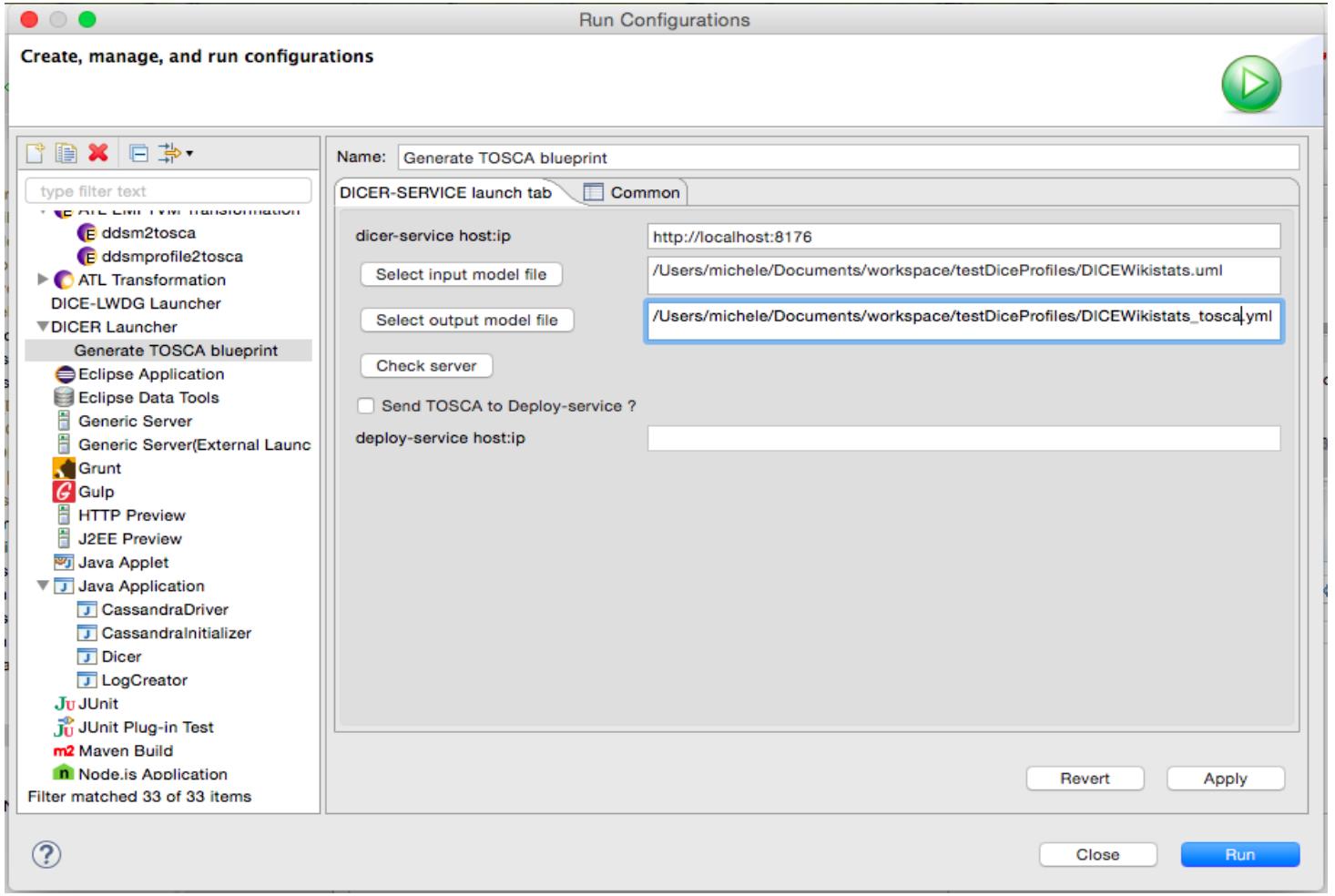


Max>> Everything seems to be ok. Let's focus on optimization…

Let's identify the optimal configuration for Amazon reserved instances

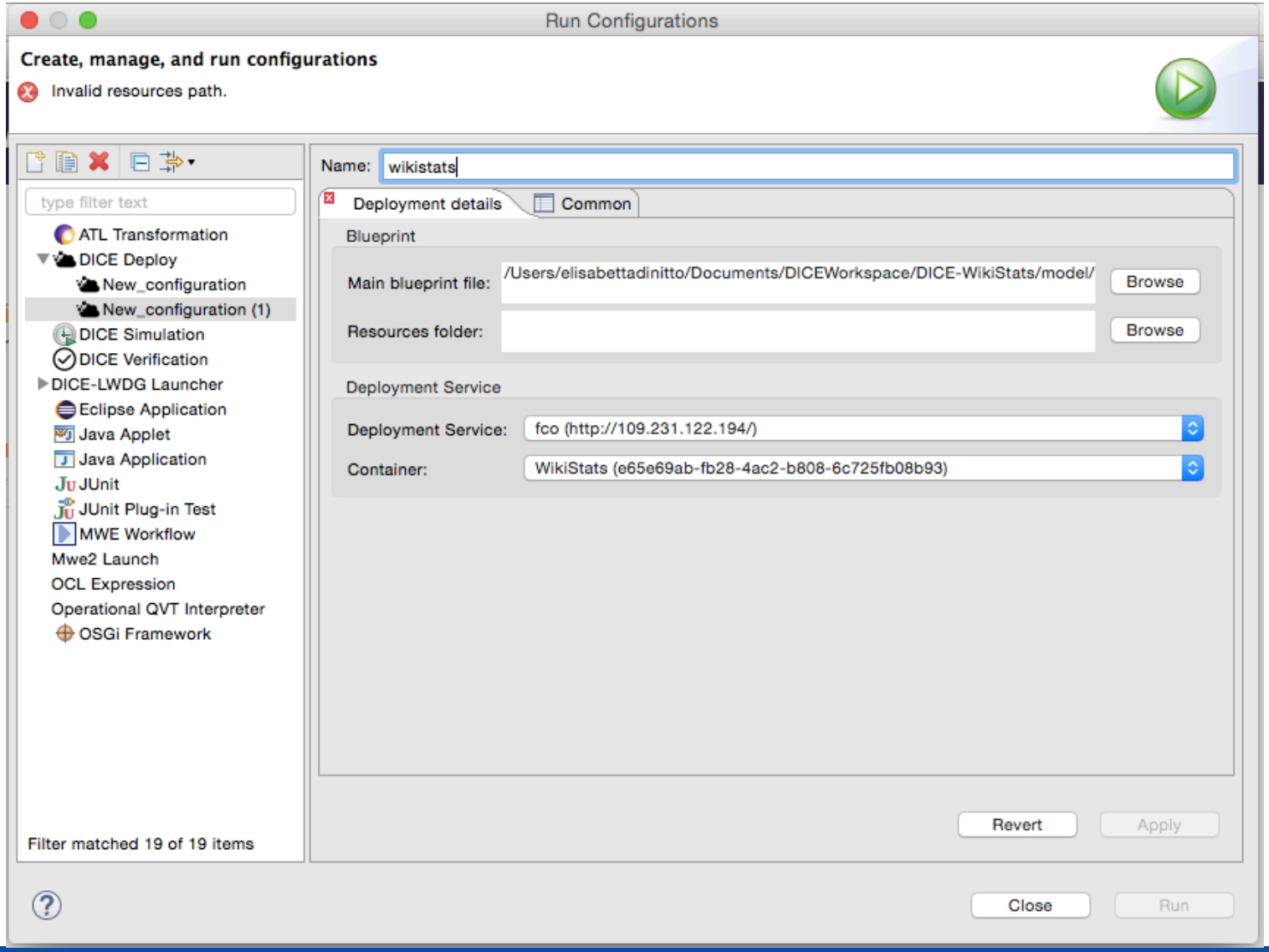
Deployment step– creating the TOSCA YAML

I have got
all I need to create
the blueprint!

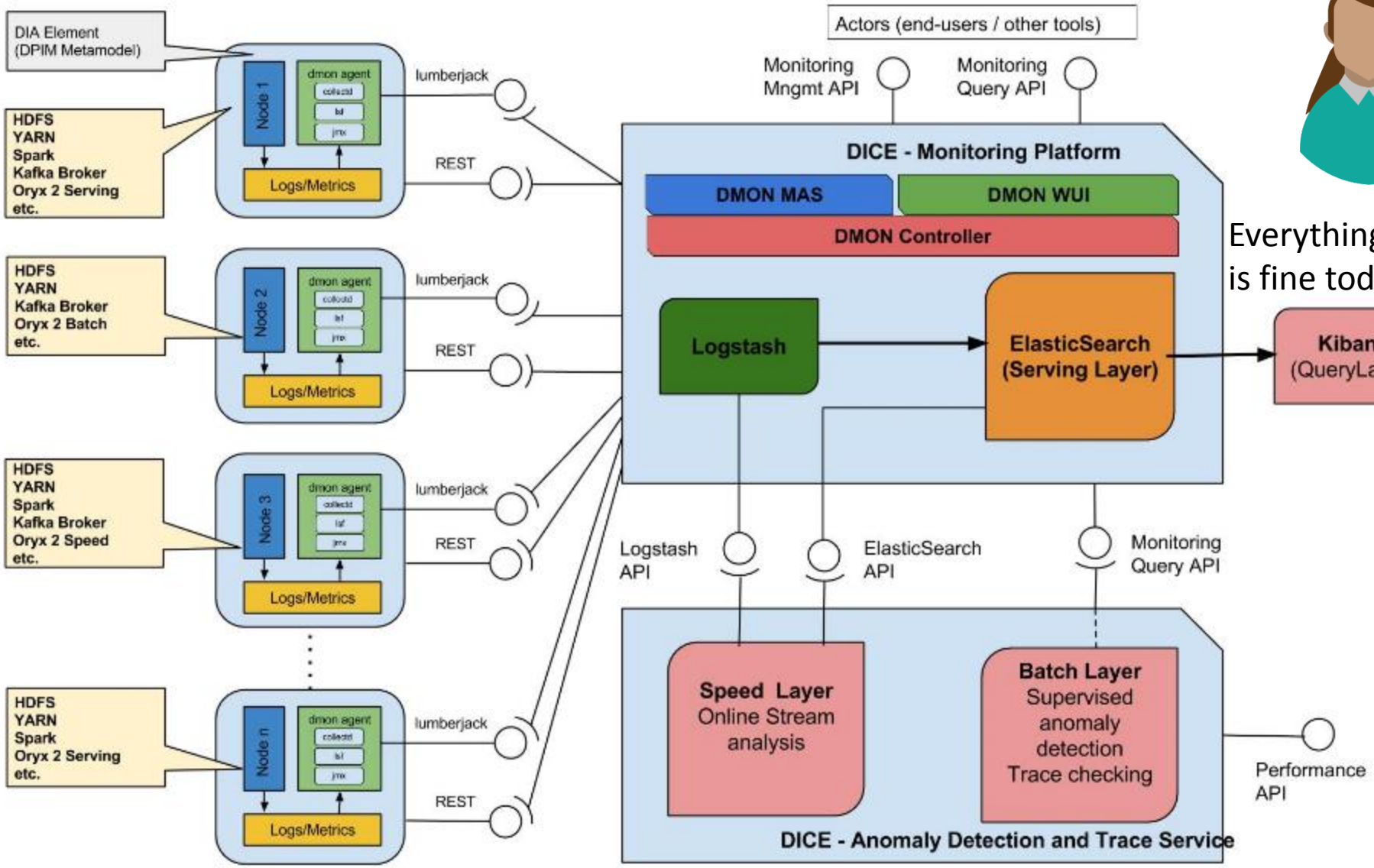


Deployment step – sending the TOSCA YAML to the deployment service

… and now let's deploy!



Monitoring and detecting anomalies



Everything is fine today!

Quality testing

Features:

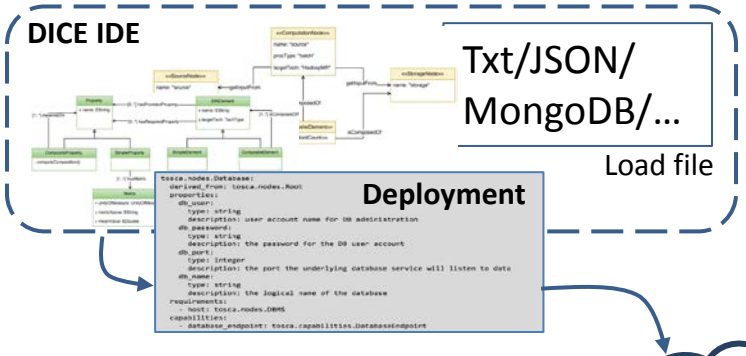
- Load generation for stream processing systems
- Load replay
- Load scaling (via Hidden Markov Model representation)

Load injection into the DIA

QT Tool workflow:

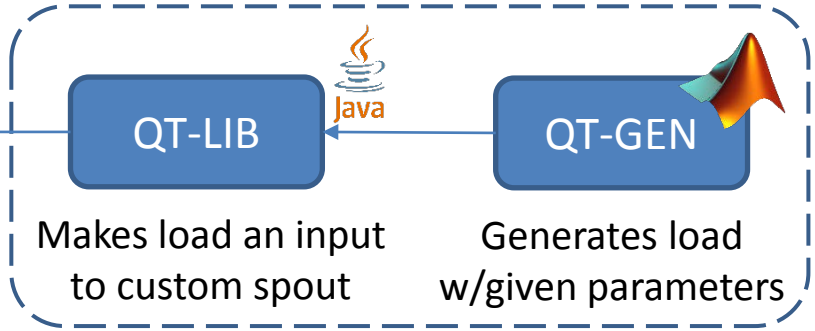
1. Max & Ophra add QT tool as a spout, provide (initial) load & deploys DIA

Let's stress our DIA!

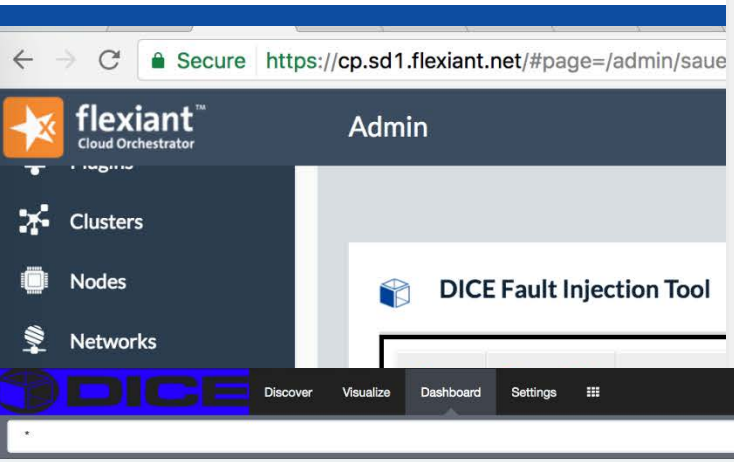


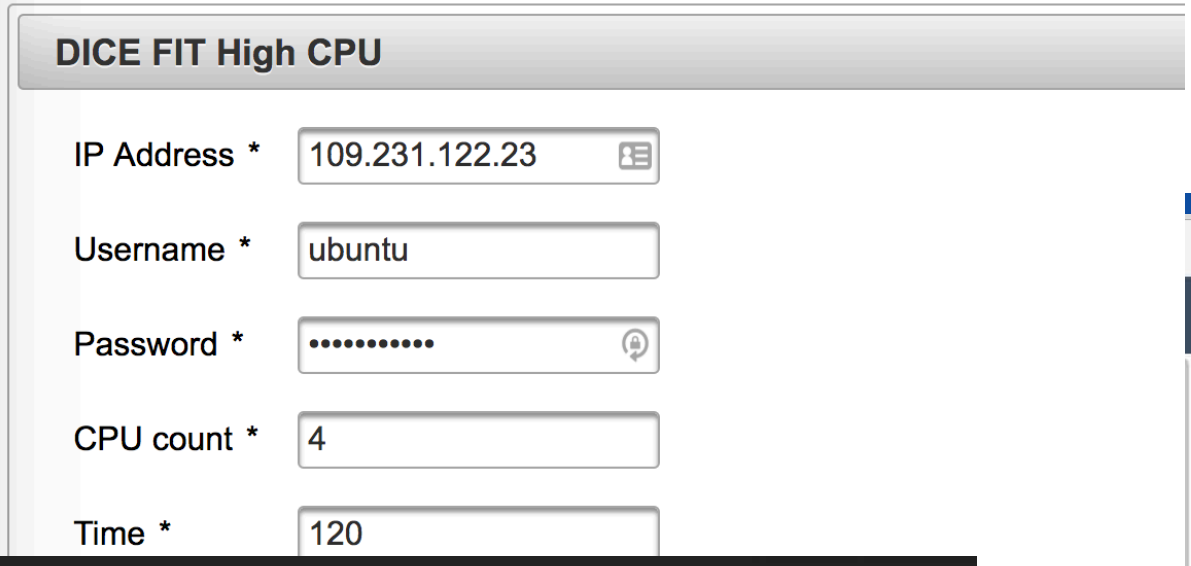
2. QT tool starts simultaneously with DIA

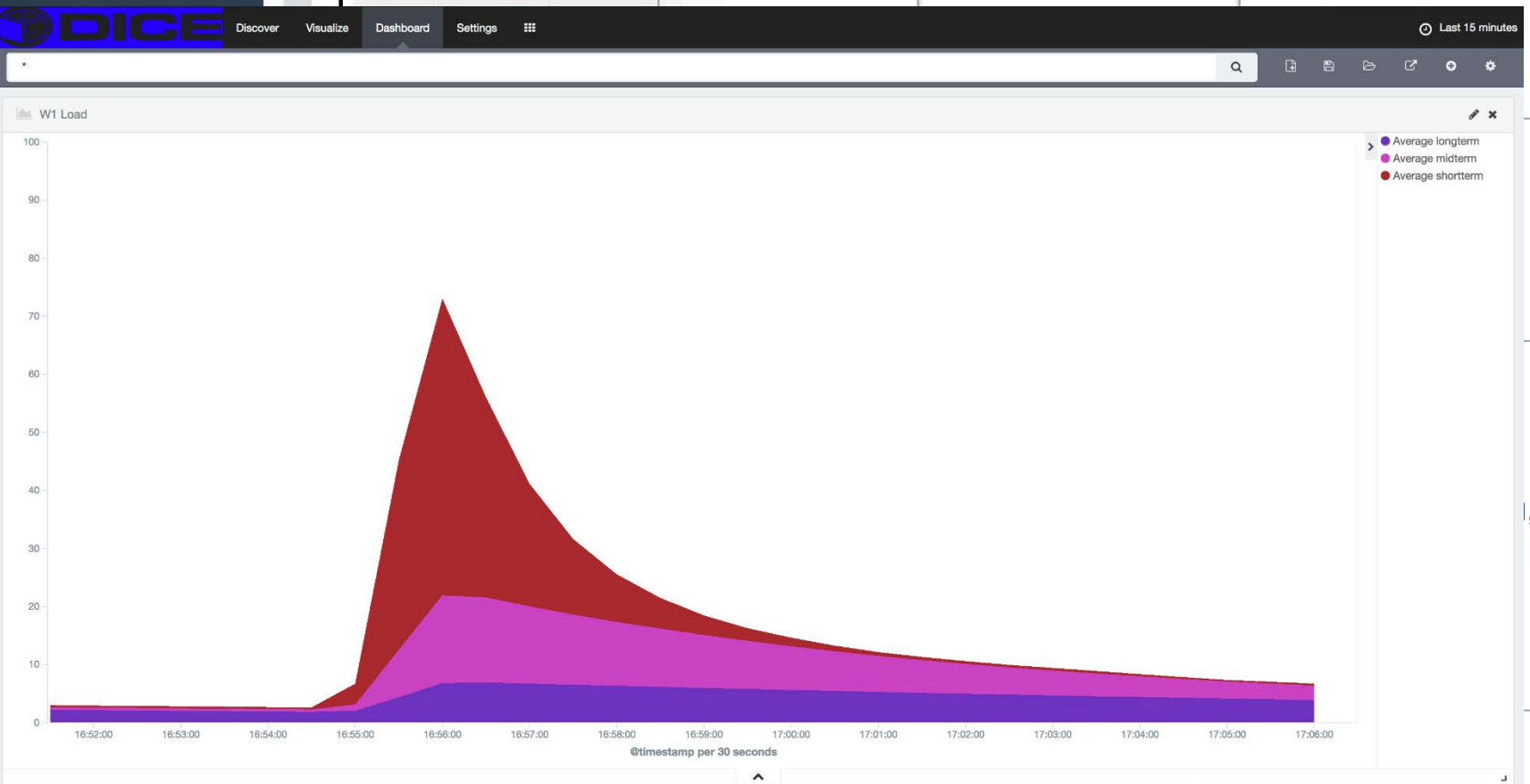
3. Injects load



Fault injection



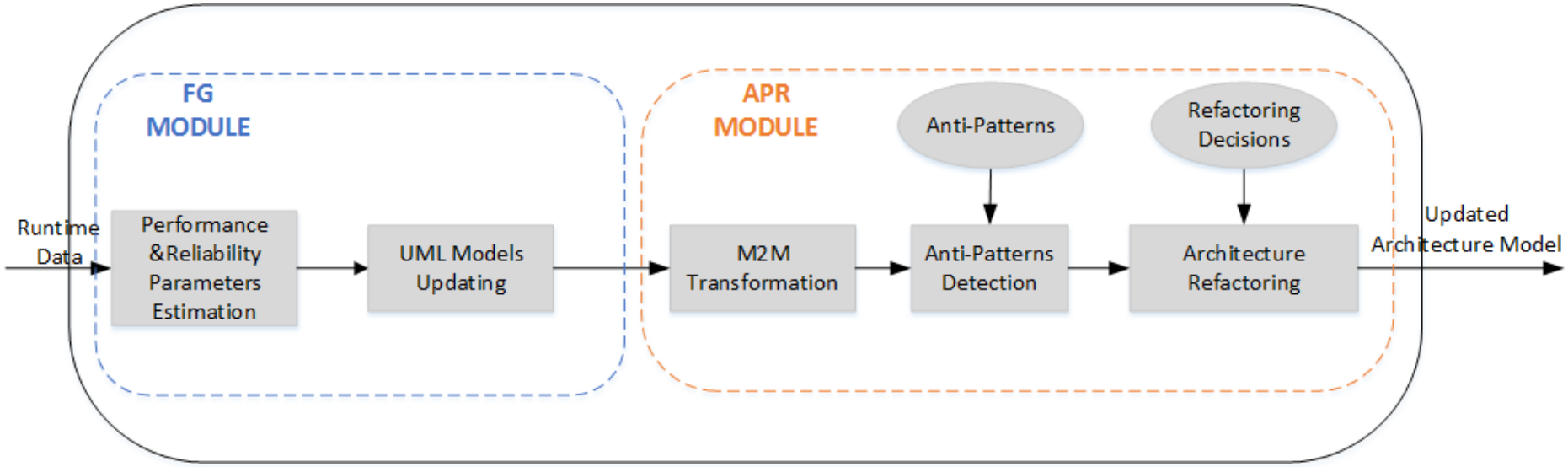




Enhancement tool

- Objective: designed for iteratively enhancing the DIA quality
- Functions:
  - Providing a performance and reliability analysis
  - Updating UML models with analysis results
  - Anti-patterns Detection
  - Refactoring the design model

let's consider the proposed improvements!



Summing up

Following the DevOps paradigm, DICE delivers a unified toolchain for the enterprise team

DICE Eclipse IDE

Enhance

Enhance

Deploy

Quality Testing

Fault injection

Configuration Optimization

Design

Transform to Formal Models

Simulate & Verify

Optimize

Demonstrators

News Asset

News&Media Market

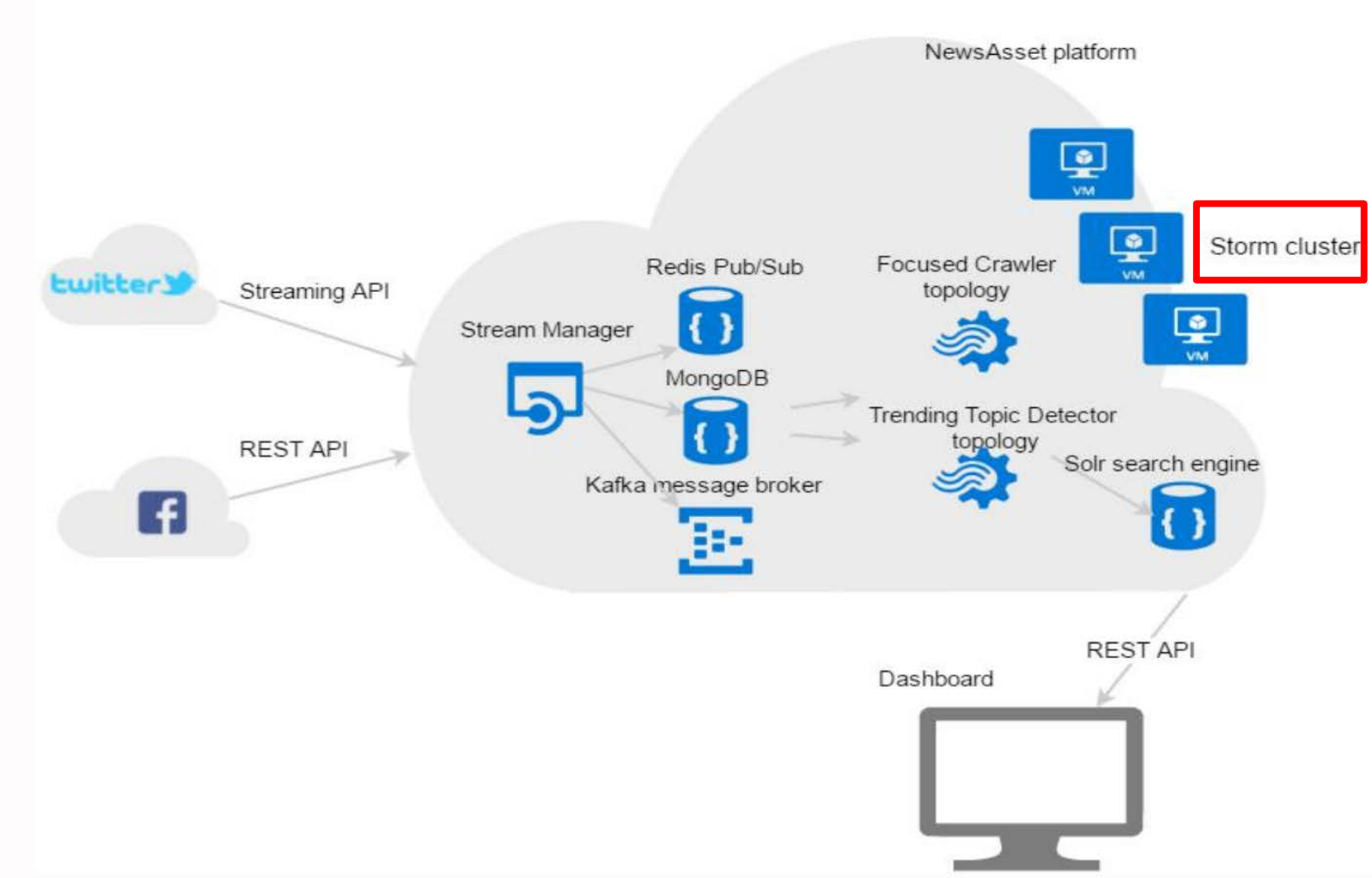
Tax Fraud Detection Application

e-Government Market

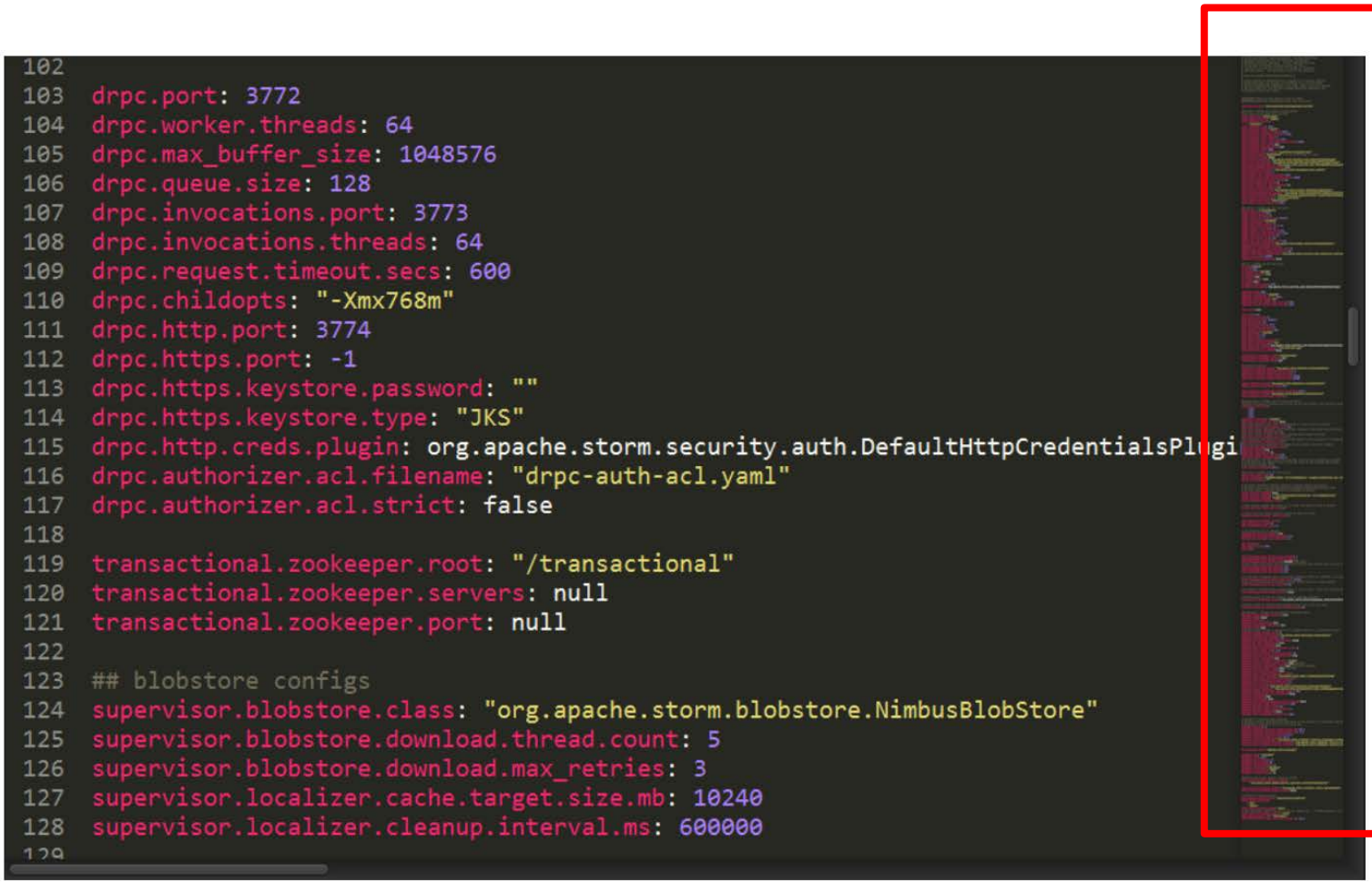
Posidonia Operations

Maritime Sector

News analysis demonstrator

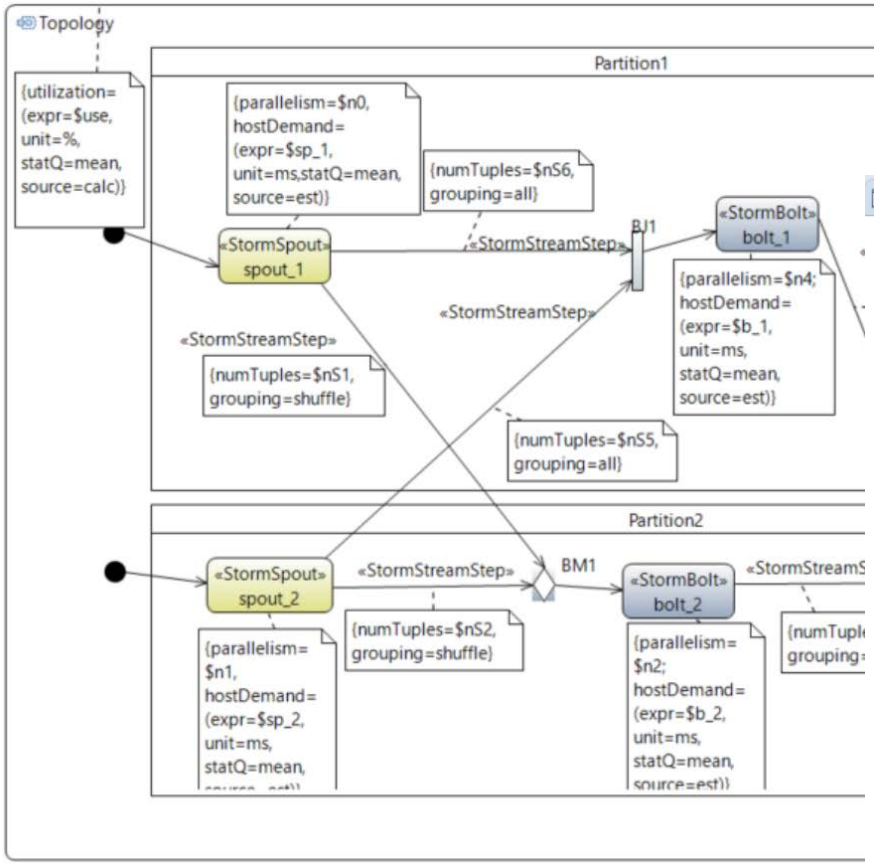


Configuring a Big data system



!

DICE configuration optimization



Properties | Console | SVN Repositories | Git Repositories | Invocations Registry | DICE Configuration View

Plugin Config | Parameter Selection | Service Config | Experiment Config | App Config | Experiments

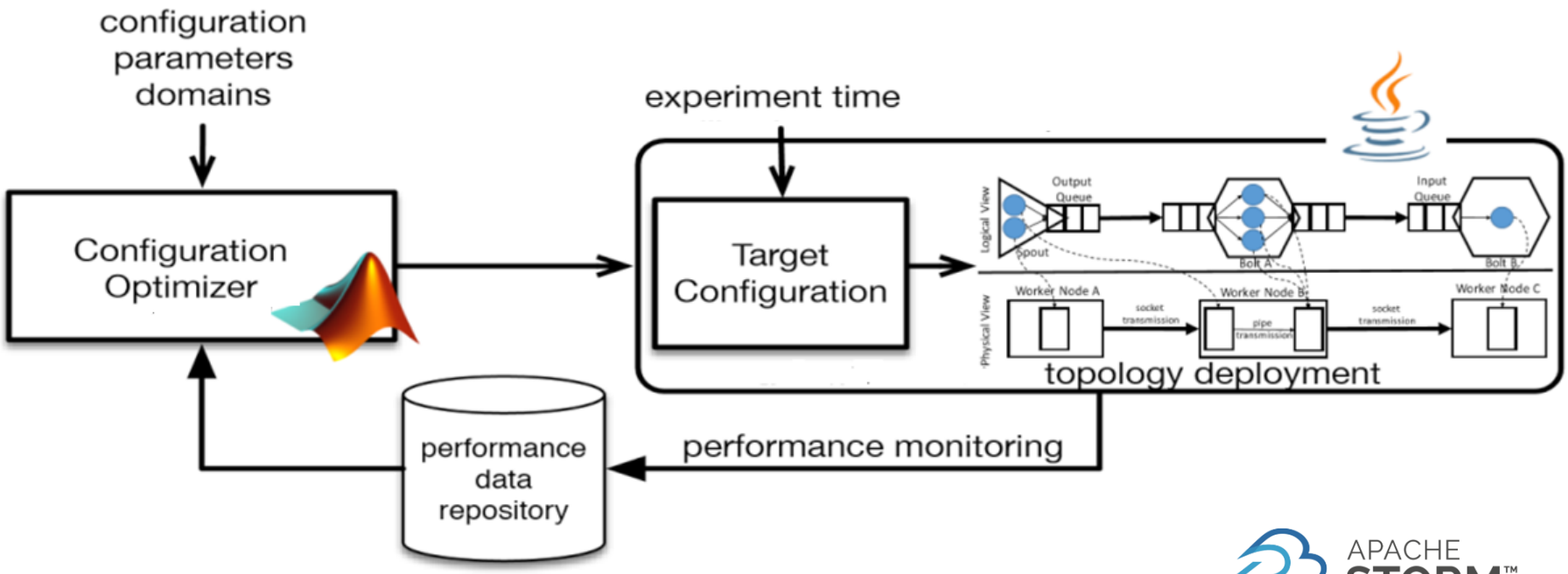
storm

Parameter	Description
topology.error.throttle.interval.secs	
topology.trident.batch.emit.interval.millis	
topology.disruptor.wait.timeout.millis	
topology.disruptor.batch.size	
topology.disruptor.batch.timeout.millis	
topology.disable.loadaware.messaging	
topology.state.checkpoint.interval.ms	
topology.max.spout.pending	
topology.acker.executors	
topology.tick.tuple.freq.secs	

Add Parameters

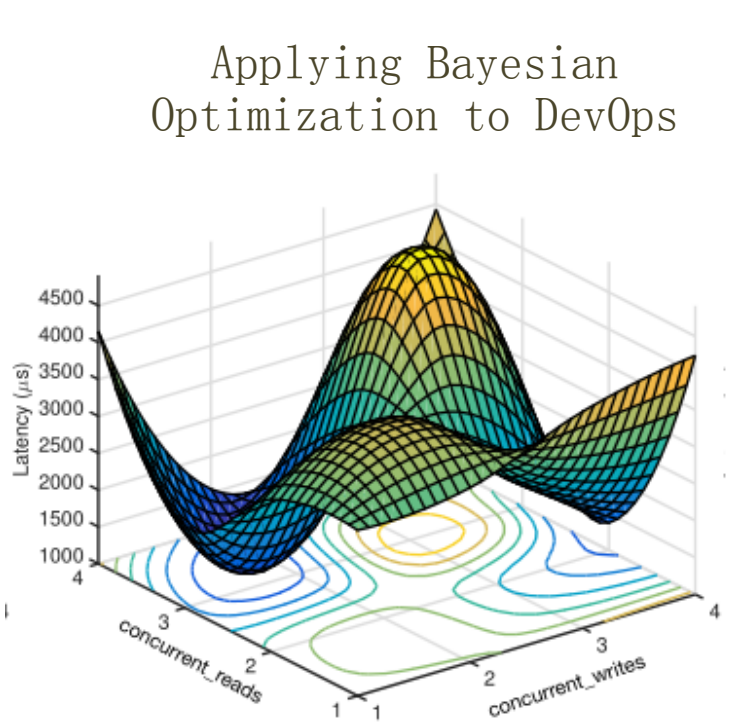
Parameter	Type	Min	Max	Step	Options
topology.executor.receive.buffer.size	Integer	1024	2048	1	
topology.min.replication.count	Integer	1	10	1	
topology.worker.shared.thread.pool.size	Integer	1	20	1	
topology.max.task.parallelism	Integer	1	50	1	

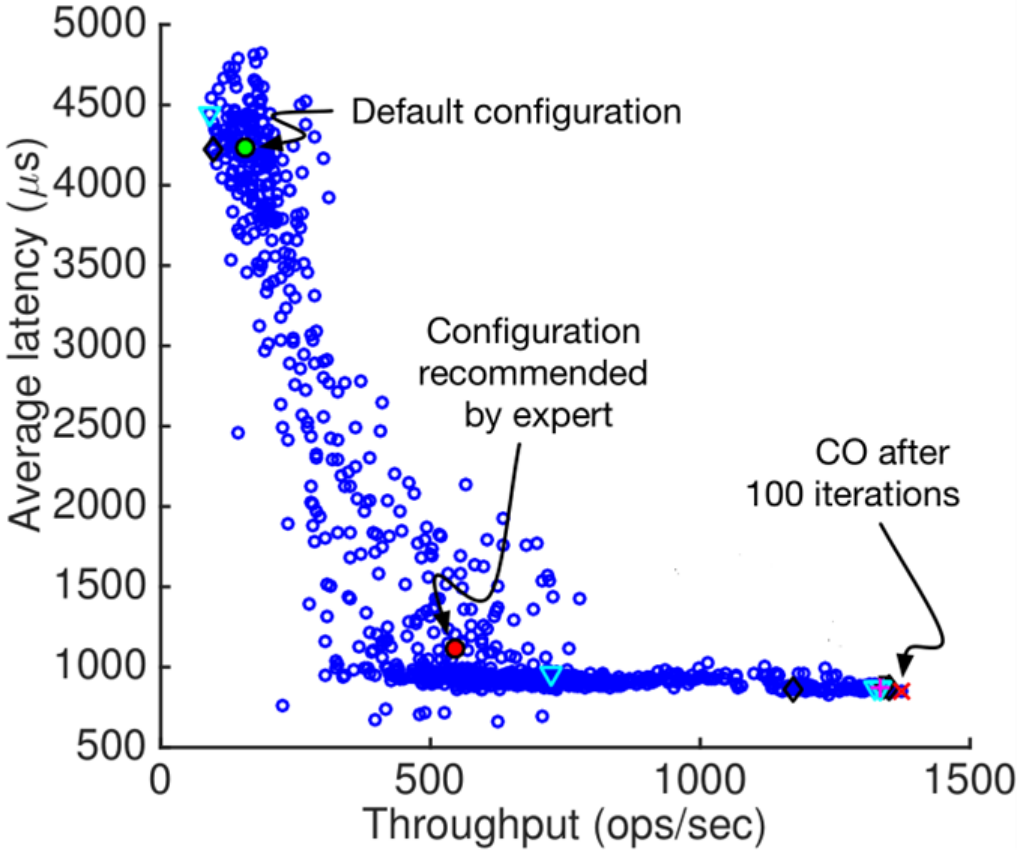
DICE configuration optimization



APACHE STORM™
Distributed • Resilient • Real-time

DICE configuration optimization





Thanks!

DICE

WWW.DICE-H2020.EU

DEVOPS
FOR
BIG DATA