



# Towards Quality-Aware Development of Big Data Applications with DICE

*DICE: Developing Data-Intensive Cloud  
Applications with Iterative Quality  
Enhancements*

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



Funded by the Horizon 2020  
Framework Programme of the European Union



# DICE Project

- Horizon 2020 Research & Innovation Action (RIA)
  - Quality-Aware Development for Big Data applications
  - Feb 2015 - Jan 2018, 4M Euros budget
  - 9 partners (Academia & SMEs), 7 EU countries

Imperial College  
London



Universidad  
Zaragoza



POLITECNICO  
DI MILANO



# Motivation



- Software market rapidly shifting to Big Data
  - 32% compound annual growth rate in EU through 2016
  - 35% Big data projects are successful [CapGemini 2015]
- ICT-9 call focused on SW quality assurance (QA)
  - ISTAG: call to define environments *“for understanding the consequences of different implementation alternatives (e.g. quality, robustness, performance, maintenance, evolvability, ...)”*
- QA evolving too slowly compared to the technology trends (Big data, Cloud, DevOps ...)
  - DICE aims at closing the gap
  - Still crucial for competitiveness!

# Quality Dimensions



- Reliability
  - Availability
  - Fault-tolerance
- Efficiency
  - Performance
  - Costs
- Safety & Privacy
  - Verification (e.g., deadlines)
  - Data protection

# High-Level Objectives



- Tackling skill shortage and steep learning curves
  - Data-aware methods, models, and tools
- Shorter time to market for Big Data applications
  - Cost reduction, without sacrificing product quality
- Decrease development and testing costs
  - Select optimal architectures that can meet SLAs
- Reduce number and severity of quality incidents
  - Iterative refinement of application design

# Some Challenges in Big Data...



- Lack of quality-aware development for Big Data
  - How to described in MDE Big Data technologies
    - Spark, Hadoop/MapReduce, Storm, Cassandra, ...
    - Cloud storage, auto-scaling, private/public/hybrid, ...
- Today no QA toolchain can help reasoning on data-intensive applications
  - What if I double memory?
  - What if I parallelize more the application?

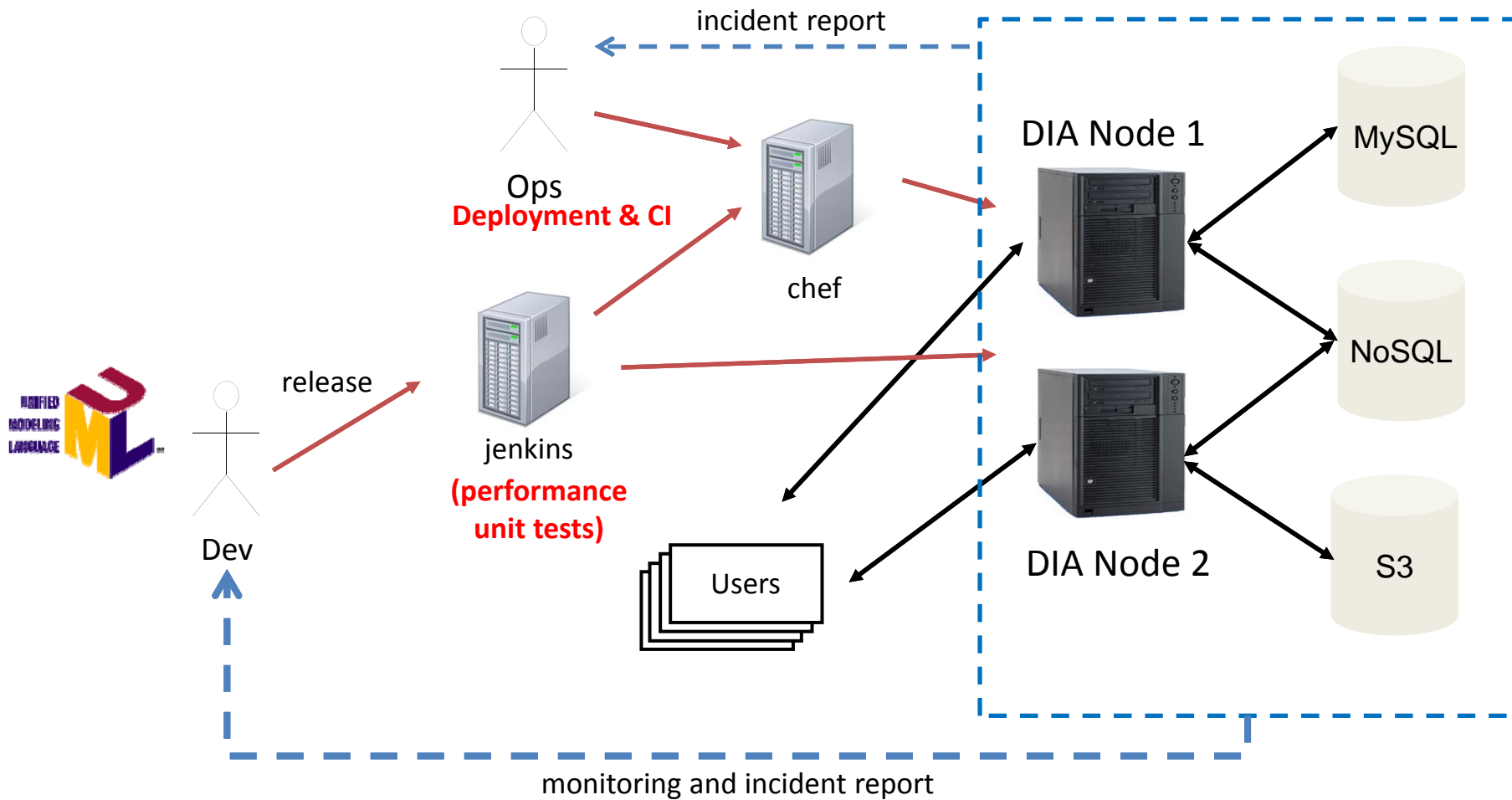
# ... in a DevOps fashion



- Software development methods are evolving
- DevOps closes the gap between Dev and Ops
  - From agile development to agile delivery
  - Lean release cycles with automated tests and tools
  - Deep modelling of systems is the key to automation

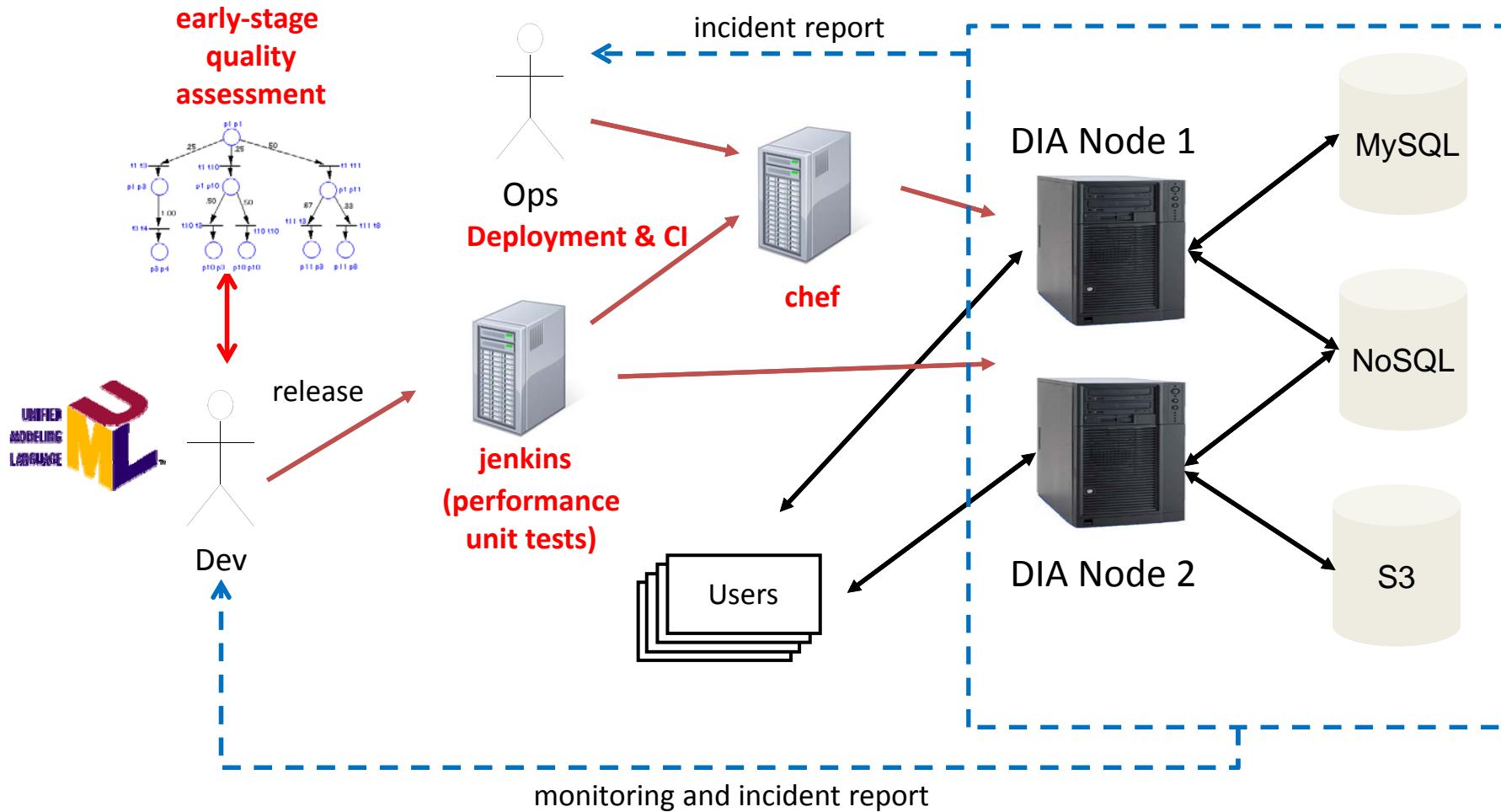


# DevOps in DICE: Measurement



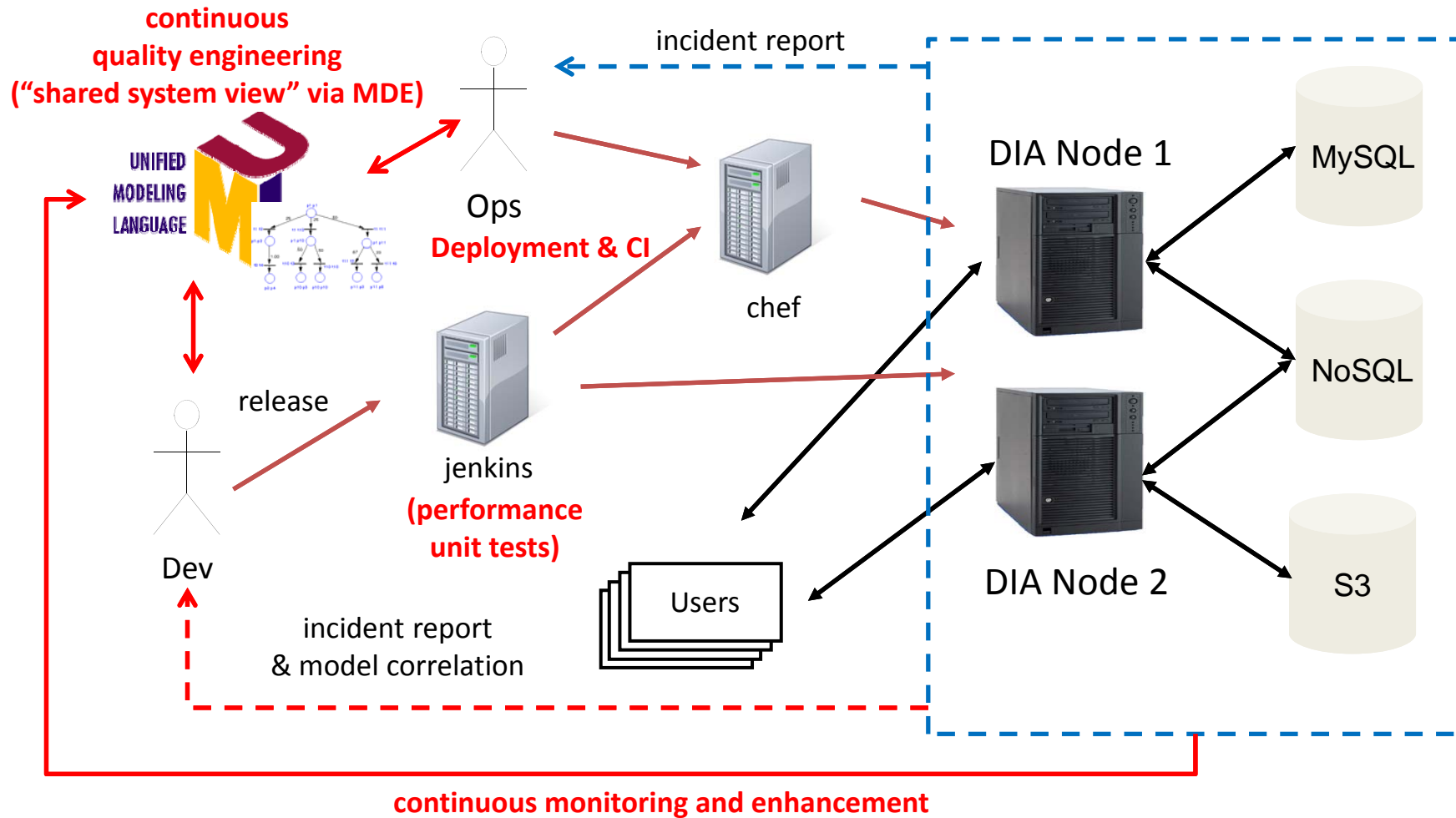


# DevOps in DICE: Early-stage MDE

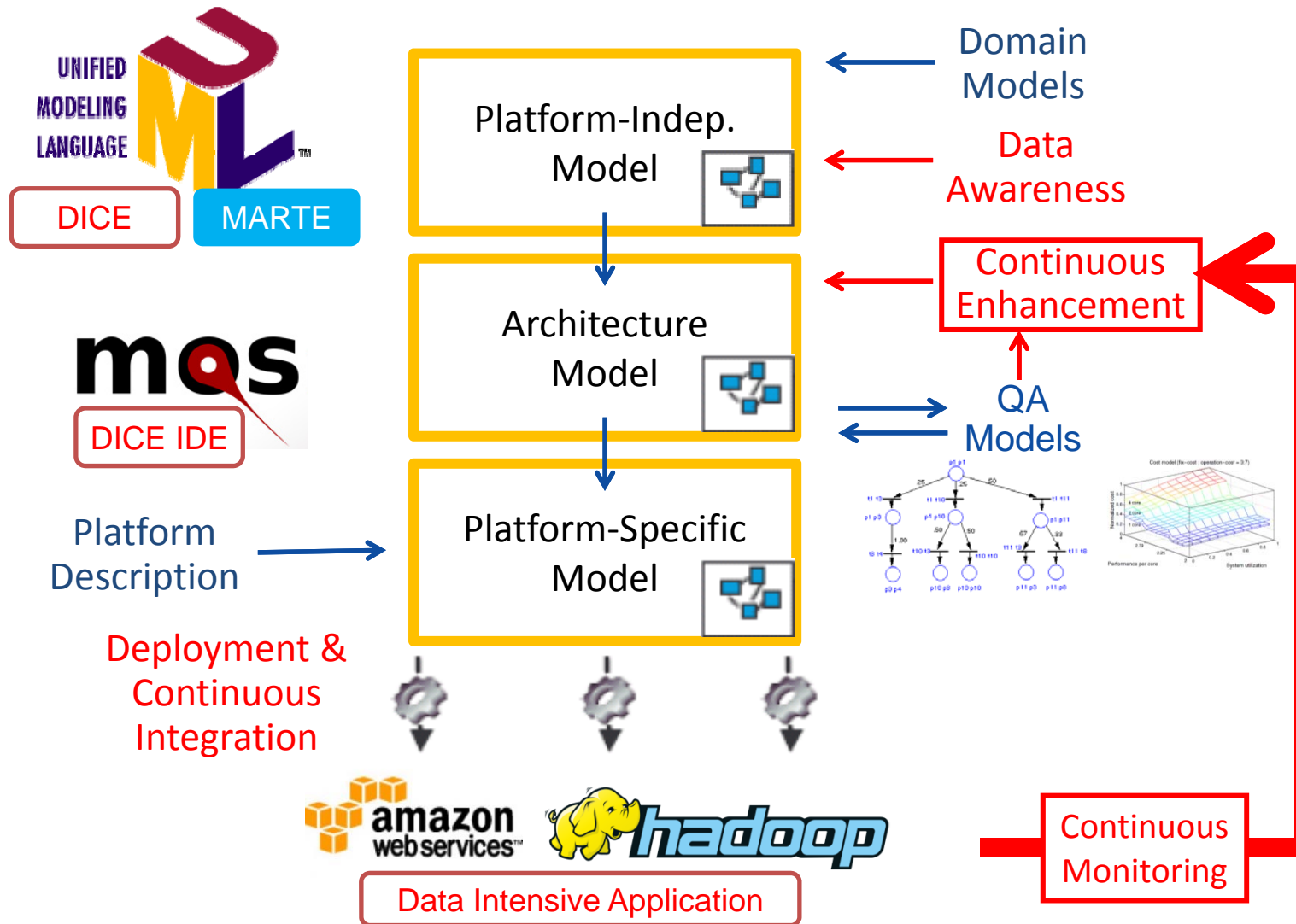




# DevOps in DICE: Enhancement



# DICE Integrated Solution



# Demonstrators



Case study	Domain	Features & Challenges
Distributed data-intensive media system (ATC)	<ul style="list-style-type: none"><li>• <b>News &amp; Media</b></li><li>• <b>Social media</b></li></ul>	<ul style="list-style-type: none"><li>• Large-scale software</li><li>• Data velocities</li><li>• Data volumes</li><li>• Data granularity</li><li>• Multiple data sources and channels</li><li>• Privacy</li></ul>
Big Data for e-Government (Netfective)	<ul style="list-style-type: none"><li>• <b>E-Gov application</b></li></ul>	<ul style="list-style-type: none"><li>• Data volumes</li><li>• Legacy data</li><li>• Data consolidation</li><li>• Data stores</li><li>• Privacy</li><li>• Forecasting and data analysis</li></ul>
Geo-fencing (Prodevelop)	<ul style="list-style-type: none"><li>• <b>Maritime sector</b></li></ul>	<ul style="list-style-type: none"><li>• Vessels movements</li><li>• Safety requirements</li><li>• Streaming &amp; CEP</li><li>• Geographical information</li></ul>

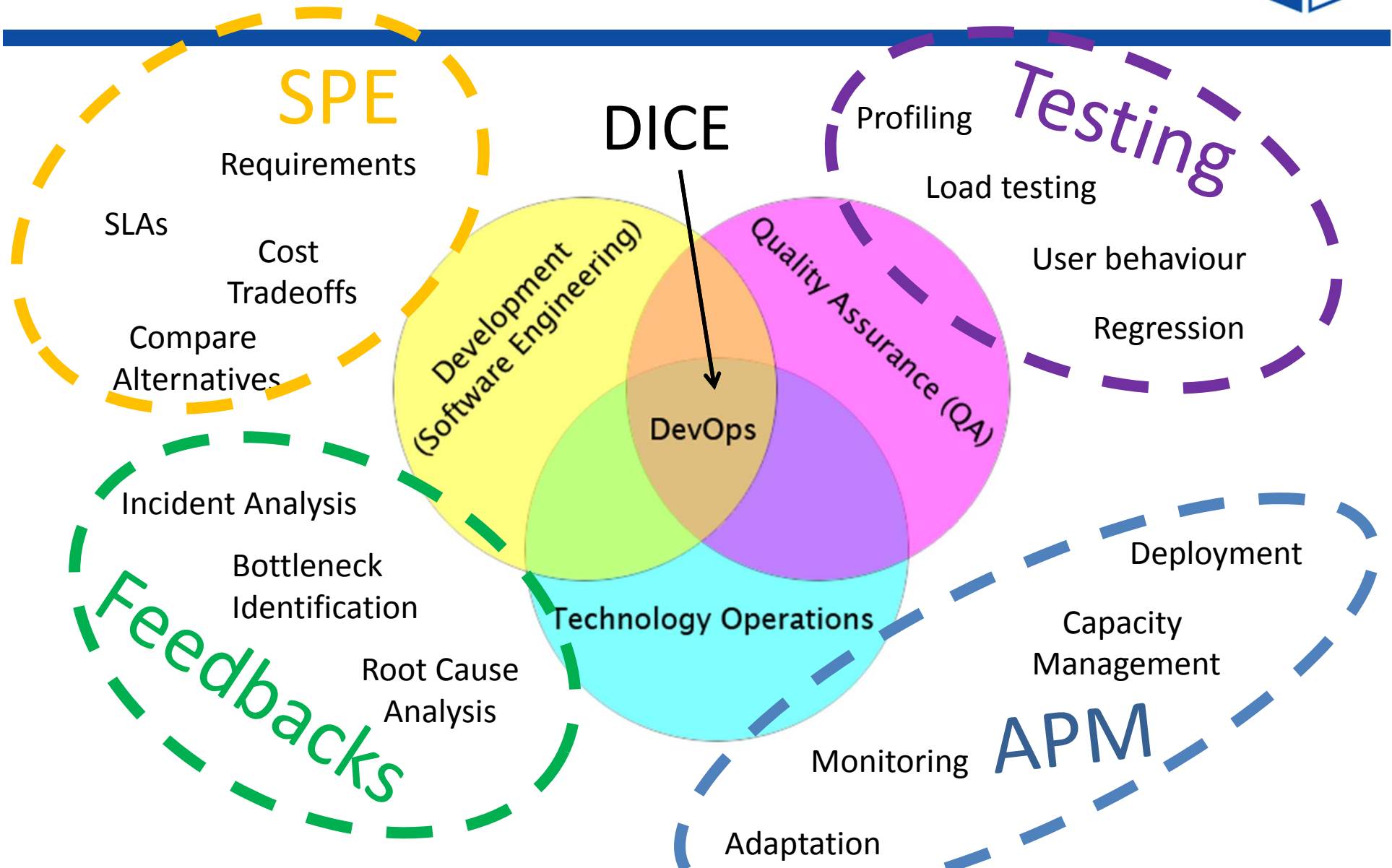


---

# Thanks

[www.dice-h2020.eu](http://www.dice-h2020.eu)

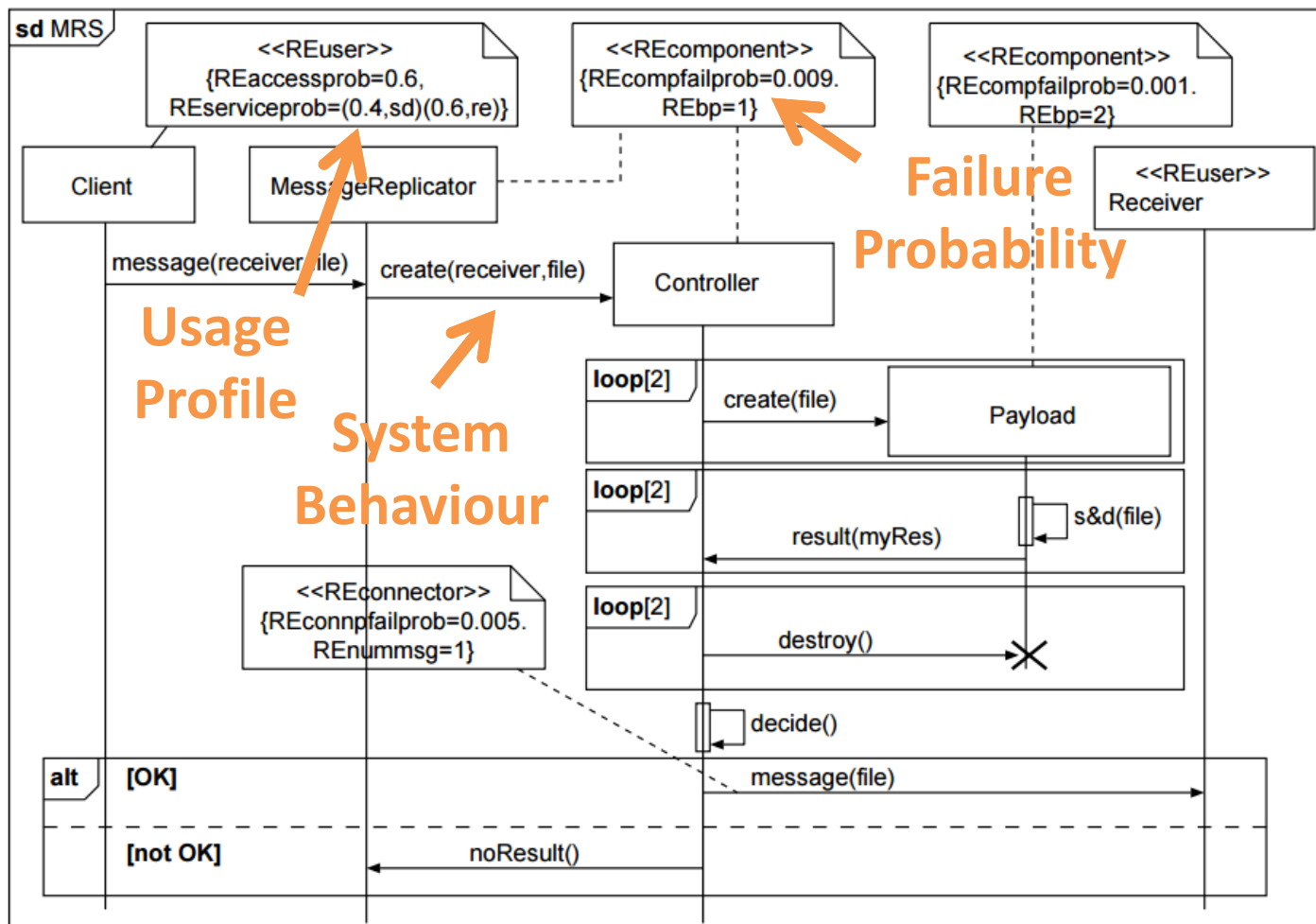
# Bringing QA and DevOps together



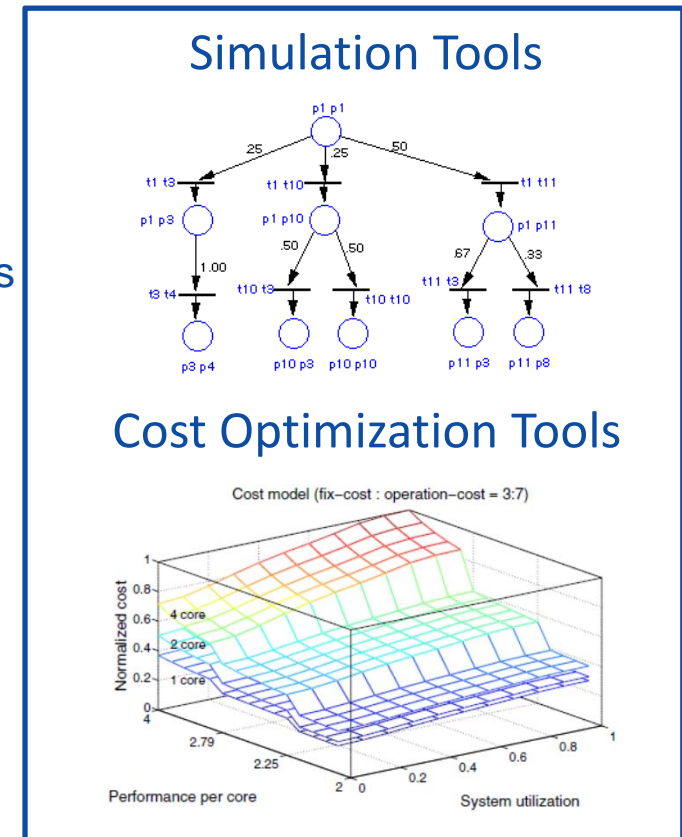
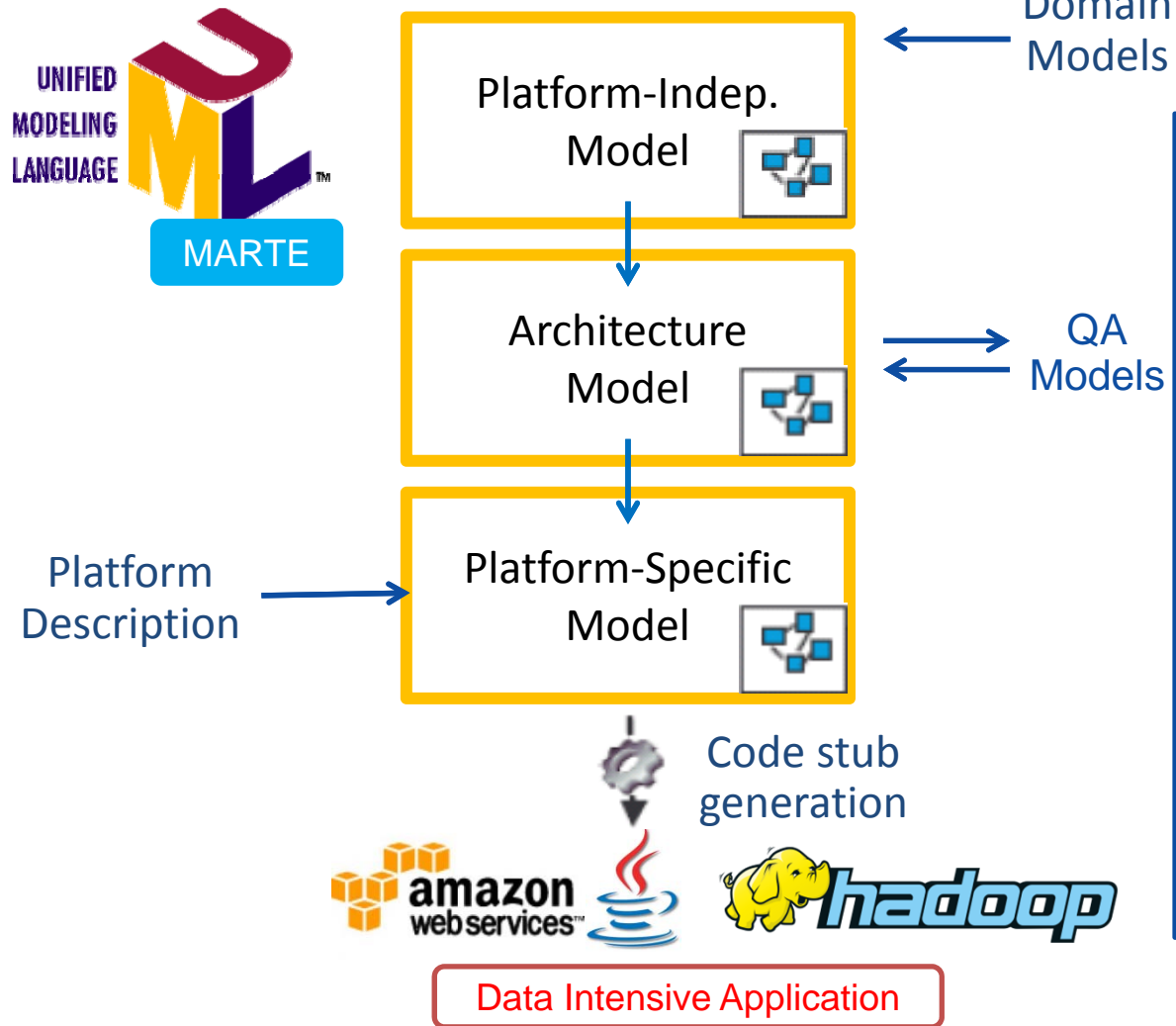


# Quality-Aware MDE

- UML MARTE profile, UML DAM profile, Palladio, ...



# Quality-Aware MDE





# Year 1 Milestones



<i>Milestone</i>	<i>Deliverables</i>
Baseline and Requirements - July 2015 <b>[COMPLETED]</b>	<ul style="list-style-type: none"><li>• State of the art analysis</li><li>• Requirement specification</li><li>• Dissemination, communication, collaboration and standardisation report</li><li>• Data management plan</li></ul>
Architecture Definition - January 2016	<ul style="list-style-type: none"><li>• Design and quality abstractions</li><li>• DICE simulation tools</li><li>• DICE verification tools</li><li>• Monitoring and data warehousing tools</li><li>• DICE delivery tools</li><li>• Architecture definition and integration plan</li><li>• Exploitation plan</li></ul>