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

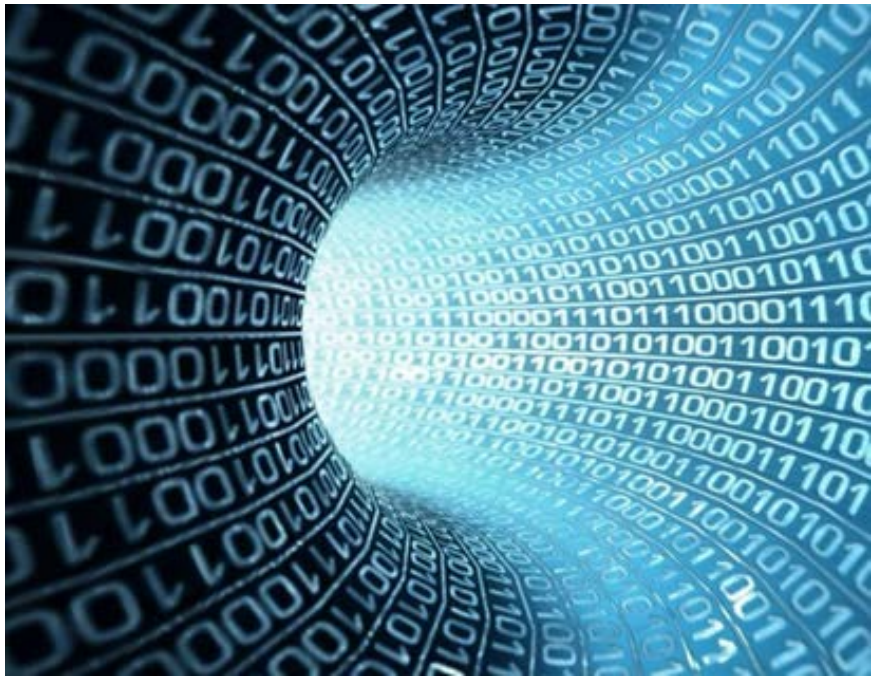
Giuliano Casale
Imperial College London
Project Coordinator

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

Funded by the Horizon 2020
Framework Programme of the European Union

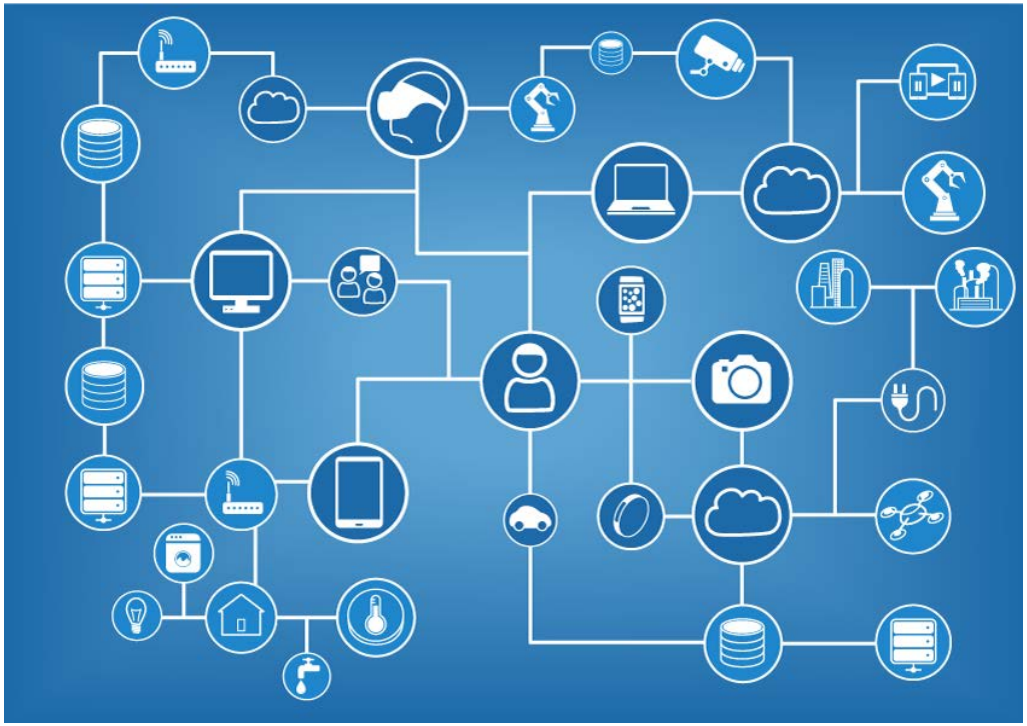
The Rise of Big Data

- Software market rapidly shifting to Big data
  - Expected 32% compound annual growth rate in EU through 2016
  - Just 35% of Big data projects are successful [CapGemini'15]



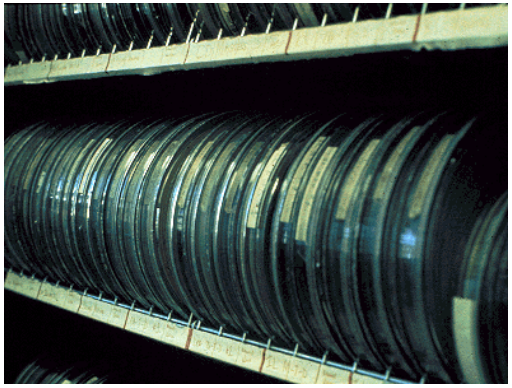
Big Data in Application Domains

- Predictive Analytics
- Environmental Monitoring
- Science & HPC
- E-Government
- Media & news
- Smart cities
- Smart manufacturing
- Healthcare
- Real-time finance
- Advertising
- Fraud detection
- Maritime operations
- …



Example: Tax Fraud Detection

Legacy Tax Data



NETFECTIVE
Creative Technology

Learning curves

Initial prototype

Risk of failure

Big Data Analytics Adoption

hadoop MapReduce

Spark

cloudera

(+ others…)

DICE Mission and Partners

Mission: support SMEs in the development of high-quality cloud data-intensive applications

- ICT 9 Call – Software engineering
- 9 partners (Academia & SMEs), 7 EU countries

Imperial College London

XLAB NOT IDLE

IeAT

Universidad Zaragoza 1542

NETFECTIVE Creative Technology

ATC ATHENS TECHNOLOGY CENTER

POLITECNICO DI MILANO

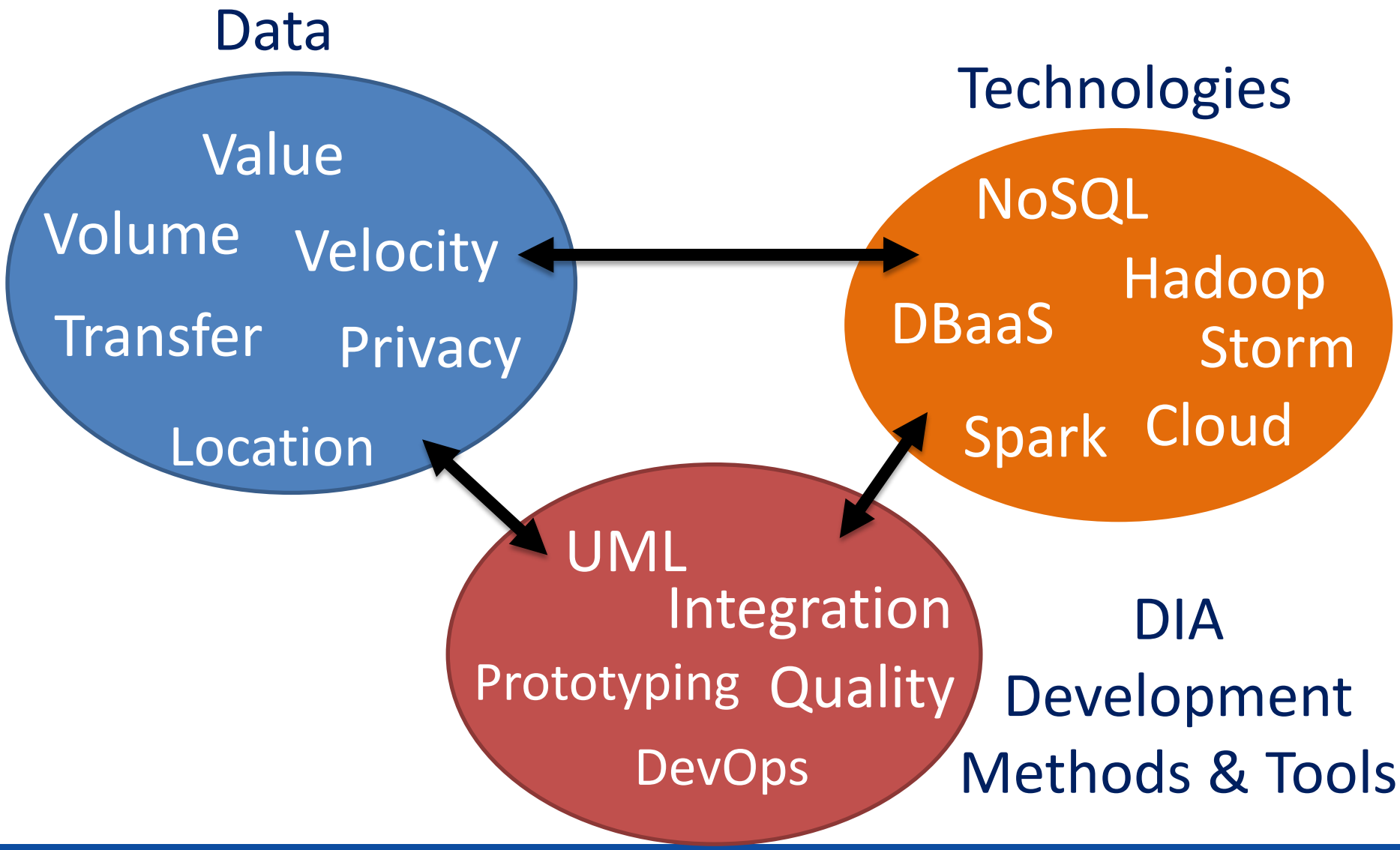
prodevelop Integración de tecnologías

flexiant™ your cloud simplified

DICE Objectives

- Tackling skill shortage and steep learning curves
  - High-degree of automation via DICE tools
- Shorten time to market for DIAs
  - Push out new products, without sacrificing quality
- Decrease development and testing costs
  - Fast iterative definition of application prototypes
- Reduce number and severity of quality incidents
  - Iterative refinement of application design

Data-Intensive Applications (DIA)

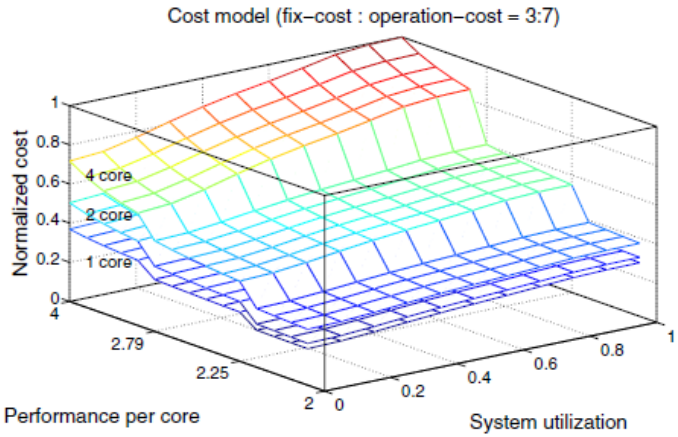


ICT-9 Call for Quality

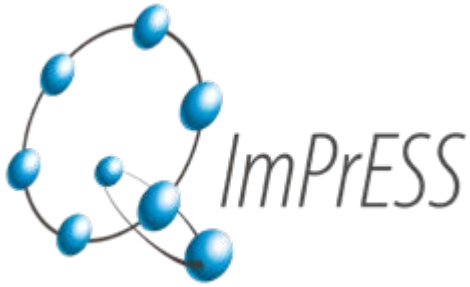
- QA evolving too slowly compared to the technology trends (Big data, Cloud, IoT/CPS, ...)
- ICT-9 call focused on SW quality assurance (QA)
  - ISTAG: call to define environments "*for understanding the consequences of different implementation alternatives (e.g. quality, robustness, performance, maintenance, evolvability, ...)*"
- Early stage issues of data-intensive applications
  - Current drive is on delivering functional capabilities
  - Lack of quality-oriented design poses long-term threats

Quality Dimensions

- Reliability
  - Availability
  - Fault-tolerance
- Efficiency
  - Performance
  - Costs
- Correctness
  - Verification metrics
  - Anomaly detection metrics



Improving Sw. Eng. in Europe



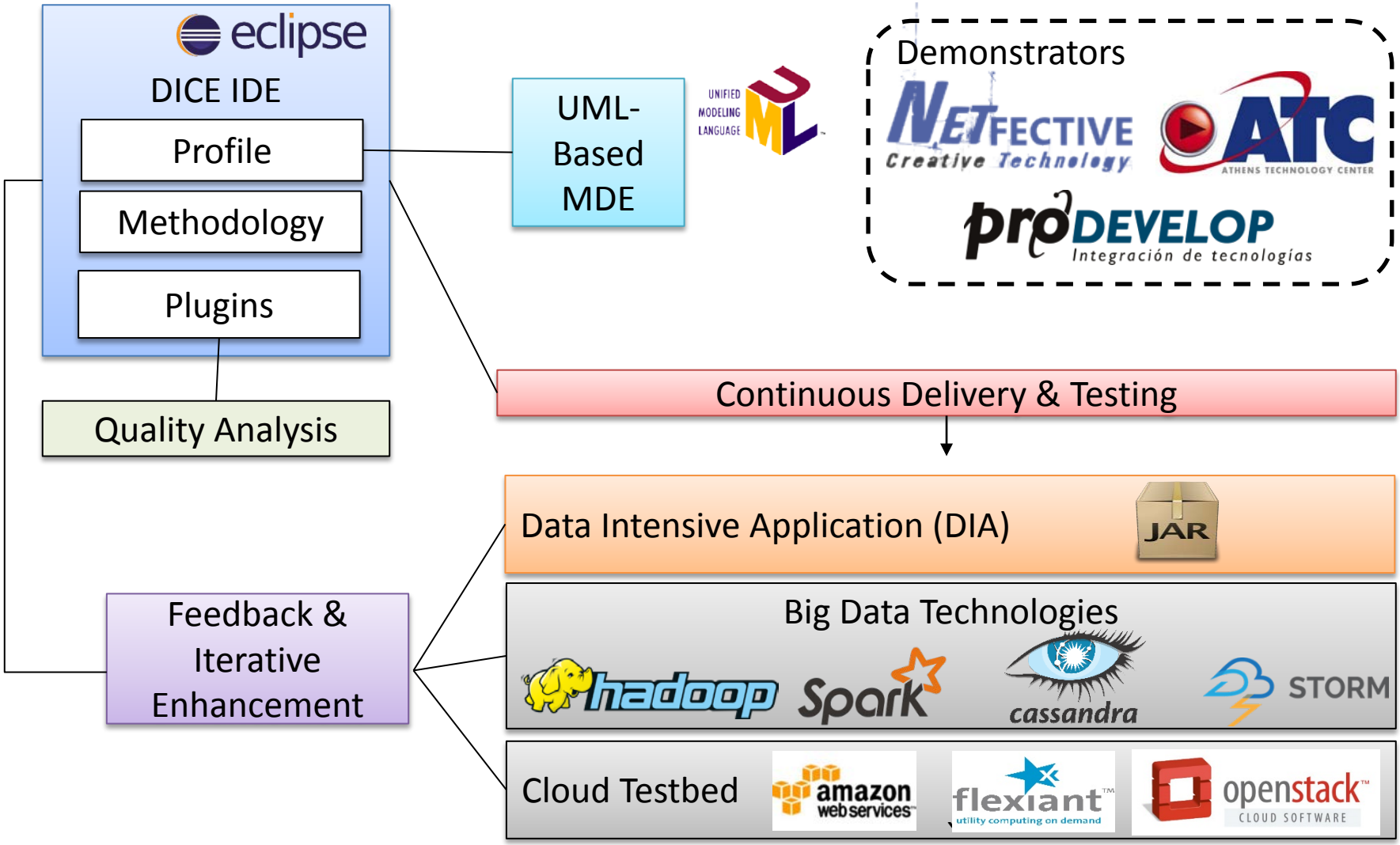
DICE pathways to innovation:

- MDE for Big Data technologies
- Big data technologies are first-class citizens
- Add continuous integration/agile delivery
- TOSCA standard
- Emerging architectures (e.g., Lambda)

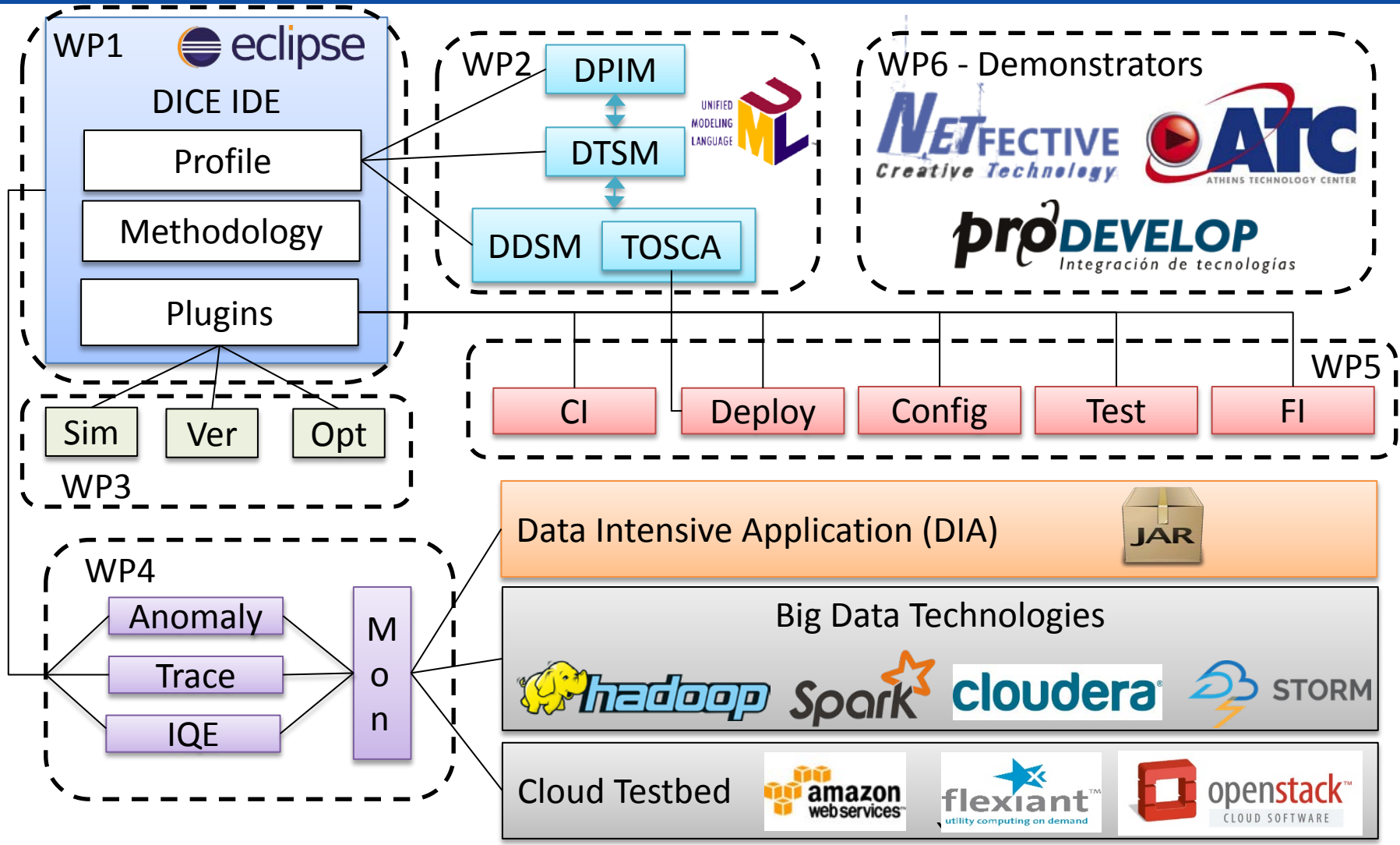
Key Technical Innovations

- DICE Contributions:
  1. Requirements for data (volume, velocity, …) WP2
  2. Requirements for data technologies (e.g., HMR) WP2
  3. Rapid prototyping of DIAs via UML, IDE & DevOps methodology WP1/WP2
  4. Model-based quality analysis and verification WP3
  5. Rapid & optimized deployment via TOSCA WP5
  6. Semi-automated feedback analysis of monitoring and test data to improve design WP4

DICE Framework



DICE Framework



Year 1 - Milestones

Milestone	*Deliverables*
Baseline and Requirements / M6	• State of the art analysis [D1.1] • Requirement specification [D1.2] • Dissemination, communication, collaboration and standardisation [D7.1] • Data management plan [D8.4]
Architecture Definition / M12	• Architecture and integration plan [D1.3] • Design and quality abstractions [D2.1] • DICE simulation tools [D3.2] • DICE verification tools [D3.5] • Monitoring and data warehousing [D4.1] • DICE delivery tools [D5.1] • Exploitation plan [D7.5]

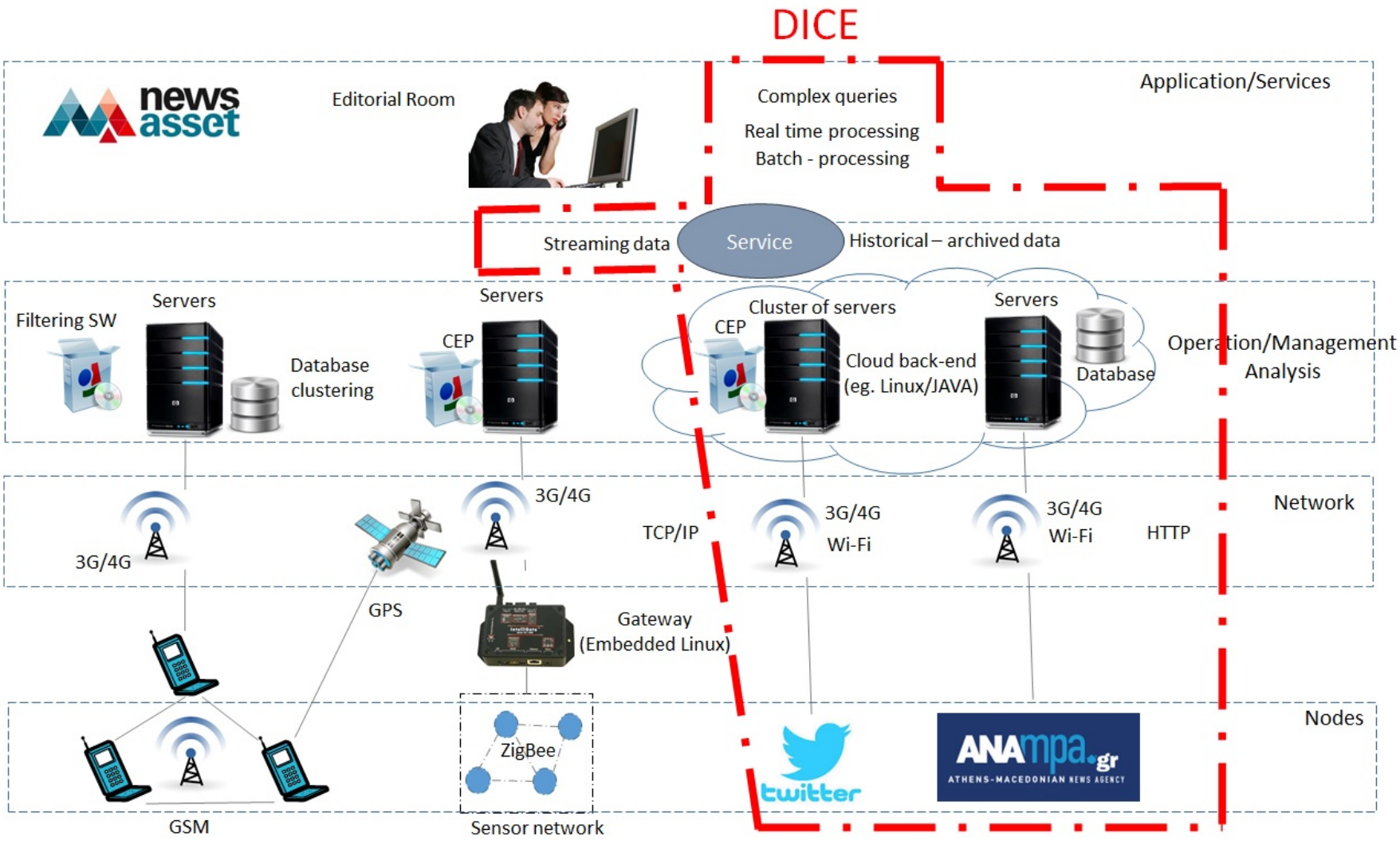
DICE Demonstrators

Case study	Domain	Features & Challenges
Distributed data-intensive media system: NewsAsset (ATC)	• News & Media • Social media	• Large-scale software • Data velocities • Data volumes • Re-engineering
Big Data for e-Government (Netfective)	• E-Gov application	• Data volumes • Legacy data • Data consolidation • Data stores • Privacy • Forecasting and data analysis
Geo-fencing (Prodevelop)	• Maritime sector	• Vessels movements • Safety requirements • Streaming & CEP • Geographical information

Demonstrators

Case study	Domain	Features & Challenges
Distributed data-intensive media system: NewsAsset (ATC)	• News & Media • Social media	• Large-scale software • Data velocities • Data volumes • Re-engineering
Big Data for e-Government (Netfective)	• E-Gov application	• Data volumes • Legacy data • Data consolidation • Data stores • Privacy • Forecasting and data analysis
Geo-fencing (Prodevelop)	• Maritime sector	• Vessels movements • Safety requirements • Streaming & CEP • Geographical information

ATC: NewsAsset



Demonstrators

Case study	Domain	Features & Challenges
Distributed data-intensive media system: NewsAsset (ATC)	• News & Media • Social media	• Large-scale software • Data velocities • Data volumes • Re-engineering
Big Data for e-Government (Netfective)	• E-Gov application	• Data volumes • Legacy data • Data consolidation • Data stores • Privacy • Forecasting and data analysis
Geo-fencing (Prodevelop)	• Maritime sector	• Vessels movements • Safety requirements • Streaming & CEP • Geographical information

Netfective Technologies

Software prototype to demonstrate the capabilities of Big Data in e-government applications: **Fraud detection.**

Example: Social aids should be received only for the living location unless the family moved form one region to another one. However there is currently no process to compare if records are duplicated and to compare with address information in tax declaration.

Demonstrators

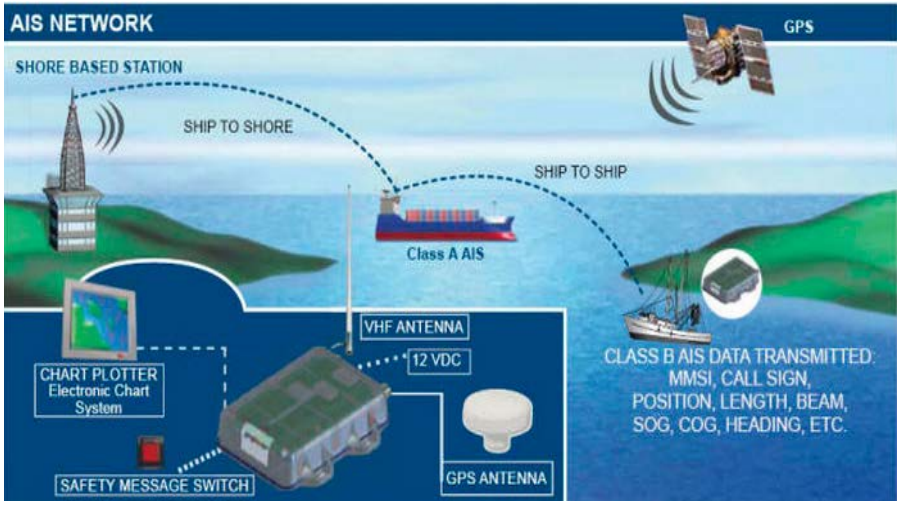
Case study	Domain	Features & Challenges
Distributed data-intensive media system: NewsAsset (ATC)	• News & Media • Social media	• Large-scale software • Data velocities • Data volumes • Re-engineering
Big Data for e-Government (Netfective)	• E-Gov application	• Data volumes • Legacy data • Data consolidation • Data stores • Privacy • Forecasting and data analysis
Geo-fencing (Prodevelop)	• Maritime sector	• Vessels movements • Safety requirements • Streaming & CEP • Geographical information

Prodevelop – Posidonia

Posidonia
Suite de Soluciones Portuarias

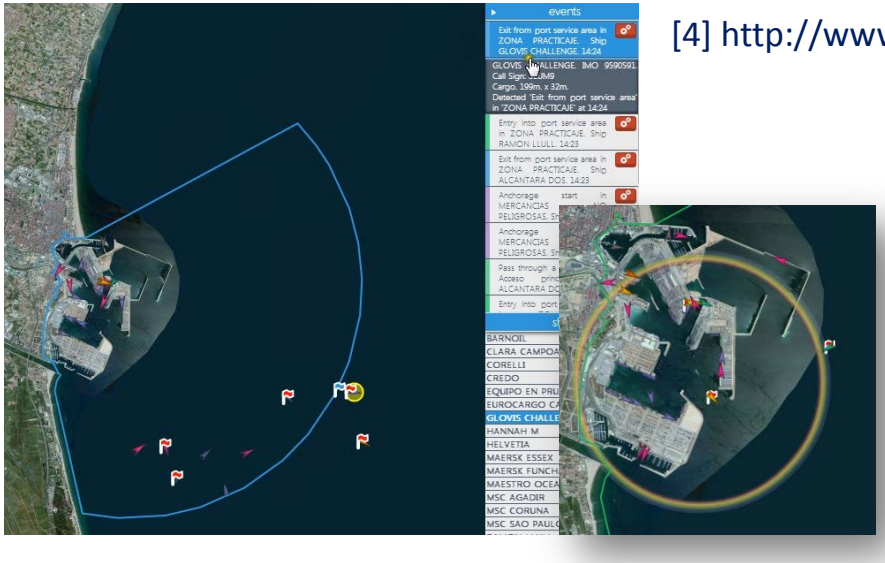
Operations

prodevelop



Example of an AIS network operation [4]

[4] http://www.hollandmarinehardware.nl/index.php?page=Beschrijving%20AIS



	1 port	3 ports	10 ports	100 ports	AISHub (filtered)
Velocity	100 msg/second	200 msg/second	500 msg/second	3000 msg/second	300 msg/second
Volume - 1 day	200MB	400MB	1GB	5GB	600MB
Volume - 1 month	6GB	12GB	36GB	150GB	20GB
Volume - 1 year	70GB	140GB	400GB	2TB	200GB

Thanks!

www.dice-h2020.eu