



Aplicando Ingeniería al desarrollo de aplicaciones de Big Data

José Merseguer
EINA, Zaragoza



**Escuela de
Ingeniería y Arquitectura**
Universidad Zaragoza

Algunos datos

- El mercado del Software se posiciona en Big Data
 - 32%, tasa de crecimiento anual en la UE en 2016
 - 35%, proyectos de Big data tienen éxito [Capgemini 2015]

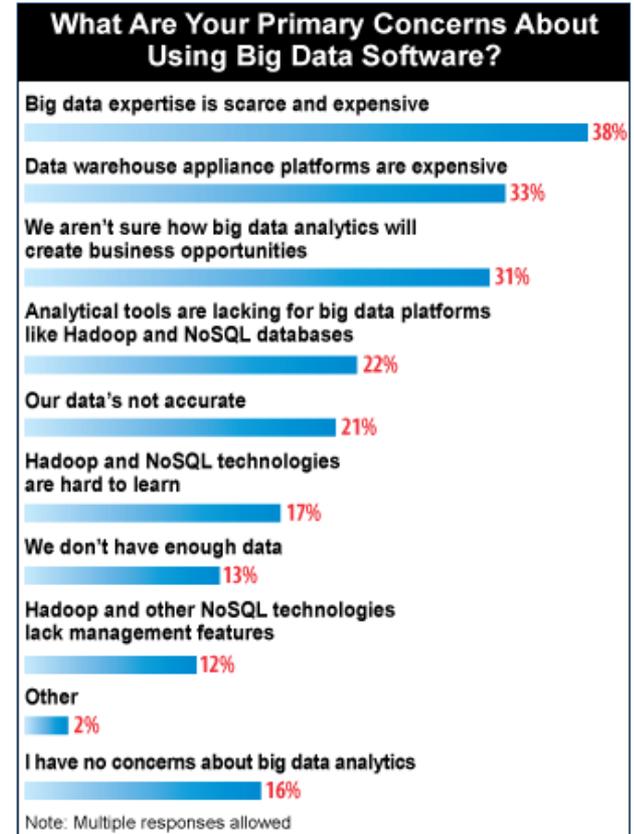
Alrededor de 140.000 a 180.000 ofertas de empleo para científicos de datos permanecerán vacantes alrededor de 2018.

— McKinsey Global Institute



Algunos datos ...

- Capgemini informa que sólo el 13% de las organizaciones consigue producir aplicaciones intensivas en datos a gran escala



InformationWeek CONNECTING THE BUSINESS TECHNOLOGY COMMUNITY



Oportunidad

- Necesidad urgente en proporcionar a las empresas metodologías y herramientas capaces de desarrollar AID con éxito
- Aplicaciones Intensivas en Datos (AID)
 - Procesan grandes cantidades de información
 - Usan tecnologías de Big Data
 - Apache Storm, Apache Spark, Hadoop MR, Cassandra, Tez, ...



Retos y desafíos

- Las tecnologías de Big Data no son fáciles
- La curva de aprendizaje es muy dura
- Las infraestructuras deben ser escalables
- Escasez de expertos
- Las aplicaciones deben ser fiables y cumplir requisitos de calidad
- La seguridad y la privacidad no son opcionales
- Procesos iterativos y basados en DevOps
- Escasez de expertos

Tecnología

Desarrollo



Visión

- Falta desarrollo guiado por la calidad para Big Data
 - ¿Cómo modelar aplicaciones que usan tecnologías de Big Data?
- No existen herramientas que ayuden a razonar sobre los requisitos de calidad de las AID
 - ¿Qué sucede si duplico el número de máquinas virtuales?
 - ¿Qué sucede si *paralelizo* determinados módulos de mi aplicación?

Ingeniería del Software



Calidad AID

○ Reliability

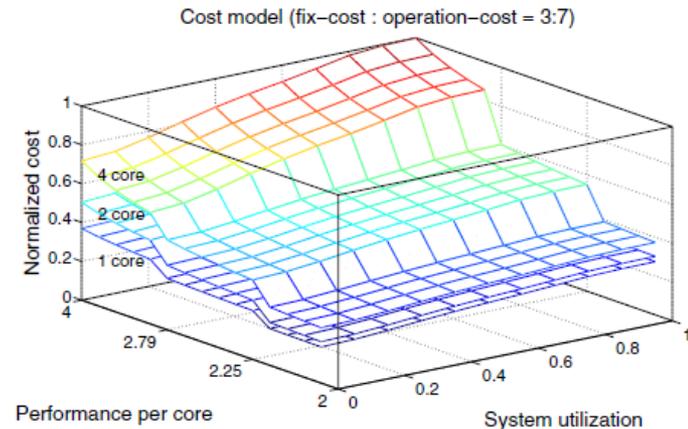
- Availability
- Fault-tolerance

○ Efficiency

- Performance
- Costs

○ Safety & Privacy

- Verification (e.g., deadlines)
- Data protection



¿Cuál es la propuesta de la EINA?

- Participación en el Proyecto DICE
 - DICE - Developing Data-Intensive Cloud Applications with iterative Quality Enhancements
 - Horizon 2020 Research & Innovation Action (RIA)
 - Desarrollo de aplicaciones con Big Data teniendo en cuenta la calidad
 - Feb 2015 - Jan 2018, 4M Euros budget
 - 9 socios (Academia & SMEs), 7 países EU



Objetivos de alto nivel

- Hacer frente a la falta de experiencia y a las curvas de aprendizaje
 - Métodos, modelos y herramientas conscientes de los Datos
- Acortar el tiempo de comercialización de las aplicaciones de Big Data
 - Sin sacrificar calidad en el producto
- Reducir costes en el desarrollo y en las pruebas
 - Seleccionar las arquitecturas que cumplen con las SLAs
- Reducir el número y la importancia de los incidentes en la calidad
 - Refinamiento iterativo del diseño



... in a DevOps fashion

- DevOps cierra la brecha entre Dev and Ops
 - Del desarrollo ágil a la entrega ágil
 - Ciclos de entrega rápida con herramientas y tests automáticos
 - La clave reside en un modelado en profundidad del sistema que permita la automatización



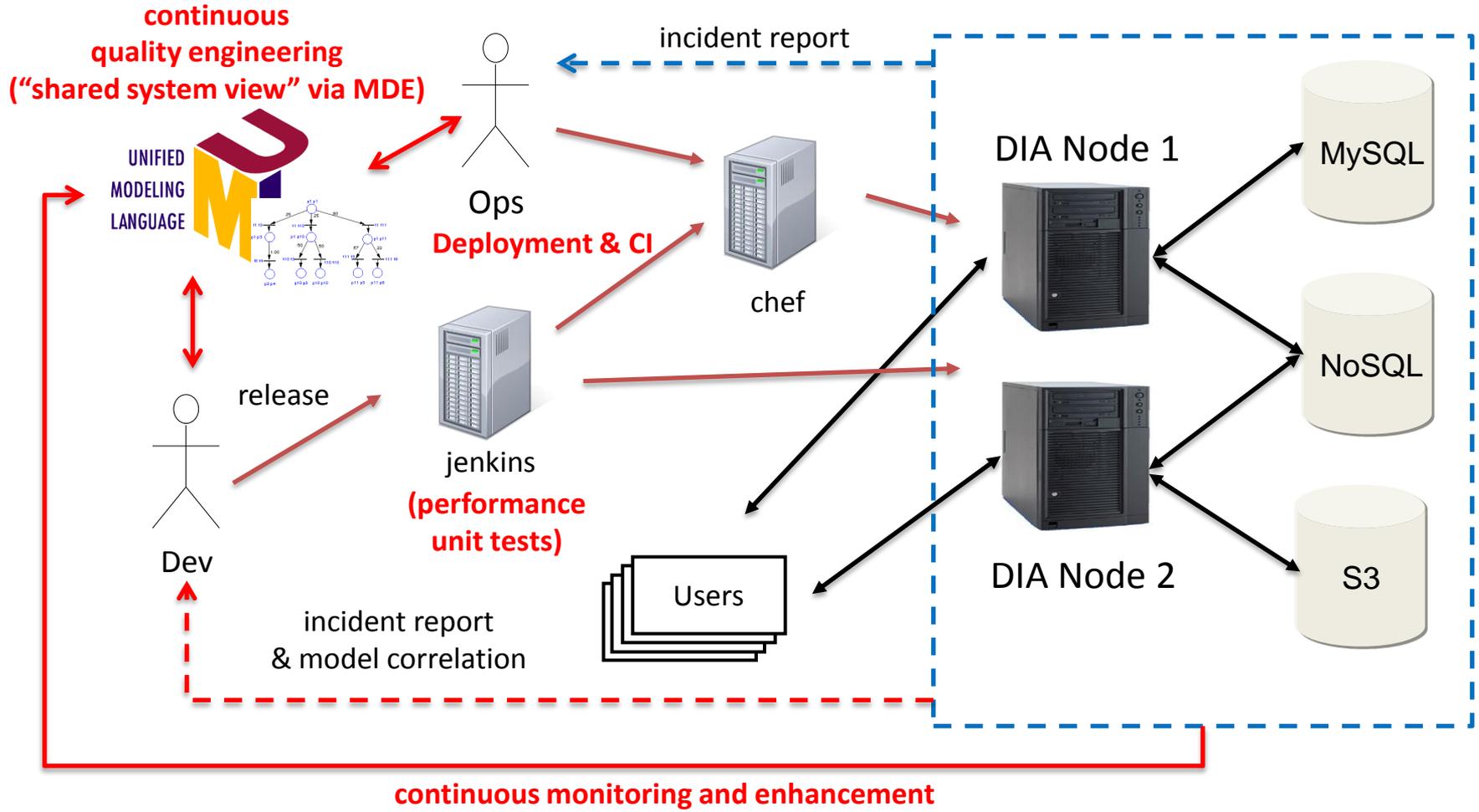
Claves de la innovación técnica

○ Contribuciones de DICE:

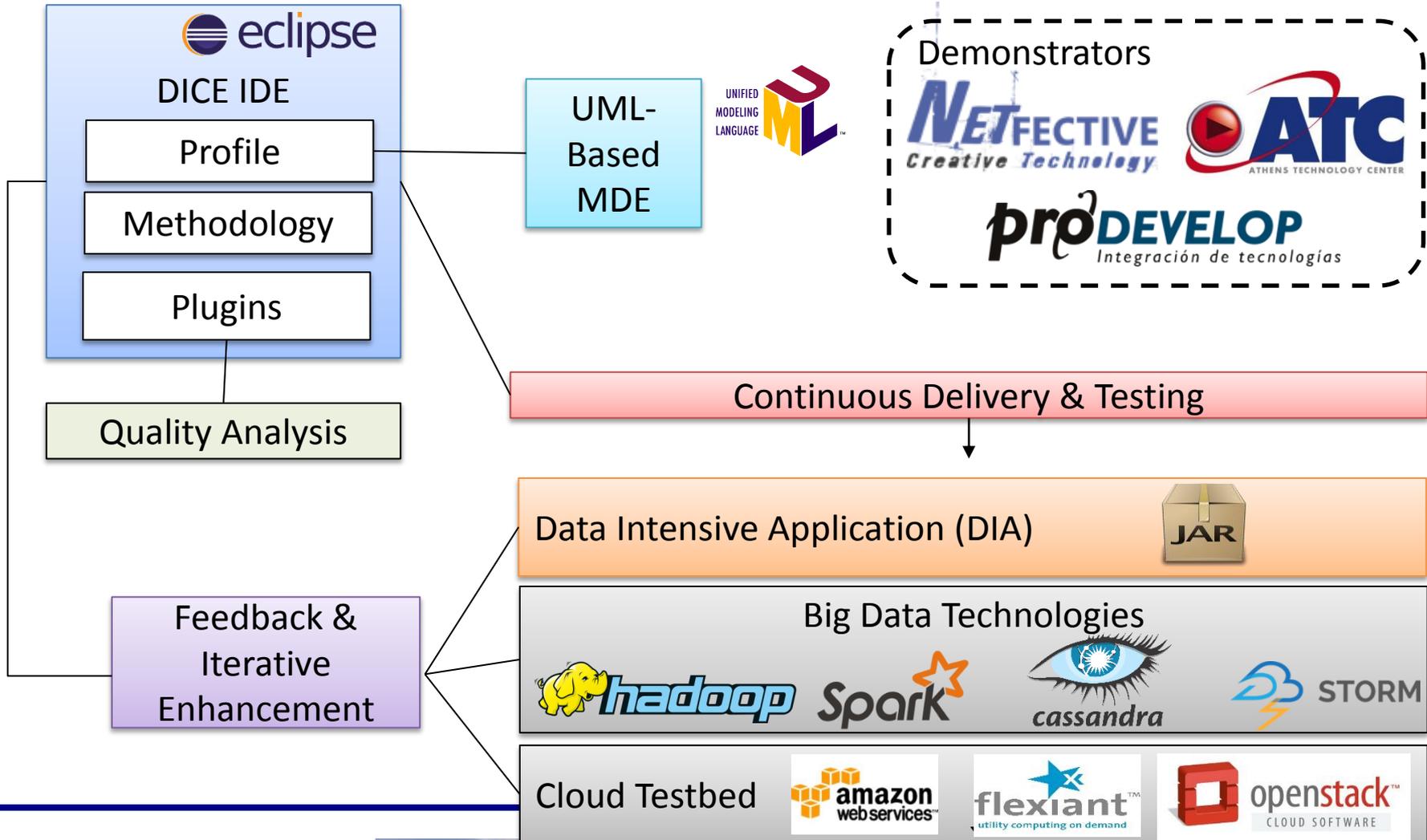
1. Requisitos para datos (volumen, velocidad, ...)
2. Requisitos para las tecnologías de datos (e.g., HMR)
3. Prototipado rápido de AID con UML, IDE & metodología DevOps
4. Verificación y análisis de la calidad basada en modelos
5. Despliegue rápido y optimizado basado en TOSCA
6. Realimentación semi-automática basada en la monitorización y los test para mejorar el Diseño



DevOps en DICE



Framework de DICE



Solución de DICE

