

Uso del Cloud Computing GARR nel laboratorio NeMIS dell'ISTI-CNR di Pisa

Andrea Dell'Amico

2017



InfraScience Research Group @ CNR-ISTI

Group leader: Donatella Castelli

- Istituto di Scienza e Tecnologie dell'Informazione (ISTI)
 - Networked Multimedia Information System Laboratory (NeMIS)
 - InfraScience Research Group



D4Science.org

Integrated technologies that provide elastic access and usage of data and data-management capabilities

D4Science web site: <https://www.d4science.org>

- Virtual Research Environments (VRE) that give access to multiple services
- Data discovery, accessing, analysis, and transformation in standard format
- Powered by gCube: <https://www.gcube-system.org/>



OpenAIRE

The point of reference for Open Access Science in Europe.

<https://www.openaire.eu>

- Operate a pan-European (and global) network for Open Science to articles and research data across countries and across research communities
- Definition and dissemination of guidelines for sharing scholarly products and links between them
- provide services for populating and provide to the public an information graph of interlinked scholarly entities
- provide services for assessing research impact of funders (the Commission in primis) and monitoring of open access trends
- Powered by D-Net:

<http://www.d-net.research-infrastructures.eu/>



Uso attuale delle risorse

- 40 IP pubblici
- 396 vCPU
- RAM: 770 GB

D4Science.org, uso attuale delle risorse

- Dataminer: esecuzione di algoritmi per l'analisi di dati (R, java, C, Fortran)
- Scrittura ed esecuzione interattiva di codice R tramite varie istanze di Rstudio Server

Gli algoritmi hanno bisogno di quantità importanti di risorse di calcolo e memoria. La loro esecuzione può durare anche alcuni giorni, perciò l'affidabilità delle risorse è un requisito fondamentale. I risultati vengono poi salvati in un workspace remoto.



OpenAIRE, uso attuale delle risorse

- Slave server per il servizio di Continuous Integration (Jenkins)
- Cluster Elasticsearch per le attività di sviluppo dei servizi di indicizzazione



Possibili scenari futuri

- Attivazione automatica di istanze per gestire la scalabilità orizzontale: dataminer, rstudio server, sharelatex, shinyproxy
- Storage cluster per le attività di sviluppo (mongoDB, Cassandra, orientDB, ecc.)
- Mirror dello storage di produzione per avere ulteriore HA (la cifratura dei dati at rest è un prerequisito)

Possibili scenari futuri

- Un sistema dinamico di Continuous Delivery
- Un cluster Hadoop/Spark complementare a quello on site
- Federazione dei servizi, appena avremo un'installazione OpenStack dal nostro lato

