



Contribution ID: 61

Type: **Presentazione orale**

## Approccio AI per la selezione dei provider nel sistema di orchestrazione INDIGO PaaS di INFN Cloud

*Tuesday, 21 May 2024 15:10 (5 minutes)*

INFN Cloud fornisce alle comunità scientifiche supportate dall'Istituto un'infrastruttura Cloud federata e un portafoglio di servizi dinamico basato sulle esigenze dei casi studio supportati. Il middleware federativo di INFN Cloud si basa sul sistema di orchestrazione INDIGO PaaS, costituito da microservizi open source interconnessi. Tra questi, l'INDIGO PaaS Orchestrator riceve richieste di deployment di alto livello e coordina il processo di creazione dei deployment sulle piattaforme IaaS messe a disposizione dai provider federati.

Nella configurazione di default, l'INDIGO PaaS Orchestrator determina il provider dove sottomettere la richiesta di creazione del deployment a partire da una lista ordinata di provider, selezionati in funzione del gruppo a cui l'utente appartiene. Tale lista viene fornita dal servizio Cloud Provider Ranker il quale applica un algoritmo di ranking utilizzando un set ristretto di metriche relative ai deployment e ai Service Level Agreements definiti per i provider. L'INDIGO PaaS Orchestrator sottomette il deployment verso il primo provider presente nella lista e in caso di fallimento scala al provider successivo fino ad esaurimento della lista.

In questo contributo viene presentata l'attività finalizzata al miglioramento del sistema di ranking e all'ottimizzazione nell'uso delle risorse attraverso un approccio basato sull'utilizzo di tecniche di intelligenza artificiale. In questo contesto, un importante lavoro preparatorio di identificazione di metriche ritenute più significative è stato messo in essere, identificando anche le sorgenti dalle quali poter reperire le metriche stesse. La successiva preparazione del dataset ci ha permesso di studiare al meglio il caso in esame, e di identificare e confrontare diverse tecniche di intelligenza artificiale. L'approccio proposto prevede la creazione di due modelli: uno predittivo per la classificazione del successo/fallimento del deployment e uno di regressione per il tempo di creazione dei deployment. Una combinazione lineare dell'output dei due modelli, unito ad un training su finestre temporali mobili e recenti, permette di definire una lista ordinata di provider che l'orchestratore può utilizzare per la sottomissione dei deployment.

**Primary author:** GIOMMI, Luca (Istituto Nazionale di Fisica Nucleare)

**Co-authors:** COSTANTINI, Alessandro (Istituto Nazionale di Fisica Nucleare); Mr DEBIASE, Francesco (INFN Sezione di Bari); DONVITO, Giacinto (Istituto Nazionale di Fisica Nucleare); VINO, Gioacchino (INFN (IT)); Dr SAVARESE, Giovanni (Istituto Nazionale di Fisica Nucleare); ANTONACCI, Marica (Istituto Nazionale di Fisica Nucleare)

**Presenter:** GIOMMI, Luca (Istituto Nazionale di Fisica Nucleare)

**Session Classification:** Sessione "Infrastrutture ICT e calcolo distribuito"

**Track Classification:** Infrastrutture ICT e Calcolo Distribuito