



Contribution ID: 55

Type: **Presentazione orale**

Il DataLake nell'infrastruttura INFN: applicazioni e prospettive future

Tuesday, 21 May 2024 15:15 (15 minutes)

Nell'ambito dei nuovi esperimenti di fisica, così come in altri ambiti scientifici e del privato, la gestione dati assume un ruolo cardine e l'aumento esponenziale della quantità di dati prodotti quotidianamente pone una sfida a chiunque voglia costruire una infrastruttura funzionale e all'avanguardia. In questo contesto nasce l'idea del DataLake, nel quale storage fisici geograficamente distribuiti vengono federati, fornendo un livello di astrazione che permette all'utente di interagire con i dati in maniera trasparente, senza dover conoscere i dettagli infrastrutturali.

Nell'ambito delle attività del WP6 di DataCloud, abbiamo dimostrato come sia possibile federare istanze di storage eterogenee, distribuite in diversi centri di calcolo dell'INFN, sulle quali gli utenti possono interagire coi dati in maniera dichiarativa (dichiarando ad esempio il numero di repliche, il QoS, il ciclo di vita, etc...).

Abbiamo implementato questo modello di DataLake utilizzando gli strumenti di Data Management (RUCIO e FTS in primis) già in uso in alcuni grandi esperimenti.

Oltre a testare diverse configurazioni (ad esempio rispetto al modello autorizzativo), abbiamo svolto un'attività di on-boarding di alcune comunità di utenti (ovvero quelle degli esperimenti Cygno e DarkSide), implementando le policy e personalizzazioni necessarie.

Viste le competenze e esperienze maturate, crediamo che i tempi siano maturi per la realizzazione di una infrastruttura DataLake nazionale "catch-all", disponibile per tutti gli utenti, non specifica ad un singolo esperimento.

Primary authors: ITALIANO, Alessandro (Istituto Nazionale di Fisica Nucleare); RENDINA, Andrea (Istituto Nazionale di Fisica Nucleare); TROJA, Antonino (Istituto Nazionale di Fisica Nucleare); SPISSO, Bernardino (Istituto Nazionale di Fisica Nucleare); SPIGA, Daniele (Istituto Nazionale di Fisica Nucleare); CIANGOTTINI, Diego (INFN Perugia); FANZAGO, Federica (Istituto Nazionale di Fisica Nucleare); MORGANTI, Lucia (Istituto Nazionale di Fisica Nucleare); VERLATO, Marco (Istituto Nazionale di Fisica Nucleare); BIASOTTO, Massimo (Istituto Nazionale di Fisica Nucleare); SGARAVATTO, Massimo (Istituto Nazionale di Fisica Nucleare); DELLI VENERI, Michele (Istituto Nazionale di Fisica Nucleare); MARCELLI, Nadir (Istituto Nazionale di Fisica Nucleare); STALIO, Stefano (Istituto Nazionale di Fisica Nucleare)

Presenter: TROJA, Antonino (Istituto Nazionale di Fisica Nucleare)

Session Classification: Sessione "Infrastrutture ICT e calcolo distribuito"

Track Classification: Infrastrutture ICT e Calcolo Distribuito