

Data Management e Storage Federation: verso la realizzazione del Datalake

Tuesday, 27 May 2025 18:30 (15 minutes)

Il WP6 di DataCloud da tempo conduce un'attività di R&D sul data management e storage federation con l'obiettivo di verificare la fattibilità dell'adozione di un modello basato su tecnologie ben note (RUCIO+FTS) per la gestione dello storage presente nei centri di calcolo dell'ente. L'obiettivo prefissato era duplice: una verifica della funzionalità del sistema insieme ad una valutazione dell'impatto non solo sull'infrastruttura HW, ma anche sui servizi e sulle operazioni.

La prima fase di R&D si è conclusa con successo ed è stato implementato un testbed di Datalake nel quale abbiamo federato storage eterogenei sparsi nel territorio nazionale. Questa infrastruttura permette ad un utente di interagire con i dati in maniera trasparente e dichiarativa (specificando, ad esempio, il numero di repliche necessarie di un data set, il ciclo vita, etc..). Un aspetto chiave di questa prima fase è stato l'on-boarding di alcune comunità scientifiche dell'INFN (Darkside, Dampe ed Euclid) che ha permesso di dimostrare con use case concreti la funzionalità del modello proposto. Queste tecnologie sono inoltre utilizzate nel proof-of-concept ICSC e Terabit, nel quale abbiamo in programma di federare anche uno storage esterno all'infrastruttura INFN (CINECA). In aggiunta, abbiamo testato diverse policy di accesso ai dati, visto l'altissimo livello di personalizzazione del sistema.

In questa presentazione mostreremo lo stato dell'arte dell'infrastruttura e le evoluzioni che prevediamo nel breve futuro. Infine mostreremo i modelli di deployment che pensiamo possano essere messi in produzione a supporto degli esperimenti impegnati nella costruzione del loro computing model e che hanno bisogno di supporto rispetto alla gestione dei dati.

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); AGOSTINI, Federica (CNAF-INFN); FANZAGO, Federica (Istituto Nazionale di Fisica Nucleare); DI NUCCI, Francesco (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); MICCOLI, Roberta (Istituto Nazionale di Fisica Nucleare); STALIO, Stefano (Istituto Nazionale di Fisica Nucleare)

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

Session Classification: Calcolo distribuito

Track Classification: Calcolo distribuito