

Analisi quasi-interattiva per big data con alto throughput per la Fisica delle Alte Energie

Friday, 5 April 2024 12:10 (15 minutes)

L'analisi dei dati nell'ambito della Fisica delle Alte Energie presenta le esigenze tipiche dei *big data*, come la grande mole di dati da processare in maniera efficiente e veloce. Il Large Hadron Collider nella sua fase di alta luminosità produrrà circa 100 PB di dati all'anno, ponendo gli esperimenti di fronte alla necessità di adottare nuove strategie per l'analisi dei dati.

Lo sviluppo di nuovi paradigmi di analisi e di gestione delle risorse, unitamente all'accesso a strumenti open source quali *Jupyter*, *Dask* e *HTCondor*, creano i presupposti per riprogettare l'analisi dei dati a LHC, tradizionalmente basata su *batch*, con un approccio interattivo con alto *throughput* che sfrutta il "Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing" (ICSC) su modello data-lake.

In questo contributo descriveremo il lavoro svolto per portare diverse applicazioni di analisi dati - provenienti da varie collaborazioni per una vasta gamma di processi fisici - da un approccio tradizionale a un approccio interattivo basato su soluzioni dichiarative, come ROOT RDataFrame. Queste applicazioni utilizzano un'infrastruttura cloud, distribuendo in carico di lavoro su più nodi *worker*, laddove i risultati sono prodotti su una singola interfaccia in maniera trasparente per l'utilizzatore. Verrà inoltre fornita una valutazione delle prestazioni, tramite metriche indicative riscalate alle risorse di ICSC, al fine di individuare eventuali svantaggi dell'approccio interattivo ad alto throughput che proponiamo e fornire indicazioni utili alla sua implementazione su ICSC.

Primary authors: D'ONOFRIO, Adelina (Istituto Nazionale di Fisica Nucleare); TARASIO, Alessandro (LNF); CAGNOTTA, Antimo (Istituto Nazionale di Fisica Nucleare); SPISSO, Bernardino (Istituto Nazionale di Fisica Nucleare); SIMONE, Federica Maria (Istituto Nazionale di Fisica Nucleare); GRAVILLI, Francesco Giuseppe (Istituto Nazionale di Fisica Nucleare); SABELLA, Gianluca (Istituto Nazionale di Fisica Nucleare); BARTOLINI, Matteo (INFN, Sez. Firenze); ANWAR, Muhammad Numan (Istituto Nazionale di Fisica Nucleare); MASTRANDREA, Paolo (INFN); DIOTALEVI, Tommaso (Università e INFN, Bologna); TEDESCHI, Tommaso (Università e INFN Perugia)

Presenter: BARTOLINI, Matteo (INFN, Sez. Firenze)

Session Classification: Nuove tecnologie