



Contribution ID: 250

Type: Orale

Sviluppo e applicazione di sistemi di Operational Intelligence per le operazioni di calcolo distribuito in ATLAS

La continua evoluzione dei modelli di calcolo distribuito dei moderni esperimenti di Fisica delle Alte Energie comporta un incremento del livello di complessità delle operazioni di processamento e gestione dei dati. Per questa ragione l'esperimento ATLAS sta valutando l'utilizzo di sistemi di Operational Intelligence per automatizzare le procedure di anomaly detection nell'ambito delle operazioni di processamento e distribuzione dei dati nei centri di calcolo afferenti alla World-Wide LHC Computing Grid. Le informazioni relative alle suddette operazioni possono essere attinte da diverse fonti, sia localmente nei centri di calcolo (sistemi di storage e di farming), sia nei sistemi centrali che gestiscono il data processing e management di ATLAS (PanDA, Rucio, FTS).

Un primo testbed per la collezione dei dati locali dei quattro Tier2 italiani di ATLAS è stata realizzato a Roma. I dati vengono raccolti in tempo reale ed indicizzati in una infrastruttura ridondata di Elasticsearch. L'analisi dei dati, in fase di sviluppo, è prevista essere effettuata con tool di DeepLearning tramite autoencoder per determinare le anomalie. L'esecuzione degli algoritmi di DL potrà anche utilizzare una infrastruttura basata su GPGPU NVidia in fase di allestimento a Roma. I dati sono già disponibili dall'inizio di marzo 2019 e comprendono varie metriche a livello di rete e di processi per diverse tipologie di nodi (WN, SE, servizi, etc). Un primo utilizzo previsto di questi dati sarà l'anomaly detection a livello di sicurezza informatica, fino alla realizzazione di un transparent firewall distribuito multisito, capace di generare allarmi in caso di intrusione o anche solo di cattivo utilizzo delle risorse di uno dei componenti.

Per quanto riguarda invece i servizi centrali di ATLAS, le metriche estratte dalle varie fonti (principalmente con procedure di parsing dei log dei sistemi) vengono aggregate in una apposita piattaforma (la ATLAS Open Analytics Platform). Tali metriche saranno poi utilizzate per addestrare un sistema di Operational Intelligence, il cui scopo sarà di identificare potenziali problemi in fase di processamento dei dati e di informare preventivamente gli operatori e gli esperti, sia centralmente che nei siti. Il sistema di Operational Intelligence sarà successivamente configurato per inviare segnalazioni automatiche, le quali saranno rese disponibile anche su piattaforme di visual analytics. L'obiettivo è di aumentare il livello di automazione delle procedure di problem detection, al fine di accelerare le procedure di risoluzione dei problemi e di incrementare conseguentemente l'efficienza nell'utilizzo delle infrastrutture di calcolo.

Primary authors: DORIA, Alessandra (NA); DE SALVO, Alessandro (ROMA1); REBATTO, Davide (MI); VILUCCHI, Elisabetta (LNF); RINALDI, Lorenzo (BO); CLISSA, Luca (BO); GIAGU, Stefano (ROMA1)

Presenter: RINALDI, Lorenzo (BO)

Session Classification: Tecnologie Software e ML

Track Classification: Machine Learning