Contribution ID: **33** Type: **Presentazione orale**

Ambienti virtuali di analisi dati in cloud creati on-demand e compliant con i requisiti tecnici e legali per la gestione di dati genetici e sanitari"

Tuesday, 24 May 2022 11:30 (20 minutes)

Gli aspetti relativi alla sicurezza dei dati e i requisiti legali ed etici sulla conservazione e la gestione dei dati genetici e sanitari stanno diventando sempre più stringenti.
Alcuni ostacoli normativi possono rappresentare un gap alla condivisione dei dati e all'applicazione dei principi di Open Science e Open Acces. In questa prospettiva la Task 6.6 del progetto EOSC-Pillar mirava ad analizzare la compliance normativa del servizio (Laniakea) di analisi dati di livello PaaS integrato e interoperabile per ELIXIR e la comunità di Life Science, frutto di un'interazione tra i servizi Galaxy e i repository di dati.
Partendo dall'attività di ricerca svolta nell'ambito della Task 6.6, il lavoro si pone l'obiettivo di definire i requisiti etici e giuridici che devono essere rispettati al fine di garantire un adeguato bilanciamento tra tutela dei dati e della vita privata ed effettiva applicazione dei principi FAIR, OS e OA.
Dal punto di vista tecnologico, abbiamo implementato le misure necessarie per migliorare la sicurezza dell'intero servizio. In particolare, l'obiettivo è garantire la creazione di ambienti isolati e sicuri per svolgere le analisi dati. Per far questo, ci siamo concentrati su due aspetti critici: la gestione dei dati e il controllo dell'accesso al servizio.
L'isolamento dei dati degli utenti avviene criptando l'intero volume di storage associato alla macchina virtuale, usando il modulo di criptazione del kernel Linux. Il livello di crittografia del disco è completamente trasparente per le applicazioni software, in questo caso Galaxy. La procedura è stata completamente automatizzata attraverso la Dashboard web del servizio di orchestrazione, sfruttando Hashicorp Vault per la conservazione delle passphrase degli utenti. Dopo l'autenticazione sulla Dashboard, l'utente abilita la crittografia dei dati durante la configurazione di una nuova istanza. La Dashboard contatta Hashicorp Vault per ottenere un token che possa essere usato solo una volta. Il token viene passato allo script di crittografia sulla macchina virtuale, viene generata una passphrase casuale, il volume viene crittografato, sbloccato e formattato. Infine, lo script di crittografia accede a Vault usando il token monouso e memorizza la passphrase che sarà accessibile, in qualsiasi momento, solo all'utente tramite la Dashboard. Questa strategia consente di creare chiavi di criptazione sicure ed allo stesso tempo di evitare che le credenziali dell'utente o la passphrase di criptazione siano trasmesse in chiaro all'infrastruttura virtuale, compromettendone la sicurezza.
Il sistema di orchestrazione delle risorse cloud (PaaS) che consente il deployment automatico del servizio di analisi dati è stato esteso per poter gestire la creazione di ambienti virtuali su rete privata , sfruttando l'isolamento delle reti virtuali L2 a livello di tenant garantito dal cloud provider e configurando automaticamente opportuni gruppi di sicurezza che controllano il traffico di rete. In questo modo l'accesso ai vari ambienti di analisi risulta bloccato dalla rete esterna e anche il traffico tra macchine virtuali istanziate sulla stessa rete privata ma appartenenti a deployment diversi risulta filtrato. La modalità di accesso al servizio fornita agli utenti è tramite VPN. Per migliorare l'esperienza dell'utente, l'autenticazione alla VPN è stata integrata con il sistema di autenticazione e autorizzazione, INDIGO IAM, utilizzato dall'intero stack PaaS/IaaS e basato su OpenID Connect. In questo modo, gli utenti non devono creare ulteriori account/credenziali, ma possono utilizzare l'autenticazione federata. In particolare, la soluzione implementata per la VPN si basa sul software opensource OpenVPN e su un modulo PAM sviluppato ad-hoc per permettere l'autenticazione via IAM.

Le soluzioni descritte sono state testate e validate sulla cloud ReCaS-Bari e sull'infrastruttura distribuita multi-sito INFN-Cloud.

Primary authors: DONVITO, Giacinto (Istituto Nazionale di Fisica Nucleare); ANTONACCI, Marica (Istituto Nazionale di Fisica Nucleare); FOGGETTI, Nadina (Istituto Nazionale di Fisica Nucleare); TANGARO, MARCO ANTONIO (BA)

Presenter: ANTONACCI, Marica (Istituto Nazionale di Fisica Nucleare)

Session Classification: Infrastrutture ICT e Calcolo Distribuito

Track Classification: Infrastrutture ICT e Calcolo Distribuito