

Il progetto pilota Cloud Storage GARRbox

Fabio Farina, Fulvio Galeazzi, Luca Prete, Mario Reale, Cristiano Valli, Simon Vocella – Consortium GARR

{fabio.farina, fulvio.galeazzi, luca.prete, mario.reale, cristiano.valli, simon.vocella}@garr.it

Andrea Biancini - INFN Milano Bicocca

andrea.biancini@mib.infn.it

Introduzione a GARRbox ed obiettivi

GARR sta valutando la possibilità di mettere a disposizione un servizio di Cloud Storage per la sua comunità di utenti: obiettivo di tale servizio e' quello di consentire loro di diventare provider di un sistema di storage cloud basato sulla federazione di identità nazionale GARR IDEM.

A tale scopo GARR ha lanciato un progetto pilota nell'ambito del quale ha studiato ed implementato un prototipo di servizio di Cloud Storage denominato GARRbox. Il sistema è attualmente in fase di beta testing ad opera di un ristretto numero di utenti.

Duplici obiettivi di GARRBox e' rispondere ai requisiti fondamentali degli utenti, con particolare riferimento alla comunità biomedica (interfacce e protocolli di accesso multipli, metodi per la condivisione dei files tra utenti e gruppi, autenticazione ed autorizzazione sicure basate sulle federazioni di identità, protezione e tutela dei dati, semplicità di utilizzo, continuità del servizio), e quelli dei provider (alta affidabilità del servizio, disaster recovery, crittazione dei dati e loro replica locale e geografica, semplicità di deployment, integrazione con gli Identity providers della federazione IDEM).

Architettura del sistema

L'architettura implementata attualmente si basa sulla suddivisione in strati funzionali distinti, separati tra loro ed indipendenti: la logica di presentazione lato utente, l'aspetto di aggregazione Cloud delle risorse di storage attraverso un filesystem Cloud, e – al livello più basso – le risorse fisiche integrate nel sistema esposte da una serie di *head nodes* attraverso interfacce POSIX.

Le interfacce lato utente attualmente implementate sono rappresentate da:

- interfaccia Web basata su Ajaxplorer
- interfaccia Amazon S3 basata su Cumulus/Nimbus
- interfaccia WebDAV / http

Lo strato di aggregazione delle risorse è rappresentato dal filesystem cloud GlusterFS che fornisce la replica, la distribuzione locale e lo *striping* dei files. GlusterFS permette anche la georeplica su WAN dei files, fornendo la possibilità di implementare disaster recovery geografico asincrono relativamente ad interi siti.

Lo strato fisico è rappresentato da macchine server sia fisiche che virtuali ospitate presso GARR Direzione e Milano Università Bicocca.

Stato del prototipo e funzionalità attualmente fornite

Le funzionalità attualmente fornite da GARRbox all'utente sono le seguenti: creazione automatica del repository basata sugli attributi utente forniti dall'Identity Provider IDEM, upload/download di files, condivisione files e repository tra utenti, accesso federato all'interfaccia Web, versioning dei files, interfaccia S3 accessibile da S3-client con generazione automatica delle chiavi di accesso, meccanismi di controllo delle quote utente soft ed hard, documentazione e sito di supporto.

Il prototipo fornisce attualmente Cloud Storage ad un gruppo di circa 40 beta tester, basato sull'accesso fornito tramite Identity Provider di GARR Direzione.

Sviluppi futuri

I prossimi sviluppi di GARRbox riguarderanno la migrazione del sistema informativo da MySQL ad un DB nosql (Apache Cassandra) con l'obiettivo di migliorare la scalabilità e le performances complessive del sistema, l'integrazione con il sistema di invio files FileSender, la possibile implementazione di un client di sincronizzazione installabile per Linux e Windows (o l'adozione di un client esistente Open Source), l'encryption, nonché – a lungo termine – la possibile realizzazione (o adozione/customizzazione) di un client su piattaforma mobile (Android, iOS).

In questa presentazione verranno descritti i requisiti utente, lo stato attuale del prototipo GARRbox, la sua architettura, e gli sviluppi futuri in termini di affidabilità, scalabilità, integrazione con IDEM, clients, come anche verrà brevemente accennato alle nuove collaborazioni che il team GARRbox sta iniziando con iniziative analoghe.

Riferimenti

GlusterFS: <http://www.gluster.org>

Ajaxplorer: <http://ajaxplorer.info>

GARRbox: <http://garrbox.dir.garr.it>

Amazon S3 <http://aws.amazon.com/s3>