



Newsletter ReCaS-Bari - N.7 - Novembre/Dicembre 2016

www.recas-bari.it

Questo e i numeri precedenti della Newsletter ReCaS sono disponibili a questo [link](#).

In questo numero:

- Review INDIGO-DataCloud
- Partecipazione di ReCaS-Bari a SuperComputing 2016
- Galaxy on the Cloud: lo use case ELIXIR-IIB in INDIGO-Datacloud
- Rilasciata in produzione la nuova procedura di richiesta delle credenziali per l'accesso alle risorse e ai servizi ReCaS-Bari
- Crediti

Review INDIGO-DataCloud

Nei giorni 7-8 Novembre si è svolta a Bologna (nella sede dell'INFN-CNAF) la prima review ufficiale del progetto europeo INDIGO-DataCloud.

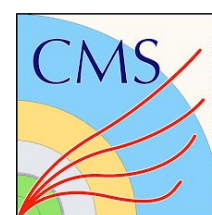


Il gruppo di revisori internazionali chiamati a valutare il progetto era composto da esperti appartenenti a vari contesti del mondo IT: Mr. Tomer Hagay Nevel (Tintri LTD), Prof. Christian Ohmann (ECRIN), Ms. Maria Pilar Santamaria Hernando (Microsoft/European University of Madrid), Dr. Ola Spjuth (Uppsala University/Karališka Institutet), coordinati dal Project Officer che segue il progetto INDIGO per conto della Commissione Europea: Dr. Enrique Gomez.

Sono state mostrate sei *demo*, relative ad altrettante applicazioni in sei differenti domini scientifici, a testimonianza della gran mole di lavoro prodotto (tre di queste *demo*, quella per Lifewatch, Elixir e CMS, facevano uso dell'infrastruttura cloud di

ReCaS-Bari). In particolare:

- Per **Lifewatch** è stata mostrata la possibilità di usare il framework di Mesos chiamato Chronos per sottomettere una grande quantità di job basati su container Docker a qualsiasi infrastruttura cloud, sfruttando la gestione dei dati su rete geografica usando l'infrastruttura di storage chiamata OneData.
- Per **Elixir** è stata mostrata la possibilità di istanziare *on-demand* un intero servizio web di gestione di workflow specificamente orientato alla bioinformatica, con la gestione automatizzata di un cluster di nodi di calcolo connesso al frontend.
- Per **CMS** è stata mostrata la possibilità di istanziare *on-demand* un cluster Tier3-like in cui qualsiasi utente della comunità possa usare i tool di esperimento per eseguire l'analisi di un dataset. In questo caso il cluster viene realizzato usando Mesos insieme a Marathon, un framework in grado di gestire processi *long-running*. Questa demo ha dimostrato la possibilità di usare qualsiasi tipo di risorsa cloud in modo opportunistico per eseguire analisi dell'esperimento CMS.
- Per **Climate Change** è stata mostrata la possibilità di istanziare dinamicamente e in modo automatico un cluster di *big data analytics* per lo studio di modelli di simulazione dei cambiamenti climatici. A differenza degli esempi precedenti, tali algoritmi richiedono risorse tipicamente HPC, e la demo ha dimostrato la flessibilità della piattaforma realizzata dal progetto INDIGO.
- Per **Theoretical Physics** sono state presentate le funzionalità di uDocker, un tool in grado di permettere all'utente non privilegiato di un cluster HPC di eseguire job che abbiano bisogno di utilizzare un particolare Docker container, anche nel caso in cui non fosse possibile installare realmente l'applicativo Docker.
- Per **Molecular Dynamics** è stata dimostrata la possibilità di eseguire calcoli ottimizzati per l'uso di GPU anche in ambiente cloud, grazie alla possibilità di eseguire docker container nativamente sull'hardware fisico in un ambiente cloud basato su OpenStack o OpenNebula.



Alla fine della review la commissione di esperti è risultata positivamente impressionata dalle attività svolte e dai risultati raggiunti nella prima metà del progetto.

Giudizi favorevoli sono stati espressi sul software implementato dal progetto, sulla documentazione messa a disposizione della comunità di utenti, e sulle attività di gestione, che sono risultate adeguate e ben strutturate.

Partecipazione di ReCaS-Bari a SuperComputing 2016

A Novembre si è svolta a Salt Lake City, Utah, USA la ventinovesima edizione di **SuperComputing** (SC16), che ha registrato ben oltre 10000 presenze. Questo evento, caratterizzato da una parte espositiva ed una congressuale, offre l'opportunità a produttori ed utilizzatori di hardware e software per il calcolo ad alte prestazioni di confrontarsi e mostrare lo stato dell'arte delle soluzioni disponibili nell'ambiente HPC e HTC. Tra i produttori presenti citiamo Intel e NVIDIA, dominatori indiscussi dell'Industria HPC (in particolare NVIDIA per la parte di "accelerator/manycore processor"). Tra gli utilizzatori, molti dei Computing Center in pianta stabile nella "TOP500" erano presenti per illustrare i risultati ottenuti grazie alle loro infrastrutture di calcolo super potenti.



"Nella foto il booth dell'INFN con al centro il poster di ReCaS-Bari"

Per la prima volta quest'anno anche il data center ReCaS-Bari era presente ad SC16 con un poster esposto nel booth organizzato e gestito dall'INFN. In particolare, quest'anno al booth era stata assegnata una buona posizione nella zona espositiva, vicino a partecipanti importanti che calamitano la maggior parte dell'attenzione del pubblico, e questo ha fatto sì che le attività relative al computing all'interno dell'INFN, descritte nei vari poster presenti, godessero di una buona visibilità. In particolare, per ReCaS-Bari ha destato particolare interesse l'attività relativa all'estensione geografica del Tier1 sulle risorse ospitate nel data center barese.

Poster link: <https://www.recas-bari.it/images/posters/SC16-ReCaS-Last.pdf>

Galaxy on the Cloud: lo use case ELIXIR-IIB in INDIGO-Datacloud

ELIXIR-IIB (elixir-italy.org), il nodo italiano di ELIXIR, raggruppa al suo interno molte tra le più importanti istituzioni italiane nel campo delle scienze della vita e ha come missione quella di consolidare e sviluppare l'infrastruttura per la bioinformatica del nostro paese, integrandola nello stesso tempo in quella continentale.

A tal fine, ELIXIR-IIB è interessata a tutte quelle soluzioni che permettano di integrare i dati biologici e le piattaforme computazionali in ambito europeo, assicurando ai ricercatori un facile e comodo accesso e sfruttamento di queste risorse. Un esempio in questa direzione è rappresentato da Galaxy (galaxyproject.org), una piattaforma dedicata alla gestione di workflow per analisi bioinformatiche, open source e basato su interfaccia web.

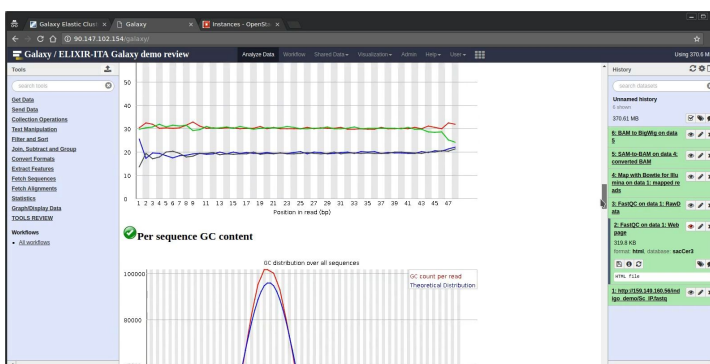
ELIXIR-IIB sta sviluppando all'interno del progetto Horizon 2020 INDIGO-DataCloud (indigo-datacloud.eu) una piattaforma che consente di automatizzare la creazione di uno spazio di lavoro *on-demand* basato su Galaxy, sfruttando tecnologie di virtualizzazione sviluppate dal progetto. Lo spazio di lavoro virtuale sarà a disposizione di ricercatori che non dispongono delle risorse necessarie (umane e/o computazionali) per la configurazione di una propria istanza Galaxy personalizzata.

Una demo del servizio, ancora in fase di prototipo, raggiungibile a questo link: <https://www.youtube.com/watch?v=TdaZxCbDRb8&t=14s>, è stata presentata alla prima revisione periodica del progetto INDIGO-DataCloud ai revisori della Comunità Europea, che hanno espresso un giudizio complessivamente positivo sul progetto, ivi compreso lo use case ELIXIR-IIB.

I tool FastQC (il cui output è mostrato in figura), Bowtie, sam_to_bam e bam_to_bigwig, sono stati installati e configurati automaticamente all'interno di una istanza Galaxy, attraverso il servizio sviluppato da ELIXIR-IIB. Utilizzando questi tool, è stato creato un workflow, i cui risultati sono stati mostrati durante la presentazione.

Il prototipo, attualmente ospitato sul testbed INDIGO installato nel data center ReCaS-Bari, consente di lanciare una macchina virtuale o un Docker container, basato su CentOS 7 o Ubuntu 14.04, con all'interno Galaxy e tutte le applicazioni ausiliarie (PostgreSQL, Nginx, uWSGI, proftpd,) già installate, utilizzando il servizio di orchestrazione PaaS di INDIGO. La configurazione delle risorse virtuali a disposizione dell'utente e la personalizzazione dell'istanza Galaxy, attraverso l'installazione di tool bioinformatici e di dati di riferimento, avviene attraverso un'interfaccia web, creata usando il portale FutureGateway sviluppato nel progetto INDIGO.

L'obiettivo finale del progetto è garantire agli utenti accesso alle risorse computazionali e infrastrutturali, e allo stesso tempo accelerare lo sviluppo di analisi e workflow complessi e l'analisi di dataset di grandi dimensioni. Un'istanza privata di Galaxy su cloud consente infatti agli utenti di personalizzare l'ambiente di lavoro secondo le proprie necessità, rimuovendo le difficoltà iniziali dovute alla



configurazione.

Rilasciata in produzione la nuova procedura di richiesta delle credenziali per l'accesso alle risorse e ai servizi ReCaS-Bari

Fra i lettori di questa newsletter, coloro che hanno avuto la necessità di richiedere le credenziali per l'accesso ai servizi offerti da ReCaS-Bari hanno subito la farraginosità della procedura: il processo di registrazione e autorizzazione di nuovi utenti prevedeva una lunga serie di azioni, talvolta complesse, che venivano gestite attraverso scambi di mail tra gli operatori e i nuovi utenti.

Con il nuovo sistema, gli utenti vengono invitati a compilare un form presente sul sito <https://www.recas-bari.it/index.php/it/recas-bari-servizi/richiesta-credenziali>

attraverso il quale, dopo aver inserito i propri dati anagrafici, possono richiedere l'accesso a risorse di calcolo e/o l'assegnazione di una quota di risorse cloud. Tali richieste vengono prese in carico dal sistema che, dopo aver effettuato un controllo sulla correttezza formale dei dati e sull'unicità delle credenziali (al fine evitare duplicati o doppie registrazioni), suddivide ed inoltra per competenza le richieste ai responsabili dei vari servizi; quelli al grid computing richiedono la creazione di un account LDAP, mentre le risorse di cloud computing fanno capo alla piattaforma OpenStack ospitata da ReCaS-Bari.

Lo scambio di comunicazioni fra il sistema e gli utenti o gli amministratori avviene mediante e-mail contenenti le informazioni necessarie e i link alle pagine dalle quali effettuare le operazioni di gestione; in tal modo, per i responsabili è sufficiente collegarsi a una pagina web, selezionare le risorse da autorizzare e confermare il tutto utilizzando le proprie credenziali - il sistema provvederà in modo automatico all'inserimento del nuovo utente nel server LDAP, all'inserimento nel server OpenStack ed alla creazione delle risorse cloud richieste.

La nuova procedura, essendo sperimentale, necessita di un periodo di rodaggio. Pertanto, ci scusiamo con i primi utenti che la utilizzeranno per gli eventuali disagi che potranno verificarsi. Suggesti e consigli finalizzati a una ulteriore semplificazione della procedura sono benvenuti, e possono essere indirizzati a recas.bari@gmail.com o support@lists.ba.infn.it.

Crediti

Hanno collaborato a questo numero della newsletter:

Giacinto Donvito, Alessandro Italiano, Giorgio Maggi, Stefano Nicotri, Michele Perniola, Marco Tangaro.

Richiesta di accesso ai servizi ReCaS
I campi contrassegnati da * sono OBBLIGATORI

1 Dati personali
Inserisci i tuoi dati personali

2 Selezione servizi
Seleziona i servizi richiesti

3 Selezione risorse
Seleziona le risorse Cloud

4 Inoltro richiesta
Invia la tua richiesta

Dati personali dell'utente

Nome *
Rossi

Cognome *
Mario

Email *
rossi.mario@infn.it

Username desiderato (max 10 caratteri) *
mrossi

Indirizzo abitazione *
via dei gelati, 1

Codice Fiscale *
RSSMRR70P12A662H

Affiliazione *
INFN

Specificare affiliazione

Ruolo richiesto *
Ricerca

Specificare Ruolo

Nome gruppo ricerca *
RECAS

Persona di riferimento

SELEZIONA FILE DOCUMENTO File del documento (documenti) (formato) ... Solo file .pdf

Buone Feste



ReCaS-Bari