



RECaS BARI



Newsletter ReCaS-Bari - N.9 - Aprile/Maggio 2017

www.recas-bari.it

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

In questo numero:

- L'interruzione dei servizi ReCaS-Bari nella settimana dell'11 aprile
- ReCaS-Bari alle Masterclass degli esperimenti ALICE, CMS e FERMI
- TensorFlow@ReCaS-Bari
- Gli utenti del cluster HPC di ReCaS-Bari
- Corso su Parallel Computing and GPU Programming using CUDA
- ReCaS-Bari ad EGI Conference 2017 e INDIGO Summit 2017
- Secondo Corso di formazione: "Uso del software INDIGO-DataCloud in ambito scientifico"
- Crediti

L'interruzione dei servizi ReCaS-Bari nella settimana dell'11 aprile

Approfittiamo dell'uscita di questa newsletter per scusarci con tutti gli utenti del disservizio causato dalla inattesa interruzione dell'alimentazione elettrica del data center verificatasi il 11/04/2017.

Nell'ultimo anno, superata la critica fase iniziale, le attività si sono svolte in maniera positiva, fino a questa interruzione: si sono superate senza alcuna difficoltà numerose interruzioni di energia elettrica (inclusa l'ultima avvenuta nella notte fra domenica 21/05/2017 e lunedì 22/05/2017), dimostrando che l'infrastruttura e il sistema di supervisione rispondono in maniera corretta e riescono a gestire le situazioni di emergenza.

L'interruzione dell'11 aprile scorso è stata causata da un errore umano. In futuro si cercherà di evitare con ogni mezzo che ciò avvenga nuovamente. A questo scopo, si stanno rivedendo le tutte le procedure del data center (anche quelle umane), cercando di renderle più robuste e aumentare la resilienza di ReCaS-Bari.

Cogliamo infine l'occasione per informarvi che è stato attivato sul sito www.recas-bari.it un servizio di "news" ed il relativo sistema RSS. Anche questo canale sarà utilizzato per informare gli utenti su quello che accade all'interno del data center. Invitiamo quindi tutti ad iscriversi al servizio RSS e farci pervenire suggerimenti e critiche per migliorarlo.

Il team di ReCaS-Bari

ReCaS-Bari alle Masterclass degli esperimenti ALICE, CMS e FERMI

Il 29 e il 30 marzo 2017 si è tenuta presso la Sezione di Bari dell'Istituto Nazionale di Fisica Nucleare (INFN) la tredicesima edizione delle *International Masterclasses*.



L'iniziativa è organizzata dall'International Particle Physics Outreach Group (IPPOG) ed è coordinata in Italia dall'INFN. Le masterclass, che si tengono in simultanea presso 210 tra i più prestigiosi enti di ricerca e università di tutto il mondo, coinvolgono ogni anno quasi 3000 studenti delle scuole superiori italiane e oltre 10000 loro coetanei di 47 nazioni diverse.

Nelle due giornate, 120 studenti provenienti da diverse Scuole Superiori di Bari e Provincia hanno vissuto in prima persona l'esperienza quotidiana dei ricercatori che partecipano agli esperimenti ALICE e CMS al Large Hadron Collider (LHC) del CERN di Ginevra.

I ragazzi sono stati accompagnati dai ricercatori in un viaggio alla scoperta delle proprietà delle particelle elementari e dei segreti del grande acceleratore di particelle LHC, che ha permesso la scoperta nel luglio 2012 del bosone di Higgs. Ciascuna delle due giornate, organizzate localmente dal prof. Donato Creanza, dal dott. Vito Manzari e dal prof. Salvatore My, è iniziata con alcuni brevi seminari introduttivi sugli argomenti fondamentali della fisica delle particelle, tenuti dalla dott.ssa Fulvia De Fazio, e sugli esperimenti che consentono di studiarne le proprietà ed è proseguita nel pomeriggio con esercitazioni al computer utilizzando dati raccolti dagli esperimenti CMS e ALICE. Nel corso delle esercitazioni gli studenti hanno avuto accesso, con il contributo e il supporto del team di ReCaS-Bari, ai dati reali provenienti dai due esperimenti e hanno ripercorso in un viaggio accelerato nel tempo tutte le tappe di elaborazione ed analisi tipiche dell'attività dei ricercatori. Le analisi sono state condotte su macchine virtuali (VM) con sistema operativo Linux, operanti sull'infrastruttura Cloud@ReCaS-Bari, a cui gli studenti si sono connessi tramite laptop messi a loro disposizione. Al termine delle esercitazioni, proprio come in una vera collaborazione scientifica internazionale, gli studenti si sono collegati in videoconferenza con il CERN e con i coetanei di altre università italiane e straniere che hanno svolto in simultanea gli stessi esercizi, per confrontare e discutere

insieme i risultati ottenuti.

Nello stesso spirito, il 5 aprile si è svolta a Bari, organizzata localmente dal Dott. Fabio Gargano, la prima edizione della Fermi Masterclass, una analoga iniziativa della sezione di Bari dell'Istituto Nazionale di Fisica Nucleare rivolta alle scuole, organizzata in collaborazione con il Dipartimento Interateneo di Fisica "M.Merlin" dell'Università e del Politecnico di Bari. 60 studenti delle scuole superiori di Bari e provincia hanno potuto affacciarsi alla scoperta dei segreti del nostro Universo.

Anche questo evento ha avuto respiro nazionale ed internazionale, poiché si è svolto in contemporanea a Bari, Perugia, Torino, Trieste e Nova Gorica (SLO), dove oltre 200 studenti hanno vissuto la stessa esperienza.

La giornata è iniziata con alcuni seminari introduttivi sugli esperimenti nello spazio come Fermi, il satellite NASA in orbita dal 2008, dedicato allo studio della radiazione gamma di altissima energia, a cui l'Italia partecipa come INFN, INAF e ASI, e sull'analisi dei dati e altri argomenti di fisica delle astroparticelle. Nel pomeriggio si sono svolte delle esercitazioni pratiche in cui i ragazzi hanno studiato la natura di una delle sorgenti più brillanti del cielo gamma, il nucleo galattico attivo (AGN) 3C 454.3, distante 7,7 miliardi di anni luce da noi e contenente un buco nero di massa superiore a 1 miliardo di volte quella del nostro Sole, costruendone la "curva di luce" e la "mappa dei conteggi" e scoprendone la natura estremamente variabile della sua emissione gamma, la cui interpretazione, è ancora argomento di discussione nella comunità scientifica.

Anche queste analisi sono state eseguite su un ambiente con accesso desktop remoto via web, progettato e realizzato dal personale di ReCaS-Bari sulle infrastrutture di calcolo a disposizione nel data center.

Alla fine della giornata, gli studenti si sono collegati con i loro colleghi nelle altre città in videoconferenza per discutere i risultati emersi dalle esercitazioni, proprio come succede quotidianamente in una vera collaborazione scientifica internazionale.

TensorFlow@ReCaS-Bari

TensorFlow è una libreria open source sviluppata da Google finalizzata alla analisi di dati, segnali e immagini di grande complessità, e si pone come strumento alternativo ad un'altra libreria open source con cui condivide le finalità, di nome Theano.



Il paradigma su cui TensorFlow si basa riguarda l'elaborazione dei dati attraverso la

declinazione più recente degli algoritmi data driven: le reti neurali profonde e gli algoritmi cosiddetti di deep learning.

Un ulteriore strumento denominato Keras permette di sviluppare le reti neurali profonde con un'interfaccia di alto livello, senza la necessità di approfondire algoritmi di calcolo ed elaborazione dei dati.

La diffusione di queste librerie a livello mondiale è molto elevata, e una vastissima ed eterogenea platea di attori ha manifestato e manifesta interesse nei loro confronti, dal mondo aziendale a quello accademico.

I ricercatori della Sezione di Bari dell'Istituto Nazionale di Fisica Nucleare e del Dipartimento Interateneo di Fisica "M. Merlin" utilizzano TensorFlow in diversi ambiti della ricerca di base e applicata, in particolare nel settore della fisica medica, nelle neuroscienze computazionali e in fisica delle astroparticelle.

In fisica medica le reti neurali profonde sono impiegate, utilizzando il software dedicato *DeepMedic*, per lo studio di biomarker, in particolare la morfometria dell'ippocampo, per la diagnosi precoce della malattia di Alzheimer. Questo argomento permette una valutazione stringente sulle prestazioni delle reti neurali profonde, essendo lo studio dell'ippocampo già stato affrontato con una moltitudine di algoritmi data driven alternativi e più tradizionali: esiste quindi una vasta letteratura scientifica con indicazioni quantitative precise sulle prestazioni ottenibili, ad esempio, nella precisione della stima del volume ippocampale.

Nel settore delle astroparticelle TensorFlow viene utilizzato per affrontare problemi di classificazione di dati acquisiti dal calorimetro elettromagnetico dell'esperimento DAMPE. Attualmente le sperimentazioni prevedono l'uso di dati simulati e l'utilizzo della rete open-source *Inception v3* già allenata su dati standard generici. L'attività vede coinvolti, oltre che i ricercatori che partecipano all'esperimento DAMPE, anche laureandi del Dipartimento di Informatica.

Il progetto è nato dall'idea di testare l'utilizzo di un modello di classificatore basato su rete neurale quasi completamente pre-allenata su un campione estremamente generico di immagini (come *Inception v3*).

Il sistema messo a punto per la discriminazione tra cascate elettromagnetiche e adroniche, seppur preliminari, permettono di identificare correttamente il 90% delle cascate elettromagnetiche, ottenendo falsi positivi solo nello 0.8% dei casi.

Un ultimo aspetto da evidenziare è che la libreria TensorFlow è utilizzabile proficuamente se vi è la disponibilità di unità di calcolo parallelo, come le GPU.

Attualmente, nel data center ReCaS-Bari è in corso una sperimentazione finalizzata ad un utilizzo automatico e trasparente di TensorFlow sulla distribuzione CentOS di Linux. Per la sperimentazione viene impiegato un server del cluster HPC con GPU Tesla K40m, e risultati positivi sono stati ottenuti usando il software Docker Community Edition, e il wrapper *nvidia-docker*, che permette la fruizione della GPU della macchina host dall'interno di un container, senza preoccuparsi di allineare le versioni dei driver. L'immagine Docker con un'installazione funzionante di TensorFlow è liberamente scaricabile da Docker Hub come `medphys/tensorflow`.

Gli utenti del cluster HPC ReCaS-Bari

Il data center ReCaS-Bari ospita un cluster HPC che fornisce risorse di calcolo parallelo ad alte prestazioni per numerose attività, di ricerca e non, fra cui:

- analisi di dati derivanti da esperimenti di Next Generation Sequencing, di genomi e trascrittomi di pazienti tumorali
- analisi dei dati raccolti negli esperimenti di fisica delle alte energie dell'acceleratore LHC al CERN
- simulazioni di energia libera per studiare il meccanismo di trasporto nei carrier mitocondriali
- simulazioni Monte Carlo e di dinamica molecolare per l'analisi del diagramma di fase di modelli statistici per particelle browniane attive
- simulazioni di fluidi applicate allo studio di cristalli liquidi e alla separazione di fase di fluidi con più componenti
- ricerca e utilizzo operativo di modelli numerici per previsioni meteorologiche

Nel mese di dicembre 2016 è stato chiesto agli utenti del cluster HPC di rispondere ad un questionario di valutazione del servizio. I risultati sono stati lusinghieri, e ne riportiamo alcuni estratti nel seguito.

“La mia esperienza col cluster è molto positiva e la mia attività ne ha tratto forte giovamento. I problemi incontrati sono stati quasi sempre risolti tramite il supporto.”

“Il cluster è stato decisamente necessario e ci ha aiutato ad ottenere i risultati prefissati in tempi ragionevoli. Il supporto è sempre stato puntuale e preciso nelle risposte anche se si sono palesati pochi casi in cui ne abbiamo richiesto l'intervento, in quanto il sistema è molto stabile ed affidabile.”

“Il cluster HPC Recas-Bari ha fortemente contribuito alle ricerche scientifiche sviluppate, permettendo una realizzazione efficiente delle simulazioni necessarie”

“Il servizio è ottimo, ben gestito, e di valido supporto alle nostre attività di ricerca”

L'indagine ha portato alla luce anche alcune criticità del servizio offerto, come la necessità di acquisire un compilatore Intel aggiornato con le rispettive librerie, o saltuari problemi di freezing in lettura o in esecuzione di job, e anche utili suggerimenti, come un potenziamento del sistema di scheduling, che includa la possibilità di assegnazione di un maggior numero di risorse nel caso in cui non siano utilizzate da altri utenti.

Corso su *Parallel Computing and GPU Programming using CUDA*

Come annunciato nella precedente newsletter, nell'ambito del XXXII ciclo della Scuola di Dottorato del Dipartimento Interateneo di Fisica "M. Merlin" si è tenuto dal 18 al 21 aprile un corso dal titolo *Introduction to Parallel Computing and GPU Programming using CUDA* tenuto dal Dr. Felice Pantaleo (CERN).

Il corso ha affrontato argomenti riguardanti la programmazione parallela attraverso l'utilizzo delle schede grafiche e del linguaggio CUDA e ha riscosso un notevole interesse da parte dei partecipanti (oltre che un elevato numero di richieste).

I discenti hanno utilizzato le risorse del cluster HPC di ReCaS-Bari per le esercitazioni e gli *hands on* proposti.

Il corso è stato valutato tramite un questionario sottoposto agli studenti al termine delle lezioni.

Dai risultati è emerso che una gran maggioranza dei discenti aveva poca o nessuna esperienza con sistemi di calcolo parallelo, ma mostrava un grande interesse per l'argomento. La platea era eterogenea e includeva persone che si occupano di fisica teorica e delle alte energie, previsioni del tempo, deep learning, online data selection, fluidodinamica, dinamica molecolare, analisi di immagini mediche, modellistica numerica per l'ingegneria costiera, data mining, bioinformatica, metodi di raccomandazione, computer vision.

Il corso è stato valutato complessivamente come molto utile e ben organizzato, così come l'infrastruttura ReCaS-Bari utilizzata, e in molti hanno manifestato il desiderio di maggiori approfondimenti, soprattutto nella parte pratica.

"un nuovo inizio ed un nuovo avvincente approccio alla programmazione"

"Il corso è stato molto bello. Sono state affrontate le basi del linguaggio con spunti per riuscire ad approfondire autonomamente. Inoltre, la parte teorica è stata ben organizzata con esempi di difficoltà crescente."

"Di utilità indiscussa come incipit a coloro i quali vorrebbero affacciarsi su questa realtà"

ReCaS-Bari ad EGI Conference 2017 e INDIGO Summit 2017

Dal 9 al 12 maggio 2017 si è svolta a Catania l'annuale conferenza di EGI (European Grid Infrastructure), collegata quest'anno all'INDIGO Summit 2017. INDIGO-DataCloud è un progetto HORIZON 2020 che si sta occupando dello sviluppo di una piattaforma cloud

aperta per il calcolo e dati ("DataCloud"), abilitante per le comunità scientifiche multidisciplinari (biologia strutturale, scienze della terra, fisica, bioinformatica, patrimonio culturale, astrofisica, scienze della vita, climatologia) su infrastruttura cloud ibrida. Il Summit è servito per illustrare alla comunità di utenti le soluzioni offerte attraverso la seconda "release" del software INDIGO denominata *ElectricIndigo*.

La comunità di ReCaS-Bari ha partecipato all'evento presentando quattro poster (uno dei quali, dedicato a descrivere il layer PaaS sviluppato nel contesto del progetto INDIGO-DataCloud, è stato votato come il secondo miglior poster presentato alla conferenza).

il dott. Giacinto Donvito della Sezione di Bari dell'INFN, direttore tecnico del progetto, ha commentato positivamente l'ottima presenza al Summit, non solo da parte di persone afferenti al progetto, ma anche di ricercatori e esperti esterni. Questo ha dato una buona risonanza per diffondere i risultati del progetto.

Secondo Corso di formazione: "Uso del software INDIGO-DataCloud in ambito scientifico"

Dal 19 al 21 Giugno a Bari si terrà il Corso di formazione sul software sviluppato dal progetto INDIGO-DataCloud, per utenti e sviluppatori in ambito scientifico.

Il corso, organizzato da ReCaS-Bari, darà la possibilità ai ricercatori di varie discipline (fisica e di altre scienze) di utilizzare le soluzioni sviluppate nell'ambito del progetto INDIGO per implementare gli use-case degli esperimenti di cui fanno parte.

Verranno mostrati diversi esempi di successo dell'uso di INDIGO nelle comunità già supportate, insieme ai dettagli tecnici su come implementare nuovi casi d'uso.

Alla scuola parteciperà (come docente) personale di varie sedi INFN direttamente coinvolto nelle attività di sviluppo del progetto.

Ulteriori dettagli possono essere reperiti al link:

<https://agenda.infn.it/conferenceDisplay.py?confId=13574>

Crediti

Hanno collaborato a questo numero della newsletter:

Roberto Bellotti, Domenico Diacono, Giacinto Donvito, Fabio Gargano, Giorgio Maggi, Salvatore My, Stefano Nicotri, Alexis Pompili, Giacomo Volpe.