



In questo numero:

- Trasferimento dei servizi Batch dalla farm Bc<sup>2</sup>S al nuovo datacenter ReCaS
- Testimonianze degli Utenti: La Genomica Mitocondriale Umana nell'Università di Bari
- Testimonianze degli Utenti: Pipeline per l'analisi massiva di dati genomici e metagenomici prodotti attraverso piattaforme di High Throughput Sequencing.
- Concluso il "Corso intensivo di programmazione di schede grafiche utilizzando CUDA"
- Crediti

### **Trasferimento dei servizi batch dalla farm Bc<sup>2</sup>S al nuovo datacenter ReCaS**

Come annunciato nella Newsletter di aprile, proseguono le azioni per trasferire le attività e i servizi batch dalla farm Bc<sup>2</sup>S al nuovo datacenter ReCaS-Bari.



Passando alla nuova farm, gli utenti troveranno alcuni cambiamenti che contriguiranno a migliorare stabilità e prestazioni:

- È nuovo il sistema di gestione delle code batch: si è passati da Torque+Maui a

HTCondor. Agli utenti è richiesto di familiarizzare con questo nuovo prodotto, imparare i comandi per mandare i job in esecuzione e modificare gli script di sottomissione dei JOB. Un manuale utente che descrive i comandi di HTCondor usati più di frequente è reperibile a questo [link](#).

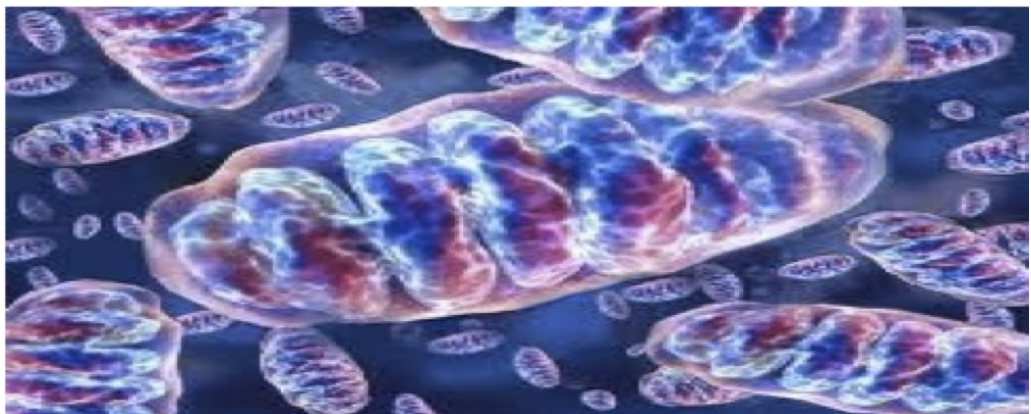
- Il filesystem Lustre è stato rimpiazzato da nuovi prodotti: l'esperimento ALICE usa XrootD nativo, uniformandosi al resto dei suoi siti; tutti gli altri utenti utilizzano GPFS su cui vengono memorizzate le cartelle home, con doppia copia online. Nel copiare i file da una farm all'altra si è cercato di mantenere gli stessi alberi di directory: le radici delle directory sono rimaste /lustre e /lustrehome anche se sulla nuova farm i file sono gestiti da GPFS.

Per completare la migrazione è necessaria una grande collaborazione da parte degli utenti. Un incontro degli Utenti con il team ReCaS il giorno 25 Maggio alle 15:00 nella Aula ex sala Consiglio, durante il quale sono stati illustrati gli aspetti finali della procedura. Le slide che sono state presentate da Giacinto Donvito in occasione dell'incontro con gli utenti sono reperibili a questo [link](#).

È importante precisare che la migrazione interessa soltanto gli utenti che utilizzano il datacenter ReCaS sottomettendo job al batch system, mentre gli utenti che usufruiscono di risorse cloud e/o quelli che utilizzano il cluster HPC potranno continuare la loro attività senza alcuna interferenza.

## Testimonianze degli Utenti: La Genomica Mitochondriale Umana nell'Università di Bari

### Mitochondri



Il laboratorio di Bioinformatica attivo nel Dipartimento di Bioscienze, Biotecnologie e Biofarmaceutica dell'Università di Bari, coordinato dalla Professoressa Marcella Attimonelli, svolge attività di ricerca nell'ambito della genomica mitocondriale umana. I mitocondri costituiscono la centrale energetica della cellula e sono coinvolti in un gran numero di processi cellulari in condizioni fisiologiche e/o patologiche. Da qui l'interesse scientifico per il ruolo del DNA mitocondriale (mtDNA), in vari campi della biologia e medicina umana. Tale interesse si è tradotto in una grande quantità di dati di sequenza mitocondriali prodotti con ritmi sempre crescenti grazie all'avvento delle tecnologie *Next Generation Sequencing*

(NGS), che richiedono soluzioni bioinformatiche adeguate per una corretta e successiva interpretazione dei dati. La banca dati *HmtDB* (*Human mitochondrial DataBase*, <http://www.hmtdb.uniba.it/>), che raccoglie più di 30000 sequenze di mtDNA e relative annotazioni, realizzato e gestito dal laboratorio Attimonelli a sostegno di studi sia in campo medico sia di genetica popolazionale è ospitata dal data center ReCaS-Bari.

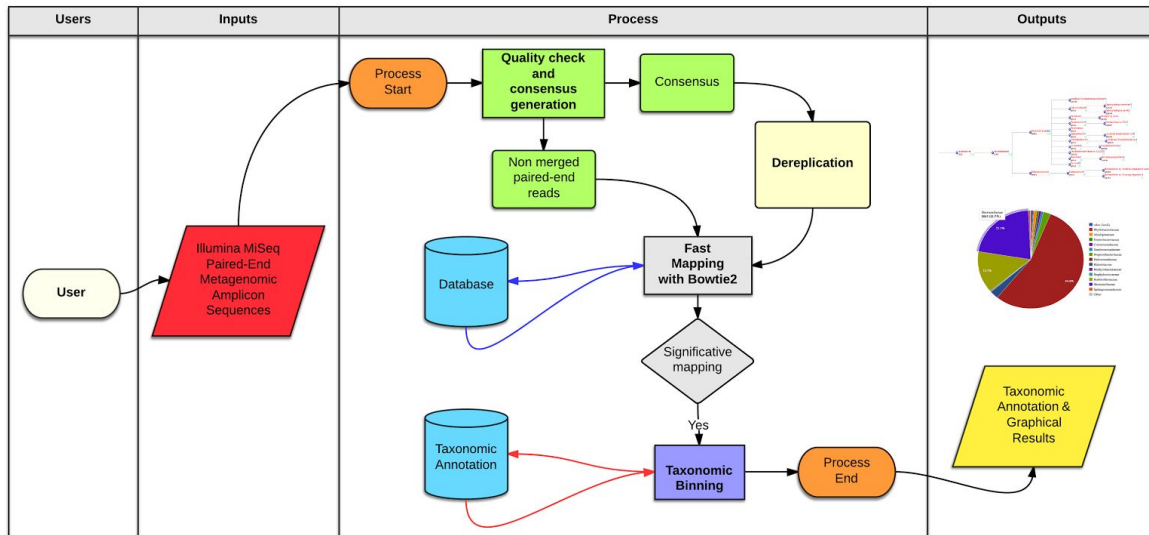
Il progetto ReCaS, con la realizzazione di una infrastruttura di calcolo e storage, ha reso disponibile al laboratorio di Bioinformatica ciò di cui aveva bisogno. Tale soluzione ha consentito l'ottimizzazione della gestione di *HmtDB*, intervenendo in particolar modo sulla riduzione dei tempi di aggiornamento. Inoltre, il lavoro di calcolo per lo sviluppo e la manutenzione di *MToolBox*, *pipeline* per l'analisi delle varianti mitocondriali umane identificate da sequenziamento NGS e/o classico (sviluppato nello stesso laboratorio, PubMed PMID: 25028726), è usualmente eseguito sulle risorse IT messe a disposizione da ReCaS.

Infine, i dottorandi e i tesisti del laboratorio possono accedere alle piattaforme cloud e cluster HPC di ReCaS per sottomettere al batch system analisi di dati genomici NGS (in particolare da Whole Exome e Whole Genome Sequencing), usufruendo dei principali applicativi, tool e software disponibili sulla farm di calcolo (ad es. R Studio, Python, Perl, Java, Bash, ecc.) e lo storage di grandi quantità di dati. I primi risultati conseguiti dal gruppo di Bioinformatica usufruendo dei servizi offerti dal data center ReCaS-Bari, sono già stati resi disponibili alla comunità scientifica internazionale (PubMed PMID: 25668474, 25028726, 26621530).

### **Testimonianze degli Utenti: Pipeline per l'analisi massiva di dati genomici e metagenomici prodotti attraverso piattaforme di High Throughput Sequencing.**

Sostanziali innovazioni nel campo della salute umana, della tutela dell'ambiente, e delle applicazioni industriali sono state ottenute negli ultimi anni grazie a diversi approcci metodologici di biologia molecolare tra i quali citiamo qui la Genomica e la Metagenomica. Quella che oggi sta assumendo i connotati di una vera e propria rivoluzione scientifica ha visto come protagoniste le piattaforme di sequenziamento massivo del DNA (*HTS*, High Throughput Sequencing) che hanno consentito un accesso senza precedenti alla conoscenza di singoli organismi e di intere comunità fino ad oggi soltanto in minima parte indagabili.

L'analisi accurata dell'enorme quantità di dati prodotta negli esperimenti di HTS richiede lo sviluppo di sistemi di analisi altamente performanti. Tra di essi vi sono MSA-PAD (Multiple Sequence Alignment Based on PFAM Accessed Domains) [Balech et al. 2015] e BioMaS (Bioinformatic analysis of Metagenomic ampliconS) [Fosso et al. 2015], due pipeline sviluppate dall'Istituto di Biomembrane e Bioenergetica (IBBE) del CNR di Bari in collaborazione con il Dipartimento di Bioscienze, Biotecnologie e Biofarmaceutica dell'Università degli Studi di Bari "Aldo Moro" e con il supporto computazionale dell'Istituto Nazionale di Fisica Nucleare (INFN) di Bari. Tali applicazioni bioinformatiche consentono, rispettivamente, di comparare sequenze genomiche multiple e di identificare la quasi totalità dei microrganismi viventi in complesse comunità procariotiche e fungine. L'esecuzione di tali strumenti di indagine è estremamente impegnativa, sia in termini della mole di dati che fluiscono lungo gli interi workflow sia per il livello di accuratezza richiesto ad ogni step dell'analisi.

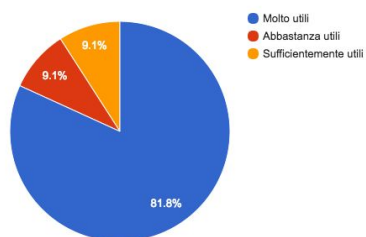


In questo scenario il data center ReCaS-Bari, nel quale i due strumenti sono attualmente ospitati, ha fornito l'infrastruttura ottimale, offrendo non solo una potenza di calcolo (anche parallelo) adeguata ma anche la possibilità di un utilizzo contemporaneo da parte di diversi utenti. Inoltre, grazie all'interfaccia *user friendly* di ReCaS, BioMaS e MSA-PAD possono essere agevolmente e diffusamente utilizzati anche da ricercatori non esperti in bioinformatica. BioMaS è stato già applicato con successo all'analisi di comunità batteriche residenti in campioni ambientali, agroalimentari e animali, questi ultimi prelevati da modelli murini di patologie umane (tumori e patologie autoimmuni), mentre MSA-PAD è stato applicato in studi di filogenesi di organismi.

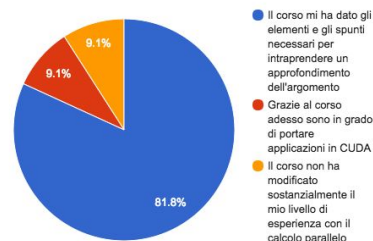
## Concluso il “Corso intensivo di programmazione di schede grafiche utilizzando CUDA”

Il corso, organizzato nell'ambito delle attività connesse al data center ReCaS con il supporto della Sezione INFN di Bari e del Dipartimento Interateneo di Fisica di Bari, si è tenuto nei giorni 11, 12 e 13 Maggio 2016, nell'Aula Multimediale del Dipartimento Interateneo di Fisica di Bari. Le lezioni sono state tenute dal Dr. Felice Pantaleo (CERN - Universität Hamburg).

Come valuteresti l'utilità degli argomenti trattati?



Quale ritieni che sia il livello da te raggiunto sull'argomento grazie al corso?



Il corso è stato accolto positivamente dai discenti, come si può evincere dai risultati del

questionario sul livello di soddisfazione sottoposto loro alla fine del corso, illustrati dai grafici a torta mostrati in figura.

Considerato il numero di richieste pervenute, si sta pensando di riproporre il corso in autunno, con un formato che sarà definito anche sulla base dei risultati ottenuti nella presente edizione.

Chi fosse interessato alle tematiche trattate può reperire il materiale delle lezioni al seguente [link](#).

### **Crediti**

Hanno collaborato a questo numero:

Marcella Attimonelli, Giacinto Donvito, Giorgio Maggi, Stefano Nicotri, Monica Santamaria.