

Chi è BioDec

Nata nel 2002 come *spin-off* di un gruppo accademico, con la missione di fornire software e consulenza in **bioinformatica**, dal 2006 l'offerta dei servizi si è ampliata nell'ambito dell'*Information Technology (IT)*, poiché le tecnologie sviluppate e le competenze acquisite si sono rivelate proficue oltre la bioinformatica e applicabili per risolvere molti problemi dell'IT moderno.

La componente di servizi di consulenza IT è successivamente cresciuta negli anni e ad oggi il supporto e lo sviluppo di infrastrutture IT sono tra le attività principali dell'azienda: esse vengono portate avanti utilizzando quell'insieme di tecniche e di pratiche che vanno sotto il nome di **DevOps**.

Fin dalla fondazione, BioDec ha scelto di usare **esclusivamente software libero e standard aperti**.

I servizi BIODEC per le infrastrutture IT

Automazione di server e servizi

Obiettivi: *Automatizzare le interazioni ripetitive con i sistemi software.* La disponibilità sempre maggiore di server e di servizi ha reso le infrastrutture IT sempre più grandi, più distribuite e quindi sempre più complesse. Molte attività che potevano svolgersi manualmente, per la loro sporadicità ed eccezionalità, sono diventate sempre più frequenti e sempre più normali. Si pensi solo all'esempio della creazione di un server, partendo dall'hardware: coi moderni sistemi di virtualizzazione, quella che era un'attività rara e che poteva durare una giornata è diventata un'operazione di routine quotidiana, e la spinta delle nuove architetture basate su container accelererà questa tendenza.

Perché: Per impiegare meno tempo nello svolgimento di compiti ripetitivi, per lasciare più tempo ed energie per i compiti creativi, per eliminare gli errori.

Come: Scrivendo dei programmi in linguaggi di programmazione di alto livello come Python, con librerie apposite o direttamente in *domain specific languages (DSLs)* pensati per coordinare più calcolatori contemporaneamente (Puppet, Ansible, CFEngine3). Utilizzando dei software che continuamente ripercorrono le azioni necessarie ad un'attività (ad esempio usando Jenkins o Gitlab per automatizzare il processo di *test*, di *build* e infine di *deployment* di un software). Adottando soluzioni di orchestrazione di container come Kubernetes.

Progettazione di infrastrutture informatiche

Obiettivi: Progettare e realizzare infrastrutture con funzionalità complesse, dove si richiedano alte prestazioni di calcolo (HPC, *high performance computing*), o la gestione di grandi moli di dati, o la

gestione di cluster di server virtuali (ad esempio per realizzare *private cloud* o sistemi basati su *container*) o infine si richiedano alte soglie di continuità di servizio (sistemi in HA, *high availability*).

Perché: L'infrastruttura informatica di un'azienda deve rendere al massimo, senza tempi morti, e senza sprechi, poiché essa – in un'impresa moderna che sfrutti al massimo le possibilità offerte dall'automazione dei processi – è l'unico tramite per erogare beni e servizi.

Come: Utilizzando sistemi basati su software libero e standard aperti: GNU/Linux (nelle distribuzioni Debian, Ubuntu e CentOS) come sistema operativo, software per la realizzazione di infrastrutture di virtualizzazione (Ganeti, KVM, Proxmox), software per la gestione di grandi moli di dati (GlusterFS, Samba, NFS e ZFS) e software per la creazione delle necessarie componenti intermedie (OpenVPN, Openswitch, OGS e altro).

Le soluzioni basate sulla necessità di *database transazionali* ad alte prestazioni sono realizzate con PostgreSQL, che BioDec utilizza fin dalle prime versioni, e sui sono stati erogate attività di formazione in importanti realtà italiane legate al mondo della finanza e bancario.

Manutenzione e monitoraggio delle infrastrutture

Obiettivi: Gestire e mantenere le infrastrutture basate su GNU/Linux e Unix, integrando le applicazioni, i servizi, i server e la rete, ed effettuando le necessarie attività di monitoraggio e telecontrollo, di gestione degli allarmi e delle notifiche.

Perché: Per garantire la continuità dei servizi e per reagire ai cambiamenti necessari, imposti dall'accadere di errori, di guasti, di imprevisti e di cambi di contesto (ad esempio, per gestire l'evoluzione di un sistema di e-commerce al crescere del volume di affari) è necessario che ogni componente dell'infrastruttura sia gestita.

Come: Usiamo Nagios, Icinga2, CheckMK e Shinken come sistemi di allarmistica, e per fornire *dashboard* di alto livello. Collectd, Statsd e simili per raccogliere ogni tipo di metrica necessaria, e *time-series database* come Influxdb e Prometheus per la loro persistenza o sistemi come Grafana per la visualizzazione. Graylog2, Logstash ed Elasticsearch sono la nostra scelta per realizzare i sistemi centralizzati di analisi dei log e Kibana come interfaccia di ricerca. Abbiamo inoltre un'ampia esperienza nell'uso di sistemi di versionamento distribuiti e non, per gestire le configurazioni di sistema (Etckeeper, Git, Mercurial, Subversion) e sistemi server come Gitlab per quanto riguarda le architetture di *continuous integration* e *continuous delivery*.

Agile IT

Obiettivi: Realizzare soluzioni per la gestione agile di progetti, e per l'integrazione con applicazioni *business*. Svolgere formazione e consulenza sul tema dell'*Agile IT*.

Perché: L'approccio tradizionale, che prevede una separazione funzionale tra sviluppatori, amministratori (di sistema, di rete, di database) e addetti all'operatività non genera le economie di scala desiderate, ed è fonte di problemi di comunicazione e di coordinamento che portano a

progetti fuori budget, non allineati con le necessità dell'impresa, sovente in cronico ritardo di esecuzione e spesso di scarsa qualità. Esistono tecniche e strumenti mutuati dalle *Agile Methodologies* che possono offrire una soluzione a molti di questi problemi e che vanno sotto il nome di DevOps.

Come: Offriamo consulenza sugli aspetti metodologici e di processo, ed integriamo nel processo produttivo aziendale strumenti flessibili di gestione delle attività come Redmine e Gitlab, e strumenti di *business report* come Jasper. Svolgiamo attività di formazione, di *mentoring* e *coaching* sui progetti per conto del cliente. Con l'iniziativa 4DevOps (<http://www.4devops.it/>) siamo in grado di coprire le necessità del cliente andando a pescare da un *pool* di competenze ampio e trasversale.

Bioinformatica

Obiettivi: Progettare e realizzare le piattaforme di *large data management* e di analisi computazionale per la bioinformatica.

Perché: Le moderne tecniche di *Next Generation Sequencing (NGS)* permettono di accumulare una grande quantità di dati genomici, proteomici e genetico-medici con facilità e costi contenuti: questa mole di dati permette di svolgere analisi e simulazioni altrimenti impossibili, e quindi avere dei risultati e delle indicazioni funzionali alla ricerca farmaceutica, alla ricerca di vaccini ed alla diagnostica medica in generale.

Come: Usando la **piattaforma di calcolo bioinformatico BioDec (BIP)**. I programmi della BioDec BIP sono fruiti in diversi modi (nativamente sui server fisici, o come macchine virtuali, o come *container*) al fine di formare una piattaforma di servizi che eroghi le seguenti funzionalità:

- un ambiente di virtualizzazione,
- uno *storage* distribuito iperconvergente ad alte prestazioni e ad alta affidabilità,
- un servizio di *identity management*,
- i servizi di code di calcolo sui nodi di computazione,
- l'installazione delle principali applicazioni di calcolo bioinformatico e NGS,
- un servizio di analisi genomica per la gestione dei dati in formato BAM e VCF,
- un'applicazione per svolgere analisi computazionali bioinformatiche tramite un'interfaccia *web*,
- un sistema di *backup*,
- un servizio di archiviazione,
- un'ambiente per il versionamento del codice di calcolo,
- i servizi di *continuous integration* e *continuous deployment* (CI e CD),
- un servizio di *issue tracking*,

- un servizio di monitoraggio (*log management*, allarmistica, metriche),
- i servizi di rete infrastrutturali (*VPN, proxy, firewall*),
- i servizi di automazione,
- una *meta-package manager* proprietario per l'archiviazione e la riproduzione delle applicazioni di calcolo bioinformatico e NGS, denominato **LTSAR system** (*Long-Term Software Archive and Reproduction System*): in particolare quest ultima applicazione BioDec consente di riprodurre a distanza anche di anni le medesime pipeline software che potevano essere in produzione anni addietro, riferendo pacchetti o librerie che potrebbero anche non essere più disponibili su Internet,
- una piattaforma di *container* Linux da utilizzare come servizio (*PAAS*).

In particolare, la piattaforma di calcolo BioDec BIP mira a realizzare **un ambiente contenuto ed autonomo per tutte le attività di un gruppo di ricercatori** che debbano sia eseguire delle *pipeline* di analisi che gestire successivamente grandi moli di dati.

La novità della piattaforma BioDec BIP risiede nell'**integrazione di numerose componenti che coprono l'intero stack tecnologico**, partendo dalle applicazioni installate sui server fisici (ad esempio per gestire gli ambienti di virtualizzazione di storage) e arrivando agli ambienti nelle macchine virtuali in cui si trovano i pacchetti utilizzati dai ricercatori, fino alla gestione del codice, del ciclo di vita del medesimo e di tutte gli stadi dei dati (dati caldi, freddi, archiviati, eccetera). In particolare, l'integrazione di *best practices* tipiche del mondo dello sviluppo software, come il versionamento del codice e i sistemi di Continuous Integration e di Continuous Delivery, permette di creare degli artefatti e degli ambienti di calcolo validati, i quali consentono di raggiungere obiettivi come:

1. la **ripetibilità delle analisi**,
2. la **verifica della qualità** dei risultati ottenuti,
3. il **controllo nel tempo**, anche a distanza di decenni, della qualità di quanto prodotto - in particolare questa funzione è abilitata dall'uso dello *LTSAR system* come sistema di archiviazione del codice, ed è una funzione unica della BIP,
4. il **riesame di dati vecchi con nuovi metodi di calcolo**,
5. **tempi di sviluppo e di messa in opera estremamente rapidi**, anche per *pipeline* di calcolo di centinaia di componenti.

Principali soluzioni realizzate per i nostri clienti

Cluster di calcolo per NGS e bioinformatica

BioDec ha realizzato un cluster di calcolo per un importante ospedale nazionale, per un totale di **384 core di calcolo, 2TB di RAM complessiva e 720TB di storage raw**. Il risparmio ottenuto dall'adottare una soluzione cosiddetta di **iperconvergenza basata su software libero** ha permesso di ottenere **un risparmio di più del 50% rispetto a soluzioni commerciali basate su approcci tradizionali** (separazione fra una SAN commerciale e nodi di calcolo).

Un altro cluster di minori dimensioni è stato realizzato per un'azienda tedesca di bioinformatica che ha come clienti alcune tra le maggiori multinazionali farmaceutiche: in questo caso al cluster di calcolo è stato associato anche **un sistema di archiviazione e cifratura** progettato per mantenere **uno storico di quindici anni**.

Inoltre, la soluzione BioDec permette una **continua evoluzione** del cluster di calcolo e di storage, poiché è possibile sostituire i singoli server, man mano che l'obsolescenza dell'*hardware* lo richieda, senza costose migrazioni e sostituzioni periodiche di ampie parti dell'architettura.

Continuous deployment / delivery di servizi gestionali in cloud

BioDec ha realizzato il codice di automazione che permette ad un proprio cliente (azienda *leader* a livello mondiale) di gestire la **creazione ex-novo di un sito di produzione in cloud sul quale viene erogato as a service il sistema gestionale** del cliente. Il software prodotto, in un misto di Ansible, Python e tecnologie proprietarie AWS quali Cloudformation, consente di definire i parametri operativi del sito di produzione ed eseguire la creazione del medesimo su cloud.

Il sistema permette di realizzare un ambiente che contenga anche alcune **decine di server diversi**, con *high availability*, configurazione dei server e dei servizi, agganci ai database, ai servizi di rete, le politiche di sicurezza eccetera. Successivamente, sempre lo stesso software di automazione, permette di gestire gli aggiornamenti, le modifiche all'infrastruttura e ogni altra variazione operativa.

In seguito a questo progetto il tempo di *delivery* di un sito operativo è passato (*fonte: dati interni del cliente*) **da diverse settimane ad alcune ore**.

Rete di telecontrollo per impianti fotovoltaici

È stata realizzata una soluzione software, basata su un sistema di VPN e di PC industriali con GNU/Linux. Il sistema permette il consolidamento **dei dati di funzionamento degli impianti**, in modo sicuro, utilizzando normali connessioni a Internet, anche di bassa qualità (*i.e.* GPRS). La soluzione è installata in **centinaia di impianti, in tutto il mondo**, e garantisce la **continuità del servizio di telecontrollo e monitoraggio 24x7** – il cliente è un'azienda italiana, leader nel campo degli apparati elettrici per gli impianti di potenza.

Automazione di processo – azienda grafica

È stata realizzata una soluzione software, basata su software *custom* sviluppato in Python per **gestire automaticamente il flusso di lavoro relativo alla lavorazione delle immagini grafiche**, rimpiazzando lavorazioni manuali estremamente ripetitive e onerose, e che sarebbero rapidamente diventate ingestibili all'ulteriore crescere del numero di immagini gestite (l'ordine di grandezza è di **migliaia di fotografie al giorno**). Il cliente è un'azienda italiana, che opera nel campo della grafica e della fotografia per il mercato della moda (tradizionale e digitale) e che ha come propri clienti i principali *e-commerce* di moda.

Ingegneria di prodotto per la teleassistenza e la sicurezza

È stata realizzata una soluzione hardware e software, basata su un sistema di VPN e di PC *commodity* con GNU/Linux. Il sistema permette di offrire un **servizio di sicurezza informatica gestita e di teleassistenza remota 24x7** in modo sicuro, utilizzando normali connessioni ad Internet. La soluzione è stata installata in **centinaia di centri, in tutto il mondo**, e ha garantito la continuità del servizio di teleassistenza, la gestione degli aggiornamenti software, antivirus, eccetera, **dal 2005** – il cliente è un'azienda multinazionale, leader mondiale nel campo delle soluzioni per i bowling.

Gestione di server fisici e virtuali conto terzi

BioDec ha costruito un'infrastruttura per la gestione e la manutenzione dei server GNU/Linux e Unix, dei propri clienti, che permette le attività di monitoraggio e telecontrollo, di gestione degli allarmi e delle notifiche, sia in modo centralizzato, che lasciando ai clienti BioDec la possibilità di azione autonoma.

Ad oggi il nostro sistema contiene – solo contando i server propriamente detti – **più di 600 host (tra macchine fisiche e virtuali) e circa 15000 servizi controllati**. La crescita dell'infrastruttura gestita da BioDec è cresciuta di circa il **30%** all'anno negli ultimi anni. Considerando anche le altre attività svolte da BioDec il **rapporto fra addetti e server è di circa uno a trecento**, ben al di sotto della media del settore.

Contatti

Internet

- <http://www.biodec.com/>
- info@biodec.com

Fonia

- Telefono (+39)-051-0548263
- Fax (+39)-051-7459582

Indirizzo

- BioDec S.r.l. Via Calzavecchio 20/2, I-40033, Casalecchio di Reno (BO), Italia.