



Exploring machine learning approaches for predictive maintenance and log analysis at CNAF

Introduzione

Barbara Martelli

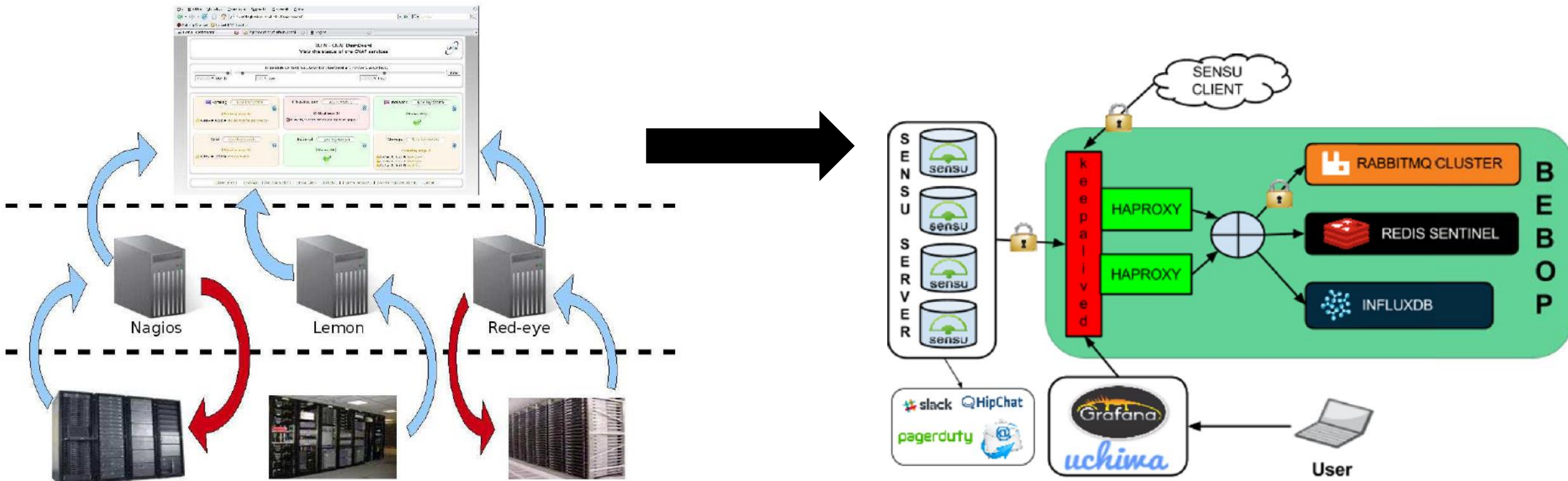
Seminar Outline

- This introduction
 - CNAF monitoring and logging system (Tier1)
 - CNAFR&D activities in the field of Machine Learning (SDDS)
 - Possible technology transfer activities in the field of ML (PETT)
 - Overview of summer students work plan
- Luca Giommi: “Towards predictive maintenance at the INFN-CNAF computing centre” – Tutor Elisabetta Ronchieri
- Simone Rossi Tisbeni: “*StoRM metrics and logs parsing and retrieval*” – Tutor Barbara Martelli
- Tommaso Diotallevi: “ELK stack & log parsing” – Tutor Diego Michelotto e Antonio Falabella

CNAF monitoring and logging system

Evolution of monitoring system at INFN-CNAF

- Antonelli et al, “INFN-CNAF Monitor and Control System”, CHEP 2010
- Bovina S., Michelotto D. “The evolution of monitoring system: the INFN-CNAF case study”, CHEP 2017
- Dal Pra, Falabella, Fattibene et al “Evolution of monitoring, accounting and alerting services at INFN-CNAF Tier1”, CHEP 2018



Tier 1 Goals

- Raccolta e correlazione log file e metriche del Tier-1 al fine di facilitare debug problematiche servizi e (possibilmente) effettuare analisi predittiva tramite tecnologie di machine learning.
- Obiettivo contingente
 - Ottimizzazione manpower anche in prospettiva del big bang del tecnopolo
 - Liberare sperabilmente personale per sviluppo
- A lungo termine
 - Creare un centro di competenza su ML
 - attrazione laureandi/dottorandi/AR con possibile percorso di uscita verso realtà produttive
- Auspicabile progetto comune/sinergia/convergenza con iniziative metodologicamente analoghe
 - Interne...
 -ed esterne al CNAF (es. proposal «Gemini2»)



Cosa serve

- **Raccolta dati**

- Predisposizione ambiente per memorizzazione dati
 - Disegno infrastruttura ELK sulla base dei requisiti del centro
 - Implementazione filtri di base infrastruttura ELK
- Basandoci su esperienza CERN, servono ~10 TB di spazio disco (per 6 mesi di memorizzazione) e 3 server per filtro e prima elaborazione log
 - Storage in dismissione per primo pilot
- Pilot con raccolta log subset server, affinamento filtri sui log, prime dashboard e analisi
- Espansione infrastruttura ELK
- Seconda fase: valutare import dati da influxdb (da valutare porting totale a ELK per monitoring)

- **Analisi dati**

- TensorFlow?
- Eventuale uso di GPU K40 per primi test ML





INFN-CNAF Machine Learning R&D

(some examples, thanks to D. Salomoni)



DEEP-HybridDataCloud (<https://deep-hybrid-datacloud.eu>), focusing on intensive computing services exploiting specialized hardware components, such as GPUs, low-latency interconnects, and others resources



ATTRACT proposal submitted som sottomesso qualche settimana fa, basato sulla BondMachine (<http://bondmachine.fisica.unipg.it>) « is that its processors forms an heterogeneous set of computing units where each of them can be configured and specialized to do a specific task implementing different architectures based on the specific problem», che dovrebbe venire utilizzata come soluzione low-power, low-cost, customizable per eseguire workload TensorFlow ("TensorFlow on FPGA")

- Submitted a Proof of Concept proposal to Google in order to test TPU and GPU
- INDIGO-Next proposal will focus a relevant part of the project on AI an ML
- Perugia PhD School on Open Science Cloud 2018 is one of the INFN initiatives aimed at exploring Machine Learning

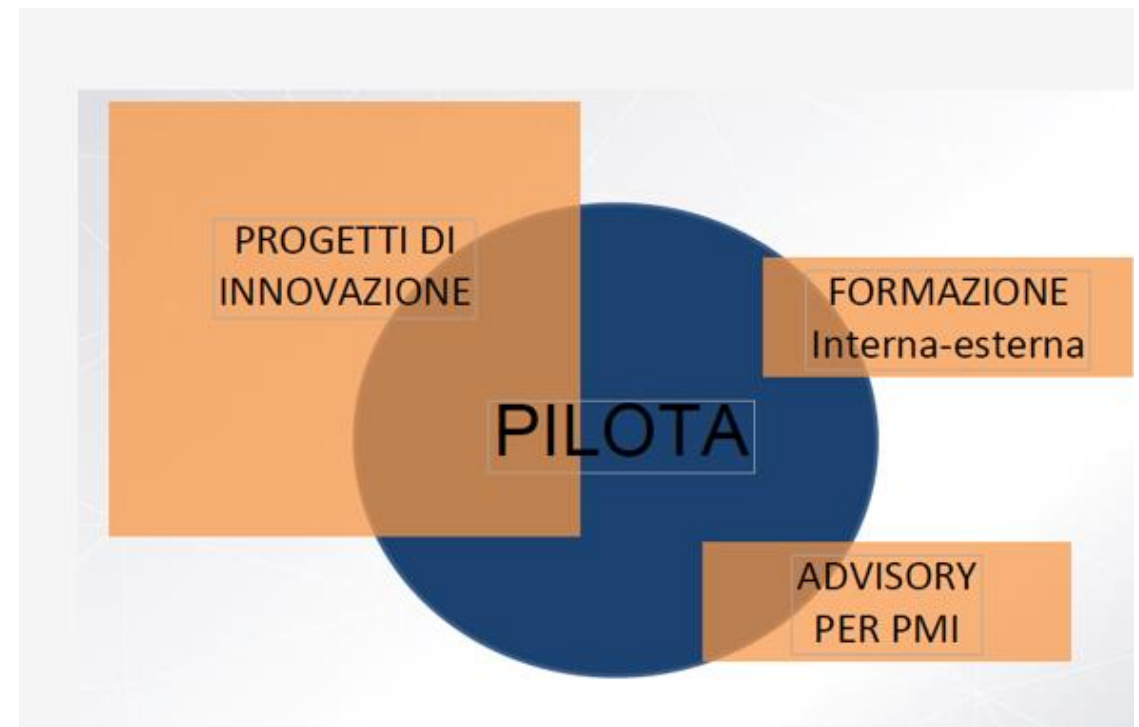


SOSC 2018 Second International PhD School Open Science Cloud

<https://agenda.infn.it/conferenceTimeTable.py?confId=15534#all.detailed>

Possible Technology Transfer of ML Competence Center BI-REX

- Public-Private Partnership aimed at founding an Industry 4.0 Competence Center in Bologna (MISE)
- **BI-REX (*Big Data Innovation & Research Excellence*)**
 - **49 companies**
 - **7 Universities and Research Institutes** (UNIBO, UNIMORE, UNIFE, UNIPR, UNICATT, CNR, INFN)
 - **5 other institutions** (ASTER, Bologna Business School, CINECA, Fondazione Golinelli, Istituto Ortopedico Rizzoli)
 - 20 Meuro total budget in 3 years





Technology Transfer

BI-REX Competence Center

An example of ML relevant project

Integrazione di Piattaforme per Gestione di Big Data da Processi Produttivi

- Integrazione di piattaforme IoT e big data processing per gestione di flussi di dati di grandi dimensioni da impianti produttivi e da sorgenti esterne non strutturate. Implementazione di algoritmi di analisi offline, capaci di generare modelli per esecuzione di monitoraggio e controllo online locale con event detection.
- TRL 6->8; Pilota

Diagnostica Predittiva basata su Tecniche di Data Analytics e Machine Learning

- Sviluppo di tecniche e piattaforme integrate per data analytics e machine learning a fini di interpretazione di dati affidabilistici. Monitoraggio, filtraggio e riconoscimento di eventi, anche tramite modelli ottimizzati offline lato cloud e poi eseguiti localmente.
- TRL 5->8; Pilota



Tech Transfer: Fellowship on Regional Funds and Harmony project

- Emilia Romagna regional project «Creazione di valore per imprese e societa' con la gestione e l'analisi dei big data»
 - Fellowship title: “L'analisi di dati nel campo della fisica, dell'astrofisica e della scienza nello spazio”
 - Among other things there is the topic: «strumenti di gestione ed algoritmi di analisi partendo dai dati raccolti dai sistemi di supervisione»
- Harmony alliance Big Data (analytics) to enable better and faster treatment for Patients with Hematologic Malignancies



HARMONY

Summer Students Work Plan (E. Ronchieri)

1. Identify a specific application to be analyzed: StoRM chosen
2. Log Parsing: StoRM frontend, StoRM backend, heartbeat, messages, GridFTP, GPFS (not covered in our study, but potentially interesting)
-> produce csv and json files
3. Data Enrichment: (enhance, refine or improve raw data). Data fusion, entity recognition, disambiguation, segmentation. i.e. add information like UTC timestamp
4. Add metrics data (from Tier1 InfluxDB)
5. Create a Dictionary and add a class (data categorization)
6. Apply Machine Learning techniques in order to extract insights



Let's give the floor to our summer students!

