Big Data Platform del CNAF

Enrico Fattibene

Workshop sul calcolo dell'INFN

Loano (SV) - 25 maggio 2023

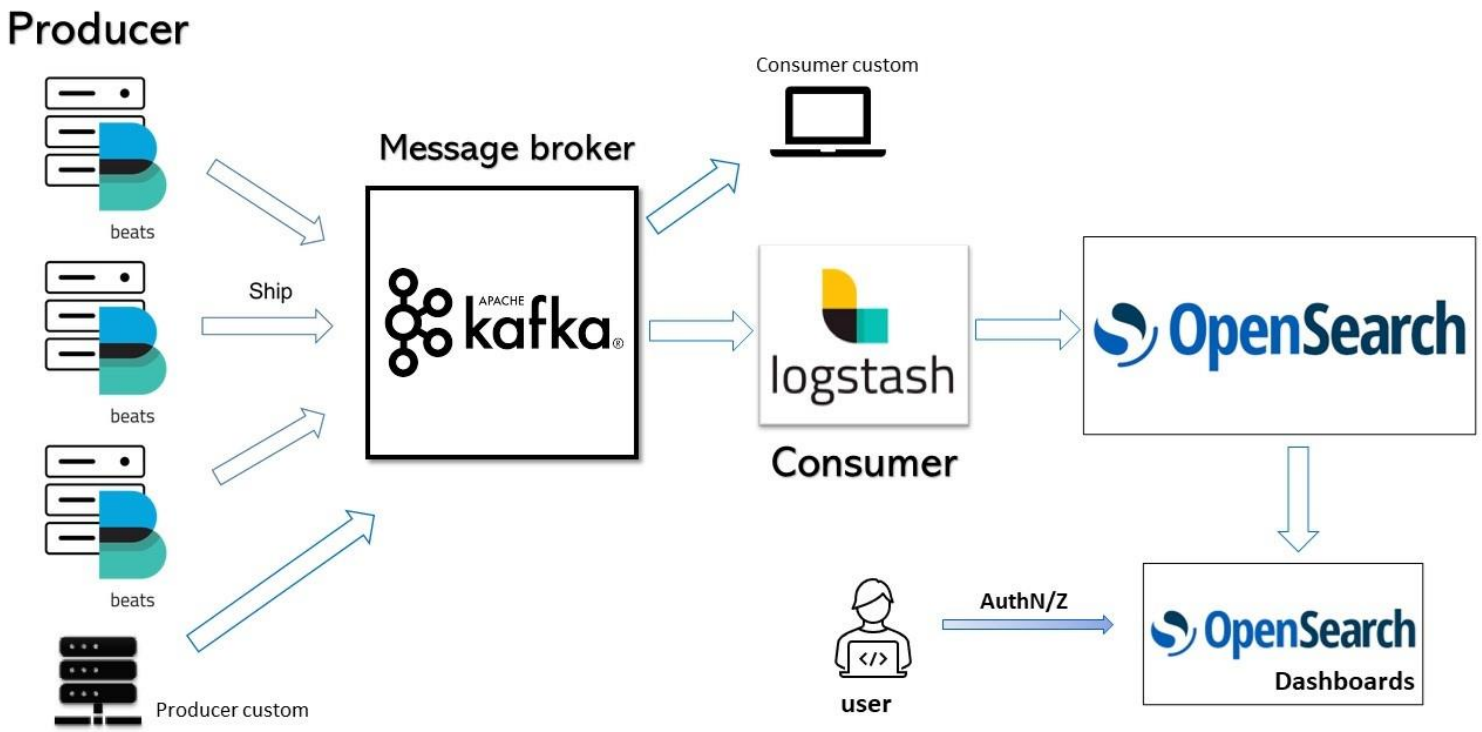
Sommario

- Motivazioni e scopo del lavoro
- Architettura e strumenti utilizzati
- Configurazione servizi
- Integrazione con servizi esistenti
- Casi d’uso

Motivazioni e scopo

- Infrastruttura **modulare** di gestione e analisi dati eterogenei a servizio di amministratori ed utenti del CNAF
  - Monitoraggio e troubleshooting attraverso analisi dei log
  - Anomaly detection
  - Analisi dati con algoritmi di Machine Learning
- Attività nell’ambito del progetto CNAF Reloaded
  - Task Operations

Schema semplificato

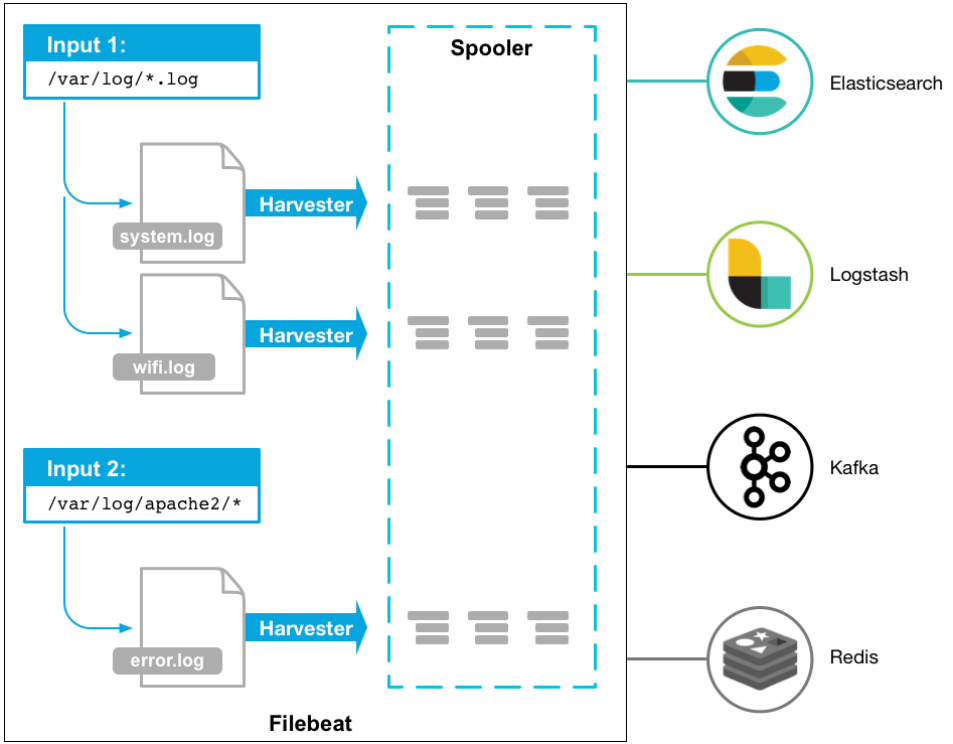


Hardware

- Tre nodi virtuali per **Kafka** (load balancing + HA) su infrastrutture separate
  - 1.8TB spazio totale
- Tre nodi virtuali per la gestione del **cluster OpenSearch**
- Quattro nodi fisici per i **dati di OpenSearch**
  - 8TB spazio totale su SSD
  - 16TB su Ceph per snapshot di dati offline
- Un nodo virtuale **OpenSearch Dashboard**
- Un nodo virtuale **Logstash**

Filebeat

- Producer specifico per spedire **log** (log shipper), nel nostro caso verso Kafka
- Configurazione semplice **via Puppet**
  - Output verso Kafka criptato via SSL
  - Autenticazione su Kafka via username e password
  - Ogni reparto/progetto scrive solo su topic concordato con amministratori
  - **Tag** usato per indicizzare su OpenSearch



Kafka

- Piattaforma open source per lo **streaming di eventi**
- Può essere usato come servizio a se
- **Topic** diversi per reparti/progetti diversi
  - Suddivisione logica dei dati
- Numero **partition**, **repliche** e **retention** configurabili
  - Di default partition 3, replica 3 e retention di una settimana
- Dati possono essere letti da diversi **consumer**
  - Eventualmente organizzati in consumer groups
- L’autorizzazione è gestita tramite **ACL**
  - Di default solo admin accede a tutti i topic
  - ACL sui topic e sui consumer group

Logstash

- **Filtraggio/arricchimento dei dati**
  - Estrazione info rilevanti che diventano campi degli indici OpenSearch
  - Costruzione nuove info in seguito ad operazioni su info esistenti
  - Rimozione record non rilevanti
  - Sincronizzazione orario entry nell’indice con orario del log
  - Possibile invio dei record arricchiti su nuovo topic Kafka
- Invio ad OpenSearch via SSL
  - Autenticazione con user e password

Opensearch - overview

- Strumento per aggregare, visualizzare e analizzare dati di vari tipi
  - Fork open source di Elasticsearch gestito da Amazon
  - Plugin analoghi a quelli forniti a pagamento con Elasticsearch
  - **Plugin Security** per la gestione di tenant, ruoli, utenti, permessi
- Configurata in HA (tre nodi master e quattro nodi data)
- Dati caldi su SSD
- Dopo una certa retention, i dati possono essere spostati su **repo Ceph**

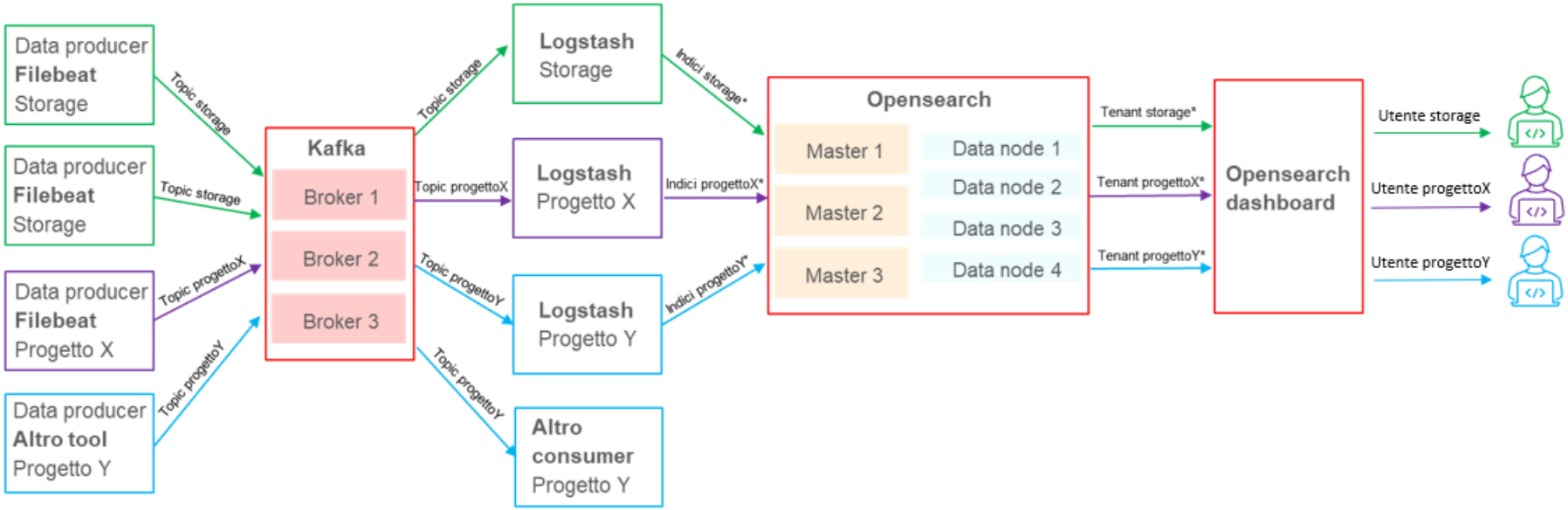
Opensearch - autenticazione/autorizzazione

- Via dashboard o API
- Doppio metodo di autenticazione
  - **Locale** : user e password (admin e utenti di servizio)
  - Via **IAM CNAF** (iam.cnaf.infn.it)
    - Utenti locali IAM
    - Utenti **AAI INFN**
- Autorizzazione tramite **gruppi IAM**
  - I gruppi IAM dell'utente vengono mappati in automatico su *backend roles* di OpenSearch
  - Ad ognuno di essi sono associati permessi per operazioni sugli indici
  - Admin IAM crea gruppi dei reparti/progetti che usano la piattaforma
    - gruppi con permessi read-write o read-only sui dati di ciascun reparto

Wazuh

- Piattaforma open source che utilizza OpenSearch
- Specifica per monitorare e allarmare in caso di eventi legati alla **sicurezza** e di incidenti
- Possibilità di effettuare azioni automatiche in caso di minacce
- In fase di integrazione nella piattaforma BDP
    - Dashboard separata che visualizza dati di Wazuh gestiti dal servizio OpenSearch

Schema piattaforma



HA e scalabilità

- Kafka e OpenSearch cluster **multinodo**
- Piattaforma **costruita per poter scalare** orizzontalmente in maniera semplice e modulare
  - Numero dei broker Kafka
  - Numero delle istanze di Logstash
  - Numero dei nodi master / data di OpenSearch
- Lo spazio storage può essere aumentato

Ridondanza e retention dei dati

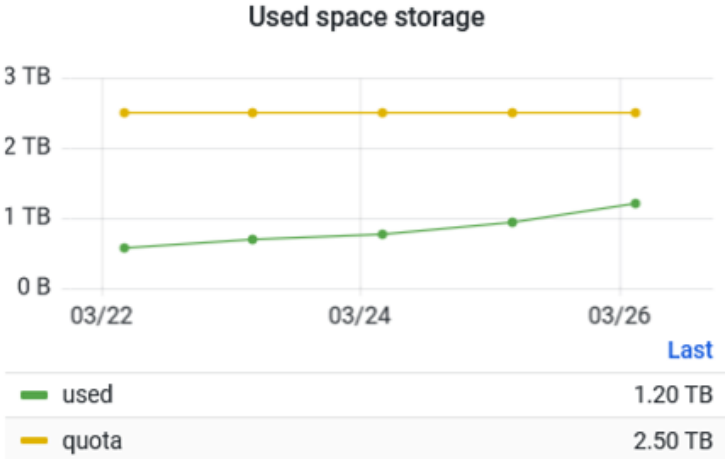
- I dati su Kafka sono replicati tre volte, sui dischi di tre nodi diversi
- Su Kafka una diversa **retention per topic**
- Gli indici su OpenSearch
  - Sono di default in **replica 2**
  - Vengono messi offline via **snapshot** su storage Ceph in base dell'età
  - Snapshot periodiche degli indici di servizio (definizioni dashboard, configurazione ruoli, permessi, etc.) su storage Ceph

Segregazione dei dati e multi-tenancy

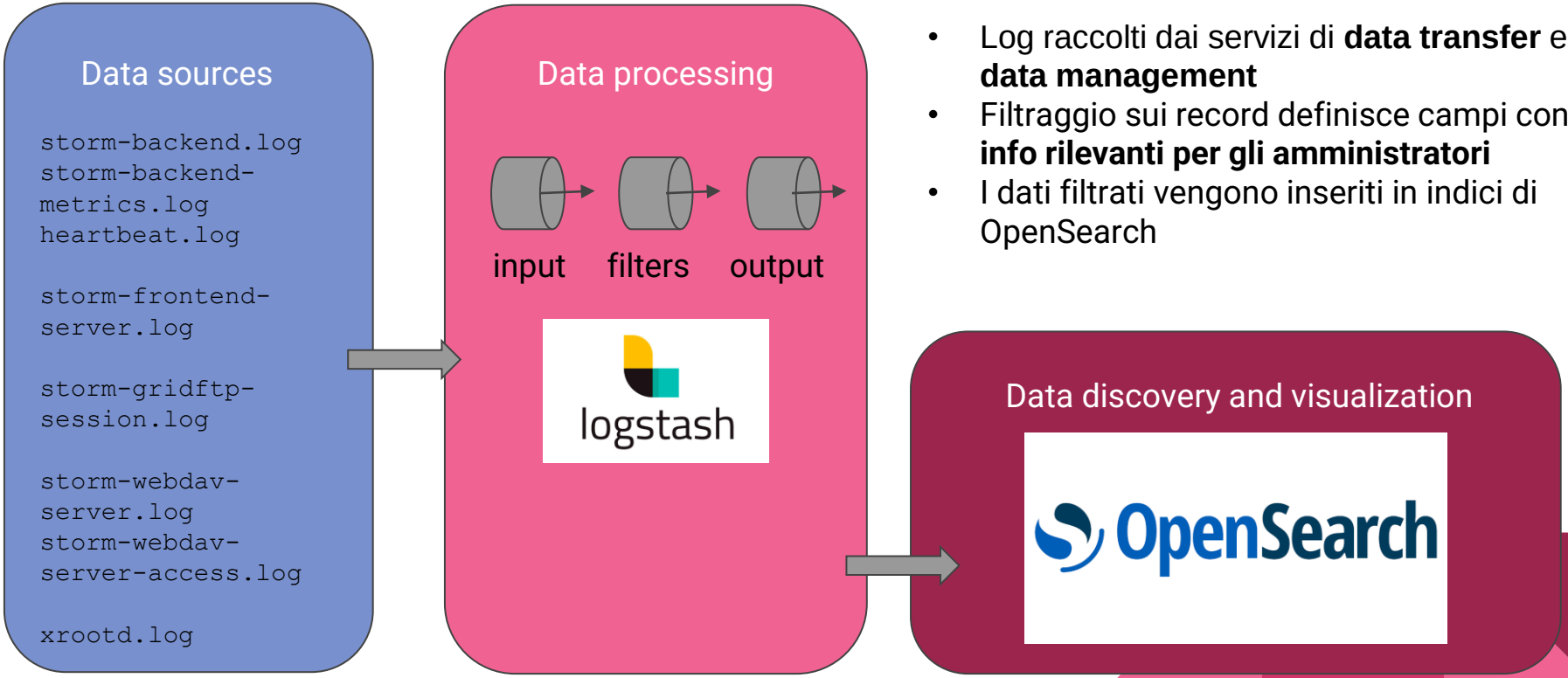
- Ogni reparto/progetto può scrivere/leggere solo sul/dal proprio topic Kafka e accedere ai propri indici OpenSearch
- Ogni utente può agire sui dati del proprio reparto/progetto
  - Possibilità di accesso read-only ad altri dati
- In OpenSearch ciascun utente può accedere a:
  - Un tenant **pubblico** comune a tutti gli utenti
  - Un tenant **privato** non accessibile da altri
  - Uno o più tenant di **reparto/progetto** con dashboard specifiche (se appartiene ai gruppi IAM che ne permettono l'accesso)

Integrazione con servizi esistenti

- Tutti nodi integrati in **Foreman/Puppet**
- Autenticazione/autenticazione tramite **IAM** CNAF
- **Monitoring** con servizi CNAF
  - Sensu / InfluxDB / Grafana
  - Check su stato dei servizi, occupazione spazio
  - Utility che controlla spazio occupato da indici di ciascun reparto e lo confronta con una quota
- **Documentazione** su Confluence
  - Descrizione infrastruttura
  - Doc per utenti
  - Doc per amministratori

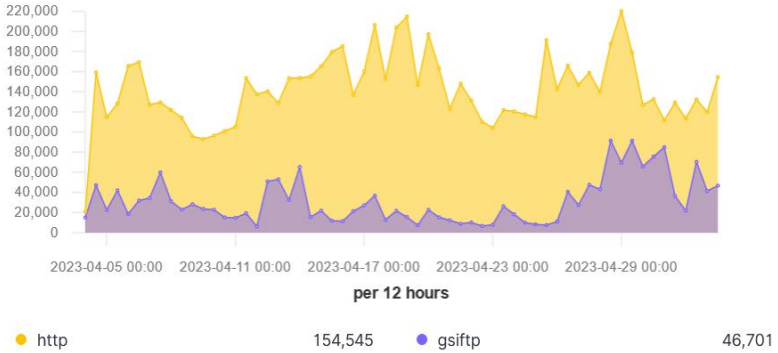


Caso d’uso: analisi log dei servizi storage

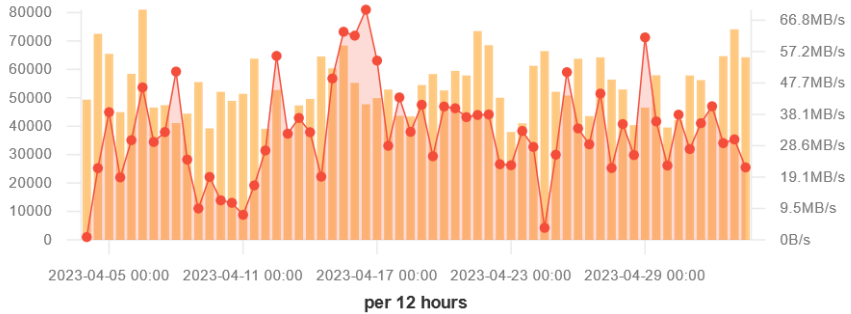


- Log raccolti dai servizi di **data transfer** e **data management**
- Filtraggio sui record definisce campi con **info rilevanti per gli amministratori**
- I dati filtrati vengono inseriti in indici di OpenSearch

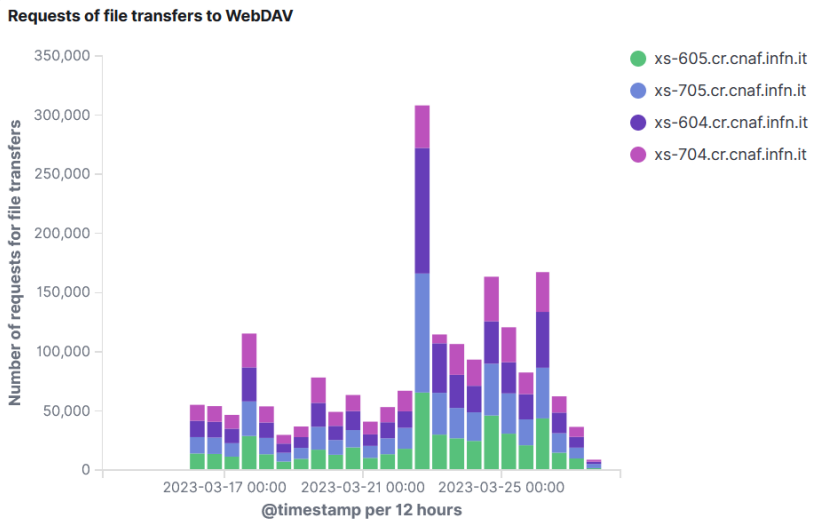
Caso d'uso: analisi log dei servizi storage



Numero di file trasferiti per protocollo (per monitorare il passaggio da GridFTP a StoRM WebDAV)



Numero e rate di trasferimenti http di ATLAS



Distribuzione delle richieste di trasferimento file tra le 4 istanze di StoRM WebDAV dedicate ad ATLAS al CNAF

Caso d’uso: visualizzazione log di più nodi

- Possibilità di visualizzare log relativi ad un dato evento che coinvolge servizi diversi che girano su nodi diversi
- Allo scopo di facilitare il **troubleshooting**
- Esempio
  - Indagine su problemi relativi al trasferimento di un file
  - Analisi di log di vari servizi (StoRM Frontend e Backend, WebDAV, GEMSS) ciascuno dei quali può girare su istanze multiple

Conclusioni

- Infrastruttura **modulare** e **scalabile**, utilizzabile per gestione e analisi dati di reparti/progetti CNAF
- Già presenti log da diversi nodi del CNAF
  - Lavori in corso per inserire / parsare altri log
- Integrazione con Wazuh in corso
- Lavoro in collaborazione con personale di diversi reparti del CNAF
  - A. Falabella, E. Fattibene, F. Fornari, L. Morganti - Storage
  - S. Antonelli - SSNN
  - D. Michelotto - Farming
  - D. Lattanzio - User Support
  - V. Ciaschini, F. Amori - Security