

BDP@CNAF

Big Data Platform at CNAF

Diego Michelotto (diego.michelotto@cnaif.infn.it)

et al.

Obiettivi

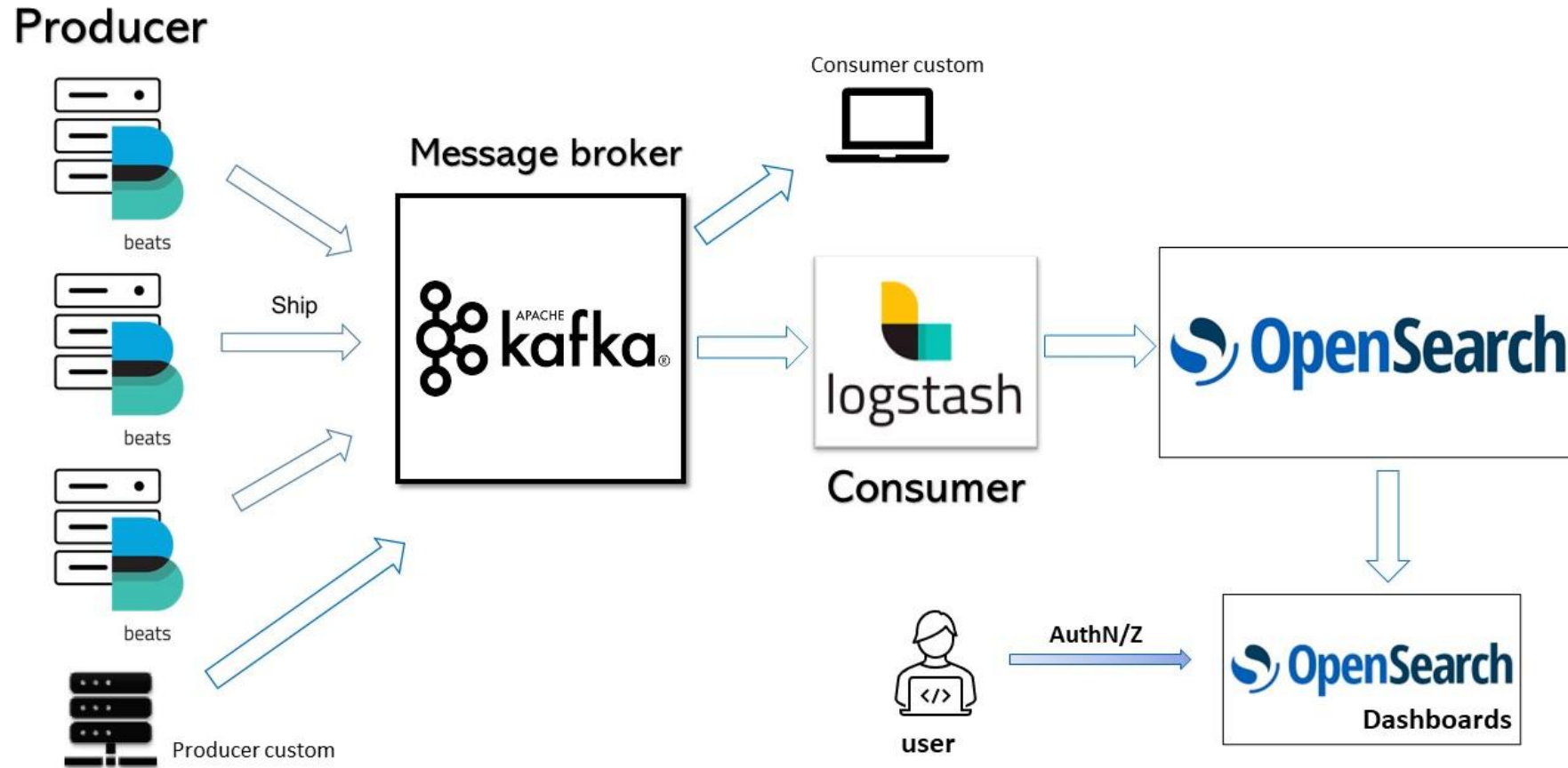
- **Predictive maintenance:**

- Gemini dice: «Predictive maintenance (PdM) is a proactive strategy that uses sensors and data analysis (AI/ML) to monitor equipment health in real-time, predicting potential failures before they happen, allowing maintenance to be scheduled precisely when needed to maximize uptime, extend asset life, and reduce costs compared to traditional time-based or reactive methods. It moves beyond simply fixing things after they break (reactive) or on a fixed schedule (preventive) by analyzing vibration, temperature, and other data to forecast the optimal time for service, making it a cornerstone of Industry 4.0»

Obiettivi

- Servizio per indicizzazione informazioni
 - Servizio per ingestion di dati
 - Servizio per trasformazioni dei dati
- Servizio per consultazione delle informazioni
- Archiviare e conservare le informazioni
- Usare software opensource
- Multitenancy
- Encryption delle comunicazioni
- Servizi riusabili

Soluzione



Soluzione

- Indicizzazione: cluster [Opensearch](#)
 - 3 master node (VM) + 4 data node (BM) con 32TB SSD + 1 dashboard (VM)
- Ingestion:
 - Cluster [Apache Kafka](#): 3 nodi con 600GB disco
 - 1 topic per ogni gruppo/reparto/esperimenti dedicato
 - Ogni topic 3 partizioni e 3 repliche, durata 1 settimana
 - Manipolazione con Logstash (fork Opensearch)
 - 1 o più per ogni gruppo/reparto/esperimenti filtri gestiti in autonomia
- Produzione:
 - [Filebeat](#): soluzione consigliata per i log
 - Qualsiasi strumento in grado di comunicare con Kafka e che produca documenti JSON

Applicazioni

- Storage
 - Monitoraggio WebDAV
- Farming
 - Monitoraggio K8s/IAMs
 - Monitoraggio Cloud@CNAF (TB)
 - Monitoraggio job HTC (TB)
- Security
 - Wazuh (XDR - SIEM)
- Esperimenti
 - Atlas



Futuro

- Dai reparti:
 - Storage: monitorare nginx, xrootd, StoRM Tape REST Api
 - Farming: completare monitoring cloud e HTC
- Rinnovo/espansione data node
- Puppettazione dell'infrastruttura + aggiornamento
- ... Predictive maintenance
 - standardizzazione/parsing log
 - real time data processing (Spark?)

Creditis

- Diego Michelotto, Lucia Morganti, Antonio Falabella, Ahmad Alkhansa, Alessandro Pascolini, Aksieniia Shtimmerman, Daniele Lattanzio, Enrico Fattibene, Francesco Sinisi, Giuseppe Misurelli, Giusy Sergi, Giacomo Levrini, Marco Lorusso

- Per collaborare contattatemi