



Un'architettura multi-tenant ad alte prestazioni per la messaggistica server-to- server: Apache Pulsar

Presentata a Ismam Abu



Argomenti

Sistemi a scambio di messaggi

Apache Pulsar

Features più interessanti

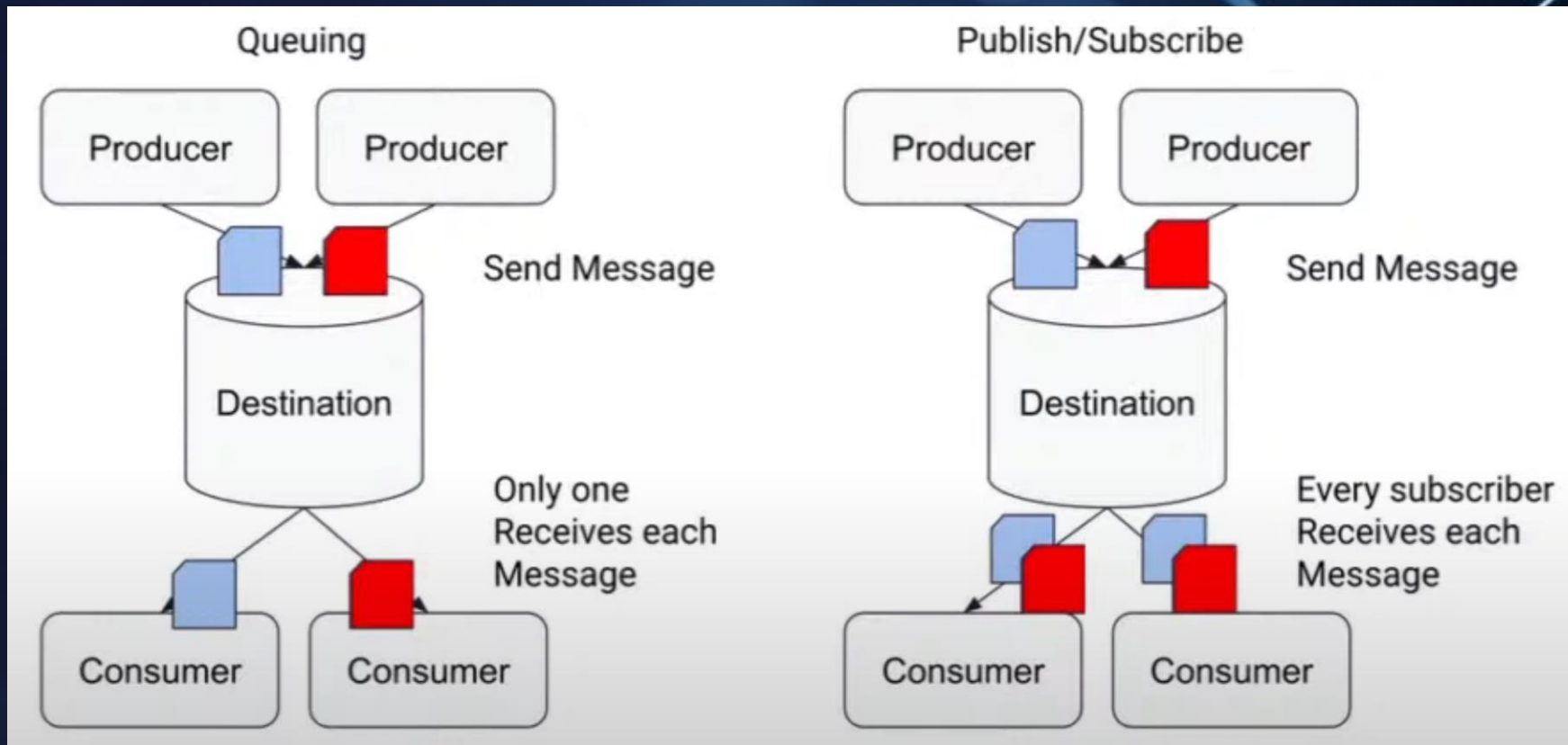
Architettura di Pulsar

Modello dei dati



Sistemi a scambio di messaggi

Queuing vs Publish/Subscribe





Apache Pulsar

Pulsar è un progetto Open Source nato a Yahoo! E successivamente donato ad Apache Software Foundation nel 2017.

Evoluzione delle tecnologie message oriented:

1. ActiveMQ & RabbitMQ
2. Kafka
3. Pulsar

Features aggiuntive:

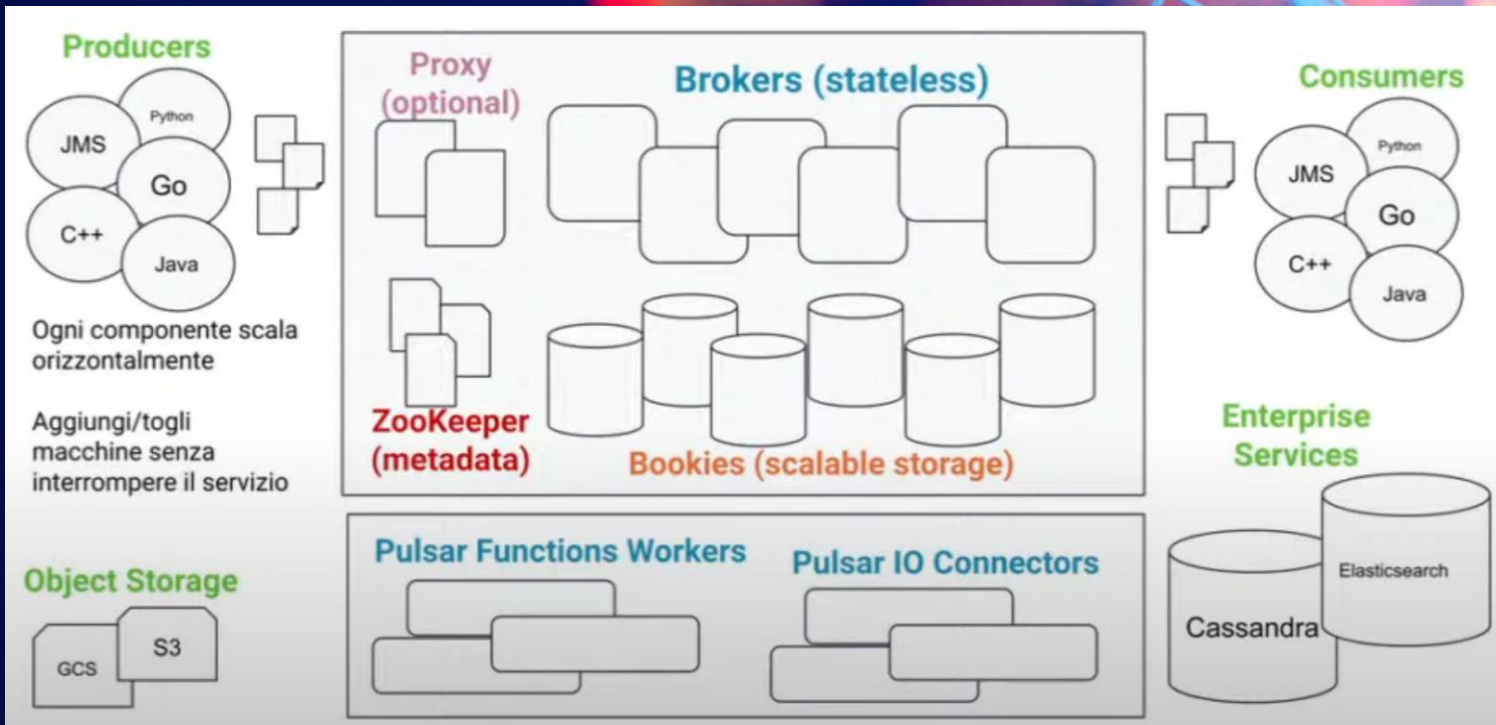
- Scalabilità (storage dei dati fatto con bookkeeper)
- Multi Tenancy
- Geo Replication
- Offloading dei dati vecchi
- Supporto a Kubernetes
- Schema Registry nativo (AVRO, JSON ecc)
- Connectors
- Pulsar Functions
- Supporto a svariati linguaggi

Apache Pulsar - Highlights

- Performance: milioni di messaggi al secondo con latenza $\leq 10\text{ms}$ (i dati sono persistenti, quando si scrive vengono inseriti sui dischi)
- Scalabile orizzontalmente: lo storage e i broker sono scalabili orizzontalmente, se ti serve più potenza colleghi semplicemente più macchine
- Message Replay: fare tornare indietro nel tempo il sistema(per quanto riguarda i processi dei consumer, per esempio dopo aver fixato un bug si possono consumare nuovamente dati già consumati)

Apache Pulsar - Highlights

- Off loading: possibilità di spostare i dati vecchi su storage a basso costo e tenerli per sempre (oppure anche spostare i dati su AWS, Google Cloud Storage o Azure)
- Consolidamento: mettere più applicazioni nello stesso cluster, non ci sono conflitti e si possono dare risorse ad applicazioni specifiche
- Geo-Replication: interconnettere più cluster in regioni diverse, sia per fare backup che per ottenere performance migliori



Architettura Apache Pulsar

Applicazioni Producer che scrivono
Applicazioni Consumano che consumano
Cluster:

- Broker (smistano messaggi)
 - Bookie (storano dati)
 - ZooKeeper (gestisce i metadati)
- Database (leggere/scrivere)

Topics e Subscription 1/3

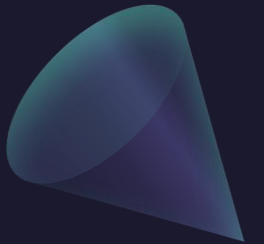
MODELLO UNIFICATO PER SCAMBIO DI MESSAGGI

- Topics

- Persistenti / non persistenti (perdo i dati riavviando il broker)
- Partizionati (distribuzione del carico migliore) / non partizionati

- Tenants e Namespaces

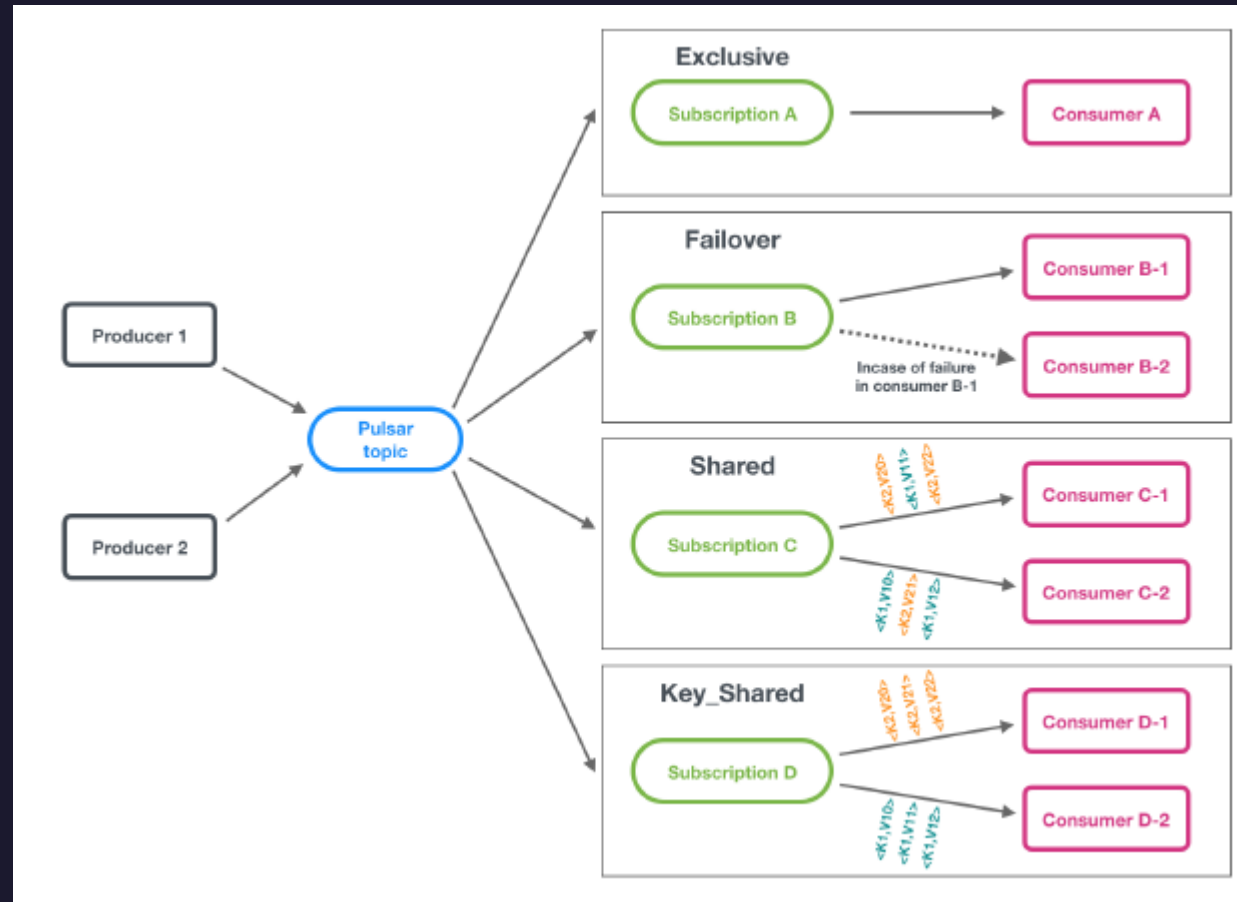
- Isolamento logico e fisico delle risorse
- Configurazione a tutti i livelli: topic/namespace/tenant/system



Topics e Subscription 2/3

- Modalità subscription

- Exclusive
- Failover
- Shared
- Key Shared



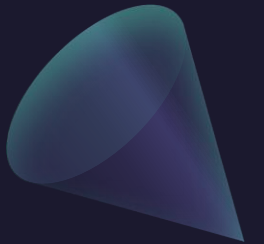
Topics e Subscription 3/3

- Tipi di subscription

- Durable (sullo storage vengono salvati i messaggi già processati, quando ti connetti non li riprocessi)
- Non-durable (lo stato è salvato solo sul client)

- Modalità per i producer

- Normal (tutti i producer possono mettere messaggi sul topic)
- Exclusive (un solo producer può scrivere alla volta, utile nell'event sourcing)





Riepilogo

Apache Pulsar:

- E' cloud native, nasce con multi tenancy e geo replication
- Tutti i componenti sono scalabili orizzontalmente
- E' open source, ma con una comunità molto attiva

Grazie

Ismam Abu

UNIBO – DISI

Sistemi Distribuiti

