

APACHE
kafka[®]

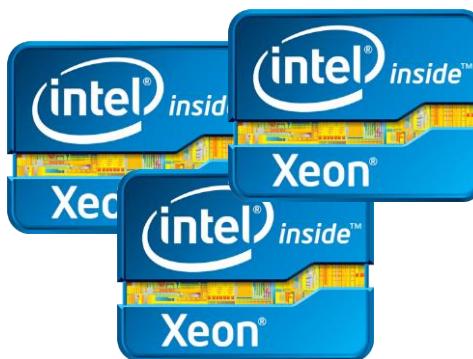
UNIT

Prerequisiti

- «Apache Kafka» è un'applicazione compilata in «bytecode Java» e quindi può essere eseguita su tutti quei sistemi operativi per i quali sia disponibile la Java Virtual Machine (JVM).
- Ciò include Windows, MacOS, Linux e altri.
- L'uso di Kafka in un ambiente **Linux** è consigliato poiché si tratta del sistema operativo più comune su cui è installato.
- Linux è anche il sistema operativo **consigliato per la distribuzione di Kafka**.



- Per un utilizzo efficiente dei cluster di Kafka è necessario lavorare con l'hardware adeguato.
- Il team di sviluppatori di Kafka consiglia l'utilizzo di processori **Intel Xeon** con Quad-Core e **24 Gigabyte** di memoria RAM.
- Importante è, in generale, avere una **memoria sufficiente** per salvare in un buffer i messaggi letti e scritti da tutte le applicazioni che hanno accesso attivo al cluster.



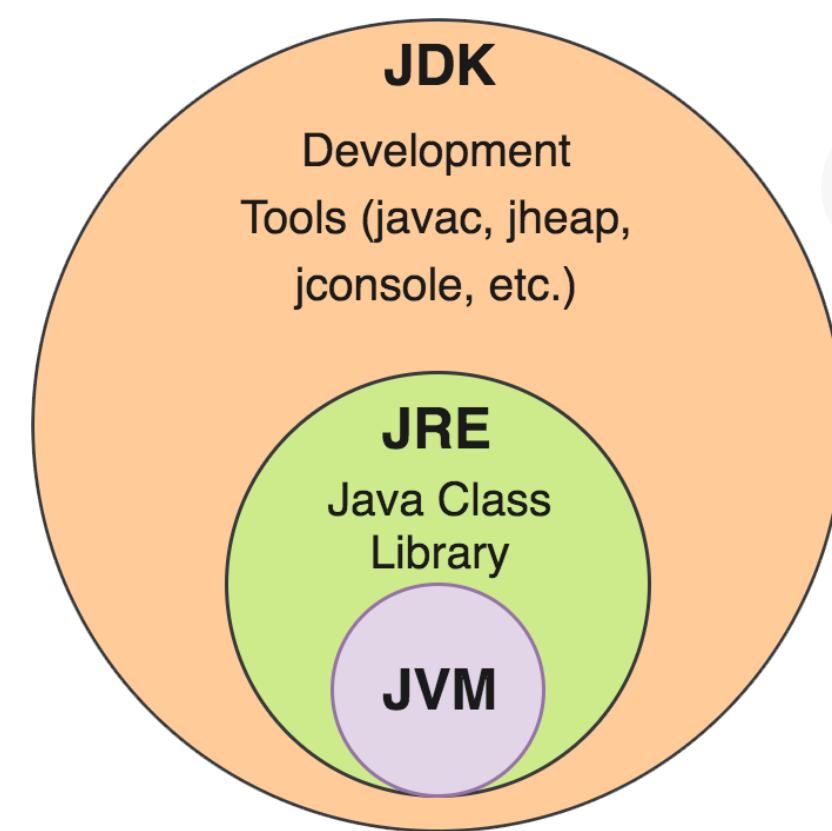
- Uno dei vantaggi di Apache Kafka è l'elevato flusso di dati, motivo per cui **è di estrema importanza scegliere dischi rigidi adeguati.**
- La Apache Software Foundation consiglia unità **hard-disk SATA (8 x 7200 UpM)**; per prevenire carenze di prestazione si applica il principio «**più dischi rigidi si hanno, meglio è**».



- Anche per quel che concerne il software bisogna rispettare alcune condizioni per utilizzare Kafka ai fini della **gestione di flussi di dati in entrata e in uscita**.
- Per la scelta del sistema operativo, ad esempio, conviene prediligere un sistema Unix come **Solaris** o una distribuzione **Linux** (così come accennato in precedenza), visto che le piattaforme di Windows offrono un supporto limitato.



- Poiché Kafka è scritto nel linguaggio **Scala** compilato in «bytecode Java» è **indispensabile** che sia installata sul proprio sistema una versione più attuale possibile del Java SE Development Kits (**JDK**) o del Java Runtime Environment (**JRE**) al fine di disporre della JVM di Java.



- Un'altra componente obbligatoria è il servizio «Apache ZooKeeper», che permette la sincronizzazione di processi ripartiti.



UNIT

Installazione e preparazione di un ambiente di sviluppo

Installazione

- <https://www.confluent.io/installation>
- Local
- Distributed
- Community



Installazione – Confluent – Prerequisiti

Component	Nodes	Storage	Memory	CPU
Control Center	1	300 GB, preferably SSDs	32 GB RAM (JVM default 6 GB)	8 cores or more
Broker	3	<ul style="list-style-type: none">• 12 X 1 TB disk. RAID 10 is optional• Separate OS disks from Apache Kafka® storage	64 GB RAM	Dual 12 core sockets
Connect	2	Only required for installation	0.5 - 4 GB heap size depending on connectors	Typically not CPU- bound. More cores is better than faster cores.
KSQL	2	Use SSD. Sizing depends on the number of concurrent queries and the aggregation performed.	20 GB RAM	4 cores
REST Proxy	2	Only required for installation	1 GB overhead plus 64 MB per producer and 16 MB per consumer	16 cores to handle HTTP requests in parallel and background threads for consumers and producers.
Schema Registry	2	Only required for installation	1 GB heap size	Typically not CPU- bound. More cores is better than faster cores.
ZooKeeper	3-5	<ul style="list-style-type: none">• Transaction log: 512 GB SSD• Storage: 2 X 1 TB SATA, RAID 10 <p>Each write to ZooKeeper must be persisted in the transaction log before the client gets an ack. Using SSD reduces the ZooKeeper write latency.</p>	32 GB RAM	2-4 Cores