

# Flume Vs. Kafka

Omid Vahdaty, Big Data Ninja



+



Apache Kafka

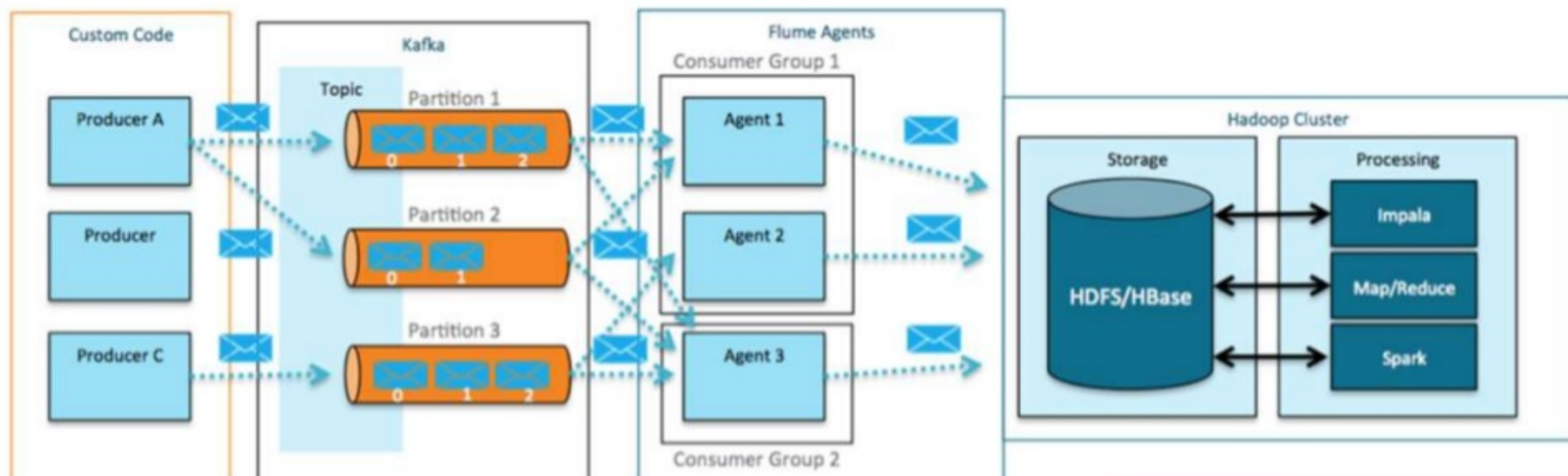
	Flume	Kafka
Original Motivation	Apache Flume is a distributed, reliable, and available system for efficiently collecting, aggregating and moving large amounts of log data from many different sources to a centralized data store. Built around hadoop ecosystem	. A general purpose distributed publish-subscribe messaging system Multi-consumer ultra-high availability messaging system.
Data Flow	push	pull
event availability	JDBC Databases Channel, file Channel. Loose flume agent = losing data.	replication of your events data by design.
Commercial support	Cloudera	Cloudera
Collectors built in	Yes.	just the messaging

	Flume	Kafka
Choose when you desire	No need for customization. Need out of the box components such HDFS sink	Need a custom made high availability delivery system
Velocity	high	higher
Event processing		

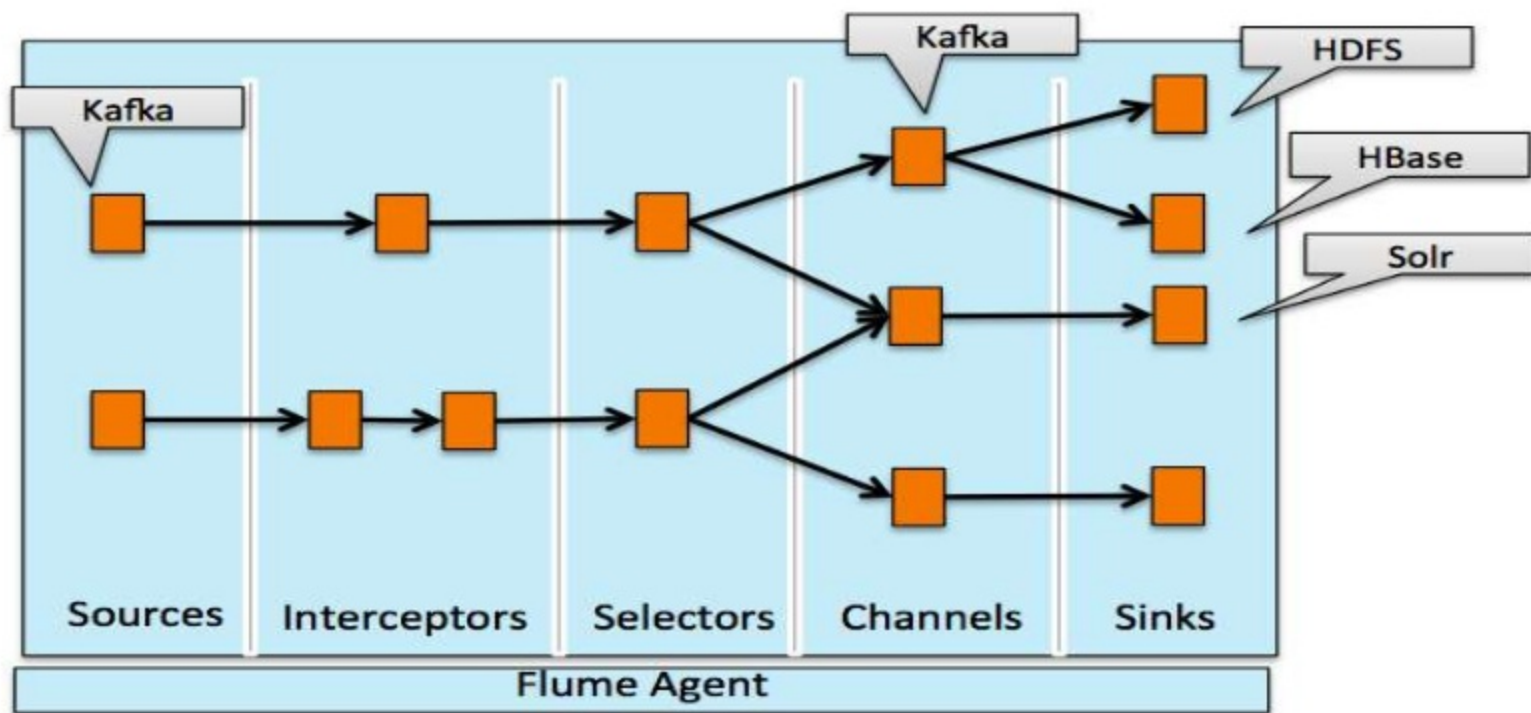


# Kafka and Flume combined

- Flume supports: Kafka source, Kafka channel, Kafka sink
- So, take the advantage of both and combine them to your needs.



# Kafka as a Channel




# Sources:

- <https://www.quora.com/What-are-the-most-significant-differences-between-Flume-and-Kafka>
- [https://www.cloudera.com/documentation/kafka/latest/topics/kafka\\_flume.html](https://www.cloudera.com/documentation/kafka/latest/topics/kafka_flume.html)



# Stay in touch...

- [Omid Vahdaty](#) 
- +972-54-2384178



 **HALO**  
ANALYTICS



The logo for Jajah Telefonica. It features the word 'jajah' in a stylized, pink, sans-serif font, with a small grey icon above the 'j'. Below it, the word 'Telefonica' is written in a black, cursive script font.