

# My Experience with Apache Pulsar



I have been working on Apache Kafka for a pretty long time & since last few months, I have been following Apache Pulsar. Finally I decided to move my personal project to Apache Pulsar to find out more ( It will be mid level data project but in complex side)

I will avoid explaining any technical details here as there are many resources already available, rather sticking to my personal **experience** with Pulsar in *brief*.

## How was the setup ✓

I found it **Complex**. It has some new components which I had no experience with, so finally I needed to learn them like Apache Bookkeeper.

Knowing Apache Bookkeeper was totally worth it as Pulsar persists data on Bookkeeper & Bookkeeper has many interesting features which, for a multi-cloud environment, is extremely helpful.

Read Bookkeeper

<https://bookkeeper.apache.org/docs/4.6.1/getting-started/concepts/>

Coming back to set up, I first tried with Multi Cluster on bare Metal, & if you are already used to setting up multiple components, one will get used to it.

*I always have problems with Zookeeper and communication, so I had it here as well, so my case may be a bit different. I needed to keep my ZK pods separate from the worker.*

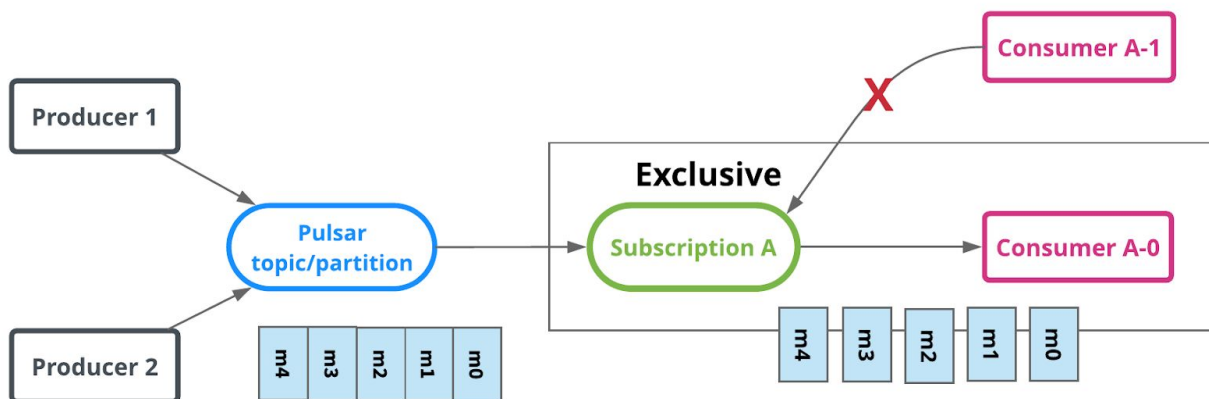
Setting up on [Kubernetes](#) was arguably easier. For K8s, I just followed the documentation & it worked. This is seriously amazing and I will highly recommend using K8s version ( if setup permits).

## What was different in Pulsar

Pulsar is definitely exciting for message Subscription. Pulsar provides 4 kinds of subscription which just makes Apache Kafka boring.

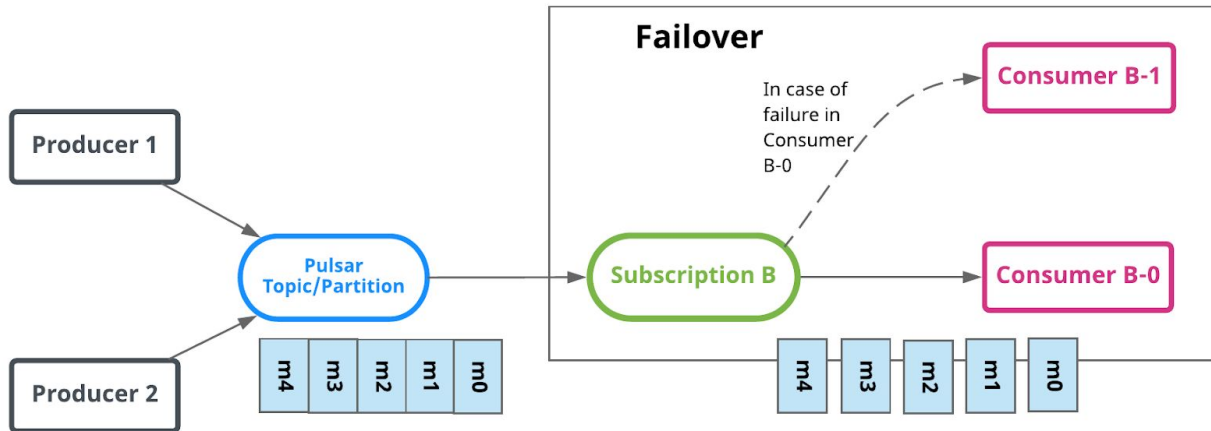
### Apache Pulsar Subscription Mode

#### 1. Exclusive



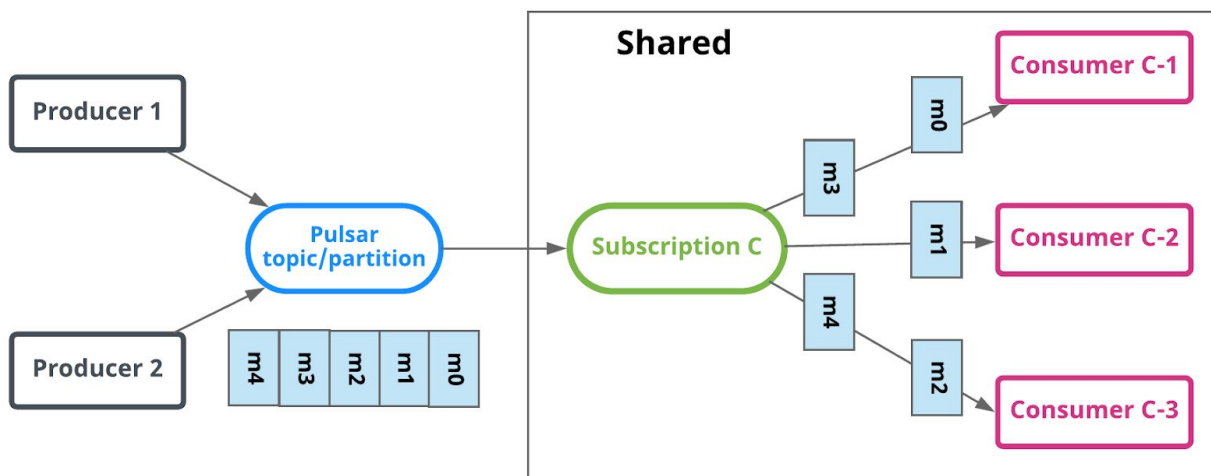
Only Single Consumers are allowed to subscribe.

#### 2. Failover



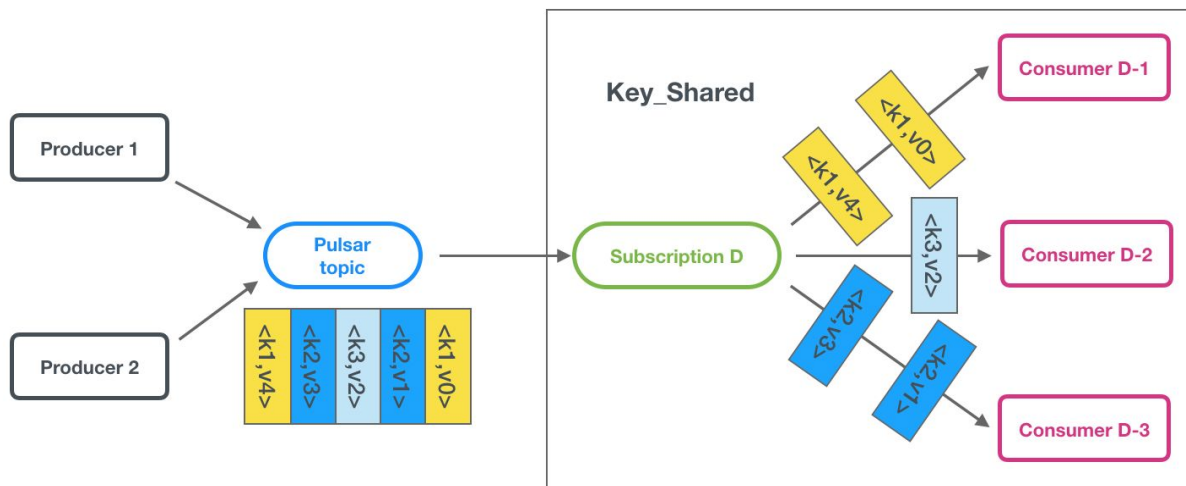
Multiple Consumers attached to a single topic and sorted lexically by names.

### 3. Shared



Round-Robin message delivery.

### 4. Key\_Shared



Multiple consumers can attach to the same subscription and delivery distributed across consumers & messages with the same key or same ordering key are delivered to only one consumer.

## What am I missing in Pulsar ?

Documentation and working example !

Well with any new technology, Pulsar suffers the same problem, there were many loopholes in Documentation & not enough examples.

Pulsar is not designed for High Throughput. ( remember this)

Conceptually, if one understands the architecture properly, it's easier to debug and find a solution, but that brings a very complicated setup and I am pointing it again here !

Even Pulsar doesn't do well with large data size ( I am not surprised) !

## What am I liking ?

- Resource Management , Data Organization ( Shared Consumption) & Geo-Replication along with Data Replication over multiple clusters are great.
- Pulsar Topic scalability is more mature compared to Kafka.

- I like Bookkeeper, so that brings another layer of usability. As I noted earlier, Pulsar architecture has more components, so that brings many improved features which Kafka can't. As it's new ( technically), it has more improved features compared to Kafka.
- Pulsar on Kubernetes is just natural, kafka can't stand!
- Finally, Queueing & Streaming, both in a single system, who doesn't like that.
- *Infinite Retention Policy is great on paper. It's easy to set up (credit Bookkeeper), but I didn't explore it but I am currently trying it.*

## The Wow Factor

After setup, I am finding Pulsar works pretty well and Auto Scaling is Amazing. It really Auto Scales without bothering. I specifically like the Bookkeeper part.

## Conclusion

Apache Pulsar is really an amazing technology & there are many promising features. I am still not finding any problem in the workflow/pipeline and it definitely has more features compared to Kafka, but that being said, Kafka is mature and has a pretty large customer base.

But does the Pulsar really have better performance compared to Kafka?  
No, I didn't experience anything like that.

But if one needs streaming and queueing both, I will definitely push Pulsar without a second thought & if you already have Kubernetes setup, testing pulsar is really really easy.