



Kafka on AWS with Raid-0 striping

0  [AWS confluent quickstart](#) configures Kafka `log.dirs` with 4 512GB EBS block devices with RAID-0 striping for higher throughput and also helps bypass the 1TB limit of block devices without provisioned IOPS. I have just learned that losing a block device in a RAID-0 group will cause all other devices in that group to fail, can someone help clarify this 

★ Now that Kafka allows multiple directories under `log.dirs`, can we mount each block device under a different mount point and configure them as a list of directories under `log.dirs`?

If that is possible(which it is, I guess), what are the trade-offs?

amazon-web-services


apache-kafka

confluent

raid

aws-ebs

 A couple things to note.



2  First, there isn't a 1TB limit on EBS volumes. As of the moment, Amazon st1 volumes can be as big as 16TB. These are the kind of volumes you want to use in your Kafka deployment because they're optimized for sequential writes, which is what Kafka does best.



✓ Secondly, yes--Kafka allows for multiple log directories. This allows you to spread storage across disks so that you're not overtaxing a single disk with all of your io. That said, having multiple log directories is going to be better than having a single directory, especially if you're dealing with large amounts of data--but there are other factors to keep in mind, too, when dealing with EBS. If you're opting for smaller st1 volumes rather than a monolithic st1 volume, that means you'll have a smaller burst bucket and a lower iops baseline per volume. Once you go over your iops baseline, you'll start consuming iops from your bucket--[see details here](#). It's important to monitor your burst balance in CloudWatch to make sure it's not being routinely depleted, which usually results in your whole cluster slowing down and your broker's request and response queues filling up, which could lead to catastrophic failures across consumer and producer applications.

As for RAID striping, if you enable it on each of your EBS volumes, all of your mounted volumes will be in the same RAID group, which means that Kafka log files will be spread across devices in the group rather than residing on a single device, the consequence of which is that if one of those devices fails, the other devices in the group will fail, too. This is supposed to be more performant than other setups, however.

Before Kafka 1.0 there was no operational difference between a single disk failing on a broker and every disk failing on that broker--both would result in the broker going down. [See discussion here](#).

Update: As of Kafka 1.0, a failed disk will not bring down the broker ([see docs](#)). Thanks to @RobinMoffat for pointing out. Ultimately with RAID-0 striping, you're trading the ability to quickly recover from a failed disk for overall io performance. That is, all partitions on a broker with a single failed disk will need to be reassigned with striping, but without striping, only those partitions on the failed disk will need to be reassigned.

 If we have each ebs volume mounted to a different directory and configured `log.dirs`, this should help with distributing IO across multiple devices, but incase of a failed volume will it help in recovering a crashed broker, since we have partitions in other volumes unaffected and only the partitions in the failed volume will need to be replicated from other brokers – Somasundaram Sekar Feb 20 '18 at 16:36 

1  As of Apache Kafka 1.0 / Confluent Platform 4.0 : per [doc](#) *A single disk failure in a JBOD broker will not bring the entire broker down; rather, the broker will continue serving any log files that remain on functioning disks.* – Robin Moffatt Feb 20 '18 at 16:40 

 Also general reference for sizing on AWS etc: [confluent.io/whitepaper/...](#) – Robin Moffatt Feb 20 '18 at 16:40



@RobinMoffatt thanks for that link. Will edit answer to reflect. – kellanburket Feb 20 '18 at 16:45
