

Is Data locality overrated?

By Deba Chatterjee, Director Products | May 12, 2017 | Application Virtualization, Applications, Big Data, Blog, Containers

“Yes, we do see a lot of underutilization in our Hadoop infrastructure,” said the chief architect of one of the largest financial institutions and then with a pause, he added, “It is getting worse every year”.

I was surprised and asked him to explain “Why is this getting worse?”

“Well, we need to ensure predictable performance and keep our end users happy” and then continued to add, “but as the storage requirements grow, we add a new set of machines every year. See we have standardized hardware spec and as we add storage we automatically add more computing power”.

“We also need to ensure data locality i.e. run the MR jobs on the same machine where the data is !!!”

There was a lot of anguish when he concluded saying that he needs to solve this problem and he wished he could run multiple clusters on the same machine.

“Only if I had a way to virtualize Hadoop”

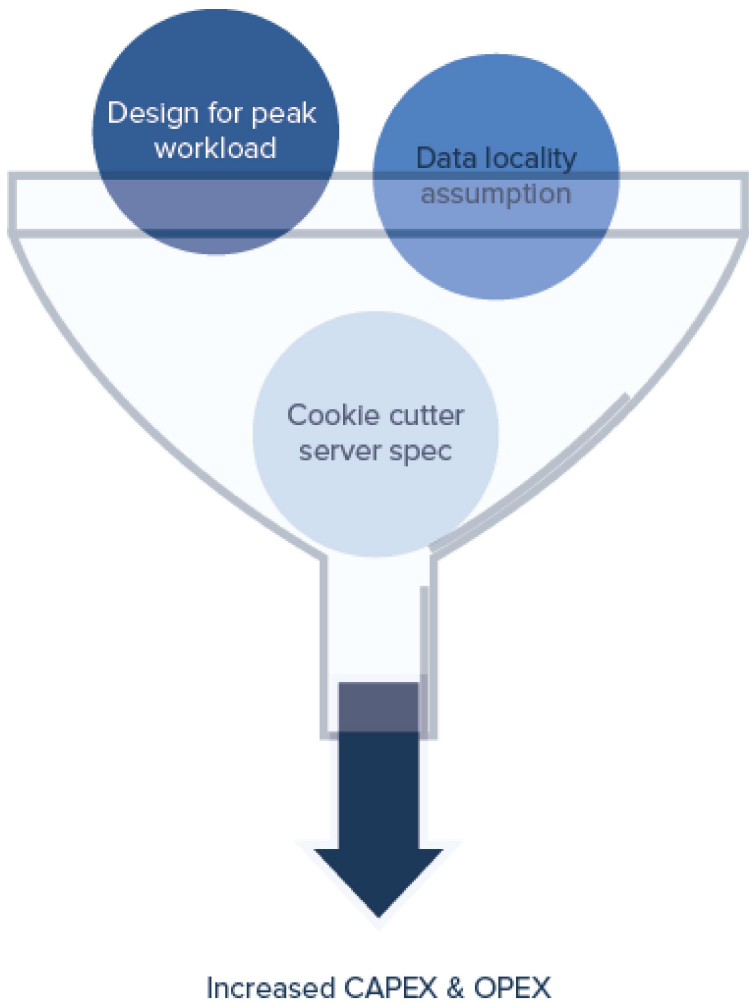
I came out of the meeting pondering on the question “Is data locality really that critical?”

The Perspective – Data Locality is no longer mandatory

If we look at the increasing cost of managing a Hadoop implementation, the three things that stand out are the decision to design based on peak workload, the data locality assumption and cookie cutter server sizes.

Inspecting Data locality

Computing frameworks such as MapReduce split jobs into small tasks that are run on data nodes. A popular approach to improve the efficiency of these jobs in Hadoop is to implement data locality – an approach to make sure the tasks that run on the data nodes have the input data stored on the local disk.



Sign up for email updates

Database, Container, DevOps Insights Delivered to Your Inbox

[Subscribe now](#)

But as this paper from Amplab (at UC Berkeley) points out disk locality for Hadoop is given too much importance when on a typical hardware setup - reading from local disk is only about 8% faster than reading from the disk of another node in the same rack. Another paper, on a similar topic from UC San Diego also questions the data locality assumption and proposes the use of Super Data Nodes in conjunction with virtualization to speed up Hadoop job execution time significantly (by 54%) .

The history behind the data locality assumption is quite simple. The time when Hadoop was becoming popular, network access speeds were levels of magnitude slower than local disk access. Hence putting data in local disks was almost common sense. But today improvements in network bandwidths have outpaced disk access speed and disk locality no longer has to be a mandatory requirement. On the other hand, the need to save more and more data in clusters today has created a great demand for affordable storage and more often than not it is no longer feasible to have all the data on the same disks where the compute is. Also, improvements in compression technology has made sure more data transferred over the network is much less than the actual data volumes and can significantly reduce the impact of network access times.

The business impact of decoupling

We will go over a case study on a Hadoop deployment at a large financial institution. In this first table below, based on the anticipated storage growth the bank calculated the number of nodes they would require for the next 3 years. As the bank uses the same server specs for all nodes, along with the increase in storage capacity, addition of new nodes also results in an increase in total compute capacity irrespective of whether it is needed or not.

| Year | 1st | 2nd | 3rd |
|--------------------------|------|------|------|
| # of Machines | 200 | 278 | 358 |
| Total usable memory (TB) | 84 | 116 | 148 |
| Total usable cores | 4600 | 6394 | 8234 |

This bank, like the architect of the other financial company, will end up adding about 1800 cores after the first year and then another almost 2000 cores in the second. There is no need to grow their CPU or memory footprint at this rate, yet in order to comply to the data locality dictum, they have no way but to grow their infrastructure and suffer from underutilization.

Now, let us look at the same scenario in a decoupled mode implementation.

Robin’s container-based cloud platform for Big Data and Databases decouples the compute and storage tiers and allows its customers to scale each tier independently. Let us look at the same setup on Robin. By splitting the nodes into two groups (same number of nodes to begin with), 100 as compute and another 100 as storage, we can now scale only the storage tier as the customer data grows. This not only results in fewer number of nodes being added every year but also addresses the issue of underutilized servers. Here is how the growth in nodes look like in the decoupled scenario

| Year | 1st | 2nd | 3rd |
|---------------------------------|------|---|-----|
| # of Machines | 100 | 138 | 178 |
| Total number of compute servers | 100 | 100 | 100 |
| Total number of nodes | 200 | 238 | 278 |
| Total usable memory (TB) | 54 | <div>Sign up for email updates</div> <div>Database, Container, DevOps Insights Delivered to Your Inbox</div> <div>Subscribe now</div> | |
| Total usable cores | 4000 | | |

Conclusion

We used sample server pricing for both the converged and decoupled mode (using thinkmate.com) for configurations that the financial company wanted to use. We find the decoupled architecture leads to a **17% decrease** in total hardware cost and **22% fewer nodes each year**. This results in significant cost saving for the customer.

About the Author

Deba Chatterjee, is part of the product management team at Robin Systems. Before joining Robin, Deba was a product manager in the Oracle database product management team . Deba has extensive experience working with multiple database and big data technologies.

Deba has also worked for the performance services team in Oracle Product Development IT, where he was responsible for the performance of large data warehouses. He has previously worked at Oracle Consulting, Michelin Tires in Clermont-Ferrand, France, and Tata Consultancy Services. Deba has a Master’s in Technology Management – a joint program by Penn Engineering and Wharton Business School.

Sign up for email updates

Database, Container, DevOps Insights Delivered to Your Inbox

Subscribe now