

AlignAlytics benchmarks Elasticsearch, sees 200% performance improvement

bigstep[®]

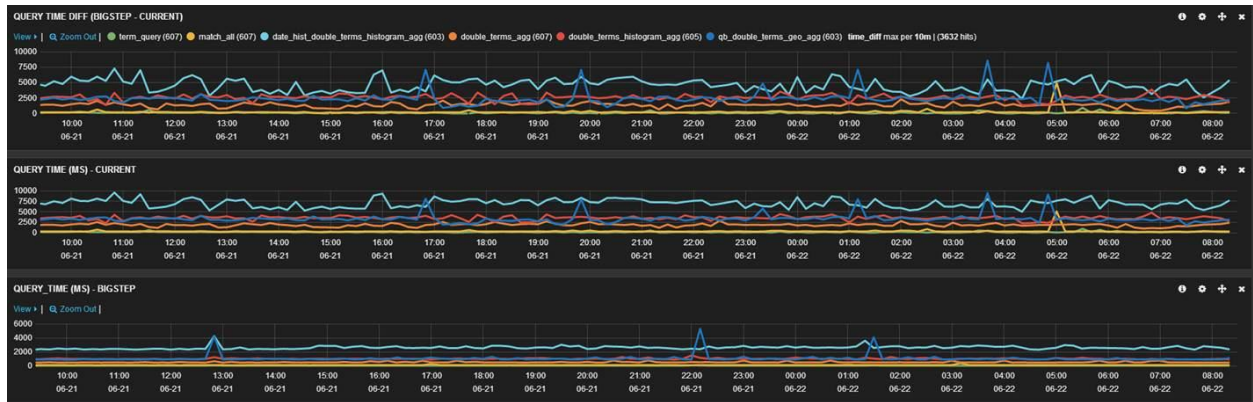
Running queries in no time with Elasticsearch on Metal Cloud

Every so often we come across a use case that makes every hour of work put into our Metal Cloud worth it many times over. Today we're in the happy position of sharing one of those use cases with you. **AlignAlytics** ran Elasticsearch queries on 10 million documents (approx. 4 GB of compressed data) and consistently saw with Bigstep a 100-200% performance improvement over their existing dedicated servers. The specs on their machines were quite similar to those of our Metal Cloud Compute Instances, which makes this one of the closest "apples to apples" comparisons we've done.

In fact, here's what the AlignAlytics team had to say about it:

"We were expecting better performance in the bare metal infrastructure compared to traditional cloud based dedicated servers, but it was incredible to see that performance was twice as good throughout and in some cases even better when dealing with highly complex queries like geo distance calculations."

Amit Talhan – Senior Developer at AlignAlytics



[Click to enlarge](#)

Test results

Data size: 10 million documents, approx. 4 GB with Elasticsearch compression.

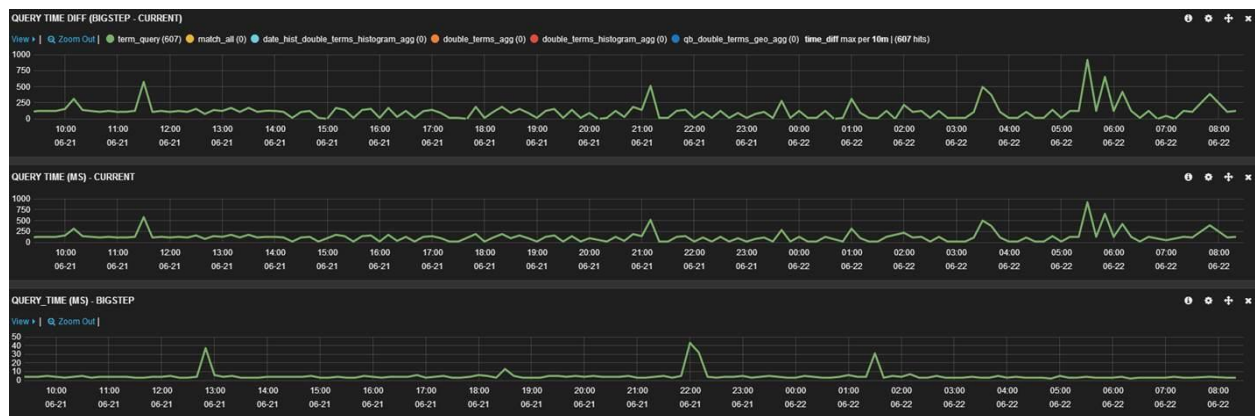
Existing AlignAnalytics cluster

	Node 1	Node 2&3	Node 4
ES Allocated RAM	6 GB	6GB	6GB
Total RAM	8 GB	8 GB	8 GB
Disk	750 GB (SATA)	250 GB (SSD)	250 GB (SSD)
CPU	Intel Xeon E3-1230 3.3 GHz (4 cores, 8vCores)	Intel Xeon X3440 @ 2.53GHz (4 cores, 8vCores)	Intel Xeon E3-1230 3.3GHz (4 cores, 8vCores)
Data Node	No	Yes	Yes
Search Node	Yes	No	No
Network Speed	100mbps	100mbps	100mbps

Metal Cloud Cluster

	Node 1 - 4
ES Allocated RAM	6 GB
Total RAM	16 GB
Disk	200 GB (SSD)
CPU	3.3 GHz, 4 Cores
Data Node	No
Search Node	Yes
Network Speed	4 GbE ports

Multiple Terms Search



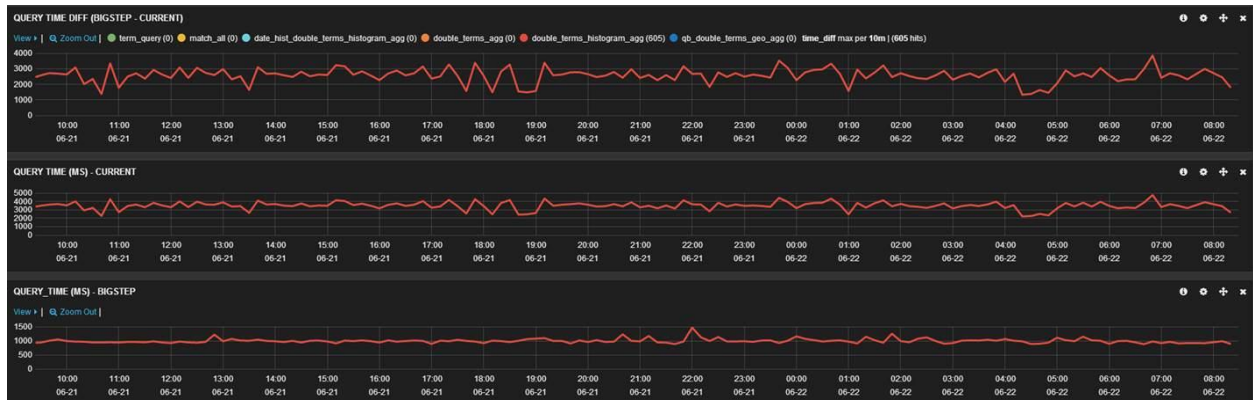
[Click to enlarge](#)

Multiple Terms Aggregation



[Click to enlarge](#)

Multiple Terms Aggregations and a Numeric Histogram



[Click to enlarge](#)

Multiple Terms Aggregations and a geo Hash Aggregation (precision 5)



[Click to enlarge](#)

Four Tier Aggregation with Date, Term, Term and Numeric



[Click to enlarge](#)

The Story behind the results

"As an analytics solutions provider our team of data scientists performs various types of analysis on a variety of large amounts of data. This deep and wide-ranging analysis is what facilitates our discovery of actionable insights for our clients in order to solve their most critical business challenges and enable confident decision making. To be able to fulfil these analysis requirements and deliver the best results, we had to move away from traditional SQL to unstructured data, where Elasticsearch was best suited. As the data size and complexity of the queries increased, it was clear to us that infrastructure mattered and we needed to ensure the best performing setup for running our Elasticsearch cluster. This led to the performance benchmarking exercise which confirmed that Bigstep's Metal Cloud can provide more than twice the performance of regular dedicated servers and therefore empower us to better execute our analysis and more rapidly deliver valuable insights to our clients. "

Amit Talhan – Senior Developer at AlignAlytics

Because the results were consistently 100-200% better than their existing infrastructure, AlignAlytics's technical team came back asking for an explanation. They might have expected that in a virtualized environment, where hardware is oversold and there are noisy neighbors. But they were working with dedicated servers specifically to avoid those problems and they were using SSD local storage to avoid any I/O bottlenecks. So how could a bare metal cloud provide more performance than dedicated servers with local SSD drives, they asked.

Here are what we consider the usual suspects responsible for the difference in performance:

- **Wire-speed network**

Our wire-speed bare metal network ensures that clients have the smallest physically possible network latency – as all switching happens at the hardware level. This means that connectivity between machines and to the storage is excellent, so much so that even working with local disks might not compensate for the difference

- **Hand-picked components**

Even with hardware, components are not created equal. Memory frequency can vary greatly and, although usually underestimated, takes quite a toll on performance. Up to 20% more performance can be achieved from the same setup, simply by increasing memory frequency as shown in one of our previous [performance benchmarks](#).

- **Hall-SSD storage based on enterprise drives**

As in the case of memory, not all SSD drives perform equally. For instance, lower end drives provide good performance for reading but not for writing. In fact, it is well **documented** that writing to SSDs can be quite slow. That's why even some SSD based systems can achieve sub-optimal performance overall.

The conclusion

The main takeaway from **AlignAlytics** findings is to never take anything for granted. Especially due to the cloud's pay-per-hour billing model, it has become affordable to test several providers and setups before deciding where you want to invest your infrastructure budget. Of course these tests take time and these comparisons aren't always like for like. But, if nothing else, you'll have a much better understanding of the strong and weak points of the system you're building. That's very precious knowledge when you find yourself having to scale or having to predict infrastructure costs realistically.

As we found in our testing with AlignAlytics, not everything labeled SSD really improves performance, local drives aren't always better and what's apparently the same 8 GB of RAM can perform very differently across providers. Nothing compares to getting your hands on a setup and testing it with your applications.

If you have any questions, don't hesitate to contact us at hello@bigstep.com