

Infinitely Scalable Clusters

GRID COMPUTING 101

Christine Johnson (Global Head of Business Development): We wrote this article back in 2016 titled “Infinitely Scalable Clusters: Grid Computing 101” - How has grid computing evolved since Hentsū brought public cloud to market back at that time?

Marko Djukic (CEO): It's still a huge topic, if not even more so. Data has been at the forefront of many firms, and there are many challenges with analyzing and pulling out answers to solve these challenges. We definitely see it as an evolving landscape and an ever-increasing one as well.

It is becoming even more widespread and accessible, barriers to entry are coming down. Naturally, there is a growing need for understanding the data, which opens up new ways of processing data, managing it, adding more data on top, and handling it.

CJ: The article mentions technologies like MATLAB, Datalab and BigQuery, are these technologies still front and center?

MD: There is a fundamental shift that's happened in the last few years, and that's the big push for serverless computing. It's basically all about serverless. Code that is distributed across ephemeral compute that handles the analysis and then churns out the answers without having to deal with what's actually the underlying compute.

CJ: Let's talk about platform as a service, this is something that really was born in the public cloud. How have you seen the Financial Services sector adopting this model over the past few years? And what's coming next?

MD: Yeah, the public cloud has seen a massive jump in 2020. What we've seen is clients experiencing another big push, which consists of putting PaaS first as their cloud adoption strategy. Of, course, that has its pros and cons. The obvious pros are that this is a quick entry and access to some pretty powerful tools. Thanks to that, things are made easier at the beginning, so all the building blocks are there. Workloads can be handled

within a day or two, rather than wasting too much time to set up the foundation. The downsides that need to be considered are the platform lock-in, because obviously the PaaS offered by Microsoft will be different from the one offered by Amazon or Google. The questions regarding around portability become important – whether it’s portability for technical reasons or for business strategy reasons and trying to figure out what balance you should have of locking vs. functionality is key around choosing PaaS.

Putting the traditional private cloud completely out of the picture. There is a bit of a hybrid cloud that is still out there; this entails some of the cloud providers offering an extension of their cloud back into the private data center. We’ve seen that both with Microsoft and AWS. This allows you to run your workloads in your own premises as if they are part of the public cloud. It’s giving you the same API, the same functionality, but if you want all that data computing in your data center you can still have it. Mind you, that’s a minor request, we only have one client exploring that.

CJ: For those less technical folks out there, can you help us understand the difference between PaaS and IaaS? How have you seen the adoption of IaaS?

MD: IaaS is seen as a lift and shift, so the power and the evolution of PaaS has taken center stage. IaaS comes “only as a last resort.” The obvious difference is that level of managing is greater; you’re getting into the OS and you’re getting into the actual virtual machine that needs to be managed, secured, upgraded, patched and scaled with demand. It’s an overhead that nobody really wants that and the end of the day.



The key thing that we see around the technologies is the serverless approach. It’s basically all about serverless. Code that is distributed across ephemeral compute, and also we’ve seen that handles the analysis and then churns out the answers without having to deal with what’s actually the underlying compute.

CJ: Between the three major players: Google, AWS & Azure, who is the best suited for big data?

MD: That’s a really tough question to answer, it boils down to individual use cases. There is a lot of similarity between the three platforms, but there are a lot of nuances and differences. Of course, we have to consider the type of data, the type of processing workloads, the cost, the compute availability and it can even come down to specific GPU and CPU sets we want to see utilized – maybe it needs to be available in limited fashion, or not available at all for some platforms or in some regions. That can really steer decision whether to go one or the other.

PaaS functionality that can be leveraged on top. Mobility is also a huge factor. We’ve had solutions where we had to deploy Kubernetes clusters across multiple cloud providers. That has been an enabler to the mobility and flexibility and the freedom to pick where you want to run your workloads. The biggest challenge is data handling. Massive data sets are not easy to move around between cloud providers – that needs forethought on where it needs to be located and what kind of compute can access it. Because once you shove hundreds of terabytes into one cloud provider and move it to another, you know it can be done, but you don’t want to be doing that on a monthly basis.

CJ: So, the foreplanning is very, very important at this stage.

MD: Yes, definitely.

CJ: We can't not bring up COVID...how is remote working impacting research processes & workloads with everyone dispersed and away from their office servers and networks?

MD: Hm, I'm not sure how directly relevant COVID is to data remote working. However, it does highlight the cloud's enablement of "work anywhere" mentality for staff as well as shifting a lot of the processing from the desktop to cloud services.

CJ: Can you dive in a bit to what data challenges Hentsū clients are specifically facing right now and how we're helping to solve them?

MD: Sure, there are various challenges involved, although from our perspective, we basically have two approaches. The first one is how Hentsū are handling our own internal data. The second is how we are handling things for client solutions?

Speaking of internal data, what's really encouraging is adopting a lot of the same data science methodologies for some of the mundane workloads we have internally; everything from billing, to support and CRM and internal reporting. Everything is quicker and better, and that's something that has rapidly grown within our own ecosystem and sphere of data.

As for client solutions, that comes down to individual data sets. Some of the interesting challenges with breaking down the barriers, and the data handling, has created another problem, in terms of what you pick. Do you go with Google, or Amazon or Microsoft as your platform? How do you figure that ahead of time so you don't get stuck with a choice that is not the most reliable one? Picking that apart and making the choice on what is the valid direction for the underlying challenge they need to solve.

CJ: You've mentioned many times that the vision for Hentsū is to bring technological agility to financial services through whatever means necessary, right now it just happens to be through the public cloud as that's the most powerful and equipped technology on the market to do the job. How do you see big data and data management in general growing with the public cloud evolution and is it here to stay?

MD: Well, the public cloud evolution is definitely here to stay, or more that concept of utility computing. There are some much more advanced data science concepts coming from quantum computing, which even 12 months ago seemed so abstract, yet only a few months ago is already materializing with services such as AWS Bracket.

CJ: If a client just wanted to test the waters with research in the cloud, what would be some easy projects you'd suggest to just get going?

MD: We've found that there is always a few internal projects which can be easily made into pilots for testing the waters. Those don't necessarily need to be massively complicated. We've done pilots in days or weeks, and it's a valid use case that can test how it can work and what benefits it can bring to the firms. That step up is very small, relative to how it was a few years back. There's no massive CapEx, there's no massive spend – if it doesn't work out, you are not stuck with it, you can turn it off and choose another path. Data and analysis are often the best candidates, and this can be anything from risk calculations, to portfolio analysis.