

# Google Cloud Platform



## Case Study

“Quest for a better investment platform leads to Google Containers, BigQuery and Sheets”

## Presentation Download

- SenTai Canaccord Quest Migration – EN (Google Slides)
- SenTai Canaccord Quest Migration – EN (ppt)
- SenTai Canaccord Quest Migration – EN (pdf)
- SenTai Canaccord Quest Migration – FR (Google Slides)
- SenTai Canaccord Quest Migration – FR (ppt)
- SenTai Canaccord Quest Migration – FR (pdf)

## Introduction

As a case study for Google Cloud Platform, we highlight the use of this cloud provider for the development and hosting of the Canaccord Quest cloud computing and web app we developed.

## App Production Cycle

The application is based on a production cycle involving:

- Core Valuation Model: a core tree-like computational model represented in a vast **Google Sheet** with 80+ stages both at the company level as well as the Aggregate level (Industry & Geographical region).

- Data Import: Raw Financial Accounting Data from S&P
- Data Integrity, validation, correction, filling, restatements and approximations
- **Google Sheets**: expression of the 7000+ business rules in custom languages and rules used by a team of 10+ business analysts working on the same model (possibly simultaneously).
- Engine: A Java engine parsing the model rules from **Google Sheet** and dynamically compiling to raw bytecode for maximum performance
- Data Storage Optimization: for multi-dimensional data storage and high performance access.
- Web: 300+ dashboards, charts and dynamic searching within the billions of results done with AngularJS and CoffeeScript.
- Dynamic **Google Sheet**: dynamic representation of pre and post calculation data using our API.

Each one of these components had to be optimized both on the software side, as well as on the infrastructure side, taking into account the various stages and their requirements in terms of RAM, CPU and Storage. The ease of deployment and dynamic rescaling of these cloud ready options made the management of the web application extremely agile.

## **Persistent Storage and Application Performance**

We have extensive needs for SSD persistent storage during the computational phase to optimize the performance of the application. Thus the ability to gradually ramp up the SSD persistent storage to 4Tb per machine has proven key to reduce our overnight app run time by a factor of 2.5. What once required 10 hours+ for significantly less data, we can now compute in the new platform in less than 30 minutes (see

Performance & Scalability Gains for more details). The time increase not being linearly proportional to the number of companies, we hope to be able to compute up to 50K companies overnight, on a single 32 core CPU VM instance with 200+GB RAM.

## Collaborative Challenge

See Google Apps.

## Google Cloud Platform Tools

- Google Big Query (we dump everything in Big Query so the BA can play with the data without the dev team. We were very surprised by the speed at which it would ingest our huge amount of data, 100m+ rows would take just a few minutes)
- Google Cloud Storage (we store our backup there, which is important as a single run produces 200+ Gb of data)
- Google SQL (we tried it, but the sheer size of our data was not a good fit)

## Big Data

Using Google Big Query allows us to easily connect third party Big Data Analysis tools such as TableauSoftware. We are able to pull from Big Query the entire universe of companies (ie up to 10K companies), thousands of Metrics, across 20 years of historical Data. A huge three dimensional Matrix, manipulated at will via powerful Big Data software. A key ingredient to analyzing the Quest® valuation engine calculations, spot errors, and perform post calculation statistical analysis. Setting up Google Big Query was simple and extremely useful.

# Technical benefits

## Performance & Scalability Gains

In the legacy system, a full VE Run took 10 hours to manage 2,500 companies, way fewer industrial sectors (only the first two MSCI Industrial sector levels), as well as a limited number of geographical regions (10). In the new system, we are running 8,500+ companies, the full 275 MSCI industrial sector classification, as well as 25 different geographical areas. Finally the model accommodates for a much greater number of company and aggregate metrics. All done in under 30 minutes. The performance and scalability gains are staggering (more than 200x).

## Price Competitiveness

Frequent price drops and most competitive offerings are driving the whole industry. Commoditization is driven by Google's know-how, and benefits the end users such as Canaccord. We've been able to start small, and scale up only when ready. The benefits compared with previous hosting providers is substantial. Canaccord was used to in-house infrastructure, with a very slow pace of change: they were amazed at the speed at which we could create/delete/update instances to meet our requirements. The affordability of both GCP and GA4B made them the right combination for the dev team, as well as the business team.

## New Feature releases

Frequent release of new useful features. The remote developer team often meets and works in a business center where the routers are misconfigured: our security tools on our servers would ban us (fail2ban). The GCP Web ssh console

was of HUGE help at such times.

## Google cloud console

Very nice and effective tool, everything is easily toolable, very dev oriented. Very good point. The local *gcloud* tool is also very simple to use, the headaches of security issues are taken care of for us. The first steps in the GCP world were very smooth. (Very far from the experience we had with other vendors).

## Docker support

Our application is decomposed into multiple docker containers. Deploying Docker containers on Google Compute Engine is very easy and was available very early. It was one of the key factors in choosing Google Cloud Platform. What works on our development machines is exactly what is deployed in production.

## Elasticity during development

We started with small machines during the initial phase of development, and as the product was growing, we could easily evolve from small instances to bigger ones.

We kickstarted the project using a single dual core cpu dev instance with a small hard drive. We now have multiple infrastructure (test, UAT, prod) all using SSD for optimal I/O.

That elasticity comes with no downside. We did not have to reserve any instance nor did we have to plan for future needs. We simply increased capacity whenever necessary.

We really changed the way we work with our servers. We delete servers every day, create new ones with different

characteristics to test our architecture choices. We could try on-disk persistence, fully managed SQL, Cloud Datastore, Cloud Storage, to end up with a mix of those which fit our needs. The whole process is tried in an iterative manner, with no strings attached.

## **Fully automated deployment with the API**

Everything from instance provisioning to deployment is automated, thanks to a combination of Google's Cloud SDK and Ansible. Our previous platform had no API that we could leverage. That is a "must have". Going from zero server to a fully running environment is one script away. Updates to the production are as simple as sending a chat message to our Hubot.

## **Fast VMS**

We use 32 core VMs with 200+ Gb of memory: the Quest® valuation engine is very CPU intensive. The virtual machines that GCE offers are really fast. You can really compare their performance to a high end on premise server.

## **DB in managed PaaS**

It is great to be able to use managed SQL and NoSQL instances in the form of a PaaS, and not to have to worry about managing these DB instances, and backups. The hybrid PaaS-IaaS services available via GCP make it a highly customizable cloud platform.

The sheer size of the data, and performance requirement, forced us to create our own custom solution with zero copy – flow down to disk, to squeeze maximum performance.

## **Google Sheets (via REST APIs) integration w/ GCP**

Heavy use of **Google Sheet** worked on collaboratively, and imported upon every run of the web app via API.

## **REST APIs queried in Google Sheets integration (client side)**

Heavy use of our own REST APIs called upon via a Google Managed Library of Google Scripts (js) into GS allows us to efficiently provide third party analysis tools to the Business Analysts.

## **Backup and disaster recovery**

We have multiple strategies to ensure the durability of our system, we take snapshots of our remote drives, make backups of our data and copy them to Google Cloud Storage using specific lifecycles.

We also use the Google Cloud Storage rsync capability to have a live mirror of our most important files. We can therefore have multiple levels of disaster recovery.

## **Conclusion**

The hybrid PaaS-IaaS GCP, although quite recent compared to industry standards such as Amazon, or various PaaS, provides us with all the tools, benefits, and pricing advantages which could and should be made available to Cloud platforms. It is redefining cloud industry standards.

GCP, with the collaborative use of GA4B, is a magical combination which enables us to create a vastly improved

equity valuation platform, together with an unrivaled agility and ease of use, whose benefits are passed on to both end users and the internal Canaccord business analyst team.