

COMPUTE THE FUTURE

SINGLE SOURCE OF TRUTH

CASE STUDY

COMPANY PROFILE:

Fortune 500 medical device, merger and acquisition, sales based organization

REVENUE:

> 10 Billion USD

EMPLOYEES:

> 30 000

**LARGE ACQUISITIONS,
OF DATABASES:**

> 20

CHALLENGES WE TOOK

QUESTION 1:

Can we ‘de-silo’, bring all sales transaction data from all of our acquired division data bases into one ‘Single Source of Truth’, able to identify (match across) all customers overs all divisions and begin to predict the next sale / transaction, as well as, identify other opportunities?

QUESTION 2:

Furthermore, can we also cross-compare the latter ‘Single Source of Truth’ (continue to scale it) to our sales pipeline data (CRM) and on-site inventory (customer) data, further leveraging this information for future questions / optimizations?

OFFERED SOLUTIONS

‘Big 5’ Consultant (project withheld only to Q. 1)

- COST:** Bid ~40 million USD, not inclusive of third party licenses, vendors and cloud computing platform fees to operate solution going forward.
- TIME:** Ask 2 years.

Top 3 Cloud Provider

- COST:** Cumulative exponential cost growth curve for foundation-ally needed compute resources. Additional “platform” and API costs not included. Organization required to hire and train or outsource a development team to use cloud product, such as the latter consultant
- TIME:** Unknown. During sales cycle, it took a team of 6 client acquisition specialists 5 days to copy 2 raw database dumps to cloud storage, further products/platforms were then up-sold.

CTF (Compute the Future)

- COST:** Adjustable 150,000USD/month adjustable to needs subscription, covers statement of work pipeline development and teaching of existing data facing personnel within their organization. One-time, tax-deductible 600,000USD infrastructure fee. No licenses or axillary fees, 100% Open computation from the data center floor to an answered question.
- TIME:** 90 day sprint for 80% of data in Question 1, 30 days per additional silo/division thereafter, including data sets from Question 2. Proved up-front the core work and cost-to-scale of Question 1 in person during a 3 day sprint.

VALUE DELIVERED

- ✓ Organization now owns 100% of their solution.
- ✓ Robust, better than commercial cloud up-time, clusters (2x) peered and deployable to new specifications within minutes. 100% of compute resources directed for the organization versus cloud shared across many organizations. This added 60% efficiency utilization of machines.
- ✓ Automatically updating from every silo, distributed time-series and complied files, encompassing all data across all divisions.
- ✓ Core graph computations for multi-level fuzzy and direct matching across all clients, all divisions.
- ✓ 100Gb/s distributed-memory, global convex optimizations focused on prediction of new transactions and new questions.
- ✓ Distributed quality control jobs, cleaning, standardizing and checking their inbound data.
- ✓ Data facing personnel are now able take new questions from hypothesis testing to production at no extra cost, building their own Machine Learning Pipelines, using any compute framework, with CTF empowering/teaching in the background.
- ✓ A cumulative linear cost-to-scale all technology needs going forward.
- ✓ Best known (time to produce and compute efficiency) statistical/analytical dash-boarding (reporting) services monitoring data quality and top-down questions regarding data sets and ML Pipelines.
- ✓ Machine Learning Pipelines (compute jobs) portable to any 3rd party cloud on-demand, if ever needed. This gave the organization a strong negotiation position for excess compute time, complete freedom.