

Not fit for purpose – Machine-learning and the cloud



Posted on Tuesday, September 19, 2017

Most businesses have come to terms with the fact that big data will be crucial to identifying and capitalizing on new opportunities. An industry-wide shortage of experienced data scientists, and the sheer volume of information that needs to be processed in real time, means that big data systems must be automated to cope.

Key to this automation will be machine learning, training systems to develop and apply algorithms to the information being collected. Well-configured machine learning can be used to compensate for a lack of human operators, and to (eventually) process information faster than a human – in real time.

Unlimited storage – but it's all too slow

As part of continued efforts to drive down capital spend and increase operational agility, many businesses are adopting “cloud first” strategies. Instead of upgrading on-site infrastructure, they are choosing to build new systems on hosted cloud platforms, taking advantage of unlimited processing and storage capacity.

Unfortunately, the cloud and machine learning are not fully compatible. The cloud is designed to store infinitely large data sets, and can even provide the processing grunt required for machine learning “epochs” (each cycle used to train and tweak the automated algorithms).

But when it comes to real time processing, there is a serious problem – network latency. Uploading data to the cloud

takes time. And when dealing with potentially hundreds of gigabytes of data being generated from IoT sensors every day, uploads have to be completed in batches. Which means that real time analytics are impossible.

The on-site data center lives again

The reality is that real time machine learning and big data analytics is still best performed locally. There are obvious cost implications – increased processing provisions, additional storage, specialist skill requirements – but the benefits of real-time will help to justify the spend.

Indeed, machine-learning epochs can only really be completed adequately on-site. A return to batch processing as demanded by cloud bandwidth constrictions will only serve to reduce the potential benefits provided by real time analysis.

The good news is that some of these additional costs can be offset by re-deploying existing storage assets. Post-warranty storage arrays provide the perfect base for an in-house private cloud platform for use with your machine-learning program, and without the latency issues inherent in public/hybrid cloud deployments.

Which means your big data program may cost even less than originally thought.

To learn about extending your data center using post-warranty storage solution – and securing the support you need – please [get in touch](#).