



CIONET

DISCUSSION SUMMARY

FEEDING THE BEAST: PROGRESSING FROM CPU AND GPU TO IPU PROCESSORS

This article was written by Roger Camrass, of CIONET UK. Stephan Ehman of Hitachi Vantara and Jonathan Martin of WekaIO participated as sponsors of a virtual wine tasting event held in November 2021 on the topic of 'In a data-driven world, time to market is everything'

Providing a context for the discussion

Jonathan Martin described two mega trends that are converging rapidly. The first is the dramatic improvement in processing power with the widespread adoption of Graphic Processing Units (GPUs) and the recent introduction of Intelligence Processing Units (IPUs). Concurrently, there has been an explosion in object-orientated data induced by the widespread use of image and video devices such as cameras on the Tesla car. Providing the input, processing and output channels to ingest, model and act upon such data in real time is one of IT's major tasks.

Every firm is at a different stage of its data journey here. Hitachi Vantara invited a group of CIOs from a range of different sectors to discuss their approaches to high performance digital architectures, with a focus on the use of AI to harness object related data assets. The roundtable event illustrated just how different such a compute environment has become compared to traditional systems of record.

How to feed the beast

According to delegates, current IT architectures have evolved around large scale transaction processing that ingests structured data from both internal and external sources such as mobile banking and point of sale transactions. ERP is a good example of a traditional processing environment. Densities are light in comparison to non-structured data such as image and video.

The explosion of non-structured data calls for radical changes in the way IT architectures may evolve, especially where super computing techniques combined with AI are needed. In the words of Jonathan Martin, modern digital architectures must be designed to ingest data at rates of tera and petabytes, in real time. Thousands of GPUs are needed to model this data using AI techniques to produce vital insights in cases such as drone activity and autonomous vehicles.

This introduces a new dimension of choice in architectural design. Should processing take place at the core (e.g., on premises), in the cloud or out at the edge

(e.g., embedded in an intelligent product). Next generation architectures will need to combine on-premises, cloud and edge-based computing in a seamless fashion to cope with a truly digital environment.

Which sectors are most advanced in data ingesting and AI modelling?

The sectors represented during the event presented a wide range of applications and experiences:

For an international mining company, drones are being used to supervise activities on the ground in real time to prevent accidents and to optimise production. The delegate described this as 'mining from the sky'

For a global energy company, large volumes of data are being extracted from drilling platforms and processed at high performance computing hubs. Low orbit satellites will facilitate petabytes of data transfer in the coming years.

For a global insurance group, super computing and AI techniques are in frequent use to assess risks and to monitor IT infrastructure. The company is already examining Quantum as the new generation of computing.

For a global defence and aerospace group, AI has become the key tool in the war against cyber-attacks. AI is also used to help frigates make informed decisions about enemy movements. Drones are employing edge-based processing techniques.

For a global pharmaceutical company, the need to compress drug approval times from ten years down to ten months during the pandemic has required massive data transfers between the manufacturer's R&D teams and the Federal Drug Administration.

Where do other sectors sit today?

Financial services such as banks and stock exchanges rely heavily on external data sources for much of their work. However, such data is typically structured, for example Bloomberg market data. One such exchange is moving to the cloud and seeking new ways of 'landing and integrating' the data from multiple sources.

Two potential areas for AI that were discussed within financial services included high

rekuensi trading and fraud detection. One banking delegate mentioned that fraud is providing a promising area to pilot AI modelling applications.

Construction firms are beginning to build giant data lakes, especially in mega projects such as HS2. The main challenge is to integrate data from hundreds of legacy applications.

A media company mentioned that it is only at the early stage of its AI journey. At present it is experimenting with AI as an advance on current Business Intelligence (BI) tools.

Challenges on the AI journey

Jonathan Martin described two challenges to be met on the AI journey:

Challenge 1 – how to experiment with new AI techniques and build to enterprise scale. This will include a transition from giga and terabyte data rates to hundreds of petabytes. It will also require a move way from CPUs to GPU and IPU processing units

Challenge 2 – AI modelling presents a very different set of workloads to traditional applications such as ERP. These new workloads require massive movements of data through input and output (IO) channels

The main question facing digital leaders is where to place different workloads, both historic and current in the emerging hybrid world of on-premises, public cloud and edge-based computing to optimise cost, complexity an speed.

What can we take away from the event?

CIOs need to recognise the growing duality of data processing environments and build the appropriate high-performance platforms and tool sets to cope with both traditional and modern workloads. In the view of Jonathan, WekaIO and Hitachi Vantara can assist in the seamless integration of data ingestion and analysis across vastly different computing terrains.



About CIONET

CIONET is the leading community of more than 10,000 digital leaders in 20+ countries across Europe, Asia, and the Americas. Through this global presence CIONET orchestrates peer-to-peer interactions focused on the most important business and technology issues of the day. CIONET members join over a thousand international and regional live and virtual events annually, ranging from roundtables, programs for peer-to-peer exchange of expertise, community networking events, to large international gatherings. Its members testify that CIONET is an impartial and value adding platform that helps them use the wisdom of the (IT) crowd, to acquire expertise, advance their professional development, analyse and solve IT issues, and accelerate beneficial outcomes within their organisation.

cionet.com