

WHITE PAPER

Analytics Workloads for the new Era of AI Applications

Sponsored by

CLOUDERA

DELL Technologies

intel

Business Trends and Challenges

Data is fundamental to creating efficient business operations, discovering new revenue opportunities, and delivering exceptional customer experience. As the world becomes increasingly technology-centric, with an ever-growing number of data sources and streams, enterprises are facing the challenge of making sense of the staggering volume and diversity of data using traditional data processing application software. One effective and most proven way of dealing with such large and complex volume of data is with the help of modern Big Data Analytics workloads solutions using Artificial Intelligence (AI) techniques, implementing Machine Learning (ML) algorithms. In fact, enterprises across the world are embracing an “AI-first” approach to mature their analytics capabilities and become more data-driven.

Big Data workloads and ever-growing data require big compute and storage. Many organizations equated that need to public cloud. The public cloud allowed for high availability and rapid scaling of compute resources to enable Big Data AI/ML applications while reducing infrastructure complexity and enhancing business agility.

Many organizations discovered that relying solely on the public cloud can get expensive fast. With advances in compute, storage, networking, and containerization technologies, running Big Data analytics in the public cloud is no longer the de-facto option. Sure, for small-scale applications (less than 1 petabyte of data), the public cloud typically presents a viable and cost-effective option. However, large-scale environments with many petabytes of data can benefit more from an on-premises solution.

The trend most prevalent in the industry today is the adoption of a hybrid cloud approach. According to 451 Research, 85% of enterprise leaders would like

to be able to run their analytics workloads in multiple cloud/datacenter environments¹. These findings are echoed by IDC, which predicts that by 2022, over 90% of enterprises worldwide will be relying on a mix of on-premises private clouds, public clouds, and legacy platforms to meet their infrastructure needs². Why hybrid cloud? Because it provides the optimal solution to universal pain points experienced by enterprises worldwide.

Customer Pain Points

Business

“Business” pain point refers to the problem an organization is trying to solve using Big Data Analytics. The pain point here is how to deploy data and analytics to more end-users to drive more value from data faster – to innovate and differentiate.

Financial

Once the analytics to support business use cases are figured out, an organization needs to decide where it will run its AI applications to maximize cost savings. Business imperatives drive the most optimal mix of infrastructures. Traditionally, the options were the public cloud or on-premises data center, with organizations trying to figure out the trade-offs between CAPEX and OPEX.

Risk

Regardless of where deployed, the deployment should ensure the organization as a whole remains within the boundaries of corporate, industry, or regulatory compliance. Enterprises need to weigh the risks of implementation complexity, cloud vendor lock-in, data and system security, data locality/sovereignty, resource contention, and governmental compliance.

References:

1. 451 Research, The Future of Enterprise Data and Analytics is Hybrid, <https://www.cloudera.com/content/dam/www/marketing/resources/analyst-reports/the-future-of-enterprise-data-and-analytics-is-hybrid.pdf.landing.html>
2. IDC, IDC Expects 2021 to Be the Year of Multi-Cloud as Global COVID-19 Pandemic Reaffirms Critical Need for Business Agility, 31 March 2020, <https://www.idc.com/getdoc.jsp?containerId=prMETA46165020>

Solution – Cloudera/Dell/Intel Private Cloud

The good news is that there is a solution to these business pain points, and that solution is private cloud. The private cloud reduces the financial burden of the public cloud, while also addressing the shortcomings of a monolithic, on-premises data center. For example, the monolithic data center approach introduces resource contention and upgradability challenges, which lead to low resource ROI, increased risk, operational inefficiencies, and “long time to value” for deploying new analytics use cases. If only it were possible to turn your enterprise data center into a private cloud – now it is!

In partnership, Cloudera/Dell/Intel provides a solution for running Big Data AI/ML analytics workloads in a private cloud. This joint solution enables the flexibility of public cloud on-premises while also providing a stepping stone for enterprises to adopt hybrid cloud architectures in the future. The solution consists of Cloudera Data Platform (CDP) Private Cloud running on Dell PowerEdge servers interconnected with Dell EMC PowerSwitch hardware for cluster networking. Compute is provided by Intel CPUs, featuring Intel's latest technologies supporting complex AI and data analytics applications. Additionally, the orchestration layer is handled by OpenShift containerization technologies to enable application and management flexibility, high availability, and resource optimization. The joint solution benefits are:

- Faster time to value
- Improved cost
- Reduced risk

Ultimately, Cloudera/Dell/Intel's private cloud platform enables agility and flexibility of the public cloud while also providing the security and control of the enterprise data center.

Use Cases for Cloudera CDP Private Cloud running on Dell/Intel

Financials – Anti-Money Laundering (AML)

Financial crime is on the rise. As strategies of attack evolve over time and risks become increasingly sophisticated, banks and financial institutions are under mounting pressure to develop intelligent risk-control systems. A rising area of concern is money laundering.

Detecting this kind of fraud requires sifting through a massive quantity of financial transactions and using AI/ML to establish patterns. China UnionPay was one of the early adopters of Intel's technology-based intelligent, artificial neural-network risk-control system. Running a platform based on Cloudera and Apache Spark, China UnionPay demonstrated the value of using machine learning to drive proactive, highly efficient, and accurate risk identification and mitigation workflow, boosting the accuracy of AML detections by 60%.

 **Success story reference - [UnionPay](#)**

Sales

Big Data AI analytics is revolutionizing the field of sales. Increasing the quality of sales leads, providing insights into which content is most effective at each stage of the sales cycle, inventory planning, product review sentiment analysis, store foot traffic optimization, revenue forecasting – the list goes on. These are all use cases where Big Data is making a contribution to the world of sales. Shoppermotion is another successful use case that leveraged the power of Cloudera, Dell, and Intel. They created an IoT architecture for their clients using Cloudera to ingest, store, and analyze streaming sensor data. In each Shoppermotion-supported physical retail store location, small beacons are attached to the shopping carts and baskets. These beacons transmit a signal every second to Bluetooth sensors in the ceiling to track the shopper while in-store. The sensors capture the customers' behavior in real-time – in which aisles they shop, which direction they approach the aisle, and even how long they dwell in front of different items.

 **Success story reference - [Shoppermotion](#)**

Healthcare

A big use case for data science in healthcare is in pharmaceuticals. More specifically, pharma companies rely on AI analytics to find participants for drug trials and help interpret findings. The exponential growth of new data types and data sources are propelling pharmaceutical companies to rethink their approach to data. Pharma organizations receive data from various sources – electronic health records, insurance claims, real-world evidence, clinical trials, patient monitors, and wearables – all of which enter the data ecosystem from disparate providers, facilities, and organizations. These diverse data sets help researchers better understand how diseases develop, uncover patterns

and risk factors based on demographics and genetics, and reveal potential new uses for existing drugs. However, if not properly managed, the same data can create inefficiencies that cause delays in clinical, operation, and market access. IQVIA and Cloudera/Dell/Intel partnered together to help pharma organizations make better use of their data. IQVIA's Cloudera CDP implementation enables complex ML and AI on petabytes of data to deliver actionable intelligence back to the point of care.

 **Success story reference - [IQVIA](#)**

Scientific Research & Development

The scientific community increasingly relies on data science to tackle the world's biggest challenges. NASA's Center for Climate Simulation, or NCCS, is one such example. As the availability and volume of data grow about the Earth, researchers spend more time downloading and processing their data than doing science. NCCS has developed the Reanalysis Ensemble Service (RES) to tackle this problem. RES is a high-performance, Big Data analytics framework built on Cloudera/Dell/Intel to analyze large datasets hosted by NCCS and accessible over the web via REST API. The RES interface supports web service access to consumer applications through a graphical user interface for interactive requests, a command-line interface for users familiar with basic RES commands, and advanced programmatic access for Python-savvy users. RES allows users to run compute close to the data without having to download large input datasets.

 **Success story reference - [NCCS RES](#)**

IoT Enabled Predictive Maintenance

IoT is fundamentally transforming the maintenance model paradigm from repair-and-replace to predict-and-prevent. By capturing and utilizing data streaming from sensors and devices, businesses can now gain visibility into the condition of their equipment in real-time. Using IoT and sensor data, predictive maintenance enables organizations to effectively predict when and how an asset might fail by detecting variances, understanding warning signals, and tracking any patterns that may indicate a potential breakdown. Heavy equipment manufacturer, Komatsu, has implemented an Industrial IoT analytics platform powered by Cloudera to help prevent mining equipment failures. Komatsu's solution allows

customers to optimize machine performance using machine data and AI analytics. It ingests, stores, and processes many types of data collected from mining equipment worldwide, even in remote locations. With Cloudera, Komatsu can now use the vast amounts of data to help its mining customers manage increasing pressure to be environmentally smarter and more productive at a lower operation cost.

 **Success story reference - [Komatsu](#)**

Summary

Enterprises rely on analytics and machine learning applications to gain a competitive edge in today's fast-moving world, allowing companies to innovate fast and differentiate themselves from the competition. The private cloud platform running Cloudera on Dell infrastructure, powered by Intel chips, is the optimal solution for your data-centric, AI-first enterprise. We invite you to reach out to us for a personalized benefit realization session that will showcase how Cloudera, Dell, and Intel allow enterprises to monetize their data by helping enterprises understand their customers, markets, products, and operations.

For more detailed information on how Cloudera/Dell/Intel come together to enable analytics workloads for the new era of AI applications please consider:

- [Cloudera CDP Private Cloud Reference Architecture for AI & Data Analytics](#)
- [Cloudera CDP Private Cloud Base Cluster Hardware Configurations](#)