SOLUTION BRIEF

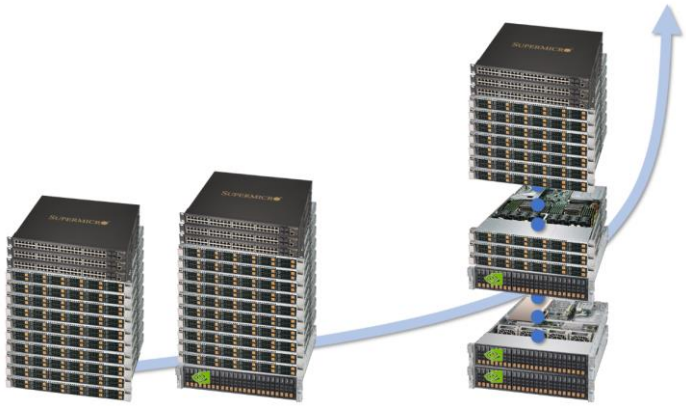
SUPERMICRO, CLOUDERA, & NVIDIA ACCELERATE ENTERPRISE DATA ANALYTICS

How Supermicro Ultra Servers and Cloudera Data Platform Extract Deep Insights from Massive Amounts of Data

TABLE OF CONTENTS

SUPERMICRO

As a global leader in high performance, high efficiency server technology and innovation, we develop and provide end-to-end green computing solutions to the data center, cloud computing, enterprise IT, big data, HPC, and embedded markets. Our Building Block Solutions® approach allows us to provide a broad range of SKUs, and enables us to build and deliver application-optimized solutions based upon your requirements.



Executive Summary

Enterprises have embraced digital transformation to extract insight from their data to improve business processes, increase productivity with the same investment envelope, reduce costs, and create a new offering to generate higher revenue and profit. Across industry sectors, enterprises are collecting more data, increasingly applying machine learning and deep learning to leapfrog their competition and grow faster.

Supermicro has partnered with Cloudera and NVIDIA to create a platform consisting of Supermicro's Ultra Servers, Cloudera Data Platform Private Cloud Base (CDP-PVC-Base), and NVIDIA RAPIDS on GPU. Cloudera provides enterprise-level support for Apache Spark 3.x and other open-source software. Supermicro Ultra servers are powered by NVIDIA A30/NVIDIA A100 GPUs and NVIDIA Networking to accelerate Apache Spark 3.x workloads significantly – 5X faster than CPU based computing.

SUPERMICRO SERVER CLUSTERS SUPPORT CLOUDERA DATA PLATFORM PRIVATE CLOUD BASE AND SPARK 3.X WITH NVIDIA GPU/RAPIDS ACCELERATION

Supermicro server cluster with NVIDIA A30/ NVIDIA A100 GPU delivers scalable machine learning processing using Cloudera CDP Private Cloud, NVIDIA RAPIDS, and Spark 3.x. The Supermicro Super Cloud Composer manages the cluster using Redfish v1.8 and IPMI.

Glossary	
Apache Spark 3.x	Software to build resilient, scalable cluster for machine learning, with programming interface. Spark 3.x provides the latest feature and performance enhancement, adding GPU acceleration
Cloudera CDP Private Cloud Base	Software platform that delivers integrated Apache Spark and other functions with enterprise support
NVIDIA A30	NVIDIA GPU for DL/ML, 3584 CUDA cores, 24GB HBM2
NVIDIA A100	NVIDIA GPU for DL/ML, 6912 CUDA cores, 40/80GB HBM2e
NVIDIA RAPIDS	NVIDIA suite of software libraries, built on CUDA-X AI, to execute data science and analytics pipelines using GPUs.
Supermicro Ultra server	1U or 2U volume server supporting GPU. Choice of 3rd Gen Intel Xeon Scalable Processors and 3rd Gen AMD EPYC™ CPUs.
SuperCloud Composer	Supermicro management software providing a single-pane-of-glass monitoring of server clusters, in-band and out-of-band control, network OS booting.

In addition, the platform can also prepare, cleanse, and feed data into deep learning training pipelines. This solution brief shows the Apache Spark 3.x cluster's benefits and critical components for deployment in enterprise hybrid data centers. The platform accelerates business insights from customer data to optimize the business and deliver higher revenue.

Supermicro Ultra Servers and NVIDIA GPUs

Supermicro Ultra servers are high performance systems that support GPUs. They support either 3rd Gen Intel® Xeon® Scalable processors or 3rd Gen AMD EPYC™ processors, using the latest PCI-E 4.0 interfaces and NVMe drives. Each server supports up to two NVIDIA A30 or NVIDIA A100 GPUs, which run NVIDIA RAPIDS to drive Apache Spark 3.x acceleration. In addition, high performance system memory supporting up to 8TB is available to support the in-memory processing for Apache Spark.

SYS-220U-TNR

SYS-120U-TNR

Additionally, a Supermicro 1U Ultra server may be used as part of the cluster for tasks that are not accelerated with the NVIDIA A30 or NVIDIA A100 GPUs.

Addressing Customer Problems using Insights from Data

Enterprises across all industries have seen the emerging power of analytics and machine learning from their data. For example, credit analysts minimize loan loss provision in the finance industry using analytics and machine learning based on customer data, macro-economic information, and transaction data. Likewise, in media entertainment, marketers can better maneuver through GDPR rules by applying analytics and machine learning to the customer, location, and country-specific information. These examples are early indicators of how machine learning can improve business results.

Table 1 shows examples of different industry sectors, their specific data, and the consequential results from applying machine learning and deep learning to the data.

As part of their digital transformation initiatives, enterprises are collecting more data. Data come from customer interactions, industry information, databases, and even sensors. In addition, enterprises now adopt machine learning and deep learning, as these technologies have matured for business deployment.

Cloudera has evolved its offering for machine learning pipelines in the past decade for enterprises. Cloudera Data Platform Private Cloud Base's (CDP-PVC-Base) broad capability enables implementing the entire data pipeline, from data collection using Kafka, to retention and access on unstructured and structured data, to data analytics running on Apache Spark 3.x.

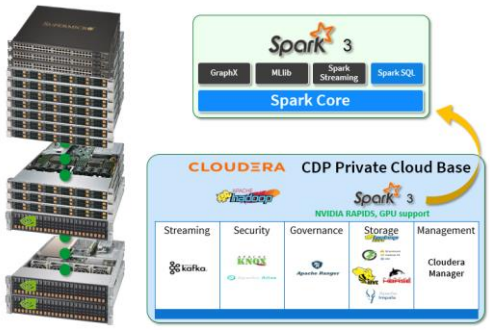
CDP-PVC-Base can also provide data preprocessing to front-end the deep learning pipeline using deep neural networks running in intensive GPU clusters. Using data processing algorithms in Apache Spark 3.x in CDP-PVC-Base, enterprises can optimally clean and prepare data for deep learning training. The resulting deep learning inference models then process new incoming data to make predictions or decisions. Thus, the deep learning system enhances business operations for the enterprise.

Supermicro offers Ultra server clusters running CDP-PVC-Base and NVIDIA GPUs to accelerate Apache Spark 3.x, delivered with Cloudera software. In addition, the Cloudera Data Platform Private Cloud software has incorporated the NVIDIA RAPIDS™ libraries, which enables compute acceleration on the NVIDIA GPU for components of Spark 3.x. Supermicro's Ultra Server Clusters support NVIDIA A30/NVIDIA A100 GPUs, and they are validated for performance, manageability, security, and scalability backed by enterprise-grade support. This platform enables enterprises to increase revenue and productivity by gaining insights from their data.

Industry	Selective Data	Selective Results
Finance	Asset prices, policies, call recordings, compliant logs, claims, economics, credit, loan, regulatory information.	Credit risk analysis, fraud detection, return on equity
Telecom	Subscribers, cells, connections	Network quality, OSS, Geospatial
Healthcare	Confidential patient records	Patient flow forecasting, nurse staffing levels
Life Sciences	Clinical trials, reports	Health diagnosis, data extraction automation
Media Entertainment	Mobile data, search data, customer viewing data	Advertising fraud detection, Enforcing GDPR, User behavior analytics
Transportation	Public transportation operation data, passenger ride data	Efficient operation of a transportation system
Travel	Hotel data and pricing	Customer trip planning, Customer recommendations
Public Sector	Intelligence data, public records data (tax, welfare claims, etc.)	Crime prevention, cyberattack prevention, reduced fraud costs
Retail, E-commerce	Merchant and user interactions	Targeted offers, enhanced customer experience
Energy	Sensor data, consumption data	Optimal energy generation in Smart Grid
Manufacturing	Picture and video data, equipment data	Efficient manufacturing flow
Others	Large amount of data, time-series data	Trend identification, Outlier detection, Process improvements, cost optimization

Table 1. Example results of machine learning from selective data in different industries.

Delivering the Infrastructure for Data Solutions



Cloudera CDP-PVC-Base and Apache Spark 3.x deployed on a cluster of NVIDIA GPU-accelerated servers offer a very scalable data analytics solution for terabytes to petabytes of continuous data. Furthermore, Supermicro offers the latest state-of-the-art Intel or AMD CPUs and NVIDIA GPUs with the best price/performance efficiency. The server reliability with built-in redundancies ensures minimal cluster downtime.

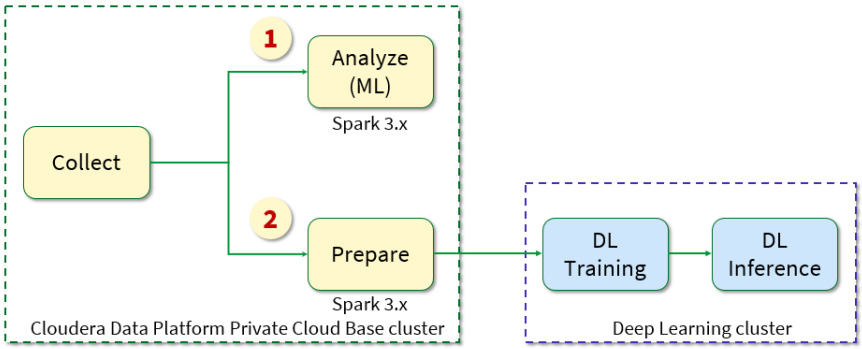
Supermicro's Super Cloud Composer (SCC) makes it easy to manage server clusters and deploy them in the hybrid data center. SCC monitors the servers, network, and storage from a single pane of glass interface. In addition, Cloudera offers a simple means to deploy the CDP-PVC-Base suite onto the servers. Overall, hardware, software, cluster setup is well organized and ensure correct cluster operations.

For enterprise level support, Supermicro offers 24x7 hardware support, while Cloudera provides enterprise class software support.

This cluster delivers many applications to process customer data. This solution brief presents two of these applications:

(1) Data analytics features in Apache Spark 3.x, delivered by Cloudera CDP Private Cloud-Base.

(2) Data preparation using Apache Spark 3.x and CDP Private Cloud-Base capabilities to process the incoming data to feed into another AI cluster to perform Deep Learning (DL) training. The DL training produces AI models, which run in DL inference to make predictions or decisions on new incoming data.



Enterprise Data Analytics / Machine Learning Applications (1)

Data analytics applies statistical and mathematical algorithms to extract insights from data. Be it linear regression on time series data or feature extraction from images or voice or complex data, these algorithms enable fast insights from continuously available data to businesses in various industries.

Cloudera CDP PVC-Base includes many data analytics capabilities, including:

- Regression to find trends
- Classification and Clustering to put data in different buckets
- Collaborative filtering to build recommender systems
- Frequent pattern mining to analyze large scale dataset
- Model selection and tuning to find the best model or parameters

By running data through these and additional algorithms in the Cloudera / Apache Spark cluster, customers can quickly get great insights and make efficient business decisions.

Data Preparation, Feature Extraction for Deep Learning Pipelines (2)

Deep Learning (DL), using neural network architectures, often enables deeper insight than machine learning. This does require additional technologies, which involve a GPU-based training server cluster and inference servers.

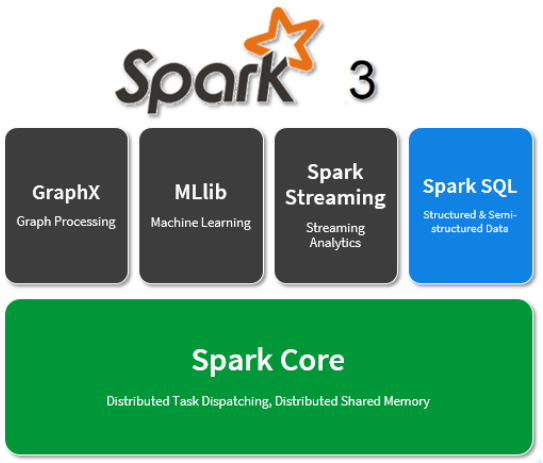
When data scientists in an enterprise determine that neural networks or other deep learning algorithms could get even more insights from the available data, Cloudera/Apache Spark has mechanisms to prepare the data to pipeline into DL training clusters. These include the following. Data scientists can combine these programmatically to perform high performance data preparation:

- TF-IDF for word frequency
- Word2Vec for word embedding
- Feature transformers to index, scale, convert data
- Feature selectors to select features using statistical rules

Whether the data consist of text, images, or other forms, customers can set up data preparation and extraction to pre-process data to pipeline into DL training, ensuring quick and efficient operations. CDP PVC-Base uses GPUs to accelerate SQL and Data Frame API processing, enabling faster data preparation.

Apache Spark 3.x with NVIDIA GPU Acceleration

Apache Spark 3.x provides the integrated components for easy deployment of tens to thousands of servers to run scalable in-memory graph processing and computation. Apache Spark 3.x has improved significantly from previous versions. Now, CDP PVC-Base with Apache Spark 3.x supports NVIDIA GPU acceleration natively. CPD PVC-Base uses the GPUs to accelerate SQL, Data Frame API processing. Using a single GPU, we see five times or more performance speedup over CPU based systems. Using multiple GPUs, we see over 43X performance speedups. NVIDIA and Cloudera have worked together to incorporate the NVIDIA RAPIDS libraries, using GPUs, into the CDP PVC-Ba software stack to achieve these performance speedups.

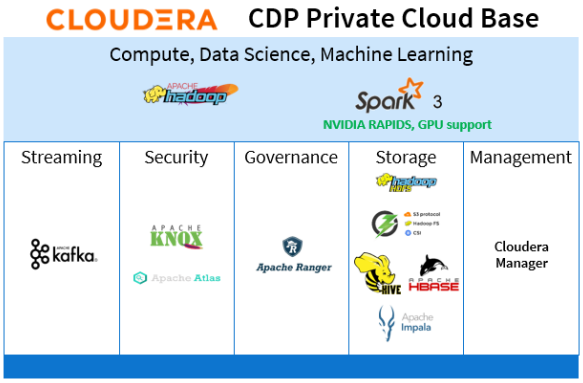


Supermicro delivers Ultra servers with NVIDIA A30 or NVIDIA A100 GPUs. When adding these GPU servers to a running CPU-based Cloudera cluster, CDP PVC-Base automatically incorporates the GPU systems to accelerate the Apache Spark 3.x workloads, especially for those running Spark SQL and Data Frame APIs.

Cloudera Data Platform Private Cloud-Base 7.1.7

Cloudera Data Platform Private Cloud Base provides an integrated platform to run data streaming, storage, and data science and machine learning in server clusters. The software can be easily deployed onto a cluster of servers and be managed using built-in in-band management. In addition, both public cloud and hybrid data center server clusters come with enterprise level support.

CDP PVC-Base 7.1.7 incorporates Apache Spark 3.x with NVIDIA RAPIDS GPU acceleration to deliver over five times more performance speedup over CPU based clusters.



Red Hat Enterprise Linux

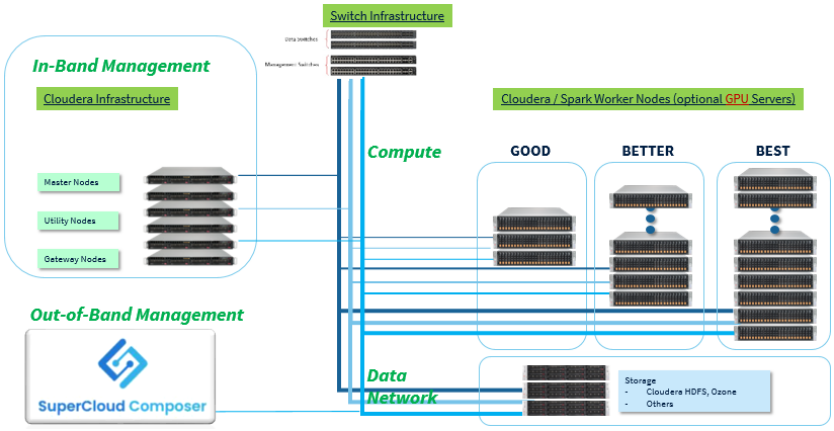
Red Hat Enterprise Linux (RHEL) is the operating system supporting Cloudera Data Platform Private Cloud Base. RHEL provides enterprise level support for the operating environment that enables reliable and scalable operations in each Supermicro Ultra server.

Supermicro SuperCloud Composer

Supermicro SuperCloud Composer (SCC) provides monitoring, management, and server operating systems deployment for server clusters. Using out-of-band IPMI and Redfish management, SCC monitors the health and operation of each server in the cluster. While server clusters are usually set up with multiple networks to support out-of-band, in-band, and data networks, SCC configures the network IP addresses on the servers. SCC can also deploy operating systems images onto one or all the servers. This makes the management and deployment of racks of servers much more manageable.

Supermicro Server Cluster running Cloudera Data Platform Private Cloud Base and Spark 3.x

The starting base Cloudera Data Platform Private Cloud Base has 10 Ultra servers (either 1U or 2U). Many more servers can be added to scale the cluster depending on the amount of data and computing needed. There are options for high availability and data storage. To accelerate Spark Apache 3.x operations, we can add Supermicro Ultra servers with NVIDIA A30 or NVIDIA A100. The Cloudera cluster automatically incorporates additional servers to scale the cluster performance. Supermicro is providing a reference architecture for this cluster in a separate document.



Conclusion

In collaboration with Cloudera and NVIDIA, Supermicro is delivering high performance server clusters supporting GPUs to provide enterprise level support to customers to turn their data into valuable insights to run their business. The GPU acceleration provides five times and more speedup over CPU operations for Apache Spark 3.x jobs. The CPU-based servers support the other Cloudera streaming, security, governance, and storage functions. Supermicro's SuperCloud Composer monitors the server hardware operations to provide high availability. SCC also deploys the appropriate operating systems to the servers in the cluster, while Cloudera offers a simplified means to deploy the Cloudera applications. Together, Supermicro, NVIDIA, and Cloudera deliver high performance clusters to process customer data into valuable insights to run their business. The entire solution comes with enterprise level support.

Contact your Supermicro Sales Team for more information.

References

1. https://developer.nvidia.com/blog/accelerating-apache-spark-3-0-with-gpus-and-rapids/

2. https://docs.cloudera.com/documentation/enterprise/6/6.3/topics/spark.html

3. https://www.infoq.com/articles/deep-learning-apache-spark-nvidia-gpu/

4. https://www.supermicro.com/solutions/Solution-Brief_SuperCloud_Composer.pdf

5. https://www.supermicro.com/en/products/ultra/

6. https://www.nvidia.com/en-us/data-center/products/a30-gpu/

7. https://www.nvidia.com/en-us/data-center/a100/