DELL Technologies

Unlock the Value of Your Data Capital

Simplify, create more value and protect your data with Dell Technologies Validated Designs for Analytics

Table of Contents

Modernize and transform IT to unlock the value of your data. . . . 2
- Dell Technologies has what you need.. . . . 2

Analytics use cases . . . 3

Overcome analytics challenges. . . . 6
- Valuable . . . 6
- Simple . . . 6
- Secure . . . 6
- Customer success stories . . . 7

Technical specifications. . . . 8
- Boomi Data Catalog and Preparation . . . 8
- Cloudera Data Platform Private Cloud Base . . . 9
- DataStax Enterprise on Apache Cassandra . . . 9
- Edge Analytics for Industry 4.0 with Confluent Platform. . . . 10
- Real-Time Data Streaming. . . . 11
- rENIAC Data Engine for Apache Cassandra NoSQL . . . 11
- Spark on Kubernetes . . . 12
- Splunk Enterprise . . . 12
- VMware Tanzu Greenplum. . . . 13

Solution highlights . . . 14

Services and financing . . . 15

Why choose Dell Technologies for analytics . . . 15
- Customer Solution Centers . . . 16
- AI Experience Zones . . . 16
- HPC & AI Innovation Lab . . . 16
- HPC & AI Centers of Excellence . . . 16
- Proven results . . . 16

Take the next step, today . . . 17

"Our partnership with Dell Technologies allows us to take advantage of the full breadth and depth of their compute, storage, networking and security solutions."

— David J. Brzozowski Jr, Chief Technology Officer, Medacist[1]

With Dell Technologies Validated Designs for Analytics, organizations across every industry and geography can unlock the value of their data capital to reach their business goals.

Validated Designs for Analytics, developed jointly with leading partners, help you create more value from data while simplifying and securing your data landscape. Consisting of high performance Dell EMC infrastructure — PowerEdge servers, PowerSwitch networking, PowerScale storage, Omnia software and best in class services — these solutions are designed to harness the power of analytics to transform how you do business.

Modernize and transform IT to unlock the value of your data.

In a world where digital transformation defines winners and losers, data is rapidly becoming an organization's most valuable asset. Data capital, or the value derived from data, has become the ultimate source of differentiation — and that means every company is now a technology company. This has led to massive IT infrastructure investments and changed how organizations treat their greatest asset — their data.

As the data landscape grows and becomes more unwieldy, it can limit the ability to leverage data and manage it effectively. To fully unlock the value of data capital, attention must be paid to where data resides, how it is managed and how it's protected. By addressing these challenges, IT can focus less effort on day to day management, and more on becoming a source of innovation that uses data to strengthen and differentiate the business.

Dell Technologies can help you unlock the value of your data capital with a solution portfolio that encompasses a wide selection of options, purpose built for today's top data transformation goals. Dell Technologies Validated Designs for Analytics are engineering tested and designed to simplify deployment and operation of analytics projects to enhance the value of your data capital while helping you maintain data protection and security.

Dell Technologies has what you need.

Expertise and guidance

Technology is evolving quickly, so your team may not have the resources to design, deploy and manage solution stacks optimized for modern analytics needs. While advanced analytics and artificial intelligence (AI) might seem like the latest IT trends, Dell Technologies has been a leader in the advanced computing space for over a decade, with proven products, solutions and expertise. Dell Technologies has a team of analytics, AI and high performance computing (HPC) experts dedicated to staying on the cutting edge, testing new technologies and tuning solutions to your applications to help you keep pace with this constantly evolving landscape.

Dell Technologies Validated Designs for Analytics

The data driven age is dramatically reshaping industries and reinventing the future. Leveraging vast amounts of data from diverse sources is both critical and transformational. Mastering analytics holds tremendous potential for dramatically growing revenue and controlling costs. But it's often difficult to know where to begin. Dell Technologies Validated Designs for Analytics can help unlock the value of data. The breadth of the Dell Technologies portfolio makes it easy to find a solution, and together with our partners, Dell Technologies also offers consulting, installation, implementation, support and education services for analytics.

Solutions customized for your environment

Dell Technologies uniquely provides an extensive portfolio of technologies to deliver the advanced computing solutions that underpin successful analytics, AI and HPC implementations. With years of experience and an ecosystem of curated technology and service partners, Dell Technologies provides innovative solutions, workstations, servers, networking, storage and services that reduce complexity and enable you to capitalize on a universe of data.

[1] Dell Technologies case study, Medacist Advances Healthcare Analytics with AI running on Dell EMC PowerEdge and PowerScale, January 2021.

Analytics use cases

The use cases for analytics are diverse, but there are common patterns across industries and verticals. Here is sampling of possible use cases.

Boomi Data Catalog and Preparation use cases

Uncovering unknown and unused data	Boomi® Data Catalog and Preparation (DCP) provides a central data repository and detailed metadata that bring unknown or underused data to the attention of business analysts.
Enabling data insights for business users	IT can enable data preparation capabilities for business users by simply adding a Boomi DCP node to a cluster. Business analysts using DCP can create powerful data transformations and summaries, with basic relational data knowledge and some common SQL commands. These jobs are then run using highly scalable Hadoop services to produce new data artifacts for analysis that replace traditional "data extract" requests. The artifacts are suitable for use with many popular reporting and visualization tools.
Enhancing data security	IT staff responsible for data governance can configure the right data access for the right users, using DCP role-based security integrated with existing source system controls.

DataStax Enterprise on Apache Cassandra use cases

Modernizing applications	DataStax® and Apache Cassandra® provide a foundation for application modernization. Reducing reliance on mainframes for running monolithic applications reduces costs and enables IT to be more responsive to the business.
Managing large data sets	The decentralized, peer-to-peer architecture of DataStax Enterprise also lends itself well to distributed applications and linear scalability for large data sets. DataStax Enterprise on Apache Cassandra enables capturing and acting on large amounts of data in real time as well as storing vast amounts of historical data with immediate access.
Enabling deeper data insights	Cassandra NoSQL databases can scale horizontally to provide fast read/write access to various structured and unstructured data types, so users can perform both real-time and batch analytics for more in-depth insights from data.
Offloading ETL	DataStax Enterprise provides integrated functionality, such as with Apache Solr™, for search and indexing with Apache Spark, which provides near-real-time processing of data streams. As a platform, DataStax Enterprise functions as a hybrid transactional/analytical processing (HTAP) architecture by transparently replicating data across Cassandra nodes without requiring costly and complex ETL processes to move data between the nodes. This replication allows individual nodes to access data instantly, while Spark and Solr provide analytics and search capabilities.

Edge analytics for Industry 4.0 use cases

IoT	Confluent enables near real-time insights into a wide variety of data by enabling inspection, correlation and analysis of massive volumes of data as they stream into databases from many different sources at an extremely high velocity. You can ingest and consolidate streaming data and events from IoT sources and extend with an extensive suite of prebuilt data connectors.
Predictive maintenance	Intelligent machine scenarios focus on preventing unplanned outages through data-model-driven predictive maintenance. This requires the integration of industrial IoT (IIoT) data management technologies with data science capabilities that are the vision of Industry 4.0.
Disconnected/offline operations	Local data processing and real-time analytics enables anomaly detection and business continuity when an edge site is offline and/or disconnected from the cloud or core.
Integrate data collection with data science	Confluent harnesses continuously streaming data with low-latency, high-throughput, fault-tolerant publish-and-subscribe pipelines that can perform distributed streaming analytics on trillions of events per day with millisecond responses to connect downstream systems with real-time data.

Hadoop and Spark use cases

Operational efficiency

Data warehouse augmentation	Reduce total cost of ownership (TCO) and increase return on investment (ROI) by offloading extract, transform, load (ETL) workloads, reducing licensing costs, enhancing data accessibility, enabling better analytics and managing performance more effectively.
Log aggregation and analytics	Prevent security breaches and threats, detect operational anomalies and increase infrastructure efficiency and automation.
Dual storage and active archive	Reduce TCO and ease compliance and reporting with lower data storage costs, better data accessibility, streamlined inquiry processes and improved business operations.
Archive-intensive and tiered Hadoop	Support storage-centric workloads with large capacity requirements and lower costs for active archive. Use for long-term tiered storage for regulatory compliance and get multiprotocol support for storage consolidation.

Business transformation

Marketing	Anticipate customer needs with 360-degree customer insights for better retention, segmentation and loyalty along with data-driven product/service launches.
Finance	Reduce risks and increase revenues and margins using advanced analytics and AI for credit scoring, customer analytics, fraud detection, risk management and regulatory compliance.
Healthcare	Improve patient care and safety, reduce costs, mitigate risks, detect fraud and better manage claims using AI and analytics.
Pharmaceutical	Enhance regulatory compliance and validation using Apache® Hadoop® and Apache Spark® for biomedical analytics, drug stability and shelf-life analysis, primary research and FDA-compliant manufacturing.
Manufacturing	Achieve continuous process improvement using advanced computing for product quality, customer insights, demand forecasting and improved operations.

Real-time data streaming use cases

IoT	Confluent® with Apache Kafka® Enterprise can help you get near-real-time insights into a wide variety of data from many different sources, ingest and consolidate streaming data and events from IoT sources, and extend capabilities with a suite of prebuilt data connectors.
Real-time feedback	Use real-time feedback to enhance inventory management and agility and set up automatic ordering/shipping of low-stock items.
Visibility via real-time sensor data	Optimize labor efficiency, machine health and equipment administration and production attainment.
Remote monitoring and alerts	React quickly to changes in demand and traffic patterns and monitor perishable goods using sensor data.
Streaming analytics	If you're currently using Hadoop as a batch process and data repository, you can easily connect to implement real-time streaming analytics. Confluent can perform distributed streaming analytics on trillions of events per day with millisecond responses to connect downstream systems with real-time data.

Splunk use cases

Application delivery	Help developers deliver applications faster with a positive user experience. Splunk® helps DevOps organizations deliver faster releases, operations teams reduce mean time to resolution (MTTR), and development teams optimize application quality, performance and costs.
Business analytics	Splunk analyzes, visualizes and monitors machine data from any source — such as applications, mobile devices and servers — to provide enhanced business insights in real time to executive, sales, product, marketing, operations and customer service teams.
Cloud	Splunk enables centralized visibility across on-premises and multi-cloud environments, so you can leverage cloud with the security, visibility and assurance that you require. Splunk delivers operational intelligence for a real-time understanding of what's happening across the business and IT, so you can make better-informed decisions.
Internet of Things (IoT)	Splunk software ingests, analyzes and visualizes real-time and historical machine data from any source — including industrial control systems and connected devices — enabling improved operations, enhanced safety and compliance, predictive maintenance and better management of the uptime and availability of industrial assets.
Operations	Splunk collects and correlates machine data so IT can quickly troubleshoot issues and outages, monitor service levels and detect anomalies. Splunk can help reduce MTTR, lower monitoring costs, improve uptime and support modernization initiatives.
Log management	Splunk can consolidate and index log and machine data, including structured, unstructured and complex multiline application logs. You can collect, store, index, search, correlate, visualize, analyze and report on any machine-generated data to identify and resolve operational and security issues in a faster, repeatable and more affordable way.
Security and fraud	Splunk enables collaboration and implementation of best practices to address modern cyberthreats. With Splunk as a nerve center, security teams can leverage statistical, visual, behavioral and exploratory analytics to drive insights, decisions and actions.

Overcome analytics challenges.

Derive more value from your data — quickly and simply — while protecting and securing data assets with Dell Technologies Validated Designs for Analytics.

"We have all this data and we're not leveraging it to make the best business decisions."

Data is critical to every aspect of running a modern business. Decision makers need data, information and metrics they can use to make the right business decisions. Data architects, analysts and scientists need fast, concurrent access to analytics results based on data from every corner of the business to attract and retain customers and outpace the competition.

Valuable

Dell Technologies Validated Designs for Analytics help you turn your data into a tangible asset and source of revenue. You can go from managing infrastructure to using data to power your digital business. Dell Technologies customers report they have been able to triple customers and revenue,[2] save hundreds of millions of dollars,[3] and collect and process 100,000 data points per second to rapidly test engineering designs without costly physical prototyping.[4] Using Dell Technologies APEX Cloud Services enables you to place your data in the optimal location to reduce costs and enhance access.

"Multiple analytics environments create more complexity."

The data avalanche, coupled with opportunities for insight and automation, means groups will continue to request different analytics environments. Those same teams may also want to experiment with AI and machine learning (ML). But it takes time to architect, procure and deploy the right infrastructure. Testing and tuning can be difficult and time-consuming. It's time-intensive and complex to manage and maintain each and every implementation while the queue for new projects continues to grow.

Simple

Don't confine your data to a single hub or cloud. Validated Designs for Analytics make all your data available to the entire business with solutions that span edge, core and multi-cloud environments. Dell Technologies customers report that it takes them hours or days to deploy new solutions instead of weeks or months,[2] that they are able to manage five billion records from 40 different systems in a single view,[5] and that they achieve 120:1 efficiency gains.[6] Dell Technologies APEX Data Center Utility combines market-leading solutions with professional services and support to fully manage your data center and its operations in a simple, single-invoice monthly payment based on your actual usage.

"Varonis reports 3,950 data breaches were confirmed in 2020."

The impact of cyberattacks on business data can be devastating. Secure operations are essential for protecting data — and the business. To keep pace with threats that organizations face today, a new approach is required that can adapt to unexpected changes, with intelligence, automation and security built into everything.

Secure

Reduce risks and secure your data assets as they move across heterogeneous environments with the advanced security of Dell Technologies Validated Designs for Analytics. Dell EMC PowerProtect Backup Service helps you enforce governance while reducing cost and complexity. Plus, Dell Technologies is #1 in Data Protection Appliance and Software.[7]

[2] Dell Technologies case study, Trintech drives 3X revenue growth with Dell EMC PowerEdge servers for Microsoft SQL, accessed November 2021.

[3] Dell Technologies case study, Advancing healthcare analytics and saving lives, January 2021.

[4] Dell Technologies infographic, In today's race, milliseconds count, April 2021.

[5] Dell Technologies case study, Informatica: Building a 360 customer view with a modern data platform, accessed November 2021.

[6] Dell Technologies case study, Automated railcar inspections increase security and revenue with AI at the edge, February 2021.

[7] Based on combined revenue from the IDC 3Q20 Purpose Built Backup Appliance (PBBA) Tracker, with select Storage Software segments from the 3Q20 Storage Software and Cloud Services Qview.

"Using Dell EMC PowerEdge servers, we are able to increase our platform capacity in a matter of weeks versus a matter of months."

— Jun Chen, Senior Vice President of Technology Services, Epsilon[8]

And whether you want to leverage cloud now or in the future, only Dell Technologies offers 100% cloud-enabled data protection and backup solutions. Our Future-Proof Loyalty Program provides peace of mind and investment protection with Dell EMC Storage and Data Protection.

Customer success stories

Medacist®

5 minutes	**Millions of dollars**	**13% savings**
instead of 24 hours for delivering analytics results	saved due to upholding 99.99% uptime service-level agreements	using Dell EMC PowerEdge for a Hadoop cluster running AI

Read the case study: Medacist Advances Healthcare Analytics with AI running on Dell EMC PowerEdge and PowerScale.

Trintech®

3X	**700%**	**Hours**
customer and revenue growth	increase in IOPS	instead of months to deploy new data centers

Read the case study: Digital transformation enables 3X customer and revenue growth.

Informatica®

5 billion	**40**	**360-degree view**
records managed	different source systems brought together in a single view	of customer drives micro-segmentation and hyper-personalization

Read the case study: Informatica: Building a 360° customer view with a modern data platform.

- McLaren® Racing uses an HPC system to rapidly study, test and prototype aerodynamic changes to race car designs.
- Duos Technologies® keeps railcars moving with AI based analytics at the edge.
- EDAG® uses AI to ensure efficient maintenance.
- The Texas Advanced Computing Center (TACC) saw up to a 50X performance increase.
- Wrangler is the most powerful data analysis system allocated in the Extreme Science and Engineering Discovery Environment (XSEDE).
- Epsilon® sends billions of emails daily, with campaign adjustments in real time.

Read more customer stories.

"The ability to understand the customer fully with all available data and attributes presents the biggest opportunity for organizations aiming to become more customer centric."

— Zahid Naeem, VP of Product Development, Informatica[9]

[8] Dell Technologies case study, Epsilon, accessed November 2021.

[9] Cloudera case study: Informatica: Building a 360 customer view with a modern data platform, accessed November 2021.

Technical specifications

Dell Technologies Validated Designs for Analytics are designed to help you throughout your digital transformation journey. They enable you to get started with your first project — quickly and simply — or to create greater value for more sophisticated analytics and AI projects. Our approach provides choice in how you acquire and deploy technology to support analytics use cases and workloads.

Boomi Data Catalog and Preparation

Accelerate business outcomes by expanding data insights to more users.

Even as enterprises produce, capture and store more data than ever before, many potential users can't find critical information — and don't have the tools they need to turn data into actionable insights. That's because most systems are designed for professional data engineers, not other users who have data-driven business use case needs.

Boomi DCP software on Dell EMC PowerEdge servers can change the game for organizations working toward becoming data-driven. Together they enable a broader user community to derive business value from practically any combination of enterprise data sources, without learning to program.

- **Connect** — Unify data and applications across hybrid IT landscapes to break down silos and boost productivity.
- **Transform** — Change how business and IT work together to deliver more satisfying experiences by improving the quality and speed of interactions.
- **Modernize** — Capture value from existing applications and data more effectively while cutting development time, reducing costs, extending investments and streamlining processes.

Configuration options

	Dell EMC PowerEdge servers			**PowerScale scale-out NAS**
	R740xd 3.5"	R740xd 2.5"	R640	
Scalability	Up to 288 nodes	Up to 252 nodes	Up to 288 nodes	Up 168 compute nodes Up to 84 storage nodes
Raw storage	64TB/node	24TB/node	24TB/node	102TB / Isilon H600 Storage
Processors	Dual Intel® Xeon® Gold 6140 2.3GHz, 18C/36T		Dual Intel Xeon Gold 6136 3.0G, 12C/24T	Dual Intel Xeon E5-2680 v4 2.4GHz 14C/28T
iDRAC network	S3048-ON			
Pod network	Z9100-ON 25GbE	S4048-ON 10GbE	Z9100 25GbE	S4048-ON 10GbE
Cluster aggregation network	Z9100-ON 100GbE	S6010-ON 40GbE	Z9100 100GbE	S6010-ON 40GbE

Cloudera Data Platform Private Cloud Base

Address numerous use cases on-premises with a multifunction analytics and data management platform.

Most organizations understand the importance of extracting insights from data. However, the considerations and requirements for data management are constantly evolving. Cloudera® Data Platform (CDP) Private Cloud Base is a scalable and customizable Hadoop platform for securely running many types of analytics workloads. When paired with Dell EMC infrastructure, the potential use cases that can be addressed through full-featured data platform are nearly limitless.

Configuration options			
Dell EMC PowerEdge servers and PowerScale Storage		**Dell EMC PowerSwitch networking**	**Software**
Master nodes	**Worker nodes**	S3100-ON (management) S5248F-ON (leaf) Z9264F-ON (spine)	Red Hat® Enterprise Linux® with Cloudera CDP including Cloudera Private Cloud Base and Cloudera Manager
R650	R750 H5600		

DataStax Enterprise on Apache Cassandra

Modernize applications and scale databases across data centers, cloud and edge.

Microservices architectures, facilitated with the use of containers and container orchestration software like Kubernetes, can help transform legacy applications into modern, scalable applications. However, modernization requires a distributed database model that can span data centers, provide high availability and deliver high-performance data processing capabilities as data continues to grow. DataStax Enterprise, built on Apache Cassandra, offers a scalable, highly available NoSQL database solution that can be scaled on-premises across data centers, across cloud providers and out to the edge.

- **Modernize applications** — Embrace a microservices approach that makes applications simpler to maintain and deploy.
- **Run anywhere** — Get the freedom to run data in any cloud — Kubernetes, bare metal, or hybrid — at global scale with no downtime or lock-in.
- **Protect uptime** — Deploy fault-tolerant, distributed databases that can manage high-velocity unstructured and semistructured data while providing high performance and 100% uptime.

Configuration options		
Dell EMC PowerEdge servers	**Management software**	**Database software**
R740xd	• DataStax Enterprise • DataStax Kubernetes Operator for Apache Cassandra • DataStax Enterprise OpsCenter	• Apache Cassandra • Apache Spark • Apache Solr

Edge Analytics for Industry 4.0 with Confluent Platform

Leverage the power of data and edge-based analytics with Kafka and Confluent to drive greater efficiencies.

The massive deployment of IoT devices that make up the industrial IoT (IIoT) produces enormous volumes of data. There is a vast opportunity to discover associations and understanding from the patterns and trends within IIoT data. Successful analytics strategies for Industry 4.0 require dynamically selecting and routing required application data from a subset of devices. In turn, this data must flow through processing at the edge, or upstream at a core or cloud facility. The Validated Design for Edge Analytics for Industry 4.0 with Confluent Platform enables leveraging the power of data to drive greater efficiencies and better decision making.

- **Leverage more data in real time** — Enable real-time data pipelines and streaming applications by integrating data from multiple sources into a single, central event streaming platform.
- **Improve efficiency** — Efficiently filter and flow IIoT data to where the right people and tools can be applied to extract useful insights.
- **Simplify IT** — Leverage validated configurations that simplify connecting data sources to Kafka, building applications with Kafka services, and securing, monitoring, and managing Kafka infrastructure.

Configuration options — edge

Dell EMC PowerEdge servers				Networking	Software
Control center nodes	**Platform node**	**KSQL node**	**Broker node**	• NVIDIA ConnectX-4 LX dual port 10/25 GbE SFP28, rDNC (cluster) • Dell EMC PowerSwitch S5248-ON (top of rack) • Dell EMC PowerSwitch S3148-ON (management)	• Confluent Platform • Apache Kafka • Red Hat Enterprise Linux
R640	R640	R640	R640		

Configuration options — core

OpenShift Cluster based on Dell Technologies Validated Design for Red Hat OpenShift using Dell EMC PowerEdge servers				Dell EMC PowerScale storage	Networking	Software
CSAH node	**Master nodes**	**Worker nodes**	**Broker node**	• Isilon H600 (dynamic and persistent storage) • Isilon H600 with Hadoop Distributed File System (HDFS) (data lake)	• NVIDIA ConnectX-4 LX dual port 10/25 GbE SFP28, rDNC (cluster) • Dell EMC PowerSwitch S5248-ON (top of rack) • Dell EMC PowerSwitch Z9100-ON (spine) • Dell EMC PowerSwitch S3148-ON (iDRAC and out-of-band)	• Confluent Platform • Apache Kafka • Red Hat OpenShift • Red Hat Enterprise Linux
R640	3x R640	3x R640 or 3x R740xd (minimum)	R640			

Real-Time Data Streaming

Optimized infrastructure allows modular integration from edge to insights.

Real-time data processing architectures are complex and time-consuming to design and implement. With a vast ecosystem and numerous moving parts, they create a high barrier to entry. Yet, many businesses need real-time data insights to be responsive, predictive and competitive. The Validated Design for Real-Time Data Streaming helps reduce the time, effort and resources spent on architecting real-time data pipelines and streaming apps.

- **Use more data** — Enable real-time data pipelines and streaming applications by integrating data from multiple sources into a single, central event streaming platform.
- **Improve productivity** — Efficiently filter and flow data to where the right people and tools can be applied to extract useful insights.
- **Reduce complexity** — Leverage validated configurations that simplify connecting data sources to Kafka, building applications with Kafka services, and securing, monitoring and managing Kafka infrastructure.

Configuration options			
Dell EMC PowerEdge servers	**Networking**	**Storage**	**Software**
R640	NVIDIA ConnectX-4 LX 25GbE Dual Port Network Daughter Card	• 2x 4TB, SATA, HDD (control center) • 2x 480GB, SAS, SSD (OS) (platform) • 2x 480GB, SAS, SSD (KSQL) • 6x 1.6TB NVMe (data) + 2x 240 GB, SATA, SSD (OS) (high-performance broker)	• Confluent • Kafka 5.0 Enterprise

rENIAC Data Engine for Apache Cassandra NoSQL

Seamlessly reduce performance bottlenecks on existing databases.

Today's applications involve ever increasing amounts of data and stringent latency requirements, putting a large burden on database infrastructures. Combining FPGAs with advanced software can provide extreme performance improvements for databases without having to make changes to the existing database software or application architecture. So, Dell Technologies and rENIAC teamed up to offer a solution that adds significant performance to open-source Apache Cassandra NoSQL databases while removing the complexity of standing up an IT solution from scratch.

- **Supercharge performance** — Increase maximum read and write capacity and achieve predictable sub-millisecond latency at heavy loads. Increase capacity by adding data engines.
- **Simplify operations** — Execute and re-execute data partitioning and cluster planning and management with little time, effort or expertise required.
- **Future proof databases** — Avoid database modifications with no need to move away from existing SQL or NoSQL databases.

Configuration options			
Dell EMC PowerEdge servers	**Server FPGA**	**Dell EMC networking**	**Software**
R740: • 3x database server • 3x rENIAC Data Engine (rDE) server • 2x client server	• Intel PAC featuring an Intel Arria 10 GX • Intel Acceleration Stack for Intel Xeon CPU with FPGAs	PowerConnect 8024	• rENIAC Data Engine • Apache Cassandra

50X performance increase

The Wrangler supercomputer at the Texas Advanced Computing Center (TACC) runs Hadoop and Spark on Dell EMC PowerEdge servers for a performance increase of up to 50X.[10]

Spark on Kubernetes

Speed up large-scale batch and streaming data processing.

The ability to process large amounts of batch and streaming data is a must-have for analytics to drive better business decision-making and power the next generation of ML applications. Apache Spark is a unified analytics engine that leverages in-memory computing for large-scale data processing. Running Spark processes that are distributed across multiple systems requires container virtualization for automating deployments and scaling. Dell Technologies makes doing this faster, easier and less risky with the Dell Technologies Validated Design for Spark on Kubernetes, a tested, validated architecture that describes the system building blocks for leveraging Kubernetes® to manage infrastructure for Spark analytics.

- **Speed time to deployment** — Reduce the time required to procure, validate and integrate components with tested, validated building blocks.
- **Simplify operations** — Build a full analytics pipeline without having to go outside the Spark ecosystem for data ingestion, cleansing, merging, model training and API development for inferencing.
- **Reduce risks** — Get expert guidance for a validated analytics solution using Spark and Kubernetes.

Configuration options

Dell EMC PowerEdge servers	Networking	Storage	Software
R640	NVIDIA® ConnectX®-4 Lx dual port 25GbE SFP 28 rNDC	2x 800GB SSD SAS mixed-use 12Gbps 2.5 in hot plug AG drive, 3 DWPD, 4380 TBW	• Apache Spark • Red Hat OpenShift® Container Platform

Splunk Enterprise

Turn machine data into actionable business insights with a high-performance, scalable analytics platform.

While machine data is one of the fastest growing sources of data, it's often one of the most underused. Splunk Enterprise software unlocks data from applications, devices, networks, IoT sensors, web traffic and more. It enables organizations to search, analyze and visualize massive streams of machine data to deliver real-time visibility across the entire business. But many organizations find it complex and time-consuming to design, architect, test and validate hardware configurations for Splunk. The Dell Technologies Validated Design for Splunk Enterprise reduces IT risk and speeds time to deployment of optimized Splunk architectures, providing a feature-rich, extensible, high-performance solution that scales to current and future needs.

- **Optimize performance** — High-performance, low-latency and high-capacity configuration options cover a range of needs.
- **Speed time to deployment** — Reduce the time required to procure, validate and integrate components with flexible design choices and guidance.
- **Reduce risks** — Based on extensive customer experience with real-world Splunk production installations and include the hardware, software, resources and services to deploy and manage Splunk Enterprise in a production environment.

[10] Dell Technologies case study, TACC Powers Wrangler HPC with Dell EMC Servers and Flash Storage, accessed November 2021.

Configuration options			
	Dell EMC PowerEdge servers	**Dell EMC VxRail**	**Dell EMC PowerFlex**
Compute	• R640 • R740xd	E560	R640
Networking	Dell EMC PowerSwitch • S4148F-ON • S3048-ON	4x 10GbE SFP+ per node	25GbE Cisco® Nexus®
Storage	2x 480GB SSD SAS mixed-use 12Gbps RAID 1	Dell EMC PowerScale Isilon X410	Isilon A200
Software	• Splunk Enterprise • Splunk Universal Forwarder • Red Hat Enterprise Linux • Dell EMC OpenManage • Dell EMC OneFS	• Splunk Enterprise • Splunk Universal Forwarder • Red Hat Enterprise Linux • VMware® vSphere® Enterprise • VMware vCenter Server® • VMware vSAN™ Enterprise • VMware vRealize® Log Insight™ • VxRail Manager	• Splunk Enterprise • Splunk Universal Forwarder • Red Hat Enterprise Linux • VMware vSphere Enterprise • VMware vCenter Server • Dell EMC Vision Intelligent operations

VMware Tanzu Greenplum

Build and deploy an all-in-one solution for analytics at scale.

To remain competitive in the digital era, businesses need to store, collect and analyze massive amounts of data at speed, and with the scalability to keep up as the business continues to collect and analyze data. However, most IT organizations spend a great deal of time managing a highly siloed legacy data center infrastructure, leading to higher costs, data inconsistency, security concerns and escalating maintenance requirements. A combination of VMware Tanzu™ Greenplum®, VMware virtualization and Dell EMC PowerScale storage, powered by Dell EMC infrastructure, can provide a flexible, all-in-one solution that answers these needs and is cost effective, easy to build and simple to manage.

- **Reduce costs** — Adopt a solution that is space-efficient, performant, and addresses several industry-standard use cases — such as landing zones for data ingestion, data lake storage, data science, data warehouses and data marts — in a single architecture.
- **Ease building** — Streamline deployment and future scalability, provide better integration into data systems and evolve systems to be more cloud-ready.
- **Simplify management** — Abstract and manage infrastructure with a robust and cloud ready approach to enterprise data needs.

Configuration options			
Compute	**Dell EMC storage**	**Virtualization**	**Software**
VxRail P570F	PowerScale Isilon H500	• VMware vSphere • VMware vSAN	• VMware Tanzu Greenplum • Apache Spark • Apache Kafka

Solution highlights

Dell EMC VxRail: A preconfigured and pretested VMware hyperconverged infrastructure (HCI) appliance, powered by industry leading VMware vSAN and vSphere software, the VxRail appliance streamlines and extends the VMware environment while dramatically simplifying IT operations with a known and proven building block for the software-defined data center (SDDC).

Dell EMC PowerFlex: These software-defined storage solutions enable transformational agility for organizations looking to modernize their data center operations. PowerFlex enables extreme flexibility, massive scalability and enterprise-class performance and resiliency while simplifying infrastructure management and operations.

Dell EMC PowerScale storage: Analytics environments require large, scalable, reliable and efficient storage. With support for multiple workloads and enterprise-grade data and file management capabilities out of the box, Dell EMC PowerScale scale-out NAS is the leading storage for analytics. You can take advantage of the high capacity of PowerScale to reduce the acquisition and ownership cost for managing and monetizing data using advanced or predictive analytics and ML.

Dell EMC PowerEdge servers: Dell EMC PowerEdge servers are engineered to deliver unmatched performance and versatile configurations to meet the demands of analytics and AI workloads. Flash storage, the latest processors, greater memory bandwidth and flexible local storage make Dell EMC PowerEdge servers a foundational choice for analytics.

Dell EMC PowerSwitch networking: Today's analytics workloads call for new thinking about network architecture. Based on open standards, Dell EMC networking frees the data center from outdated, proprietary approaches. Our future-ready networking technology helps you improve network performance, lower networking costs and remain flexible to adopt new innovations. Take control of your network's future and learn how the Dell Technologies strategy for open networking can dramatically transform your business.

Omnia software: Dell EMC Omnia is an open source, Ansible®-based software stack designed to automate the deployment of mixed-workload clusters, giving IT the agility to run AI, HPC and analytics workloads in the same environment, with a single pane of glass for cluster provisioning, deployment and management, with easy-to-use point-and-click templates for building complete environments.

"It's not just about building technology for technology's sake. It's about building technology to solve problems, to work in the real world and to make a difference... Having a partner who is truly invested in trying to change things for the better is absolutely critical. That's something we value immensely."

— James Lowey, Chief Information Officer, TGen[11]

Services and financing

Dell Technologies is with you every step of the way, linking people, processes and technology to accelerate innovation and enable optimal business outcomes.

- Analytics Consulting Services help you create a competitive advantage for your business. Our expert consultants work with companies at all stages of analytics to help you plan, implement and optimize solutions that enable you to unlock your data capital and support advanced techniques, such as AI and ML.
- Deployment Services help you streamline complexity and bring new IT investments online as quickly as possible. Leverage our 30-plus years of experience for efficient and reliable solution deployment to accelerate adoption and ROI while freeing IT staff for more strategic work.
- Support Services driven by AI and deep learning will change the way you think about support with smart, ground-breaking technology backed by experts to help you maximize productivity, uptime and convenience. Experience more than fast problem resolution—our AI engine proactively detects and prevents issues before they impact performance.
- Payment Solutions from Dell Financial Services help you maximize your IT budget and get the technology you need today. Our portfolio includes traditional leasing and financing options, as well as advanced flexible consumption products.
- Dell Technologies APEX offers a simple approach that gives you a wide range of consumption models, payment solutions and services so you can optimize for a variety of factors while realizing more predictable outcomes.
- Managed Services can help reduce the cost, complexity and risk of managing IT so you can focus your resources on digital innovation and transformation while our experts help optimize your IT operations and investment.
- Residency Services provide the expertise needed to drive effective IT transformation and keep IT infrastructure running at its peak. Resident experts work tirelessly to address challenges and requirements, with the ability to adjust as priorities shift.

Why choose Dell Technologies for analytics

We're committed to advancing analytics, HPC and AI, and we've dedicated a great deal of resources toward that goal.

- Schedule an executive briefing and collaborate on ways to reach your business goals.
- Dell Technologies Customer Solution Centers are staffed with computer scientists, engineers and Ph.D.s who are subject matter experts in a variety of disciplines.
- We are committed to providing you with choice. We want you to get what you need and have a great experience working with us. If we don't have what you need, we'll tell you who does. We believe in being open, and we publish our performance results at HPCatDell.com.
- Dell Technologies is the only company in the world with a portfolio that spans from workstations to supercomputers, including servers, networking, storage, software and services.
- Because Dell Technologies offers such a wide selection of solutions, we can act as your trusted advisor without trying to sell you a one-size-fits-all approach to your problem. That range of solutions has also given us the expertise to understand a broad spectrum of challenges and how to address them.

[11] Dell Technologies case study, Groundbreaking research with life-changing results, October 2020.

Customer Solution Centers

Our global network of dedicated Dell Technologies Customer Solution Centers are trusted environments where world-class IT experts collaborate with you to share best practices, facilitate in-depth discussions of effective business strategies, and help your business become more successful and competitive. Dell Technologies Customer Solution Centers reduce the risks associated with new technology investments and can help improve speed of implementation.

AI Experience Zones

Curious about AI and what it can do for your business? Run demos, try proofs of concept and pilot software in Singapore, Seoul, Sydney, Bangalore and other Customer Solution Centers. Dell Technologies experts are available to collaborate and share best practices as you explore the latest technology and get the information and hands-on experience you need for your advanced computing workloads.

HPC & AI Innovation Lab

The Dell Technologies HPC & AI Innovation Lab in Austin, Texas, is the flagship innovation center. Housed in a 13,000-square-foot data center, it gives you access to thousands of Dell EMC servers, three powerful HPC clusters, and sophisticated storage and network systems. It's staffed by a dedicated group of computer scientists, engineers and subject matter experts who actively partner and collaborate with customers and other members of the HPC community. The team engineers HPC and AI solutions, tests new and emerging technologies, and shares expertise including performance results and best practices.

HPC & AI Centers of Excellence

As analytics, HPC and AI converge and the technology evolves, Dell Technologies worldwide HPC & AI Centers of Excellence provide thought leadership, test new technologies and share best practices. They maintain local industry partnerships and have direct access to Dell Technologies and other technology creators to incorporate your feedback and needs into their roadmaps. Through collaboration, Dell Technologies HPC & AI Centers of Excellence provide a network of resources based on the wide-ranging know-how and experience in the community.

Proven results

Dell Technologies holds leadership positions in some of the biggest and largest growth categories in the IT infrastructure business, and that means you can confidently source information technology needs from Dell Technologies.

- #1 in servers[12]
- #1 in converged and hyperconverged infrastructure[13]
- #1 in storage[14]
- #1 in enterprise infrastructure[15]

See Dell Technologies Key Facts.

[12] IDC, WW Quarterly x86 Server Tracker, 2Q2021, Vendor Revenue & Shipments, September 9, 2021.

[13] IDC, WW Quarterly Converged Systems Tracker, 1Q2021, Vendor Revenue, June 17, 2021.

[14] IDC, WW Quarterly Enterprise Storage Systems Tracker, 2Q2021, September 9, 2021.

[15] IDC, WW Quarterly Enterprise Infrastructure Tracker: Buyer and Cloud Deployment, Vendor Revenue, October 2021.

Take the next step, today

Don't wait to harness the benefits of analytics on optimized solutions designed from the ground up to unlock the value of your data capital, simply and securely. Contact your Dell Technologies representative to find out more today.

Contact us

To learn more, visit
delltechnologies.com/analytics
or contact your local representative
or authorized reseller.

DELL Technologies

Copyright © 2022 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, and other trademarks are trademarks of Dell Inc. or its subsidiaries. Apache®, Cassandra®, Hadoop®, Kafka®, Spark®, and Solr™ are trademarks of the Apache Software Foundation or its subsidiaries in Canada, the United States, and/or other countries. Splunk® is a registered trademark of Splunk Inc in the United States and other countries. Confluent® and associated marks are trademarks or registered trademarks of Confluent, Inc. Boomi® is a registered trademark of Boomi, LP, in the United States and/or other countries. VMware® is a registered trademark or trademark of VMware, Inc. in the United States and other jurisdictions. DataStax® is a registered trademark of DataStax, Inc. and its subsidiaries in the United States and/or other countries. Medacist® is a registered trademark of Medacist Solutions Group, LLC. Trintech® is a trademark of Trintech, Inc. Informatica® is a registered trademark of Informatica LLC in the United States and other countries. McLaren® is a registered trademark of McLaren Racing Limited. Duos Technologies® is a trademark and brand of Duos Technologies, Inc. EDAG® is a trademark of the EDAG Group. Epsilon® is a registered trademark of Epsilon Data Management, LLC. Cloudera® is a trademark or trade dress of Cloudera. Red Hat®, Ansible®, and OpenShift® are trademarks of Red Hat, Inc. in the United States and other countries. Linux® is the registered trademark of Linus Torvalds in the U.S. and other countries. Kubernetes® is a registered trademark of The Linux Foundation. NVIDIA® and ConnectX® are trademarks and/or registered trademarks of NVIDIA Corporation and/or Mellanox Technologies in the U.S. and other countries. Cisco® and Nexus® are registered trademarks or trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries. Intel® and Xeon® are trademarks of Intel Corporation in the U. S. and other countries. Greenplum® is a trademark of Pivotal Software, Inc. in the U.S. and other countries. Other trademarks may be the property of their respective owners. Published in the USA 01/22 Solution overview analytics-portfolio-overview-SO-103

Dell Technologies believes the information in this document is accurate as of its publication date. The information is subject to change without notice.