

Accelerating Data Science Pipelines with Iguazio and Dell Technologies

The rise of advanced computing models in the forms of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) are becoming more prevalent in the enterprise data center. Organizations are seeking to extract value from their large unstructured data sets by applying Data Science practices. The emergence of these AI/ML/DL workloads has pushed the limits of the traditional architectures for IT organizations, uncovering the need for an integrated and optimized AI infrastructure.

PowerScale Scale-Out Storage

With Dell EMC PowerScale, large petabyte and exabyte data lakes can often hold critical data for many lines of business, such as Genomic Data for Precision Drugs or Fraud data for payment fraud risk management. Data Scientists are becoming more sophisticated in how they write algorithms and build models to query the data. Once deployed, these models can cut costs and increase efficiency by reducing the need for human intervention, and/or by assisting in decision making for a variety of use cases.

PowerScale delivers the performance and concurrency required for AI/ML/DL workloads:

- Eliminate the I/O bottleneck for AI with up to 18x more bandwidth; up to 15 GB/s per chassis, up to 540 GB/s per cluster.
- Bring the analytics to the data with in-place analytics leveraging multi-protocol support such as SMB, NFS, HTTP, S3 and native HDFS to support multiple uses
- Simple and seamless scaling to 21x more capacity 1 than the competition scaling from 10s TBs up to 10s of PBs per cluster with industry leading storage efficiency up to 85%.
- PowerScale OneFS delivers robust data protection and security options with guaranteed resiliency. Automatic tiering between All-Flash, Hybrid and Archive nodes via SmartPools.

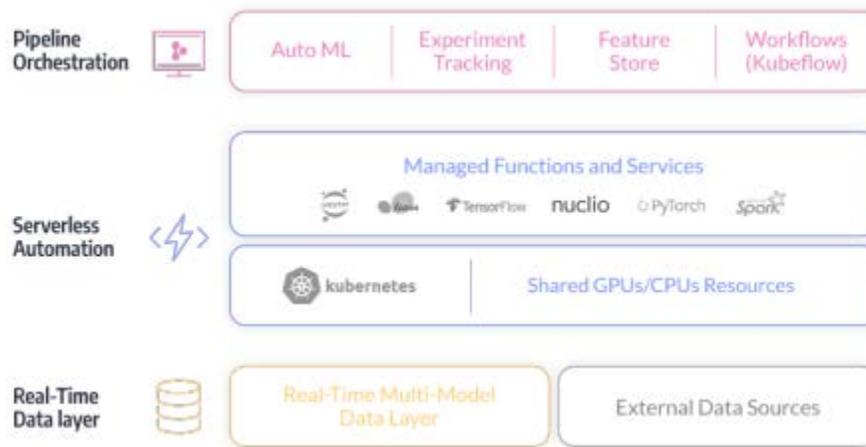
Iguazio Data Science Platform Software

The Iguazio Data Science Platform software automates and accelerates ML workflows, enabling data scientists to develop, deploy and manage real-time AI applications at scale. Iguazio speeds and simplifies deployment of AI and ML applications by building in essential frameworks, such as Kubeflow, Apache® Spark® and TensorFlow™, along with well-known orchestration tools like Docker® and Kubernetes®. The Iguazio software platform enables simultaneous access through multiple industry-standard APIs for streams, tables, objects and files that are stored and normalized once, so you can launch new projects quickly and then consume, share and analyze data faster.

Iguazio Data Science Platform Software + Dell EMC PowerScale

The Iguazio Data Science Platform software on Dell EMC PowerScale consists of four main components, all packaged and managed in an open platform.

- **Real-time data layer:** Provides fast, secure and shared access to real-time and historical data running as in-memory databases on Flash memory, enabling low costs and high density.
- **Serverless automation:** Automates DevOps to save time and resources typically spent on data collection, packaging, scaling, tuning and instrumentation.
- **Pipeline orchestration:** Provides end to-end workflow management via a full-stack, user-friendly environment featuring fully integrated workflow management, experiment tracking and AutoML tools.
- **Extreme concurrency at scale:** Dell EMC PowerScale delivers the analytics performance and extreme concurrency at scale to consistently feed the most data-hungry analytic algorithms.



Iguazio Data Science Platform software on Dell EMC infrastructure provides the following capabilities:

- **Collect and enrich data from any source.** Ingest multi-model data at scale in real time, including event-driven streaming, time series, NoSQL, Microsoft® SQL Server® and other files.
- **Prepare online and offline data at scale.** Explore and manipulate online and offline data using a real-time data layer and your preferred data science and analytics frameworks.
- **Accelerate and automate model training.** Continuously train models in a production-like environment, dynamically scaling GPUs and managed ML frameworks
- **Deploy in seconds.** Deploy models and APIs from Jupyter® Notebook or IDE to production in just a few clicks and continuously monitor model performance.
- **Manage end-to-end workflows.** Manage and automate the entire workflow, streamlining the process of data preparation, training, validation and deployment to operationalize models.
- **Run feature engineering in real-time.** The platform provides real-time feature engineering based on fast data layer and other functions. Users can leverage the software platform capabilities of key value and time series with built-in aggregation functions to calculate and analyze data in real time

This solution offers IT teams an optimal balance of cost, capacity, footprint and performance while delivering the massive datasets required by Data Science teams.

About Dell Technologies

Dell Technologies is a trusted leader in storage for data analytics and AI. Dell EMC PowerScale forms the foundation of a simple, scalable and future-proof IT infrastructure, giving organizations the agility to transform business operations and the flexibility to adapt to new AI and analytics workflows. Through innovative products and services, Dell Technologies and PowerScale accelerate analytics and AI, helping organizations unlock the value of their data to drive the business.

About Iguazio

The Iguazio Data Science Platform automates MLOps with end-to-end machine learning pipelines, transforming AI projects into real-world business outcomes. We're building a faster, easier, and smarter way for data science to flow seamlessly from raw idea to real-world impact. This is data science brought to life.

Notes

1. Compared to closest competitor based on Dell EMC internal analysis, June 2018. G17000096



To learn more about our offerings visit our website
DellTechnologies.com/StorageforAI