Technical White Paper

Dell EMC Isilon: Storage Solution for Autonomous Driving

Dell EMC Isilon H5600 and OneFS 8.2.0

Abstract

This document provides an overview on Dell EMC™ Isilon™ H5600 storage and the OneFS 8.2.0 release, and discusses the relevant use cases that support the development of advanced driver assistance system (ADAS) and autonomous driving (AD).

July 2019

H17846

Revisions

Date	Description
July 2019	Initial release

Acknowledgements

Authors: Bergau, Ryan (Ryan.Bergau@dell.com); Hu, Frances (Frances.Hu@dell.com)

Support: Baumann, Florian (Florian_Baumann@Dell.com); Sevior, Charles (Charles.Sevior@dell.com); Vivolo, Lawrence (Lawrence.Vivolo@dell.com)

Table of contents

Revisions....2
Acknowledgements....2
Table of contents....3
Executive summary....4
1 Dell EMC Isilon H5600....5
1.1 Key benefits....5
1.2 Technical specifications....5
1.3 Isilon OneFS 8.2.0....6
2 ADAS development use cases for H5600....7
2.1 ADAS development lifecycle....7
2.2 Key values for ADAS solutions....9
3 Performance benchmark for H5600....11
3.1 Test methodologies....11
3.2 NFS performance test results....11
3.3 SMB performance test results....12
3.4 H5600 storage efficiency....13
4 Dell EMC ADAS solution architecture....14
A Technical support and resources....16
A.1 Related resources....16

Executive summary

As the automotive industry approaches the reality of widely available, consumer-ready, self-driving vehicles, the need for high-resolution sensors and extremely accurate algorithms becomes more and more important. With so many variables in any given environment, from weather, to signage, to pedestrians and cyclists, tier 1 suppliers and vehicle original equipment manufacturers (OEMs) are challenged to gather, process, and store as much data as possible to account for the trillions of situational permutations their vehicles may encounter on roads around the world. To put the amount of data required to develop a car for full autonomy into perspective, today's common Society of Automotive Engineer (SAE) level 2 autonomous vehicles (with parking assist and adaptive cruise control) typically require ~200,000 kilometers of driving distance, or 4-10 petabytes of sensor data, which is then used for testing purposes. At SAE level 5, fully autonomous vehicles are expected to require over 240 million kilometers of driving distance with over 2 exabytes of sensor data. To get the most value out of this data, automotive companies must make investments in storage infrastructure that can store huge amounts of data gathered from the vehicle's sensors — including lidar, cameras, radar, ultrasonic, and GPS.

While engineering teams are looking down at the infrastructure stack today, Dell EMC is increasingly working to automate, standardize, and cloud deliver those infrastructure layers. This allows engineering teams to look up to the application layer and create business-differentiating services like advanced driver assistance systems (ADAS). Dell EMC continues to focus on delivering a balance of high performance, scalability, and ease-of-use while minimizing costs over time. This challenges the status quo of how storage technology works and provides simplicity, performance, and future-proof infrastructures to support the massive data needs for ADAS developments.

This document provides an overview of Dell EMC Isilon H5600 storage and the OneFS 8.2.0 release, and describes the relevant use cases which support the development of advanced driver assistance systems and autonomous driving.

1 Dell EMC Isilon H5600

With so much data being collected in the ADAS development space, Dell EMC Isilon H5600 storage was introduced to resolve the need for a deep, high-capacity storage data lake. It also offers high performance for the development of machine-learning and deep-learning algorithms, as well as hardware in the loop (HiL), software in the loop (SiL), and model in the loop (MiL) simulations.

Powered by the Isilon OneFS 8.2.0 operating system, Isilon H5600 solutions are designed for simplified management no matter how large the environment becomes. Enterprise-grade data protection and security options also help safeguard valuable ADAS data assets.

Figure 1 Dell EMC Isilon H5600 running the OneFS 8.2.0 file system

1.1 Key benefits

The Isilon H5600 is ideal for high-throughput, large-capacity, unstructured data workloads like ADAS developments. The key benefits of Isilon H5600 are listed below:

- **Capacity**: The H5600 can store up to 800 TB capacity in a single 4U chassis using low-cost 3.5-inch SATA HDD components. With Isilon OneFS 8.2.0, a single volume namespace can scale to 50 PB (raw) in one Isilon cluster.
- **High performance**: The H5600 can deliver up to 8 GB/s of throughput per chassis. With Isilon OneFS 8.2.0, a single H5600 Isilon cluster can deliver over 500 GB/s.
- **Efficiency**: The H5600 is designed for high density with a deep chassis to reduce the data-center footprint and related costs.

1.2 Technical specifications

The Isilon H5600 hybrid storage node running the OneFS 8.2.0 operating system provides 800 TB of capacity in a 4U form factor with 4 nodes. The Isilon H5600 is available in a dense, deep-chassis configuration. With 10 TB SATA drives, an H5600 cluster can scale up to 50 PB while optimizing the rack space used within the data center. Detailed technical specifications per chassis are listed below:

- Drive type: High-capacity 10 TB HDDs (as of June 2019)
- Capacity and performance: 800 TB and up to 8 GB/s bandwidth
- Dense: Deep 4U chassis
- HDD Drives: 80 x 10 TB 3.5” 4KN SATA
- Processor: 4 x Intel® Xeon® Processor E5-2630 v4

- RAM: 1 TiB
- Back-end network: 2 x 40 GbE for intra-cluster communication (as of June 2019)
- Front-end network: 2 x 40 GE or 2 x 10 GbE
- Operating system: Isilon OneFS 8.2.0 or later

For more detailed information, refer to the Dell EMC Isilon H5600 specifications.

1.3 Isilon OneFS 8.2.0

Isilon OneFS 8.2.0 is the latest version of the Isilon operating system that accelerates demanding workloads by enabling customers to consolidate, store, and manage their unstructured data. Customers can upgrade to this version without disruption to take advantage of these powerful features. OneFS 8.2.0 enables customers to transform their business and modernize IT while delivering the scale and performance to maximize business value of their data. OneFS 8.2.0 includes the following new features:

- Delivers a new level of performance and capacity that enables customers to scale well beyond current limits. With 75% more nodes supported per cluster, Isilon customers can scale beyond the 144 nodes limit — up to 252 nodes per cluster. With more nodes supported per cluster, customers can run more AI deep learning jobs, ADAS re-simulation tests and deliver results more quickly.
- Provides extreme and predictable performance with support for 252 nodes per cluster. A single H5600 Isilon cluster can deliver over 500 GB/s throughput performance.
- Simplified management and seamless integration with the cloud for tiering cold data. Isilon CloudPools now supports improved reporting, seamless snapshot and quota integration and compatibility with new platforms like Alibaba Aliyun, Federal C2S clouds, and Google Cloud Platform™.
- Strengthens security and data protection while simplifying data management with features like Apache® Hadoop® TDE, SyncIQ encryption, multifactor authentication, and RBAC per access zone.
- Delivers a faster and more flexible backup solution with a new integrated Network Data Management Protocol (NDMP) Fibre Channel hardware card that can decrease data center space used. OneFS 8.2.0 nodes now support two-way NDMP-based backup to Fibre Channel connected archive libraries for improved data protection.
- Seamless, non-disruptive upgrades to OneFS 8.2.0 are supported from OneFS 8.0.x, OneFS 8.1, OneFS 8.1.1, and OneFS 8.1.2.

2 ADAS development use cases for H5600

2.1 ADAS development lifecycle

The following lists the typical ADAS development lifecycle for automotive OEMs and Tier-1 suppliers leveraging the Dell EMC Isilon scale-out NAS as the central data lake. This is also illustrated in Figure 2.

1. Data acquisition: Huge volumes of sensor data are captured by a fleet of test vehicles, which may comprise video sequences, sonar, radar, lidar, GPS, with more high-resolution sensors such as 4k and even 12k video in the works, and others. Some of these sensors will be beta samples of the actual sensors planned for the production vehicle, while other sensors may be capturing high-resolution reference data around the test vehicle. Typically, we see our customers generating real-world test data of from 2 TB per hour, to 30-80 TB per car per day. Many customers are running test fleets with 50 or more vehicles. The data is stored in real-time using dedicated industrial data-logging hardware with removable solid-state storage disks. These drives are swapped out either daily or at the end of each shift – depending on the amount of data captured per shift. The drives are then either shipped directly to a centralized ingest server, transferred virtually through WAN lines or transferred locally to tape, with the tapes then being shipped to a centralized ingestion server for upload to the data lake.
2. Data ingestion: During the data ingest process, which includes moving data from the vehicle to the data lake, custom copy stations are used to apply data cleaning and lossless data compression algorithms with the goal of reducing the final amount of needed storage and costs. Typically, only a portion of recorded data is needed to train the machine-/deep-learning algorithms and further, this data can be losslessly compressed during ingest to avoid multiple copy or move operations.
3. Data preparation: Once the data has been ingested, the engineering teams will start to prepare the data which may include trimming, decoding, data enrichment (labeling or ground truth generation), processing and adding metadata such as weather and traffic conditions. This requires a vast amount of CPU and GPU resources in the HPC cluster to process the data held in Isilon storage taking advantage of the high sequential read and write performance needs.
4. Test preparation: Engineers can build test suites including designing test cases, required re-simulation and simulation to develop ADAS models. With massive raw datasets, it is very important to be able to search the metadata from data lake quickly to find the right sensor data for different test use cases. Tests are created to cover all possible corner cases, with discrepancies between the ECU validation and test driver actions identified as potential bugs during the validation phase.
5. Design and development phase: When the data is ready, the ADAS engineering teams can develop and build algorithms for smart cameras or ECU models through deep learning and iterative testing using data fusion of all the sensors, GPS, weather and road/environment data. On small projects, individual sensors and ECUs may be tested independently. Then all subsystems are tested together at the system level.
6. Re-simulations: As test cases are defined the engineering teams can schedule re-simulation jobs on the Hardware-in-the-Loop/Software-in-the-Loop computer clusters. This involves “replaying” the captured raw sensor data back through the test farm – usually with hundreds or even thousands of iterations running in parallel. This workload leverages the inherent high-concurrency benefit of the Isilon scale-out NAS architecture.
7. Analysis: Once testing is complete, engineers need to analyze the test results and determine whether additional validation is required. In-place analytics can be used to compare ECU operation to original test driver actions to quickly identify potential bugs. The algorithms can then be refined to achieve the expected output results and the revised ECU version can be uploaded to the test vehicles adopting a

continuous improvement process. All the results are sent to the data center storage to provide the engineering teams with on-demand access.

8. Archiving: Following final validation, data can be moved to lower-cost archive storage. Archiving must meet regulatory and contractual commitments, which typically span multiple decades – the "life of the vehicle". Many OEMs stipulate service-level agreements (SLAs) of 1-30 days for simulation data restoration time – for example, in the event of a safety recall – to allow quick turn-around of updates. This is a critical requirement and must be well documented as it has dramatic impact on the archive strategy.

Note: The above steps are not time-sequential but must all be conducted concurrently and continuously to ensure efficient development team progress and solution outcomes. The read and write demands on the central data lake (Isilon) are substantial.

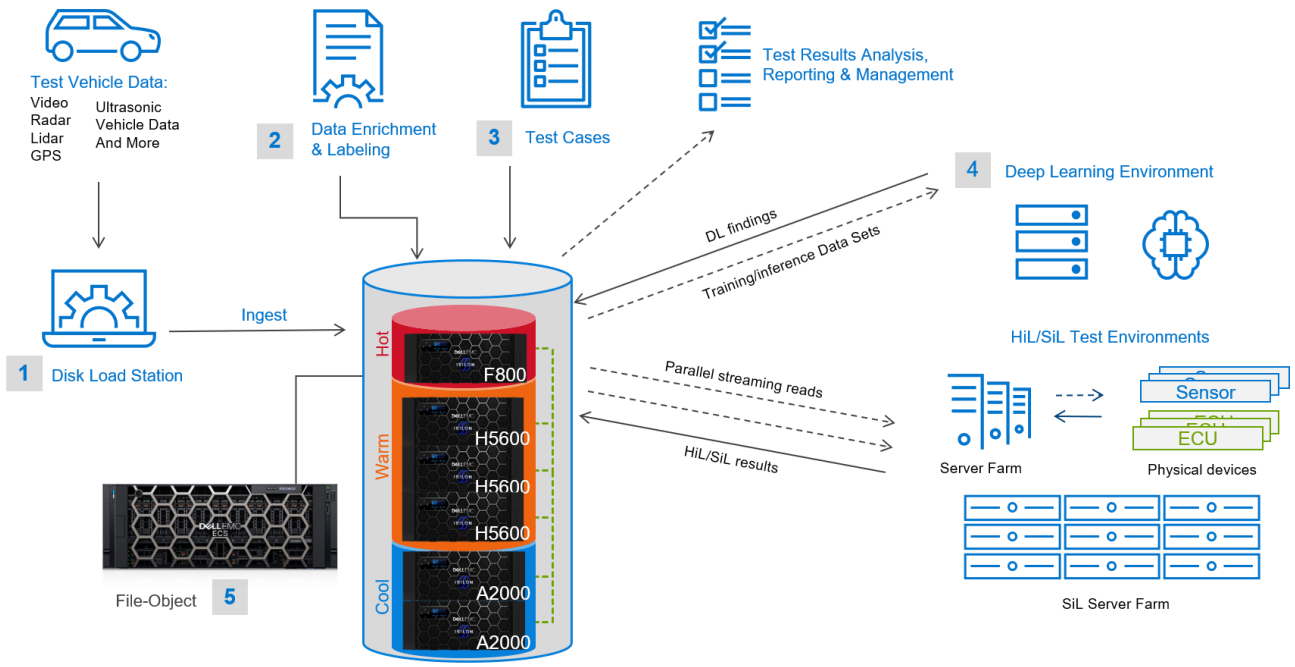

Figure 2 ADAS development lifecycle

In this development lifecycle, Isilon clusters can provide storage tiering to easily meet the different business needs on ADAS development cycle, shown in Table 1.

Table 1 Isilon storage tiering for different ADAS development workloads

Storage tiering	Isilon node type	Suitable ADAS workloads
Hot tier	• Isilon F800 • Isilon F810	• Deep learning training
Warm tier	• Isilon H5600	• 24*7 re-simulation jobs • Daily/weekly data ingestion • Sensor Data Preprocessing • Test Results Data analytics

Storage tiering	Isilon node type	Suitable ADAS workloads
Cold tier	• Isilon A2000 • ECS ES3000	• Long-term file archiving • Multi-site object storage

2.2 Key values for ADAS solutions

With mixed workload requirements and massive datasets requirement, the Dell EMC Isilon H5600 was introduced to resolve meet he performance and capacity data lake needs of ADAS development. Here are some key values that Isilon H5600 delivers to our customers for ADAS solutions.

Accelerate time to market: Isilon H5600 delivers the consistent, high throughput required to ingest data from test vehicles while simultaneously delivering test data into hundreds to thousands of concurrent streams to MiL/SiL/HiL servers, test stands and even deep learning training. H5600 can deliver up to 8 GB/s throughput per chassis. With Isilon OneFS 8.2.0, a single H5600 Isilon cluster can deliver over 500 GB/s. Isilon also scales performance near-linearly, so performance isn't degraded as additional capacity is added – critical for ADAS development where sensor data ingest rates of 2 PB+ per week are becoming common.

Scale to accommodate growing AD/ADAS Datasets: Autonomous driving (AD)/ADAS datasets are growing exponentially, with requirements ranging from petabytes to exabytes of data. Dell EMC Isilon H5600 scales predictably as your needs grow, with OneFS 8.2.0 operating system, a single namespace can easily scale to 50 PB raw data and can effectively be extended to exabytes with Isilon Data Management Service (DMS). Isilon offers truly scalable and predictable performance and an ever-expanding single namespace that eliminates data silos by consolidating all globally collected AD/ADAS data. The OneFS operating system automatically senses and adds new hardware transparently, creating a "plug and play" expansion capability that won't disrupt ongoing projects. Isilon's automated policy-based tiering within a cluster eliminates file server sprawl and performance bottlenecks while streamlining management, backup, and disaster recovery operations. Isilon CloudPools and file-object orchestration also takes this scalability into the exabyte range, allowing data to be moved between the high performance NAS and multiple lower-cost private and public cloud storage options. CloudPools fully encrypts data before offloading it to the target, which can include your own on-premises Dell EMC ECS™ object storage and third-party providers such as Virtustream, Amazon® S3, Google Cloud Platform and Microsoft® Azure®.

Multi-protocol support for centralized storage: With Isilon, you can streamline your automotive and non-automotive storage infrastructure by consolidating large-scale file and unstructured data assets, eliminating islands of storage across the enterprise. Isilon scale-out NAS includes integrated support for a wide range of industry-standard protocols, including Internet Protocols IPv4, and IPv6, NFS, SMB, HTTP, FTP as well as a REST API for file access via HTTP for your cloud initiatives — including internet of things (IoT), vehicle-to-everything (V2X), and Connected Car. With native Hadoop Distributed File System (HDFS) support, Isilon even allows you to run analytics on your manufacturing data. You can even combine this data with your computer-aided design (CAD) or computer-aided engineering (CAE) design data on a single cluster and volume. With Isilon, you can simplify workflows, accelerate business analytics projects, support cloud infrastructure initiatives, and get more value from your enterprise applications and data, all from a single storage platform.

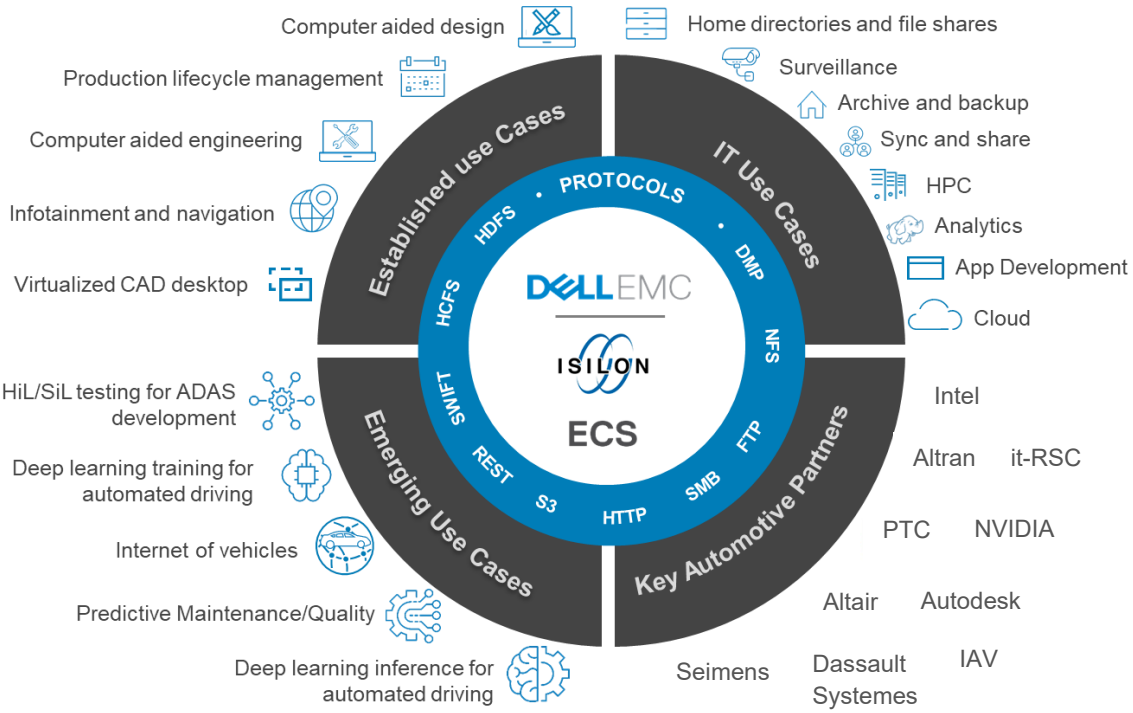

Figure 3 Multiple procotol support for Isilon data lake

Small footprint for tight spaces: When facing the reality of ADAS storage requirements, the last thing you want to worry about is physical space requirements. The Isilon modular architecture delivers maximum flexibility and capacity in a small form factor. The H5600 is designed with a high-density, deep chassis to reduce the data-center footprint and related costs. A single chassis, 4-node Isilon H5600 has 800 TB of raw storage today. A typical data-center rack holding 10 x H5600 chassis will provide an 8 PB storage building block. Coupled with the industry-leading high usable-to-raw storage efficiency of OneFS, H5600 provides massive capacity, performance and storage density with the low cost of SATA HDD technology. Thanks to its modular architecture, as higher density drives become available and certified, higher density Isilon configurations can be supported — further assuring your investment in the Isilon architecture.

3 Performance benchmark for H5600

This section lists the performance benchmark test results of H5600 for different workloads and protocols.

3.1 Test methodologies

We used FIO for sequential read and write tests. Table 2 lists the configuration that used for both Isilon and benchmark test tools.

Table 2 Isilon and benchmark test tool configurations

Isilon configurations	FIO configurations	Clients configuration
• Use OneFS 8.2.0 for tests • H5600 4-nodes cluster with 2 x 10 GbE NICs (MTU: 1500) • SmartCache is On • Read/Write Transfer Preferred on Isilon NFS export is set to default	• Sequential read workload (128 KiB) • Sequential write workload (512 KiB) • Protection policy is 2d:1n • Set “streaming” as access pattern	• NFS Clients - CentOS 7.3 • SMB Clients – Windows Server 2012 R2 Standard • At least 1:1 client: node ratio in all tests and many use more clients than nodes. • All clients used are bare metal.

3.2 NFS performance test results

Table 3 lists the 4-node H5600 throughput performance benchmark test results for NFS protocols. Different I/O type tests are used to prove the performance for different ADAS workloads.

Table 3 4-node H5600 throughput test results for NFS

Protocol version	I/O type	ADAS workloads	16 KiB	128 KiB	512 KiB
NFSv3	Sequential Read	SiL/HiL process		8.38 GBps	
NFSv3	Sequential Write	Sensor Data Ingestion			4.74 GBps
NFSv4	Sequential Read	SiL/HiL process		7.04 GBps	
NFSv4	Sequential Write	Sensor Data Ingestion			4.98 GBps
NFSv3	Random Read	• Sensor Data Preprocessing • Test Results Data analytics	188.2 MBps		
NFSv3	Random Write	• Sensor Data Preprocessing • Test Results Data analytics	56.4 MBps		

Figure 4 shows the NFSv3 read throughput performance and client latency with 128 KiB read on 4 nodes H5600. The read throughput will continue to grow after increasing the number of threads and kept stable at 8.38 GB/s. The average client latency is less than 1.6ms even with high concurrency.

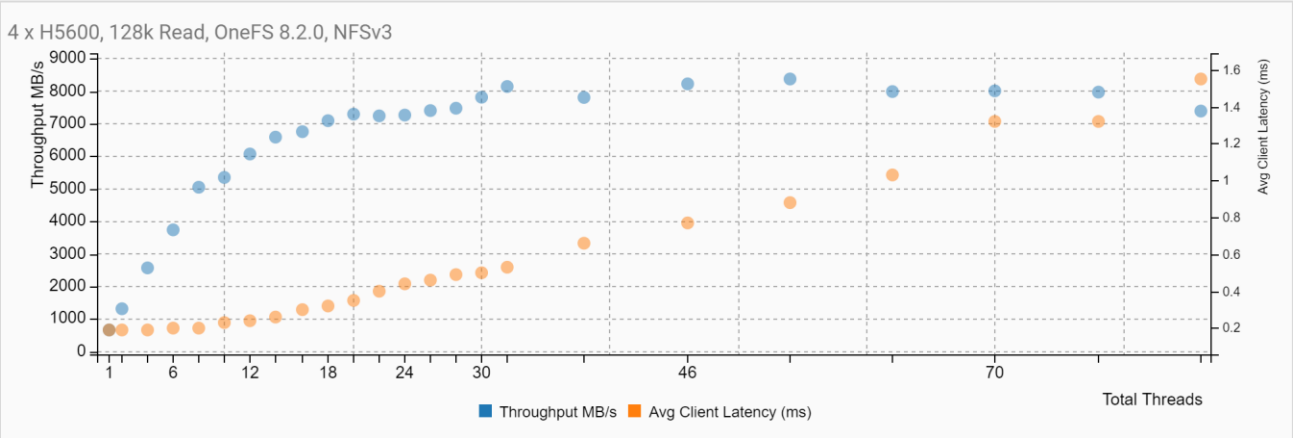

Figure 4 NFSv3 read throughput performance and client latency on 4 nodes H5600

Figure 5 shows the NFSv3 write throughput performance and client latency with 512 KiB writes on 4 nodes H5600. The average client latency is less than 18ms.

Figure 5 NFSv3 write throughput performance and client latency on 4-node H5600

3.3 SMB performance test results

Table 4 lists the 4-nodes H5600 SMB throughput performance benchmark test results. Different I/O tests are used to prove the performance for different ADAS workloads.

Table 4 4-node H5600 throughput test results for SMB

Protocol version	I/O type	ADAS workloads	16 KB	128 KB	512 KB
SMBv2	Sequential Read	SiL/HiL process		7.82 GBps	
SMBv2	Sequential Write	Sensor Data Ingestion			5.32GBps

Protocol version	I/O type	ADAS workloads	16 KB	128 KB	512 KB
SMBv2	Random Read	• Sensor Data Preprocessing • Test Results Data analytics	130.2 MBps		
SMBv2	Random Write	• Sensor Data Preprocessing • Test Results Data analytics	75.4 MBps		

3.4 H5600 storage efficiency

Figure 6 lists the H5600 storage efficiency from 4 to 252 nodes with raw and usable capacity with recommended protection policy. With a single namespace, Isilon H5600 can easily scale out to 50.4 PB raw capacity with 252 nodes.

For all current Gen 6 hardware configurations, the recommended protection levels are either '+2d:1n' or +3d:1n1d'. The recommended protection level depends on the size of the node pool and node types. For the H5600, with 10 nodes or less, the recommended protection level is '+2d:1n'. With more than 10 nodes, the recommended protection level is '+3d:1n1d'.

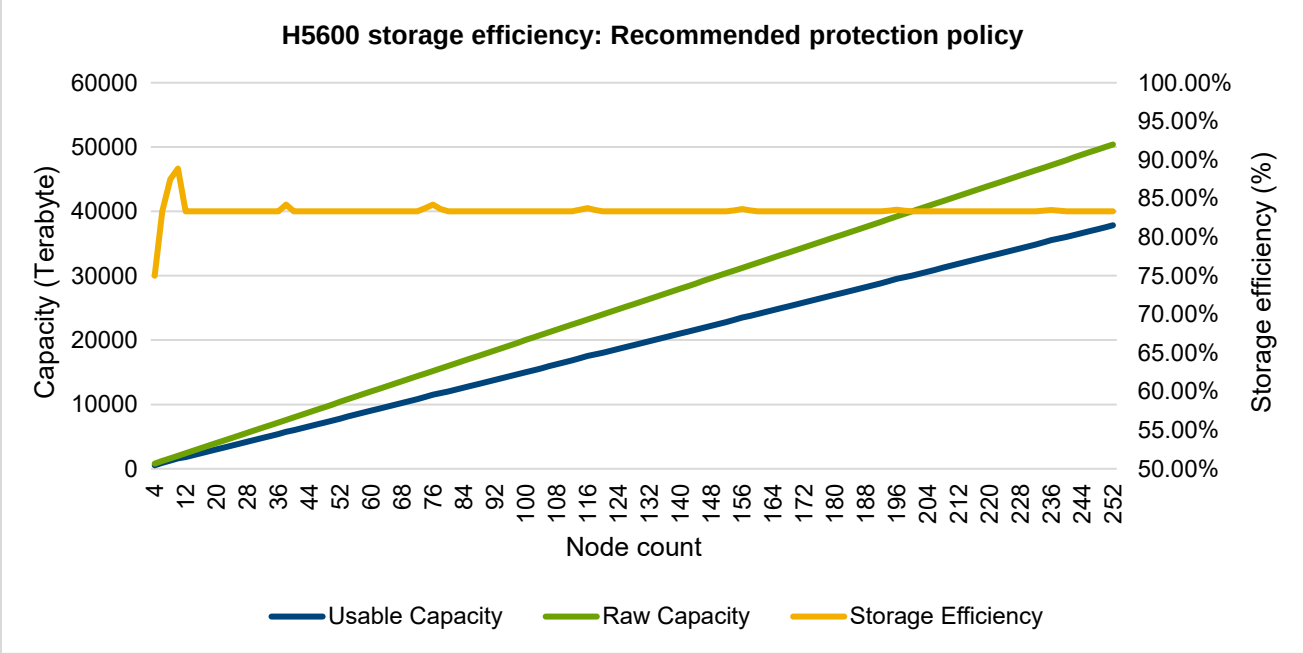

Figure 6 H5600 storage efficiency from 4 nodes to 252 nodes

4 Dell EMC ADAS solution architecture

As shown in Figure 7, Dell EMC provides a robust ADAS solution architecture for our ADAS customers. This solution has been deployed at multiple customer sites, including each stage of the ADAS development cycle. Examples include the following:

- **In-vehicle system:** Includes the ADAS logging systems and data ingestion process to gather all the sensor data from fields.
- **Data lake platform:** Stores all the sensor data for developing and validation. It will also store all the test results for future data analysis. It can also easily integrate with public/private cloud platform through Dell EMC CloudPools technologies.
- **Metadata management service:** Allows users to search ADAS sensor data through metadata info and presents a dashboard for ADAS data platform. Isilon data management system is recommended here.
- **Data analytics platform:** Contains different open source big data platforms like Hadoop, HBase, or Spark on Dell EMC servers. These integrate with Dell EMC Isilon storage through HDFS to analyze the data for business decisions.
- **HiL test cells:** Used to test real-time embedded systems by streaming data from Isilon with low latency.
- **SiL farm with HPC servers:** Used to run batches of simulation tests in parallel on Dell HPC clusters and require high throughput from Isilon.
- **App development:** Builds a continuous integration and platform as a service with our Pivotal team to deploy software on Kubernetes empowering the engineering team to development with agility and speed.
- **AI GPU platform:** Enables the develop team to train deep learning technologies on sensor data by leveraging open source AI frameworks like TensorFlow/Caffe2 on Dell EMC HPC servers with GPU enabled.
- **VDI platform:** Hosts the virtual desktop clients and allows end users and remote team to access the data with highly flexible and secure.
- **Operational management platform:** Manages all the hardware and software in the whole solutions and provide monitoring and alert system.

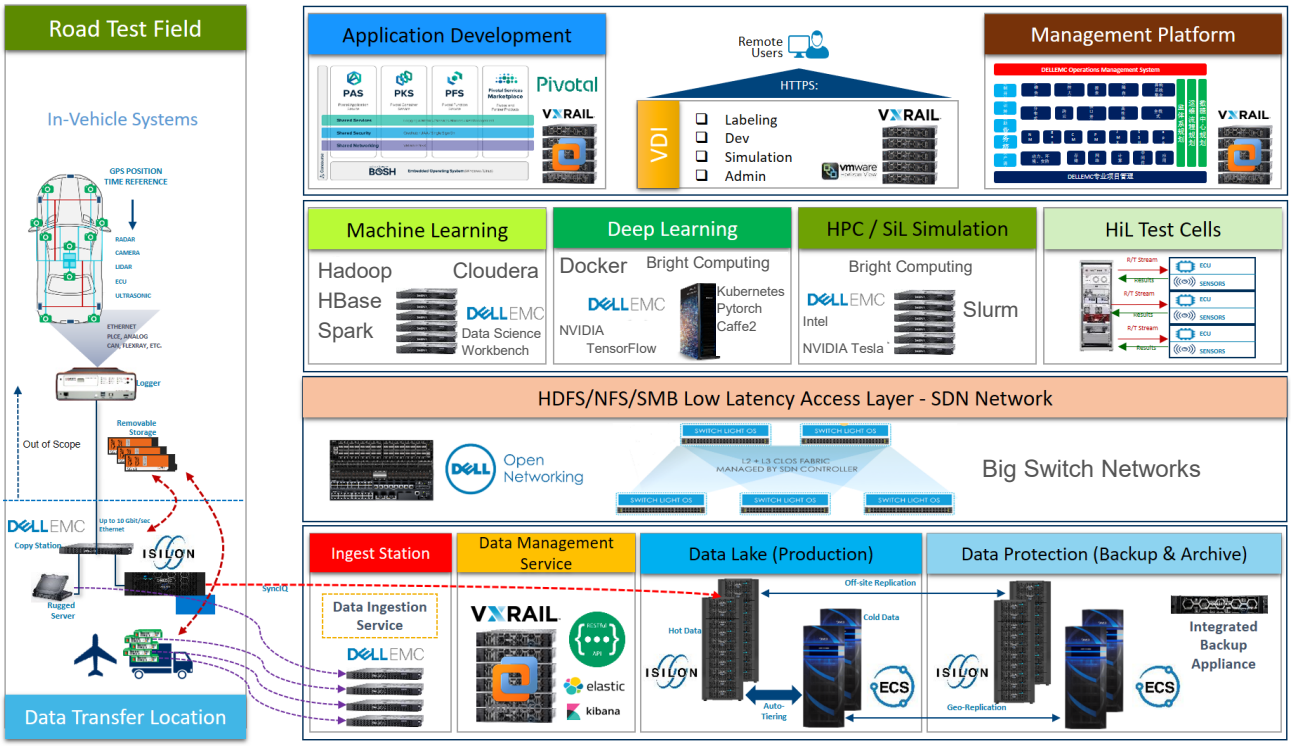

Figure 7 Dell EMC ADAS solution architecture

A Technical support and resources

Dell.com/support is focused on meeting customer needs with proven services and support.

Storage technical documents and videos provide expertise that helps to ensure customer success on Dell EMC storage platforms.

A.1 Related resources

Isilon H5600 Storage Tech Specs

Dell EMC Isilon for ADAS and Autonomous Driving

Top 5 reasons to choose Dell EMC Isilon for AD/ADAS

Solving the storage conundrum in ADAS development and validation