Solving the storage conundrum in ADAS development and validation
# Table of Contents

The need for speed in ADAS development ........................................... 3
The importance of ADAS development and validation infrastructure ........ 3
ADAS development and validation workflow and methodology ............... 4
Major challenges of ADAS development and validation infrastructure ....... 6

- **Explosive data growth** ................................................................. 6
- **Ability to deliver performance at scale is key** ........................... 6
- **Data retention, restoration and protection** .................................. 6
- **Costly data duplication** .............................................................. 7

Dell EMC Isilon for ADAS and Autonomous Driving ......................... 7
Key summary .................................................................................... 10
1. The need for speed in ADAS development

The automotive industry is in the midst of a highly competitive transitional period, with the ultimate goal of fully autonomous or "driverless" vehicles likely to be realized within a decade. The scale and intensity at which OEMs and Tier 1 suppliers must bring innovations to market – while containing costs, mitigating risks, managing product complexity and maintaining compliance – is challenging. The emergence of Advanced Driver Assistance Systems (ADAS), designed to enhance passenger, vehicle and road safety, introduces disruptive requirements for engineering IT infrastructure – particularly storage, where even entry-level capacities are measured in petabytes. This paper will explore the infrastructure challenges facing OEMs and Tier 1 suppliers in developing and validating ADAS technologies, and propose a storage solution that is optimized for such workloads, delivering high performance, high concurrency and massive scalability.

2. The importance of ADAS development and validation infrastructure

Development and validation are key strategic stages of the ADAS lifecycle. Since many of these systems directly affect safety, the robustness and reliability of ADAS functions is paramount. Today, ADAS typically includes features such as automatic braking, collision protection and emergency assistance. As the technologies mature, evolving legislation, together with recommendations for enhanced protocols from independent bodies such as Euro NCAP and ANCAP, could see such assistance systems fitted as standard equipment, particularly as vehicles shift towards greater autonomy. This demands exhaustive testing and verification to represent diverse traffic scenarios and dimensions, which might include road geometry, driver and pedestrian behaviors, traffic conditions, weather conditions, vehicle characteristics and variants, spontaneous component faults, security, and more.

Given the logistical impracticality of accounting for all these scenarios using field tests alone, laboratory development and tests are typically conducted as a precursor. These employ Model-in-the-Loop (MIL), Software-In-the-Loop (SIL) and Hardware-In-the-Loop (HIL) test systems which connect physical and simulated hardware with model-generated and real-world data captured by a fleet of test vehicles. These laboratory tests are an important method for analyzing and validating the vast amount of potential test scenarios at reasonable effort and cost.

Because of the unique value of laboratory development and tests, it is vital to have a shared, reliable and effective ADAS infrastructure for consistent development and validation. The infrastructure is used to store vast libraries of real-world scenarios including sensor and video data; train software algorithms with these data; replay with high performance to concurrently test multiple copies of the ADAS subsystems; and support data analytics and data archiving.
ADAS development and validation consists of various design, development and test stages, with multiple aspects considered. Adoption of the ISO 26262 safety standard is widespread across the automotive industry. For safety-critical functions, ISO 26262 recommends more advanced test methods to cover scenarios such as fault injection and erroneous decisions by ADAS. This adds complexity and effort to ADAS development and validation phases and, as a result, increases development time and cost.

Figure 1 visualizes the various testing levels used to carry out the development and validation of critical ADAS functions and software. The higher the desired level of automation, the greater the development and validation efforts involved in these safety-critical assistance systems.

- **Design and requirement** involves the development team understanding the market requirements, defining and proving the concept of ADAS functionality, and designing the architecture, physical layouts, and interaction with systems.

- **Coding level development and tests** allow engineers to train and develop the algorithms using deep learning technologies, and evaluate and validate functional concepts and requirements at code level. A model of the developed system is integrated in a simulation loop with models of vehicle dynamics, sensors, actuators, and the traffic environment.

- **Electronic Control Unit (ECU) systems and subsystems development and test stages** allow engineers to reuse scenario libraries, combining SIL and HIL simulation with real-world data using shared infrastructure. SIL runs in a virtual environment without any hardware, and is quicker and less costly to run than HIL, but is not deemed sufficiently accurate for government safety certification. HIL offers superior accuracy by employing a prototype of actual hardware, but is more expensive, is typically available late in the project, and can only be run in real-time. This stage is highly complex due to the large number of test cases and the response of ADAS features to the state of the vehicle and its environment. At this stage, a high performance, scalable, reliable and cost-effective infrastructure is needed for the engineering team to recreate realistic test cases and analyze test results in the lab.
• **Field validation** – there is no substitute for real-world testing. This stage allows engineers to validate ADAS functions using physical drivers in real vehicles on a test track and, eventually, in real-world situations (with heavy oversight).

Simulation, required models of ADAS, and others. Tests are driven to cover all possible corner cases, with discrepancies between the ECU validation and test driver actions identified as potential bugs.

---

Figure 2 shows the detailed workflow of the laboratory development and validation of ADAS, detailed below:

- **Data Capture** – Huge volumes of sensor data are captured by a fleet of test vehicles, which may comprise video sequences, sonar, radar, LIDAR and GPS, with more advanced sensors such as 4k and even 12k video in the works. Some of these sensors will be pre-production samples of the actual sensors intended for the production vehicle, while other sensors may be capturing high-resolution reference data around the test vehicle. The data is stored in real-time into removable onboard drives and loaded daily to a centralized ingestion server, either overnight or between shifts.

- **Data Preparation** – Once the data has been ingested, the engineering team will start to prepare the data including reviewing, labeling and adding metadata such as weather and traffic conditions. Analysis is typically a manual process but in future may be performed by AI or machine learning.

- **Design and Development Phase** – When the data is ready, the ADAS engineering team can develop and build algorithms for ECU models through deep learning and iterative testing, using fusion data from all the sensors, GPS, weather and road or environment data.

- **Test Preparation** – At the same time, engineers can build test suites including design test cases, required models of ADAS, and others. Tests are driven to cover all possible corner cases, with discrepancies between the ECU validation and test driver actions identified as potential bugs.

- **Validation** – As test cases are defined, the engineering team can schedule simulation runs on the HIL / SIL cluster. This involves “replaying” the captured raw sensor data back through the test farm – usually many iterations running in parallel.

- **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 team with on-demand access.

- **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. Many OEMs stipulate service level agreements (SLAs) of 1-7 days for simulation data restoration time – for example, in the event of a safety recall – to allow quick turn-around of updates.
4. Major challenges of ADAS development and validation infrastructure

ADAS workloads require high-performance storage at petabyte scale. As data volumes grow, the limitations and flaws of traditional legacy storage architectures become increasingly hard to ignore.

Explosive data growth
Since many ADAS systems are safety-critical, data capture requirements are high and will increase exponentially the greater the level of automation. The Society of Automotive Engineers (SAE) has defined six levels of automation in ADAS:

- **Today’s development efforts (SAE levels 2-3)** have a typical requirement for 200,000 km to 1 million km of real-world data to be captured by a fleet of test vehicles in order to simulate sufficient conditions for software development and validation. Level 4 automation will require 2+ million km of data, and as the industry moves towards level 5 (fully autonomous vehicles), this will increase to approximately 240 million km.

A typical vehicle used for data collection in the ADAS system test use case is equipped with multiple sensors and cameras generating data. While reality is impossible to predict, this high level of visibility and redundancy builds a detailed picture to enable the vehicle to make response decisions in adverse weather conditions or in the event of individual component failure. This creates a massive challenge in terms of the scale of the sensor data that must be captured and replayed to test ADAS subsystems.

To illustrate, a typical SAE level 2 ADAS project, capturing 200,000 km of driving at an average speed of 65 km/h, would generate over 3,076 hours of data, requiring approximately 3.8 petabytes of storage for one single sensor – where, depending on the specific ADAS feature, multiple sensors would be required. An SAE level 3 ADAS project, which involves capturing 1,000,000 km of driving, could generate 19.3 petabytes of raw sensor data.

**Ability to deliver performance at scale is key**
As autonomous vehicle development advances along the SAE scale, it will become imperative that storage is architected to handle ever-increasing performance demands. Performance requirements are hard to accurately forecast, with many more sensors predicted to be added to test vehicles, some of higher data rates. Therefore ADAS validation architecture needs to be future-proofed by enabling storage capacity to be upgraded seamlessly, linearly and non-disruptively, without impacting storage performance.

The data preparation phase involves intensive data processing, which imposes high performance and bandwidth demand from storage to read and write streaming raw video data and sensor binary files. To handle ever-growing sensor data, the ADAS infrastructure requires storage that can deliver consistent performance for ingesting and streaming massive volumes of raw data concurrently – as most projects will continue to ingest sensor data even after validation has started.

As more and more sensor data is captured, and project footprints grow, maximum ROI is achieved by adding storage throughout the project. This requires storage that can be expanded without incurring the delays associated with massive data migration, system downtime or interruption to data availability – all common challenges when upgrading traditional storage platforms.

**Data retention, restoration and protection**
While governance around data retention has not been standardized at an international level, most OEMs require that ADAS data must be retained for multiple decades.
most likely, the lifespan of the vehicle. The specific period depends on the prevailing country regulations and contractual stipulations. Although data is archived once a project is complete and the car moves into full production, it must be readily available for restoration: for example, in the event the ADAS system requires an update, a hardware bug is identified or, in the worst case, a safety recall is triggered. In the event of a future lawsuit due to an accident, it is also vital that manufacturers can provide on-demand access to the necessary data, to determine liability.

For these reasons and more, most OEMs require suppliers to be prepared to restore simulation environments with petabytes of data within a matter of days. This rules out tape as an archive option, as the data cannot be restored within a reasonable timeframe (nor reliably). Equally, while cloud offers inexpensive storage, the cost of data egress becomes prohibitive at this scale.

Lastly, as connected and autonomous vehicles become more ubiquitous, they may become a more interesting target for cyber threat actors. Data relating to ADAS functionality must therefore be protected against loss, human error and malicious activity, with data governance policies strictly enforced.

Costly data duplication
As data from millions of miles of real-world driving is collected, analytics can be used to query specific driving situations (corner cases) within the test data. For example, "show me all situations where the driver was accelerating while the ECU triggered Emergency Brake Assist". These exceptions between the recorded actions of the test driver and the intentions of the ADAS algorithm can represent design bugs.

Central to analyzing large-scale sets of unstructured file data is Hadoop, a powerful analytics engine that is traditionally implemented on dedicated infrastructure. This alone adds unnecessary capital costs and is management-intensive. The manual ingest of large data sets into a separate Hadoop cluster is time- and resource-consuming, which can lead to significant delays in extracting insight from the Hadoop analytics effort. It also means data and analytics results can’t be accessed or shared easily with other enterprise applications due to the lack of industry standard protocol support, further hindering development efforts.

5. Dell EMC Isilon for ADAS and autonomous driving
With an architecture that is ideally suited to ADAS development and verification use cases, Dell EMC Isilon® scale-out NAS provides petabytes of storage capacity in an ever-expanding single namespace. Isilon combines high-performance storage, 10/40GbE connectivity and a scale-out architecture to provide fast, concurrent connectivity. By utilizing Dell EMC Isilon Storage for ADAS verification, OEMs and Tier 1 suppliers can maintain reliability and time-to-market requirements in a scalable, cost-effective manner.

Dell EMC Solution Architecture for ADAS HIL validation
A simplified HIL architecture is illustrated in figure 4.

![Figure 4. ADAS Infrastructure for HIL validation](image)
As noted earlier, data from various sensors is typically captured on rugged and removable SSDs, which are installed in the test vehicles. At the end of each test driving shift, the SSDs are swapped out for fresh ones to allow the next shift to commence without delay. The full SSDs are usually copied onto some other cheap storage transport medium (e.g. linear tape open (LTO) cartridges) in the field, and then shipped overnight to a location with high-bandwidth direct local networking (e.g. 10 Gb/s or 40Gb/s) where they can be ingested to the central storage system. Dell EMC Isilon provides 10/40GbE connectivity and a simple scale-out storage platform to accelerate the ingestion of massive volumes of raw sensor data.

At the data enrichment stage, video, radar and LIDAR sequences need to be time-synchronized, labelled and indexed to enable developers to query specific video sequences. With Dell EMC Isilon, each node adds capacity, performance and resiliency to the cluster, and each node can process requests from developer clients at the same time as data is being ingested. This allows the ADAS development team to take advantage of the entire cluster’s performance, and makes Dell EMC Isilon ideally suited to the ADAS workloads of today and tomorrow.

For the simulation part of the process, the data streams are read by HIL servers that streamed the data into the devices under simulation. Relevant simulation data is generated from storage and forwarded to the devices, for instance via the Controller Area Network (CAN). In the case of a camera, this typically comes with an embedded Engine Control Unit (ECU). The simulation receives and processes output signals of the camera ECU in the form of messages about detected objects or other responses by ADAS systems.

Because the real physical device is part of the simulation workflow, this architecture uses HIL test methodology. The HIL testing is typically run in real-time; the need for accuracy is for quality assurance purposes as well as regulatory requirements. Consequently, the only way to accelerate the test and simulation time is by running dozens – and sometimes hundreds – of HIL servers and devices in parallel. Each of these streams requires a throughput of 150-400MB/s. Therefore with 80 HIL servers reading at 250 MB/s, the cluster needs to be able to handle 20 GB/s read throughput. Since Isilon distributes data and load equally across all nodes in the cluster, this is eminently achievable. At the same time, additional GB/s may be ingested from newly-loaded data.

Dell EMC Isilon gives the development team flexibility and choice by enabling the cold data to be archived to private cloud, such as Dell EMC Elastic Cloud Storage™ (ECS), another Isilon cluster, or public cloud with third party vendors. For more detail, refer to the section on data retention.

**Dell EMC Solution Architecture for ADAS SIL Validation**

A simplified SIL architecture is illustrated in figure 5.
Compared to the HIL simulation, SIL tests can be run much faster. Here, the ECU software is run in a simulator and many more compute cores are available via commodity x64 server hardware. Additionally, the simulation may be run at accelerated speed as it is not limited to the speed of physical hardware, such as a camera. SIL simulation can be run in the early stages of the development cycle and demand for this is increasing. While quality and regulatory requirements currently limit the number of simulation cycles for SIL, this is likely to change in future.

Dell EMC Isilon brings unique value to ADAS Infrastructure

In all these ADAS infrastructures and workflow, Dell EMC Isilon provides the unique value for development and validation:

- **Capacity and scalability**
  Dell EMC Isilon meets the ADAS infrastructure requirement to scale from terabytes to petabytes of capacity in a single file system. Capacity can be added to existing Isilon volumes in minutes, with no downtime or interruption to data availability, and no reduction in performance. The scale-out architecture and OneFS operating system eliminate the I/O bottlenecks that are common to legacy storage solutions when highly concurrent automotive workloads are at play.

  At the same time, the ADAS infrastructure must avoid the traditional RAID architectures that require the administrator to deal with hundreds or even thousands of RAID arrays, aggregates, and volumes and file systems. The Isilon file system, OneFS, is a node-based architecture that allows up to 144 nodes, with only a single file system to administer. The scale-out architecture of Isilon and its simplicity makes it unique in the market and the solution of choice for the ADAS industry.

- **Performance at scale**
  As autonomous vehicle development advances along the SAE scale, performance requirements will be hard to accurately forecast; it is therefore imperative that storage performance scales predictably with capacity, thereby future-proofing the solution investment. In contrast to traditional storage systems that must "scale up" when additional performance or capacity is required, Isilon enables the system to "scale out", seamlessly increasing the existing file system or volume into petabytes of capacity while simultaneously increasing performance in a linear fashion. Isilon can also optimize the access pattern for different workloads at the file or directory level. It optimizes the I/O performance for high-speed streaming workloads to enable very fast ingestion or reading with a single client.

  - **Data retention**
    To cost-effectively meet commitments for multi-decade test data retention and commonly mandated restoration and re-simulation times, Isilon's CloudPools feature supports data tiering. This policy-driven automated tiering solution allows data to be aligned to the optimal price/performance storage tier, depending on the stage of the project. Critical data, such as current ADAS sensor data, can be kept on a higher-performance tier, while less critical data, such as that from a released vehicle, can be assigned to a more cost-effective, higher-density archive tier.

    Archived ADAS data can be automatically migrated back to higher performance tiers in the event of an urgent need to restore a simulation environment. Whenever the development team needs to retrieve the data, it can be accessed as before without changes to policies and procedures.

  - **Native analytics support and Multi-Protocol Access**
    Isilon OneFS provides access to the data not only through typical NAS protocols like NFS3, NFS4, SMB2, SMB3 and FTP, but also HDFS (Hadoop Distributed File System). By combining the highly efficient storage platform with native Hadoop integration, Isilon enables ADAS developers to accelerate their analytics efforts and get results in minutes. Native HDFS support eliminates the time and expense of moving large data sets between File and Hadoop storage.
• **Artificial Intelligence and Deep Learning**
  The ADAS development team can leverage new deep learning and machine learning solutions such as TensorFlow. Deep learning algorithms can be trained and validated by labeled sensor data and test use cases within the same infrastructure, as shown in Figure 6 below.

![Figure 6. ADAS Infrastructure for Deep Learning](image_url)

• **Manageability**
  Isilon streamlines the management of petabyte-scale storage by enabling storage capacity and performance to be increased quickly and easily. Nodes can be added to the file system and be ready to use in minutes, without downtime or manual data migration. Nodes can be mixed and matched from different generations in a single cluster, while an automated background process moves data from older nodes being retired to newer ones. The Isilon OneFS file system also provides non-disruptive system upgrades to facilitate lights-out operations, with no loss of connectivity during the upgrade process.

6. **Key summary**

A key challenge facing OEMs and Tier 1 suppliers is to make the ADAS development and validation processes as seamless as possible to get innovation to market faster. Due to the evolution of sensor technology, development and simulation methodologies, ADAS imposes massive capacity and performance requirements on shared storage which can only be satisfied by a distributed filesystem architecture. It is therefore vital to underpin the ADAS development framework with a shared storage architecture that can scale out to petabytes of capacity while delivering predictable performance.

Dell EMC Isilon meets the automotive industry’s need for a game-changing storage solution optimized for high performance, high concurrency and massive scalability and, as such, is playing a key role in autonomous vehicle development. To find out more ›