Red Hat

Open source storage for a data-driven auto industry

Self-driving vehicles and data-rich cars require new development platforms

Automakers and technology companies are rushing to enter the market for self-driving vehicles. Gartner analysts predict that "many of the most powerful changes expected in the auto industry will happen in the next two to five years."[1] They also forecast that, "Longer term, electric vehicle technologies and autonomous vehicle technologies will dominate the transformation of the auto industry."[1] The right software platform is critical to research and development (R&D) efforts for creating the next generation of cars. To provide real-time, effective safety and mobility, autonomous vehicles need highly secure and reliable networks, applications, and data analytics—all of which are powered by software.

"Longer term, electric vehicle technologies and autonomous vehicle technologies will dominate the transformation of the auto industry."[1]

Gartner

Transforming the industry also requires making appropriate investments in various technologies, and underlying development platforms become a vital consideration in the investment strategy. Putting fully autonomous vehicles on the road means auto manufacturers must invest significantly in the R&D phase. At the same time, competition in the autonomous vehicle market is starting to draw players from other industries, including software and hardware. These factors put pressure on auto manufacturers to bring the best products to market ahead of the competition.

This paper describes how open source storage software can accelerate development of safer and more innovative autonomous vehicles. It also describes how IT organizations can store and use the information produced by today's data-rich vehicles to improve their companies' products.

Data is revolutionizing the automotive industry

Hundreds of sensors and microcontrollers in today's vehicles monitor everything from engine speed to suspension stiffness to the distance from the car ahead.

Some sensors, such as radar devices, create data by looking outward. Some generate data by looking inward and measuring factors such as engine power output, exhaust emissions, and suspension spring rates. Other devices monitor speed or ambient conditions and connect car-to-car and car-to-infrastructure (e.g., smart cities).

A fleet of autonomous vehicles can produce a massive "digital exhaust"—perhaps a petabyte of data in a day. That is roughly equivalent to 10 billion Facebook photos. Even conventional vehicles create up to 25 gigabytes of data every hour.

Drivers as well as microcontrollers and onboard computing devices use this data in real time to improve fuel efficiency, avoid accidents, and execute other functions. Automakers can also use collected data to improve the quality, safety, and performance of their vehicles. Among these, safety ranks as the primary concern for drivers and manufacturers. Even minor malfunctions can lead to serious accidents. If the automaker is at fault, they could suffer serious financial losses and brand damage.

Automotive data scientists and engineers are collecting, sorting, analyzing, interpreting, simulating, and iterating vast amounts of data to make a vehicle to navigate the road accurately without human input. Supporting the development process with the right IT infrastructure is essential.

1 *Gartner. "Hype Cycle for Connected Vehicles and Smart Mobility, 2020," Michael Ramsey, 17 July 2020.*

IT's role in autonomous driving R&D

For long-term success, an automaker's autonomous driving R&D IT platform needs to be:

- Open source for extensibility and integration with state-of-the-art application services and technologies, including artificial intelligence (AI) and machine learning (ML).
- Agile, scalable, and robust to accommodate the volume of applications and data needed to manage analytics for autonomous vehicles, from R&D to pilot projects and commercial adoption.
- Optimized to orchestrate the analytics, ML, and simulation workloads flexibly and at scale based on graphics processing unit (GPU), central processing unit (CPU), memory, and storage.
- Highly secure to prevent hacking that compromises driving and safety feature integrity.
- Flexible so that it can adapt to workload changes and foster collaboration, innovation, and performance.

Autonomous vehicles versus conventional cars

This explosion of data is most pronounced in self-driving vehicles. However, it is difficult to predict whether consumers will adopt autonomous vehicles. Adoption and development will likely fall along a spectrum from fully autonomous to conventional non-autonomous vehicles.

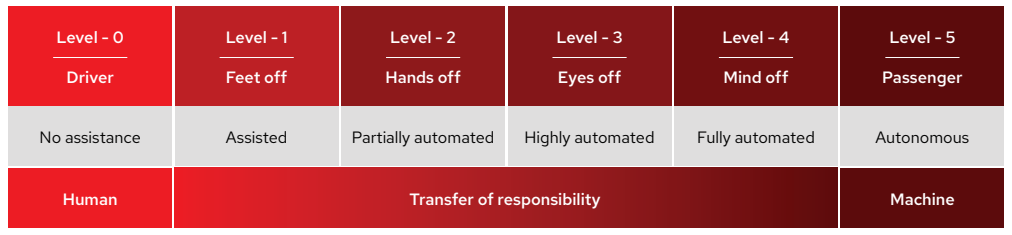

Figure 1. Spectrum of vehicular autonomy from full human control to the driver as passenger.

There is some confusion about the definition of "autonomous vehicles," which spans a range of vehicle types. The Society of Automotive Engineers defines five separate levels of autonomy, ranging from complete human control to complete machine control. Higher autonomy levels mean more technology is needed and more data needs to be stored.

- L0: Human driver controls everything except what is transparent to the driver like engine fuel injection systems.
- L1: A single-driver function is automated, such as cruise control or lane-keep assist.

- L2: Multiple driver functions—such as cruise control, lane control, steering, and braking—are digitized.
- L3: Conditional automation while the human driver remains on standby to take over in an emergency.
- L4: Fully autonomous with no human driver control.
- L5: Driver assumes the role of passenger in a fully autonomous vehicle.

The average self-driving car creates 1 to 15 terabytes (TB) of data per day. Experts differ on the exact number, but self-driving vehicles could generate as much as 450 TB data per day gathered by sensors, radar, and cameras.[2] By any measure, that represents a massive amount of field data. The data is collected under constantly changing conditions that include traffic lights, signage, pedestrians, weather, other vehicles, and navigational data. All this information needs to be quickly stored, processed, and structured in ongoing cycles. However, we may see partially automated vehicles that provide limited features like lane control and advanced radar-assisted cruise control.

Even on the non-autonomous end of the spectrum (Level 0 in Figure 1), drivers and automakers can benefit from data-rich vehicles. For example, back up and 360-degree high-resolution cameras—highly valued by consumers—typically produce data at ranges from 500–11,500 Mbit per second.

Automakers can also use the data collected by sensors in brake systems, climate control systems, suspension, and transmissions, among others, to refine future vehicles. Data collected from vehicles-in-use could potentially help manufacturers detect safety issues and initiate recalls before a software or hardware defect causes a crash.

Challenges of storing automotive data in the cloud

Using the data generated by Industrial Internet of Things (IIoT) devices, embedded microprocessors, and onboard monitoring and control systems creates a number of challenges. First, with so much data floating around, storing it locally becomes a huge challenge. Greater vehicle-to-vehicle and vehicle-to-infrastructure communication also makes data access even more critical. However, access to high-speed networks limits access to public cloud for storage. It is neither economical nor technologically feasible to store this data in the cloud because AI/ML algorithms need rapid access to the learning data sets.

Cost

The cost of storing large amounts of data in the public cloud can be prohibitive. Egress costs from major public clouds can also be exorbitant.

Security and compliance

Training datasets are considered intellectual property, and therefore need to be stored securely while allowing access to authorized users. As vehicle-to-infrastructure communication increases with the advent of 5G networks, it will become even more critical for data and analytics to be compliant with government-set regulatory standards.

Automotive data storage solutions tailored to today's vehicles and smart cities

As an open software platform company, Red Hat offers flexible storage as part of a vision for edge computing. Open source software gains its strength from diverse, global developer communities. These communities solve problems and develop new technologies collaboratively. The result is more innovative, iterative, stable, and secure technology.

Massively scalable storage for demanding applications

Red Hat® Ceph® Storage is an open source, massively scalable, simplified storage solution for modern data pipelines. Engineered for data analytics, AI/ML, and emerging workloads, it delivers software-defined storage in cloud, on-premises, or hybrid environments and offers a number of benefits.

- Decouple software from hardware to run cost-effectively on industry-standard servers and disks.
- Scale flexibly and massively to support multiple petabyte deployments with consistent performance.
- Provide web-scale object storage for modern use cases, such as cloud infrastructure, media repository, and big data analytics.
- Combine Red Hat Ceph Storage with a storage management console, deployment tools, and support services.

Red Hat Ceph Storage is suited for automotive applications because it can scale to multiple petabytes without compromising performance or cost. It also improves access because data stored in Red Hat Ceph Storage can be accessed via high-speed S3 or RADOS Gateway (RGW) interfaces. Multiple analytics clusters can access the data for faster analysis with minimal data movement.

Red Hat Ceph Storage offers the open source advantage. Retailers, banks, and government agencies have deployed it to create efficiencies and reduce operating costs.

Finally, it enhances security and compliance in complex automotive data applications. Red Hat Ceph Storage is built on Security-Enhanced Linux® (SELinux) and supports federal security standards such as federal information processing standards (FIPS). It integrates responsively with third-party security and user authentication solutions.

Contact Red Hat to learn how open source solutions can address requirements for scalability, access, cost, and security of data storage in the automotive industry.

About Red Hat

Red Hat is the world's leading provider of enterprise open source software solutions, using a community-powered approach to deliver reliable and high-performing Linux, hybrid cloud, container, and Kubernetes technologies. Red Hat helps customers integrate new and existing IT applications, develop cloud-native applications, standardize on our industry-leading operating system, and automate, secure, and manage complex environments. Award-winning support, training, and consulting services make Red Hat a trusted adviser to the Fortune 500. As a strategic partner to cloud providers, system integrators, application vendors, customers, and open source communities, Red Hat can help organizations prepare for the digital future.

facebook.com/redhatinc
@RedHat
linkedin.com/company/red-hat

North America
1 888 REDHAT1
www.redhat.com

Europe, Middle East, and Africa
00800 7334 2835
europe@redhat.com

Asia Pacific
+65 6490 4200
apac@redhat.com

Latin America
+54 11 4329 7300
info-latam@redhat.com

redhat.com
#F24748_1020