Deep Science Publishing
https://doi.org/10.70593/978-93-49307-21-6_10

Chapter 10: Designing scalable data architectures for automotive systems using cloud computing platforms

10.1. Introduction

Cloud computing is ushering in a new era of automotive applications. The automotive industry is experiencing paradigm shifts toward more sophisticated computation, connectivity and collaboration. Driven by advances in sensing and intelligence technologies, vehicles have an increasing number of onboard computers and embedded sensors for perception and data acquisition. Conventional vehicular applications, such as navigation based on on-board maps, are evolving into computation and data intensive ones, termed cloud-based vehicular applications, relying on external computing and/or data resources. Clouds providing high-end computation and storage, and big data generated by networked vehicles are newly available. Ready access to distributed information, computing and storage resources can therefore enable new vehicular applications. Heterogeneous information and data sharing among vehicles, infrastructure, and the cloud may improve computation and data intensive vehicular safety, drivability, fuel economy, and infotainment. This paper aims to stimulate innovative designs of novel and scalable data architectures for new cloud-based automotive applications on cloud platforms. Cloud-based applications, including cloud-based vehicle speed optimizer and cloud-aided comfort-based route planner, that help improve vehicular safety/comfort and fuel economy for everyday driving scenarios, are proposed. Scalable and efficient computing frameworks to execute these applications on public storing cloud service are presented. Comprehensive evaluation methods are studied to analyze the performance of the computing frameworks, including their effectiveness on service scalability and efficiency. Comparative simulations results on fleet size, vehicle density, and transmission distance parameters are reported, followed by recommendations for future systematic verification studies. Cloud computing is providing great opportunities for computation and data intensive vehicular applications.

The enabling architectures of such applications for vehicle safety/comfort and fuel economy improvement are proposed. In addition, the generic evaluation method on these and/or other new applications' scalability and efficiency on public storing cloud is presented.

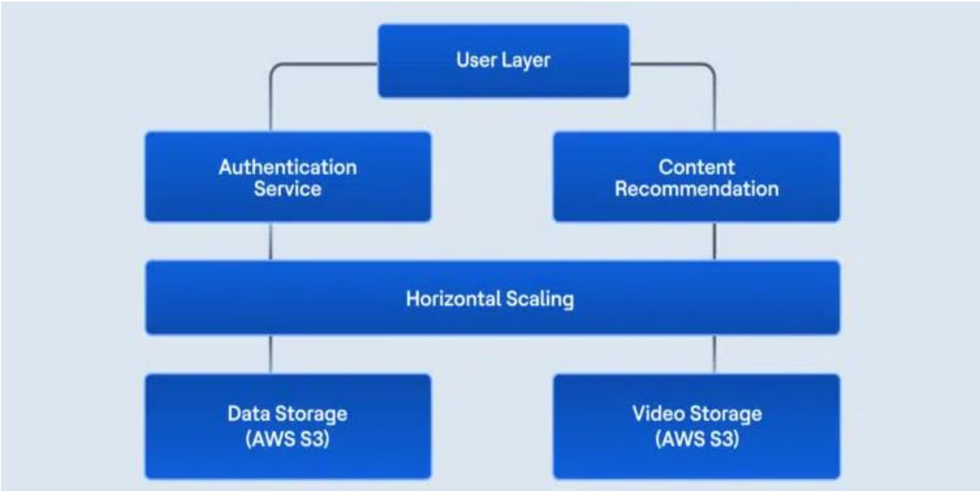

Fig 10.1: Scalable Data Architecture

10.2. Background and Motivation

The preliminary research into automotive data exploitation plans has translated into action on how this data can provide value to consumers, suppliers, and the automotive industry. The challenges of data privacy and ownership, as well as the automotive data handling company´s obligations, will be reviewed. Data-driven, edge-based applications and services such as predictive maintenance, data-driven feature updates, and insurance use cases are explored, as are the edge-to-cloud and edge-to-edge processing architectures.

Vehicle manufacturers are adding advanced sensors to their onboard ADAS to enable better environment understanding and support automated driving and safety. This competition to incorporate cutting-edge ADAS features has not only come with the understanding that the better the onboard intelligence is, the better the mobility system will be, but also with the addition of connectivity devices for telemetry and maintenance diagnostics, as well as other value propositions and communication channels to the user. OEMs and Tier-1s have identified that a significant part of their future income will come from exploiting the data captured in the vehicle, and hence as much as possible processed intelligently and in a right-time manner. On the other hand, internet giants such as Google, Apple, AWS, and Microsoft are positioning themselves in the automotive data marketplace, having a role in data value exploitation that is expected not to be under OEM or Tier-1 control. Currently, the data generated out of car sensors are being processed by either the onboard computing systems (as with ADAS processing) or a cloud system from the manufacturer (as for processing telemetry/fleet maintenance data for over-the-air updates and diagnostics), but only the manufacturer, at the end of the day, is able to access both sources of data.

Leverage data for added value to the customer, with the challenge of how such intelligence can be generated from data. To transform OEMs and Tier-1s into data-driven companies, it is proposed to decompose the overall data value business case into the various stages considered. These stages would be to create an automotive data ecosystem (as the core), which is expected to capture and connect data producers, data consumers, and data. Afterwards, to generate a data space with trusted and secure ownership and access control for producing, sharing, and processing data, both at edge and cloud requested with resources. Then, to be able to ingest and settle the data, with established data pipelines and/or on-demand services. Ultimately, to be able to exploit the processed data, upstream or downstream on the concurrent development of data-driven applications, services, and processes.

10.2.1. Research design

Knowledge about a system depends on the subscribed categories, being raw or processed data stored in cyberspace. An important aspect during the design of CPS is to address excessive data flow which may lead to infinite objectives on maximizing data transfer. This information may concern assessment of items in a manufacturing process, quality indicators of items, a set of data computed by a sensor based on data from a process set of sensors, maximum and minimum properties from a data set about items' mass, length, etc. A framework is developed to categorise knowledge as categories of common items to be manufactured, or categories of autonomous robots with comparative speeds or estimated costs. Once a limited number of categories of interest is subscribed, data transfer to a knowledge category is automatic unless the criteria of usefulness is breached.

The applicability of Cloud Computing to many possible scenarios is being researched by considering vehicles with onboard processing capabilities that send data. A Cloud server that receives requests for computing and data storage from vehicles every second is tested under hypothetical situations to assess predictable behaviour under extreme conditions and struggle with a functioning situation forever without a decision. This is an indication of the need for proper assessment of this architecture at different time scales, taking into consideration performance measures, component constraints, loss of reliability, and assessment of timing behaviour for public road situations.

10.3. Cloud Computing in Automotive Systems

The automotive industry is one of the fastest growing industries in the world. The rapid technological advances in electronic and information technologies are posing new challenges to the automotive industry. Now-a-days, a car is viewed as a sensor-rich computer on wheels. The technical requirements for the new generation cars are changing drastically. In the next several years, they are expected to make autonomous driving, roadside communication and car-to-car communication possible. In addition to these anticipated functions, ongoing and future infotainment demands call for advanced and high performance computation. The trend is to design in-vehicle systems as networks of electronic control modules mainly composed of digital signal processors (DSP) and micro-electronic circuits. However, the rapid increase in the data composition and processing demands could exceed the limitations of in-vehicle computing resources, exposing a plethora of data and causing bandwidth bottlenecks in the on-board vehicles.

Cloud computing, which takes advantage of existing internet bandwidth between cloud servers and smart devices, allows seamless deployment of data and computation intensive applications on the cloud infrastructure. In automotive systems, cloud resources can be exploited to reserve safety critical, data and computation intensive applications. Service delays, the key factor distinguishing cloud and in-vehicle applications, could be large and vary with behaviors of the cloud. To take on cloud-based functionalities, service delay can be significant. The variety of operating conditions of cloud resources raises challenges for the design and execution of cloud-based automotive applications. The service delay, which is probably the most critical factor distinguishing cloud and in-vehicle applications, could be large and vary with change of search requirements and behaviors of cloud resources.

Cloud computing is employed in many applications including social networks, video on demand service, and word processing. Recently, the feasibility of cloud-based data and computation intensive automotive applications has gained attention in both industry and academia. In this context, designers no longer consider the in-vehicle computing device and soft real-time clouds as separate entities. There is growing interest in employing cloud computing in automotive applications. Ready access to distributed information and computing resources can enable computation and data intensive vehicular applications for improved safety, drivability, fuel economy, and infotainment .

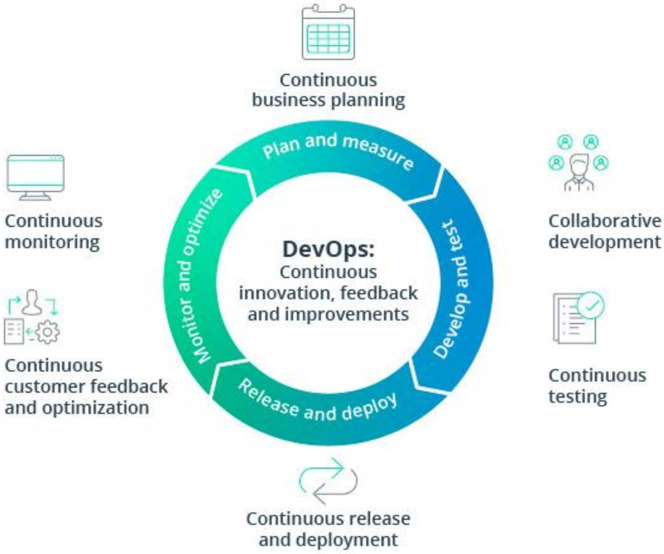

Fig 10.2: Cloud Computing in Automotive Industry

10.3.1. Overview of Cloud Computing

Cloud computing has introduced new possibilities for data processing and storage (Hernandez et al., 2022; Anderson et al., 2023; Carter et al., 2023). The paradigm provides remote services that relieve the end users from handling data processing and storage in their own devices. The demand for cloud services is growing exponentially. Efforts are made to tackle the issues of cloud computing from different perspectives including, but not limited to, service performance, energy consumption, network management, application scheduling, pricing models, and business strategies. The term vehicular cloud was coined to describe exploiting underutilized vehicular processing resources to deliver cloud services to end users. Research efforts are investigating different perspectives. From the application perspective, resource allocation in such distributed architecture becomes a new research challenge. Emerging automotive applications comprise various data acquisition, communication, processing, and storage tasks that require a preparatory study before actually implementing or deploying the applications in a specific cloud infrastructure. However, getting such design knowledge often requires significant efforts and expertise, which an application developer may not have. On a more theoretical standpoint, using the principle from network information theory, researchers have analytically examined the limits on cloud computing performance with or without task offloading, and modeled vehicular computation offloading scenarios with focus on data privacy requirements, mobility of the vehicles, and network architectures, among other parameters. A broader perspective on the principles of cloud computing is outlined. It notes research challenges and opportunities from the perspectives of a data center, resource provisioning, job scheduling, offloading, and security in cloud computing. Modern vehicles equipped with on-board units (OBU) are playing an essential role in the smart city revolution. The vehicles are increasingly becoming sources of valuable data, and hence researchers are interested in performing analytics on the data. Smart vehicles and Clouds, a vehicular cloud computing (VCC) framework, is proposed to enable distributed and scalable traffic management and to ensure safety and security in vehicle sensor networks. Cloud-based traffic signal preemption and public safety were studied with focus on safety and convenience. The approaches to identify a near field intelligent transportation system (ITS) structure that can enhance safety in merged freeway lanes and to solve the corresponding resource-allocation problem while minimizing incurred delays are investigated. To better utilize the vehicle OBUs, VCC is proposed as a means to exploit under-utilized vehicular resources and provision cloud computing services. A mixed integer linear programming (MLP) model is developed to optimize the allocation of the computing demands in the distributed architecture while minimizing power consumption. In addition to traffic management, the vehicular data are also considered useful for academic research. Cloud-based vehicle and driver analytics were reviewed addressing the non-trivial challenges in data collection and processing. Vehicle data such as speed and acceleration from onboard sensors and driver behavior data from mobile devices play a pivotal role in improving the vehicle safety and efficiency and optimizing vehicular applications.

10.4. Data Architecture Fundamentals

Automotive data-intensive software systems generate, disseminate, and analyze massive amounts of data related to vehicles, drivers, infrastructure, and the environment (Singh et al., 2024; Zhang et al., 2025). The automotive industry is currently experiencing a significant transformation due to the evolution of software-centric trends such as smart vehicles, connected vehicles, automated driving, and mobility-as-a-service. Some of the technology enablers for developing these systems include Advanced Driver Assistance Systems, automotive Ethernet networks, and vehicle-to-everything communication. These new automotive trends open up significant opportunities for multiple actors, including original equipment manufacturers, mobility service providers, public institutions, and third-party software companies. Nevertheless, it is essential to address various challenges related to systems' software complexity, safety, dependability, and the new business models' sustainability and privacy. This chapter addresses such challenges related to the design of software data architectures that scale up analytics over vehicle-generated data.

Big data applications, such as those in the automotive domain, are characterized by the volume, velocity, variety, and sensitivity of the data generated and collected in existing and future systems. Conventional data processing architectures may be unable to handle the requested data processing volume or latency. Thus, these applications are implemented using cloud or hybrid architectures to take advantage of better scalability. However, the typical cloud- and edge-based operations of these systems may also raise challenges related to data ownership, privacy, separation of concerns, and interoperability. This chapter addresses these challenges from a data architecture perspective. Evolutionary data architecture technologies are presented, which can exploit cloud computing platforms, with a particular emphasis on the ongoing green-field design of such analytics platforms for large-scale automotive systems monitoring the data evolution of wide-scale vehicle and city traffic and mobility scenarios.

10.4.1. Key Components of Data Architecture

Data architecture is one of the software architecture views describing the data collect, store, manage, and process structures and operations. It is decisions that shape these structures and associated processes. An architecture is a set of principle decisions on the components and their interconnection as a system. A data architecture is an architecture where the components are data and the interconnection is data structures and access mechanisms. One of the fundamental architecture design problems is how to design data architecture that scales as the architecture evolves. Data architecture is the architecture of data and is a critical aspect of software architecture. Unfortunately, scaling data architecture is much harder than scaling software architecture. Most data architecture considerations focus on its design, which is nevertheless crucial to prevent the architecture from becoming a bottleneck early on. Even more important is to get the design right proactively. Data architecture design needs to be in tune with software architecture design. Otherwise, there will be misalignment between the two as the systems evolve. This misalignment makes developing and evolving systems harder. It is, therefore, important to maximize aligning data warehouse, application software, and data architecture design. Key components of data architecture should be data structures responsible for physical data management for structured data, analysts only see a defined schema for data that is stored according to it, and views which provide analysts with a way to see a tailored schema in tune with their analytics capability, Domain, and several related concepts such as measures, dimensions, facts, fields, tables, and keys. It is a schema that describes domain concepts and their associations. Unlike a view, it is produced first and then created for analysts. Data is likely generated and stored according to domain concepts. Further, it is about schema decomposition, using more than one source schema to enforce analysis. Data warehouse is a connected set of data sources managed as a unit. A data warehouse is always a process that creates a data warehouse, as data changes in the systems in a very short time. Definitions will become invalid very soon unless they are executed as views of data in sources or computations on data in sources. A data cloud is a quality of a data warehouse that changes its configuration as sources and definitions change.

10.5. Designing Scalable Data Architectures

The rapid advancements in automotive data processing, transmission, storage, and exploitation capacity increase the potential for new data-driven services in the automotive ecosystem. This increases the number of data functions carried out in the data chain and the complexity of data life cycle management. In this context, it is essential to determine which entity or agent on which data function will perform each processing on the data. Decision criteria pursued include costs, processing delay, transfer bandwidth capacity, and data ownership. Data functions and their chaining are mapped to MEC infrastructures in this work. The cloud layer is defined as a hypothetical aggregation point of MEC function outputs and data generated in the device. The return of the data functions based on cost, standardized price modelling with free charge commercialisation is determined in this work.

The Quality of Service (QoS) requirements of automotive applications are extreme compared to other categories of applications. Accuracy, reliability, service continuity, computation and processing latency, completeness, timeliness, data freshness, consistency, and communication latency are examples of these extreme QoS requirements. These requirements must be ensured by the system from the design perspective to enable the introduction of cloud-based architectures into the automotive domain. Online service reconfiguration mechanisms are also essential for this introduction, especially to guarantee the safety of applications. However, considering the inherent characteristics of automotive applications, the existing approaches to ensure QoS are generally insufficient. Multi-layered architecture for automotive cloud applications is developed with mixed computational resources in general-purpose cloud computing, M2M cloud computing, Fog computing, Edge computing, and Road Side Units (RSUs), including Process-Computing, Virtual Machine, Resource Host, Cloud, Gateway, Edge, and Road Side Unit layers. A safe and high-quality service delegation mechanism is proposed as an intermediate component to this architecture and develops various design approaches for the mechanism.

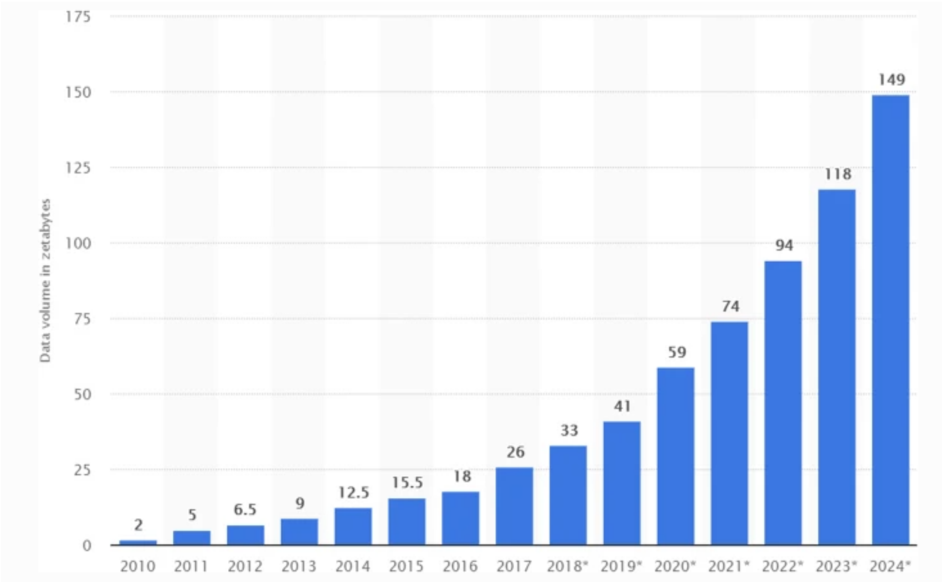

Fig : Data analytics in Cloud computing

10.5.1. Principles of Scalable Architecture Design

Cloud systems provide advanced models of processing and sharing data and information. A recent analysis of cloud computing has identified key characteristics of these systems, namely scalability, pay-per-use utility model, virtualization, and service-oriented architecture. Scalability concerns the ability of a system to increase its computing power to deal with increasing workloads. Two approaches exist: vertical scalability increases the computing capacity of each node, whereas horizontal scalability extends the system adding nodes. The latter provides a more appropriate solution to ensure high scalability in distributed systems. Mechanisms to be considered include replication.

The pay-per-use model supermarket-style pricing is prevalent in cloud computing. It allows service providers to adapt the cost of infrastructure and maintenance to the needs of on-demand large number of on-demand users. It is also a good protection against economic risk for users since they only pay for what they use. One of the key challenges in this model is how to ensure all users a fair treatment with respect to cost and access delay, particularly when the system is highly loaded. This fairness must be ensured through resource allocation schemes. Please note possibilities of tuning research efforts to focus on a certain combination of these challenges. Service-oriented architectures decouple the design of applications and the provisioning of resources. This proposal for a cloud architecture is naive and oversimplified so please provide references on design principles on cloud architectures.

Scalability on service-oriented approaches has been mainly achieved by simplification or elimination of transactions. This, in turn, reduces synchronization needs and, therefore, makes it possible to avoid full ACID guarantees. Hence, design principles for distributed systems that intend to provide full ACID guarantees cannot directly apply to these systems. Examples of systems that simplify transactions are document-oriented databases in cloud services which still provide the basic concepts of transactions on aggregate functions over an uncontrolled-level. More examples of systems that eliminate transactions are whose semantics relies entirely on counters and mutable key-value pairs, respectively. Please note design principles and common techniques for ensuring scalability in data management.

10.6. Conclusion

The automotive market is undergoing a radical transformation the likes of which have never been seen. New connectivity, functional, and business models are emerging aiming to capitalize on the data generated by vehicles. Systems-On-Chip (SoC), with hundreds of heterogeneous cores and different processing units, are used by vehicle manufacturers (OEMs) to run this myriad of functions. However, the vehicle and the driver will become progressively more data-intensive with the combination of increasingly advanced data-generating sensors and the demand for complex, data-hungry algorithms. For these reasons, cloud computing platforms are being proposed. Capable of infinitely scaling computation, storage, and communication resources, they can offer a tremendous performance boost for automotive data functions. They can complement onboard processing and take a huge burden from the vehicle. However, the deployment of automotive data functions in cloud computing platforms poses significant challenges. Network latency poses serious questions about the effectiveness of this function placement. There are also profound architectural questions about the design of the processing tiers. Core data functions, potentially needing high-speed processing resources, will need to be deployed at the network edge or hybrid edge/cloud infrastructure, while data-dry functions could be run at a central cloud.

Different architectural solutions will impact the way these functions are offered and the functional capabilities of the platforms. A hybrid cloud computational infrastructure composed of a central cloud and a large pool of edge facilities building a backend computational ecosystem is proposed. Open-Source cloud platforms are further assessed showcasing the capabilities of pre-existing system software to streamline the engineering process of cloud-computing data functions. Finally, these architectures are evaluated in terms of performance using synthetic problem simulations.

A discussion of the transformative impact of this new processing tier on automotive data functions architecture is presented. The future of multi-access edge computing (MEC) infrastructures for the automotive data marketplace is then envisioned. The main design choices of such systems and their most immediate challenges: global and local data privacy, and low-latency guidelines for functional and architectural design are discussed.

10.6.1. Future Trends

In recent years, there has been increasing research and industry interest in employing cloud computing in automotive applications. In this new collaborating paradigm, computers in the cloud as well as computers mounted on cars exchange computation and data instead of solely using on-board computers. These powerful computation and data resources can enable computation- and data-intensive vehicular applications which would not have been possible otherwise, and therefore it is beneficial for improving safety, drivability, fuel economy, infotainment, etc. There are already numerous cloud-based automotive applications. The advent of cloud computing and the wide adoption of mobile internet, plus the rapid advancement of many sensors and devices, have already brought about social and industrial changes in many domains like education, entertainment, social network, and retail. The automotive industry is poised to be next, with multiple forces affecting the effort in growing cloud-based automotive applications including competition, technology evolution, and regulation changes. This paper gives out a discussion on how cloud computing will continue changing the automotive landscape in the near future. Such a discussion covers the evolving directions of cloud-based applications in the automotive industry, where the definition, understanding, and relation among various terms are clarified.

With an ever increasing number of processing and communication units on-board, and with increasingly tighter resource limitations, vehicular embedded and on-board real-time control systems become more and more complex. Various innovative and complex processing and communication architectures are under development to sustain the ever growing demands of safety and security due to connected and automated driving. In parallel, an increased dependency on third party components widens the attack surface by introducing new security vulnerabilities. And high computational and real-time performance requirements increase the complexity for design, validation, verification and testing procedures. Multiple studies are being performed on different aspects of autonomous driving with the aim of developing novel and more efficient methods combined with spurred knowledge in safety and security. New methodologies and architectures will improve system performance, formal verification will provide more safety, whilst thorough testing with respect to functional safety and cyber-security will increase assurance of on-board systems for the correct and secure use of vehicles.

References

Anderson, C., & Clark, P. (2023). Artificial Intelligence in Automotive Manufacturing: Toward a Green Future. Green Technologies Journal, 25(2), 123-138.

Carter, E., & Miller, F. (2023). AI and Machine Learning in Marketing: Transforming Consumer Engagement and Sustainability. Journal of Marketing Innovation, 18(2), 67-79.

Hernandez, M., & Garcia, J. (2022). Cloud-Based Platforms for Enhancing Connected Mobility Solutions. Transportation Systems and Engineering, 19(3), 152-166.

Singh, V., & Kumar, R. (2024). AI and Machine Learning Integration in the Financial Sector: A Path to Sustainable Growth. Journal of Financial Technology, 58(1), 211-225.

Zhang, X., & Zhao, Q. (2025). Sustainable Solutions in Automotive Manufacturing: The Impact of AI and Cloud Technologies. Automotive Sustainability Journal, 12(1), 89-102.