Whitepaper

Convergence of IoT and real-time analytics in a hyper-connected world

Why IoT data and real-time analytics are inseparable

People, machines, and devices have never been as connected with each other as they are in today's hyperdigital world. As the number of IoT devices grows exponentially, businesses need a smart way to manage and analyze unprecedented amounts of data and derive missioncritical business insights in real-time.

Without the right tools, organizations can miss opportunities to act on time-sensitive data. But what capabilities should one look for in a modern data platform? And what approach should enterprises take to enable real-time analytics from IoT devices?

This white paper aims to answer these questions and delve into the importance of real-time analytics for today's organizations. We will outline how enterprises can process high-throughput geographically distributed events in real-time using data streaming technologies.

We will delve into real-world use cases and explore how real-time analytics can empower organizations with critical insights across their lines of business. We will also learn how modern data platforms can accelerate the entire analytics lifecycle - from data ingestion and integration to insights and action.

Enabling real-time analytics at scale

More than 75 billion IoT devices are expected to be connected to the internet by 2025. All of these devices will gather, share, and transmit data in real-time. Here are the some examples of how organizations can benefit from analyzing and acting upon this data:

Manufacturing: Sensory devices provide realtime data that can be used to analyze machine wear and tear for predictive maintenance in manufacturing.

Information security: Device data is critical for thwarting intrusions in real-time, while live log data can nip computer hacking and data breaches in the bud.

Customer 360: Real-time geospatial data from smartphones and other mobile devices can become the basis for location-based marketing and immediate customer engagement.

FACT CHECK

75.4 Billion

Connected devices projected by 2025

$11.5 Billion

Projected IoT cloud platform market growth by 2025

Common roadblocks to real-time analytics success

1. Incoming data delays

Real-time data coming in from IoT sensors presents several challenges. A large number of nodes push data through low-bandwidth network sequences. As a result, sensors could fail or malfunction, creating issues about whether data should be sent immediately or retained and sent later.

2. Inability to handle multiple streams

IoT data consists of disparate streams of event data that needs to be analyzed in real or near real-time. This data often supports rail networks, patient care systems, and electric grids, where the ability to respond in seconds is critical. Legacy data integration platforms are unable to integrate and process this data fast enough for real-time analytics.

3. Low bandwidth

IoT sensors, devices, and gateways are omnipresent in homes, retail stores, manufacturing floors, and farm fields. Yet moving one terabyte of data over a 10 Mbps broadband network takes 9 days to complete. Enterprises need to plan how to address the projected 40% of IoT data that will be processed at the edge in just a few years' time.

4. Lack of predictive capabilities

Rather than reacting to opportunities, forecasting is at the center of IoT analytics strategies - whether to predict demand/churn, anticipate maintenance, avert fraud, or classify customers. This necessitates enterprises to leverage a modern data platform with predictive analytics capabilities.

Decoding real-time analytics architecture

In the race to acquire customer and operational data, IoT sources play a major role. Real-time analytics helps organizations analyze IoT-generated data from connected devices, enabling them to proactively improve decision making.

The architecture for this type of real-time data processing must deal with data import, processing, storage, and analysis of millions of events per hour. There are 4 key aspects of an IoT real-time analytics architecture - telemetry, computation and analytics, device management, and operationalization.

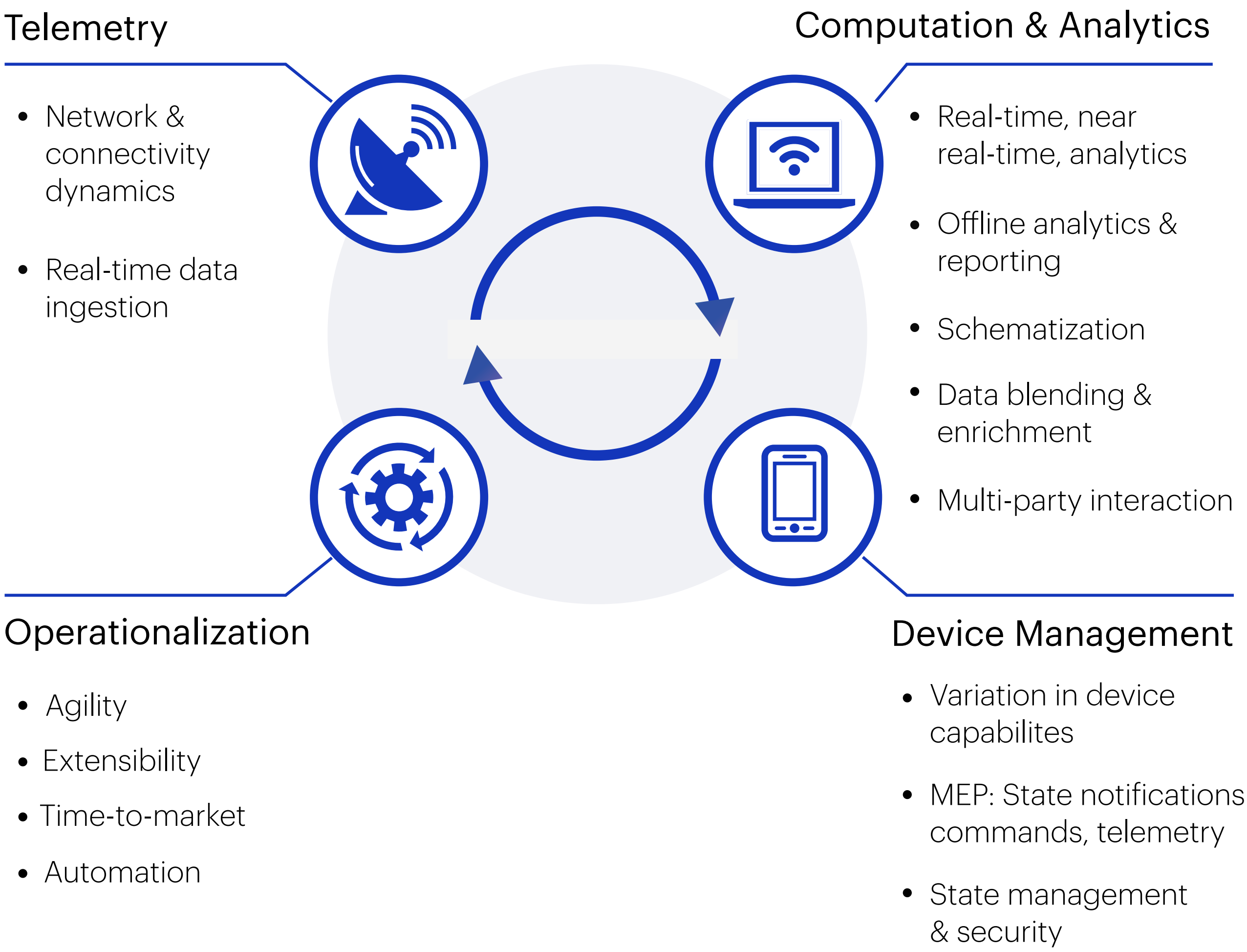


Figure 1: The foundational elements of an IoT application architecture

1. Telemetry

When it comes to real-time data analytics, the foremost requirement is to create a robust framework for collecting data. Telemetry deals with acquiring data from remote sites for analysis and investigation. It defines the data ingestion architecture of the entire blueprint of an IoT architecture.

2. Computation and analytics

To extract maximum value from your data, it's important to analyze both historical and real-time data. It is therefore imperative that the computation engine you choose must support both real-time as well as offline analytics. Also, given the large number of data representation standards and key data characteristics of devices, schematization of metadata is a key challenge. Further, due to lean transmission and bandwidth of data formats, IoT architectures and data flows must support access to various other data sources.

3. Device management

Devices vary a lot in their processing capabilities and an IoT architecture must enable simplified device management. Due to the different transmission protocols of specialized devices, there's a need to support multiple Message Exchange Patterns to generalize service through device communication.

4. Operationalization

To selectively and swiftly operationalize your IoT flows, scalability, agility, and extensibility are absolute imperatives. It is important to choose a modern, enterprise grade data integration platform that has been extensively tested for scale, security and stability.

Real-time analytics, powered by Spark

When it comes to real-time analytics, Spark's single execution engine and unified programming model provides unique benefits over other streaming systems. It also supports interactive queries and streaming along with complex processes like machine learning and graph implementation. Apart from conventional benefits like high processing speed, fault tolerance, and dynamic scaling, a Spark-based data platform also enables fast recovery from failures and better load balancing and resource usage.

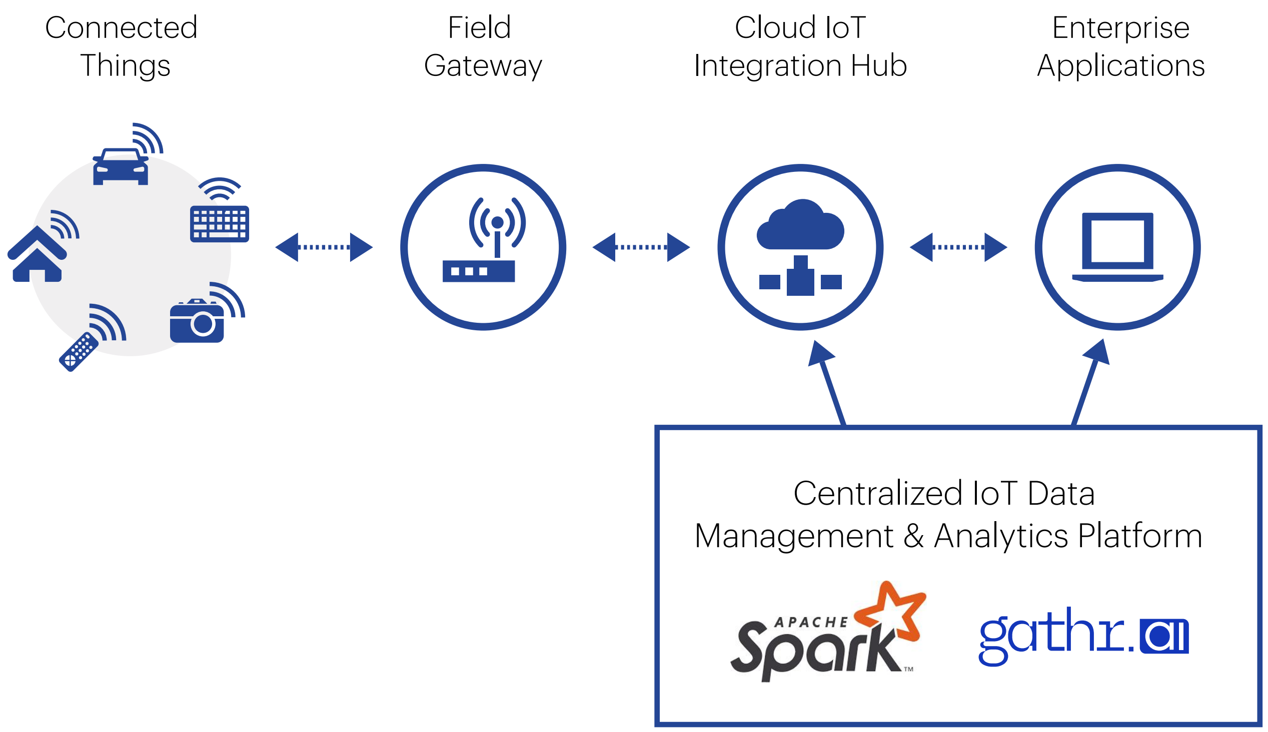


Figure 2: IoT application architecture, with a data integration platform powered by Spark

FACT CHECK

"IoT is regarded as one of the top five game-changing technologies due to its ability to create new sources of data, provide real-time performance updates and create new value propositions."

Gartner

Here are some of the top advantages of using Apache Spark for real-time data processing:

1. Massively scalable

Spark provides robust support for a variety of cluster managers, which makes it highly scalable and reliable. There's also a cloud-ready version backed by multiple vendors. In addition, it has fault-tolerance built into the framework itself.

2. Widely adopted

Spark is witnessing rapid adoption access industries. With a strong and unstoppable momentum behind it, Spark has emerged as the preferred choice for enterprises in recent times.

3. Rich set of transformations

Spark comes with a rich library of out-of-the-box data sources and data sinks. It can handle multiple data formats and statistical operators.

4. Unified programming model

Spark has a simplified unified programming model. With structured streaming, both batch and real-time uses cases are unified into a single model.

5. Supports machine learning

From a machine learning and analytics perspective, Spark provides many open standards that can be implemented. There's also a huge library of machine learning algorithms and transformations available for stock users.

6. Micro-batch capable

Spark supports micro-batch processing at small and regular intervals. Even though processing may happen as often as once every few minutes, data is still processed a batch at a time.

> IoT use cases demand real-time analytics. They demand modern data integration platforms that can ingest streams in any format in real-time and process them atr scale.

Multi-layered architecture for real-time analytics

The figure below illustrates an IoT data processing architecture to power realtime analytics. Notice how the devices layer is only the tip of the iceberg with the underlying data platform doing all the heavy lifting.

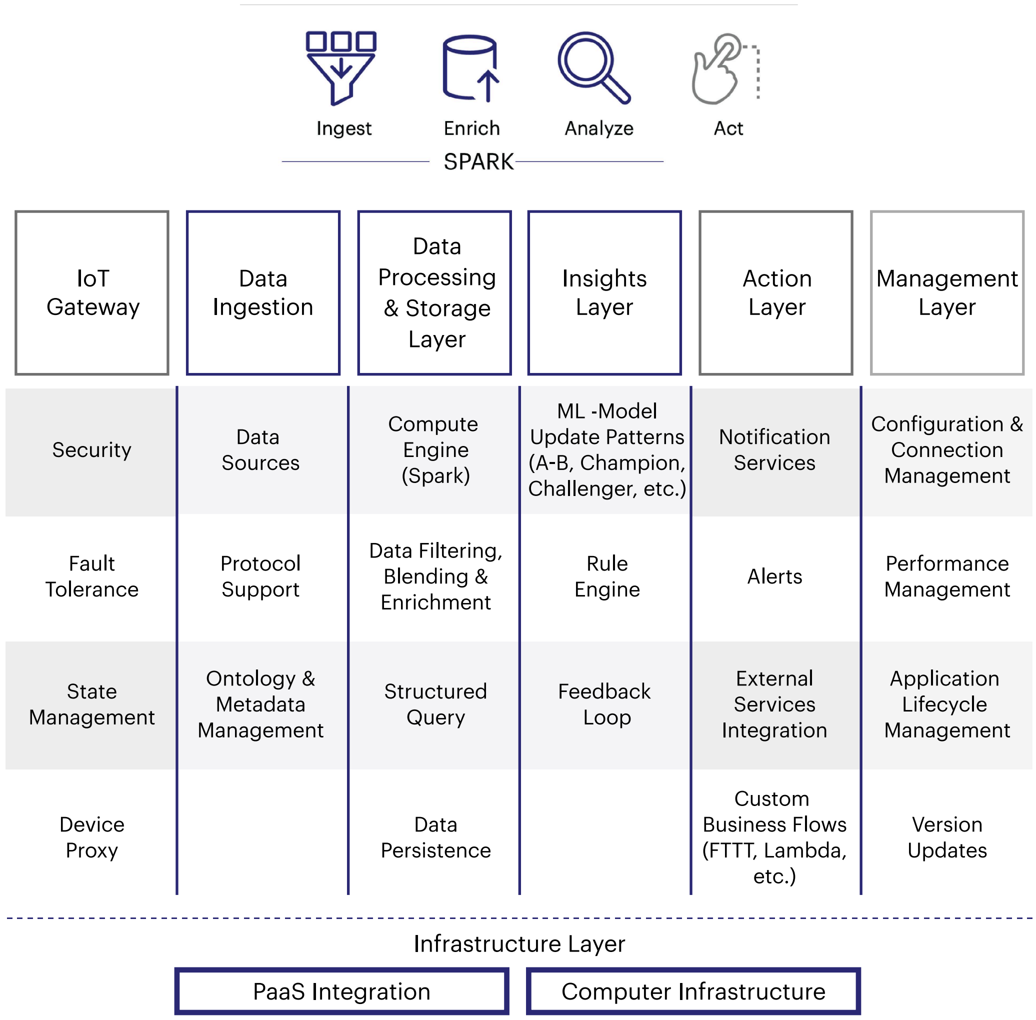


Figure 3: Layered architecture for IoT

1. Data Ingestion

This layer offers simplified APIs and connects to multiple data sources such as Kafka, EventHub, etc. Here, the data is prioritized and categorized, which enables it to flow smoothly in downstream layers.

2. Data Processing and Storage

This layer provides the basic building blocks for commonly used data manipulation and advanced statistical operations. Here, the focus is to store the data and create specialized pipelines for processing it. Users can perform necessary transformations, route the data to a target destination and establish the starting point of the analytics process.

3. Insights

This layer is built over the APIs provided by Spark, and are composed of machine learning patterns such as Champion Challenger. They utilize the services provided by the Spark engine to help build high level abstractions on top of such patterns.

4. Action

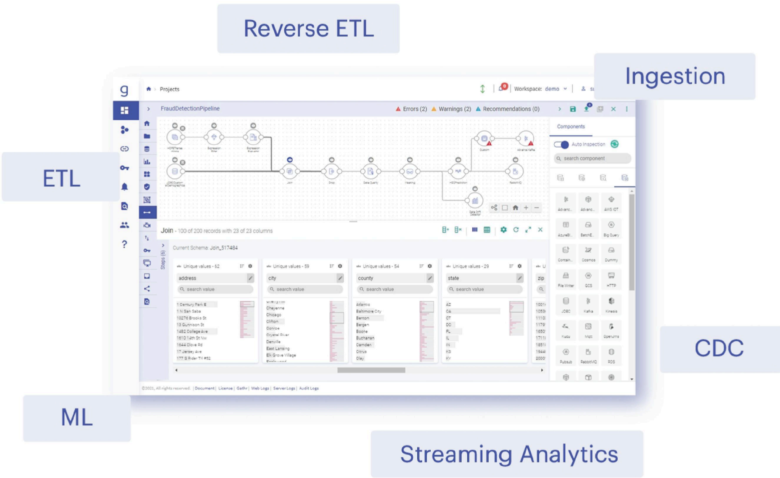
The action layer provides outbound notification services, third party integrations, etc. Just below this layer in the diagram are the infrastructure layers, which include the PaaS integration layer and compute infrastructure. The PaaS integration provides APIs and runtime for action layer services.

4. Management

This layer is the tunnel to all the other layers in the architecture. It enables rapid application development, lifecycle management, etc. for stitching together multiple IoT flows into a bigger, orchestrated IoT flow.

Meet Gathr.ai : A real-time data integration and pipelining platform

Gathr.ai is a unified data platform to handle all your data integration and engineering needs, including ETL, reverse ETL, data ingestion, CDC, data preparation, machine learning, and more. It provides an easy-to-use, drag-anddrop UI to design, develop, manage, and deploy 'data to action' pipelines.



With Gathr, businesses need not invest in multiple tools to manage different aspects of the data lifecycle. It offers all the capabilities needed to derive 'in the moment' actionable insights from enterprise data.

Gathr connects seamlessly to all popular data sources and enables scalability to process millions of events per second. Its unique abstraction architecture provides a common interface over multiple streaming engines like Apache Storm and Spark Streaming, making it a future-proof and robust option for your real-time analytics needs.

Leading companies use Gathr to dramatically cut short their cycle time and productionize streaming analytics use cases at the speed of business.

A real-time IoT analytics use case: Connected cars

he auto insurance industry is rising to meet consumer expectations of personalization and flexibility. To keep up with the new digital consumer and remain competitive, auto insurers are increasingly investing in connected car solutions to offer simplified, transparent, and flexible products and pricing.

Business needs:

- End-to-end analytics application for driver and risk profiling
- Classify drivers based on current driving pattern and historical data
- Raise alerts on behavior change
- Blend data from syndicated and open data marts and services
- Derive additional analytics through supplemental data flows

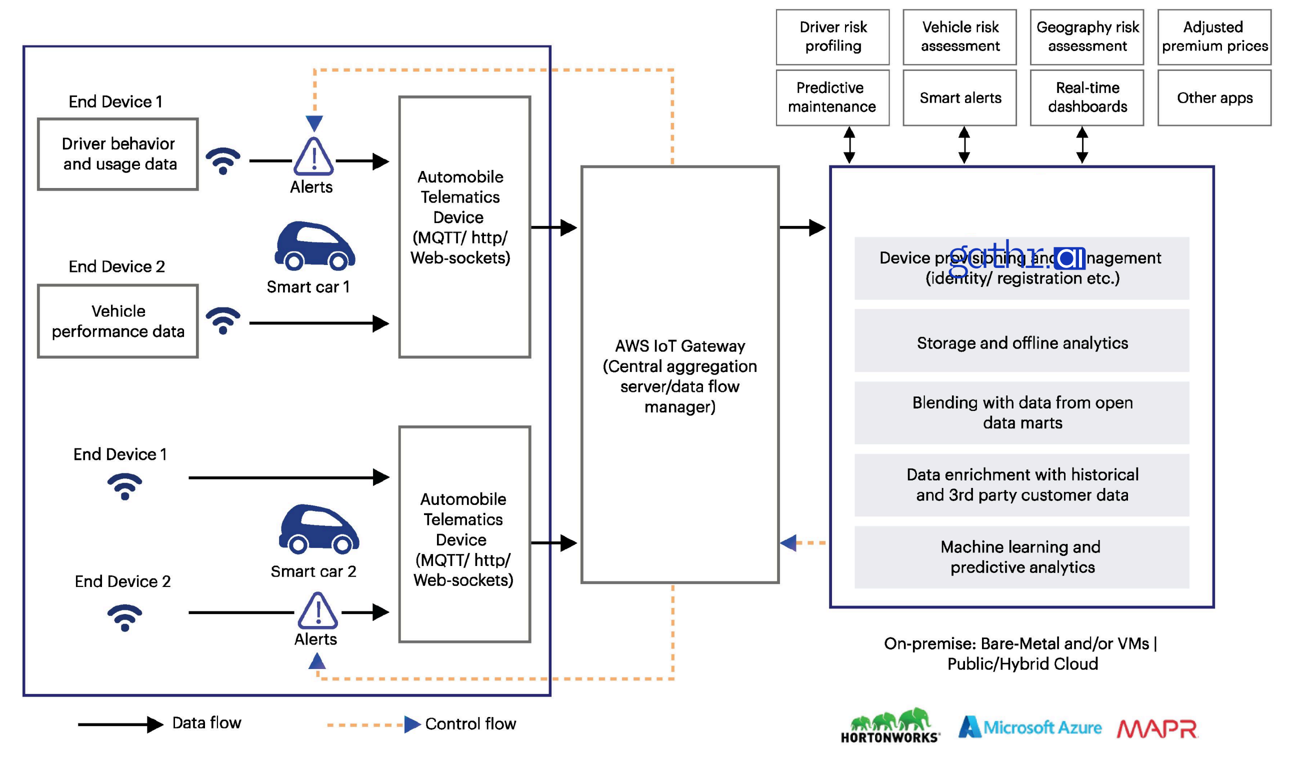


Figure 4: Connected Car Solution with Gathr.ai

Solution highlights

Gathr.ai enabled real-time ingestion of sensor data:

- Driver behavior such as rapid acceleration, hard braking, hard cornering, air bag deployment, etc.
- Vehicle sensor data such as oil temperature, engine performance, and brake wear and tear
- Usage data such as mileage, location, and routes used

In-memory data transformation, data blending, and data enrichment was used to:

- Blend driving behavior data with other real-time data sources such as syndicated public data marts and services like weather data
- Combine real-time behavior and vehicle sensor data with risk history
- Enrich data with customer information such as contact, location, age, past purchases, and past claims

Our application creates alerts to flag risks based on altered behavior patterns: `

- Alerts for vehicle health can be created to flag predictable faults and repair needs, reducing the number of claims caused by potential vehicle breakdowns
- Smart alert models are built to reduce false positives. E.g. frequent braking is not flagged as risky behavior if the driving route explains the need for altered driving
- Customers are optionally alerted in real-time on risks during driving to enable course correction and caution

Results

A real-time analytics application to enable usage-based insurance plans
Through driver profiling and individual risk scores, the auto insurer can now offer highly personalized insurance policies and pricing plans. Additionally, the insurance giant can also offer predictive maintenance services, pre-empting vehicle break-downs and repair needs.

Real-time tracking
It is now easy to track driver activity and vehicle data in real-time through a custom web UI and interactive real-time dashboards. Customers can also easily track their own driving behavior and vehicle performance in real-time through an installed mobile application and take corrective action that can impact their insurance premium prices.
insurance premium prices.

Increased customer loyalty and claims reduction
Remote vehicle diagnostics and predictive maintenance services proved to be a consumer-friendly unexpected value addition and a driver of increased renewals. The insurer's customers provided favourable feedback acknowledging that they looked forward to application predictions on component failures and break-downs.

Results

[1] https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/

[2] https://www.marketsandmarkets.com/Market-Reports/iot-cloud-platform-market-195182.html

[3] https://gartner.com/en/information-technology/insights/internet-of-things

Scan and
start free 14-day trial

Scan to
schedule a demo