

How Rivian builds real-time analytics from electric vehicles

Vidhi Taneja

(she/her)

Principal Product Manager
AWS**Anirban Kundu**

(he/him)

Principal Engineer
Rivian**Rupesh More**

(he/him)

Staff Engineer
Rivian

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS re:Invent 2023 - How Rivian builds real-time analytics from electric vehicles (ANT317)

 AWS Events
99.8K subscribers[Subscribe](#)

1

Share

Clip

Save

...

16 views Dec 2, 2023 #AWSreInvent #AWSreInvent2023

Rivian's vehicle data platform is the source of truth for their vehicle data. The platform serves as a foundational service supporting different domains including digital commerce and insurance, advance driver assistance systems, vehicle reliability, smart diagnostics, charging, and vehicle servicing. Rivian's real-time processing of data stream enables anomaly detection and quick insights into the state of their electric vehicles. In this session, learn how Rivian's vehicle data platform team uses Amazon MSK and other AWS analytics services to build a secure, scalable, observable, cost-effective, and extensible data platform that powers different teams across the organization.

Learn more about AWS re:Invent at <https://go.aws/46iuzGv>.

Subscribe:

More AWS videos: <http://bit.ly/203zS75>

More AWS events videos: <http://bit.ly/316g9t4>

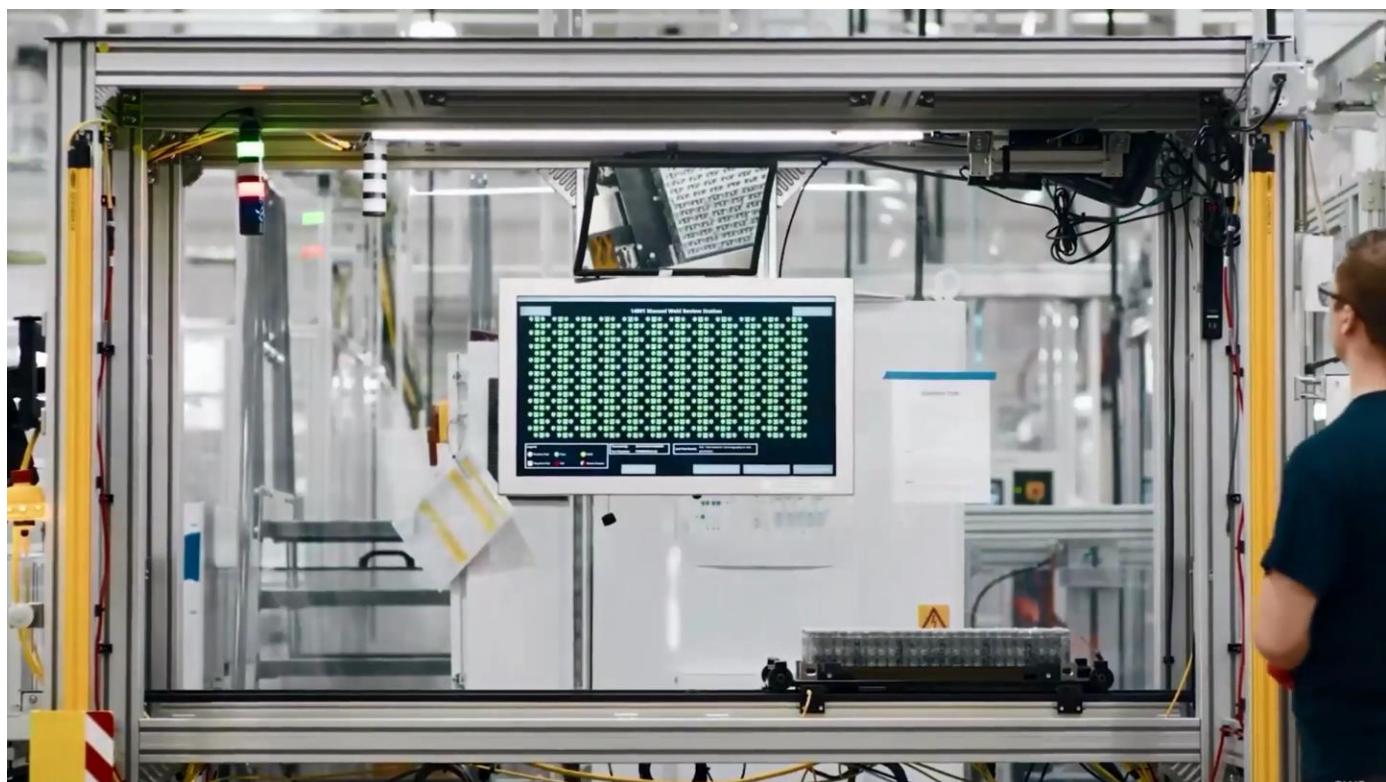
ABOUT AWS

Amazon Web Services (AWS) hosts events, both online and in-person, bringing the cloud computing community together to connect, collaborate, and learn from AWS experts.

AWS is the world's most comprehensive and broadly adopted cloud platform, offering over 200 fully featured services from data centers globally. Millions of customers—including the fastest-growing startups, largest enterprises, and leading government agencies—are using AWS to lower costs, become more agile, and innovate faster.



Machines are learning

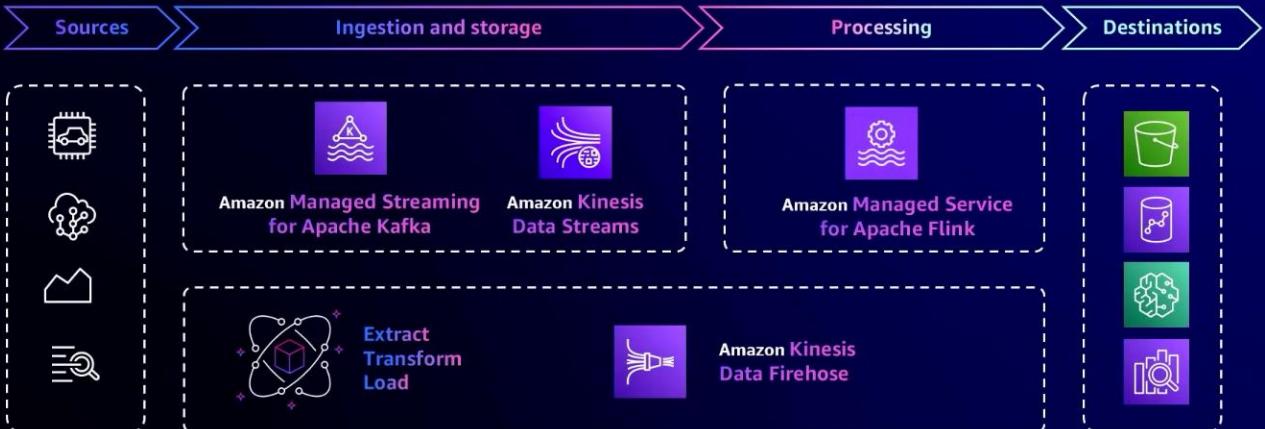




"Real-time data *drives* Rivian"



AWS Streaming Services



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Managed Streaming for Apache Kafka



A fully managed service for
Apache Kafka
and
Kafka Connect

Amazon MSK
Provisioned | Serverless

Amazon MSK



Fully compatible with open-source Kafka, connectors, and tooling

Maximize Apache Kafka performance (high throughput, low latency) at any scale. Monitoring via Amazon CloudWatch or open monitoring with Prometheus



Highly available, scalable, and secure

Multi-AZ replication. Horizontal and vertical scaling. Multiple levels of security. Encryption at rest and in transit. VPC isolation and IAM, Secrets Manager, TLS-based authentication



Deeply integrated with AWS services

Amazon Redshift and OpenSearch Service streaming ingestion, AWS Lambda as a data consumer, schema management with AWS Glue Schema Registry, stream processing with Amazon Managed Service for Apache Flink

What's new with Amazon MSK



Improved cost efficiency



Increased resiliency



Simplified client connectivity



Service integrations

Graviton 3 Instances



M7G INSTANCES



24% lower compute costs



29% more throughput per instance



60% reduction in energy consumption

Tiered storage



BASED ON KIP-405: KAFKA TIERED STORAGE
GENERALLY AVAILABLE ON MSK V3.6



Elastic storage

Seamlessly scale to
virtually unlimited
storage



Cost efficiencies

Save with low-cost tier
for longer retention,
high throughput
workloads



Simplified architectures

Use same app code
for historical and
real-time data



Enhanced performance

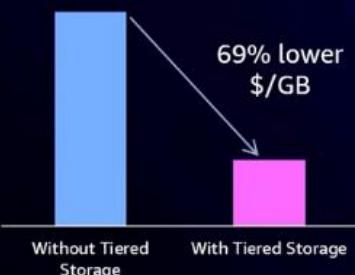
Faster partition
rebalancing and recovery



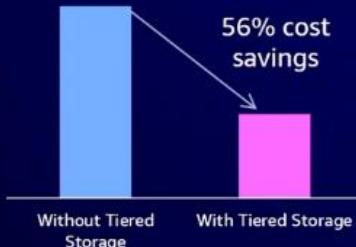
Built in

No additional
infrastructure to
manage

Elasticity and cost savings with tiered storage



Reduced \$/GB by **69%** while
increasing their retention by **3x**



Achieved **56%** savings in their
infrastructure costs

What's new with Amazon MSK



Improved
cost efficiency



Increased
resiliency

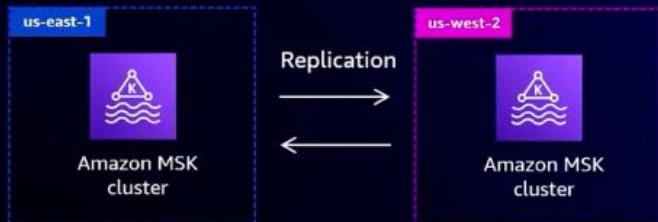


Simplified client
connectivity



Service
integrations

Amazon MSK Replicator



Build robust applications with cross-region resilience

Provide lower latency data access to different geographies

No code, infrastructure, or networking to manage

What's new with Amazon MSK



Improved cost efficiency



Increased resiliency



Simplified client connectivity



Service integrations

Simplified client connectivity



Cross-account access

Establish private connectivity between your Amazon MSK clusters and clients in different VPCs and AWS accounts

Multi-VPC private connectivity

Keep all traffic with the AWS network

Pay-as-you-go pricing

AWS IAM support for all languages

Use AWS IAM to manage authentication and authorization of Kafka clients in all programming languages

Helper libraries for Java, Python, Go, JavaScript, and .NET

Supported for MSK Provisioned and MSK Serverless

What's new with Amazon MSK



Improved cost efficiency



Increased resiliency



Simplified client connectivity



Service integrations

Deep integrations with AWS Services



Kinesis Data Firehose



Amazon Athena



Amazon OpenSearch



Amazon Redshift

Data delivery to Amazon S3

No code

Fully managed

Built-in transformations

New connector

Query real-time streaming data with SQL queries

Works with Amazon MSK Connect

Streaming ingestion

Direct ingestion for real-time use cases

No third-party connectors required

What's new with Amazon MSK



Improved cost efficiency



Increased resiliency



Simplified client connectivity



Service integrations

M7g Graviton 3 support

Tiered Storage (with v3.6)

Amazon MSK Replicator

Cross-account access

Multi-VPC private connectivity

IAM support for all languages

Simplified S3 delivery

Athena connector

OpenSearch, Redshift streaming ingestion



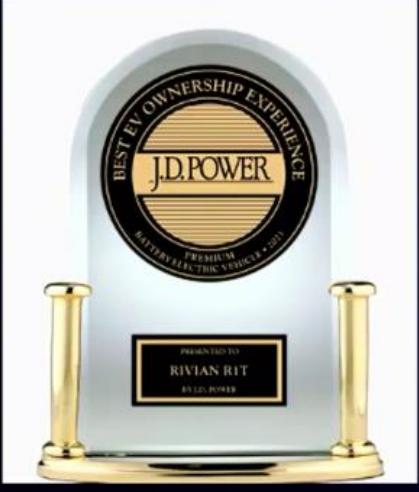
aws





 RIVIAN

Awards and recognition



**IHS
HLDI** VEHICLE RATINGS NEWS TOPICS ▾ Search website

Home / News / 2022 / **2022-23 Rivian R1T earns TOP SAFETY PICK+ award**

December 16, 2022

MOTORTREND News Reviews Buyer's Guide Watch MotorTrend+ The Future FREE TV ▾ SEARCH

Ad closed by Google

The Rivian R1T Is the 2022 MotorTrend Truck of the Year

A remarkably original take on the pickup truck redefines the genre.

Scott Evans - Writer; Renz Dimaandal - Photographer | Dec 13, 2021

What defines us



Software-defined vehicle



Vehicle data and analytics



Power of cloud

Software-defined vehicle



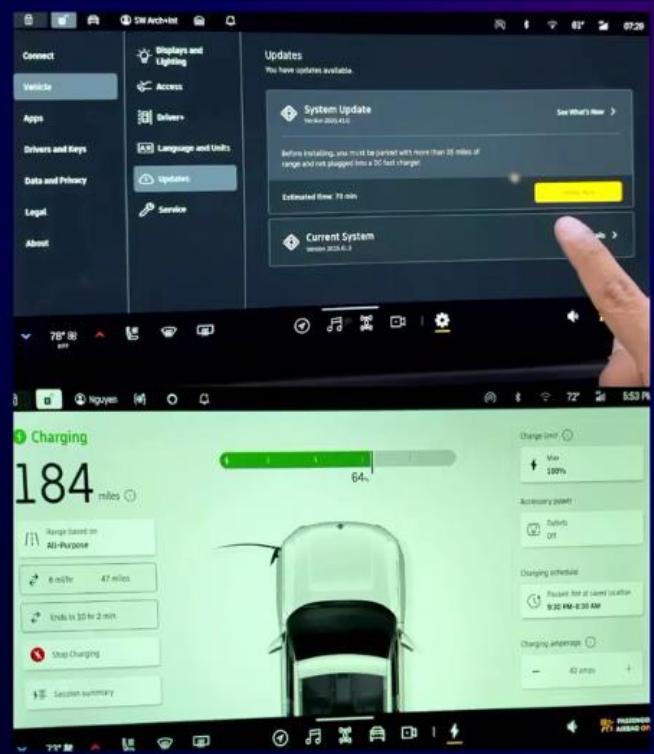
Shared mechanical and electrical platform



Over-the-air updates

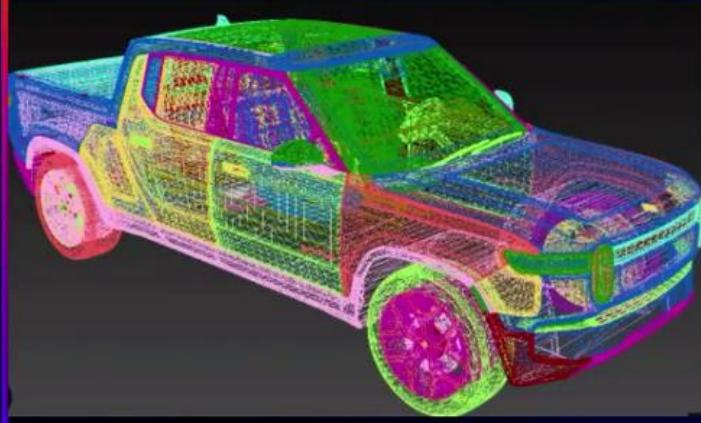


Algorithms optimizing vehicles



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Source of data



Onboard electronic control units (ECUs)



Vehicle data and analytics



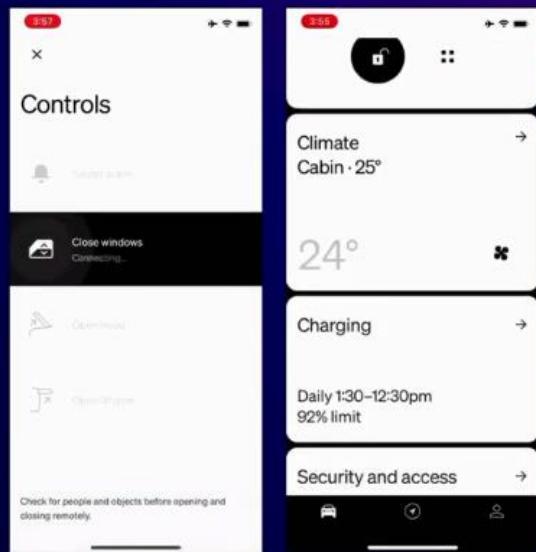
Remote commands and configuration



Vehicle data and analytics



Streaming and batch analytics



Vehicle data and analytics



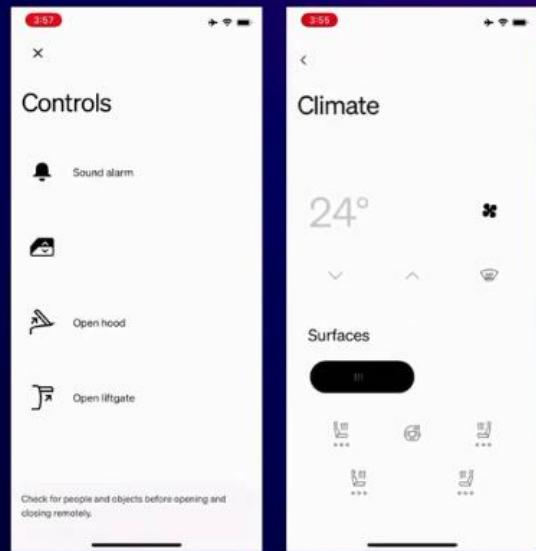
Remote commands and configuration



Vehicle data and analytics



Streaming and batch analytics



Vehicle data and analytics



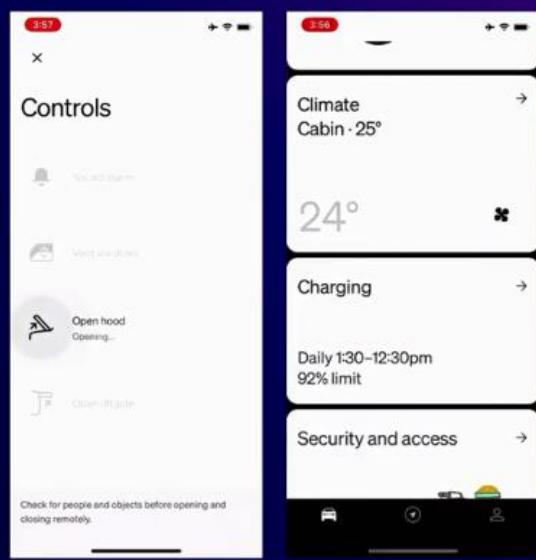
Remote commands and configuration



Vehicle data and analytics



Streaming and batch analytics



Vehicle data and analytics



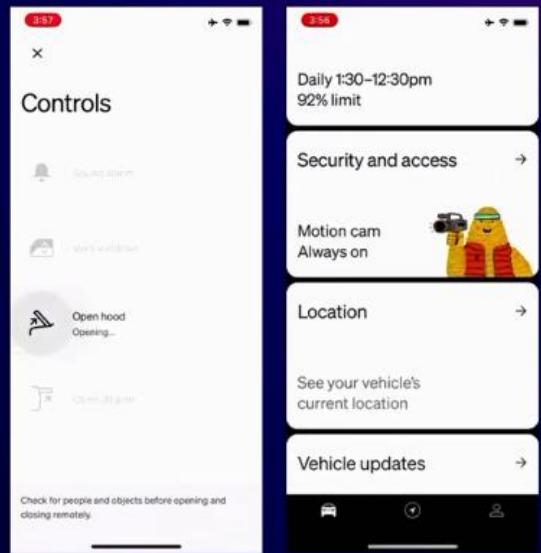
Remote commands and configuration



Vehicle data and analytics



Streaming and batch analytics



Vehicle data and analytics



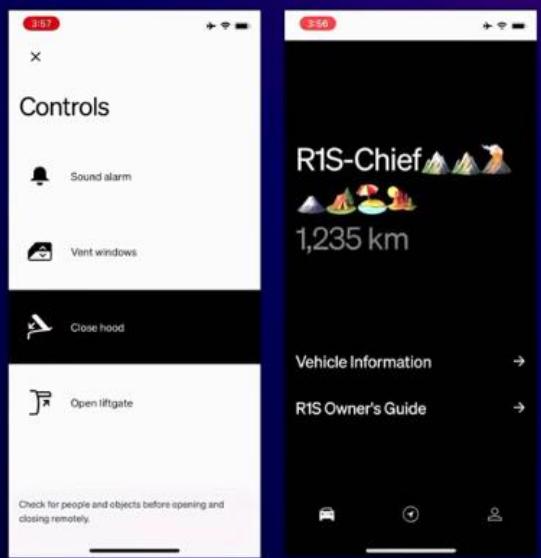
Remote commands and configuration



Vehicle data and analytics



Streaming and batch analytics



Vehicle data and analytics



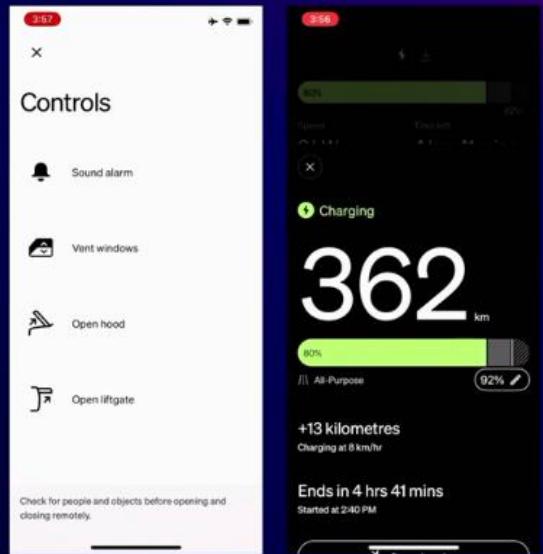
Remote commands and configuration



Vehicle data and analytics



Streaming and batch analytics



Vehicle data and analytics



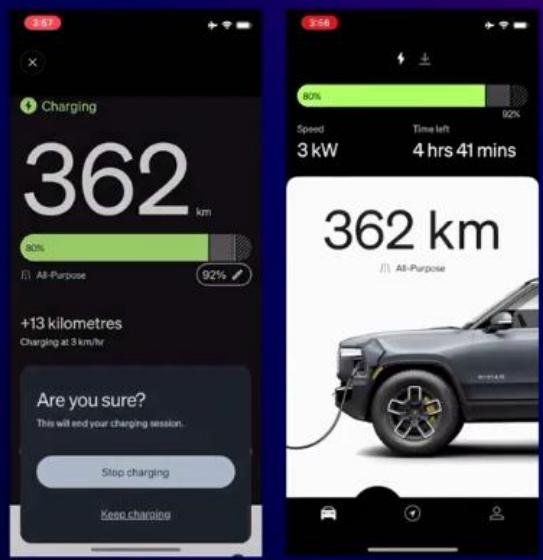
Remote commands and configuration



Vehicle data and analytics



Streaming and batch analytics



Vehicle data and analytics



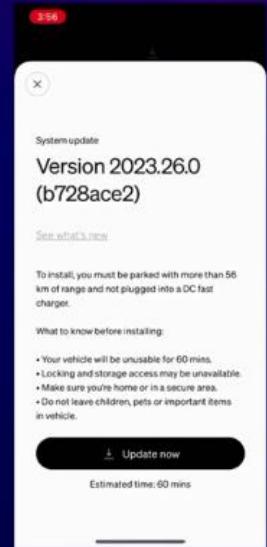
Remote commands and configuration



Vehicle data and analytics



Streaming and batch analytics



Vehicle data and analytics



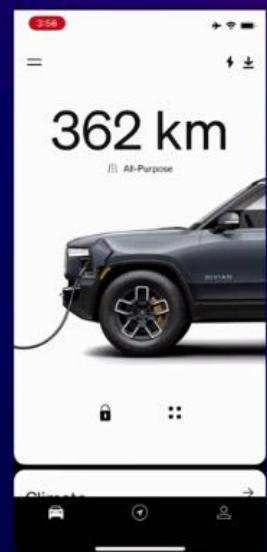
Remote commands and configuration



Vehicle data and analytics



Streaming and batch analytics



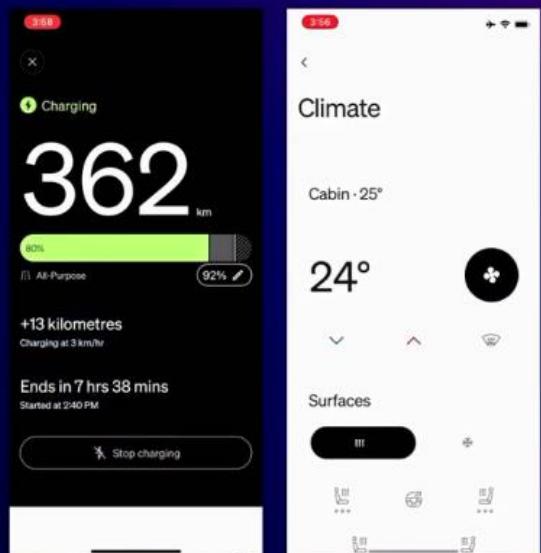
Vehicle data and analytics



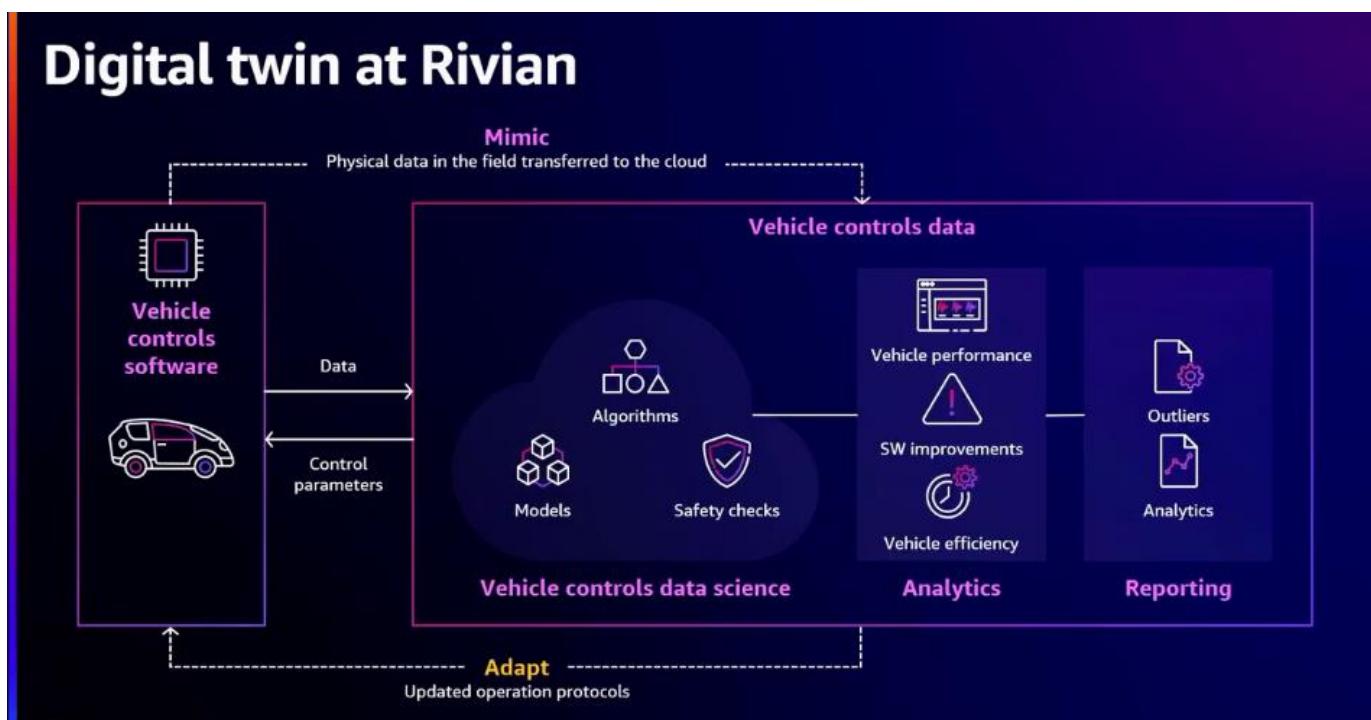
Remote commands and configuration

Vehicle data and analytics

Streaming and batch analytics



Digital twin at Rivian



Power of cloud



Scalable processing of vehicle data



Security and privacy in cloud



Standardized formats and interfaces

Ingest



Compute



Store



Connect



Connected vehicle model



Data Platform Tenets



Performance, Scalability,
and Cost



Security and Privacy



Data Democratization



Data Platform Tenets



Performance, Scalability,
and Cost



Security and Privacy



Data Democratization



Data Platform Tenets



Performance, Scalability,
and Cost



Security and Privacy



Data Democratization



Technology discovery and evolution

Choice of stream processing engine
and compute



Data backbone that enables stream
processing

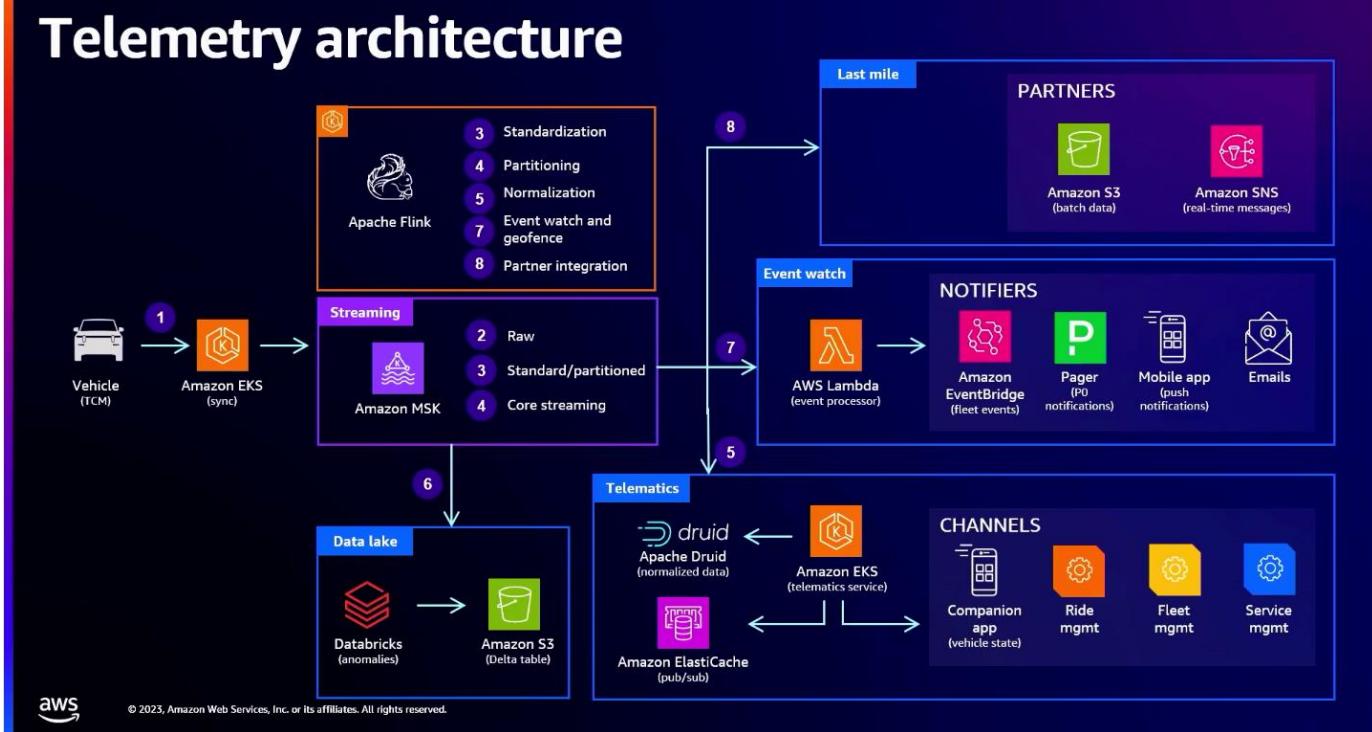


Well-supported tools & services

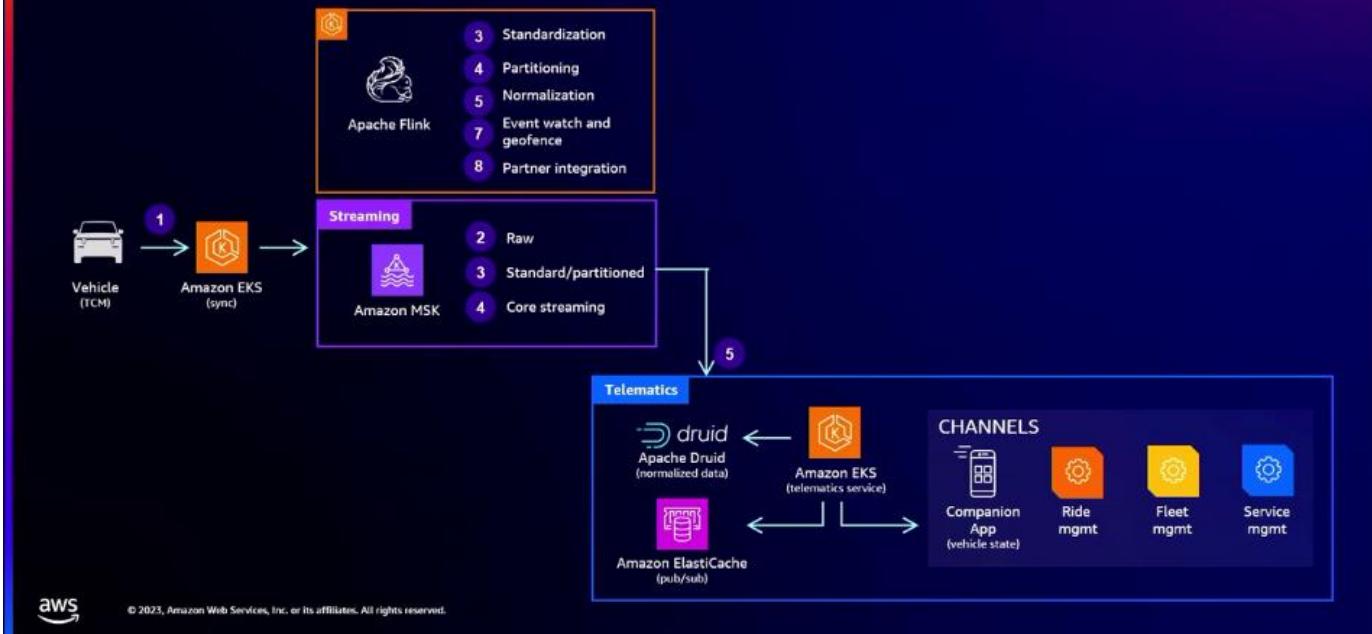


Manageability

Telemetry architecture



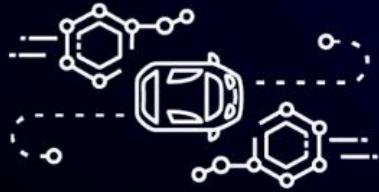
Telemetry architecture



Telematics

ACCESS TO REAL-TIME VEHICLE DATA

Enables access to sensor signal data

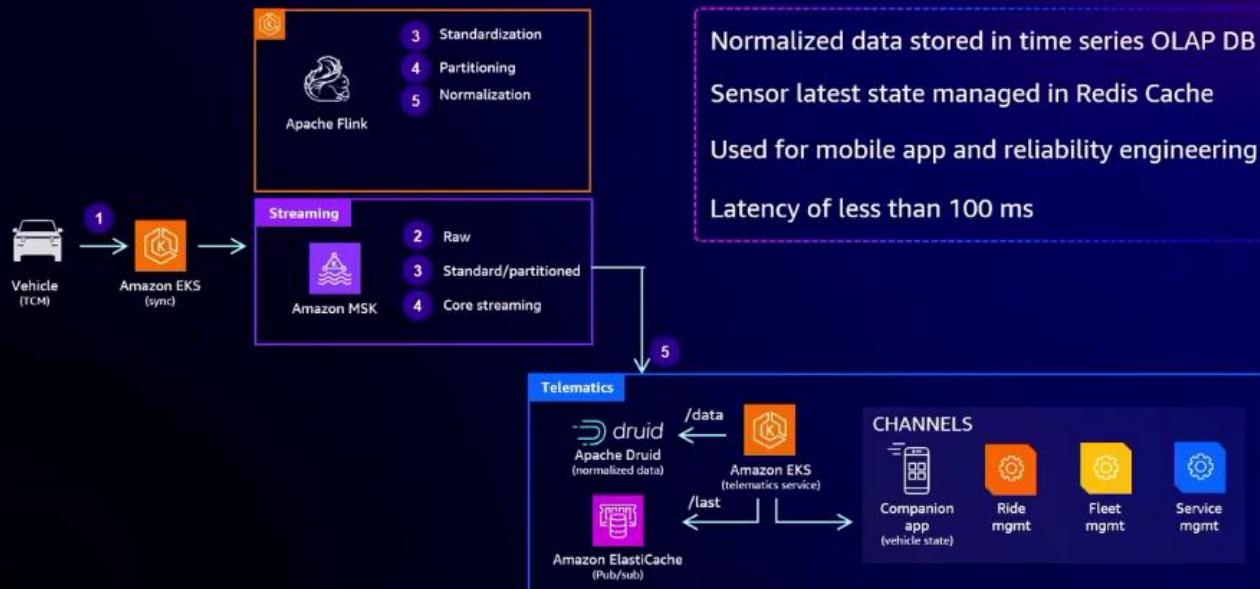


Latest and date range state

Vehicle heartbeat

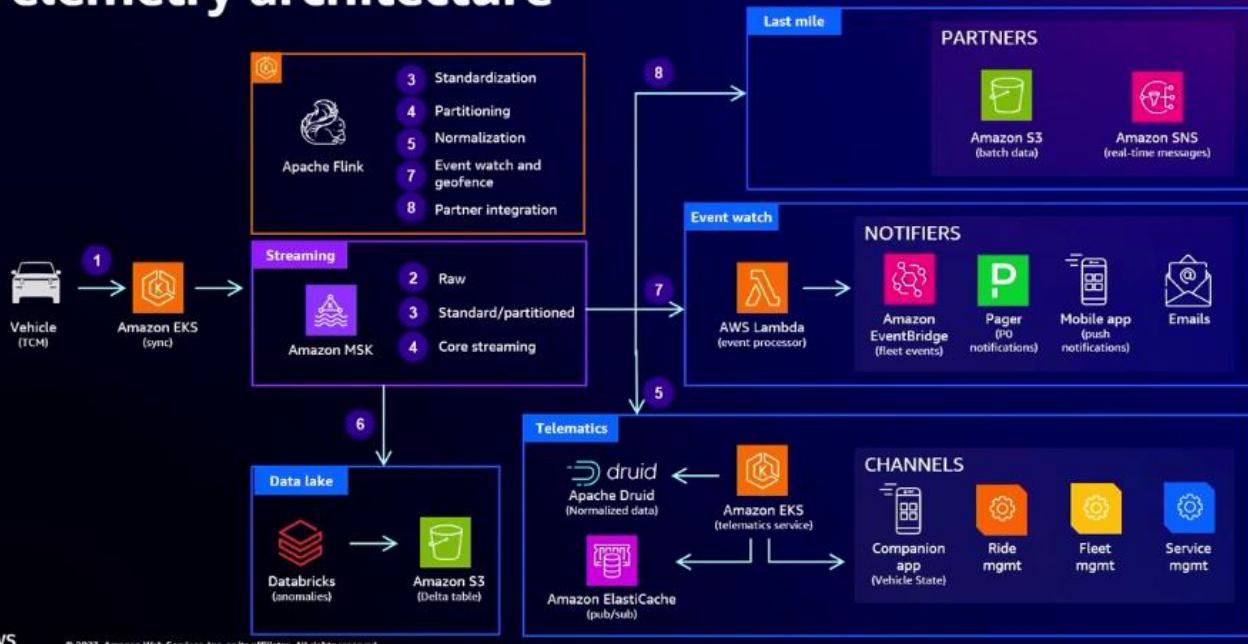
Powers dashboard and internal apps

Telematics



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Telemetry architecture



Telemetry architecture



Data Lake

ACCESS TO BATCH DATA

Deep queries spanning many months



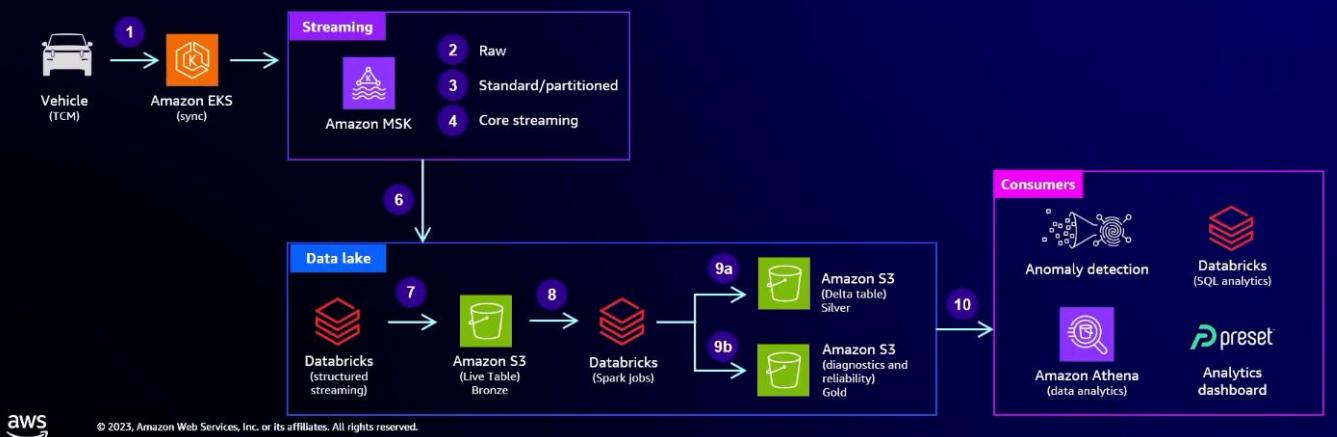
Aggregations, analytics and machine learning needs

Varied datasets accessible

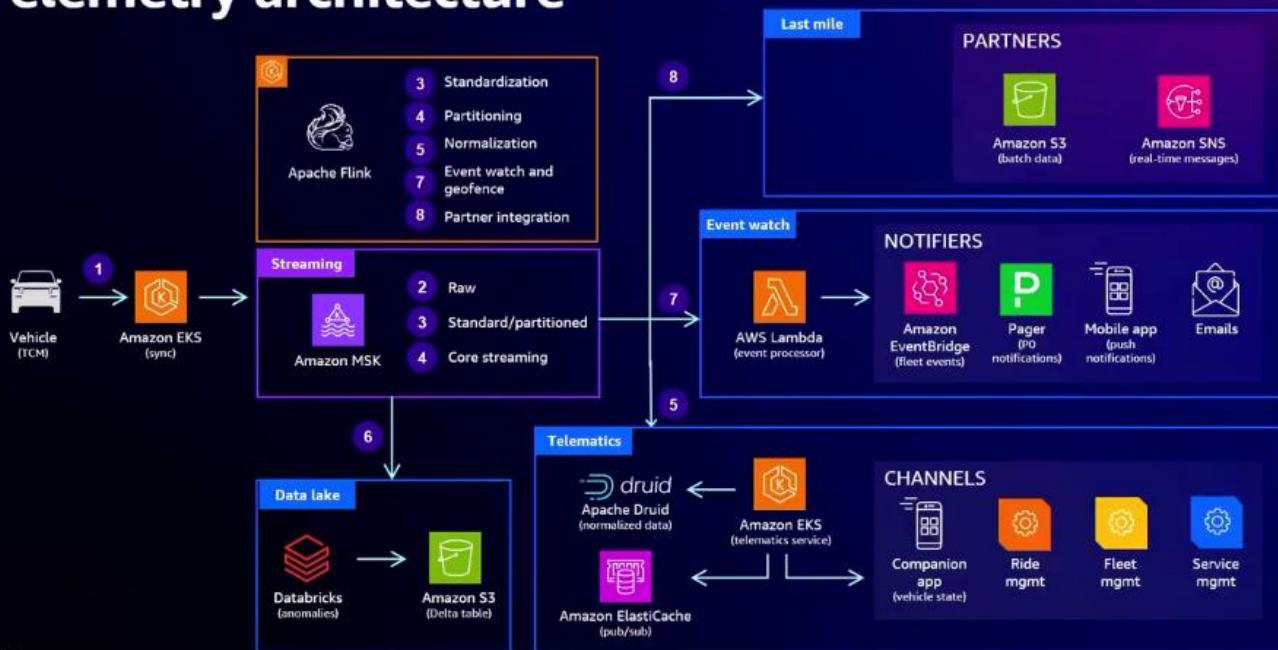
Enforces authorization, privacy objectives

Data Lake

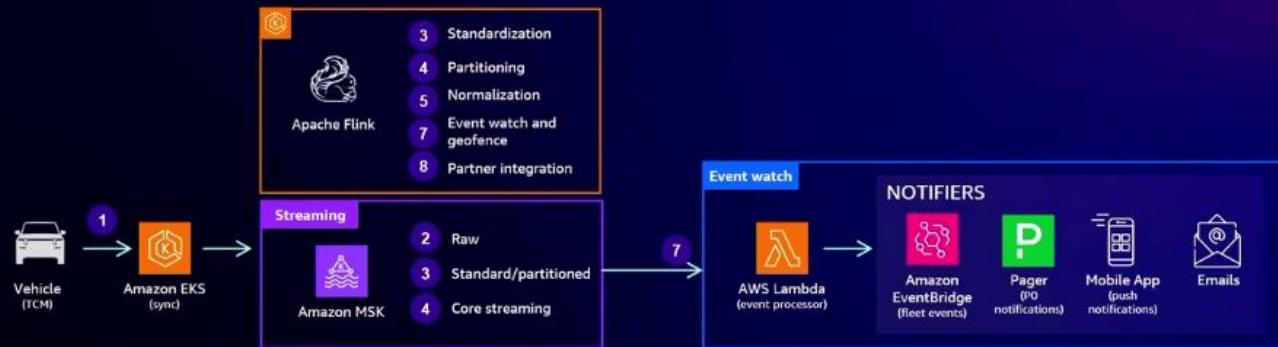
ACCESS TO BATCH DATA



Telemetry architecture

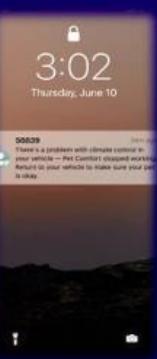
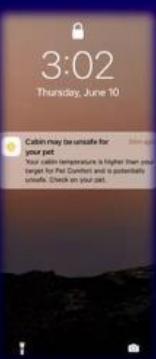
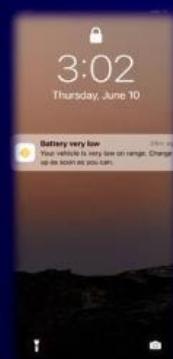


Telemetry architecture



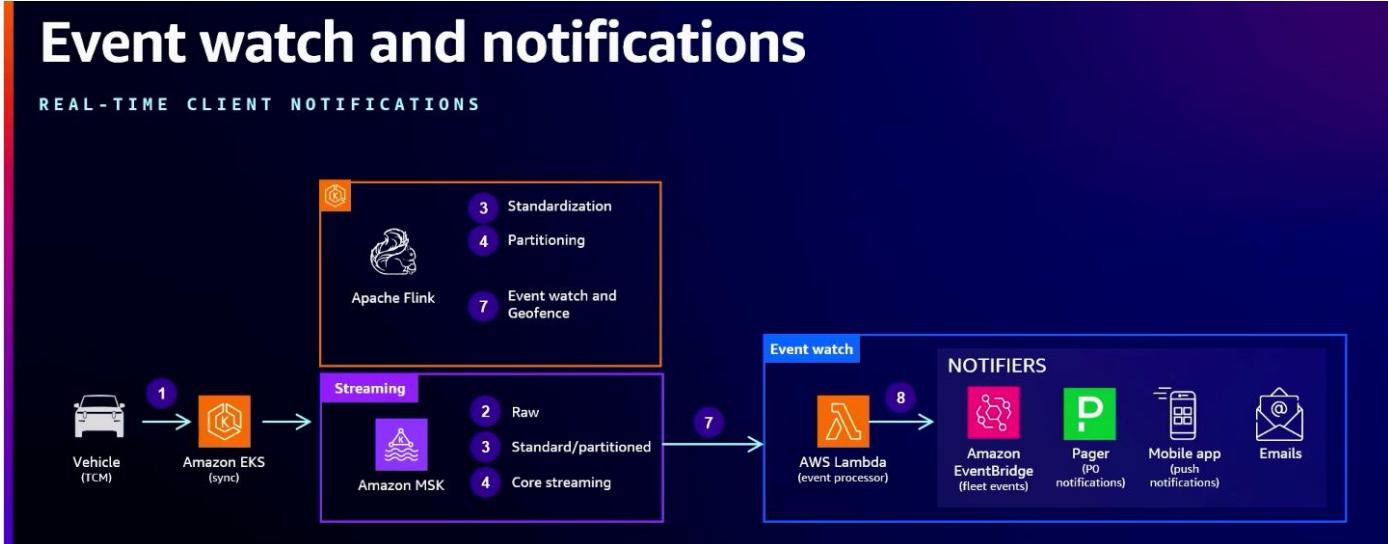
Event watch and notifications

REAL-TIME CLIENT NOTIFICATIONS

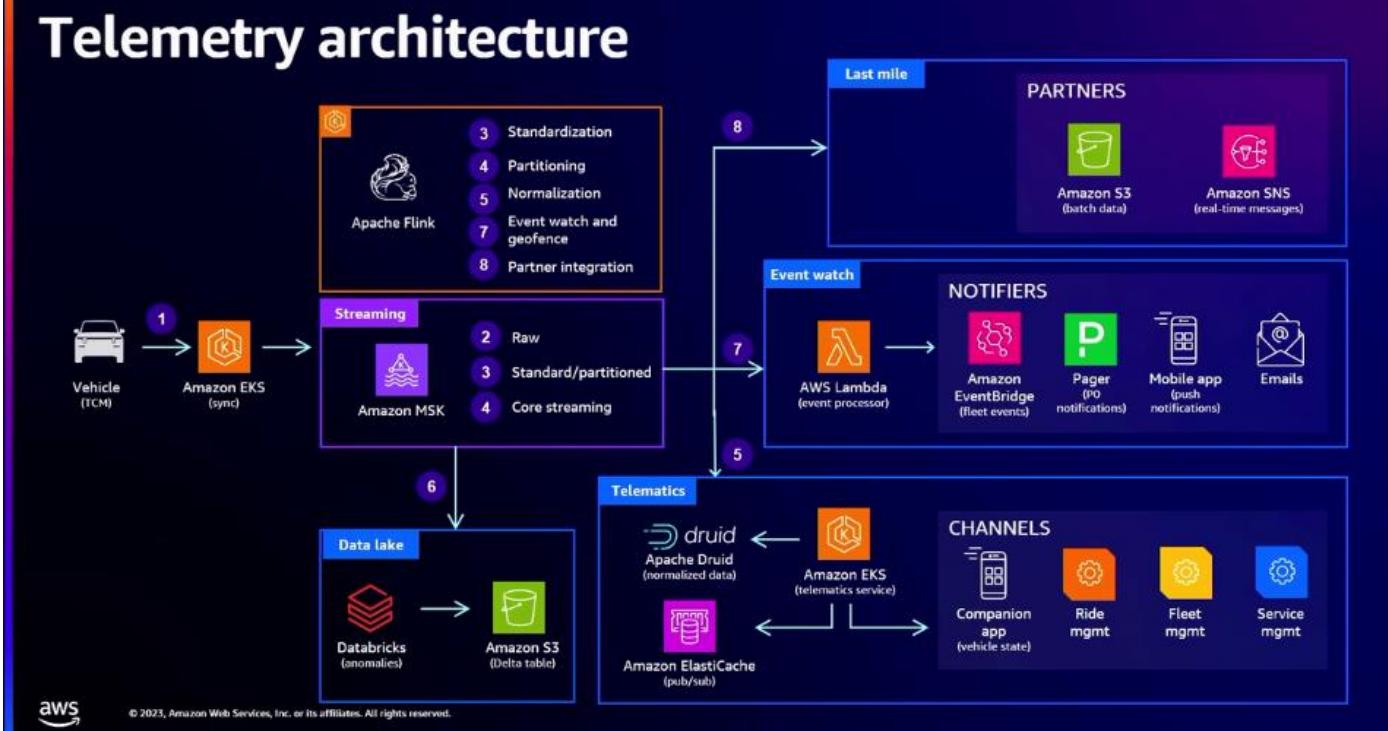


Event watch and notifications

REAL-TIME CLIENT NOTIFICATIONS



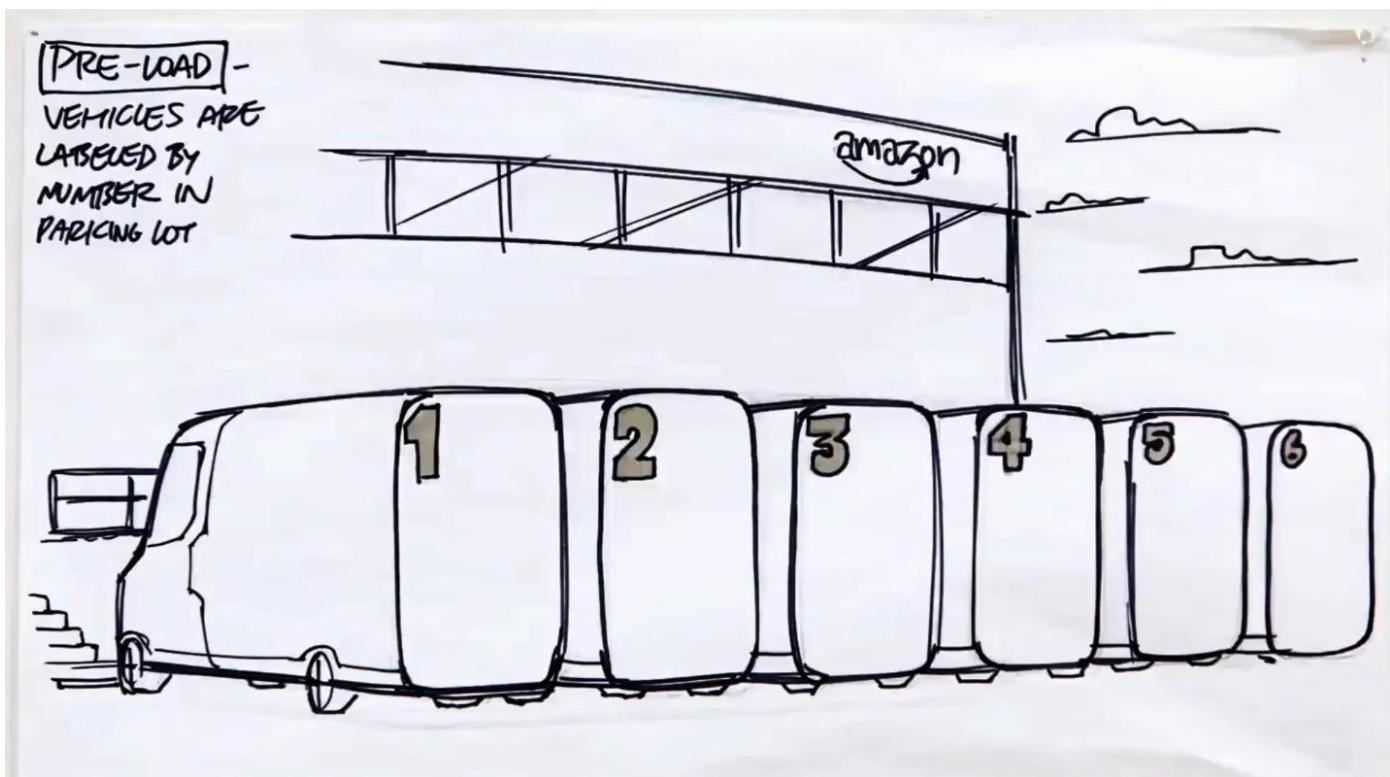
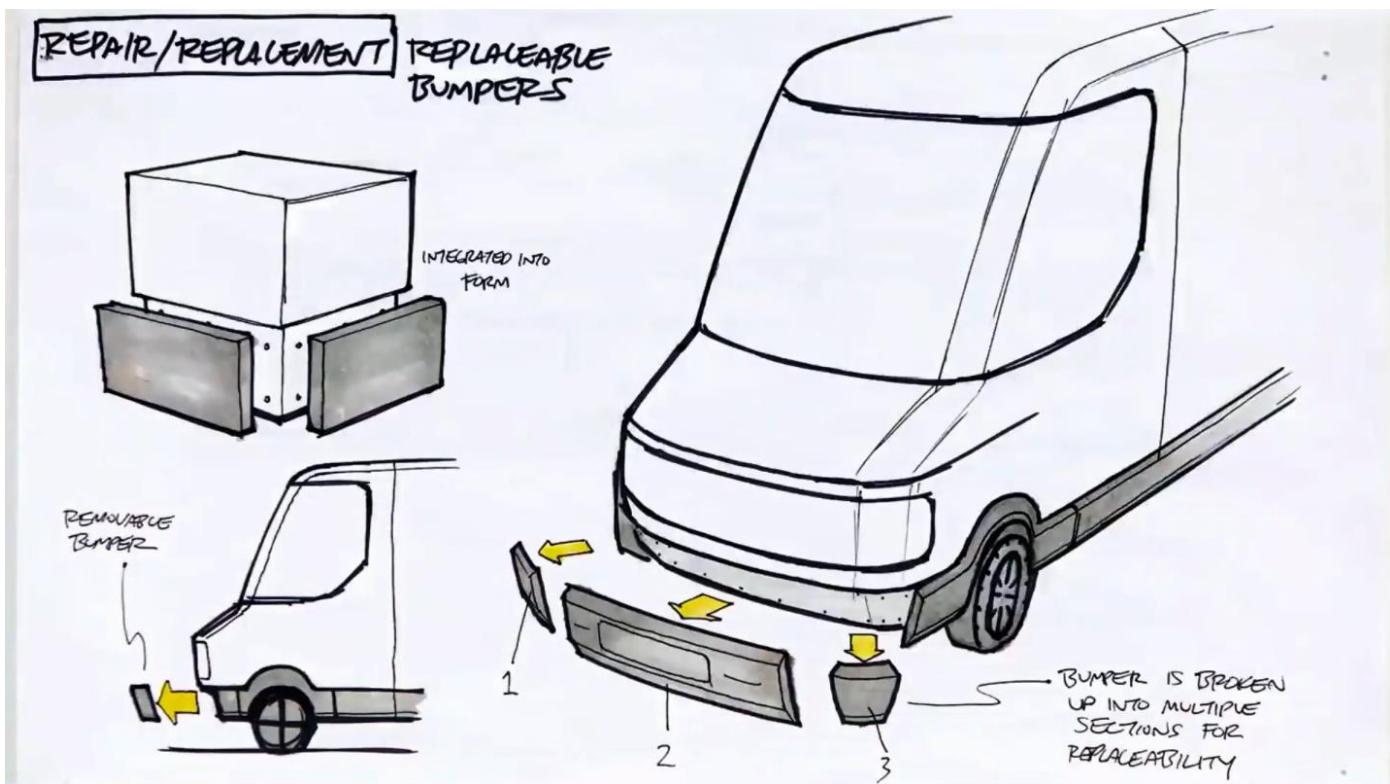
Telemetry architecture



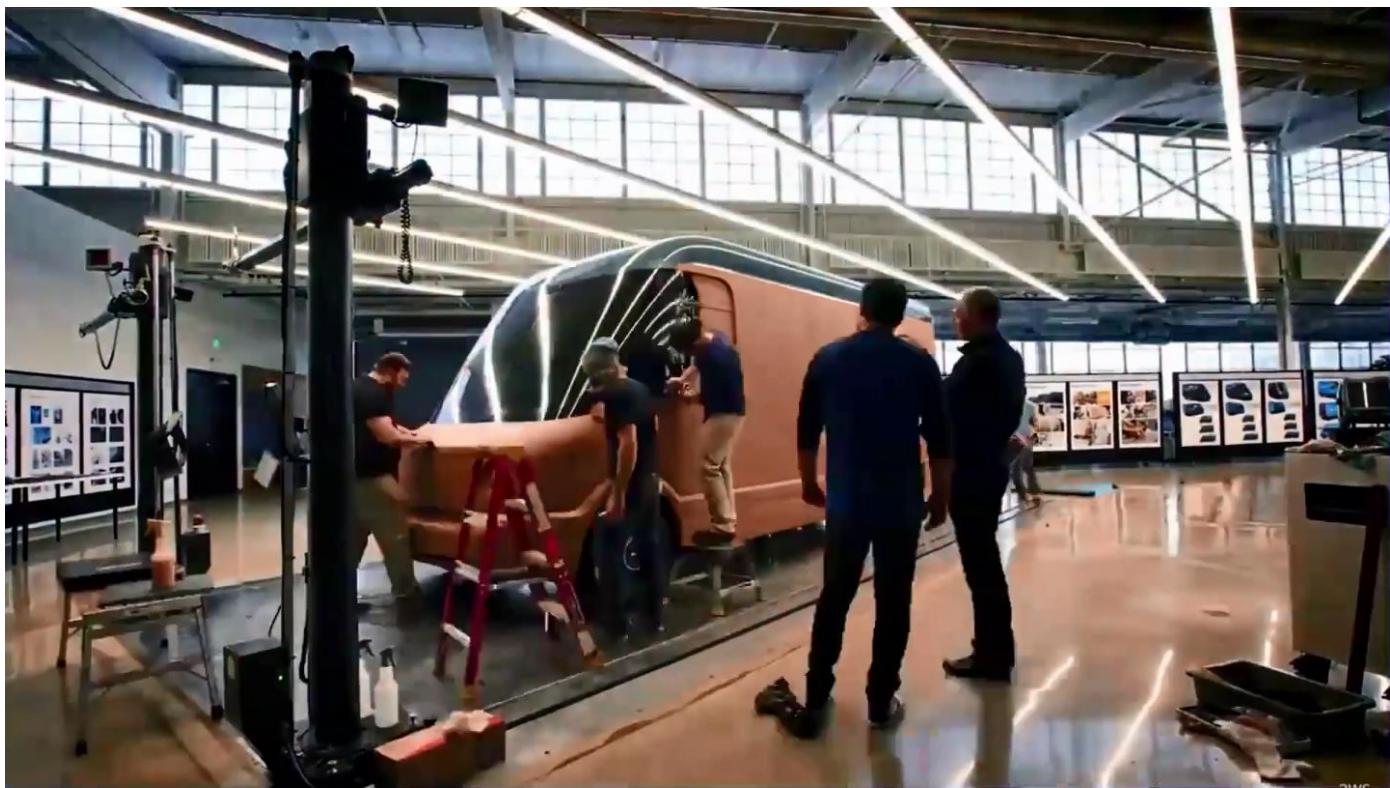
Telemetry architecture

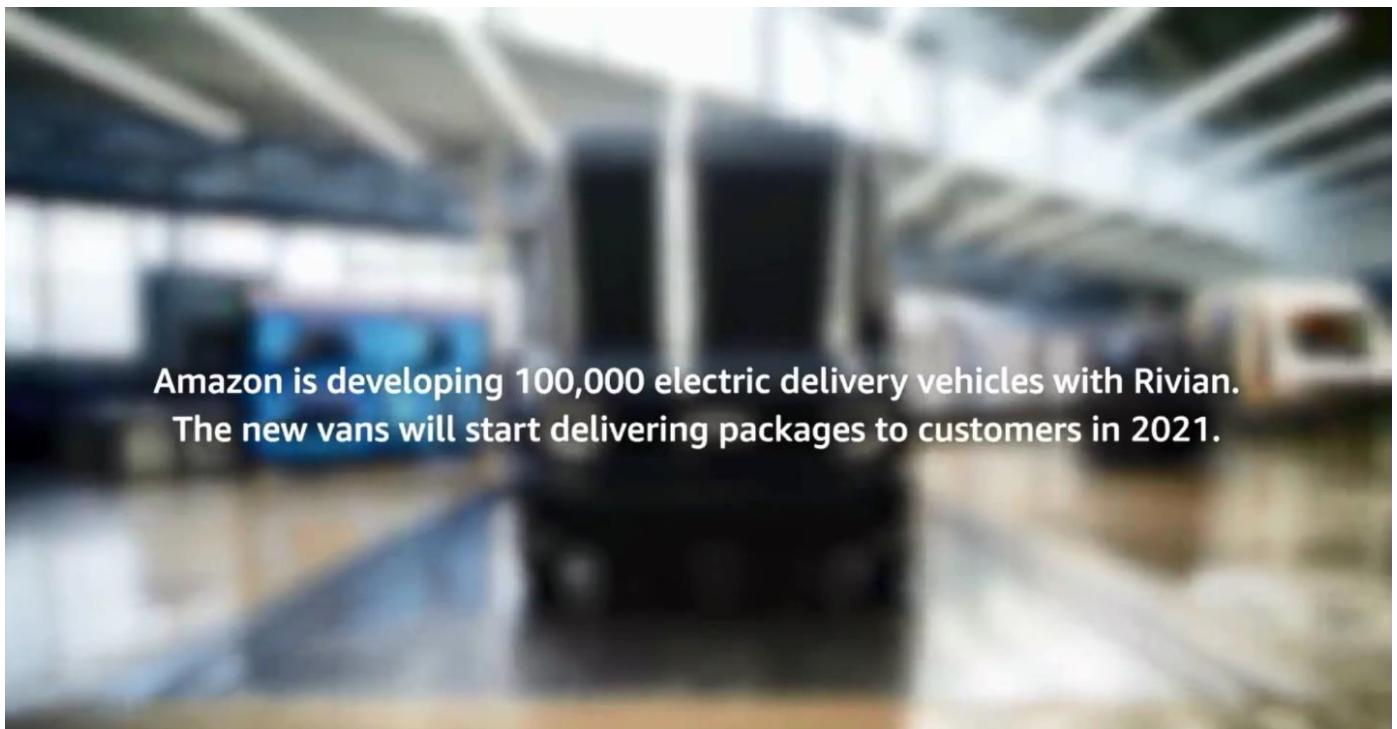
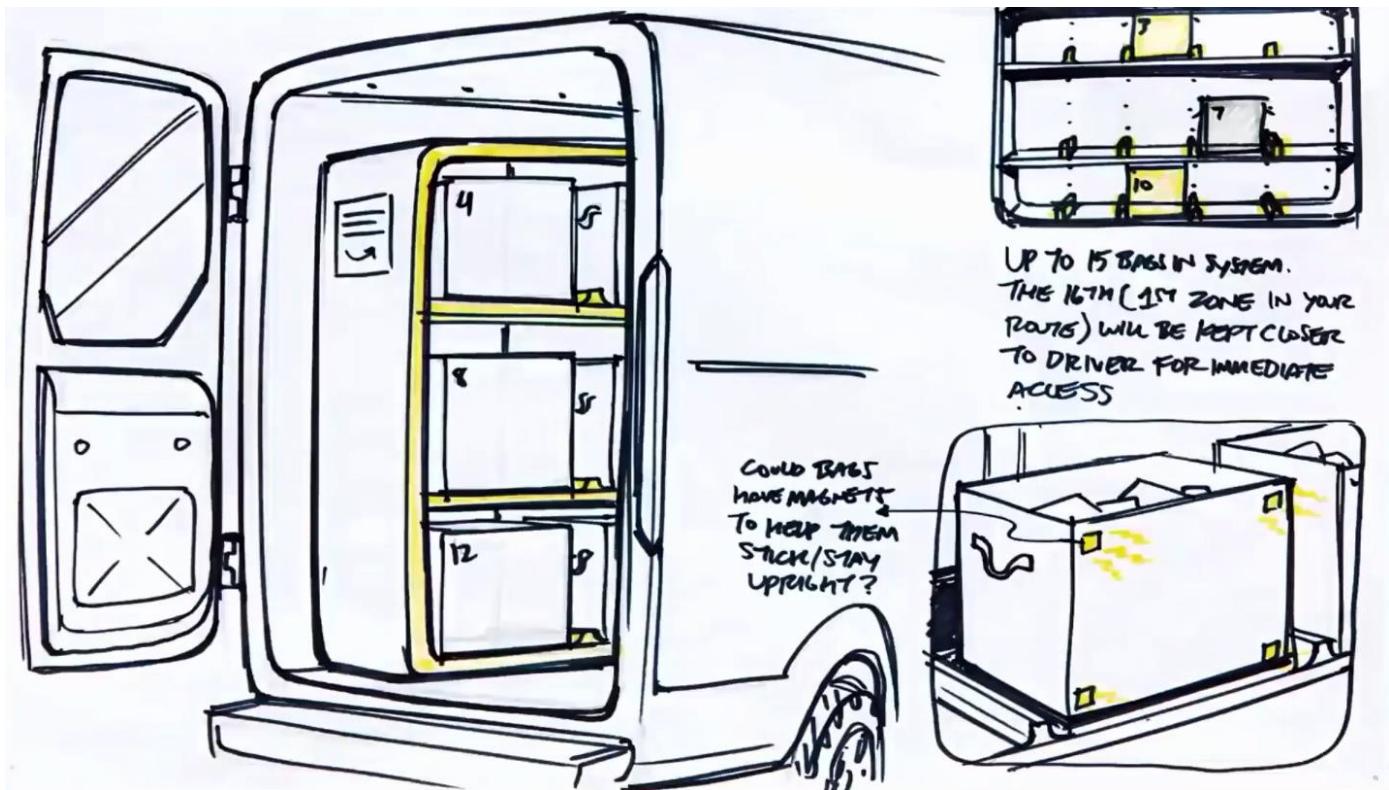












Amazon last mile

TELEMETRY SYNC FROM AMAZON EDV FLEET

Efficiency

Route planning

Safety

Collision detection, airbag deployed

Cabin/cargo temperature too hot

Driver seatbelt buckled/unbuckled status

10K

Electric delivery vans today

100K

Electric delivery vans by 2030

62.5M

Total miles driven by Oct 2023

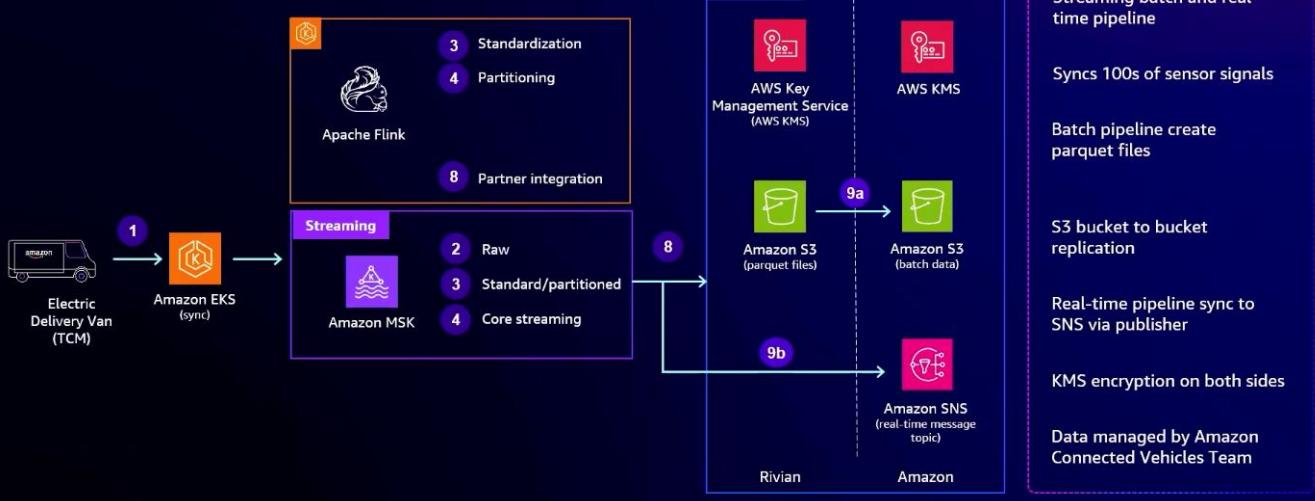
NA & EU

Electric delivery van operating regions



Amazon last mile

TELEMETRY SYNC FROM AMAZON EDV FLEET



Amazon MSK at Rivian



Ecosystem



Security



Monitoring and
observability



Cost
Optimizations

Tools and services with Amazon MSK

Tools



UI for Apache Kafka

Primary admin UI
Single sign-on (SSO) integration with Active Directory



Cruise control

Custom goals and goal violation
Anomaly detection

Services



Schema registry

Data governance and schema validation
70% reduction in bytes in



MirrorMaker 2

Mirror topics across Availability Zone and Region
Hydrate topics in new AWS account, cross Region
Plan to weigh in AWS managed MirrorMaker

Rivian

Rivian's Polyglot Team



Apache Flink



Apache Spark
Structured Streaming



Spring Boot Java



Python



Rust



.Net/C#



Go



AWS Lambda



Kafka REST Proxy

JVM based

Flink, Spark Structured Streaming, Spring Boot Java clients

librdkafka based

Python, Rust, Go and .Net/C# clients

Event based

AWS Lambda based event source mappings

Open-source components

Kafka Connect, Kafka REST Proxy

Security



Authentication

SASL_SSL/SCRAM-SHA-512 auth or IAM-based authentication
SCRAM-SHA-512 include username/password mechanisms



Authorization

API-based microservice for authorization for access control lists (ACL)



Data encryption

Encryption at rest using AWS KMS
Encryption in transit (TLS 1.2)

Monitoring and observability

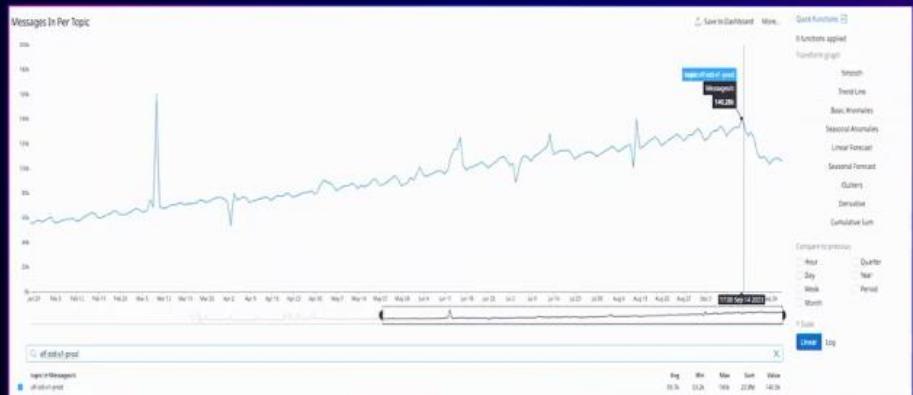
Monitoring services



DATADOG

Alerting

PagerDuty



24,337

cars produced
in 2022

140%

increase in
messages/sec

39,691

cars produced
YTD Q3 2023

DATA

Cost optimizations

Client side

Rack awareness to optimize data transfer costs

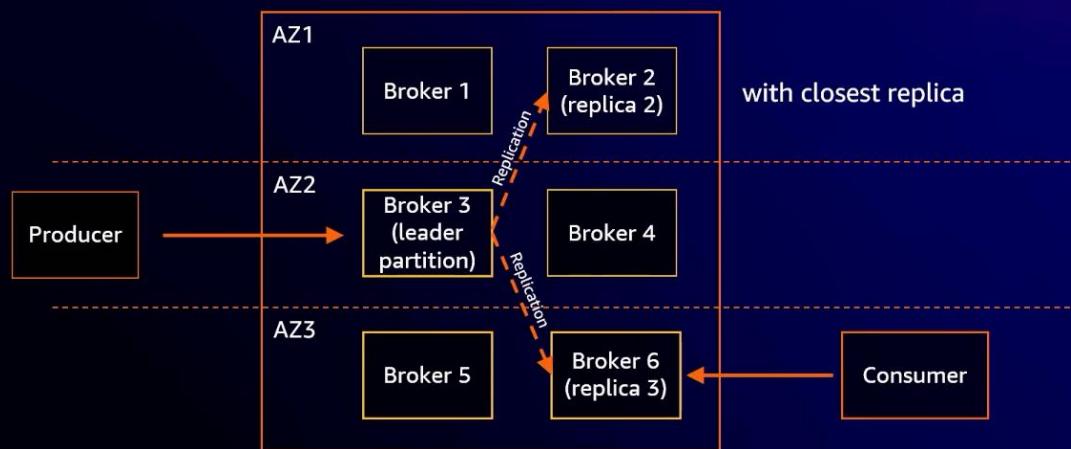
Kafka producer compression

Broker side

Tiered storage

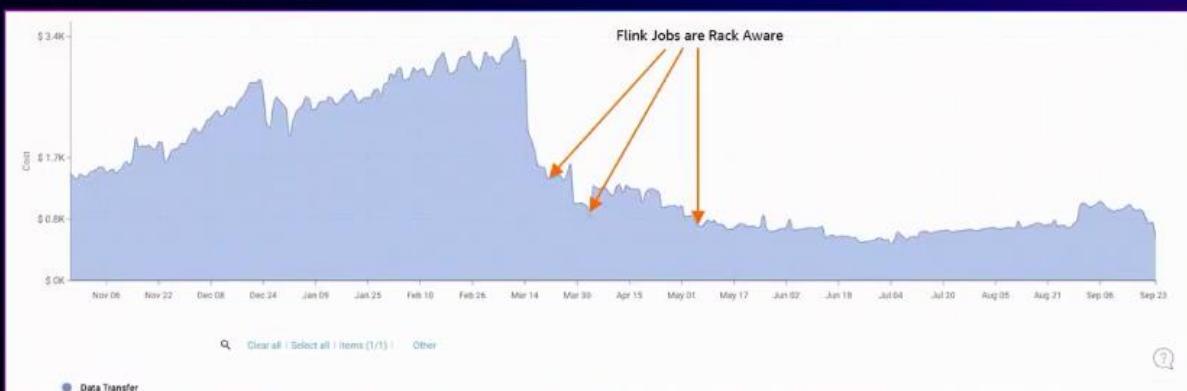
Rack awareness

OPTIMIZING DATA TRANSFER COST



Rack awareness

OPTIMIZING DATA TRANSFER COST



85%

YoY reduction in data cost

Broker settings

`replica.selector.class=org.apache.kafka.common.replica.RackAwareReplicaSelector`

Client settings

`client.rack=use1-az4`

Kafka producer compression

CLIENT-SIDE OPTIMIZATIONS

BytesInPerMessage drop

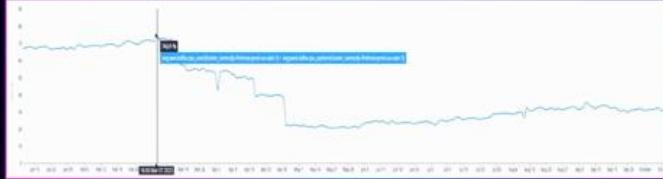


70%

Reduction in bytes in

Producer
compression.type
=snappy

YoY avg broker CPU % drop



57%

Reduction in cluster CPU

Batching
linger.ms=100

Amazon MSK Tiered Storage

SERVER-SIDE OPTIMIZATIONS

45% cost reduction with tiered storage



Load testing results look promising

Planned to deploy in production by Q1 2024

Lessons learned



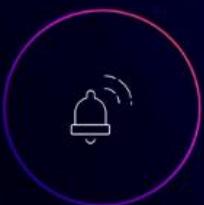
Kafka transactions



Cross-AZ data transfer



Capacity planning



Observability alarms



Access control

Opportunities ahead



Scale, Scale and Scale



New Applications



Global Deployments



Cost Benefits



Robust Toolkit



AI-based Innovations



RIVIAN

+



=



Amazon MSK

Thank you! Check out these sessions as well:

ANT201 Accelerating Innovation with Real-Time Data

ANT315 Accelerating value from data: Migrate from batch to stream processing

ANT401 Event detection with MSK and Amazon Managed Service for Apache Flink



Learn more about streaming data



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Learn more about modern
data streaming architectures on AWS