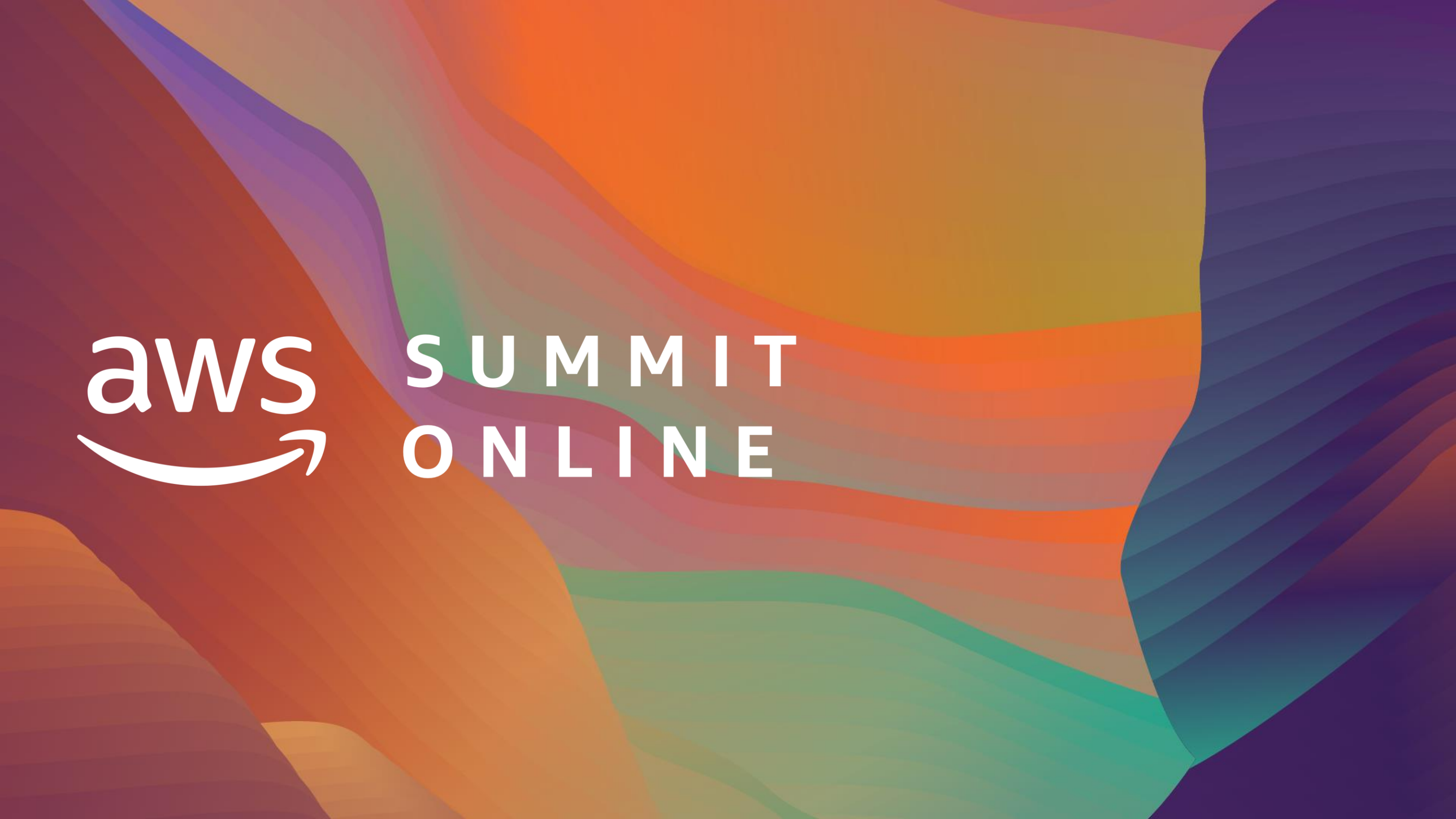

aws
SUMMIT
ONLINE

Streaming and real-time analytics

Donnie Prakoso
Senior Developer Advocate, ASEAN
Amazon Web Services

aws SUMMIT ONLINE

Agenda

- Use cases and customer examples
- Why real-time data streaming and analytics?
- Principles of data streaming
- Real-time streaming on AWS
- Video streaming with Amazon Kinesis Video Streams

EPIC GAMES

Fortnite | 125+ million players

Challenge

Create a constant feedback loop for designers

Gain up-to-the-minute understanding of gamer satisfaction to guarantee that gamers are engaged, making it the most popular game played in the world

Why real-time data streaming and analytics?

aws SUMMIT ONLINE

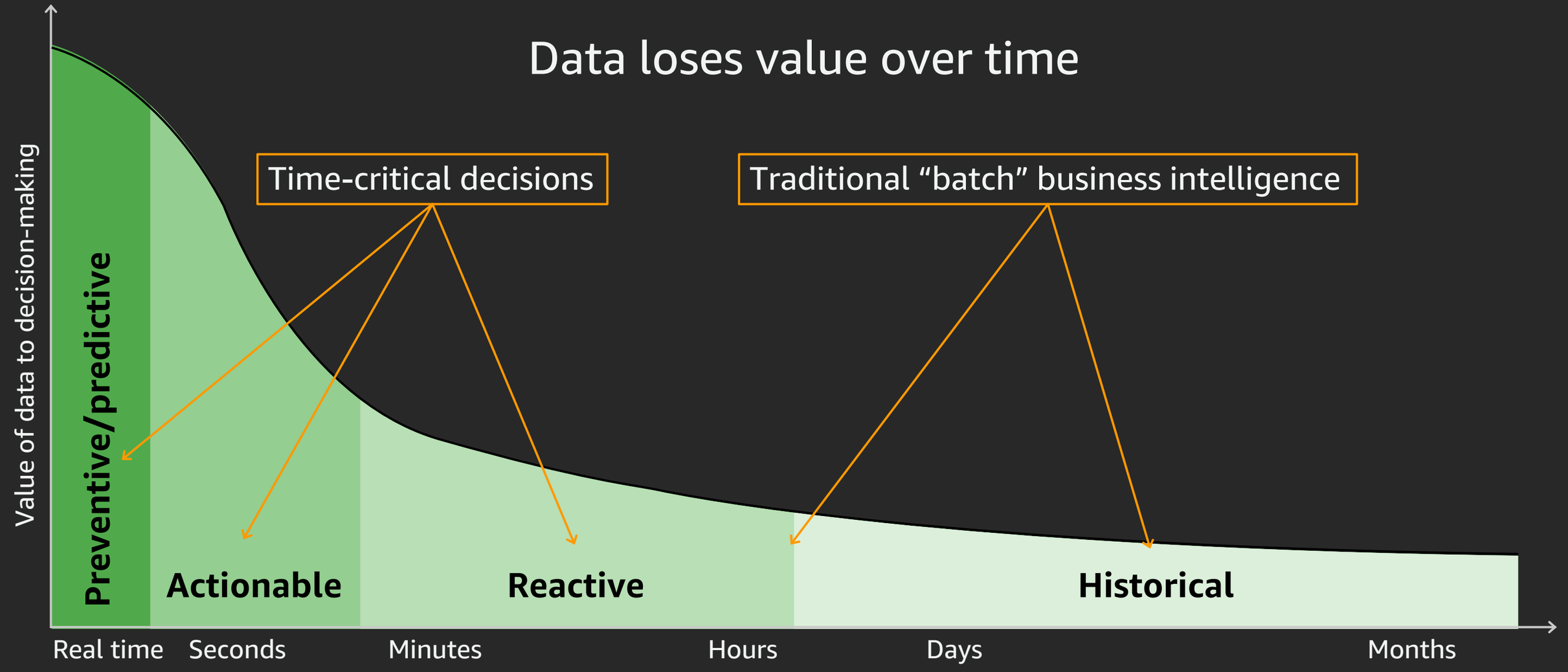
real-time

To create value, companies must derive insights from a variety of data sources that are producing data at high volume and velocity

> "Data integration requirements … now demand real-time streaming, replication, and virtualized capabilities…"

Gartner, 2019 Planning Guide for Data and Analytics

Why real-time data streaming and analytics?



Source: Mike Gualtieri, Forrester, “Perishable insights”

Current trends

Data streams are increasingly being used for messaging

Data streams are replacing batch workflows to lower analytics latency

Data streams are the event freeway for microservices

Database integrations are happening via change streams

In-stream machine learning (ML) and artificial intelligence (AI) for real-time automation

Stream new data in seconds

Get actionable insights quickly

Ingest data **as it's generated**

Process data **without interrupting the stream**

Real-time analytics/ML, alerts, actions

Principles of data streaming

The foundation of real-time analytics

1. Data can be produced, captured, and processed in milliseconds

2. Data is buffered, enabling parallel and independent I/O

3. Data must be captured and processed in the order it was produced

Data streams are for real time

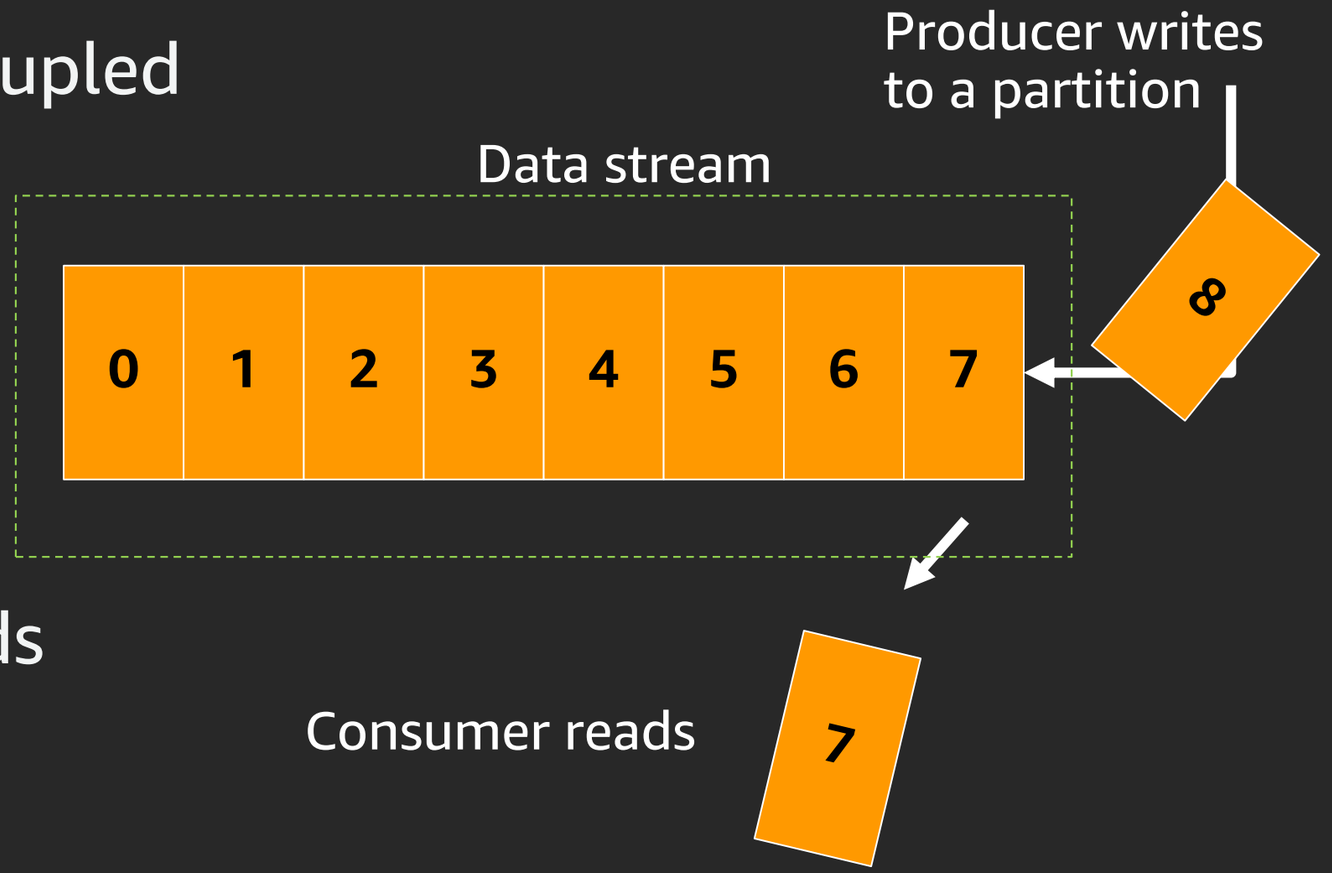
Commit log (aka partition) is the core of a data stream

Producers and consumers are decoupled

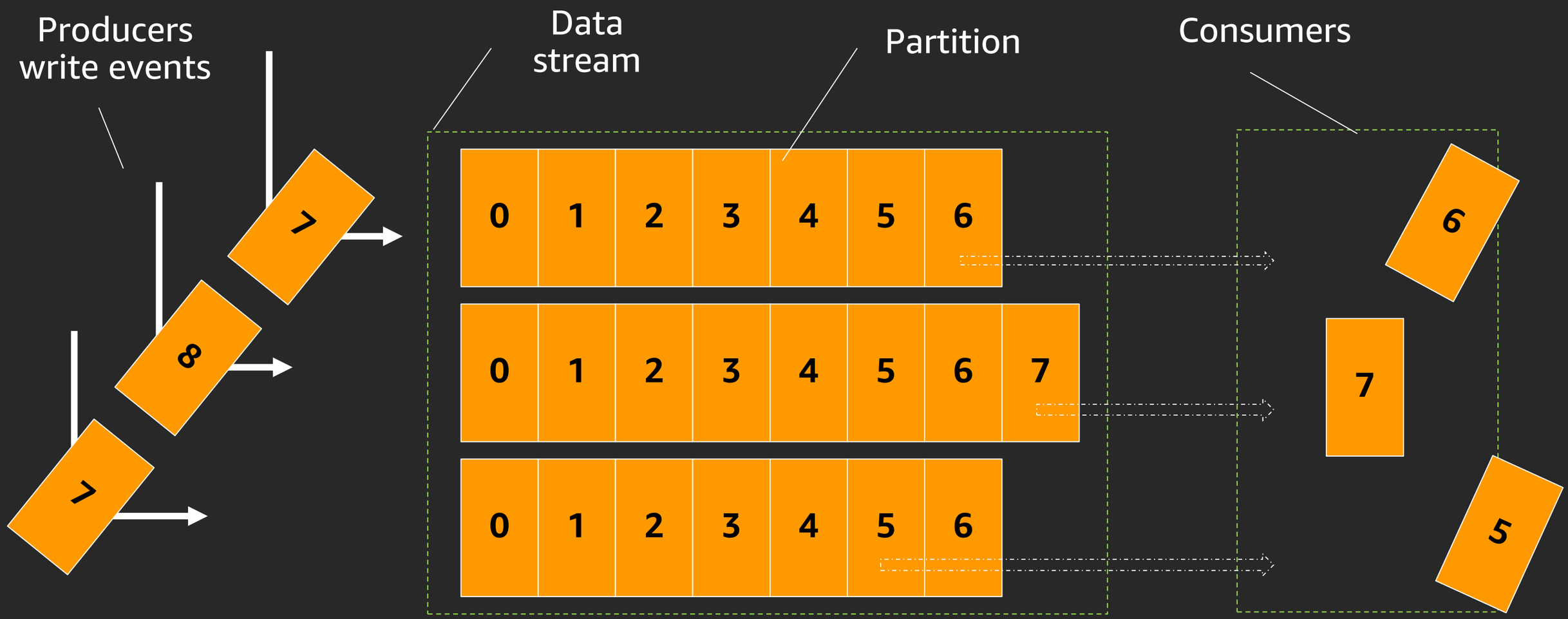
Data is retained, not destroyed

Write order is preserved

Produce → Consume in milliseconds



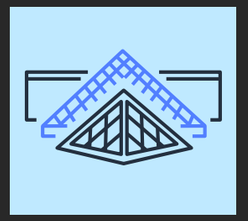
Data stream concepts



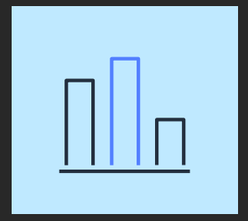
A data stream is a logical grouping of *partitions* or *shards*

Challenges of data streaming

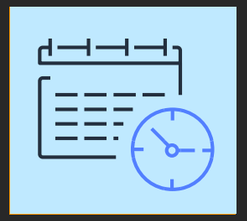
Difficult to set up



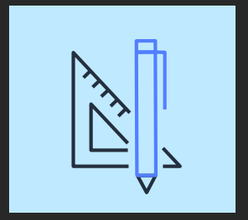
Tricky to scale



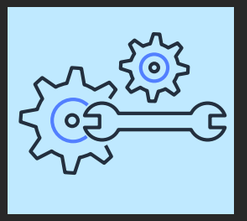
Hard to achieve high availability



Integration requires development



Error-prone and complex to manage



Expensive to maintain

Real-time streaming on AWS

aws SUMMIT ONLINE

Real-time streaming on AWS

Easily collect, process, and analyze data and video streams in real time

Easy to use

Elastic

High availability and durability

Seamless integration with AWS services

Fully managed

Pay for what you use

Real-time streaming on AWS

Easily collect, process, and analyze data and video streams in real time

Amazon Kinesis Data Streams

Collect and store data streams for analytics

Amazon Kinesis Data Firehose

Load data streams into AWS data stores

Amazon Kinesis Data Analytics

Analyze data streams with SQL or Java

Amazon Managed Streaming for Apache Kafka

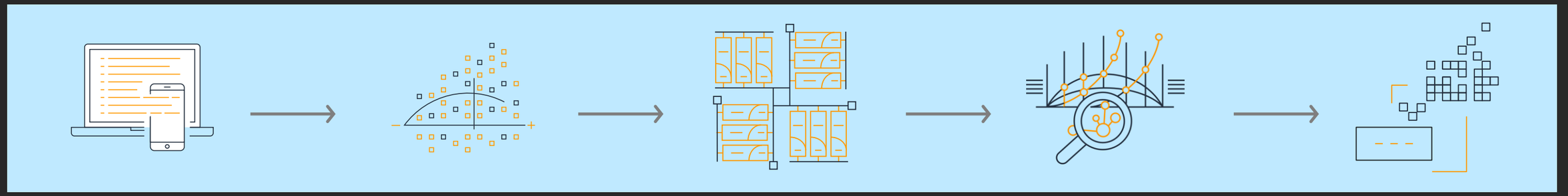
Collect and store data streams for analytics

Amazon Kinesis Video Streams

Capture, process, and store media streams for playback and analytics

Streaming data architecture

Data streaming technology enables customers to ingest, process, and analyze high volumes of high-velocity data from a variety of sources in real time



Data sources

Devices and/or applications that produce real-time data at high velocity

Stream ingestion

Data from tens of thousands of data sources can be written to a single stream

Stream storage

Data is stored in the order in which it was received for a set duration, and it can be replayed indefinitely during this time

Stream processing

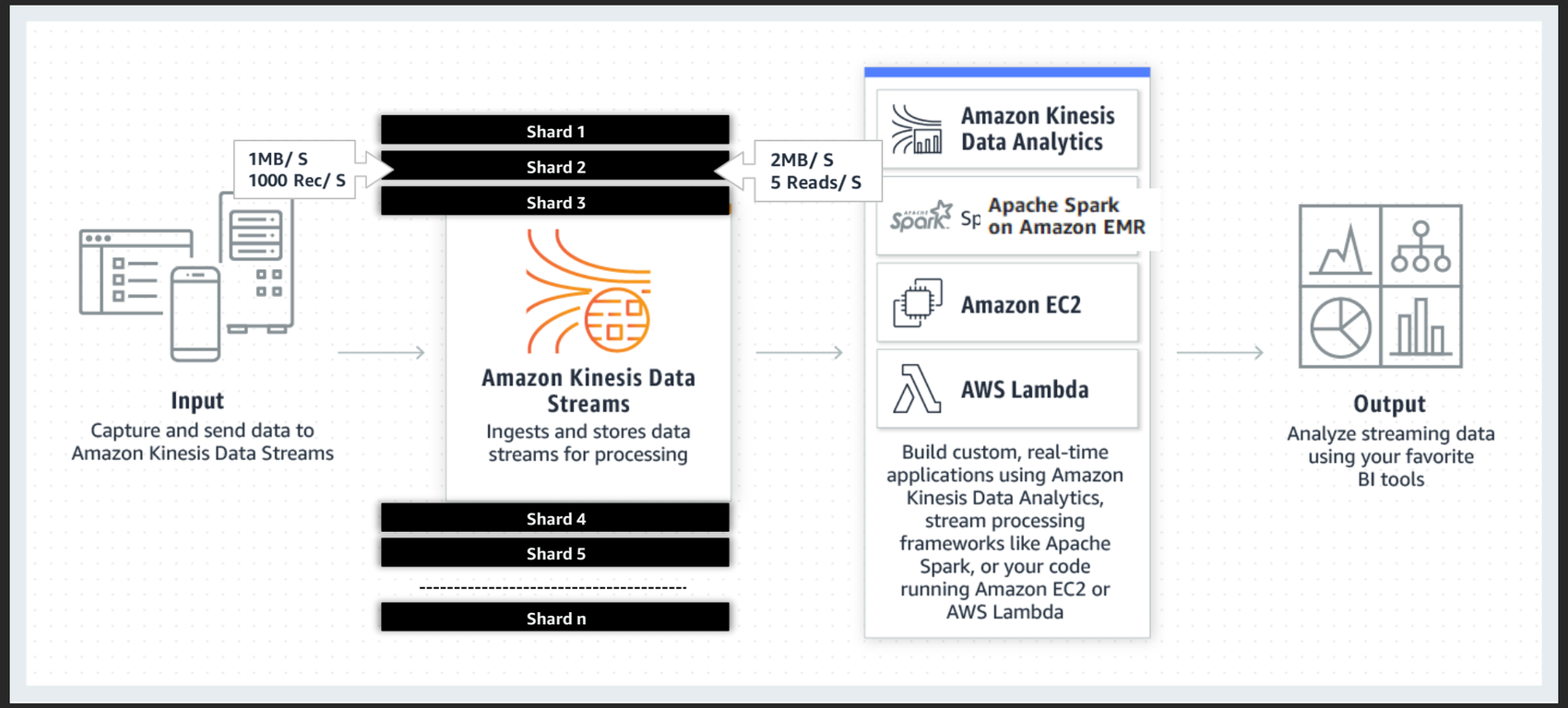
Records are read in the order in which they are produced, enabling real-time analytics or streaming ETL

Destination

Data lakes, databases, and analytics services

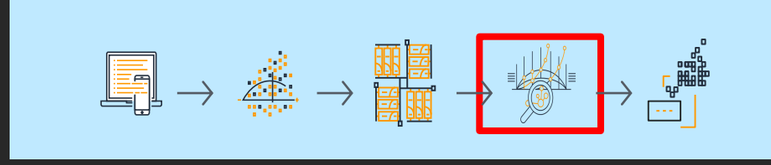
Amazon Kinesis Data Streams

Collect streaming data, at scale, for real-time analytics

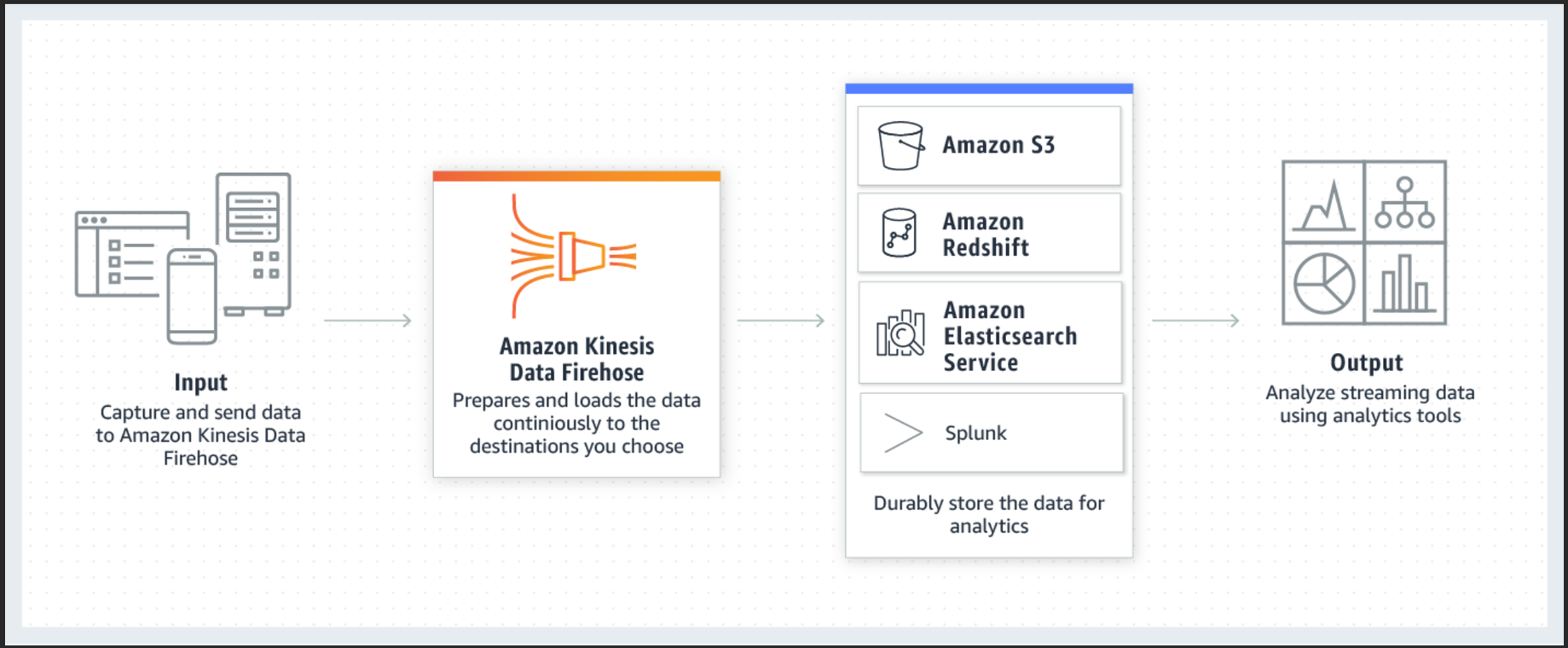


- Easy administration and low cost
- Real-time, elastic performance
- Secure, durable storage
- Available to multiple real-time analytics applications
- Average latency of 200 ms with one standard consumer
- Enhanced fan-out offers typical average latency of 70 ms

Amazon Kinesis Data Firehose



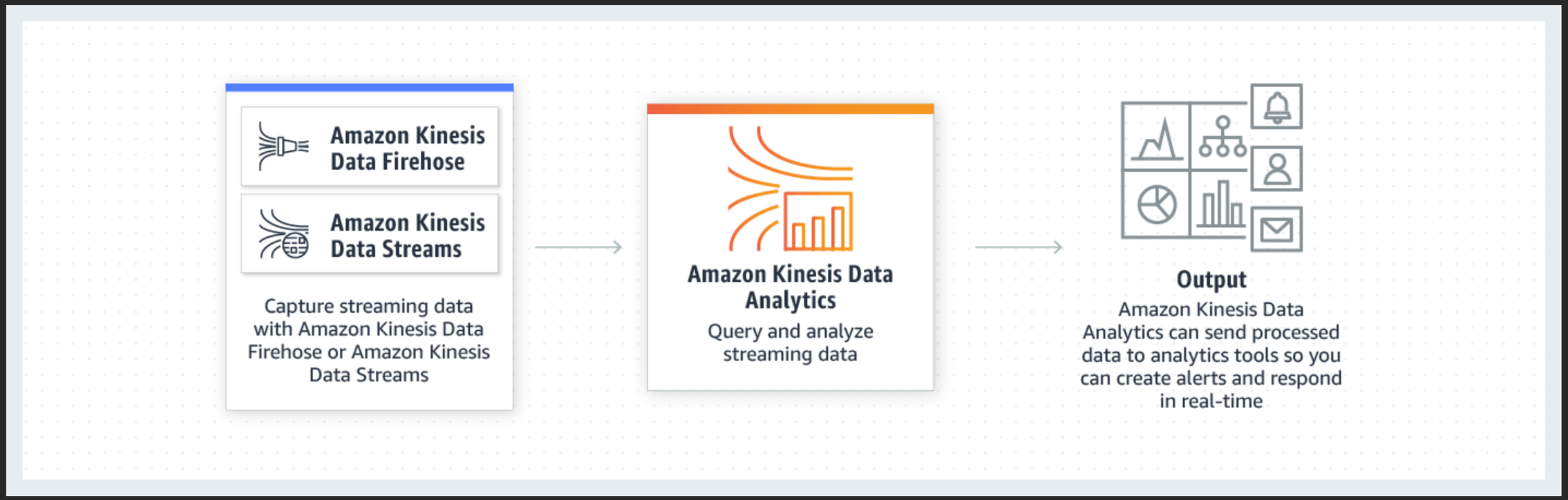
Prepare and load real-time data streams into data stores and analytics tools



- Zero administration and seamless elasticity
- Direct-to-data-store integration
- Serverless continuous data transformations
- Data format conversion to Parquet or ORC
- Near real time

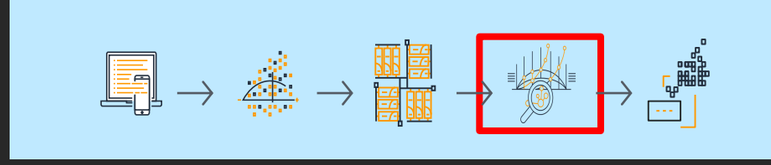
Amazon Kinesis Data Analytics

Get actionable insights from streaming data in real time



- Interact with streaming data in real time using SQL or integrated Java applications
- Build fully managed and elastic stream processing applications

Amazon Kinesis Data Analytics SQL

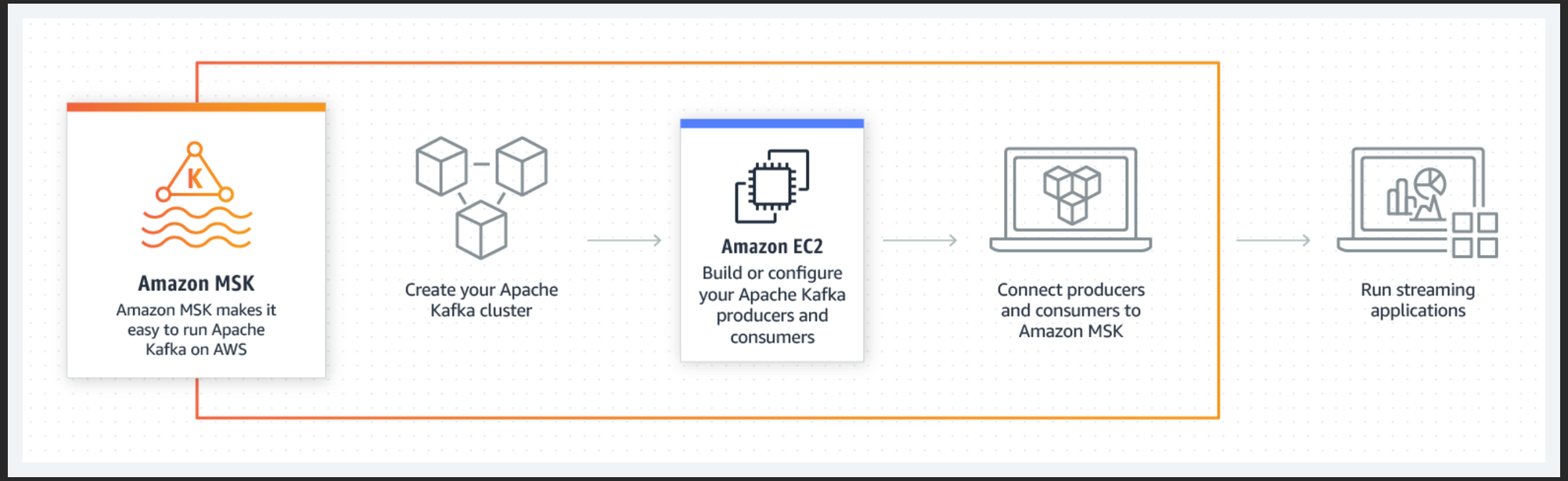


For simple and fast use cases

- Subsecond, end-to-end processing latencies
- SQL steps can be chained together in serial or parallel steps
- Build applications with one or hundreds of queries
- Prebuilt functions include everything from sum and count distinct to ML algorithms
- Aggregations run continuously using window operators

Amazon Managed Streaming for Apache Kafka (Amazon MSK)

Fully managed, highly available, secure Apache Kafka service



- Migrate and run existing Apache Kafka applications on AWS without changes
- AWS manages the provisioning, configuration, and maintenance of Apache Kafka clusters
- Continuous cluster health monitoring and component replacement
- Multiple levels of security – network isolation, authorization, encryption, and access control

Demo: Analyzing web logs

aws SUMMIT ONLINE

Demo: Analyzing web logs

aws SUMMIT ONLINE

QuickSight

Cloud-native business intelligence with pay-per-session pricing and ML insights for everyone

Amazon QuickSight benefits

Cloud native = No servers = Automatic scaling

No servers or software to manage, maintain, or deploy; start with 10s of users and scale to 10s of 1,000s

Fast, consistent performance

Fast, predictable performance every time; concurrent users or increased interactions do not slow down the system

Fully integrated with AWS

Build end-to-end analytics in AWS; secure private VPC access, fine-grained access control, ML integrations

ML insights

Contextual, relevant insights with ML-powered anomaly detection, forecasting, alerts, and customizable narratives

Secure and global

End-to-end encryption; native high availability; 10 global regions; HIPAA, PCI, ISO, SOC, and FedRAMP eligibility

Insights for everyone

Provide access to all users, pay only for usage; no upfront costs, no charges for inactive users

Easy to develop and maintain

Design with QuickSight, integrate with APIs; secure data with row-level security and authenticate seamlessly via SAML SSO

Customize and embed

Embed into applications and enable analytics in hours, not months or years; use themes to match application and corporate branding

Administrator ⟵ ⟶ End user

Benefits

Demo: QuickSight

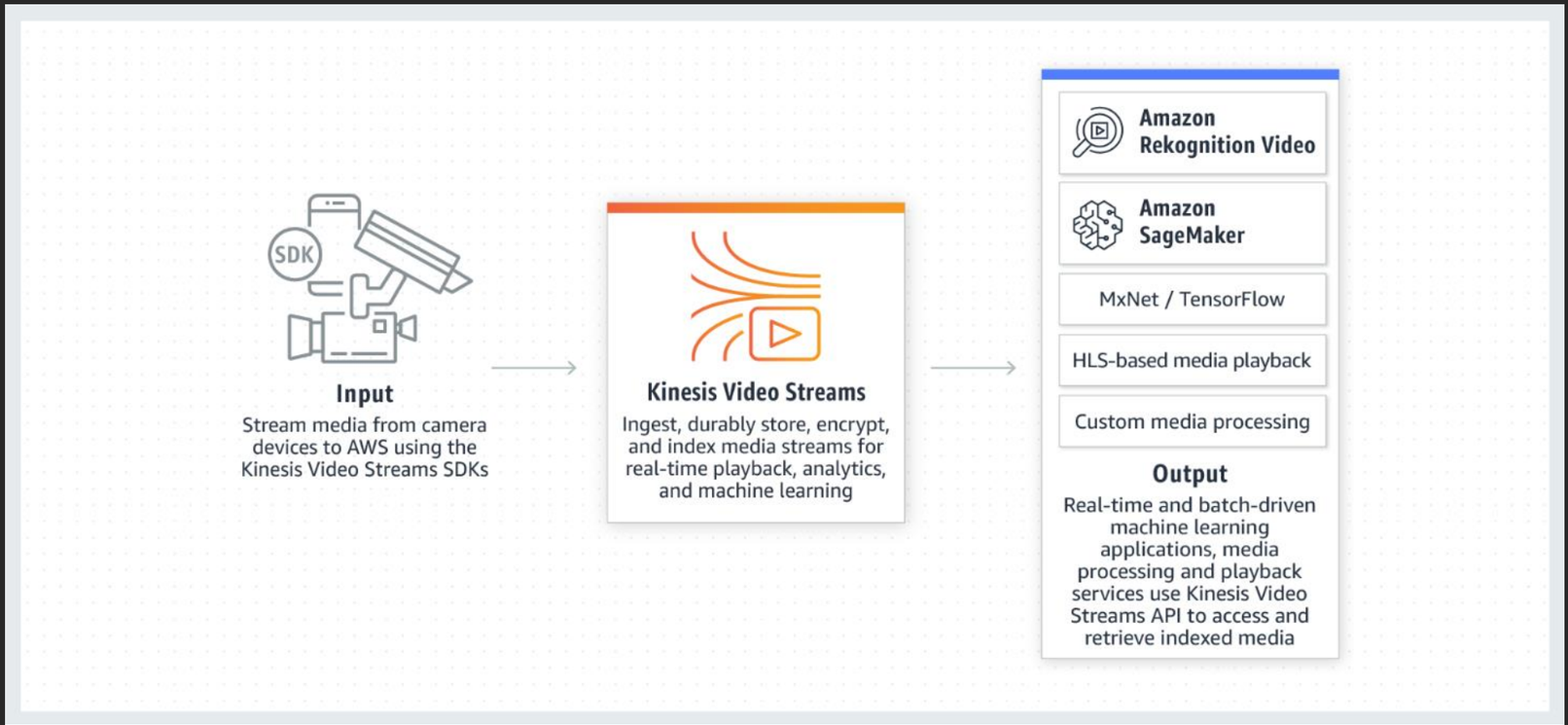
aws SUMMIT ONLINE

Video stream

aws SUMMIT ONLINE

Amazon Kinesis Video Streams

Stream video and audio in real time for playback, storage, and analytics



- Stream video from millions of devices
- Easily build vision-enabled applications
- Playback live or on demand
- Durable, searchable storage
- Secure and fully managed

Amazon Kinesis Video Streams

Kinesis Video Streams is a fully managed service that elastically scales to support the media streaming needs of millions of camera devices

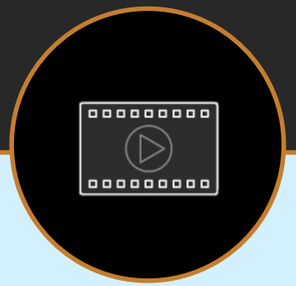
Low latency, live streaming, and two-way interactivity

WebRTC audio, video, and data

Large-scale, secure media ingestion and retention

Encrypted, durable, and time-indexed media storage

Real-time and on-demand playback and processing



HLS, MPEG-DASH, and integration with AWS ML services

Demo: WebRTC

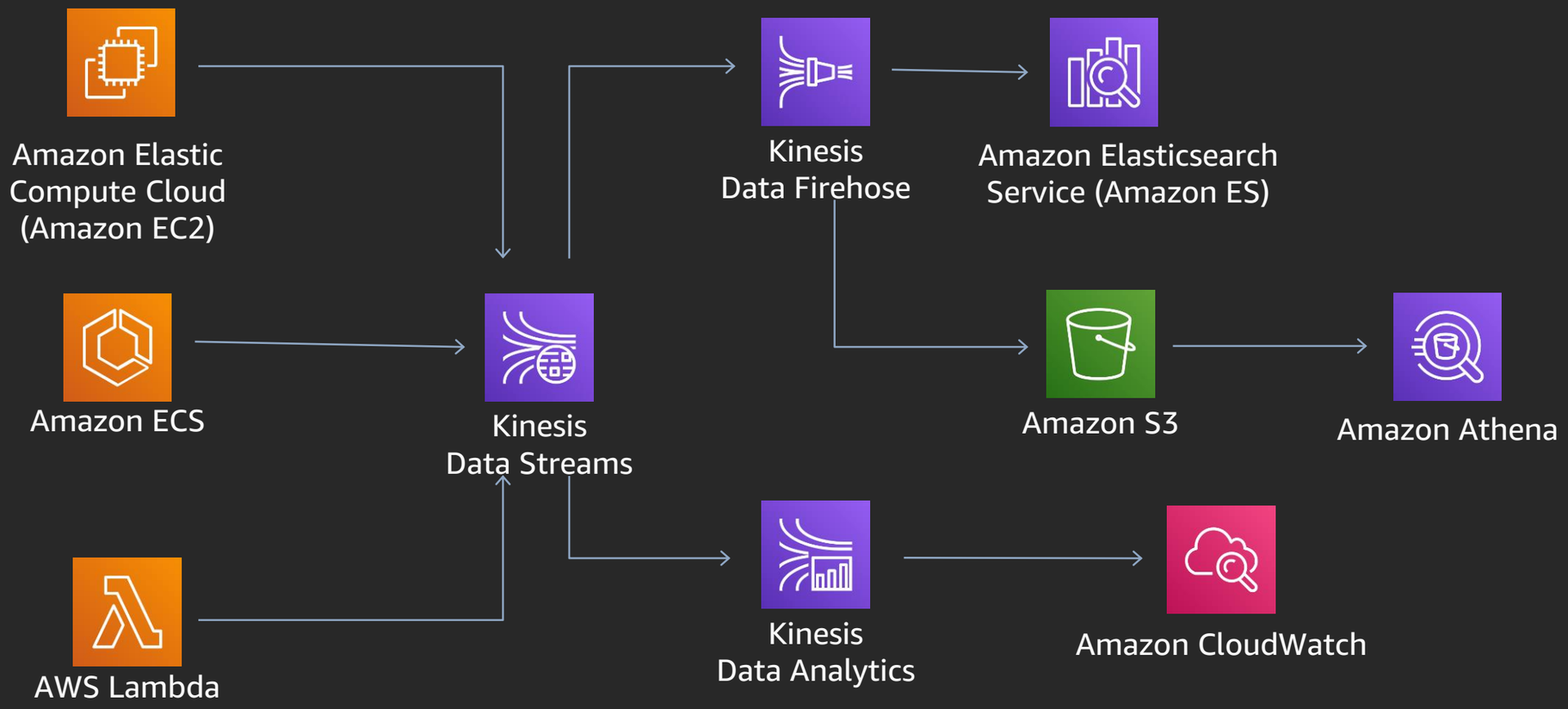
aws SUMMIT ONLINE

Moving forward

aws SUMMIT ONLINE

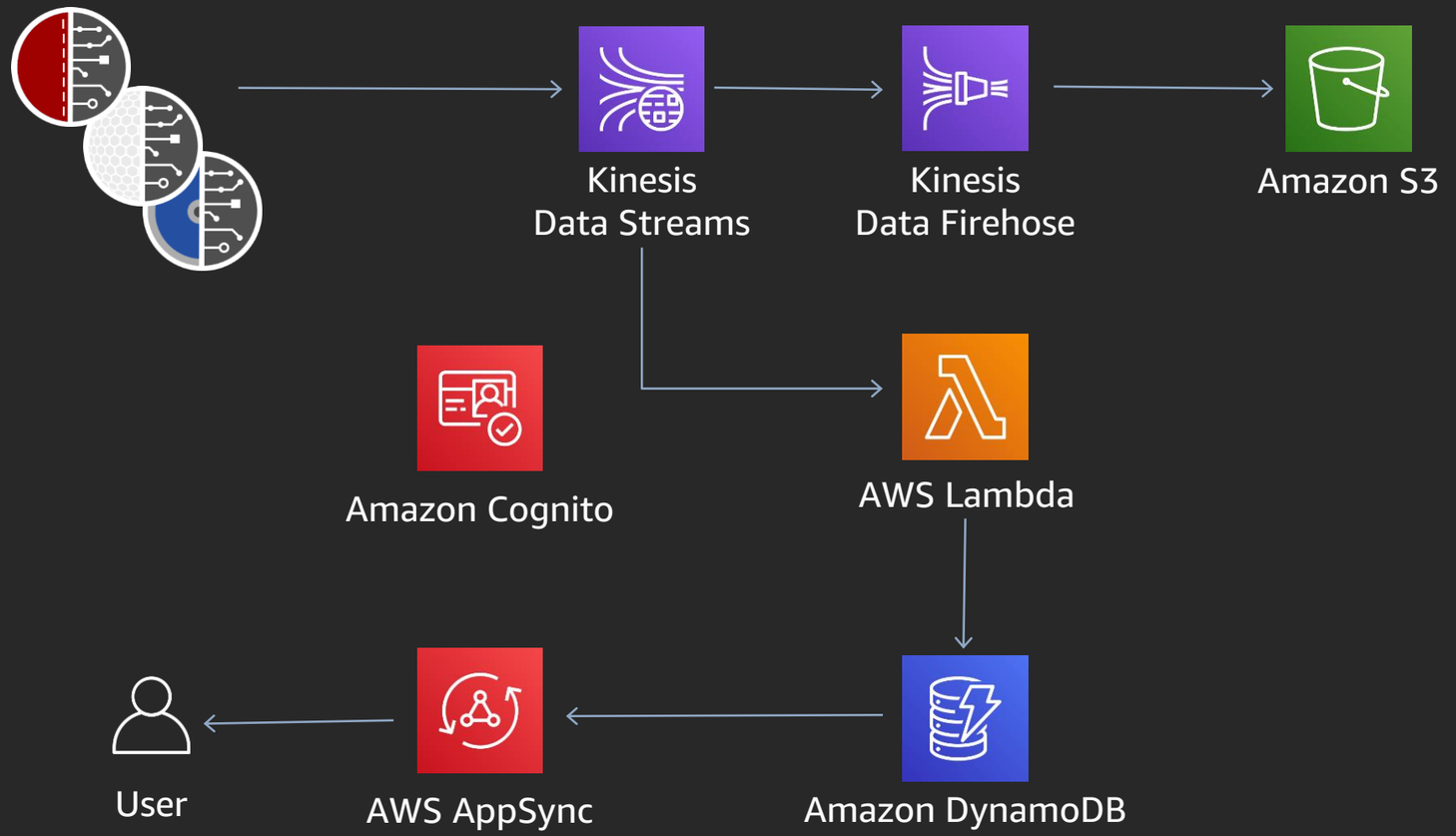
Log ingestion

Terabytes of data processed in real time every day



IoT

Sensor-powered sports: Cricket, golf, rugby, and more



Learn analytics with AWS Training and Certification

Resources created by the experts at AWS to help you build and validate data analytics skills

New free digital course: **Data Analytics Fundamentals**

Classroom offerings, including **Big Data on AWS**,
feature AWS expert instructors and hands-on labs

Validate expertise with the **AWS Certified Big Data—Specialty** exam or
the new **AWS Certified Data Analytics—Specialty** beta exam

Visit aws.amazon.com/training/paths-specialty/

Thank you!

Donnie Prakoso


@donnieprakoso


donnieprakoso

aws SUMMIT ONLINE