AWS KINESIS

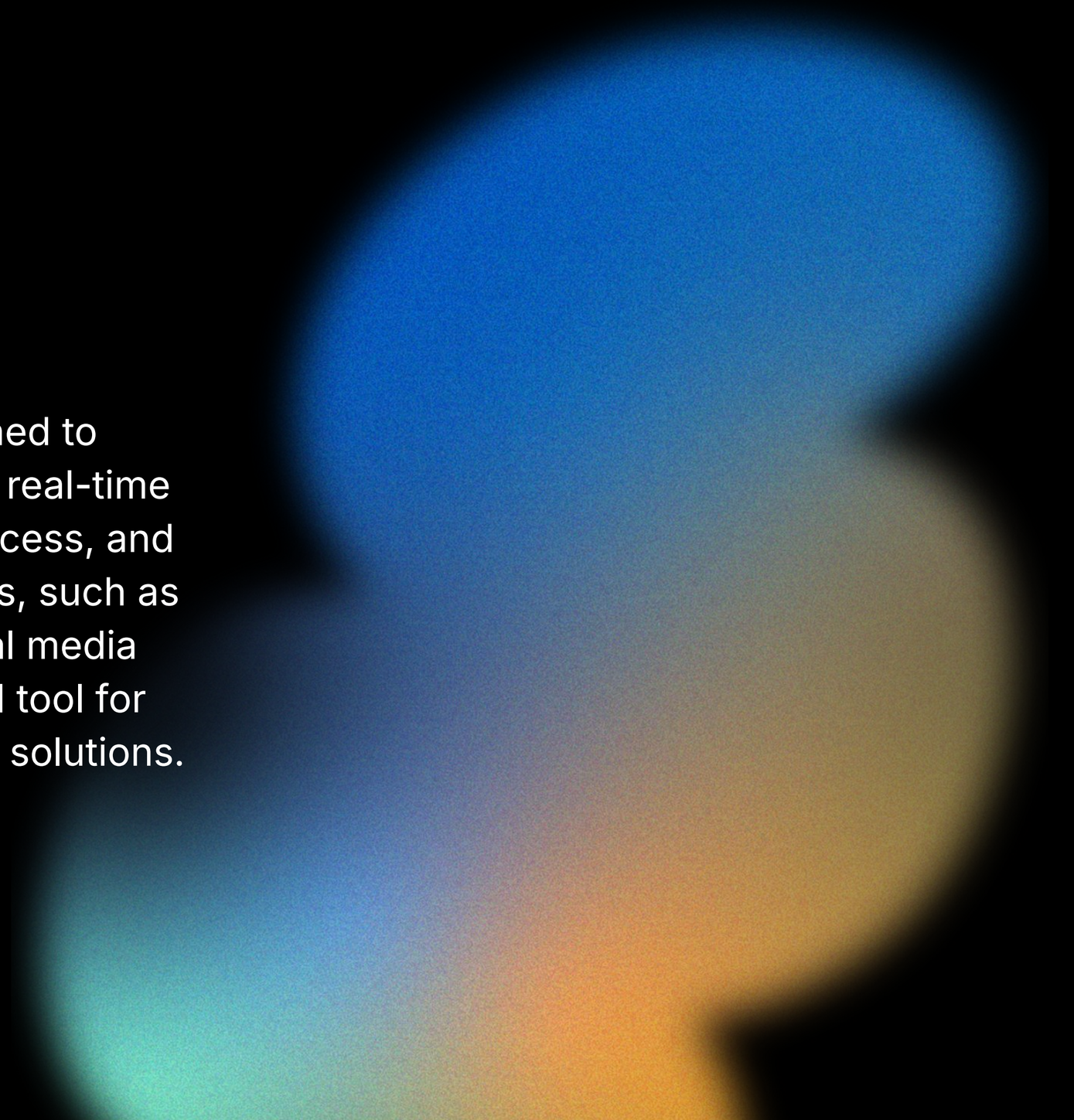
The Tech
Stuff

Table Of Contents

Introduction

Introduction

Amazon Kinesis is a suite of services designed to handle large-scale data streaming, enabling real-time data processing. It allows you to collect, process, and analyze streaming data from various sources, such as website clickstreams, application logs, social media feeds, and IoT devices. Kinesis is a powerful tool for building real-time applications and analytics solutions.



Components

Components

Core Components of AWS Kinesis

- Kinesis Data Stream
- Kinesis Fire House
- Kinesis Data Analytics
- Kinesis Video Stream


Data Stream

Kinesis Data Stream

Kinesis Data Streams is the foundational service within the Kinesis suite that enables the collection and processing of real-time streaming data. It is designed to handle massive amounts of data, supporting throughput of terabytes per hour.

Example:
A web application sends user clickstream data to a Kinesis Data Stream.

A real-time analytics application consumes the data, providing insights on user behavior within seconds of interaction.

Key Features

Shards

Each data stream consists of multiple shards, which are the basic units of parallelism. Each shard can ingest up to 1MB of data per second and read up to 2MB per second.

Producers

Applications or devices that send data records to the stream. Producers can be anything from IoT devices to web servers or mobile applications.

Consumers

Applications that consume data from the stream, processing records as they arrive. Examples include real-time analytics platforms, data lakes, or machine learning models.

Partition Key

Each data record within a shard is identified by a unique partition key, which determines the shard that will store the data.

Retention Period

Data in a stream is retained for a configurable period, ranging from 24 hours up to 7 days. This allows for replaying and reprocessing of data if needed.

Use Cases

Real-time Analytics

Analyzing user behavior in real-time from clickstreams.

Log and Event Data Collection

Ingesting log data from servers for real-time monitoring and alerting.

Fraud Detection

Analyzing financial transactions as they happen to detect fraud.


Fire House

Kinesis Fire House

Kinesis Data Firehose is the easiest way to reliably load streaming data into data lakes, data stores, and analytics services. It is fully managed, scales automatically to match the data throughput, and offers integrated data transformation capabilities.

Example:
An IoT application sends sensor data to a Kinesis Data Firehose delivery stream.

The data is transformed using AWS Lambda, compressed, and stored in Amazon S3 for further analysis.

Key Features

Compression

Supports data compression (e.g., Gzip, Zip) to reduce storage costs.

Encryption

Provides encryption at rest using AWS KMS and in transit using SSL/TLS.

Multiple Destinations

Delivers data to Amazon S3, Amazon Redshift, Amazon Elasticsearch Service, Splunk, and any HTTP endpoint.

Automatic Scaling

Kinesis Data Firehose is the easiest way to reliably load streaming data into data lakes, data stores, and analytics services. It is fully managed, scales automatically to match the data throughput, and offers integrated data transformation capabilities.

Data Transformation

You can transform data using AWS Lambda before it is delivered to the destination. This is useful for filtering, aggregating, or modifying data formats.

Use Cases

Real-time Data Loading

Stream data directly into Amazon Redshift for real-time analytics without the need for custom coding.

Log & Event Data Collection

Continuously deliver server logs to Amazon S3 for long-term storage and later analysis.

Monitoring & Alerting

Deliver security or application logs to Amazon Elasticsearch Service for real-time monitoring and search.


Data Analytics

Kinesis Data Analytics

Kinesis Data Analytics is the service that enables you to process and analyze streaming data in real-time using SQL. It is fully managed, meaning you don't need to manage the underlying infrastructure.

Example:
A Kinesis Data Stream collects clickstream data from a website.
Kinesis Data Analytics processes the stream in real-time to compute metrics like the number of clicks per minute.
The processed data is delivered to an Amazon S3 bucket via Kinesis Data Firehose for storage and later analysis.

Key Features

SQL Queries

Allows you to run SQL queries on streaming data, making it easy for those familiar with SQL to process data streams.

Real Time Processing

Perform real-time aggregations, filtering, and windowed computations on streaming data.

Key Features

Ingestions

Easily integrates with Kinesis Data Streams and Kinesis Data Firehose, allowing you to use the processed data for storage, analytics, or further streaming.

Application Templates

Provides templates for common use cases like anomaly detection, stream filtering, and time-series analytics.

Use Cases

Real-time Analytics

Continuously analyze data streams to derive insights, such as detecting trends or anomalies.

Streaming ETL

Extract, transform, and load (ETL) data in real-time before it reaches its final destination.

Monitoring Metrics

Analyze metrics in real-time, such as application performance metrics, and trigger alerts based on certain conditions.


Video Stream

Kinesis Video Stream

Kinesis Video Streams enables you to securely stream video from connected devices to AWS for analytics, machine learning (ML), and other processing. It supports live video, audio, and other time-encoded data, making it ideal for IoT applications.

Example:
Security cameras send live video streams to Kinesis Video Streams.

The streams are processed using Amazon Rekognition for real-time object detection, and the results are stored in Amazon S3.

Key Features

Secure & Scalable

Kinesis Video Streams can scale automatically to ingest streams from thousands of devices securely.

Real Time Processing

Enables real-time video processing using Amazon Rekognition or custom machine learning models.

Key Features

Playbacks

Streams can be viewed in real-time or archived and replayed on demand.

Integration with ML Services

Integrates with AWS machine learning services, allowing for advanced analytics like object detection or sentiment analysis.

Use Cases

Smart Homes

Stream video from home security cameras and analyze it in real-time for intruder detection.

Industrial Iot

Monitor industrial equipment using live video streams for fault detection or predictive maintenance.

Healthcare

Stream and analyze video data from healthcare devices, such as monitoring patient conditions in real-time.

How Kinesis Works

01) Data Producers

Devices or applications that generate data and send it to Kinesis services.

Examples include web servers sending clickstream data, mobile devices sending user activities, or sensors sending IoT data.

02) Data Streams

The core unit where data is collected and processed.

In Kinesis Data Streams, the data is divided into shards, while in Kinesis Data Firehose, the data is continuously delivered to destinations.

03) Consumers

Applications or services that read and process data from Kinesis streams.

Consumers can include real-time analytics applications, machine learning models, or data storage solutions.

04) Destinations

Final storage or processing points for streaming data.

Common destinations include S3, Redshift, Elasticsearch, and third-party tools like Splunk.

Use Cases

Use Cases

01)

Real-Time Analytics

Process and analyze data in real-time from sources like social media feeds, application logs, or clickstreams.

Use Kinesis Data Analytics to run SQL queries on streaming data to generate insights.

02)

Event Monitoring

Monitor and respond to IT infrastructure events, security logs, or application performance metrics in real-time.

Detect anomalies, trigger alerts, or automate responses using the insights gained.

Use Cases

03)

Data Ingestion for Big Data

Continuously ingest large volumes of data from various sources into data lakes or warehouses.

Kinesis Data Firehose simplifies loading this data into S3 or Redshift for batch processing or analysis.

04)

IoT Data Processing

Capture, process, and analyze data from IoT devices, such as temperature sensors, GPS trackers, or health monitors.

Use Kinesis Video Streams to handle video data and perform real-time video analytics.

Benefits

Benefits

Scalability

Automatically scales to handle large volumes of data streams.

Kinesis can handle thousands of data producers and consumers simultaneously, ensuring high throughput.

Real-Time Processing

Enables real-time data capture and processing, allowing for immediate insights and actions.

Ideal for time-sensitive applications, such as monitoring, analytics, and security.

Benefits

Flexibility

Supports a variety of use cases, from real-time analytics to video streaming and machine learning.

Allows you to choose the right tool within the Kinesis suite to meet your specific needs.

Real-Time Processing

Seamlessly integrates with other AWS services like S3, Redshift, Lambda, and SageMaker.

Supports easy setup and configuration, reducing the complexity of building real-time data pipelines.

Thank you!

Any Questions?

mayamnaizel2013@gmail.com

@the_techstuff

The Tech Stuff