# CDN Log Analytics

Business challenge

Content Delivery Networks generates huge amounts of log files as they stream video across the internet to our homes and Mobile devices. These logs contain vital information about the performance of the CDN servers and the quality of the video streaming. These logs run into terabytes of data and handling this data in real time and performing analytics to understand customer experience and network issues has its own challenges.

Approach and Solution

The approach was to build a solution to

- Manage the big data problem – a scalable platform to handle the data in real time

- Identify issues in thousands of servers in real time for quick rectification

- Understand real time customer experience and quality of video streaming
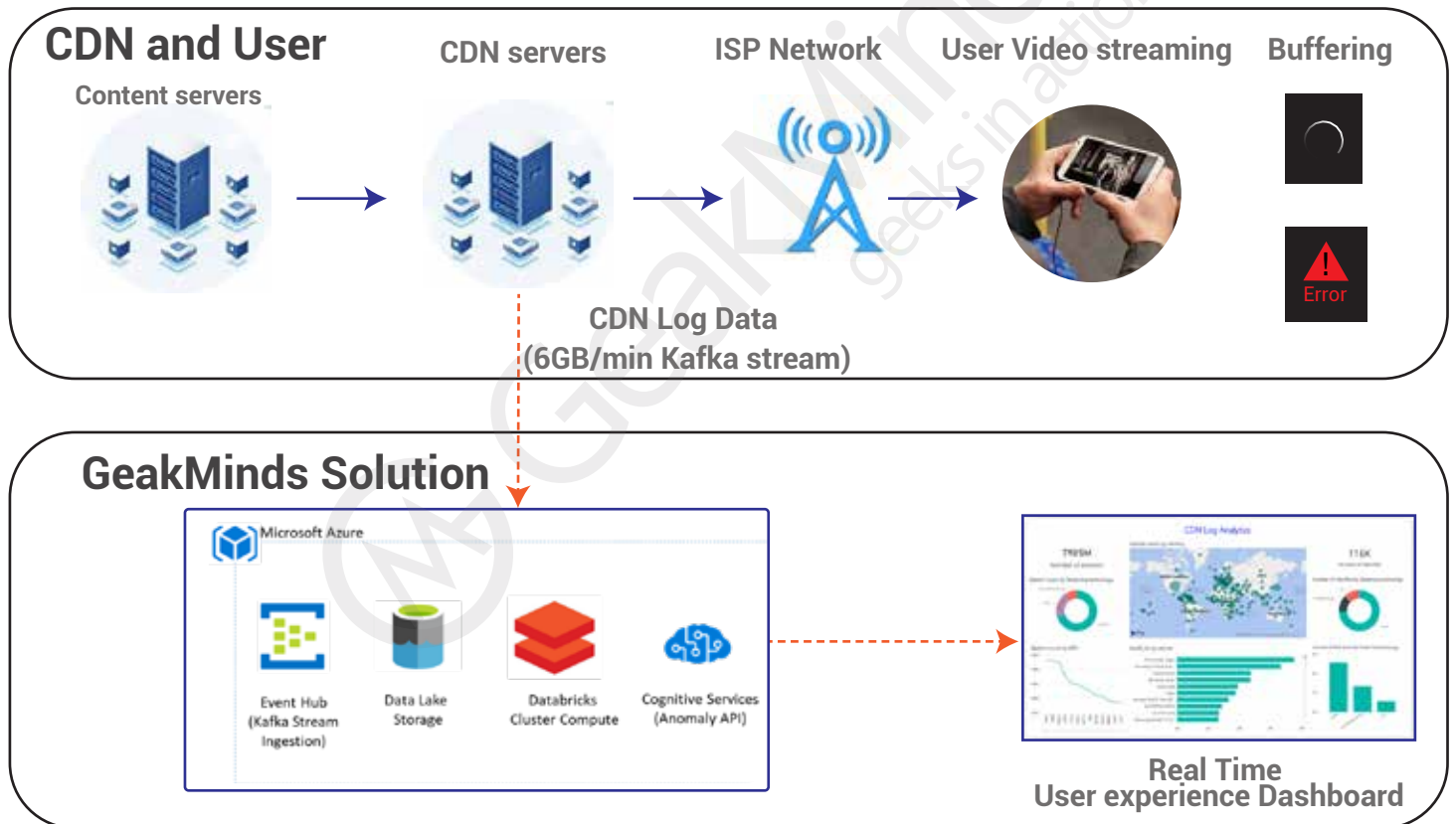
Cloud based Analytics solution was built as detailed

- A scalable platform using Microsoft Azure for Ingesting 6GB/min (8TB/day) data from various data sources

- Data Ingestion into Azure EventHub and using Kafka and spark structured streaming.

- Real time anomaly detection to identify issues in CDN servers using Azure Anomaly API

- Machine learning model using LightGBM for real time classification buffering and Normal user sessions.

- Custom build .net UI Dashboard for visualization

Benefits

- Real Time detection and hence fixing of issues resulting in reduction in buffering and improvement in video playback quality

- Improved customer experience and business