Design Analysis of Cloud-based Microservices Architecture at Netflix

03/03/2023

Netflix subscribers

- In 2019, 167M subscribers
- 5M+ subscribers each quarters
- 200+ countries
- 165M hours, 4k+ films, 47k episodes, daily

Infra transformation

- Since 2008
- Over 8 years
- From DC to Public Cloud
- From monolithic to microservices
- Chose AWS: HA DB, large-scale cloud storage, multiple DC globally, avoid undifferentiated heavy lifting work
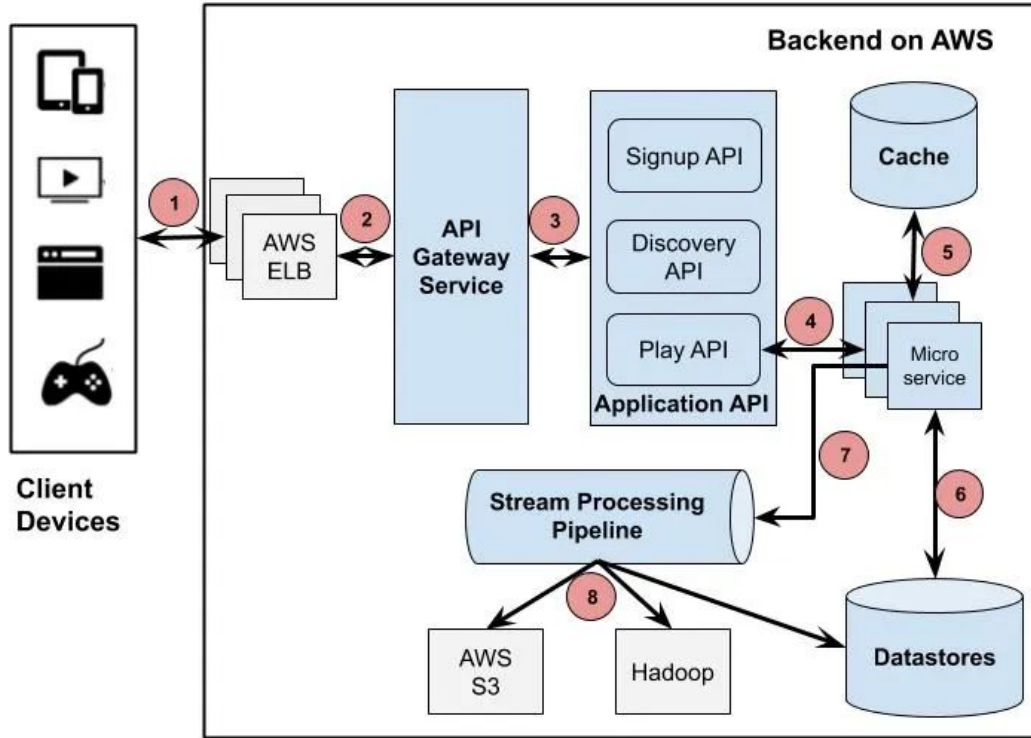
Architecture

- Client: browsers, mobile Apps, smart TVs
- Backend: handles everything not involving steaming videos
  - Scalable computing instances (AWS EC2)
  - Scalable storage (AWS S3)
  - Business logic microservices (purpose-built frameworks by Netflix)
  - Scalable distributed database (AWS DynamoDB, Cassandra)
  - Big data processing and analytics jobs (AWS EMR, Hadoop, Spark, Flink, Kafka and other purpose-built tools by Netflix)
  - Video processing and transcoding (purpose-built tools by Netflix)
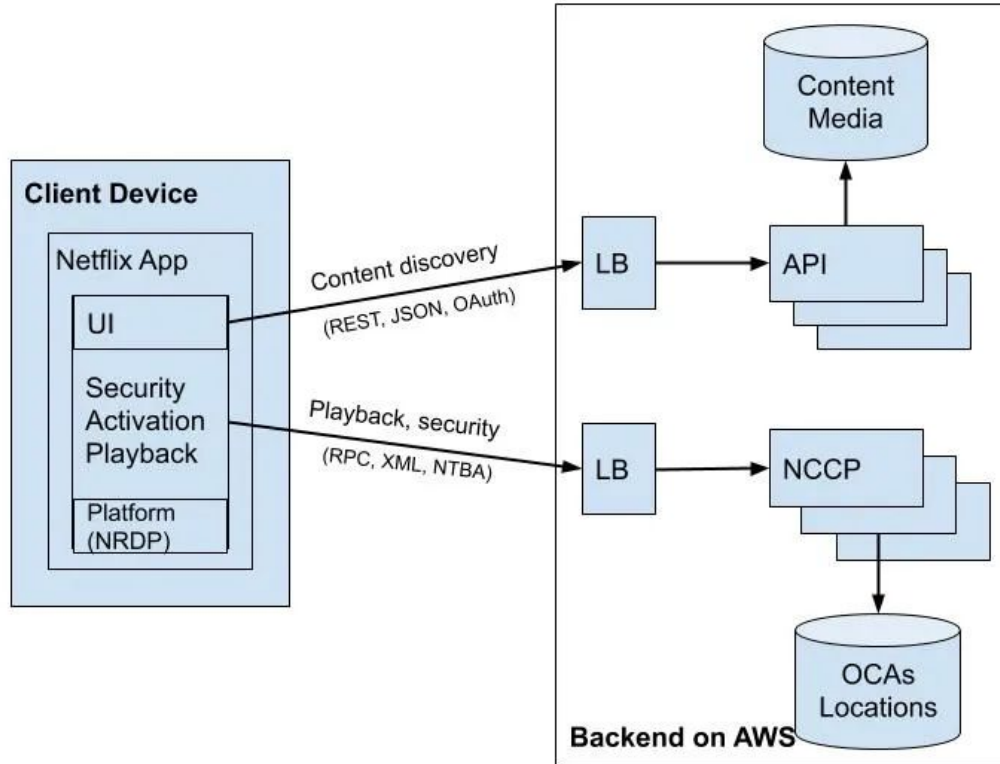- CDN: Open Connect Appliances (OCAs)

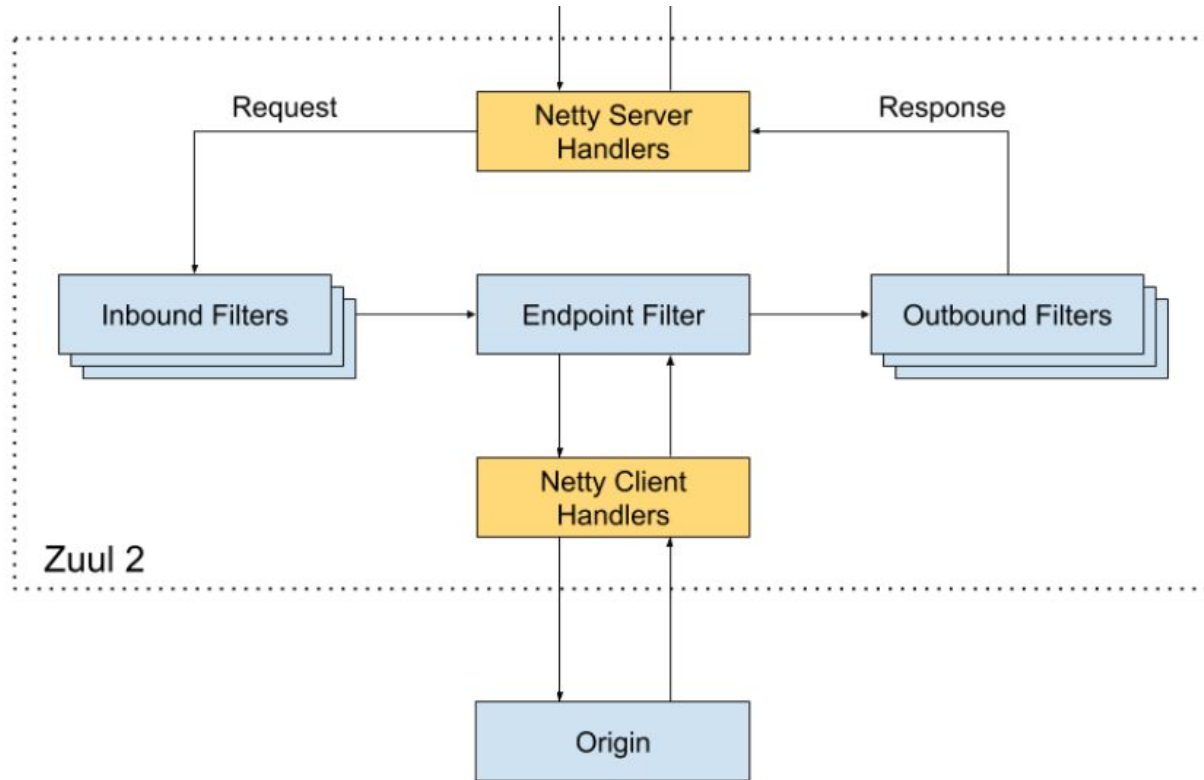# Playback Architecture - architecture about streaming videos
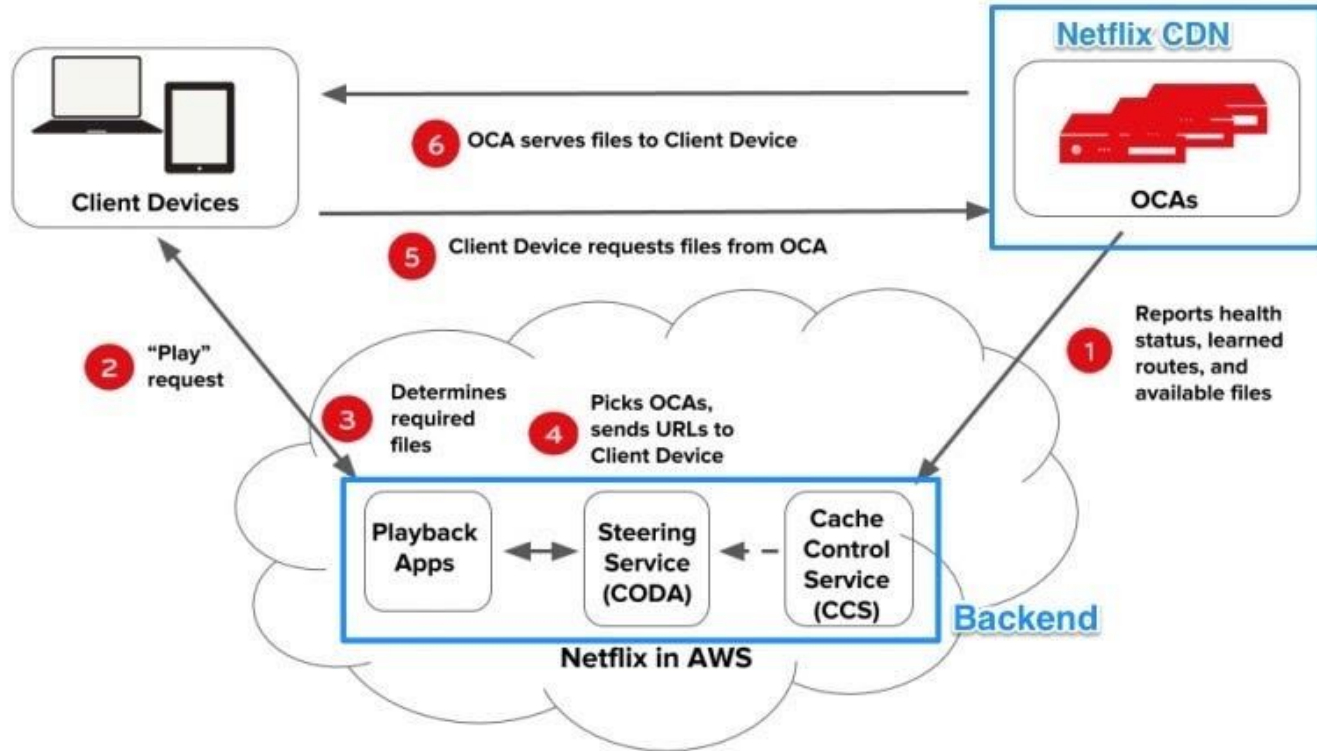
Backend Architecture
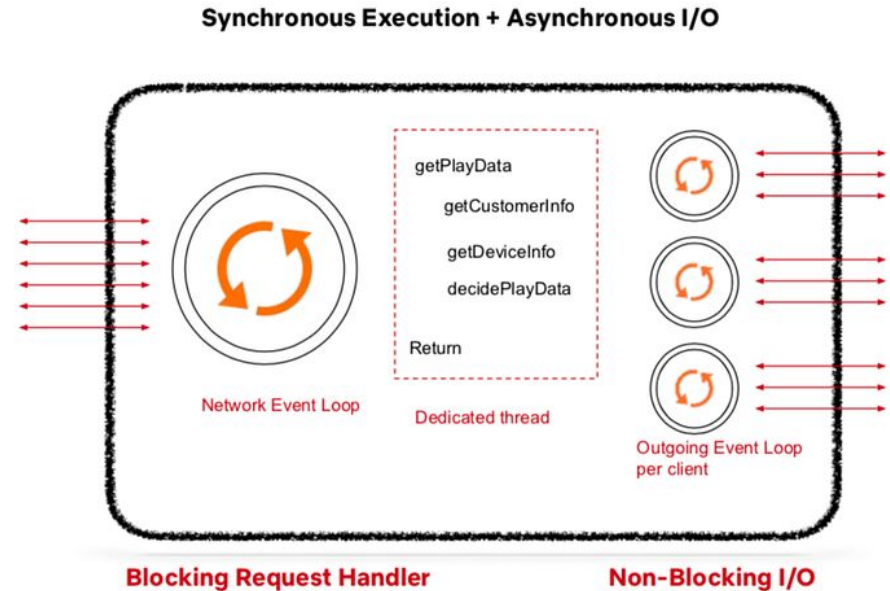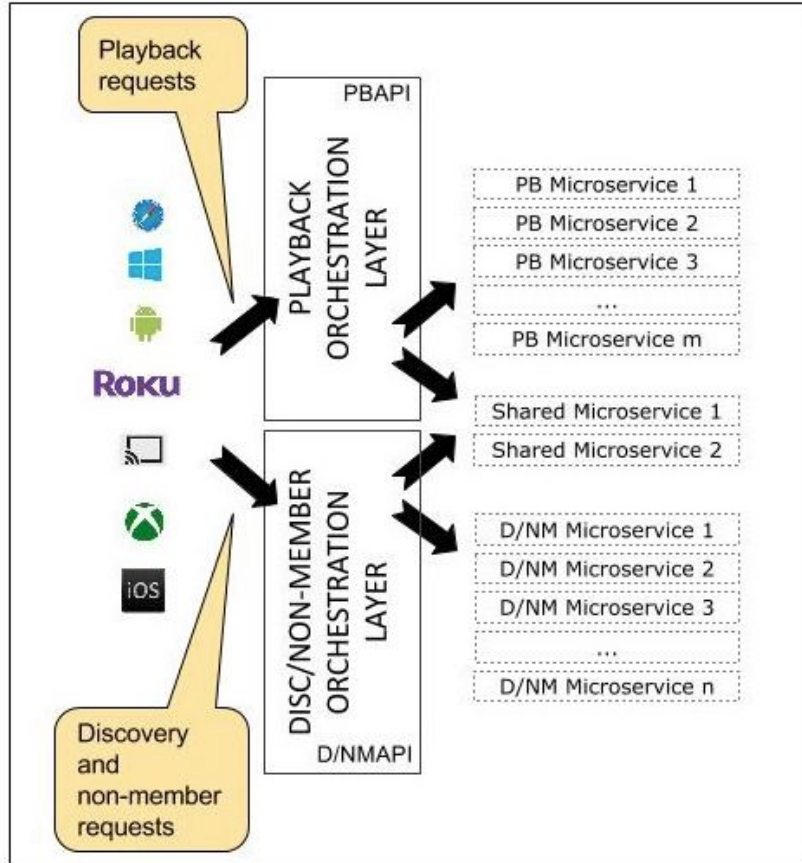
Backend Architecture - client
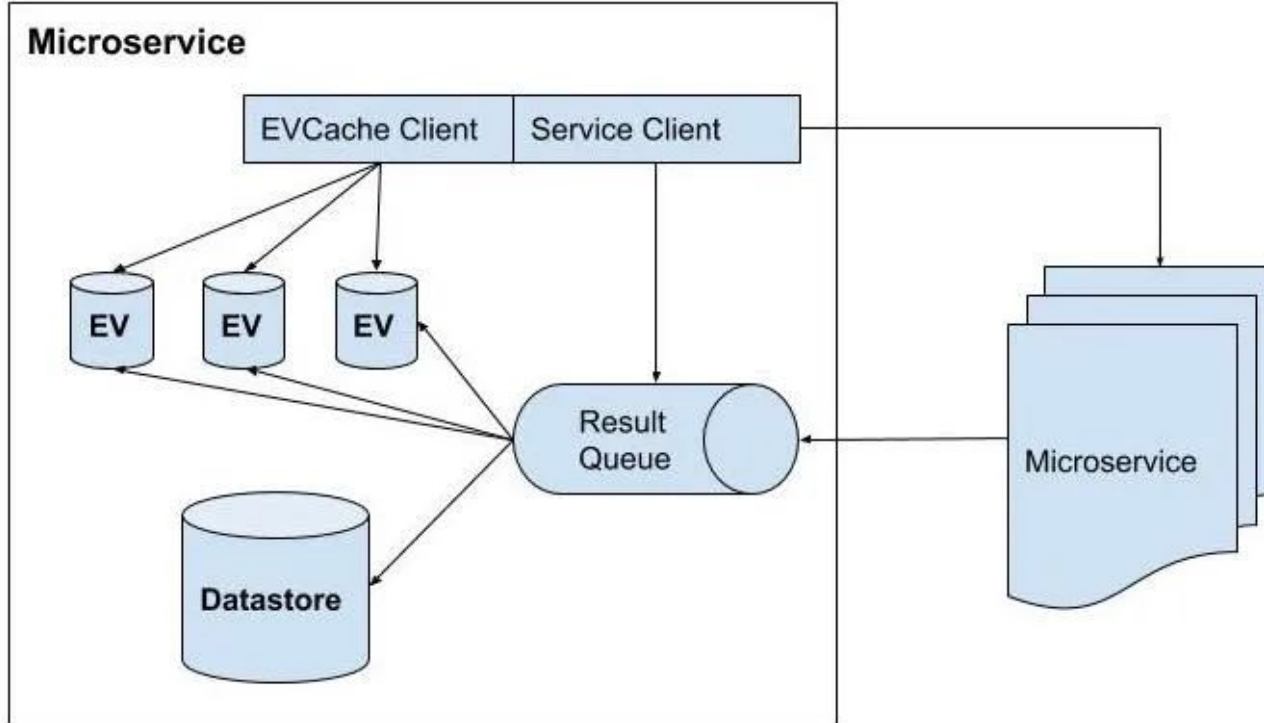
Backend Architecture - API Gateway Service

SD: Eureka
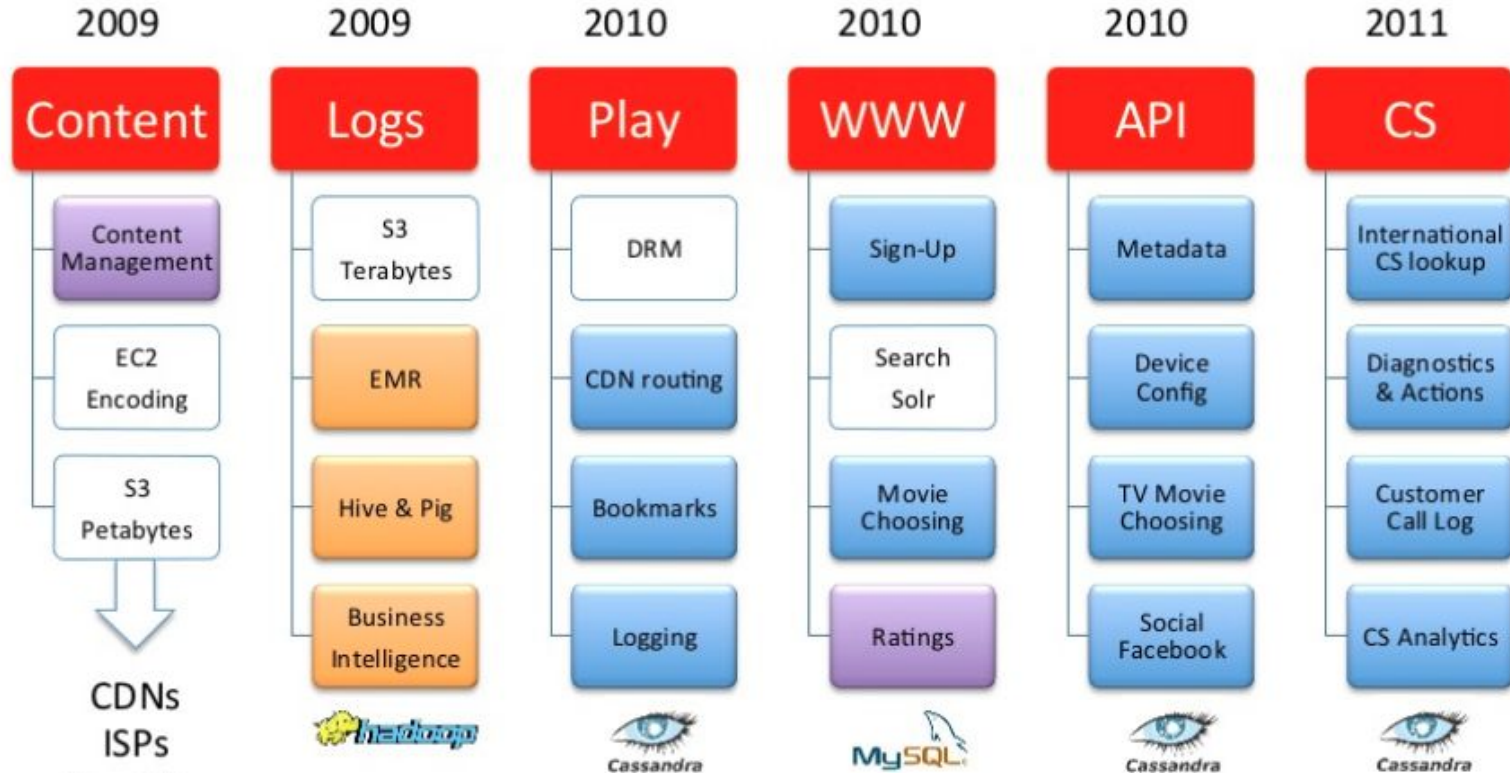
Backend Architecture - Application API
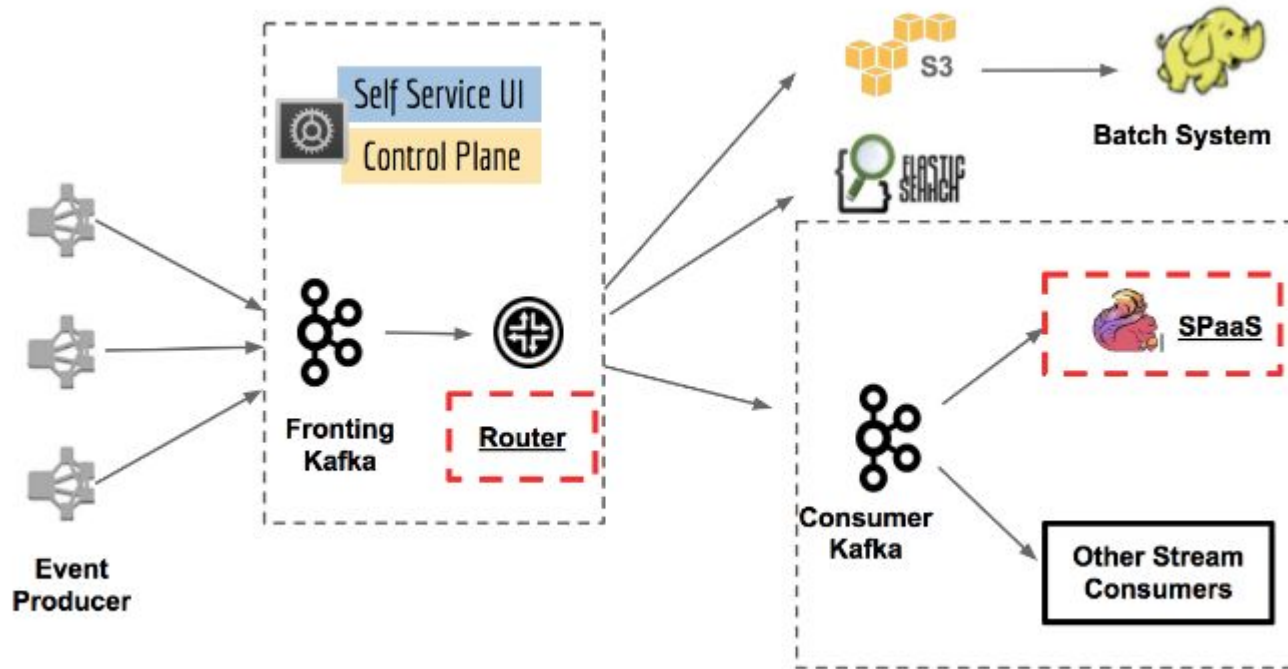
Backend Architecture - Microservice

EV Cache

Backend Architecture - Data stores

2009	2009	2010	2010	2010	2011
Content	**Logs**	**Play**	**WWW**	**API**	**CS**
Content Management	S3 Terabytes	DRM	Sign-Up	Metadata	International CS lookup
EC2 Encoding	EMR	CDN routing	Search Solr	Device Config	Diagnostics & Actions
S3 Petabytes	Hive & Pig	Bookmarks	Movie Choosing	TV Movie Choosing	Customer Call Log
CDNs ISPs	Business Intelligence	Logging	Ratings	Social Facebook	CS Analytics
	hadoop	Cassandra	MySQL	Cassandra	Cassandra

Backend Architecture - Stream Processing Pipeline

CDN - Open Connect Appliances

IX SITE

ISP SITE

Open Connect
Control Plane
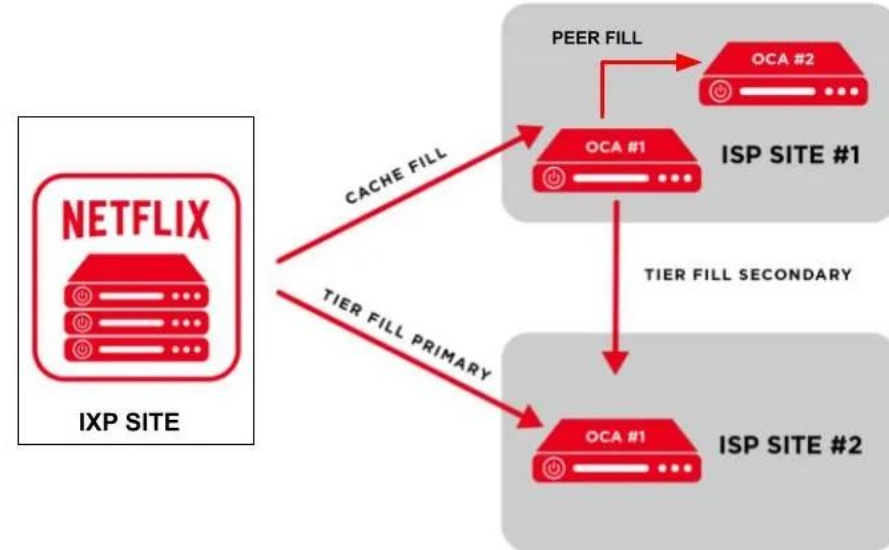on AWS
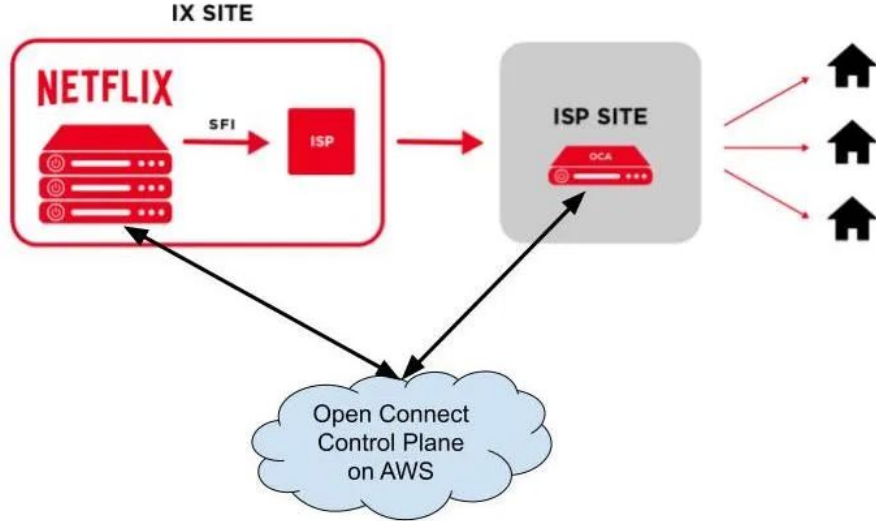
OCAs report health
metrics optimal routes

Cache Fill: from OCAs servers to ISP sites
Peer Fill: between OCAs servers in the same site;
Tier Fill: for other IP addresses

PEER FILL

OCA #2

OCA #1   **ISP SITE #1**

CACHE FILL

TIER FILL SECONDARY

NETFLIX

TIER FILL PRIMARY

IXP SITE

OCA #1   **ISP SITE #2**

Design Goals

- High Availability
- Low Latency
- Tradeoffs
  - Low latency over consistency
  - High availability over consistency
- Resilience
- Scalability

References

https://medium.com/swlh/a-design-analysis-of-cloud-based-microservices-architecture-at-netflix-98836b2da45f