Universitat de Barcelona

Arthur Font Gouveia 20222613

Eduard Vives Isern 20206970

**Bases de Datos Avanzadas**

Práctica 1

Barcelona

2021

# 1. Plataforma escogida: Netflix

# 2. Descripción del sistema de la base de datos

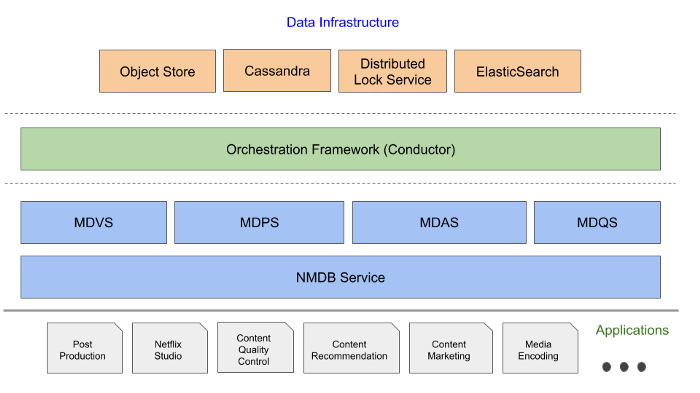
* Quines tecnologies i perquè han triat?

Netflix necesita de un sistema de base de datos que permita la reproducción de vídeo de alta resolución en tiempo real para miles de usuarios alrededor del mundo. A fin de obtener la escalabilidad, el rendimiento y la estabilidad de este sistema, Netflix ha diseñado el **Netflix Media DataBase (NMDB)**.

El NMBD és una base de datos consultable construida en la plataforma de microservicios de Netflix. Esta base permite guardar metadatos sobre archivos de media y atender consultas en tiempo real utilizando una combinación de búsquedas y recursos computacionales.

Los metadatos son estructurados en un documento JSON que permite representar metadatos estáticos así como dinámicos de los diversos tipos de media. Así, el sistema dispone de la capacidad de consultar características espacio-temporales. Un ejemplo sería: *“Get all events happening during given time interval(s) in given region(s)”*.

# 3. Infraestructura del sistema



# 

# 3.1. Tecnologías utilizadas

* Cassandra (NoSQL)
* MySQL (SQL)
* Object Store (Media)

Artigo 2: The Media Document Model

Flexible framework that can be used to represent static as well as dynamic (varying with time and space) metadata for various media modalities

NMDB is built to be a highly scalable, multi-tenant, media metadata system that can serve a high volume of write/read throughput as well as support near real-time queries.

Characteristics of a Media Database

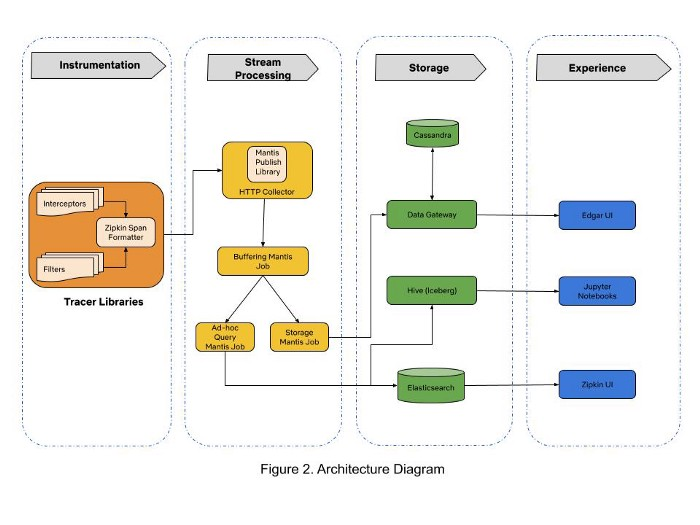
1. Affinity to structured data:
2. Efficient media timeline modeling
3. Spatio-temporal query-ability
4. Multi-tenancy
5. Scalability

NMDB as a “schema-on-write” system — data is validated against schema at the time of writing to NMDB

Artigo 4: high data capacity and high performance Cassandra (C\*) database as the backend implementation that serves as the source of truth for all our data

We started with Elasticsearch as our data store due to its flexible data model and querying capabilities. As we onboarded more streaming services, the trace data volume started increasing exponentially. The increased operational burden of scaling ElasticSearch clusters due to high data write rate became painful for us. The data read queries took an increasingly longer time to finish because ElasticSearch clusters were using heavy compute resources for creating indexes on ingested traces.

The high data ingestion rate eventually degraded both read and write operations. We solved this by migrating to [Cassandra](https://netflixtechblog.com/tagged/cassandra) as our data store for handling high data ingestion rates. Using simple lookup indices in Cassandra gives us the ability to maintain acceptable read latencies while doing heavy writes.

We use [Mantis](https://netflixtechblog.com/open-sourcing-mantis-a-platform-for-building-cost-effective-realtime-operations-focused-5b8ff387813a) for processing the stream of collected traces, and we use Cassandra for storing traces.

# 4. Referencias

<https://netflixtechblog.com/the-netflix-media-database-nmdb-9bf8e6d0944d>

<https://netflixtechblog.com/netflix-mediadatabase-media-timeline-data-model-4e657e6ffe93>

<https://netflixtechblog.com/implementing-the-netflix-media-database-53b5a840b42a>

<https://netflixtechblog.com/building-netflixs-distributed-tracing-infrastructure-bb856c319304>

The Netflix Media Database: <https://www.youtube.com/watch?v=OQK3E21BEn8&ab_channel=HPAonline>

NETFLIX system design: <https://medium.com/@narengowda/netflix-system-design-dbec30fede8d>

How netflix manages petabyte scale apache cassandra in the cloud: <https://es.slideshare.net/VinayKumarChella/how-netflix-manages-petabyte-scale-apache-cassandra-in-the-cloud>

Cassandra Serving Netflix:

<https://feathercast.apache.org/2019/09/12/how-netflix-manages-petabyte-scale-apache-cassandra-in-the-cloud-joey-lynch-vinay-chella/>

(min 28:00) <https://www.youtube.com/watch?v=psQzyFfsUGU&ab_channel=TechDummiesNarendraL>

(min 5:00) <https://www.youtube.com/watch?v=2l0_onmQsPI&ab_channel=TheLinuxFoundation>

Mantis:

<https://netflix.github.io/mantis/#:~:text=%C2%B6,bytes%20of%20data%20every%20day>.

<https://netflix.github.io/mantis/getting-started/use-cases/>

<https://netflixtechblog.com/stream-processing-with-mantis-78af913f51a6>

<https://netflixtechblog.com/open-sourcing-mantis-a-platform-for-building-cost-effective-realtime-operations-focused-5b8ff387813a>

MySQL Customer: Netflix: <https://www.mysql.com/customers/view/?id=1276>

Netflix Billing Migration to AWS: <https://netflixtechblog.com/netflix-billing-migration-to-aws-451fba085a4>

Migración de facturación de Netflix a AWS - Parte II: <https://netflixtechblog.com/netflix-billing-migration-to-aws-part-ii-834f6358126>