

The Netflix Media Database

February 13, 2019

Rohit Puri (rpuri@netflix.com),
Engineering Manager, Cloud Media Systems Team,
Content Engineering



Acknowledgments

- Arsen Kostenko (akostenko@netflix.com)
- Cyril Concolato (cconcolato@netflix.com)
- Shinjan Tiwary (stiwary@netflix.com)
- Sreeram Chakrovorthy (schakrovorthy@netflix.com)
- Subrahmanya Venkatrav (svenkatrav@netflix.com)
- Yi Guo (yig@netflix.com)

Presentation Outline

- Why a Media Database?
- The Netflix Media Database
- Data Model
- Implementation Overview

Presentation Outline

- Why a Media Database?
- The Netflix Media Database
- Data Model
- Implementation Overview

The Netflix Service Experience

- Optimized User Interface
- Personalized Content Recommendations
- Efficient Media Streaming
- Curated Content Catalog

Promotional Artwork and Video

The image displays various promotional materials for the Netflix series *Altered Carbon*, including:

- A large banner image showing a person in a suit floating in a green, glowing liquid environment.
- A mobile device screen showing the Netflix interface for the show, featuring a large image of a person in a suit and a summary text block.
- A desktop browser window showing the Netflix interface for the show, featuring a large image of a person in a suit and a summary text block.
- A Netflix landing page for the show, featuring a large image of a person in a suit and a summary text block.
- A Netflix promotional poster for Season 1, showing a person in a suit inside a clear plastic bag.
- A grid of smaller promotional images for the show, including scenes of people in futuristic settings and characters.
- A detailed promotional image of a woman with braided hair looking intensely at the viewer.
- A promotional image of a man in a white shirt holding a small device.
- A promotional image of a shirtless man standing in a dark, industrial-looking setting.
- A horizontal row of promotional images for other Netflix shows: *Unbreakable Kimmy Schmidt*, *Chef's Table*, *Altered Carbon*, and *Altered Carbon*.

Content Recommendations

NETFLIX Browse ▾ DVD Search Joshua

Top Picks for Joshua

Trending Now

Because you watched Narcos

New Releases

NETFLIX Beyond STRANGER THINGS Moana THE MIST THE BABYSITTER RIVERDALE NETFLIX Doctor STRANGE Hater Back off NEW EP

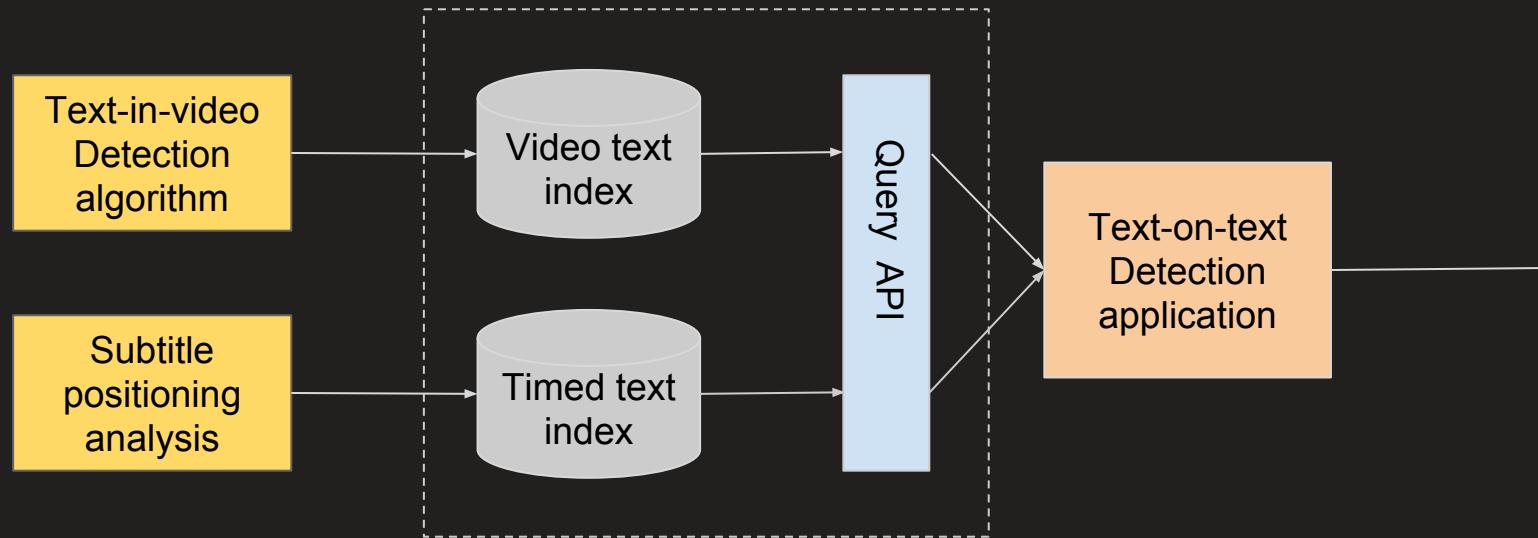
Content Quality



Detecting Text on Text (Problem)



Detecting Text on Text (Solution)



Modular analysis and persistence promotes data re-use

Presentation Outline

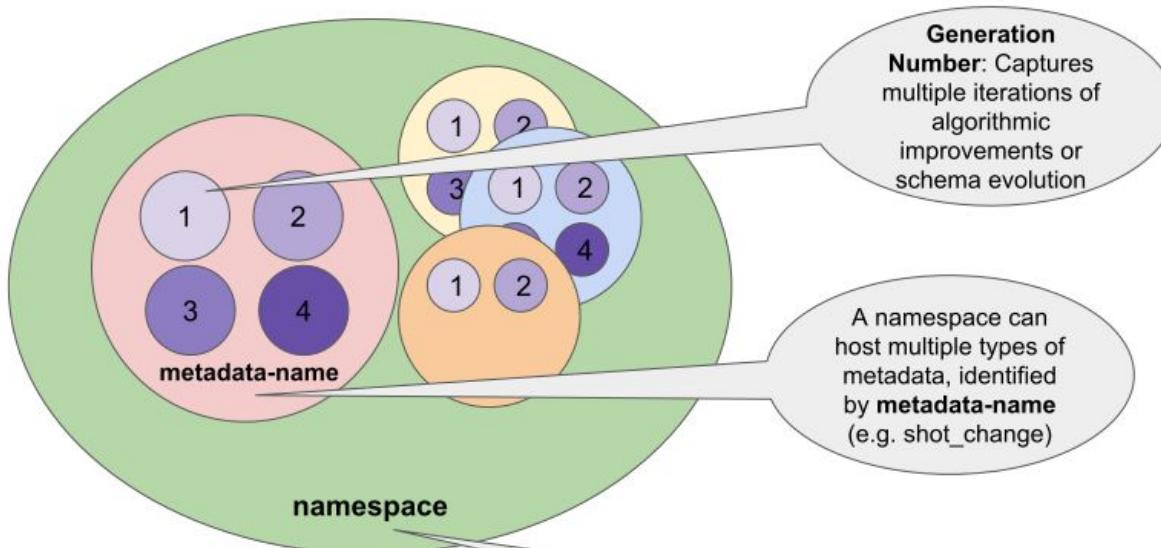
- Why a Media Database?
- **The Netflix Media Database**
- Data Model
- Implementation Overview

The Netflix Media Database

- A multi-tenant data store for dynamic (temporally and spatially varying) metadata for Netflix media assets
- Supported asset modalities include audio, video, text, images
- Answers spatio-temporal queries on the media timeline (at scale)

User View

DataStore: namespace + metadata-name + generation-number [+sandboxId]



Generation Number: Captures multiple iterations of algorithmic improvements or schema evolution

A namespace can host multiple types of metadata, identified by **metadata-name** (e.g. shot_change)

A restricted GG (e.g. cloud-media-systems, video-algorithms)

Some User Features

- Authentication and authorization
- Content vault (e.g., firewalls for pre-release content)
- “DataStore” based access control
- High read throughput

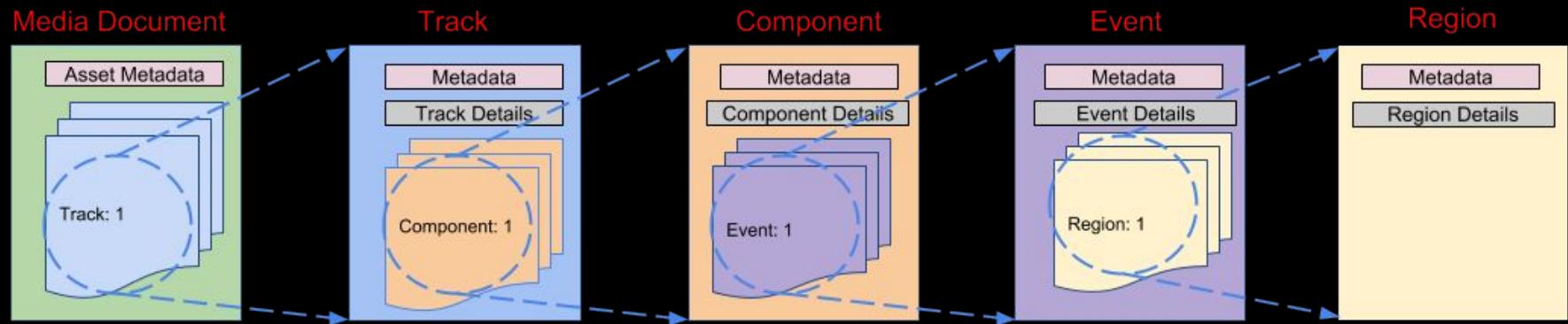
Data Schema and Modeling

- Schema oriented - every DataStore has a schema
- Schema-on-write
- Schema updates can be applied on the fly
- Capable of representing spatial (3D) and temporal metadata
- Integer units for temporal and spatial dimensions enable full precision

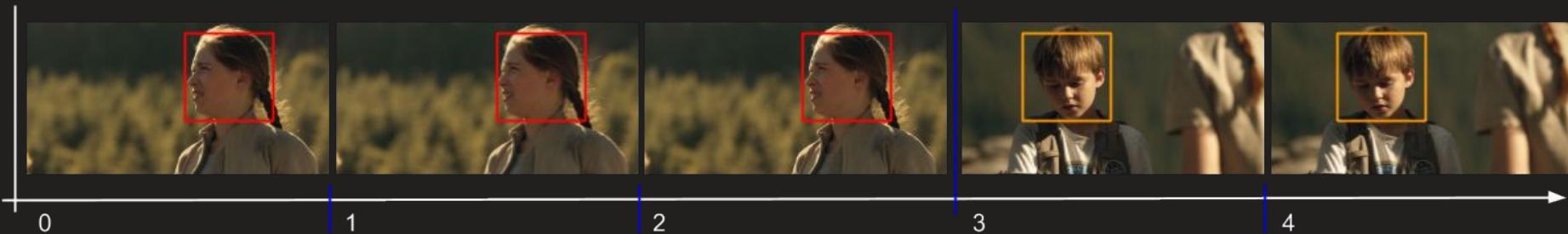
Presentation Outline

- Why a Media Database?
- The Netflix Media Database
- **Data Model**
- Implementation Overview

Media Timeline Data Model



A Media Document Instance

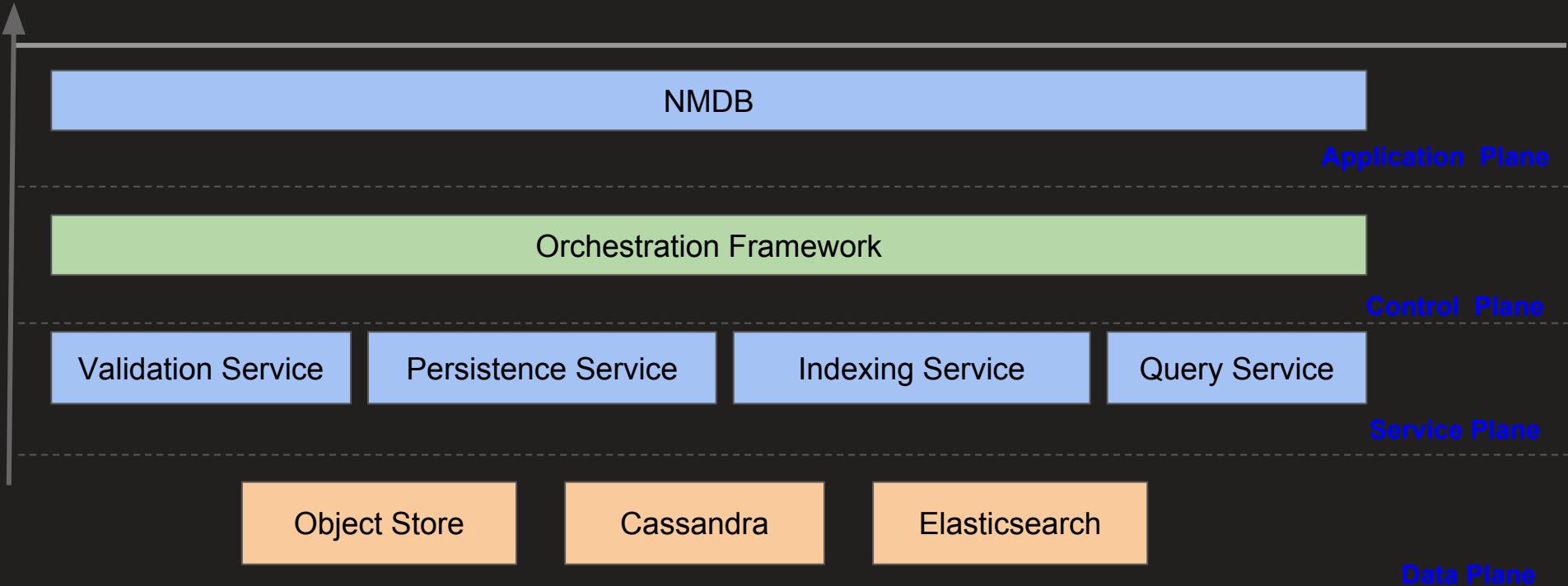


```
{"metadata": {"algorithm": "video_face_detection"},  
 "tracks": [{"id": "0",  
 "components": [{"id": "0",  
 "eventRateNumerator": 24000, "eventRateDenominator": 1001,  
 "xSize": 1920, "ySize": 1080,  
 "events": [  
     {"startTime": 0, "endTime": 2,  
     "regions": [{"xmin": 1152, "xmax": 1536, "ymin": 108, "ymax": 648}]},  
     {"startTime": 3, "endTime": 4,  
     "regions": [{"xmin": 576, "xmax": 960, "ymin": 108, "ymax": 648}]}]}]}]}
```

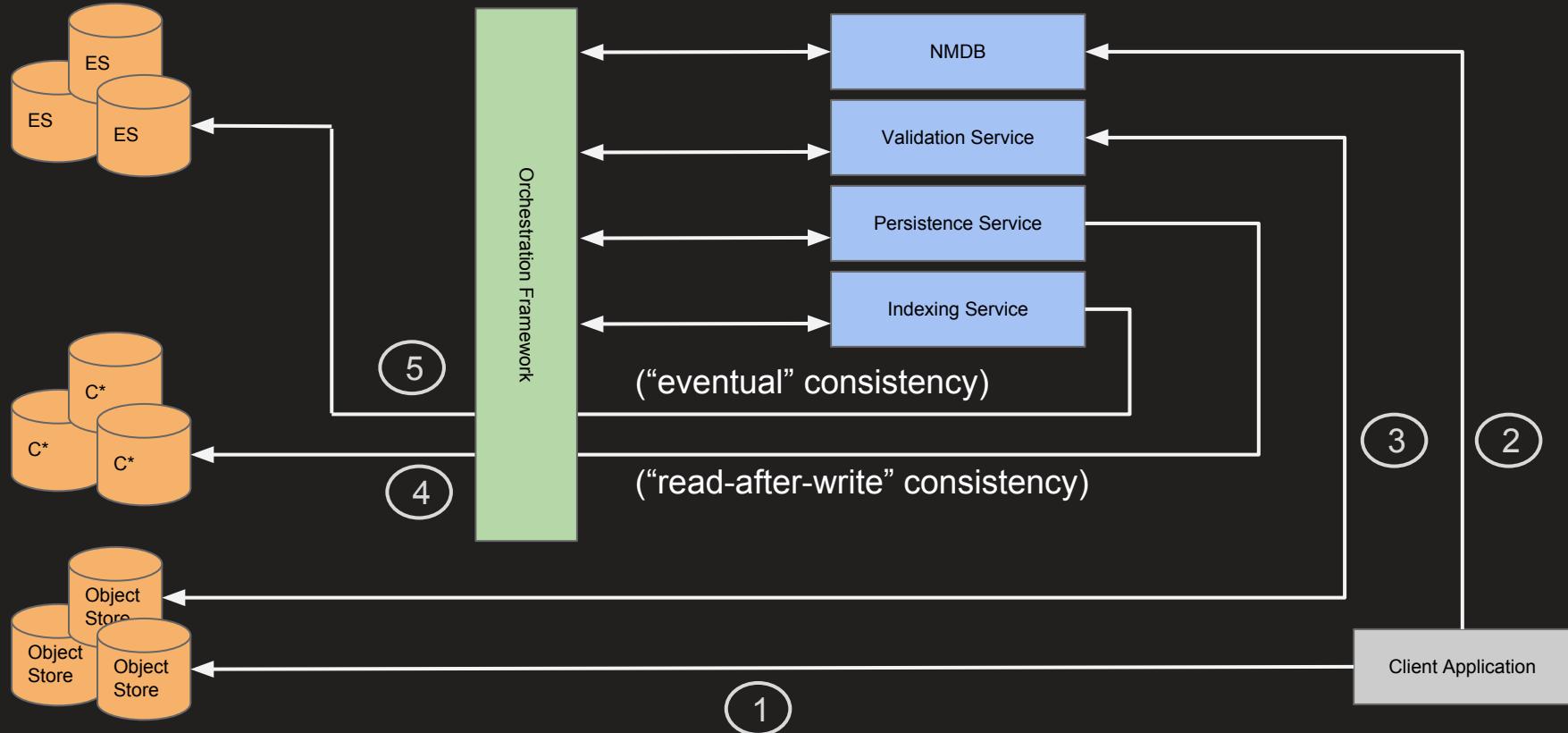
Presentation Outline

- Why a Media Database?
- The Netflix Media Database
- Data Model
- **Implementation Overview**

Architecture View



Writing Data



Challenges and Future Work

- Scaling (hierarchical) Media Document model
- Large (> 100 MB) Media Document Instances
- Complex analytical queries
- Best performance for a given cost

Links

- [AVA: The Art and Science of Image Discovery at Netflix](#)
- [The Netflix Media Database](#)
- [Netflix Media Database - the Media Timeline Data Model](#)
- [Implementing the Netflix Media Database](#)