

The Netflix Media Database

February 13, 2019

Rohit Puri (rpuri@netflix.com),

Engineering Manager, Cloud Media Systems Team,
Content Engineering

NETFLIX

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

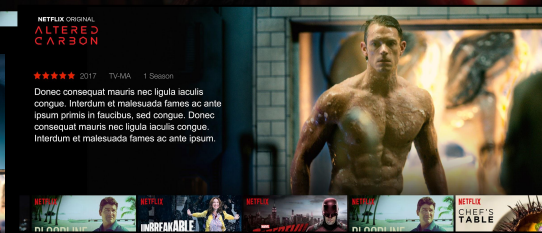
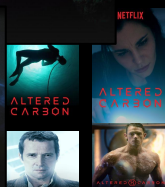
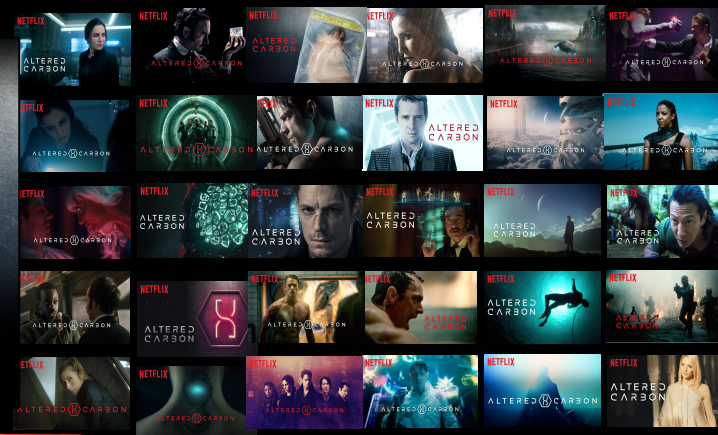
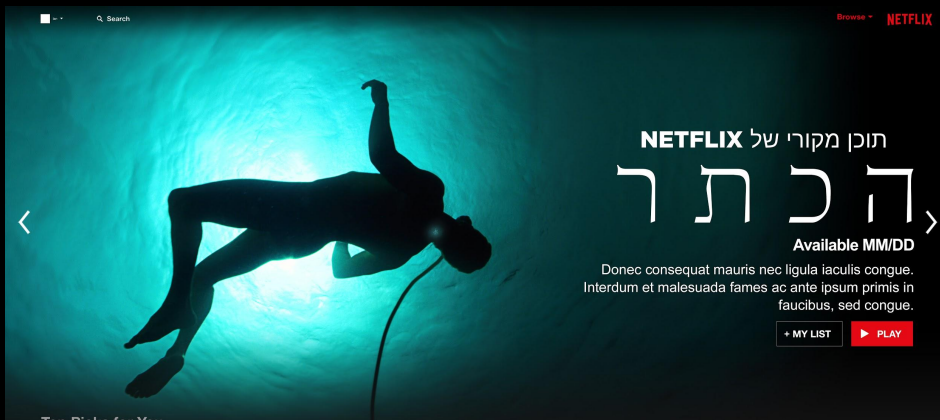
Presentation Outline

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

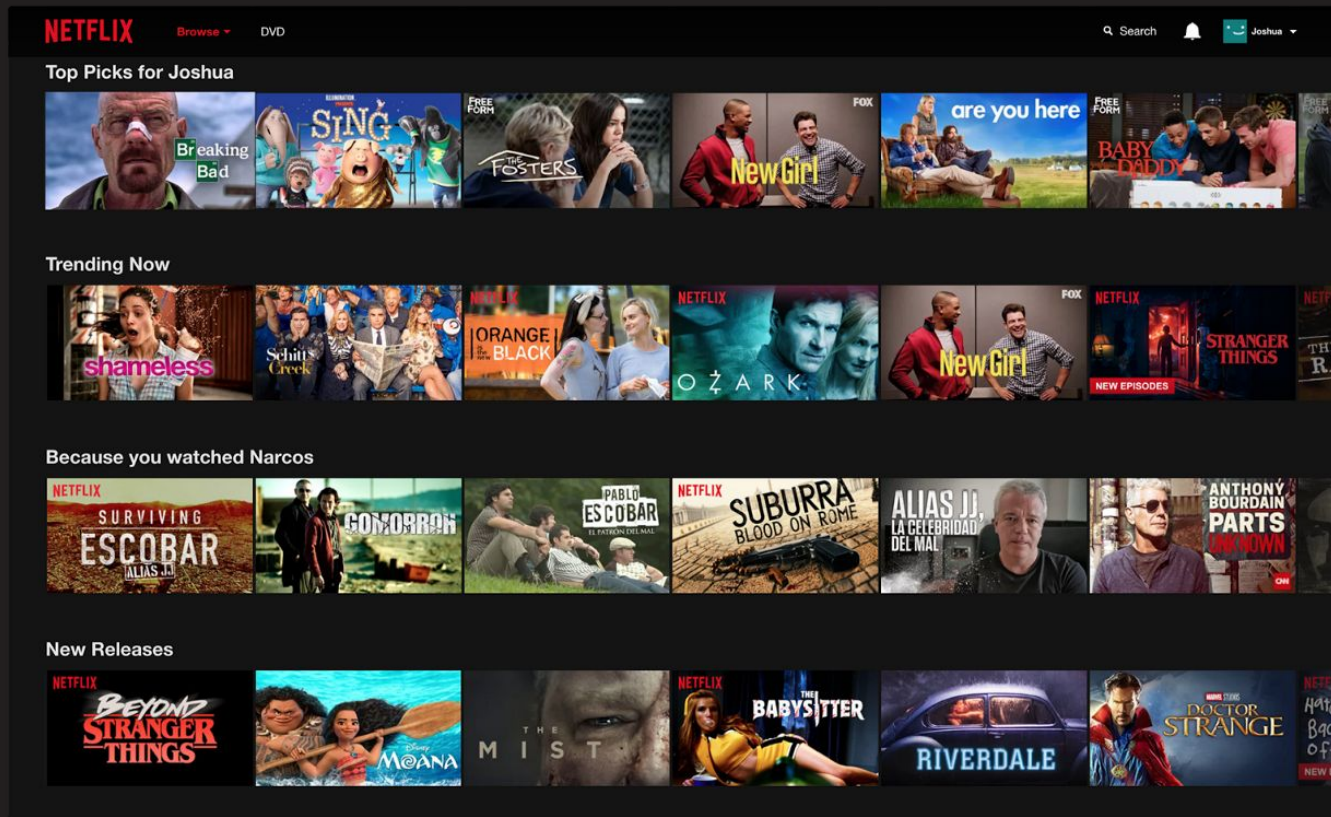
The Netflix Service Experience

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

Promotional Artwork and Video



Content Recommendations



Content Quality



Detecting Text on Text (Problem)

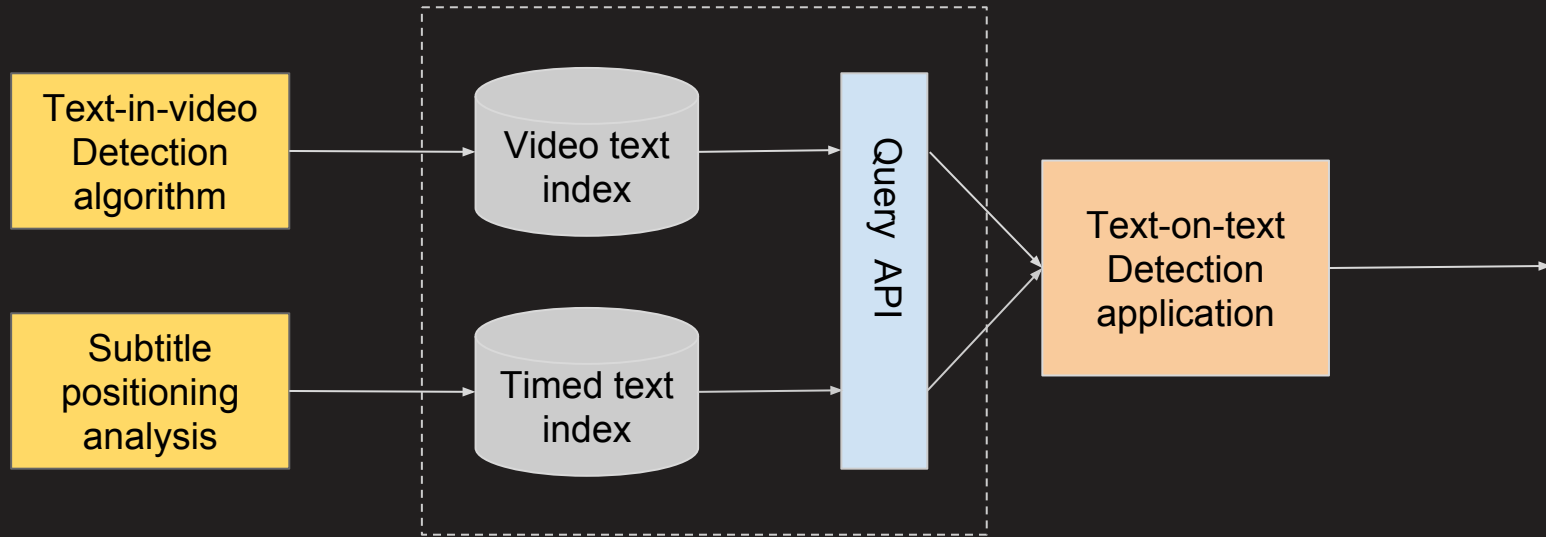


The Problem

Now remember, you're buying
for this fantastic neighborhood.

Bryan Cranston

Detecting Text on Text (Solution)



Modular analysis and persistence promotes data re-use

Presentation Outline

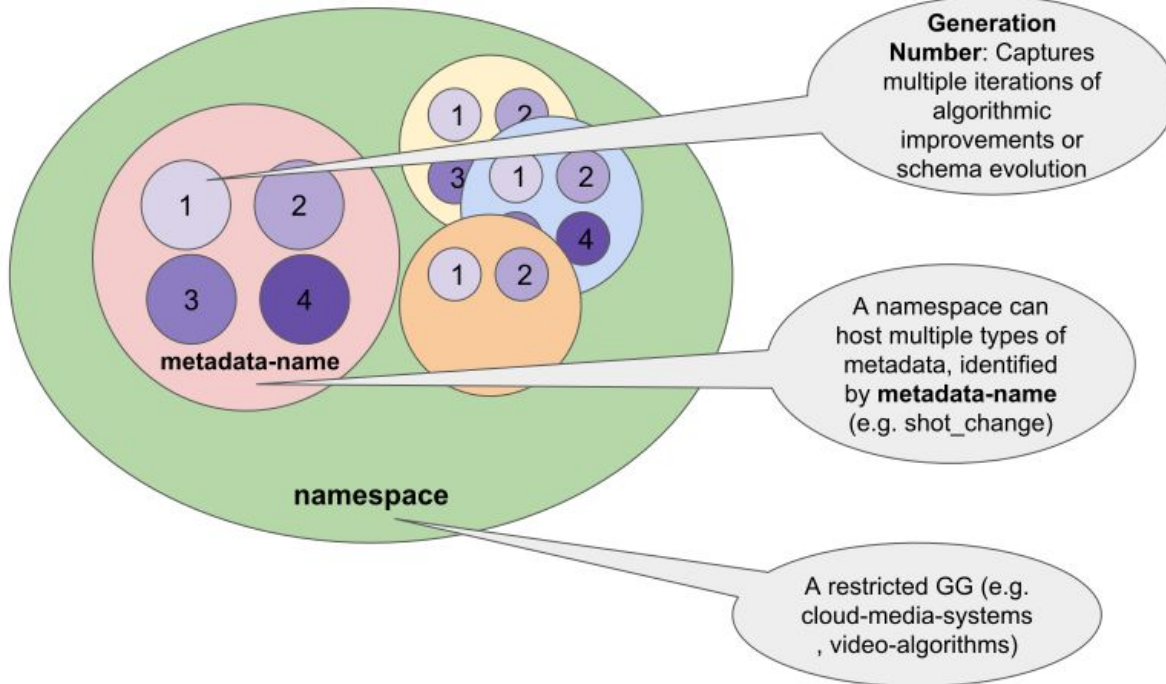
- Why a Media Database?
- **The Netflix Media Database**
- Data Model
- Implementation Overview

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]



Some User Features

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

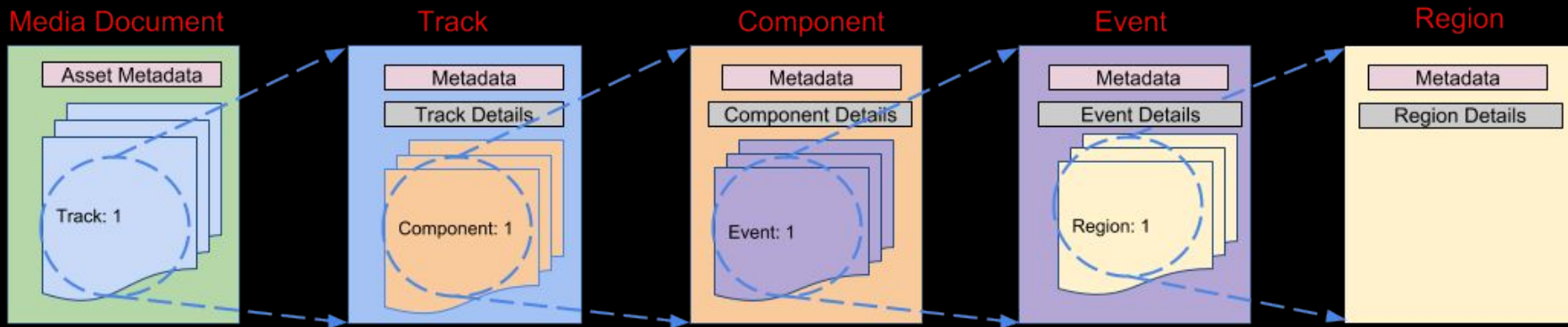
Data Schema and Modeling

- 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

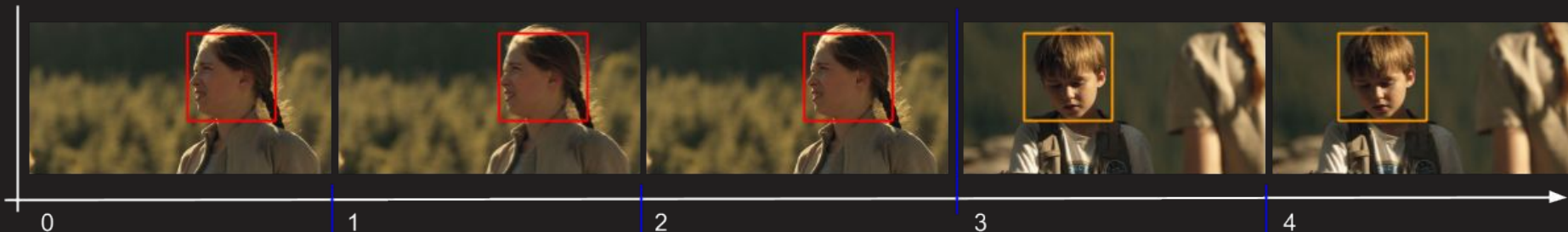
Presentation Outline

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

Media Timeline Data Model



A Media Document Instance

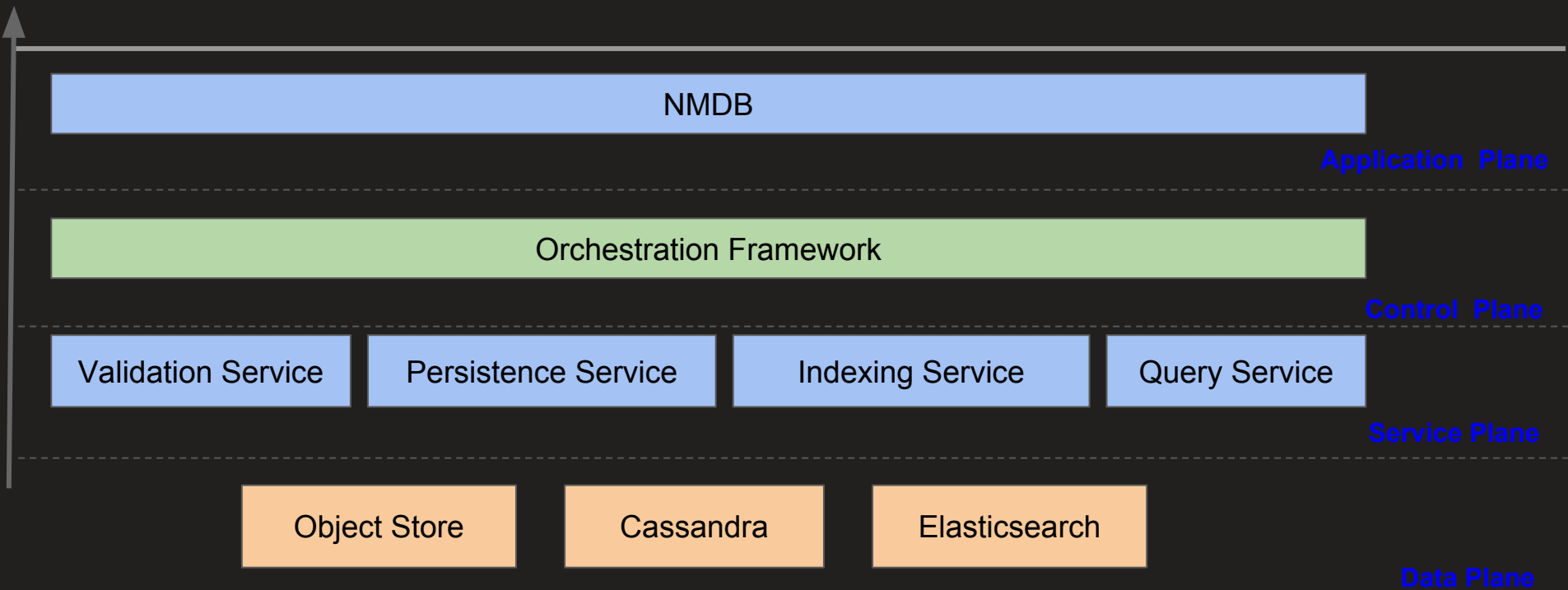


```
{"metadata":{"algorithm": "video_face_detection"},  
"tracks": [{"id": "0",  
  "components": [{"id": "0",  
    "eventRateNumerator": 24000, "eventRateDenominator": 1001,  
    "xSize": 1920, "ySize": 1080,  
    "events": [  
      {"startTime": 0, "endTime": 2,  
        "regions": [{"xmin": 1152,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

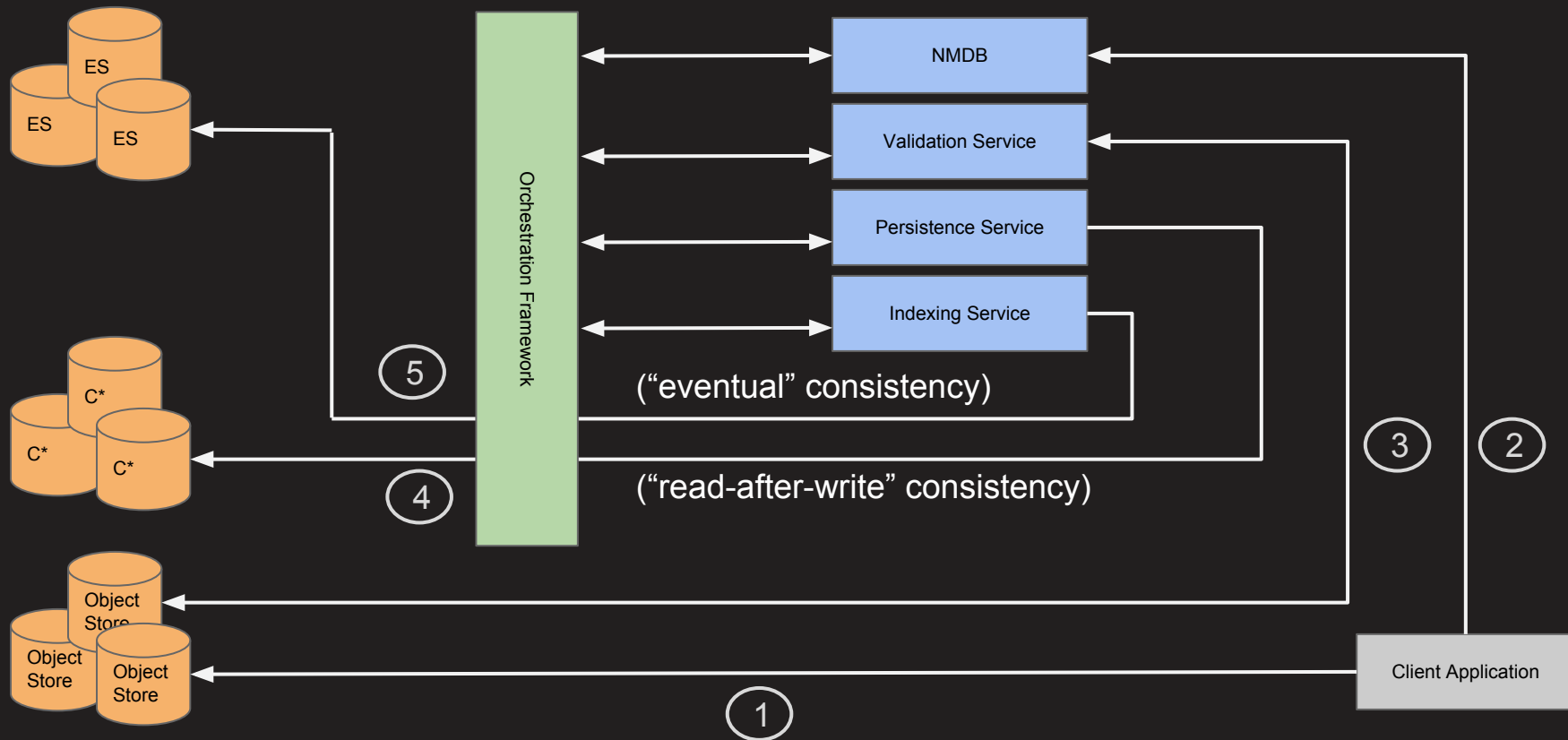
Presentation Outline

- 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](#)