

Changing the Way Viacom Looks at Video Performance

Mark Cohen
VP, Engineering, VAMP
@markcohenn
mark.cohen@viacom.com

Mike Rosencrantz
System Owner, Reporting
@mrosencrantz
mike@mtv.com

<http://blog.viacom.tech>

Agenda

- Viacom Intro
- Realtime Data Proof of Concept
- Video Performance Metrics
- Lessons Learned
- Future
- Questions

Have You Heard of Viacom?



A VIACOM COMPANY



COMEDY
CENTRAL

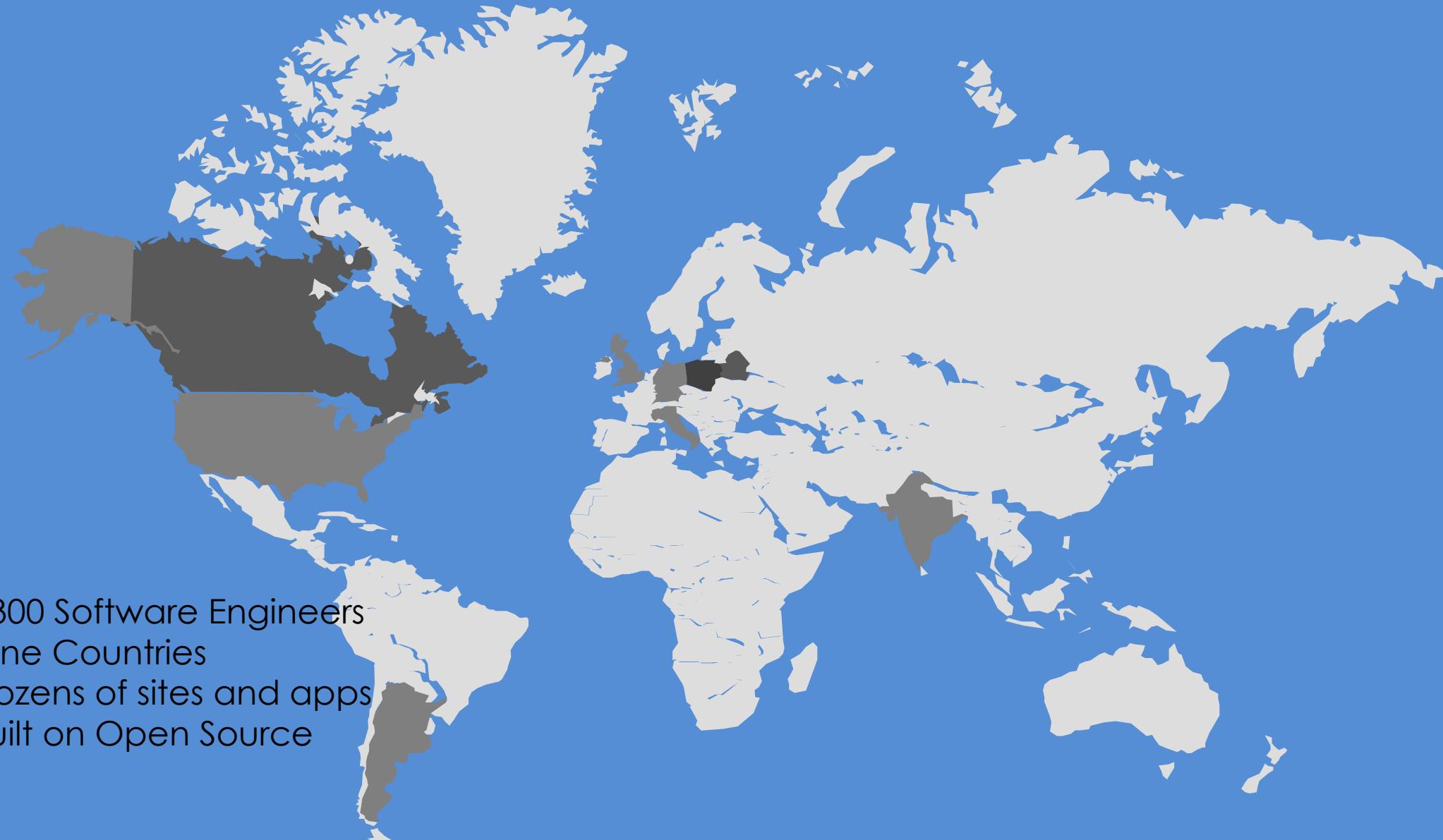


A VIACOM COMPANY

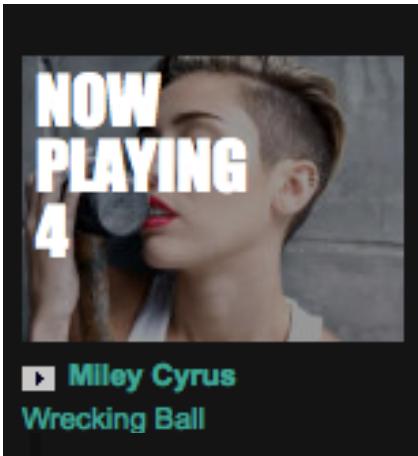


RATE MY PROFESSORS

Viacom Software Engineering



Recommendations



- People who clicked on recommendations:
 - 156% more page views
 - 69% more video than non-recommendation clickers
- 46% higher click-thru vs vendor recommendations

The screenshot shows a 'RECOMMENDED FOR YOU' section with four recommended tracks:

- Eminem** - Berzerk (Explicit)
- Katy Perry** - Roar
- 2 Chainz** - Where You Been ft. Cap 1 (Explicit)
- DJ Khaled** - I Wanna Be With You ft. Nicki Minaj, Future, Rick Ross

At the top right, there is a '1 of 2' indicator with left and right arrows.

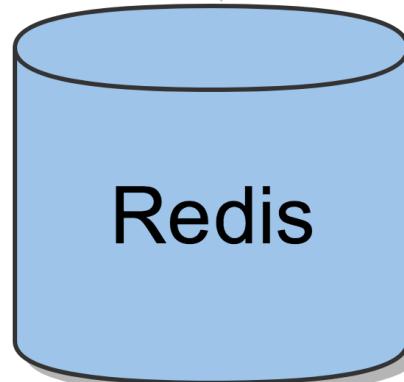
Nginx/DropWizard



Kafka



Apache Storm



Proof of Concept Infrastructure

Top Videos:

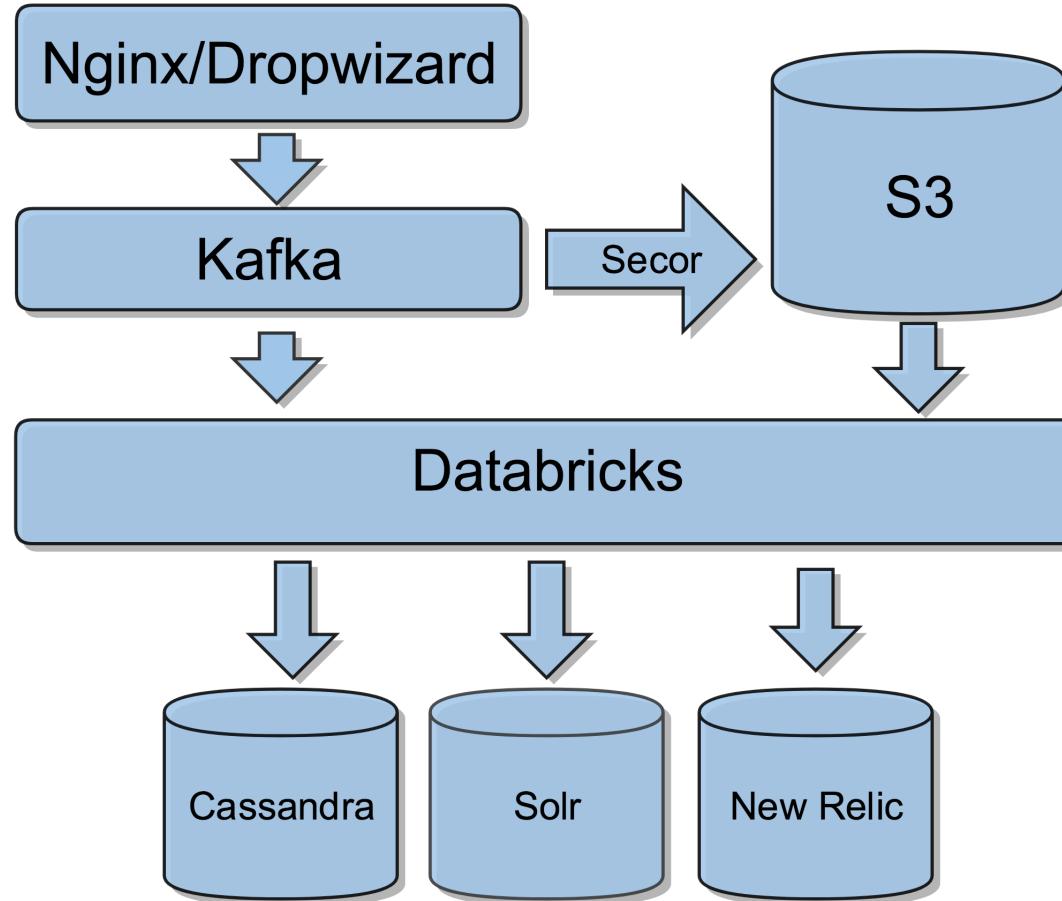
- * 30 seconds
- * 5 minutes
- * 30 minutes
- * 1 hour

Introducing Spark

- Engineer working with Spark started PoC with Databricks
- Spark SQL made it easy for analysts and engineers
 - 170 users in Databricks, many are not engineers
- Easy to Manage Spark Clusters
- Player team got access, started sending data



Current Infrastructure

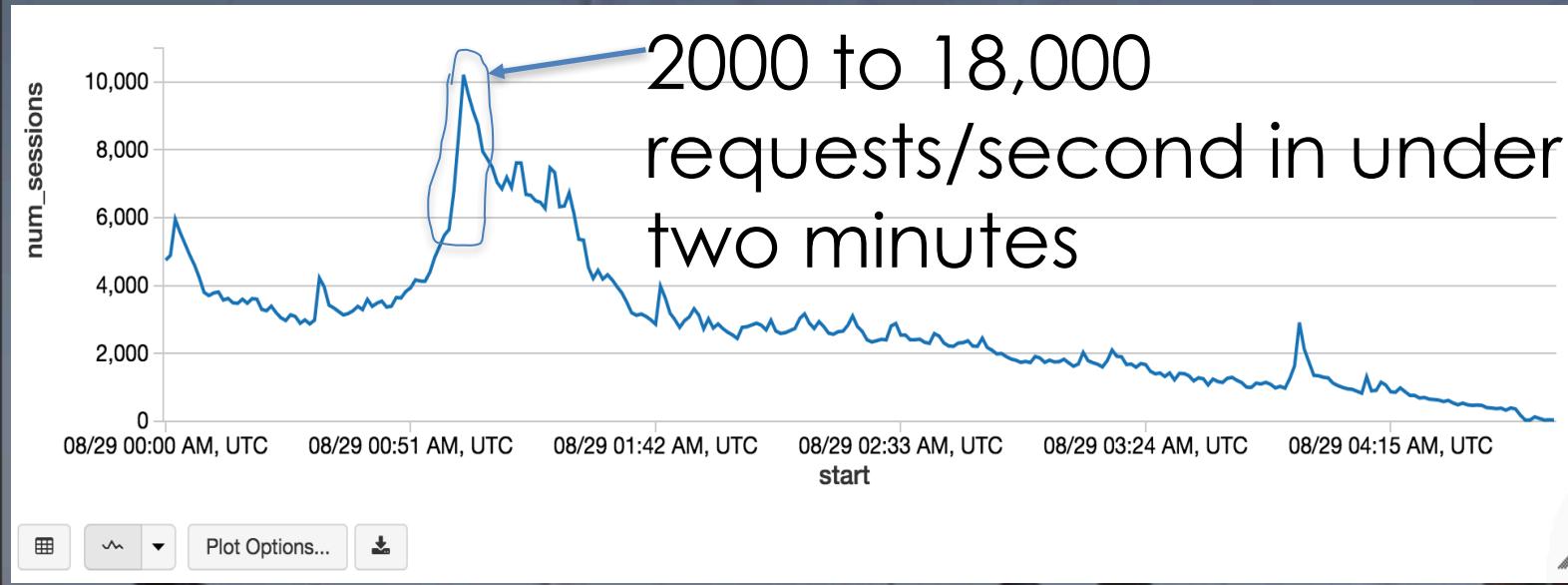


Data Statistics

- Volume
 - 1.25 TB/Month and growing
- Velocity
 - 2000-3000 messages/sec average
 - 18k-20k messages/sec peak
- Variety
 - JSON in Parquet files loaded into Hive tables of four types from Kafka



Pressure for VMAs

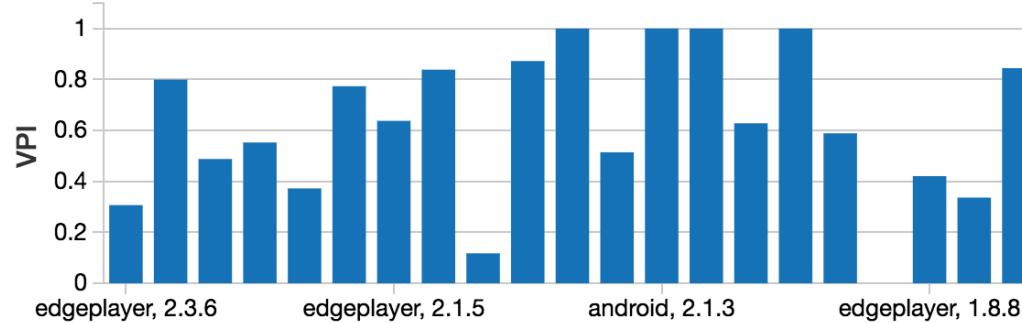


Video Performance Metrics

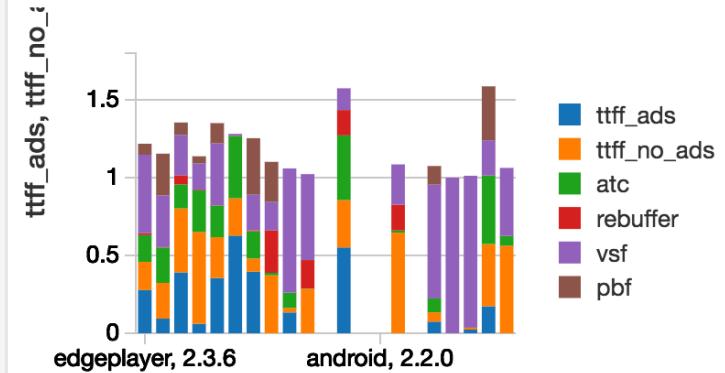
- Beacons from iOS, Android, Web, Roku
 - Errors
 - Playback beacons every 10 seconds
 - Video Start Time
 - Ad to Content

Video Performance Index

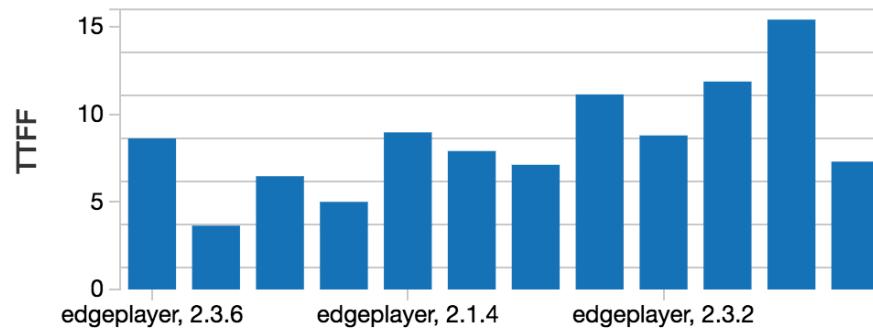
Video Performance Index (VPI, %)



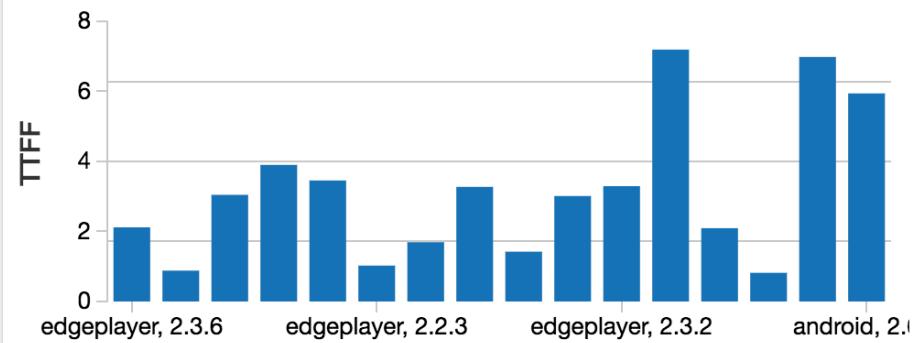
VPI Breakdown



Time to First Frame (TTFF, seconds)

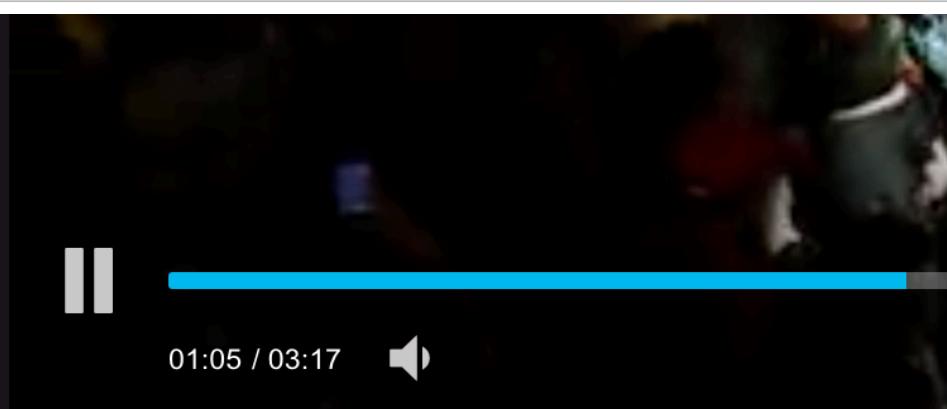
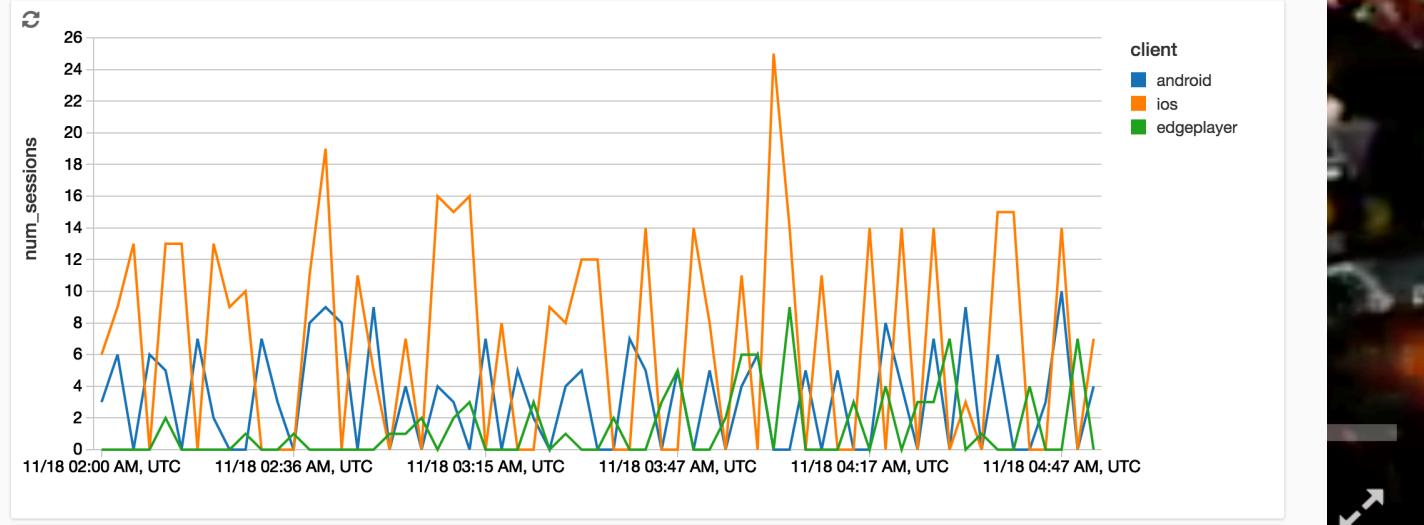
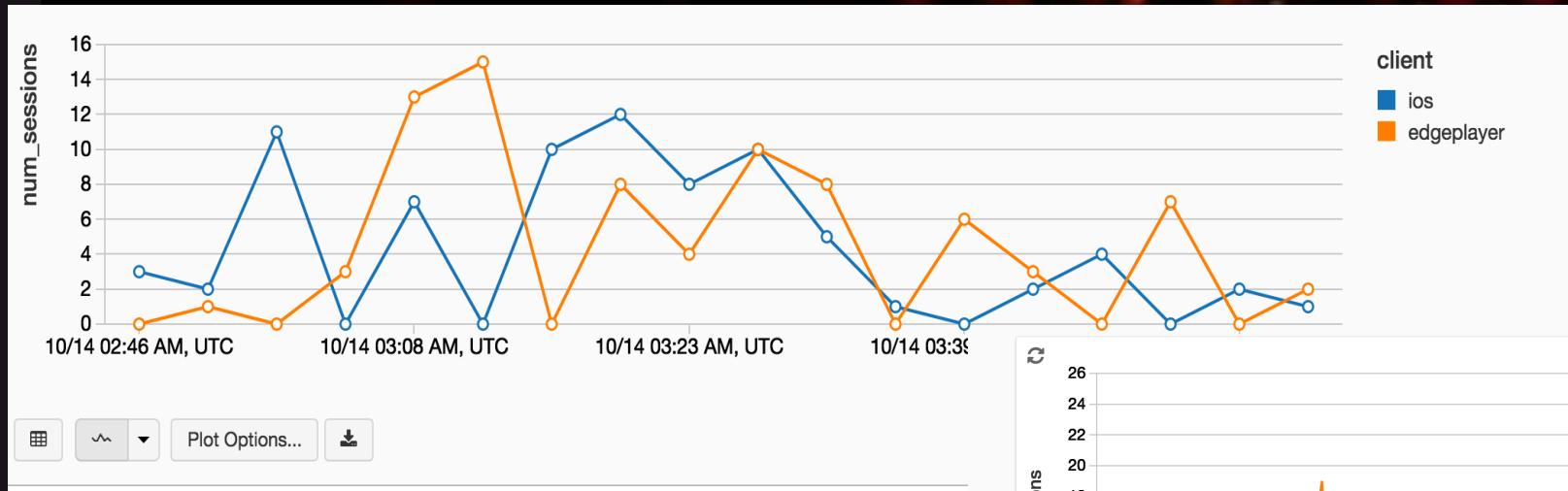


Ad-Free Time to First Frame



Test With Something Small

Wonderland: Backstage Cameras Streaming Online



MTV Movie & TV Awards



Lessons Learned

- Discover Your Aggregations
- UDFs are awesome
- Data Comparison
 - You WILL have to compare pre-existing data to your data
 - Sessions

```
1 %scala
2 sqlContext.udf.register("getBrandGroupByMgid", (mgid: String) => {
3   try{
4     var site = mgid.split(":")(3);
5     if(brandMap.contains(site)) brandMap(site) else "NotConfiguredMGID"
6   }catch {
7     case e: Exception => println("Exception occured in getBrandGroupByMgid UDF "+ e)
8     "Invalid mgid"
9   }
10 })
```

res13: org.apache.spark.sql.expressions.UserDefinedFunction = UserDefinedFunction(<function1>,Str

Lessons Learned

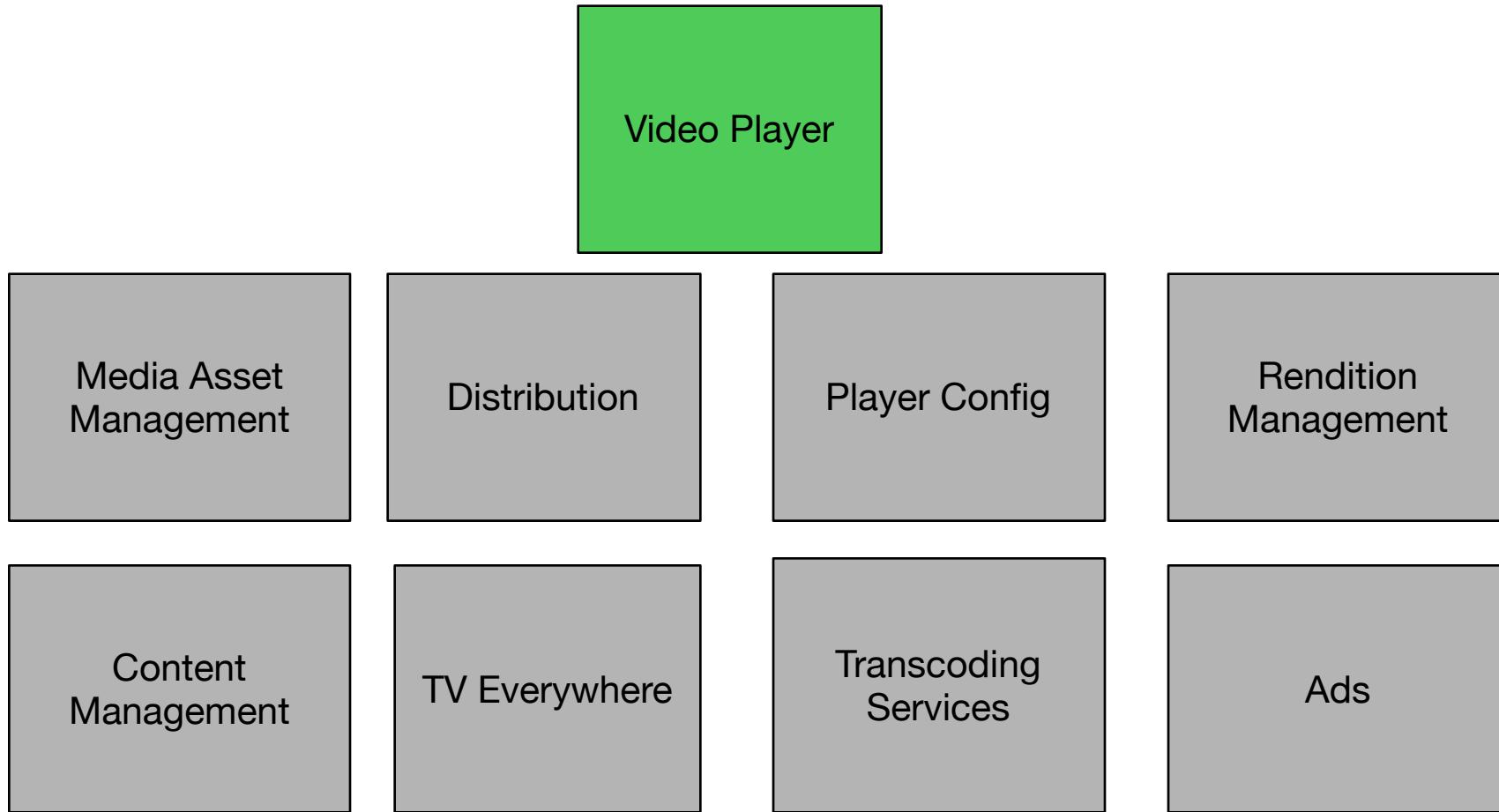
- Producer/Consumer
 - We thought “they” were going to do the reporting
 - Sure, put anything in Kafka. We don’t care.
- Cost Savings
 - Use the right kind of instances on EC2 (reserved/on-demand/spot)
 - Preprocess data for reporting as soon as possible

Lessons Learned

- Your data should go where the users want it



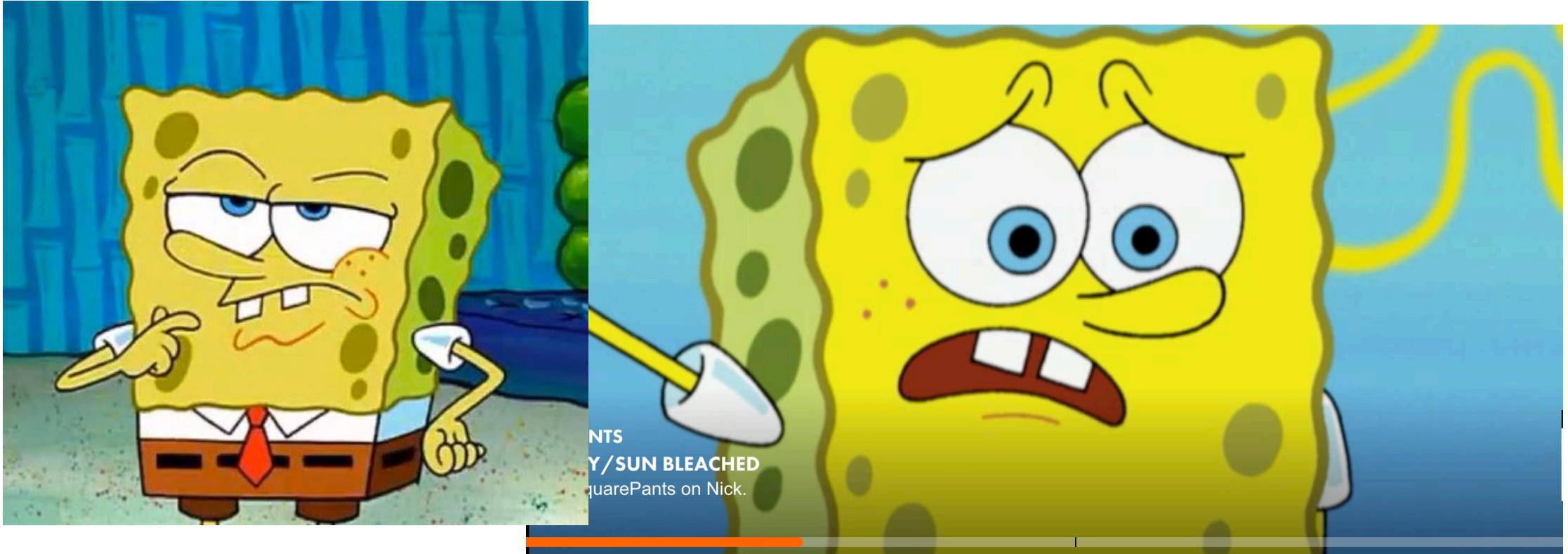
We Are Not Done



Future

- Trending Dimensions
 - like trending topics, but for metrics
- Always exploring self-service options for higher level reporting (e.g. ELK, AWS QuickSight + Athena, Druid, Kylin)
- SparkML
 - Scene Detection & Image Classification
 - Finding Root Cause of Video Problems Faster
- Continuous Delivery into Databricks

Questions?



Mark Cohen
@markcohenn
mark.cohen@viacom.com

<http://blog.viacom.tech>

Mike Rosencrantz
@mrosencrantz
mike@mtv.com