Building a Just-In-Time Data Warehouse

Dan Morris (Viacom)
Jason Pohl (Databricks)
February 18, 2016

About Viacom

- Leading Global Entertainment Content Company
- 23 Brands in 170+ Countries

Introductions

Dan Morris

- Senior Director of Product Analytics
- 12 Years @Viacom in a variety of roles
- Intersection of Product and Data

databricks™ VIACOM

About My Team

- Product Analytics team formed one year ago

- Our mission is to grow our **global** audience with the **highest quality** users possible

databricks™ VIACOM

Key Areas of Focus

- **Mobilize** efforts using growth targets

- Uncover deep **insights** using churn and cohort analysis

- Treat all ideas as **hypotheses** and test them rigorously

databricks VIACOM

Where Are We Today: App Platform

Make it *extremely simple* to build and deploy *engaging* apps around the globe

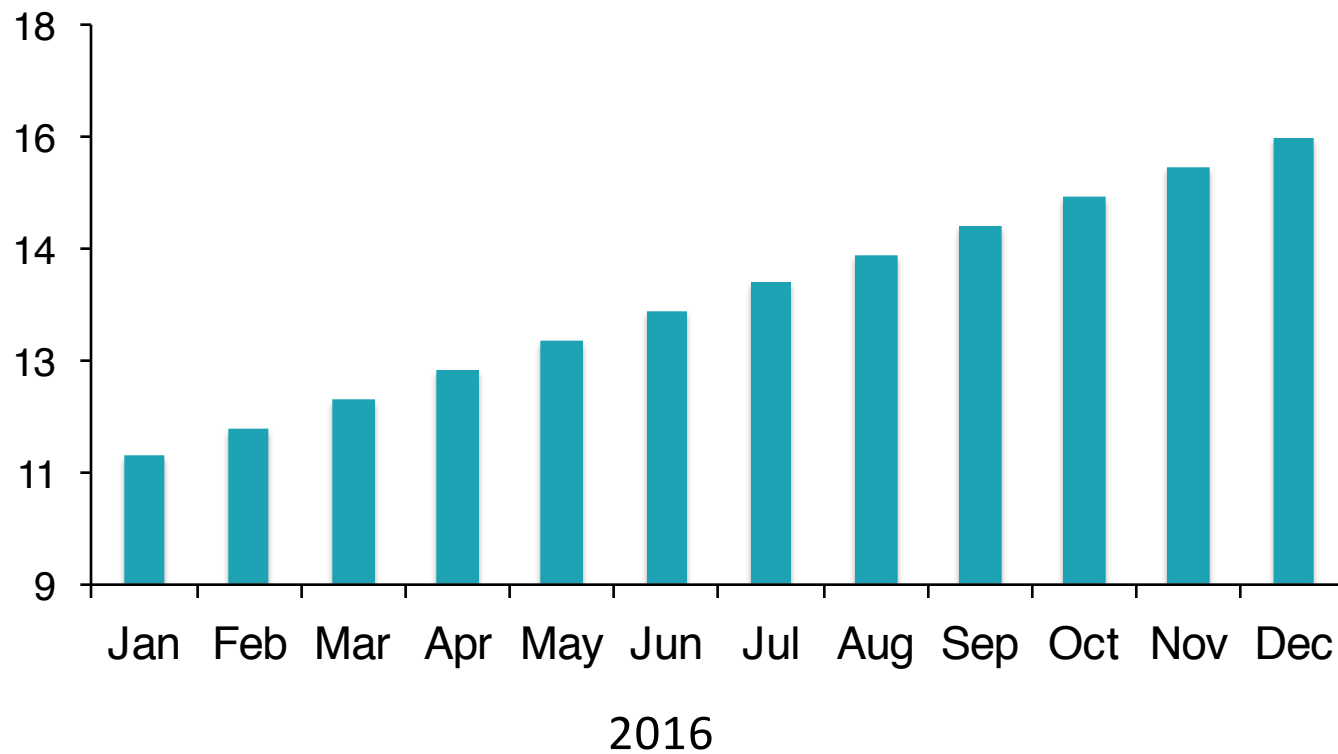| FEATURES |
| UI & ANIMATIONS |
| CONFIGURATION SETTINGS |

→

| APP BINARY |
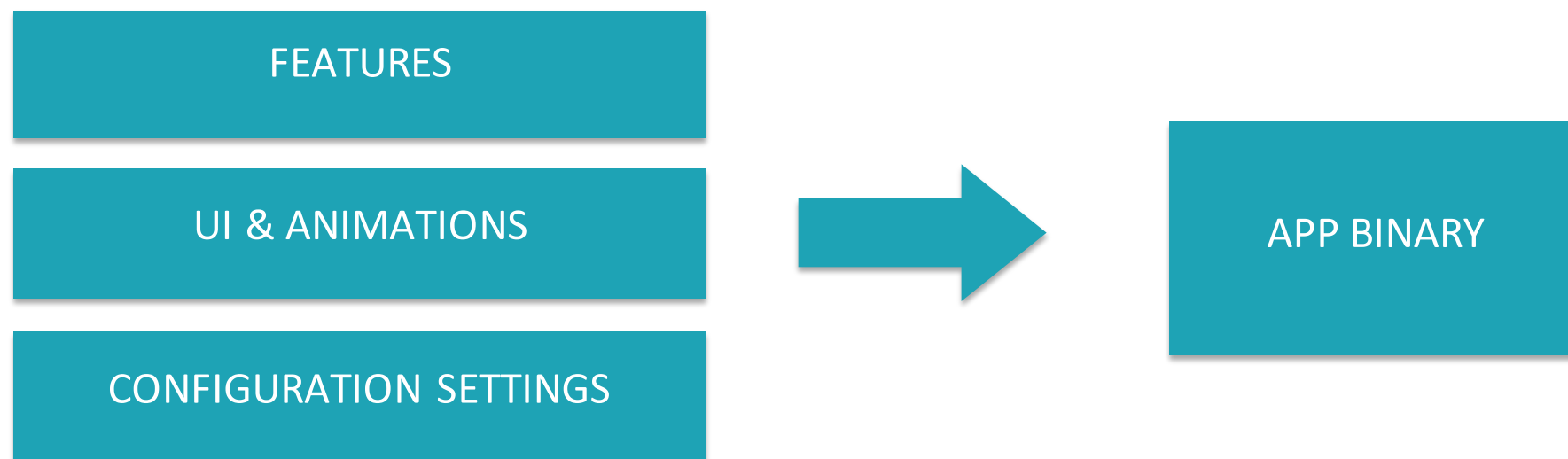
Disciplined Product Dev Approach is Key

- **23 brands in 170+ countries**

- Lots of market dynamics

- Many stakeholders

… Data is a must!

databricks™ VIACOM

Sound Data Management is Required

Expected Data Volume Growth (TB)
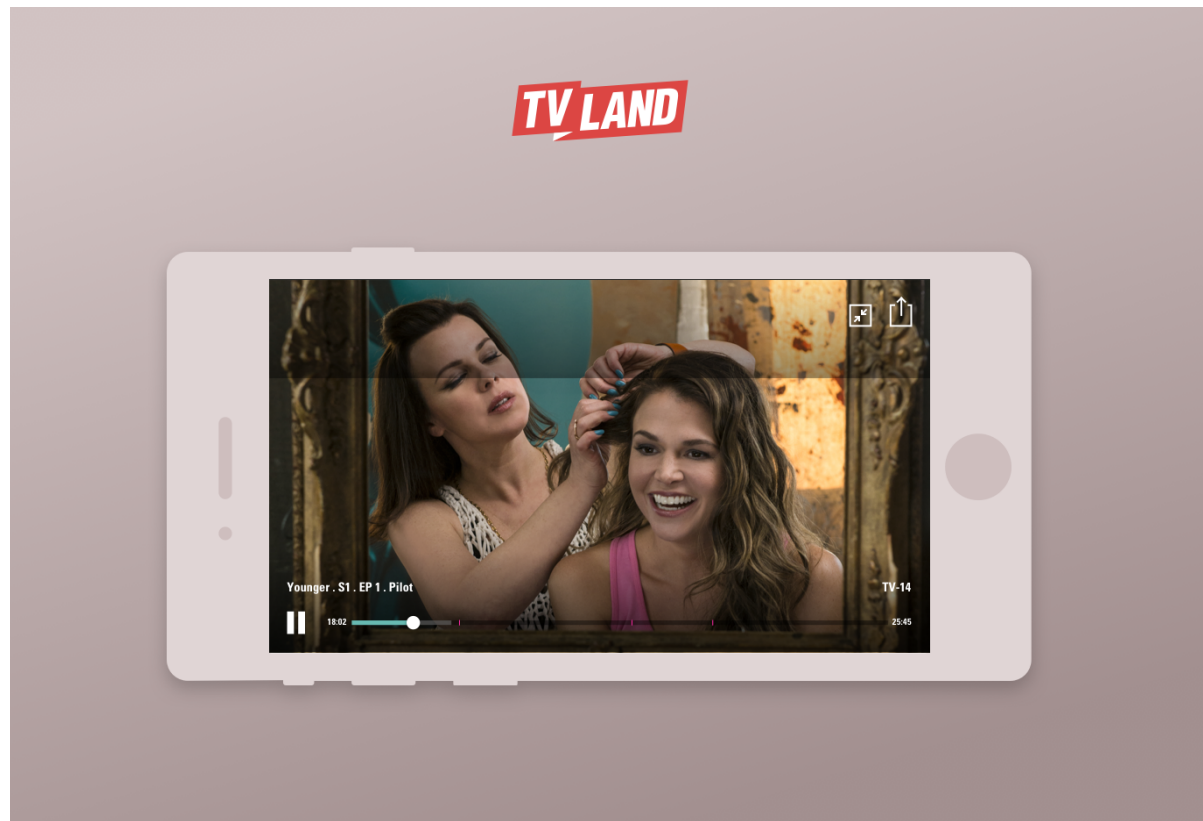

2016

Our Data Infrastructure

S3

Spark + Databricks

Redshift

Tableau

Introducing the TV Land iOS App

Applying Product Analytics to TV Land

- **Growth Targets**

- Dashboards

- Deep Dive Analyses

- A/B Testing

Baselines Used to Set Growth Targets

ETL	
S3	Spark + Databricks

Business Modeling	
Redshift	Tableau

Data Volume

- 30 sites/apps
- 11 TB

Data Volume

- 30 sites/apps
- 1 TB

Growth Targets are Monitored via Dashboards

Audience Growth by Cohort

New Users
Returning Users

| | 1/4/16 | 1/11/16 | 1/18/16 | 1/25/16 |
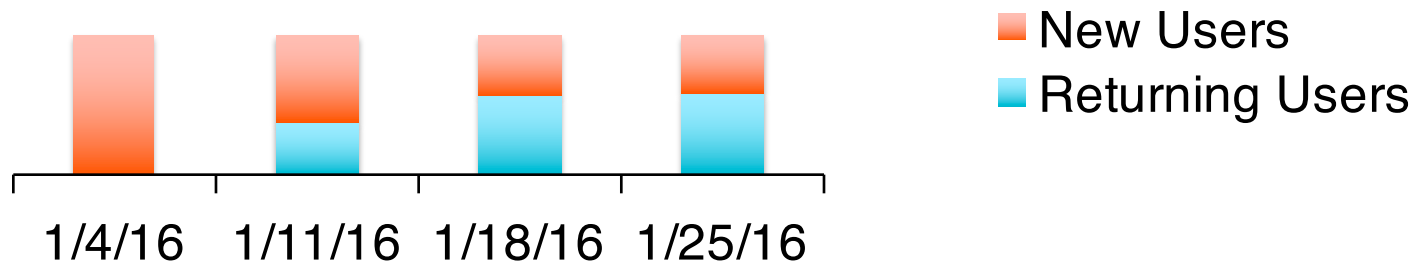
Weekly Retention by Cohort

	0	1	2	3	4
1/4/16	100%	53%	41%	33%	30%
1/11/16	100%	58%	51%	42%	
1/18/16	100%	49%	38%		
1/25/16	100%	49%			

databricks VIACOM

Dashboards Spark Deep Dive Analyses

% of users retained (y-axis: 0% to 100%)

Users will quickly churn if not activated within small window of time.

x-axis: 0 1 2 3 4 5 6 7 8 9 10 11 12

databricks™ VIACOM
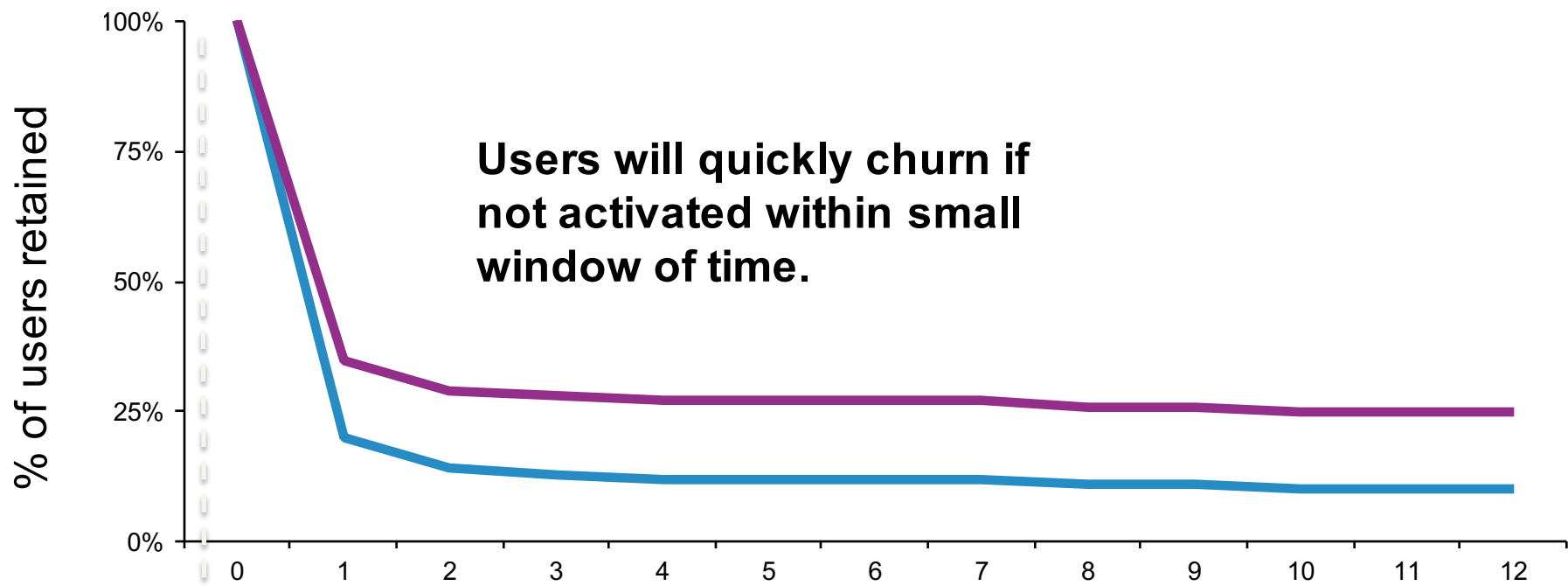
Deep Dive Analysis Requires Flexibility

Deep Dive Analysis		Not Needed	
S3	Spark + Databricks	*Redshift*	*Tableau*

- Define **schema on read** instead of write
- Work through data quality issues *just-in-time*.
- Tease out business question **iterative and interactively**.
- Use **programming language** of your choice.

Hypotheses Require A/B Testing

Statistical Analysis		Not Needed	
S3	Spark + Databricks	*Redshift*	*Tableau*

Data Sets

- Adobe Logs
- Experiment Logs

Summary of Our Setup

	Just in Time	Traditional
Primary Audience	• Product Analysts	• Product Team • Business Stakeholders
Tasks	• Exploratory Analysis • A/B Testing	• Ad Hoc Queries • Dashboards
Tools	• S3 • Spark • Databricks	• Redshift • Tableau

databricks

17

Coming Soon…

- Go live with internal A/B testing platform

- Continue to evolve our setup

- Further scale model to support Product Analytics Pan-Viacom

Questions ?

databricks™

Thank you.

Other parting words or contact information go here.

databricks™