



# Build an open data ecosystem on AWS

Marco Chiapusso  
Director Platform Engineering, adidas



adidas

**BUILD AN OPEN  
DATA ECOSYSTEM  
ON AWS**

# **AGENDA**

- 1. ADIDAS STRATEGY**
- 2. DATA DRIVEN INSIGHT**
- 3. ELASTIC BIG DATA PLATFORM**
- 4. TAKE AWAYS**
- 5. Q&A**

## WHO I AM



### TOPICS I AM WORKING ON

Building new data platform,  
Agile transformation in  
Advanced Analytics



### CAREER HIGHLIGHTS

Lead infrastructure transformation in a Large  
manufacturing company  
Pushing Cloud since 8 yrs  
Helping adidas in his journey



### MY FAVORITE THING TO DO WHEN I AM NOT WORKING

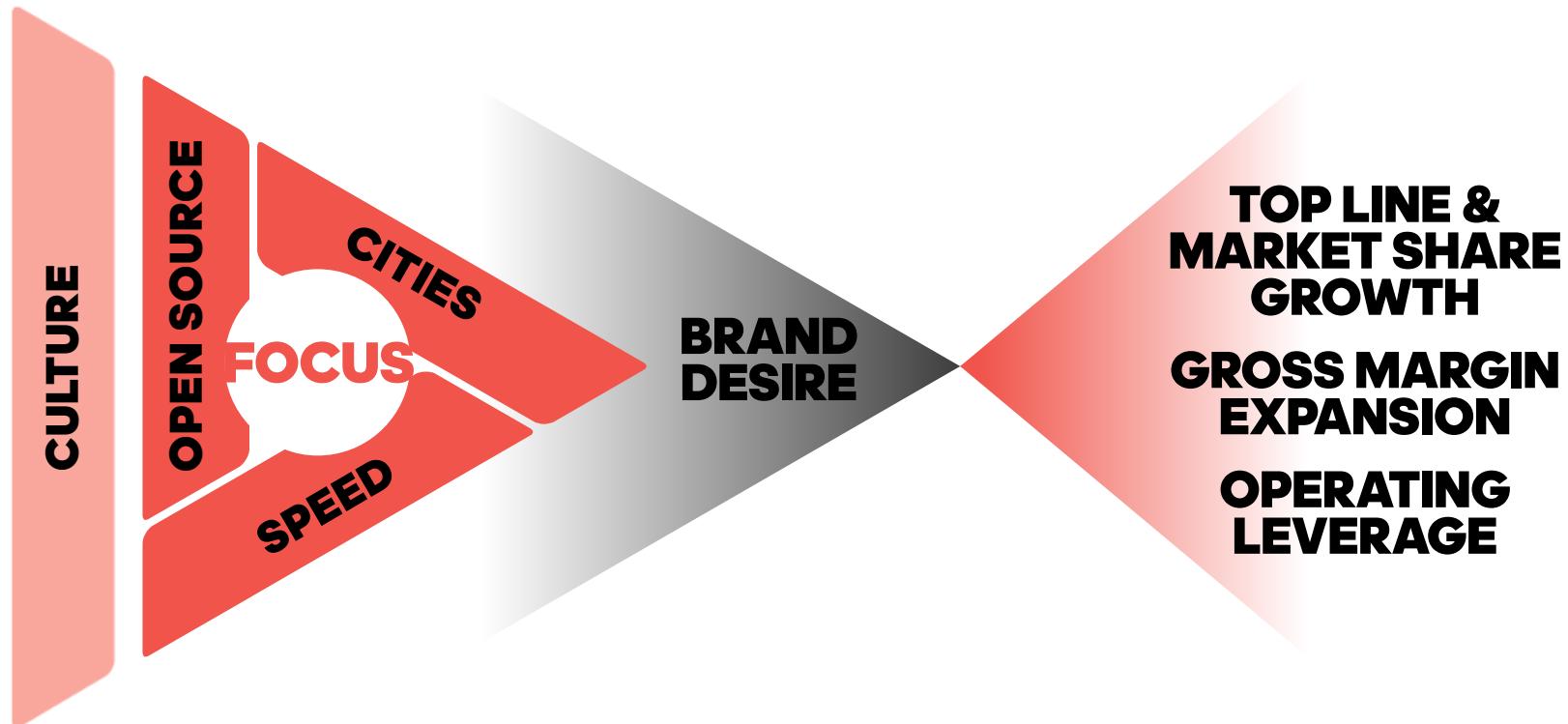
Playing with my daughter, Learning new  
things, Motorbikes, Lego Collector.



**WE ARE CREATING THE NEW**



## OUR STRATEGY FOR CREATING THE NEW



## WHY CULTURE IS A RELEVANT ASPECT



- I spent 70% of my time waiting
  - I don't know who to escalate to
- Please open a ticket
  - What IT is doing? I need my app now
- I'm constantly fire fighting
  - I don't have spare capacity

# DATA DRIVEN INSIGHT

## **TRANSFORMATION PHASE**



„THERE ARE NOT TWO IT AT  
DIFFERENT SPEED , BUT JUST ONE IT  
AT HIGH SPEED“

M.VOEGELE , ADIDAS CIO

## DATA DRIVEN INSIGHT

OUR WEBSITE IS THE MOST IMPORTANT STORE  
WE HAVE IN THE WORLD

GERMANY MARCH 14, 2018. REUTERS/MICHAEL DALDER

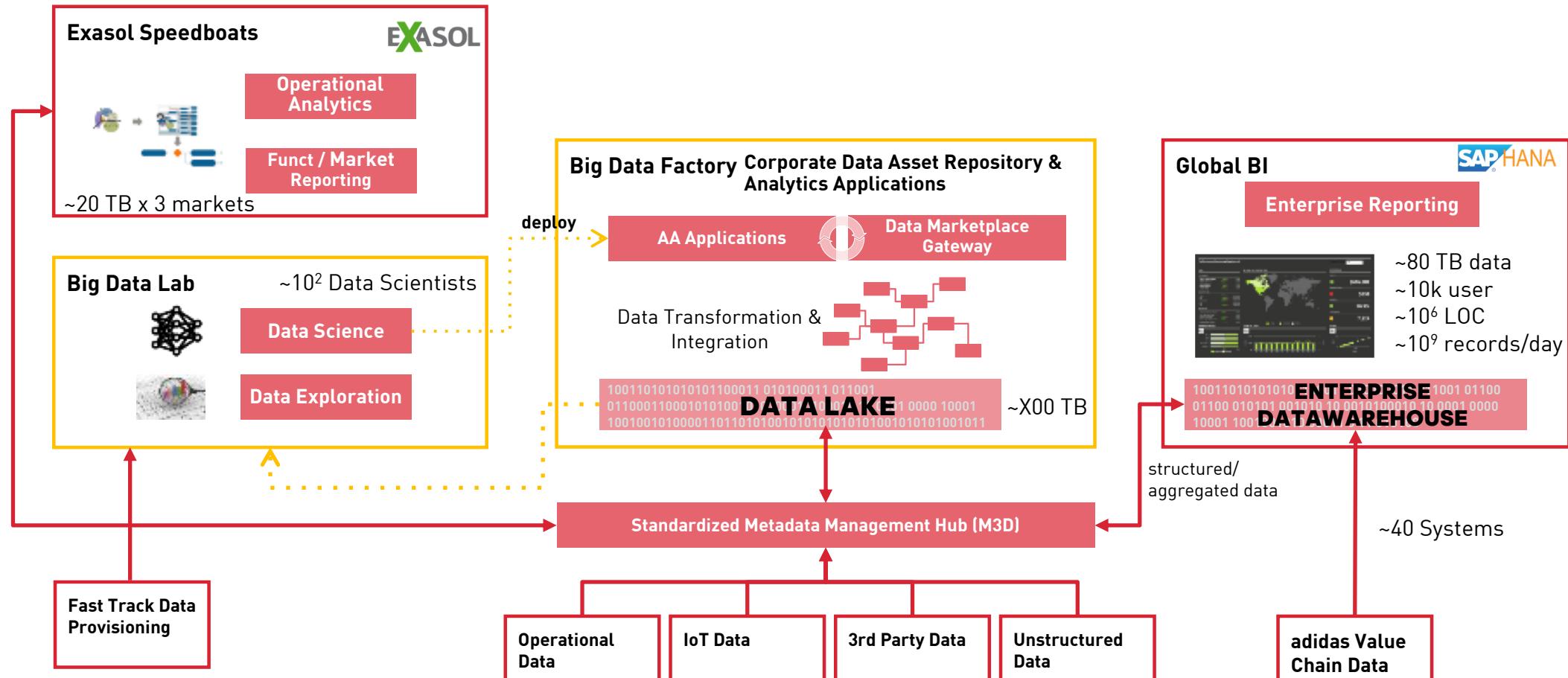


Photograph by Krisztian Bocsi— Bloomberg  
via Getty Images

## DATA DRIVEN INSIGHT

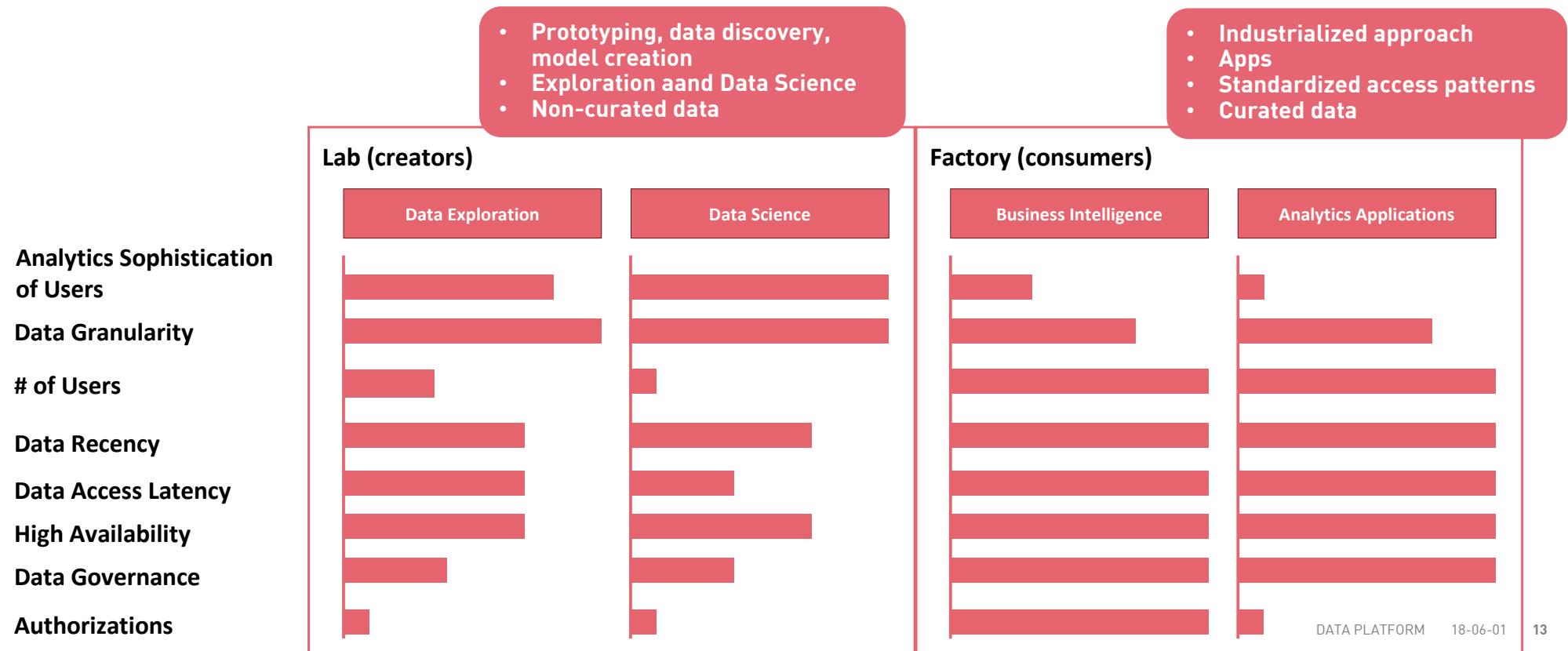


## **DATA ANALYTICS LANDSCAPE AT ADIDAS**



## BASIC TERMINOLOGIES: LAB AND FACTORY

Lab and Factory are the major terminologies to distinguish when looking at the Analytics platform. Whereas Factory is set up to enable standardized access to curated data, Lab serves the purpose of individual data exploration and investigation of new analytical methods. The following picture shows up the differences based on criteria of usage.



# EMBEDDING LEARNINGS IN FACTORY PROCESSES

## LAB



### LAB IS FOR

- Building prototypes
- Testing hypothesis
- Innovations

### LAB REQUIRES

- Reliable data foundation
- Own ad-hoc data
- Very high speed of changes
- Flexibility in tools
- High level of self-service

## FACTORY IS FOR

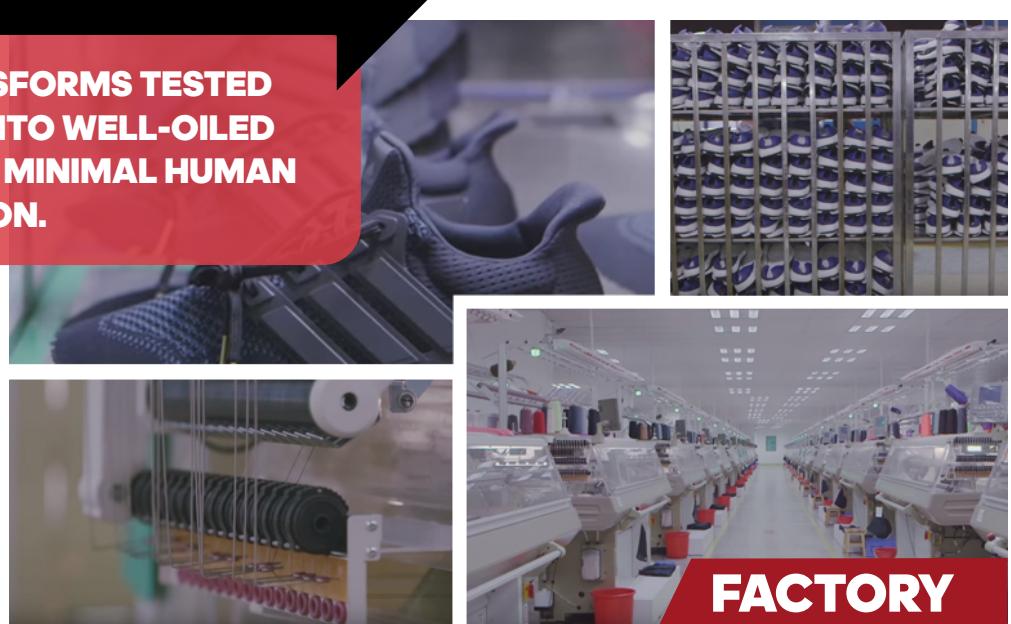
- Running decision engines
- Automated decision making
- Repeatable tasks

## FACTORY REQUIREMENTS

- Reliable data foundation
- Automated data supply
- Predictable running times
- Reliable calculation results
- E2E monitoring and support

## FACTORIZATION

**“FACTORIZATION” TRANSFORMS TESTED VALUABLE PROTOTYPE INTO WELL-OILED ENGINE, WHICH REQUIRES MINIMAL HUMAN INTERACTION.**



## FACTORY



# ELASTIC BIG DATA PLATFORM

# ELASTIC BIG DATA PLATFORM

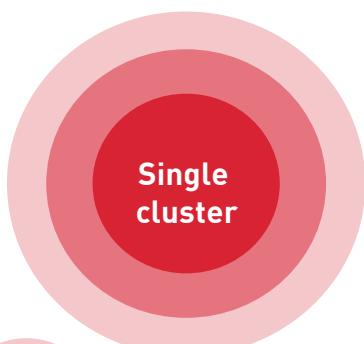
## BDP GENERATION 1

Since 2015



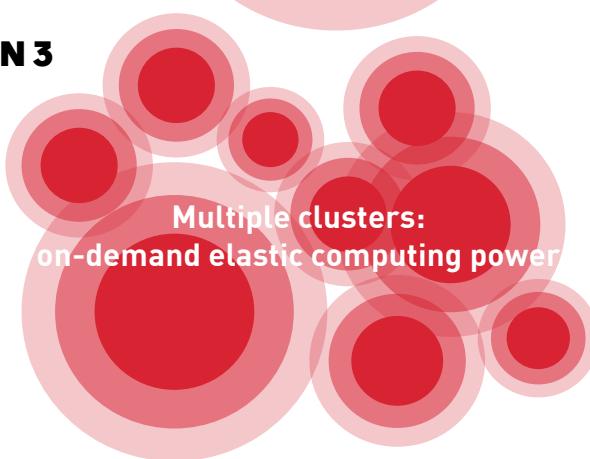
## BDP GENERATION 2

Active since H2.2017



## BDP GENERATION 3

New from Q3.2018



### Sandbox setup up to 12 worker nodes

- up to 100 TB data, up to 384 CPUs
- Manual monitoring and scaling
- Basic data curation

- First insights generated

### Productive setup up to 50+ worker nodes

- up to **Petabytes** of data, up to **1600** CPUs
- **High availability** of the core components
- **Automated** monitoring and housekeeping
- Full operations model
- **Curated** data lake
- Streamlined user management

- Productive factory with SLAs
- Impact from productive applications payed the whole system build in less than one year

### Fully flexible cloud computing

- **Elastic** scalability
- storage and CPUs, up to  $\infty$
- **On-demand** Data Labs for Data Scientists
- GPU-based computing
- **Pay-per-use** model
- De-coupled Data and Processing power

- Backbone for all Insights Driven initiatives
- Next generation of flexible workspaces for machine learning algorithms of any complexity
- Around-the-clock availability for Data Scientists across the Globe
- More data points from Global BI and other systems

### Own-your-base: full internal ownership of the technology

## PLATFORM STRATEGY

**Definition:** A platform is an open architecture, together with a governance model, designed to facilitate interactions.

- **Open architecture:** allows 3rd parties to participate, build, and innovate in standardized way.
- **Governance:** gives power to exclude bad actors, steer community behavior.
- **Interactions:** means by which value is created, finer grain than “exchange.”
- **Platform Purpose:** The goal of a platform is to consummate the match, interactions that create value.

“From PLATFORM REVOLUTION: HOW NETWORKED MARKETS ARE TRANSFORMING THE ECONOMY—AND HOW TO MAKE THEM WORK FOR YOU, by Geoffrey G. Parker, Marshall W. Van Alstyne, and Sangeet Paul Choudary. Copyright © 2016 by Geoffrey G. Parker, Marshall W. Van Alstyne, and Sangeet Paul Choudary.

## PLATFORM STRATEGY

**iPhone has email, calendar, contacts, browser, movies, etc. Apps built from underlying functions:**

- Video
- Multitouch screen
- WiFi
- Camera & Scanner
- Processor
- Accelerometer
- G3 mobile
- Audio playback
- Clock
- Geolocation

**Recombined into 2,000,000+ apps Apple never conceived!**



## **DESIGN PRINCIPLES**

### **Build decoupled systems**

- Data → Store → Process → Store → Analyze → Answers

### **Use the right tool for the job**

- Data structure, latency, throughput, access patterns

### **Leverage AWS managed and serverless services**

- Scalable/elastic, available, reliable, secure, no/low admin

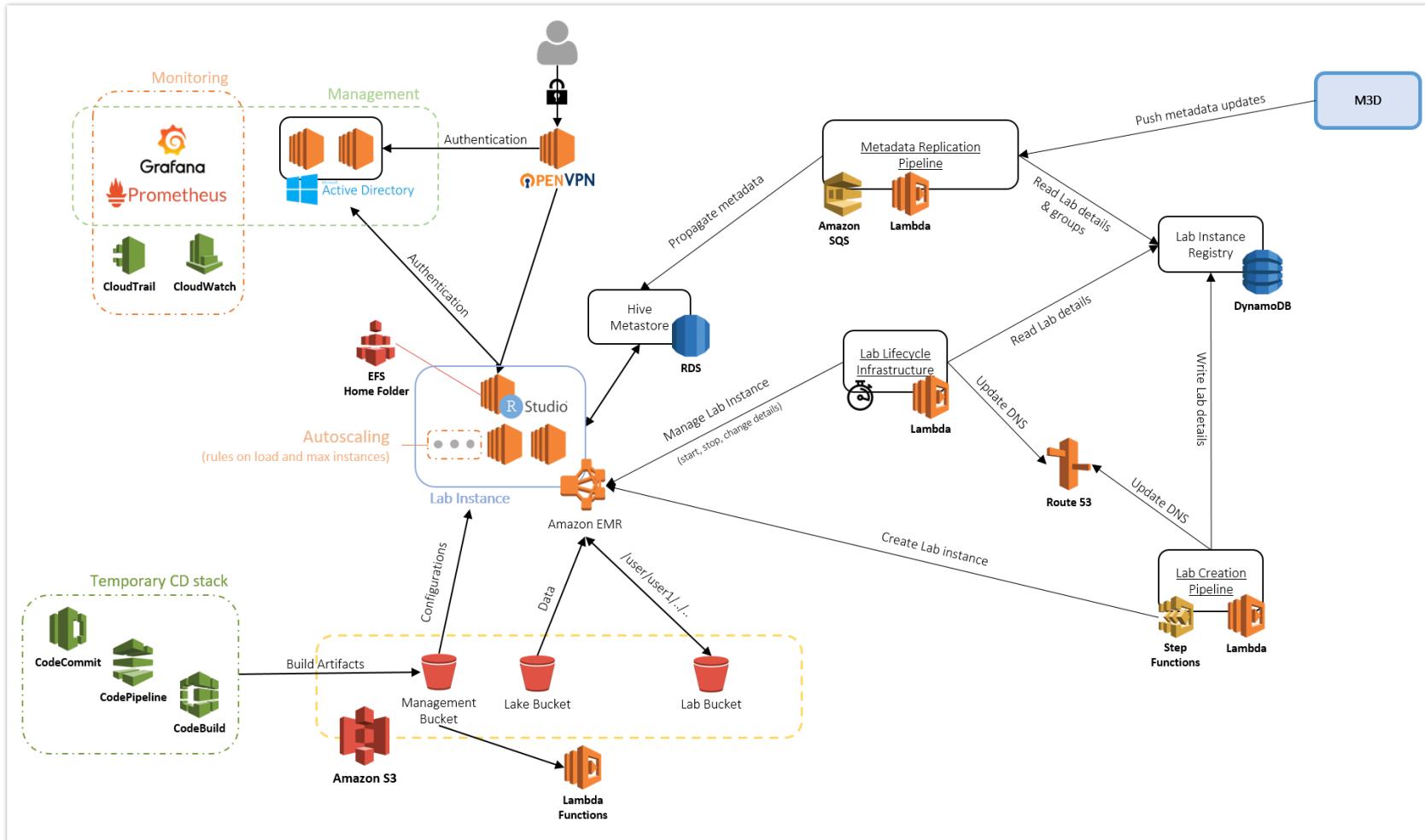
### **Use log-centric design patterns**

- Immutable logs, data lake, materialized views

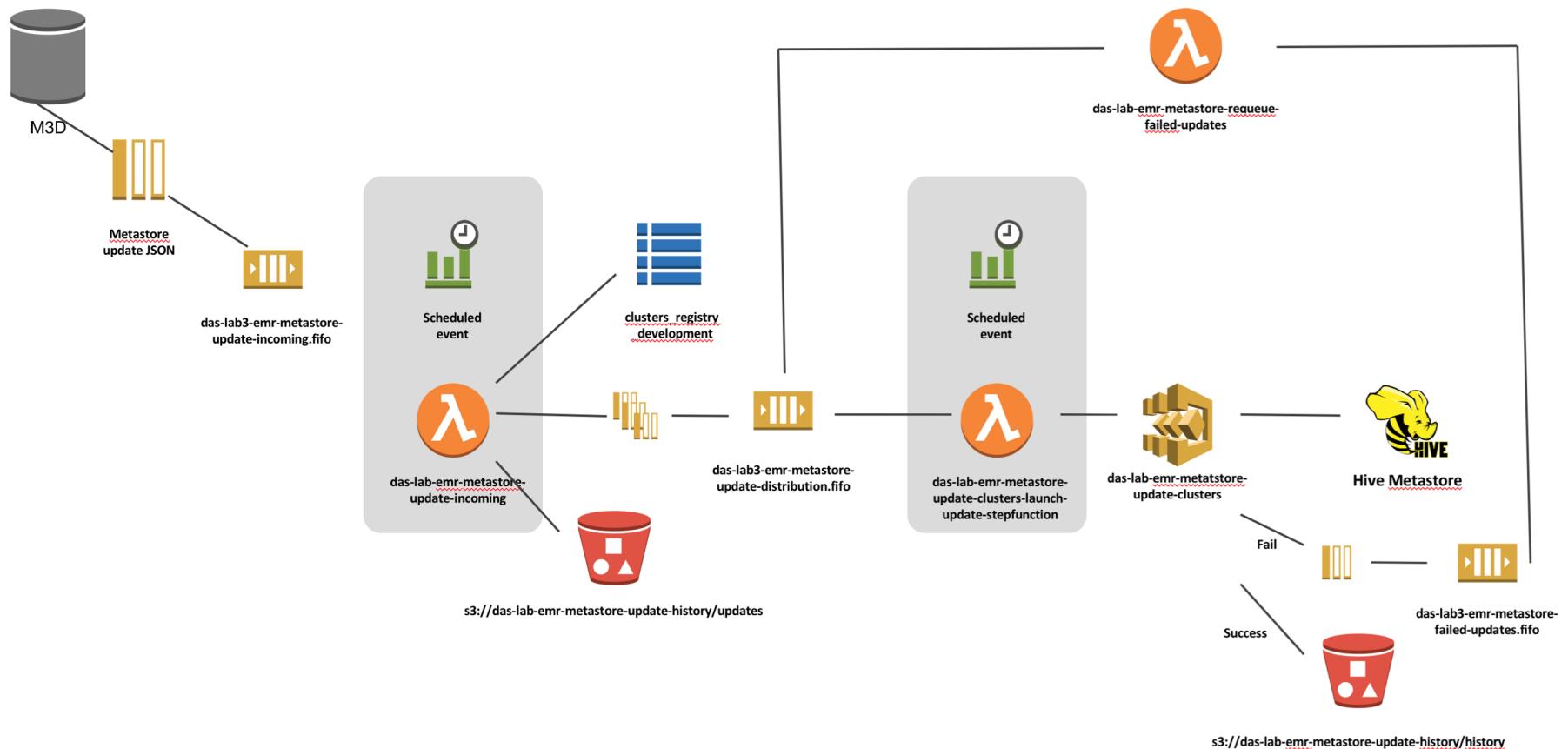
### **Be cost-conscious**

- Big data ≠ Big cost

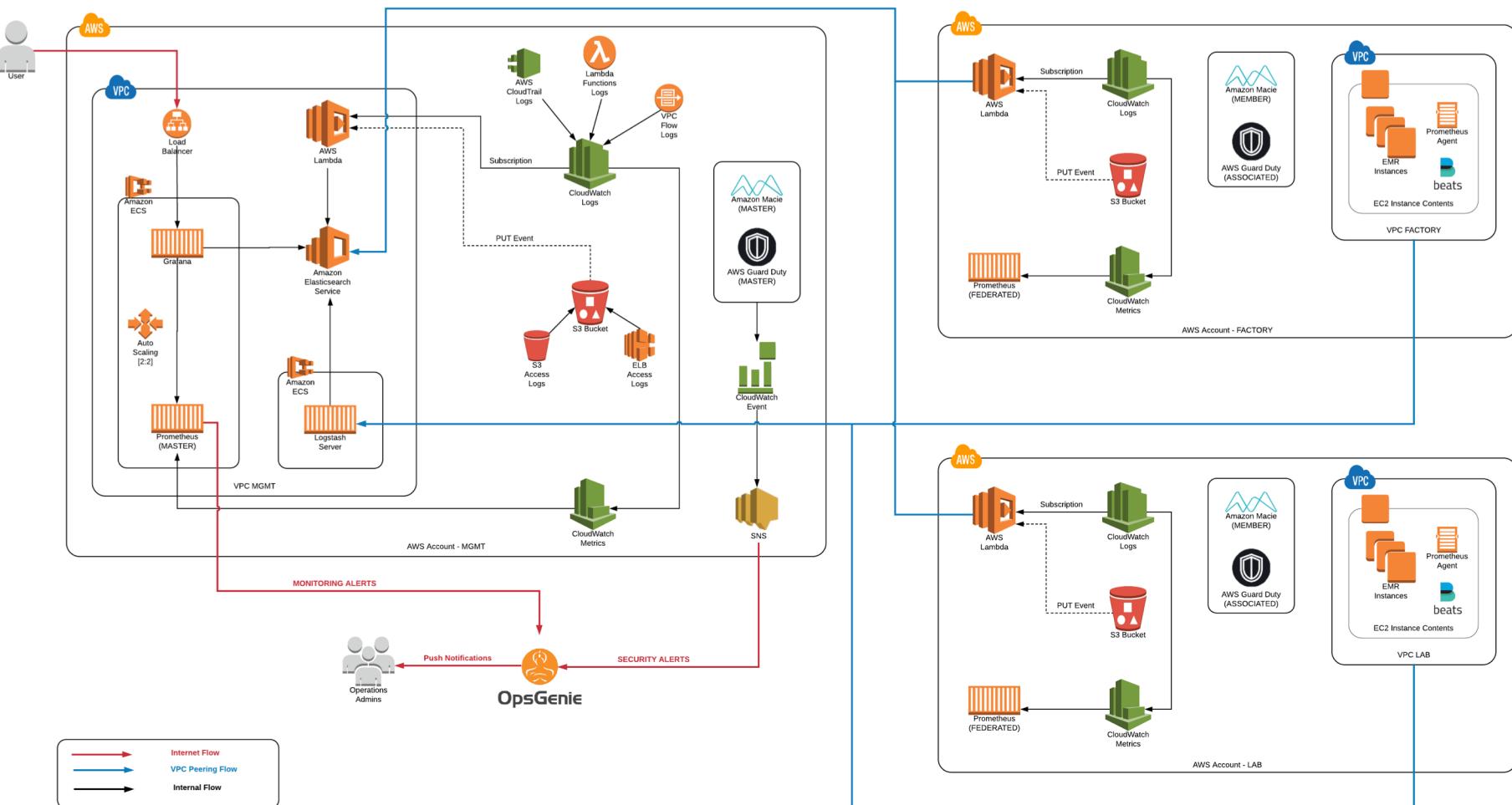
# ELASTIC ARCHITECTURE



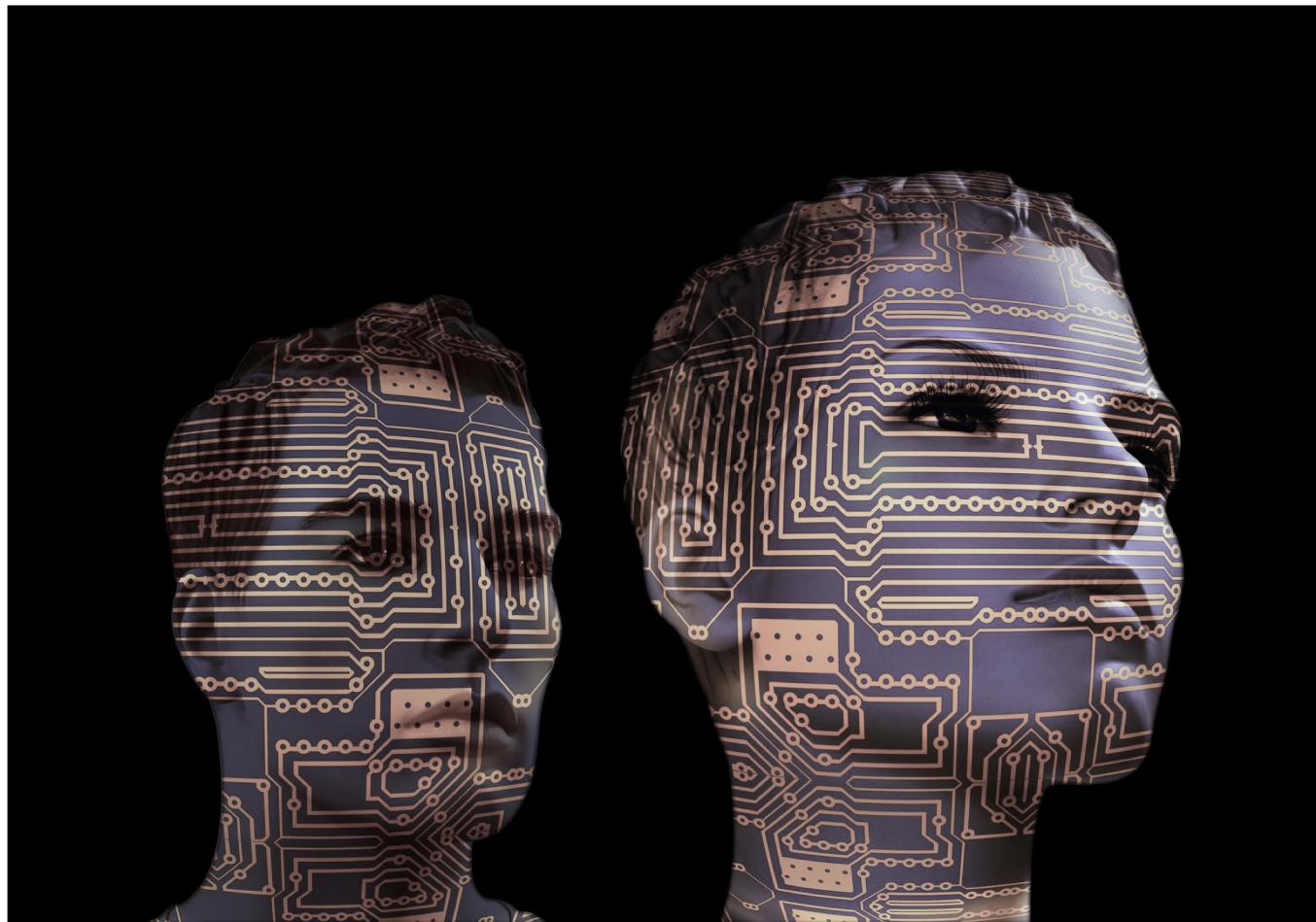
# ELASTIC ARCHITECTURE



# ELASTIC ARCHITECTURE



## AI/ML ENABLED ARCHITECTURE



## KEY TAKE AWAYS

BE READY FOR CHALLENGES AND OBSTACLES

ACCEPT FAILURES AND LEARN FROM THEM

BE FOCUSED AND PERSISTENT



---

## Q&A