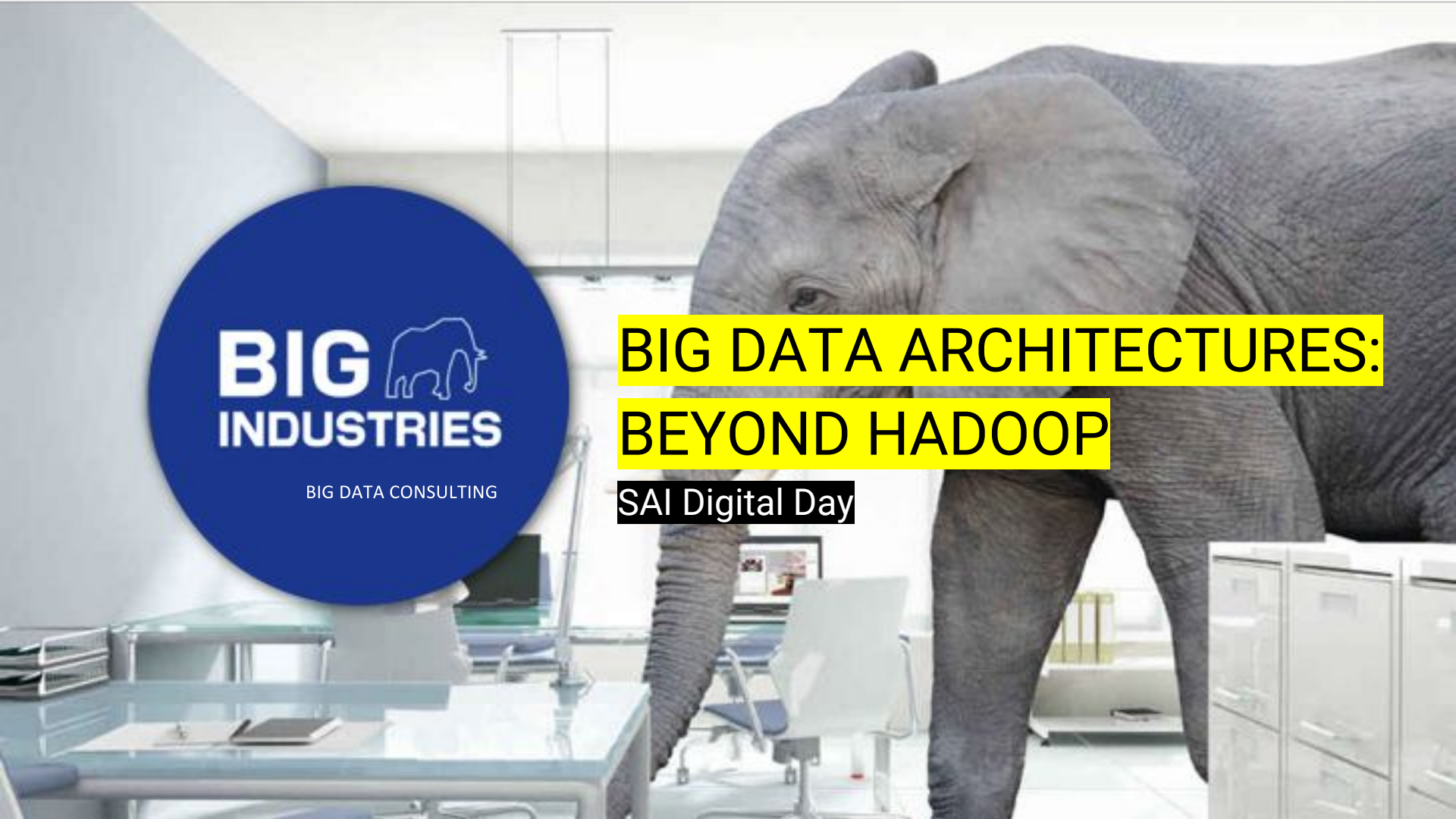
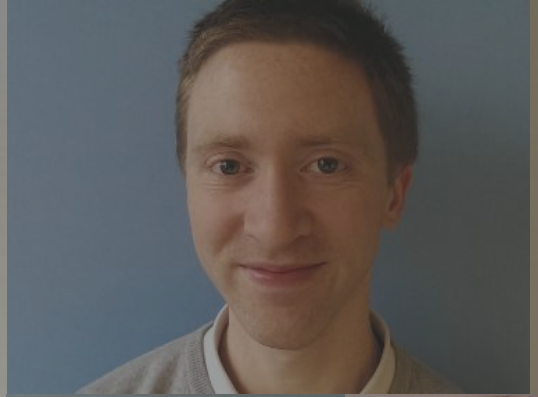
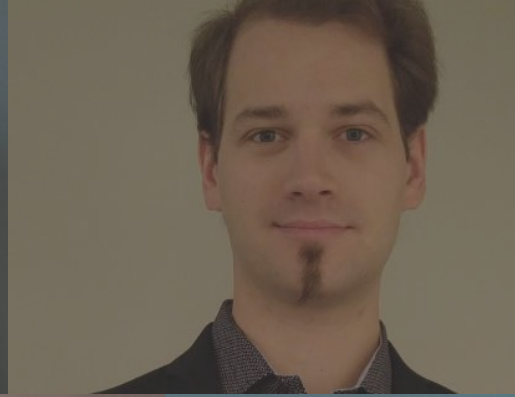


BIG DATA ARCHITECTURES: BEYOND HADOOP

SAI Digital Day

BIG INDUSTRIES

- ESTABLISHED 2014
- BIG DATA CONSULTING
- SERVING BELGIUM
- FOCUS ON SYSTEM ARCHITECTURE & SECURITY
- SERVING ALL KEY INDUSTRIES
- MEMBER OF CRONOS GROUP

BIG INDUSTRIES
Big Data Consulting

DATA SCIENCE
DATA DRIVEN APPLICATIONS
BIG DATA ARCHITECTURE CONSULTING
GOVERNANCE & SECURITY
CLOUD
BIG INDUSTRIES
Big Data Consulting

THERE HAS ALWAYS BEEN BIG DATA

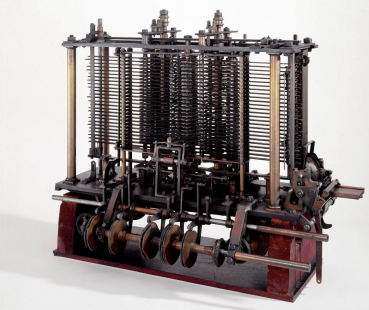
- HUMAN DEVELOPMENT THRIVES ON INNOVATION
- DEFINE “BIG”

BIG INDUSTRIES
Big Data Consulting

DATA VOLUMES KEEP GROWING

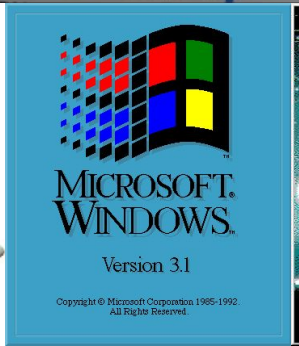
- TECHNOLOGY ADAPTS
- -OR SHOULD ADAPT-

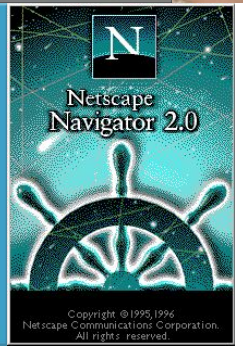
BIG INDUSTRIES
Big Data Consulting



LIVE FREE OR DIE
UNIX*
TRADEMARK OF BELL LABS*
UNIXライセンスプレート

University of Minnesota


MICROSOFT WINDOWS
Version 3.1
Copyright © Microsoft Corporation 1985-1992.
All Rights Reserved.


Netscape
Navigator 2.0
Copyright ©1995,1996
Netscape Communications Corporation.
All rights reserved.

A80581-66
SX837
intel®
pentium™
PROCESSOR
L4142580
INTEL©©1992

Google!

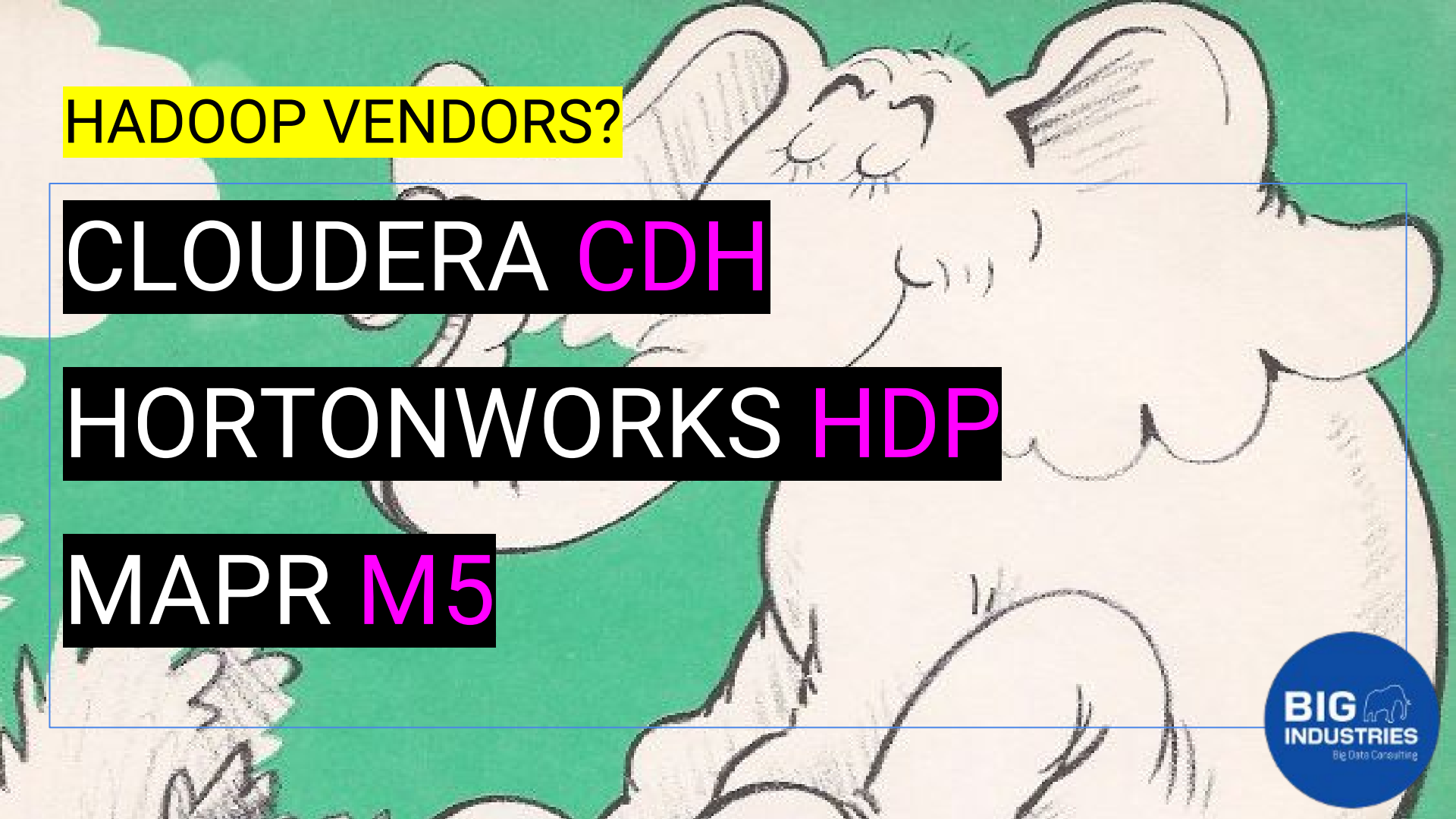
1995 - 2005: ZOMBIE IT
VENDOR DECADE
BIG
INDUSTRIES
Big Data Consulting

WHAT IS HADOOP

- A GRID COMPUTING STACK
- MODULAR, EXTENSIBLE DATA MANAGEMENT
- RELATIVELY EASY TO USE
- FREE OPEN SOURCE SOFTWARE

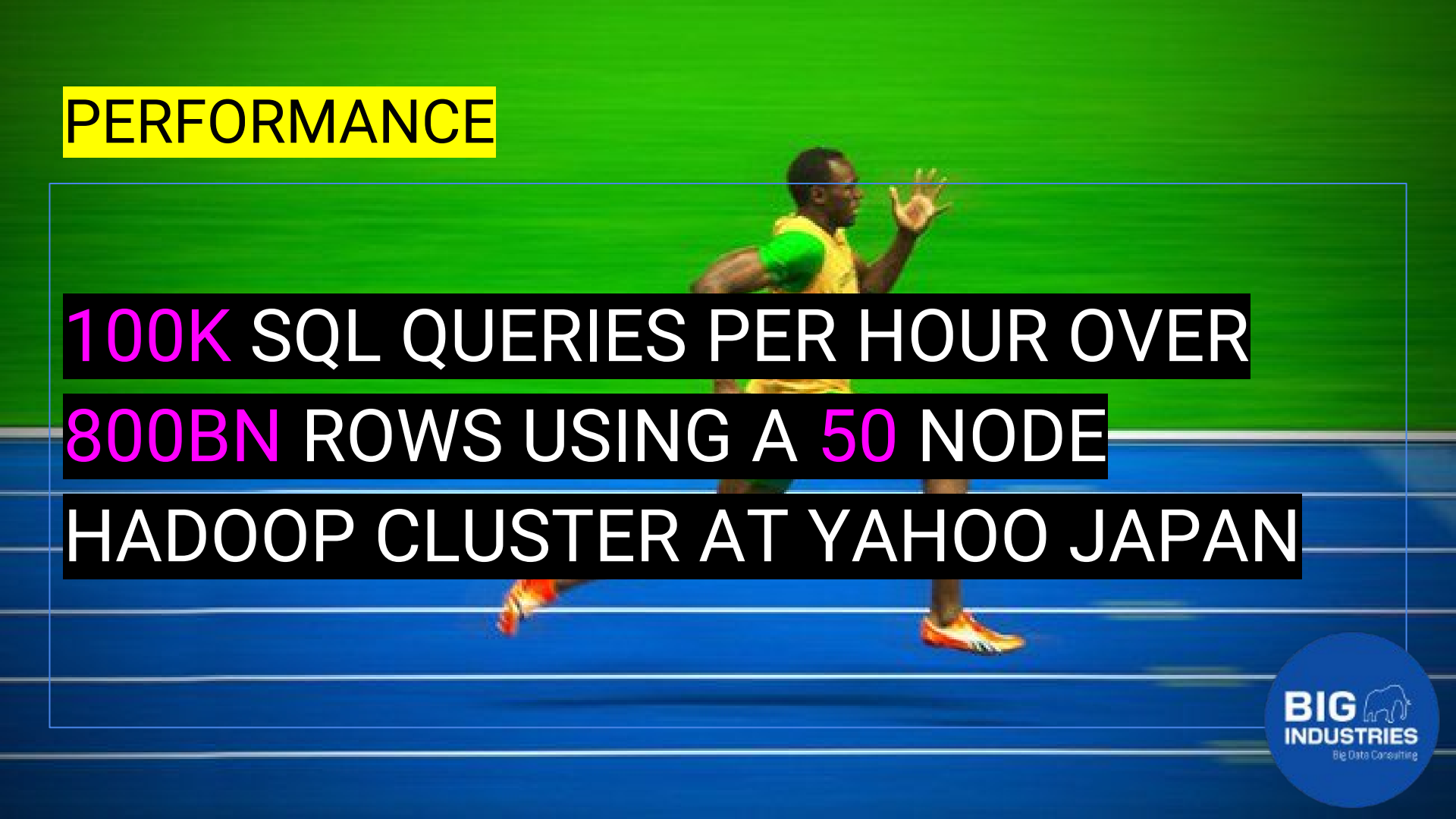
BIG INDUSTRIES
Big Data Consulting

HADOOP: ADOPTION
48% OF GLOBAL ENTERPRISES HAVE SET UP A DATA LAKE*
* FORRESTER RESEARCH
BIG INDUSTRIES
Big Data Consulting


HADOOP VENDORS?
CLOUDERA CDH
HORTONWORKS HDP
MAPR M5
BIG
INDUSTRIES
Big Data Consulting

HADOOP BENEFITS

- HORIZONTALLY SCALABLE DATA PROCESSING
- HIGH PERFORMANCE
- FLEXIBLE WORKLOADS
  - AI/MACHINE LEARNING
  - BI/DATA WAREHOUSING
  - STREAMING DATA
  - NOSQL DATABASE


PERFORMANCE
100K SQL QUERIES PER HOUR OVER
800BN ROWS USING A 50 NODE
HADOOP CLUSTER AT YAHOO JAPAN
BIG
INDUSTRIES
Big Data Consulting

WHAT’S THAT OVER THE HILL IS IT A MONSTER
BIG
INDUSTRIES
Big Data Consulting

amazon
BIG
INDUSTRIES
Big Data Consulting

CLOUD STRENGTHS

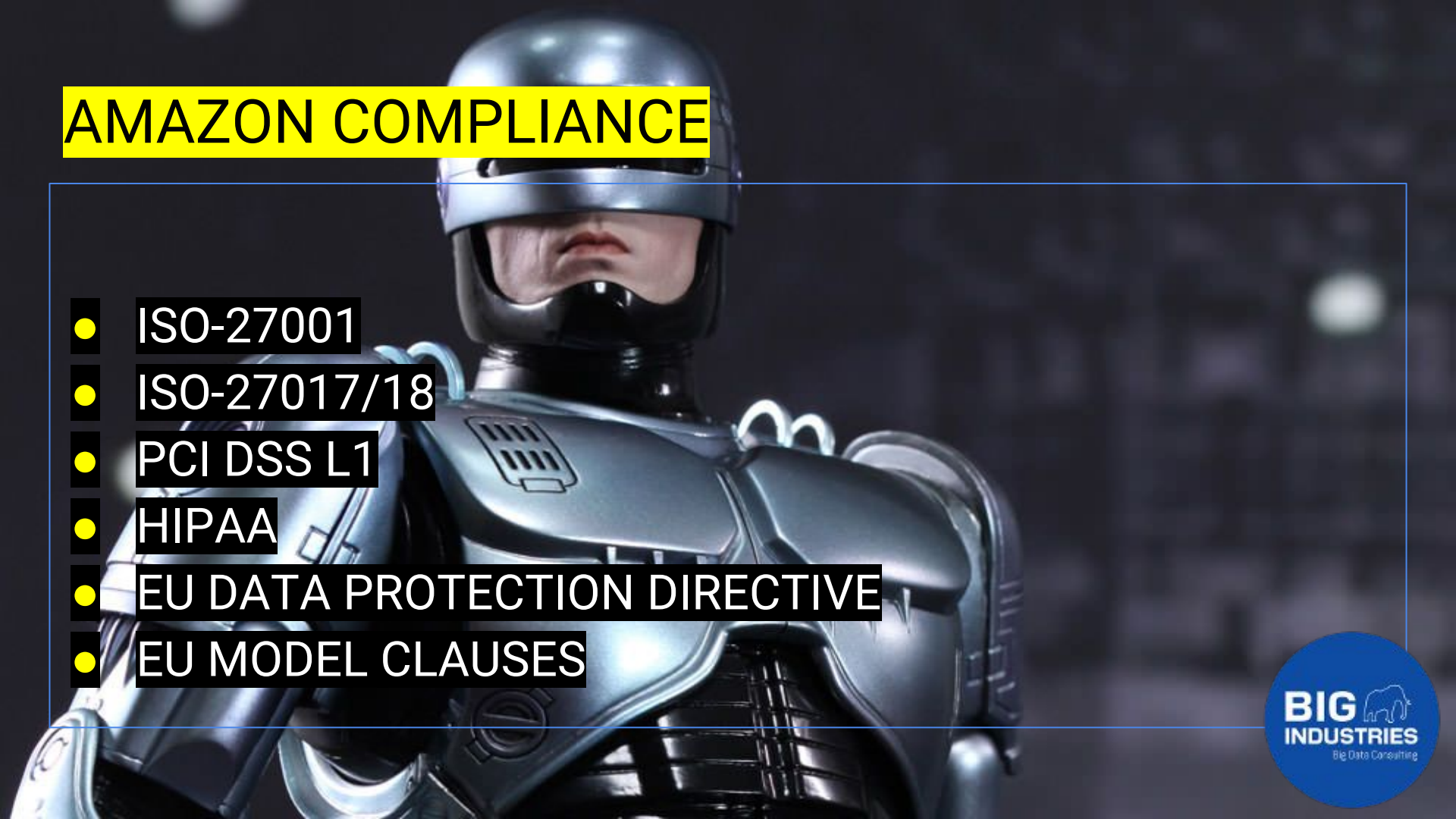
- COMPLIANT
- MATURE
- CAN BE CHEAP
- HIGHLY FLEXIBLE
- COMPREHENSIVE OFFERINGS

BIG INDUSTRIES
Big Data Consulting

CLOUD DRAWBACKS

- PROVIDER LOCK-IN
- RUNNING COSTS
- SHADOW IT SERVICES
- COMPLEXITY & RISK
- GETTING DATA BACK OUT

BIG INDUSTRIES
Big Data Consulting



AMAZON COMPLIANCE

- ISO-27001
- ISO-27017/18
- PCI DSS L1
- HIPAA
- EU DATA PROTECTION DIRECTIVE
- EU MODEL CLAUSES

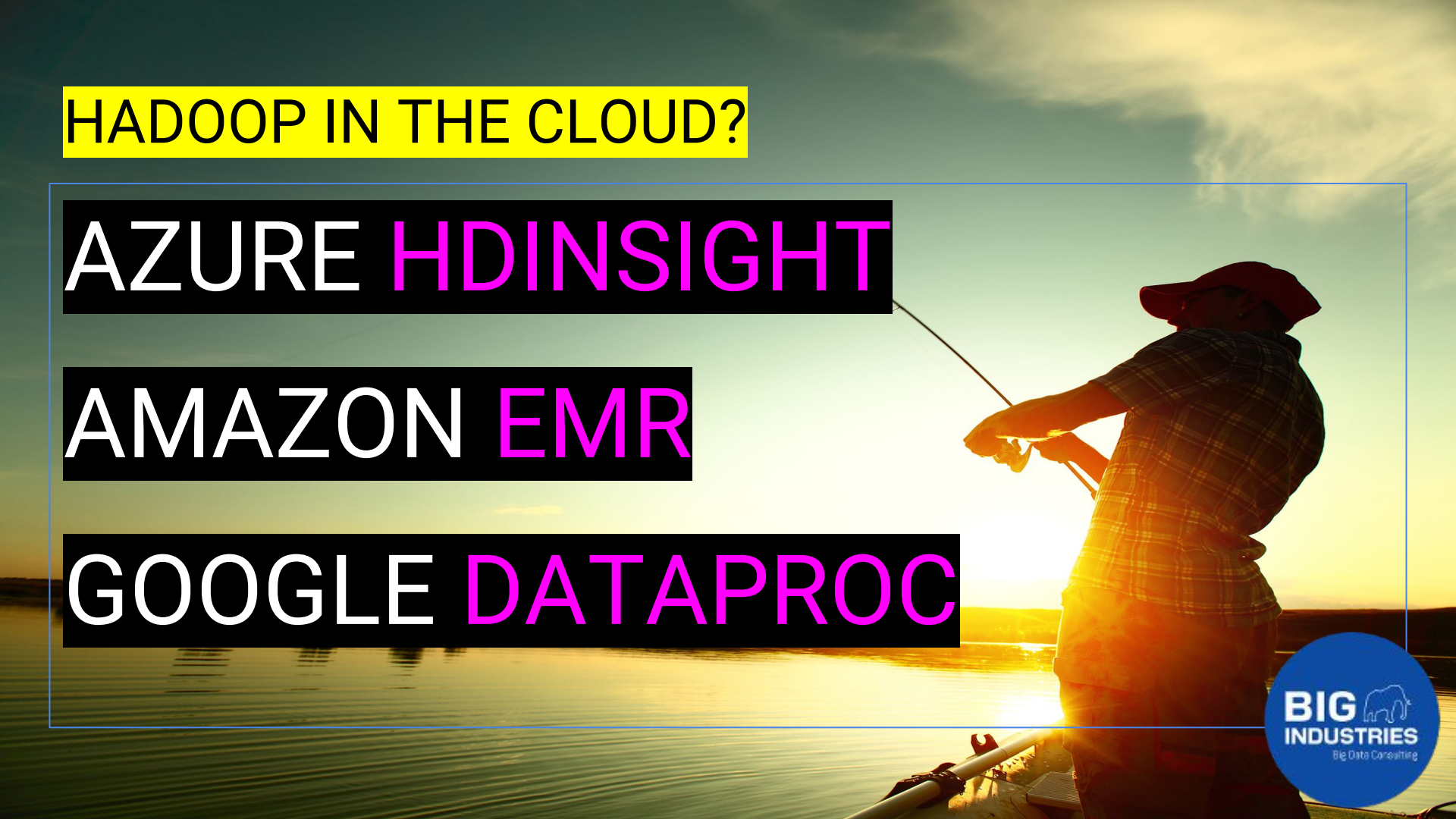
BIG INDUSTRIES
Big Data Consulting

HADOOP IN THE CLOUD?

AZURE HDINSIGHT

AMAZON EMR

GOOGLE DATAPROC

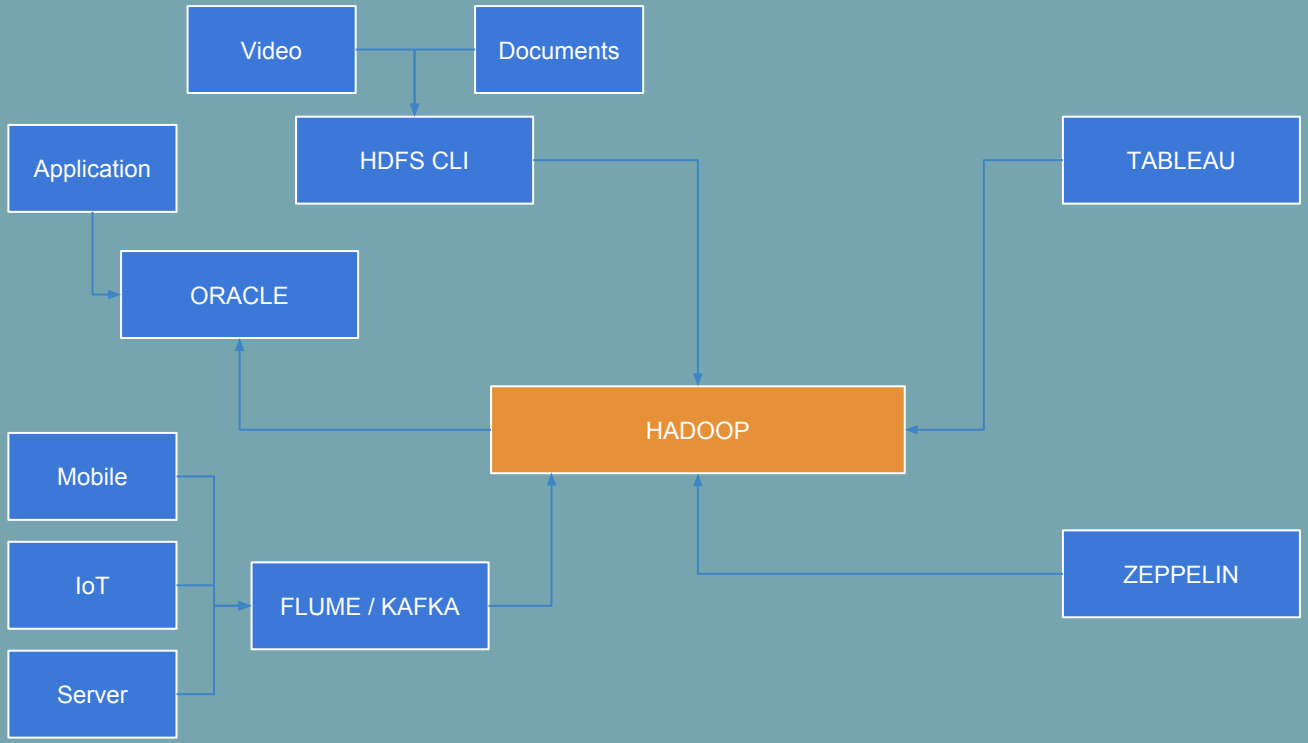


CLOUD HADOOP PLATFORMS

- + ELASTIC SCALABILITY
- + RAPID PROVISIONING
- + INTEGRATION WITH OTHER SERVICES
- - ACCESS CONTROLS
- - METADATA AND AUDITING
- - HIGH AVAILABILITY

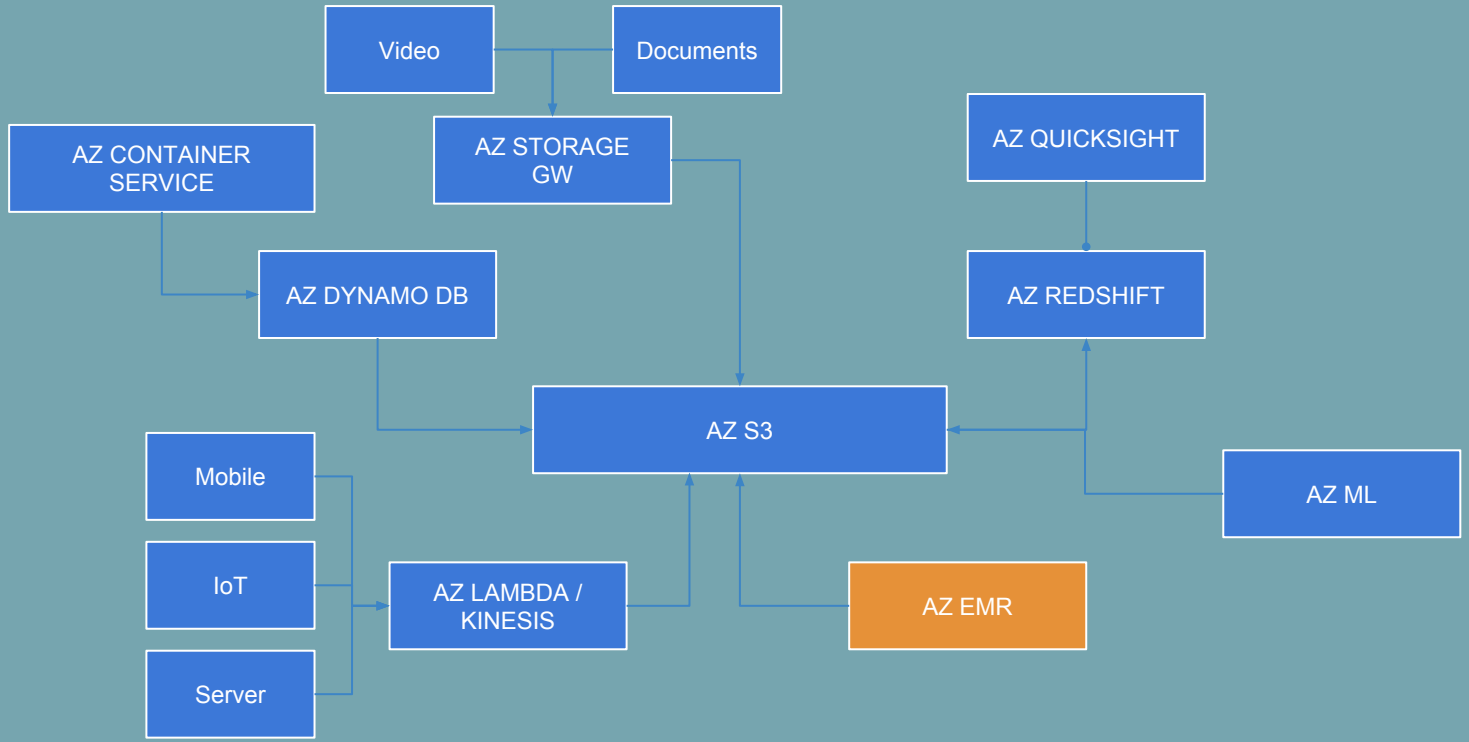
RIDING THE THIRD WAVE

CLASSIC HADOOP BIG DATA ARCHITECTURE



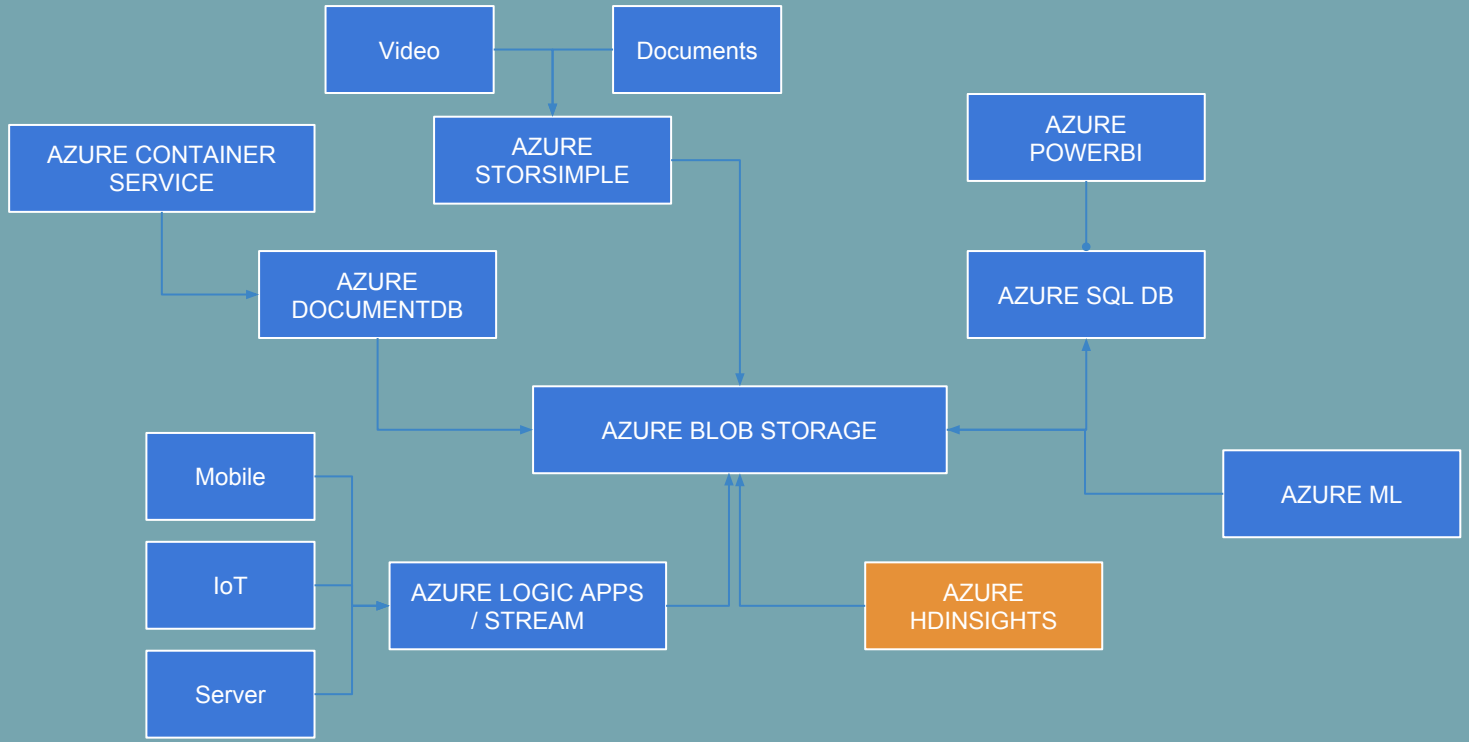
BIG INDUSTRIES
Big Data Consulting

CLOUD BIG DATA ARCHITECTURE (AWS)



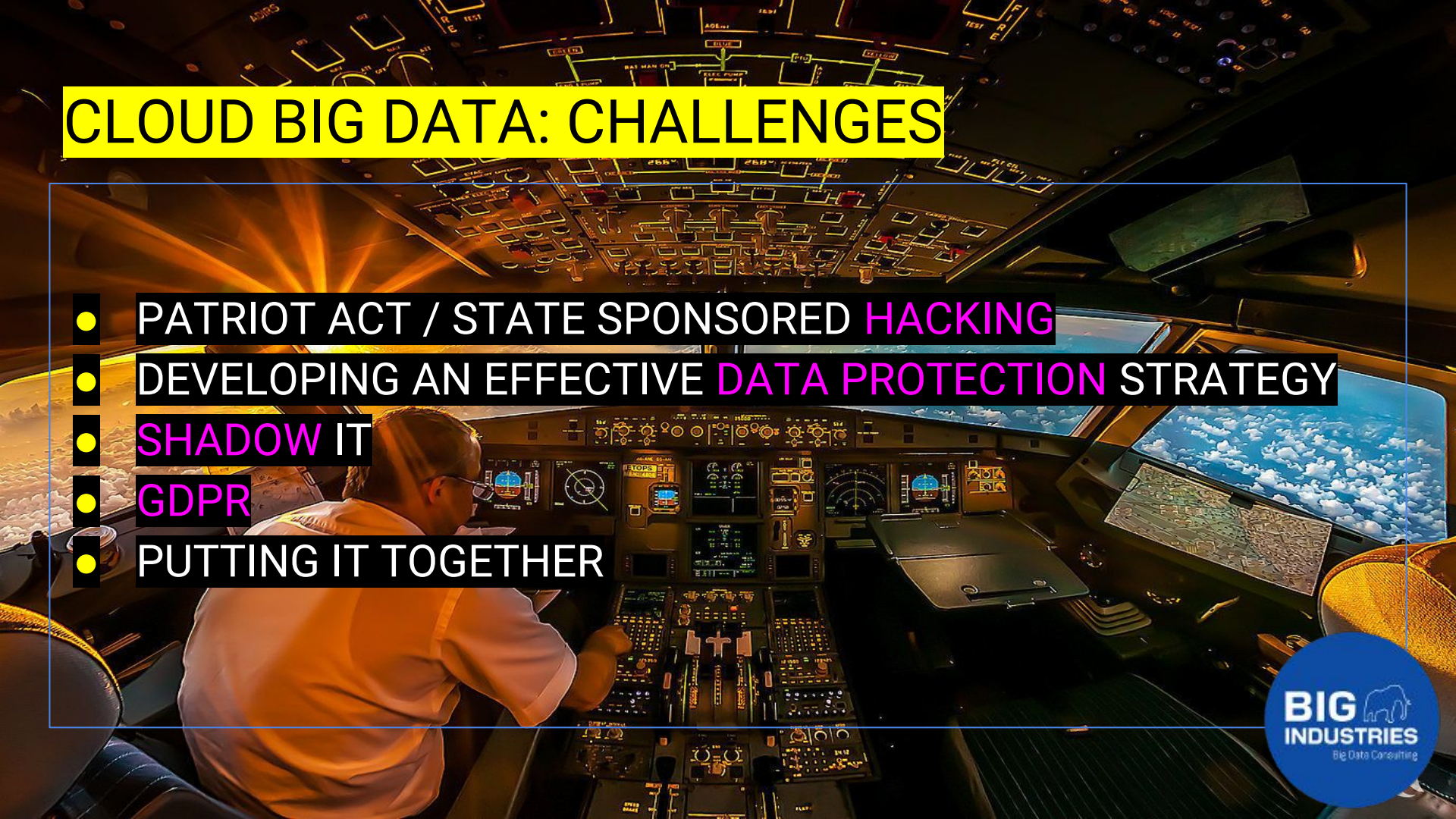
BIG INDUSTRIES
Big Data Consulting

CLOUD BIG DATA ARCHITECTURE (AZURE)



BIG INDUSTRIES
Big Data Consulting

CLOUD BIG DATA: CHALLENGES



- PATRIOT ACT / STATE SPONSORED HACKING
- DEVELOPING AN EFFECTIVE DATA PROTECTION STRATEGY
- SHADOW IT
- GDPR
- PUTTING IT TOGETHER

BIG INDUSTRIES
Big Data Consulting

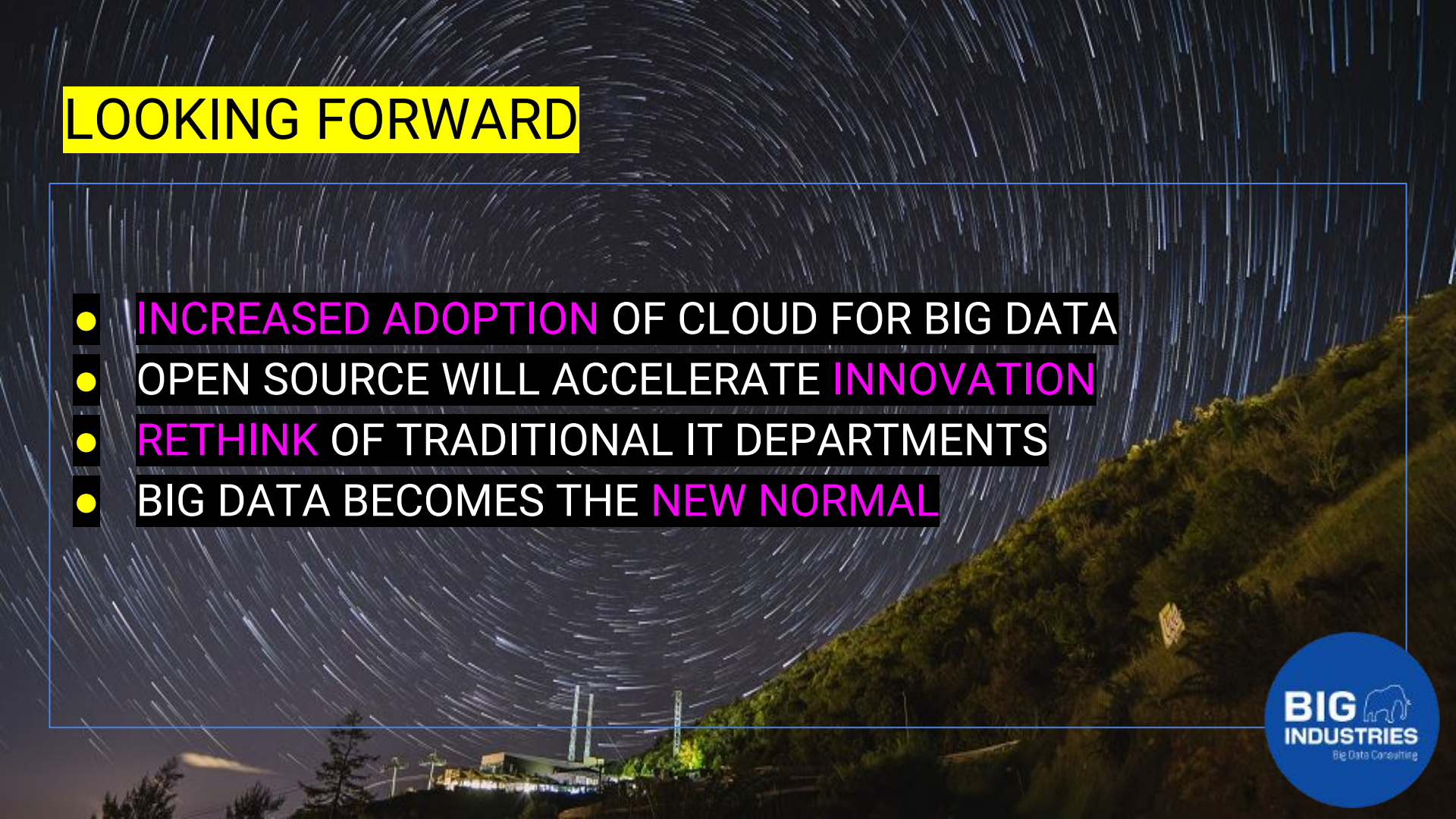
BIG DATA SECURITY RECOMMENDATIONS ON AWS

- USE AWS IAM
- APPLY VPC-ENDPOINT RESTRICTIONS
- APPLY S3 POLICY RESTRICTIONS
- USE S3 CLIENT MANAGED ENCRYPTION
- INVEST IN AWS CLOUD HSM
- USE AWS CLOUD TRAIL
- USE AWS EMR WITH A CUSTOM CRYPTO CODEC

BIG INDUSTRIES
Big Data Consulting

LOOKING FORWARD

- INCREASED ADOPTION OF CLOUD FOR BIG DATA
- OPEN SOURCE WILL ACCELERATE INNOVATION
- RETHINK OF TRADITIONAL IT DEPARTMENTS
- BIG DATA BECOMES THE NEW NORMAL



CONCLUSION

- THE RIGHT SOLUTION IS THE ONE THAT WORKS FOR YOU
- THERE ARE TRADEOFFS WHATEVER THE APPROACH
- SPECIALIST EXPERTISE REQUIRED

BIG INDUSTRIES
Big Data Consulting

THANKS FOR LISTENING

- ROBERT GIBBON
  - <be.linkedin.com/in/robertgibbon>
- BIG INDUSTRIES - BIG DATA CONSULTING
  - <www.bigindustries.be>