EXECUTIVE BRIEFING:

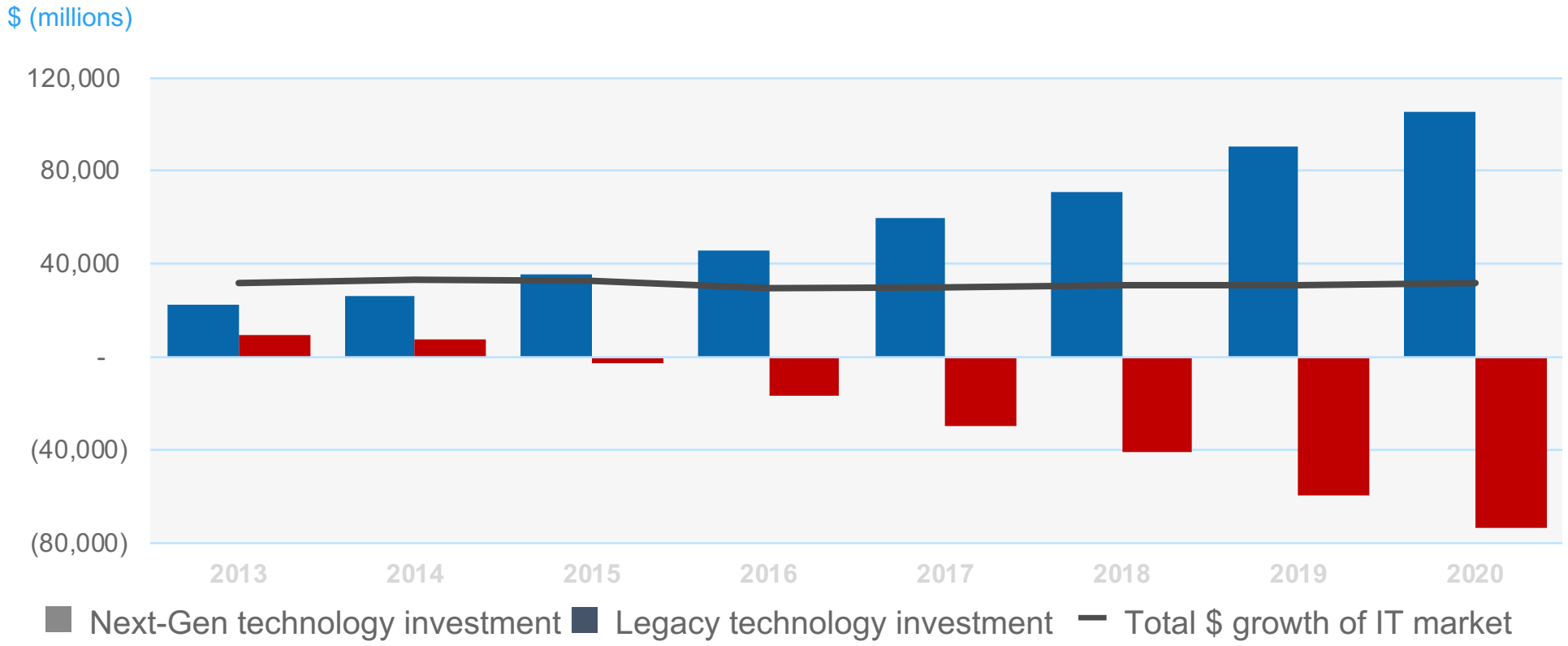
The Hidden Data Scientists Lurking in your Company

Jack Norris, SVP Data and Applications

MapR Technologies

Putting Things into Perspective

INVESTMENT IN NEXT-GEN VS. LEGACY TECHNOLOGIES FOR DATA



Source: IDC, Gartner; Analysis & Estimates: MapR
Next-gen consists of cloud, big data, software and hardware related expenses

© 2019 MapR Technologies, Inc. // MapR Confidential

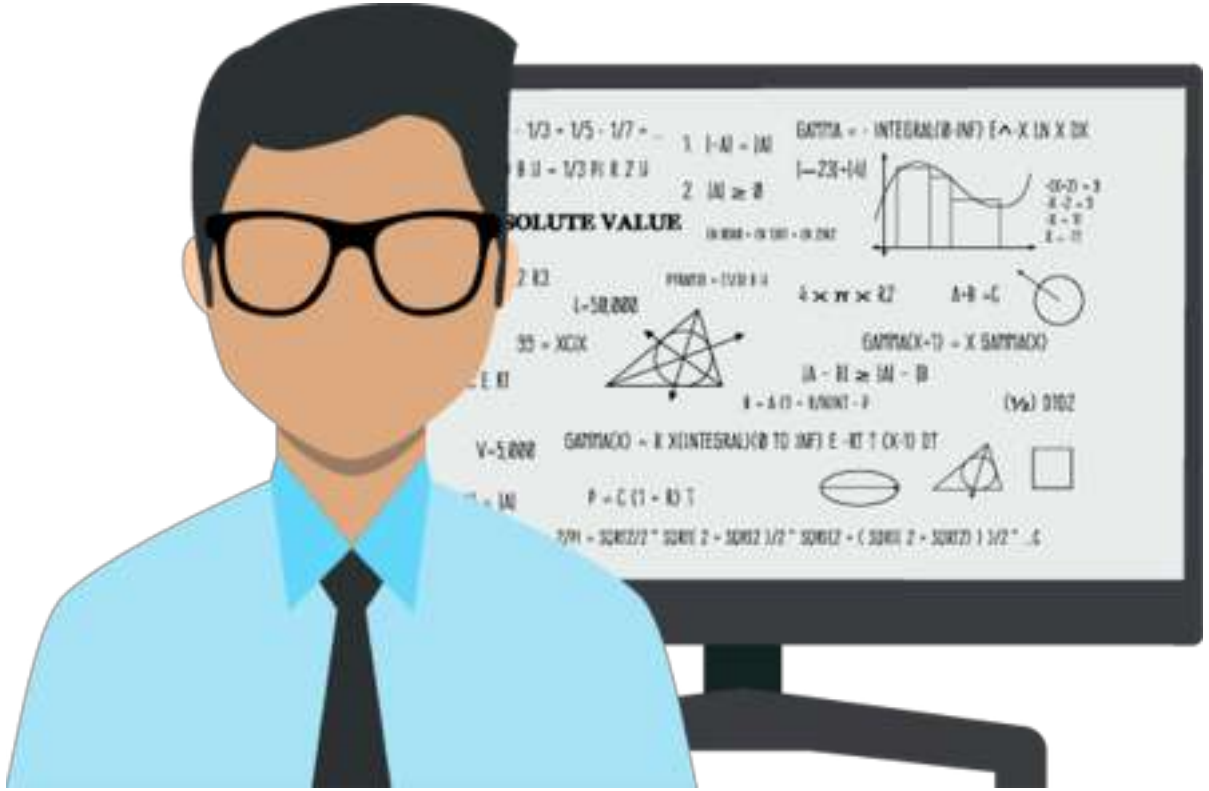
The Impact of AI

McKinsey estimates AI techniques have the potential to create between **$3.5T** and **$5.8T** in value annually across nine business functions in 19 industries.

Forrester Research predicts that by 2020, businesses adopting Machine Learning, AI, and Deep Learning, the Internet of Things (IoT), and Big Data *will take away more than $1.2 trillion from their less-informed peers*.

© 2019 MapR Technologies, Inc. // MapR Confidential

Demand for Data Scientists



29% annual Increase in Demand
344% since 2013

- Indeed Job Stats

In the UK 80% of companies are planning to hire data scientists

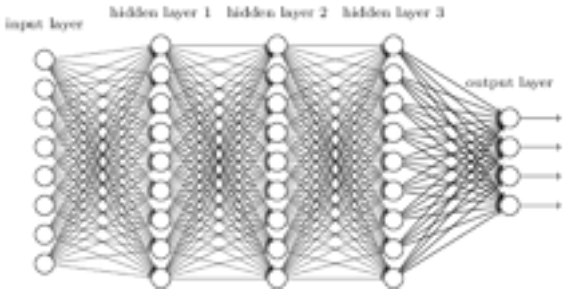
- MHR Analytics

© 2019 MapR Technologies, Inc. // MapR Confidential

Deep Learning Algorithms

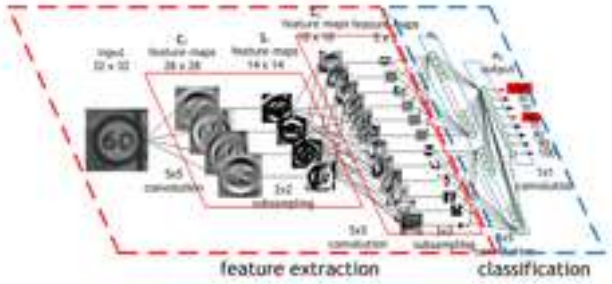
Providing lift for classification and forecasting models

Deep
Neural Networks



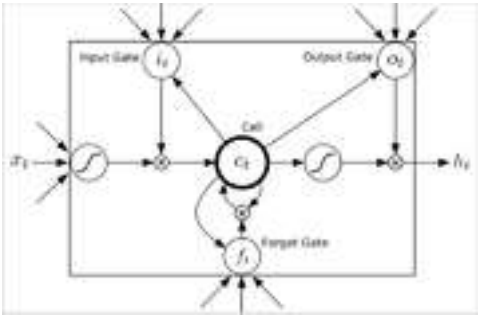
Feature extraction and classification of images

Convolutional
Neural Networks

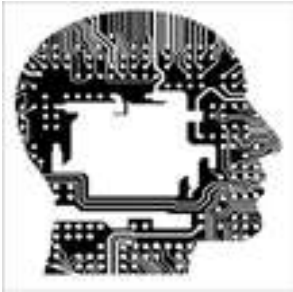


For sequence of events, language models, time series, etc.

Recurrent
Neural Networks



© 2019 MapR Technologies, Inc. // MapR Confidential



ML/AI

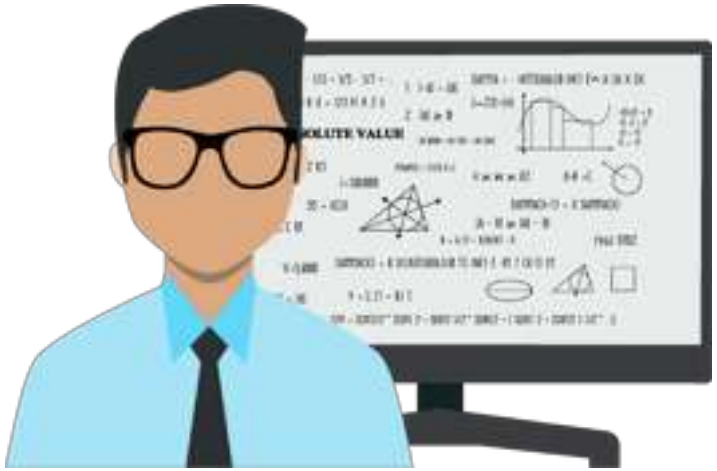
ALGORITHMS

+

DATA

+

DOMAIN EXPERIENCE



© 2019 MapR Technologies, Inc. // MapR Confidential

Focus On Machine Learning Tools

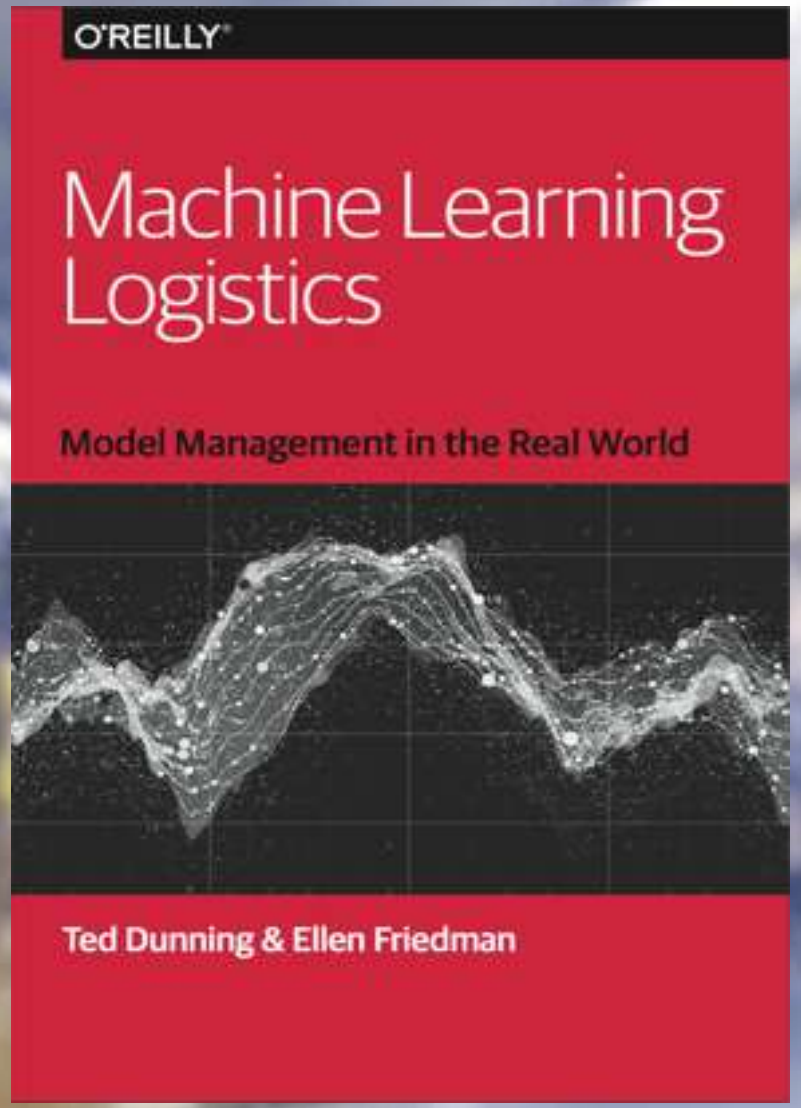
© 2019 MapR Technologies, Inc. // MapR Confidential

Curiosity

© 2019 MapR Technologies, Inc. // MapR Confidential

The Importance of Data

“90+% of Machine Learning Success is Data Logistics”



https://mapr.com/ebook/machine-learning-logistics

© 2019 MapR Technologies, Inc. // MapR Confidential

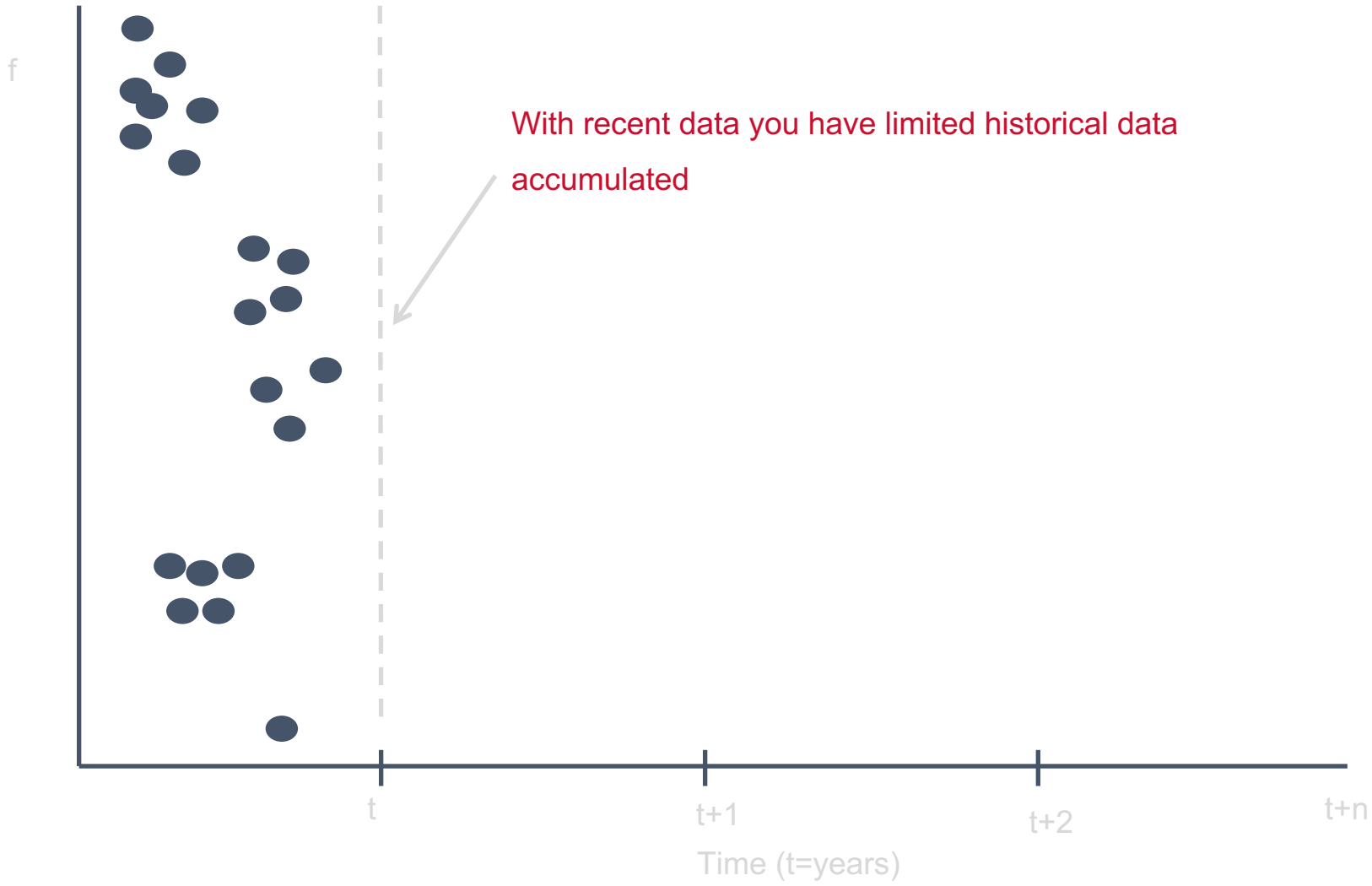
More

DATA

beats complex algorithms

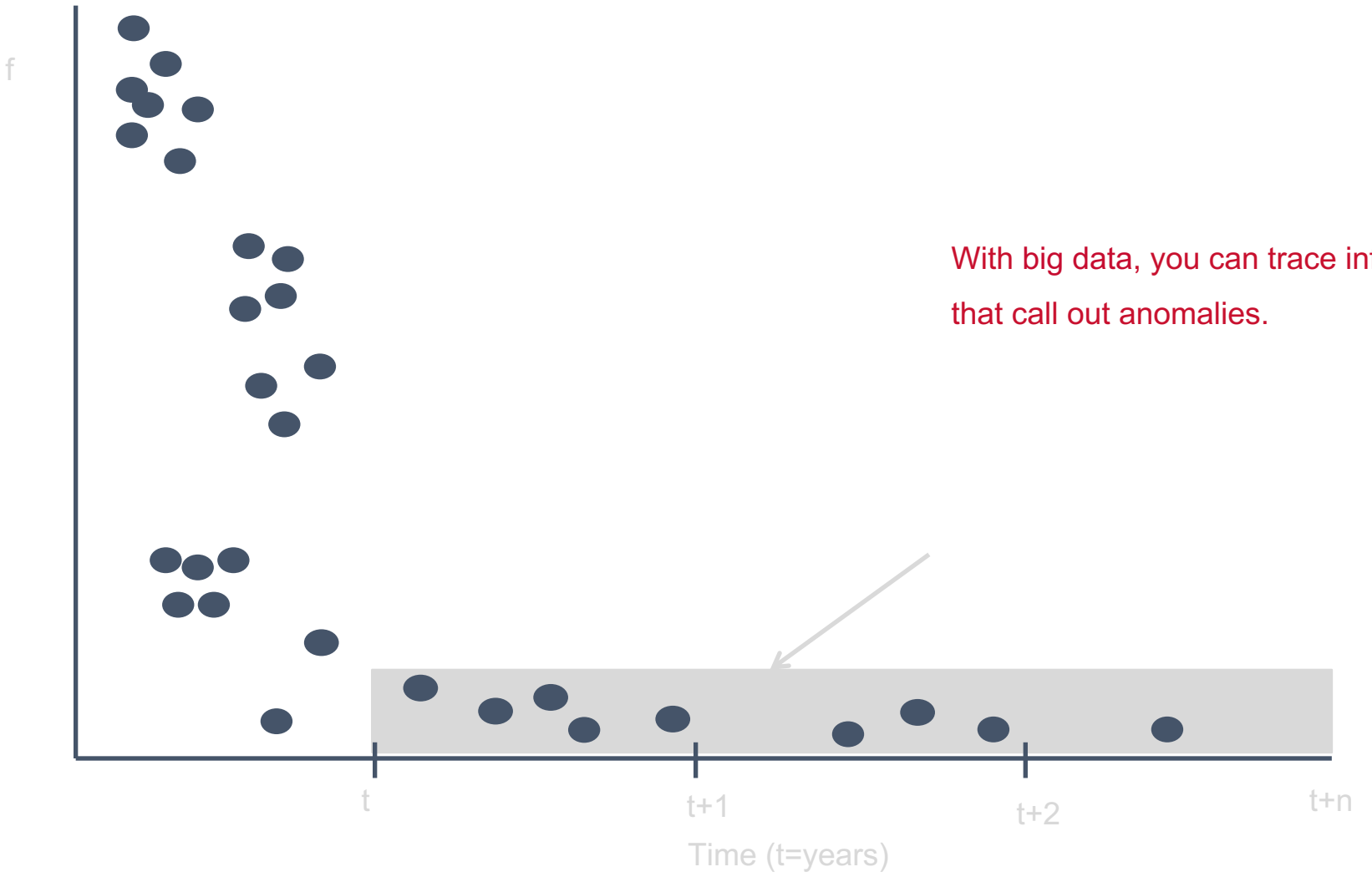
The Unreasonable Effectiveness of Data, published by Google

More Data Allows You to Spot Infrequent Behaviors



© 2019 MapR Technologies, Inc. // MapR Confidential
© 2014 MapR Technologies

More Data Allows You to Spot Infrequent Behaviors



With big data, you can trace infrequent patterns through time that call out anomalies.

© 2019 MapR Technologies, Inc. // MapR Confidential
© 2014 MapR Technologies

How to see beyond the obvious

© 2014 MapR Technologies

How to see beyond the obvious and react quickly

© 2014 MapR Technologies

The Importance of Data Logistics

"Machine learning offers a fantastically powerful toolkit for building complex systems quickly.... **it is remarkably easy to incur massive ongoing maintenance costs** at the system level when applying machine learning."

Machine Learning: The High-Interest Credit Card of Technical Debt

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young
{dsculley,gholt,dgg,edavydov}@google.com
{toddphillips,ebner,vchaudhary,mwyoung}@google.com
Google, Inc

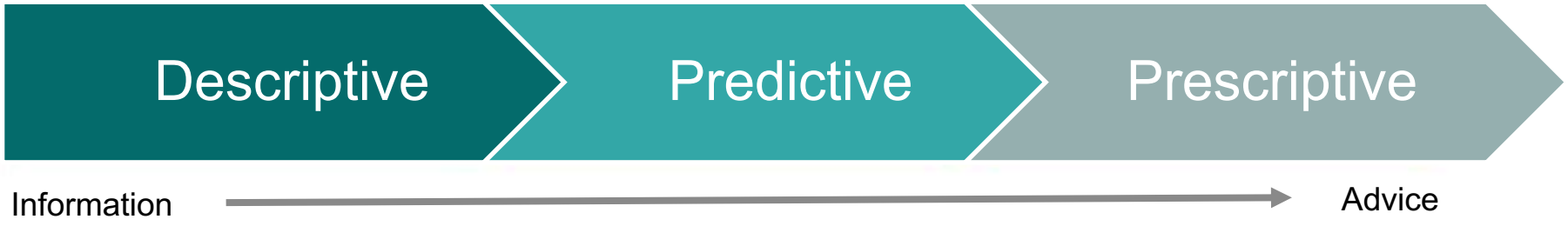
Abstract

Machine learning offers a fantastically powerful toolkit for building complex systems quickly. This paper argues that it is dangerous to think of these quick wins as coming for free. Using the framework of *technical debt*, we note that it is remarkably easy to incur massive ongoing maintenance costs at the system level when applying machine learning. The goal of this paper is highlight several machine learning specific risk factors and design patterns to be avoided or refactored where possible. These include boundary erosion, entanglement, hidden feedback loops, undeclared consumers, data dependencies, changes in the external world, and a variety of system-level anti-patterns.

© 2019 MapR Technologies, Inc. // MapR Confidential

The Impact of Analytics

Evolving Analytics at Scale

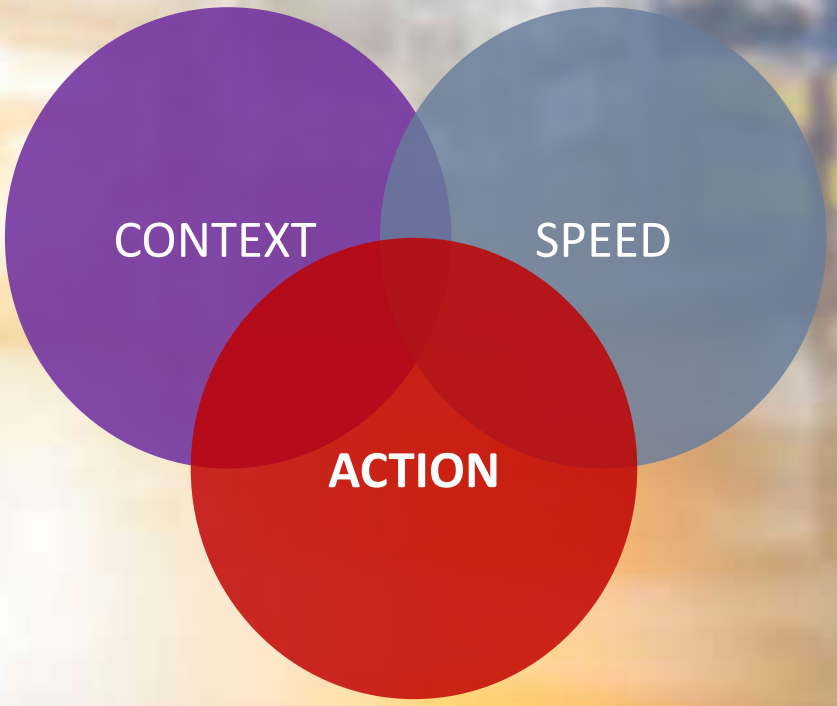


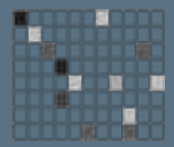
- Descriptive – What happened?
- Predictive – What will happen?
- Prescriptive – The best response to what is happening
  - ➢ Injecting analytics into operations

Global Report on Technology and the Economy, "Flow and the Big Shift in Business Models".

© 2019 MapR Technologies, Inc. // MapR Confidential

Digital Growth Requires Complex Requirements




NoSQL

Document Database

Analytics

Storage

Processing Engines

Streaming

Messaging

© 2019 MapR Technologies, Inc. // MapR Confidential

AMERICAN EXPRESS

‘Implemented big data machine learning use cases: fraud detection & prevention; new customer acquisition & recommendations for better customer experience.’

Chao Yuan, SVP & Head of Decision Science

$1Trillion

Protected Annually from Fraud

< 2ms

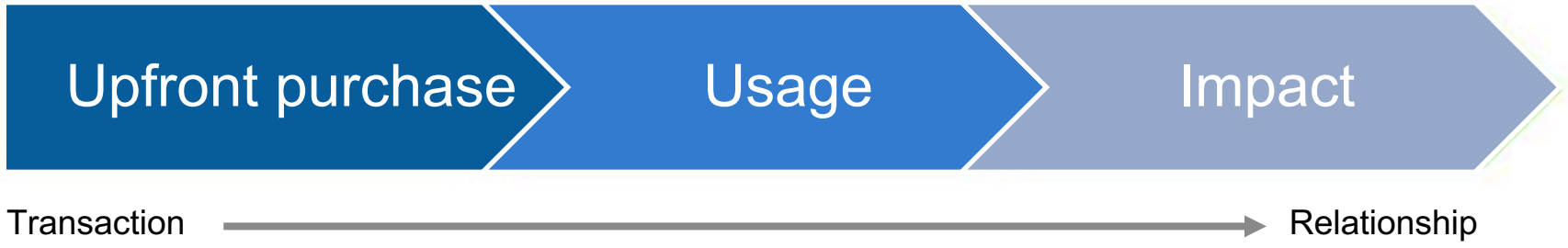
It Takes To Make A Decision

Tüpraş
Machine Learning based predictive maintenance & data intensive applications to reduce operational expenses and increase uptime
200K+
Sensors
300 Billion+
Data Points Evaluated Daily

Automation to Aid not Replace Data Scientist

© 2019 MapR Technologies, Inc. // MapR Confidential

Analytics at Scale: Impact on Business Models



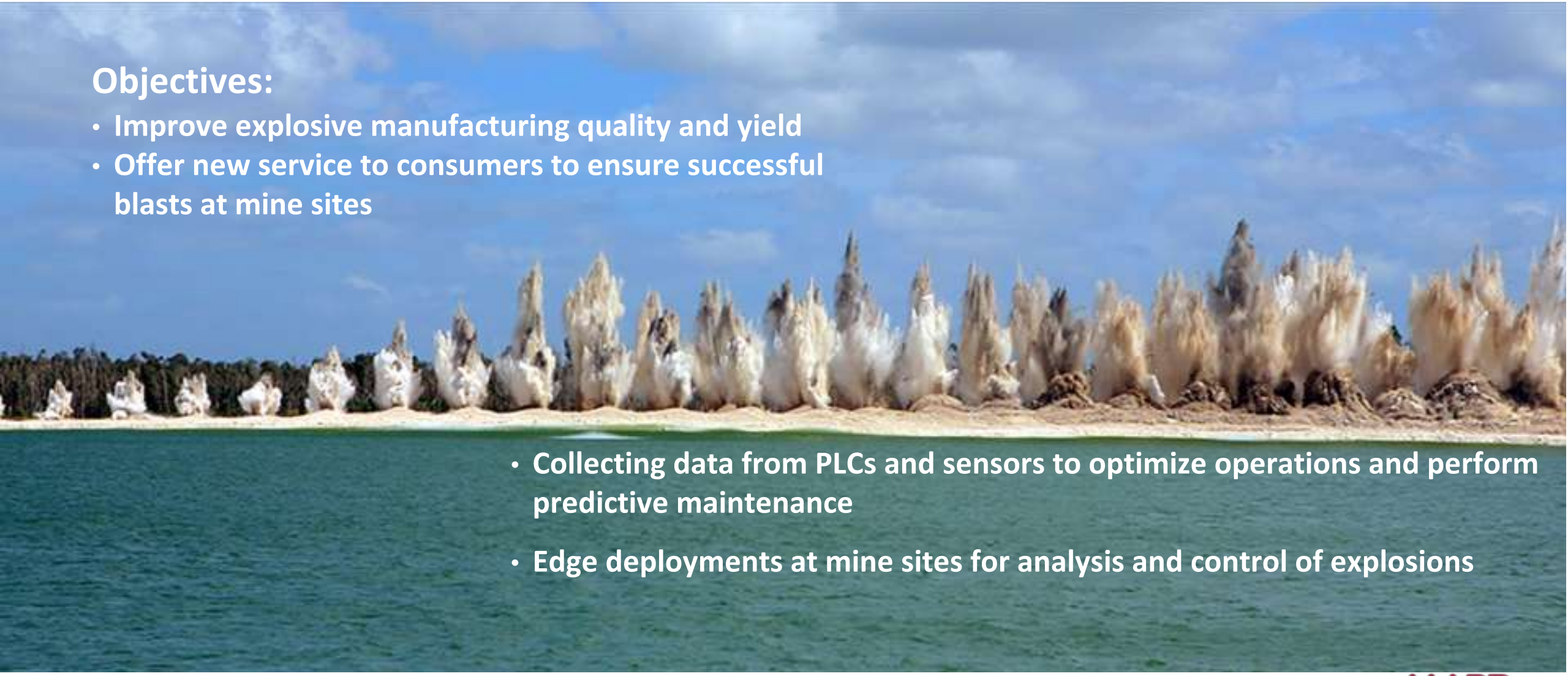
- Moving from upfront payments to usage models
- Paying for the actual value created creates new sources for competitive advantages

Global Report on Technology and the Economy, "Flow and the Big Shift in Business Models".

© 2019 MapR Technologies, Inc. // MapR Confidential

Incitec Pivot Ltd

From Manufacturing Optimization to “Explosions as a Service”



Objectives:

- Improve explosive manufacturing quality and yield
- Offer new service to consumers to ensure successful blasts at mine sites

- Collecting data from PLCs and sensors to optimize operations and perform predictive maintenance
- Edge deployments at mine sites for analysis and control of explosions

© 2019 MapR Technologies, Inc. // MapR Confidential

Training Resources

	Data Engineering		Data Science	
	Core Platform Services	**Big Data Workflows**	**Solution Design**	**Advanced Analytics**
	IT/ Infrastructure	BI / DBA	Development	Modeler / Analyst
ENGAGEMENTS	• Installation • Migrations • SLA Plans • Best Practices • Performance Tuning	• Hive/Pig/Spark • Oozie/Sqoop • Flume • MapR-DB/HBase • Data Pipeline • MapR Streams	• HBase/MapR-DB • Map/Reduce • Application Development • Integration	• Use case Discovery • Use case Modeling • POC • Workshops
SKILLS	Converged Platform Linux Networking Data Center Storage Operations	BI / ETL / Reporting Scripting / Java Hadoop MR Eco Projects (HBase, Hive, …)	Java Hadoop Developer Architectural Design	PhD Statistics/Math MatLab / R / SAS Scripting / Java BI / ETL / Reporting

© 2019 MapR Technologies, Inc. // MapR Confidential

Quick Start Solutions: Speeding Time-to-Value

- Self Service Data Exploration and BI Analytics on Hadoop
- Data Warehouse Offload, Optimization and Analytics
- Financial Services – Fraud Detection, Anti-Money Laundering
- Retail – Customer 360, Social Media Analysis, Recommendation Engine.
- Oil and Gas – Tier 2 Historian
- Complex Event Processing with Drools / Stream Processing
- Time Series Analytics, NoSQL Webstore Applications
- Deep Learning on GPUs for Image Analytics

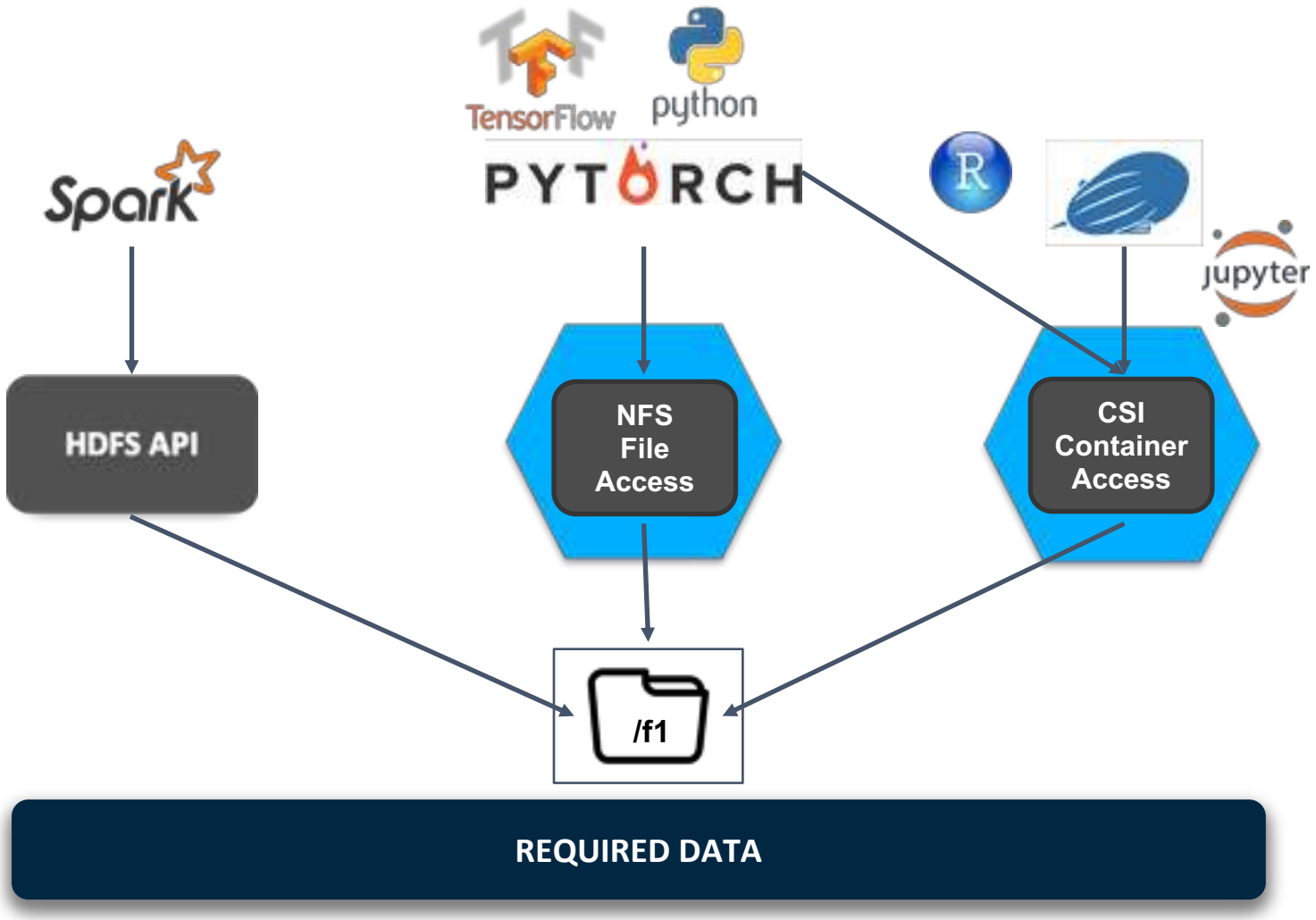
Solution Template

Deployment Architecture

Knowledge Transfer

© 2019 MapR Technologies, Inc. // MapR Confidential

Data Scientist Access To Fast Moving Innovation & Agility



© 2019 MapR Technologies, Inc. // MapR Confidential

The Challenge for IT Organizations

Innovation
Flexibility
Democratization of Data

Control
Security
Protection
Availability

AI , ANALYTICS

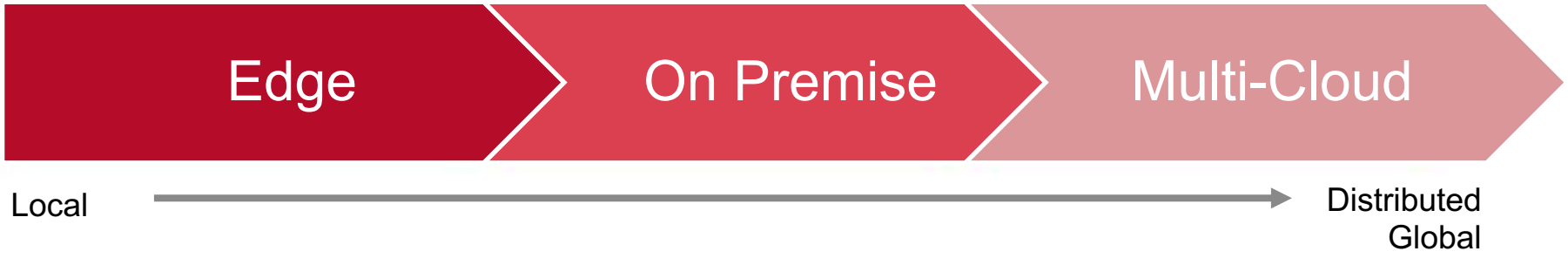
- Customer Engagement
- Fraud detection
- Predictive Maintenance
- Anomaly Detection
- ML/AI data logistics
- BI analytics
- Data science platform
- Recommendation engine

DATA LAYER IS THE LEVERAGE POINT

IoT, EDGE, CLOUD & CONTAINERIZED

- Inter-cloud data/app portability
- Cloud bursting
- Persistent app in containers
- Secure data provisioning
- Data modernization
- Scale out data lakes
- Multi-temperature data
- Data catalogs

Analytics at Scale: Importance of Location



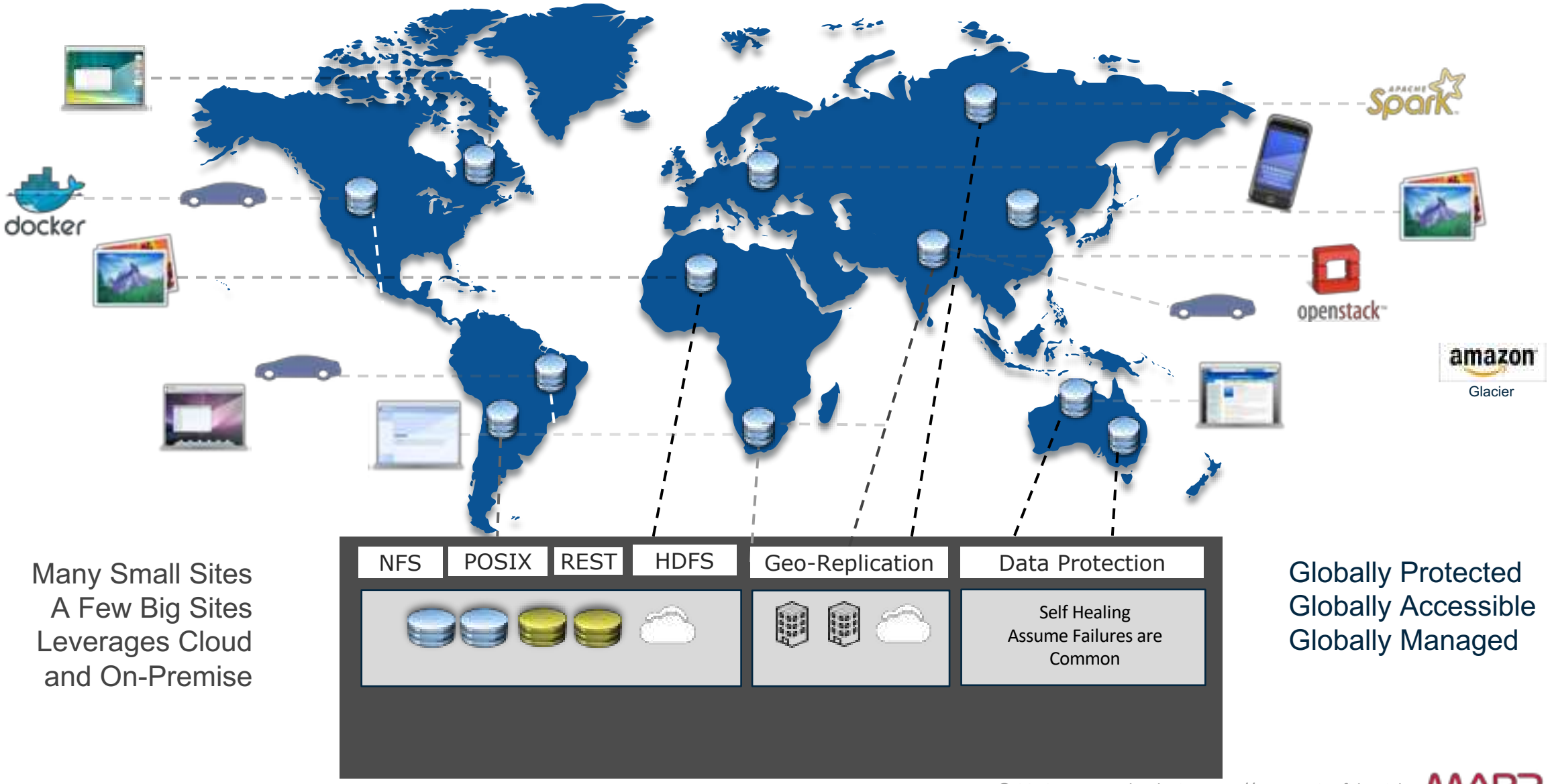
- Edge – Single Location Processing
- On Premise – Centralized Processing
- Multi-Cloud – Fully Distributed

➢ Learning Globally, Acting Locally

Global Report on Technology and the Economy, "Flow and the Big Shift in Business Models".

© 2019 MapR Technologies, Inc. // MapR Confidential

Customers Tell Us They Need A Global Data Fabric

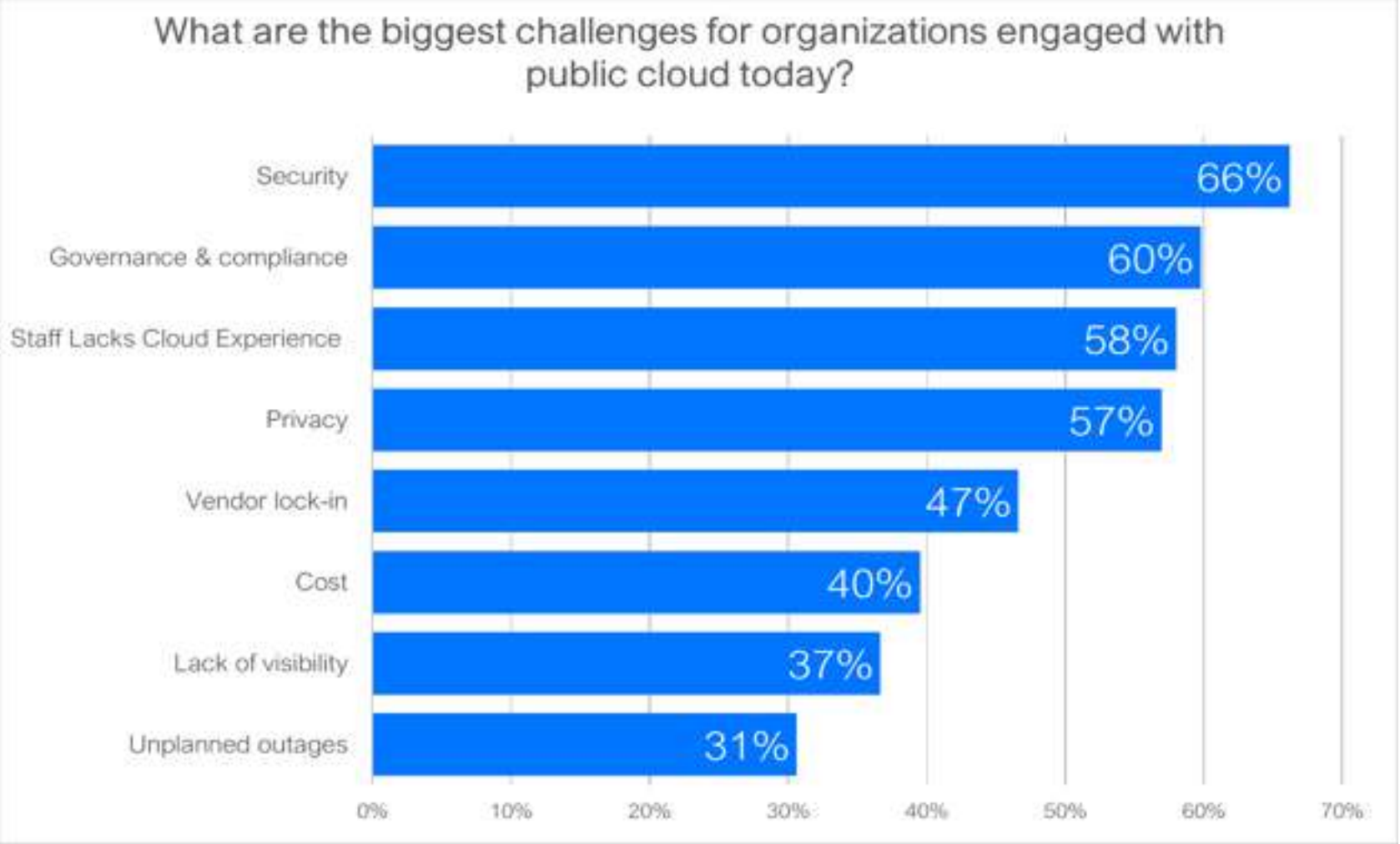


Many Small Sites
A Few Big Sites
Leverages Cloud
and On-Premise

Globally Protected
Globally Accessible
Globally Managed

© 2019 MapR Technologies, Inc. // MapR Confidential

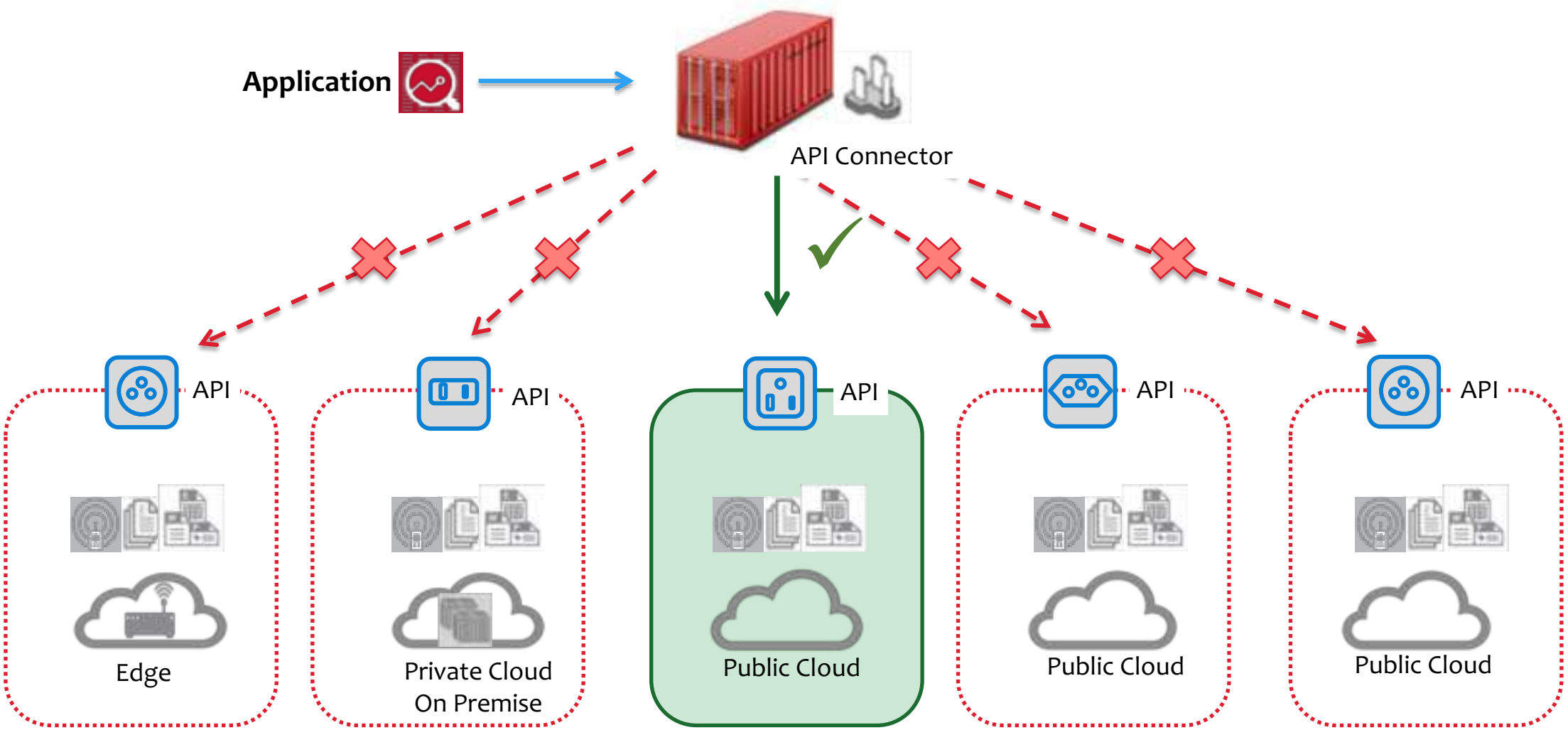
Top Issues for Cloud Today



https://www.logicmonitor.com/wp-content/uploads/2017/12/LogicMonitor-Cloud-2020-The-Future-of-the-Cloud.pdf

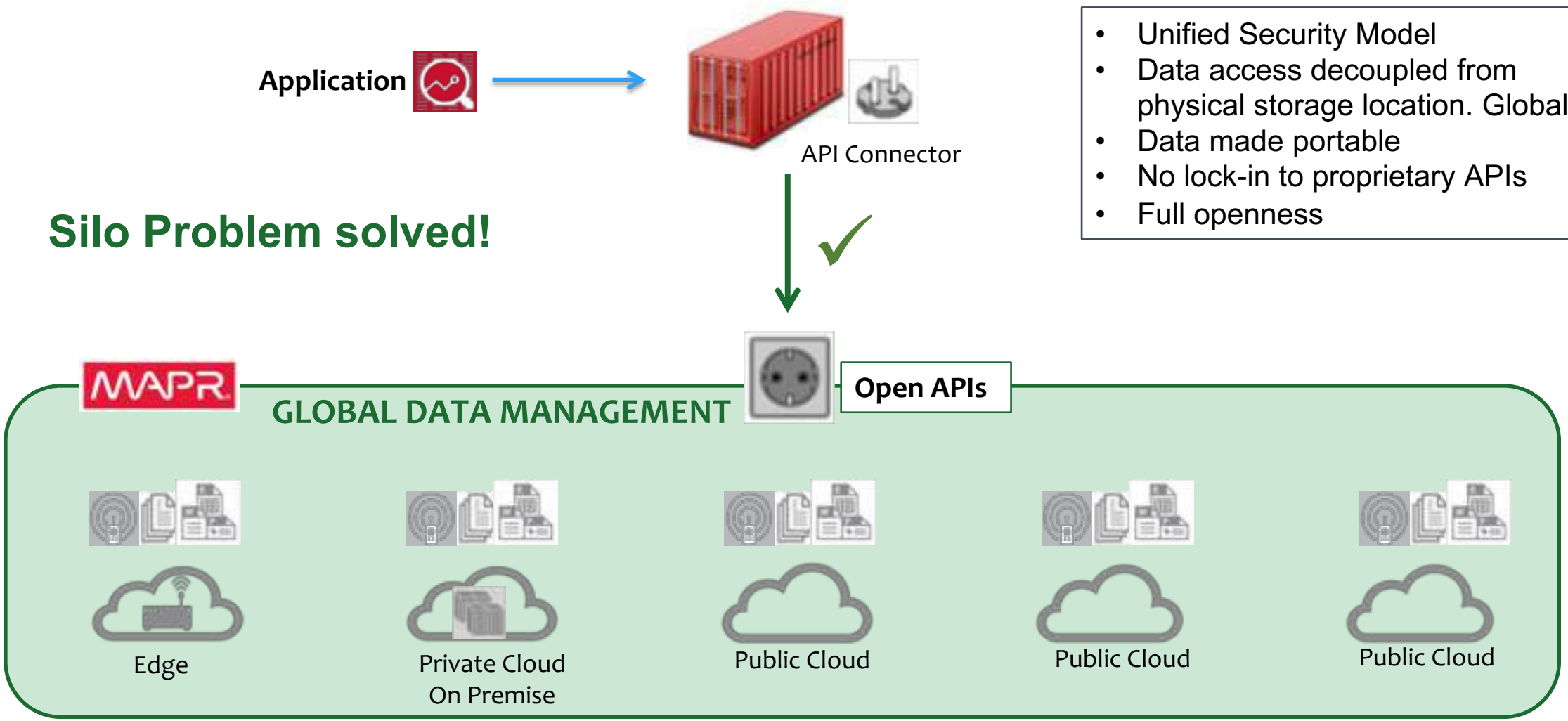
© 2019 MapR Technologies, Inc. // MapR Confidential

Cross Cloud: Difficult to Establish Unified & Secure Data Access



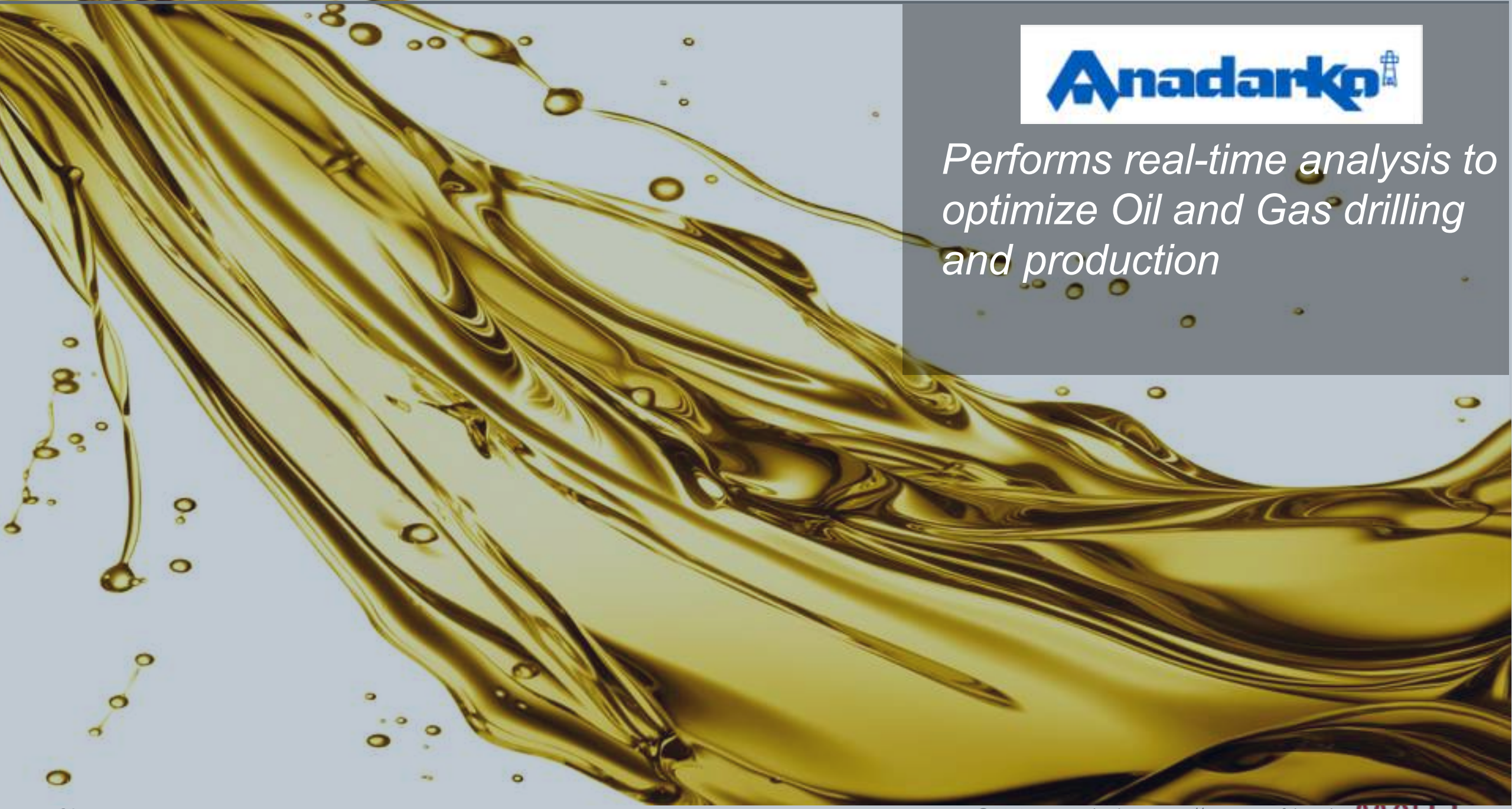
© 2019 MapR Technologies, Inc. // MapR Confidential

Cloud With A Data Platform: Portable, United Access with Security



- Unified Security Model
- Data access decoupled from physical storage location. Globally.
- Data made portable
- No lock-in to proprietary APIs
- Full openness

© 2019 MapR Technologies, Inc. // MapR Confidential

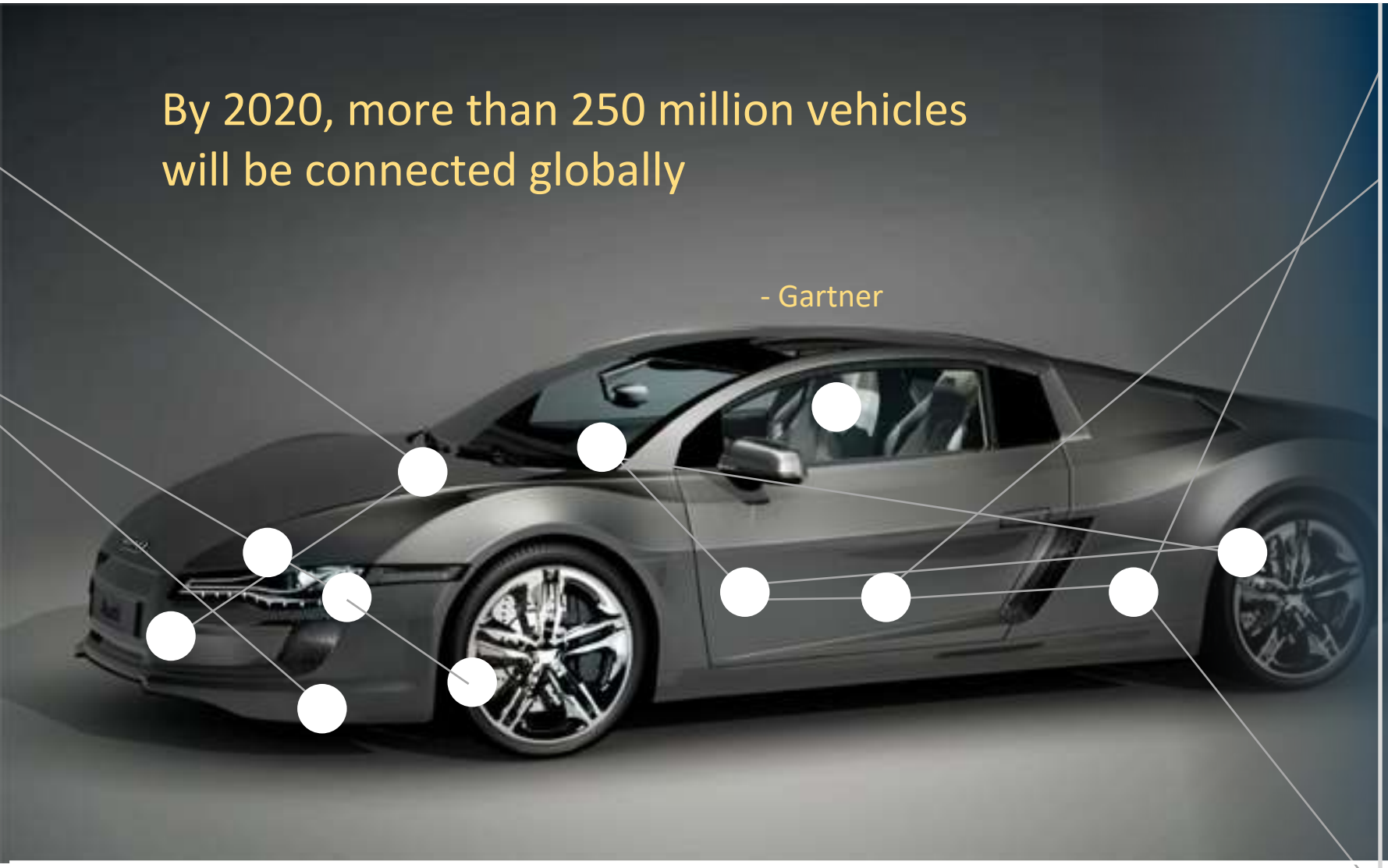


Anadarko

Performs real-time analysis to optimize Oil and Gas drilling and production

© 2019 MapR Technologies, Inc. // MapR Confidential

Connected Car: Driving the Edge Use Cases



By 2020, more than 250 million vehicles will be connected globally

- Gartner

Data Driver for:

Vehicle Efficiency and Performance

Personalized Experience

Manufacturing Optimization

Driver Assisted and Driverless Vehicles

© 2019 MapR Technologies, Inc. // MapR Confidential

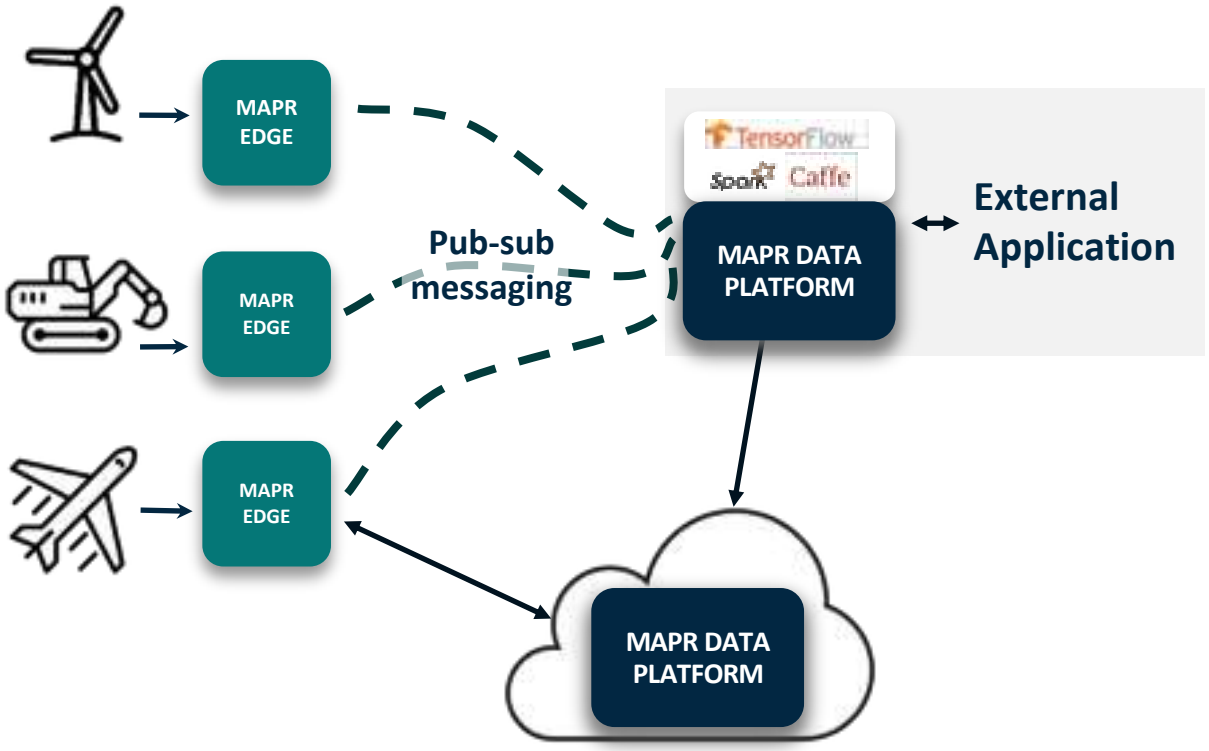
Requires Ability to Learn Globally and Act Locally

REQUIREMENTS

- Collect data in full fidelity
- Apply the latest, most accurate models – even in areas of low bandwidth, high latency, or space constraints
- Avoid the patchwork of silos that comes with historian systems

COORDINATED DATA FLOWS

- Pub-sub data streaming:
  - Data ingestion from IoT devices and
  - Edge to cloud data communication where streams are persisted for centralized analysis
- Extends the platform to the edge, allowing for applications and models to run at the edge (required for time-sensitive applications)

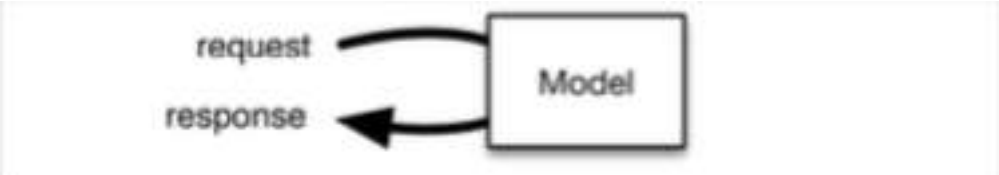


© 2019 MapR Technologies, Inc. // MapR Confidential

How to Unleash your Lurking Data Scientists

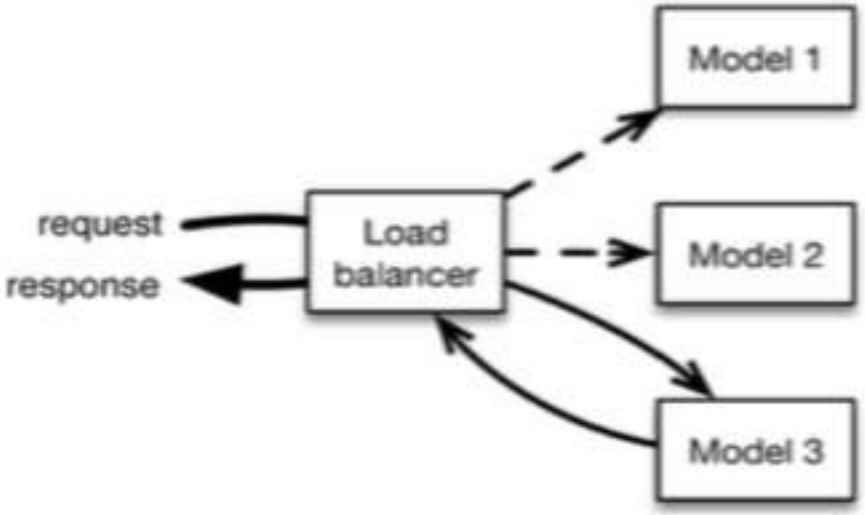
Machine Learning Logistics

Discrete Response System



Move to Support Multiple Models for accuracy, separation of concerns

Load Balancer Approach

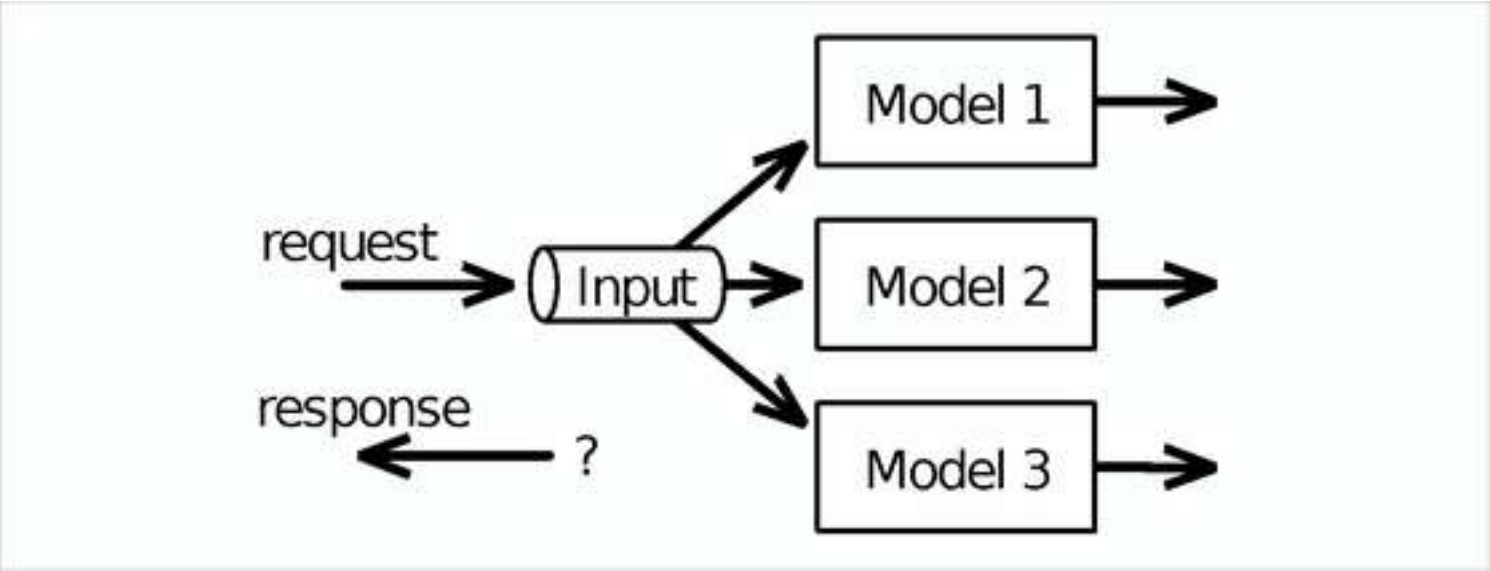


With a load balancer, you can start and stop new models pretty easily, you lack:

- Latency guarantees,
- Ability to compare models on identical inputs,
- Records of all of the requests with responses from all live models.

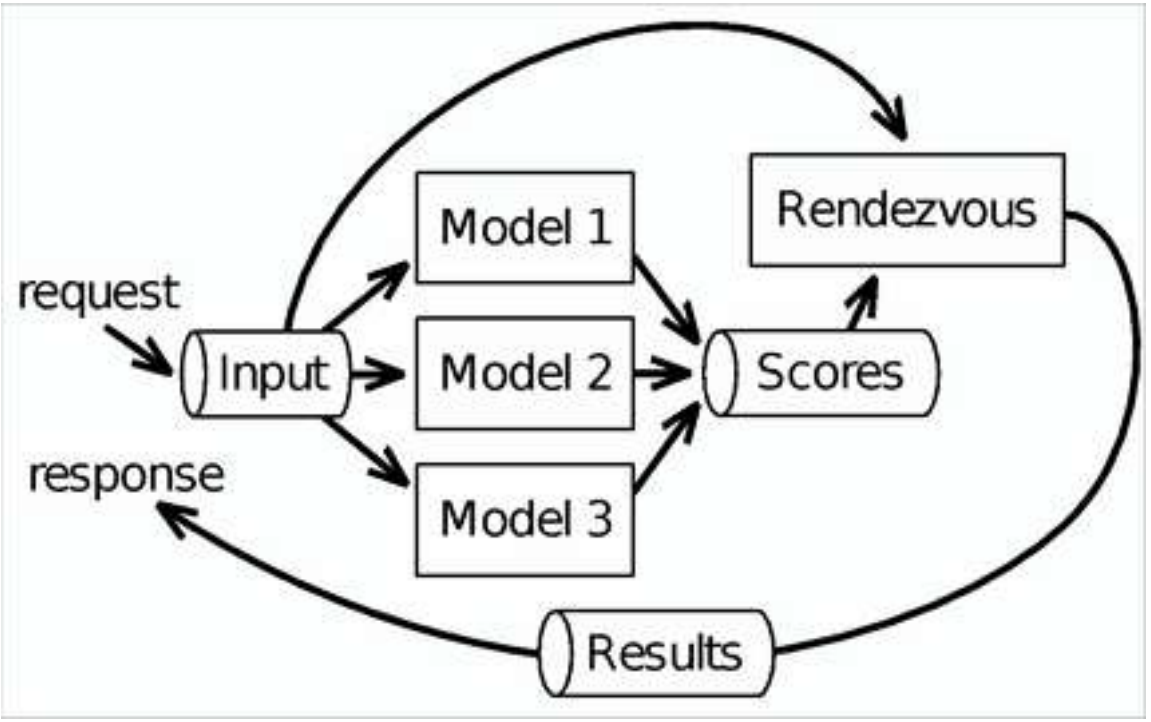
© 2019 MapR Technologies, Inc. // MapR Confidential

Stream-based Logistics



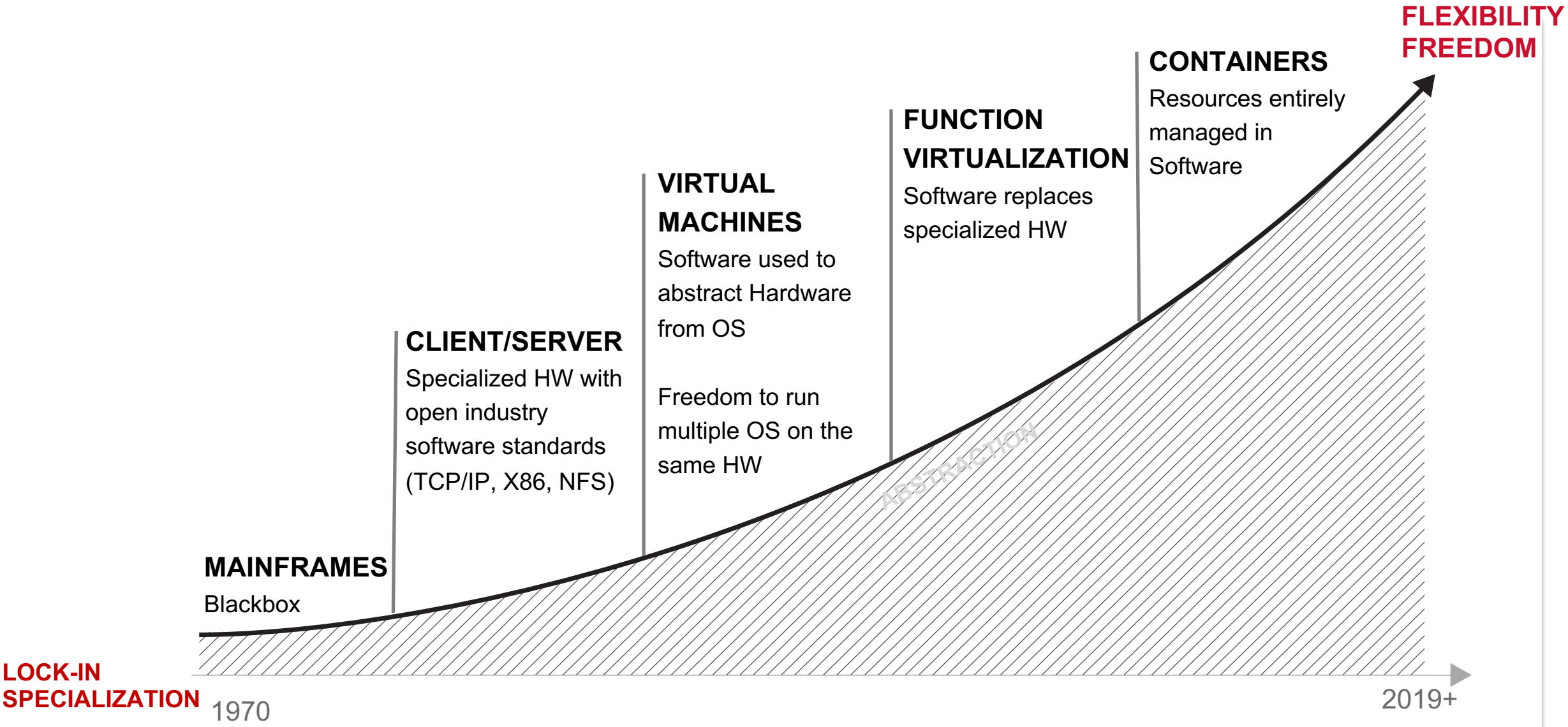
© 2019 MapR Technologies, Inc. // MapR Confidential

Rendezvous Architecture



© 2019 MapR Technologies, Inc. // MapR Confidential

Abstraction: a Short History of IT



© 2019 MapR Technologies, Inc. // MapR Confidential

Data is a resource as well

Software

Middleware

Hardware

© 2019 MapR Technologies, Inc. // MapR Confidential

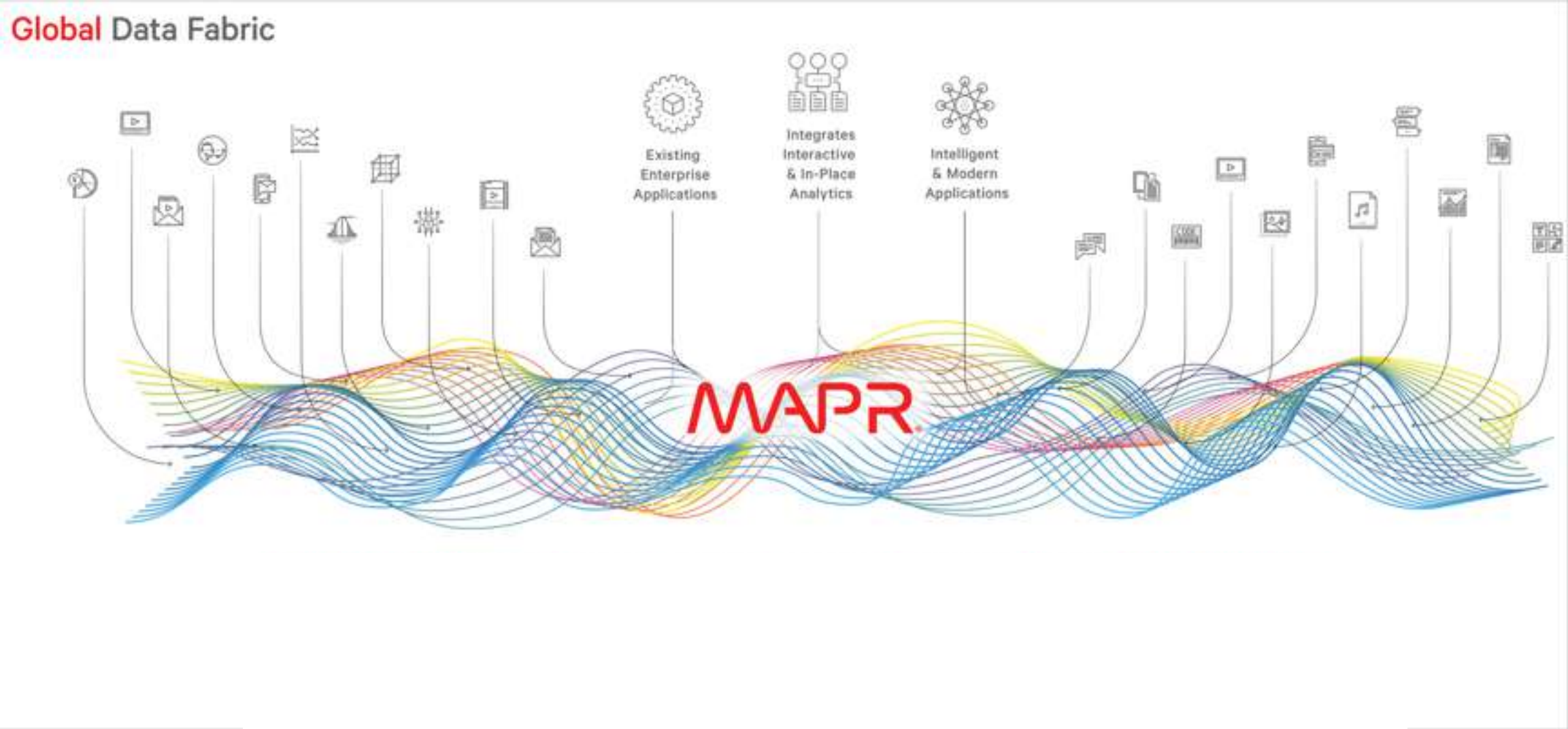
Turning Data into a Manageable Resource
All in Software

- Data Containerization
- Global Multi-Tenancy
- Data Portability
- Resource Isolation
- Workload independence
- Security
- Global Web-Scale Deployments
- Performance
- Universal Access

All Managed by Policies in One Layer

© 2019 MapR Technologies, Inc. // MapR Confidential

Global Data Fabric Supports Existing and New Applications including AI



© 2019 MapR Technologies, Inc. // MapR Confidential

The Solution for Organizations to Support Analytics

Innovation
Flexibility
Democratization of Data

Control
Security
Protection
Availability

AI , ANALYTICS

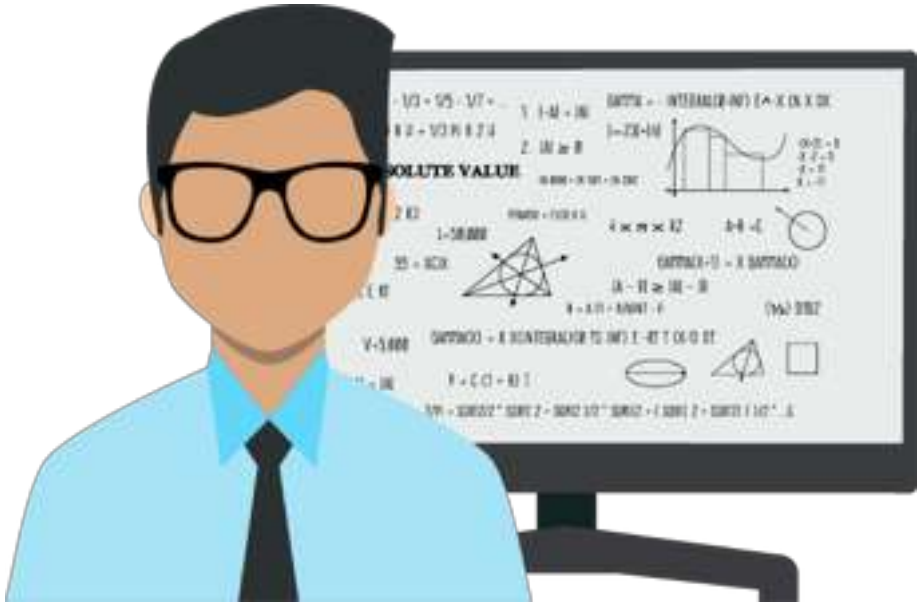
- Customer Engagement
- Fraud detection
- Predictive Maintenance
- Anomaly Detection
- ML/AI data logistics
- BI analytics
- Data science platform
- Recommendation engine

DATAWARE FOR DATA-DRIVEN TRANSFORMATION

IoT, EDGE, CLOUD & CONTAINERIZED

- Inter-cloud data/app portability
- Cloud bursting
- Persistent app in containers
- Secure data provisioning
- Data modernization
- Scale out data lakes
- Multi-temperature data
- Data catalogs

Finding and Empowering Your Data Scientists



ML/AI

ALGORITHMS

+

DATA

+

DOMAIN EXPERIENCE

© 2019 MapR Technologies, Inc. // MapR Confidential

Meetup Tonight

Online Evaluation of Machine Learning Models

Ted Dunning PhD, MapR CTO

The Microsoft Reactor London
70 Wilson St
Finsbury
EC2A 2DB
United Kingdom

Thursday, May 2, 2019 from 7:00 PM to 8:30 PM (BST)

© 2019 MapR Technologies, Inc. // MapR Confidential

Rate today ’s session

Executive Briefing: The hidden data scientists lurking in your company

Jack Norris (MapR Technologies)

14:05-14:45 Thursday, 2 May 2019

Executive Briefing and best practices, Strata Business Summit

Location: Capital Suite 13

Secondary topics: AI and machine learning in the enterprise

Rate This Session

Session page on conference website

© 2019 MapR Technologies, Inc. // MapR Confidential

Questions?

MAPR