



Keeping The Finger On Your Business Pulse @ Scale with Big Data

Jin-Goo Lee
Country Manager, MapR Korea
jinlee@mapr.com

MAPR
DATA TECHNOLOGIES



40% of today's Fortune 500 companies on the S&P 500 will no longer exist in 10 years.

Fortune Top 10 Companies

1998

[General Motors](#)

[Ford Motor](#)

[Exxon Mobil](#)

[Wal-Mart Stores](#)

[General Electric](#)

[Intl. Business Machines](#)

[Chrysler](#)

[Mobil](#)

[Altria Group](#)

[AT&T](#)

2008

[Wal-Mart Stores](#)

[Exxon Mobil](#)

[Chevron](#)

[General Motors](#)

[ConocoPhillips](#)

[General Electric](#)

[Ford Motor](#)

[Citigroup](#)

[Bank of America](#)

[AT&T](#)

2018

[Walmart](#)

[Exxon Mobil](#)

[Berkshire Hathaway](#)

[Apple](#)

[UnitedHealth Group](#)

[McKesson](#)

[CVS Health](#)

[Amazon.com](#)

[AT&T](#)

[General Motors](#)

2018 Market Capitalization Ranking

Q1	Q2
Apple Inc.	Apple Inc.
Alphabet Inc.	Amazon.com
Microsoft	Alphabet Inc.
Amazon.com	Microsoft
Tencent	Facebook
Berkshire Hathaway	Tencent
Alibaba Group	Alibaba Group
Facebook	Berkshire Hathaway
JPMorgan Chase	JPMorgan Chase
Johnson & Johnson	ExxonMobil

https://en.wikipedia.org/wiki/List_of_public_corporations_by_market_capitalization#2018



Data

Power of Digital & Data Networks



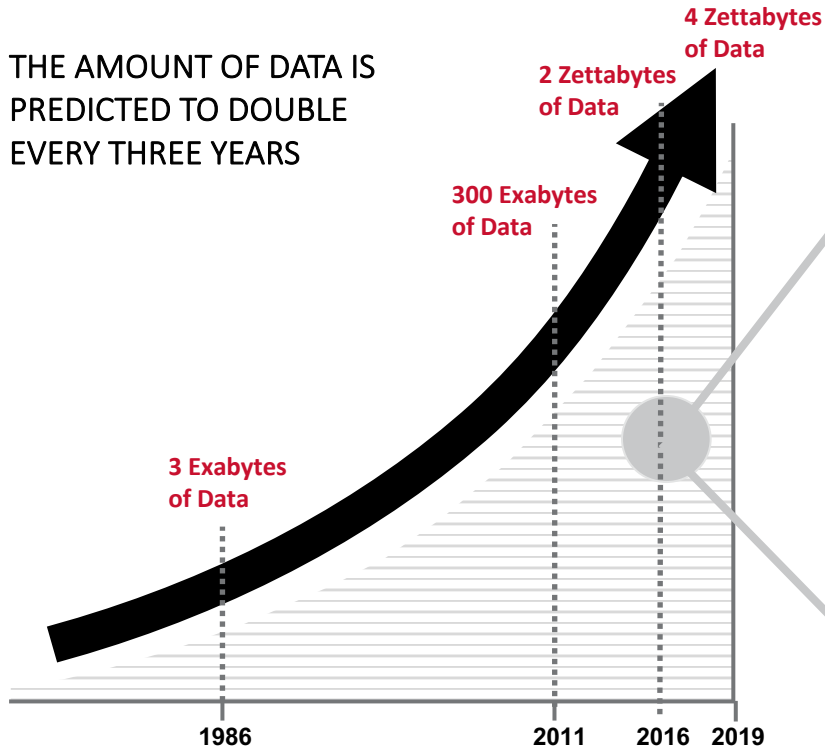
The background features a dark blue gradient with glowing orange and yellow light trails that suggest data flow. A semi-transparent pie chart is visible in the center, and binary code (0s and 1s) is scattered throughout the scene.

DIGITAL BUSINESS

extract & capture new value from data @ scale

Data Growth & Diversity Puts Premium On Efficient Leverage

THE AMOUNT OF DATA IS PREDICTED TO DOUBLE EVERY THREE YEARS



DATA DIVERSITY

A grid of 24 icons representing various data types. The icons are arranged in four rows and six columns. The top row includes Call Detail Records, Click stream, CSV, Data, Documents, and Emails. The second row includes Billing Data, Meta, JSON, Mobile, Network, and PDF. The third row includes Product Catalog, Medical Records, Text Files, Text Messages, Video, and XML. The bottom row includes Sensor Data, Server Logs, Set Top Box, Audio, Social Media, and Merchant Listings. A dark blue horizontal bar is overlaid across the middle of the grid, containing the text: "If it is ALL about the data, then it better be about ALL your data."

Source: McKinsey Global Institute: "The Age of Analytics", Dec. 2016



Nap Time



DEEPER CONTEXT @ SCALE

A high-speed photograph of a single water droplet falling into a pool of water, creating a series of concentric ripples. The background is a soft, out-of-focus blue-grey. The text is overlaid in white, bold, sans-serif font.

CAPTURING VALUE @ SCALE

TIMELY & RELEVANT

MapR Technologies

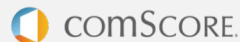
COMPANY

Silicon valley
software company

Preferred choice
for production AI &
analytics deployments

Better value,
lower TCO

SUCCESSFUL CUSTOMERS

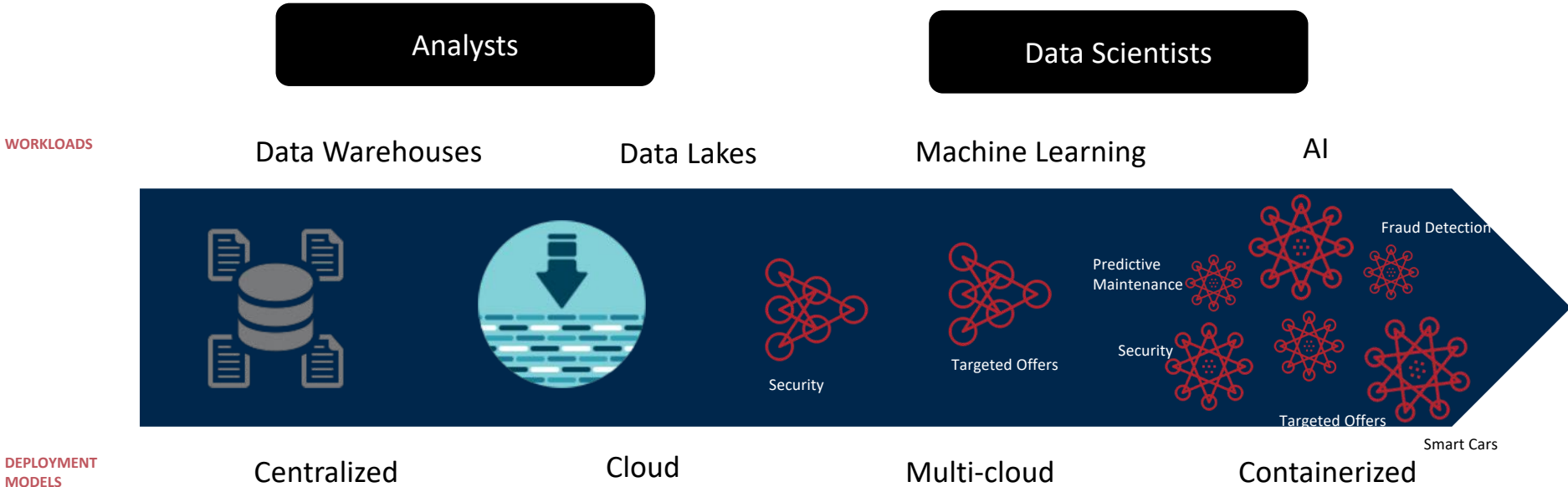


ACCLAIMED INNOVATOR

AI, Machine Learning
& Analytics at scale
IoT & Edge Analytics
Reliable distributed
file & object store

The Trend From Analytics to Active AI & Data Science

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.



A Powerful Founding Vision

VISION: A POWERFUL INNOVATIVE DATA PLATFORM

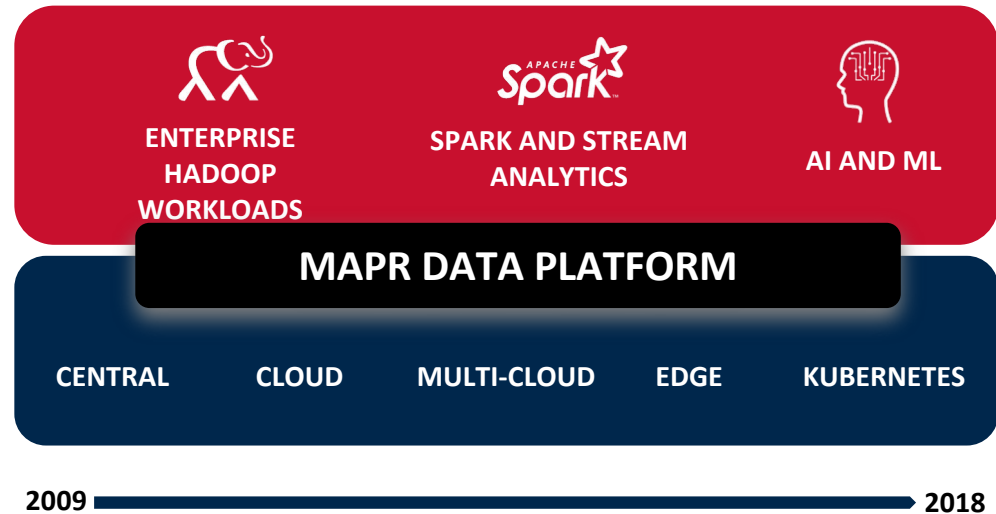
- Ingest, store, manage and use many data types
- Execute wide range of analytical and AI tools

MISSION CRITICAL CAPABILITIES

- Patented, superior distributed technology platform
- High scale, high reliability, efficient architecture

VALUE AT PRODUCTION SCALE

- Large global enterprises rely on MapR



MapR Delivers Value To Leading Companies

HOW CUSTOMERS APPLY MAPR

- Speed up AI, Analytics initiatives for more impact
- Create highly reliable scalable data fabric
- Use data streams for real time analytics

BUSINESS VALUE

- Proven to deliver transformative business benefits
- Increase revenue with higher impact of analytics
- Reduce cycle time of business processes

SUPERIOR TECHNOLOGY

- Patented software with a highly distributed architecture
- More scalable and reliable enterprise grade platform
- Runs Hadoop, Spark, ML, AI tools better

PRODUCTION SCALE CUSTOMERS

Amex
UHG
Nordea
KDDI
Ericsson
Philips
REWE
Cisco
Charles Schwab
BNP Paribas
NTT

MapR Delivers Business Value On Multiple Dimensions

GROW REVENUE

- Mega retailer: Created \$1B more revenue
- Schwab: 30% better customer engagement
- 9 Entertainment: better forecasts through ML
- REWE: Better customer engagement

REDUCE COST

- Indian Government: \$10B in avoided fraud
- United Healthcare: Saves \$150K/day
- F100 Bank: 46x savings of \$2.7M/yr

GO FASTER

- NTT: Processes 2x more volume
- Paysafe: Sped up processing time from 12hrs to 30s
- Qualcomm: 92% faster processing
- Cisco: 40x more message throughput

REDUCE RISK

- Security service providers: Apply data science to security & advanced evolving threats
- Major banks: Avoid compliance fines in banking
- India: Operate largest biometric DBMS

Big Data use cases

ADVERTISING, MEDIA & ENTERTAINMENT	FINANCIAL SERVICES	GOVERNMENT	HEALTHCARE & LIFE SCIENCES
<ul style="list-style-type: none">• Personalized recommendations• Enhanced game player engagement	<ul style="list-style-type: none">• Fraud detection• Credit risk assessment• 360 customer service	<ul style="list-style-type: none">• Cybersecurity, intelligence• Crime prediction and prevention• Defense, national security	<ul style="list-style-type: none">• Personalized treatment planning• Assisted diagnosis• Monitor patient vital signs
MANUFACTURING	OIL & GAS	RETAIL	TELECOM
<ul style="list-style-type: none">• Preventative maintenance• Monitoring product quality• Market pricing and planning	<ul style="list-style-type: none">• New oil prospect identification• Equipment maintenance• Safety and environment	<ul style="list-style-type: none">• Up-sell/cross-sell recommendations• Dynamic pricing across multiple channels• Clickstream analysis	<ul style="list-style-type: none">• Customer churn analysis• Recommendations• Network management/optimization

MapR Customers Have 5x The Success Rate For Going Into Production

- A data platform, not components
- Runs Hadoop, Spark, AI, ML on the same cluster
- Enterprise grade reliability built in
- HA, DR, data protection built-in
- Common platform wide security model
- POSIX & NFS easily integrate enterprise apps

MapR Enabling Data-Driven Business Transformation



**Data
Storage
Efficiency**



**Analytics On
More Data**



**Intelligent
Processes &
Products**



**Proactive
Actions with
Predictive
Insight**



**Machine Driven
Decisions &
Optimisation**

MapR Converged Data Platform

Keeping The Finger On Your Business Pulse @ Scale

A hand is shown in the lower-left corner, pointing towards the right. A glowing blue ECG line starts from the hand and extends across the top of the image. The background is dark blue with some bokeh light effects.

1. Importance Of Data Flows
2. Leverage An Emerging Data Fabric
3. Injecting Intelligence Into Operations



What's your data strategy? May-June 2017
Leandro DalleMule & Thomas H. Davenport

‘ The ‘plumbing’ aspects of data management may not be as sexy as the predictive models and colorful dashboards they produce, but they’re vital to high performance.

As such, they’re not just the concerns of the CIO and the CDO; ensuring smart data management is the responsibility of all C-suite executives, starting with the CEO.

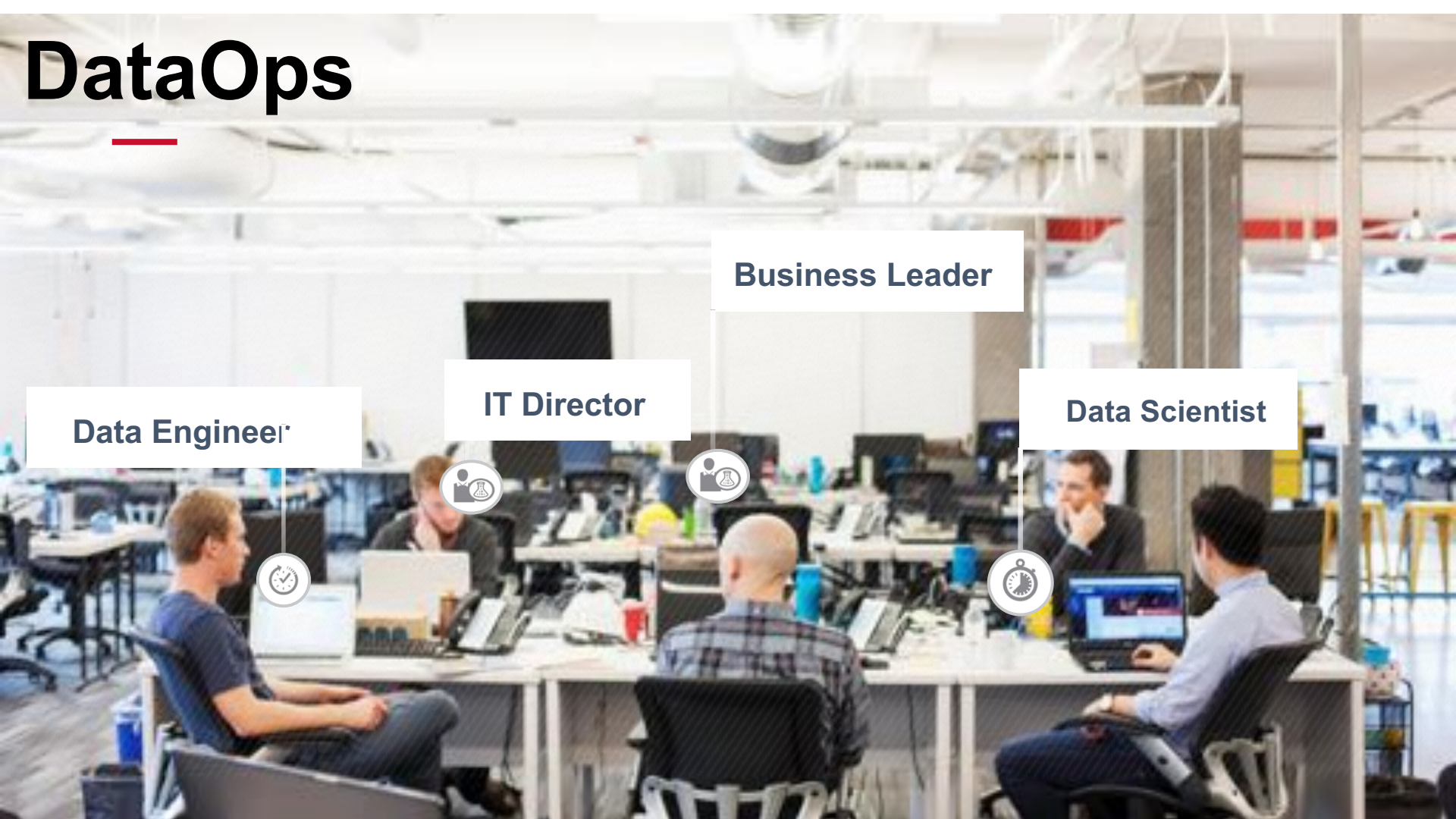
DataOps

Business Leader

IT Director

Data Scientist

Data Engineer



Importance Of Data Flows

- Data from business as it happens
- Existing applications & business processes
- New data flows from 'things' or 'products'
- Internal & external data





Data Flows Start With One 'Business Event' At A Time



"time" : "6:01.103",
"event" : "RETWEET",
"location" :
 "lat" : 40.712784,
 "lon" : -74.005941



"time": "5:04.120",
"severity": "CRITICAL",
"msg": "Service down"



"card_num" : 1234, "merchant"
: "Apple", "amount" : 50

Data Flows Are Key To Impact Business While It Is Happening

A flowing stream of water, likely from a waterfall or a dam, is shown in a dynamic, slightly blurred state. The water is a vibrant blue color and is moving from left to right. Several text boxes are overlaid on the water, each containing a different type of data source. The boxes are light blue with dark blue text and a thin dark border. The text boxes are: 'web events' (top left), 'machine sensors' (center), 'biometrics' (bottom right), and 'mobile events' (top right). The text 'etc...' is also present in the top right corner, indicating that there are more data sources beyond those listed.

web events

etc...

mobile events

machine sensors

biometrics



United Health Group

- Big Data Platform as a Service for 50+ business units
- Data lake for diverse health records, claims, hospitals, member info, call center, web traffic
- Collaborative environment for data scientists
- Graph analytics of patient-provider network
- Preventing claims fraud (captured 10% fraud in early use case, over 1.8 million claims/day)

Business value: Reduces healthcare fraud, improves customer service, reduces cost of creating innovative new services

Transforming the Online Digital Shopping Experience



Real-time Integrated Pub/Sub Streams to capture

- Stocking changes
- Orders
- Cancellations from various systems
- Process events in real-time for latest stock availability

Critical consistent SLAs for stock availability checks

- < 50 ms at any Large Scale

Global Multi-Master replication for serving users across the world

NO downtime

Lower TCO: Migrate from expensive RDBMS (DB2)

Modernize online digital shopping experiences for customers through robust acquisition, processing and serving data to customers in real-time while reducing costs

Real-Time Metrics & Predictive Maintenance



300B Data Points Daily



- ✓ Industry 4.0 transformation – ML for Predictive Maintenance
- ✓ 200K sensors, giving real-time access to refinery process data
- ✓ Increasing data collection frequency from 30s to 1s



Leverage an emerging data fabric

- Network Effect – diversity of applications & analytics on a common data platform
- Supports both data-in-motion and data-at-rest
- Stretches from the edge to the cloud



Global Data Fabric – The Common Data Platform

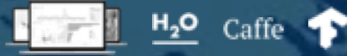
Existing Enterprise Applications



Data Processing



Business & Machine Intelligence



Modern Data Intensive Applications



GLOBAL SECURITY

GLOBAL ACCESS

GLOBAL FLOWS

GLOBAL MANAGEMENT



Files

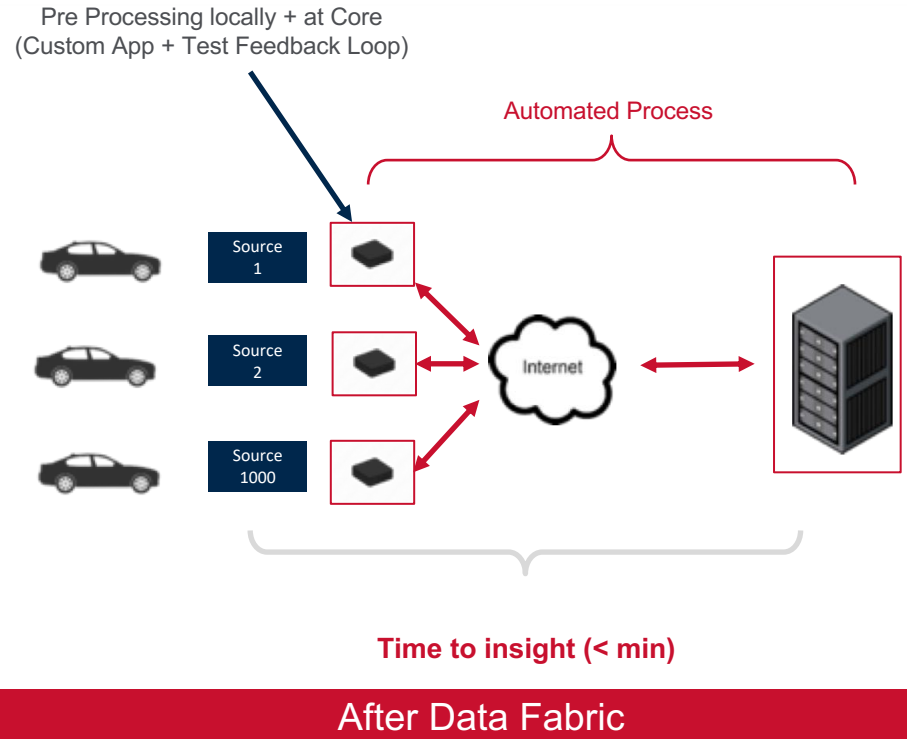
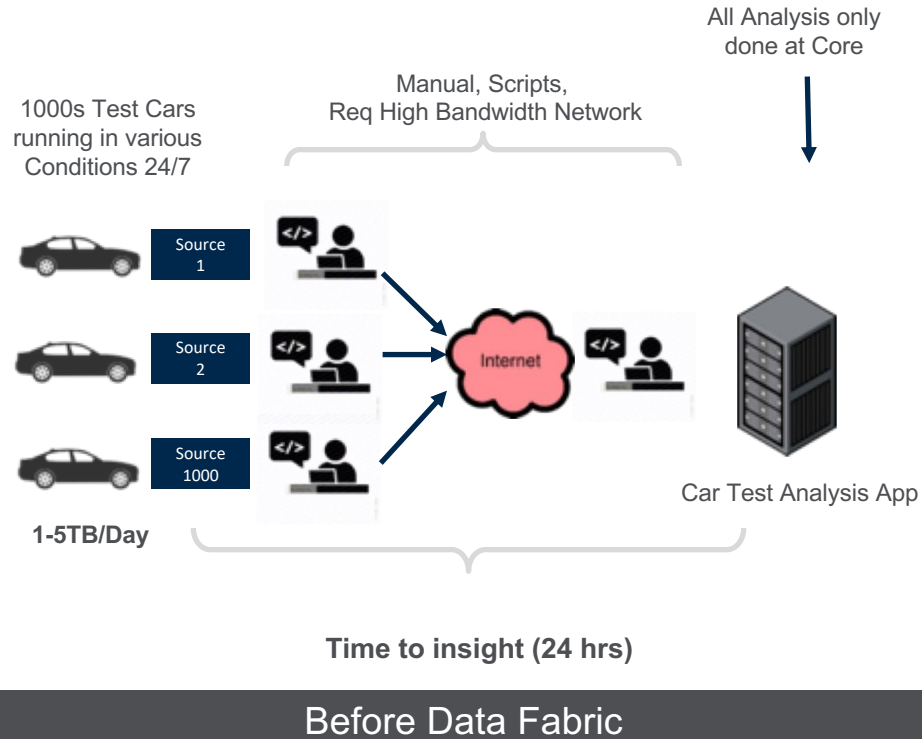


Database



Streams

Use Case: Test & Dev New Cars



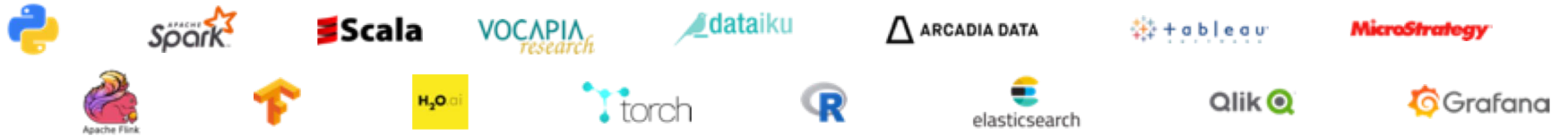
Use Case: Fraud Detection & Investigation

INSIDER TRADING

INSIDER THREAT

MONEY LAUNDERING

MARKET MANIPULATION



Fraud Detection & Investigation Data Fabric



CALL DETAIL RECORDS	SOCIAL MEDIA	EMAILS	LOG FILES
VIDEO	TEXT MESSAGES	AUDIO	DOCUMENTS

CALL DETAIL RECORDS	SOCIAL MEDIA	EMAILS	LOG FILES
VIDEO	TEXT MESSAGES	AUDIO	DOCUMENTS



CALL DETAIL RECORDS	SOCIAL MEDIA	EMAILS	LOG FILES
VIDEO	TEXT MESSAGES	AUDIO	DOCUMENTS



Integrate Analytics Into Operations

- Make adjustments faster than your competitors
- Move from batch to real-time analytics & applications
- Need for mission-critical data flows

A background image showing a close-up of a smiling woman's face on the left, wearing a striped shirt. On the right, a hand is holding a credit card over a payment terminal. The scene is brightly lit, suggesting a retail or service environment.

Major Financial Services Company

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

Chao Yuan, SVP & Head of Decision Science, American Express

Business value: Saves consumers money, drives more revenue for merchant network

Migrate Big Data Infrastructure @ Scale



250B Recommend'n/Mth

- ✓ Game sometimes changes, when systems are put in production
- ✓ Performance & stability needed to meet business SLAs
- ✓ Agility for researchers or algorithm developers to introduce something new

Transforming Society @ Scale



- 10 fingerprints, 2 iris scans, digital picture
- Over 1B entrants to date, 40+ TB
- 1M registrations (trillions of matches)/day
- 100M+ authentications/day
- 20ms response time from the MapR cluster



**20 BILLION
BIOMETRICS**

Keeping The Finger On Your Business Pulse @ Scale

A hand is shown in the lower-left corner, pointing towards the right. A glowing blue ECG line is visible in the upper-left and center, with a horizontal line extending from its right end across the top of the slide. The background is dark blue with some bokeh light effects.

1. Importance of Data Flows
2. Leverage an Emerging Data Fabric
3. Injecting Intelligence Into Operations

Job/Business opportunities

Business Leader

Data Engineer

IT Director

Data Scientist



MAPR.

Limitless Possibilities.