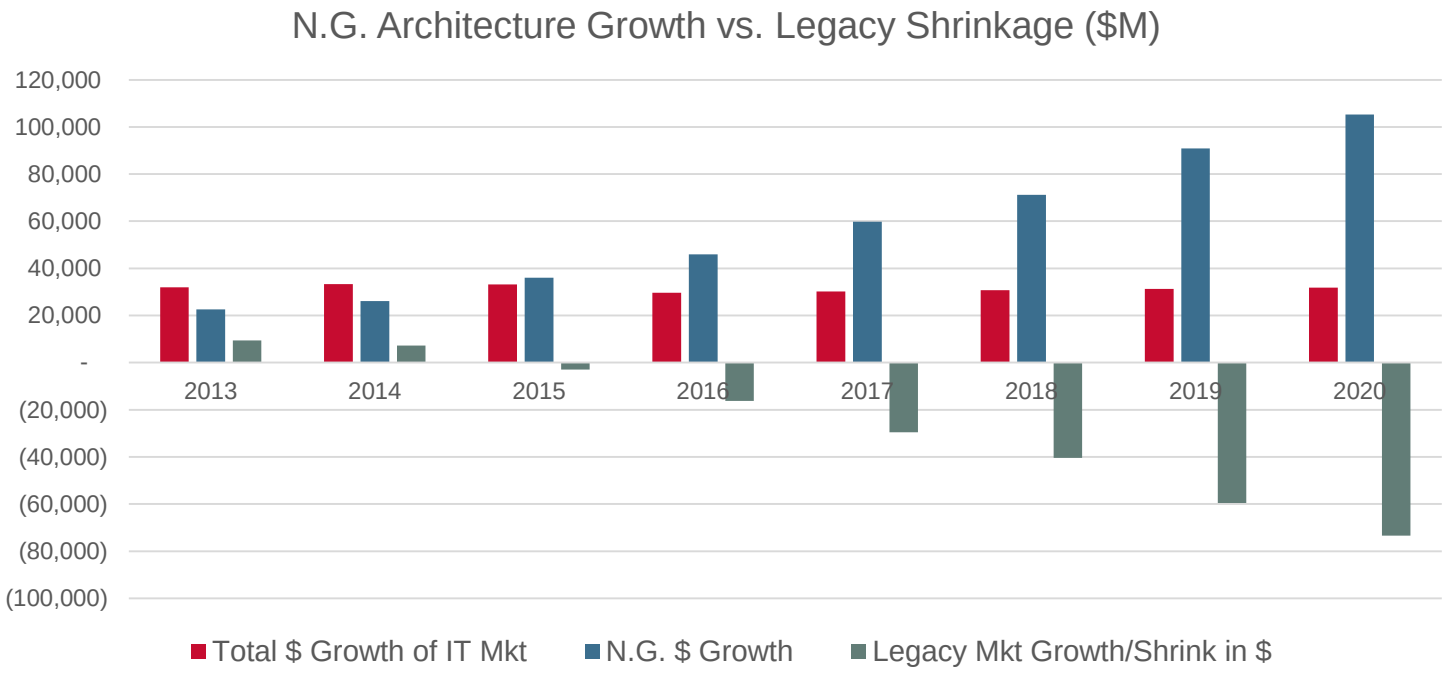
The “Next-Gen” is Now

How big data technologies are revolutionizing the application landscape

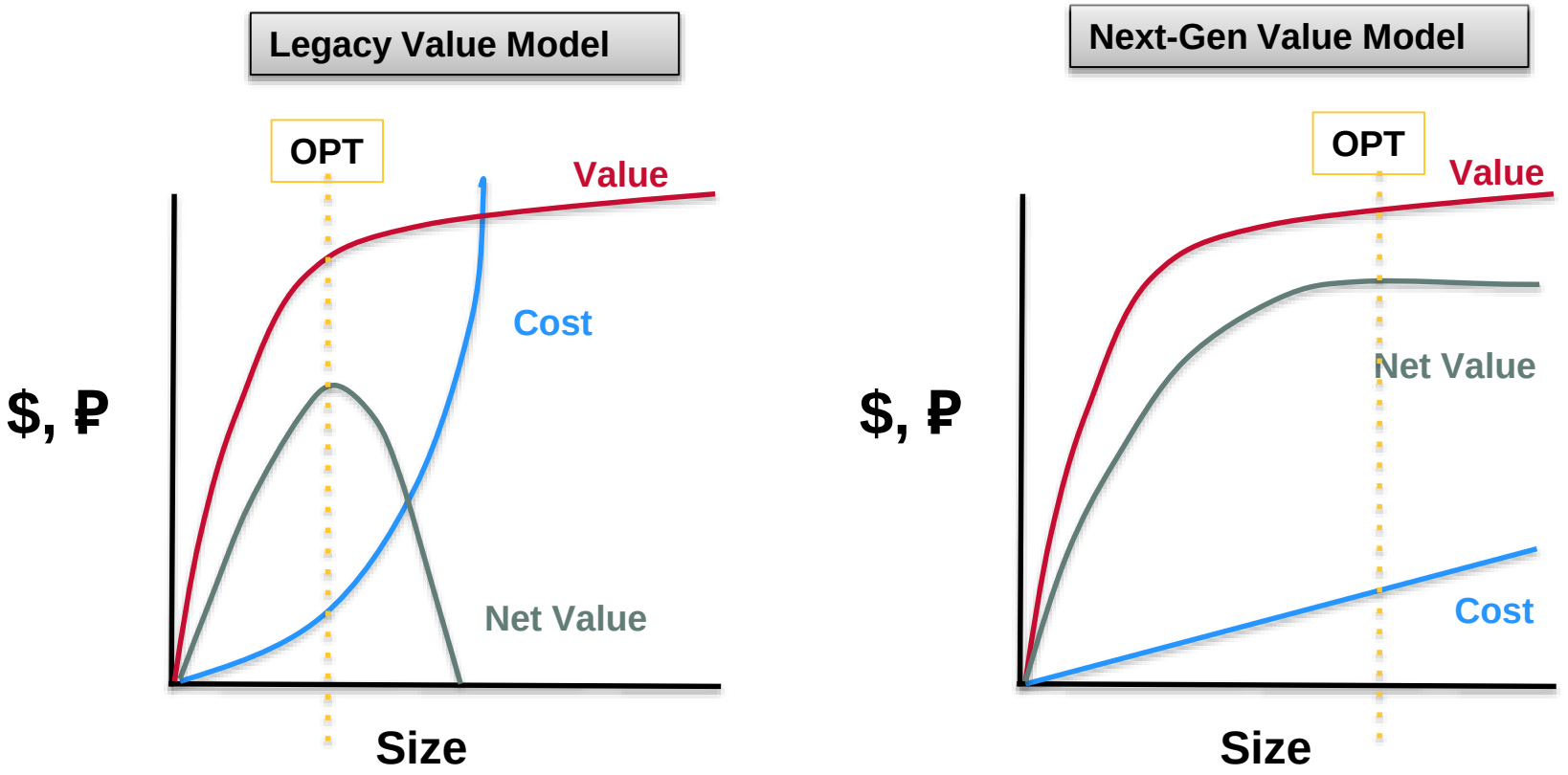
Crystal Valentine, PhD
VP of Technology Strategy

A Technology Inflection Point



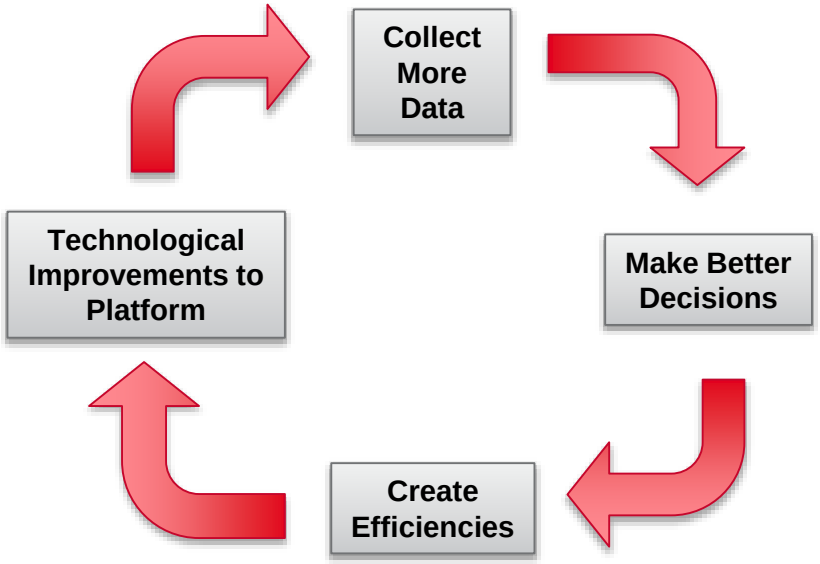
Data source: IDC, Gartner; Analysis & Estimates: MapR

The Value of Data



Data as Infrastructure

Big Data Virtuous Cycle



We are only at the beginning of a new data-first era.

The winners today are succeeding at operationalizing their data best.

The key is the platform.

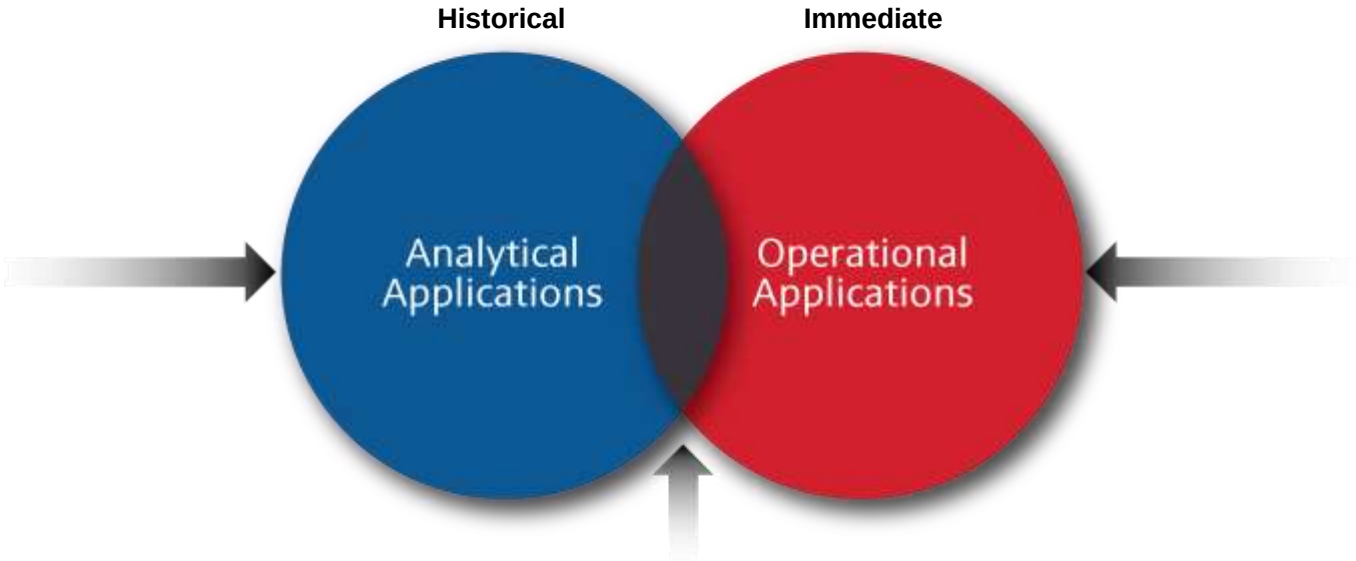
The Next-Gen is Now

In 4 years 90% of Data will be on next-gen technology

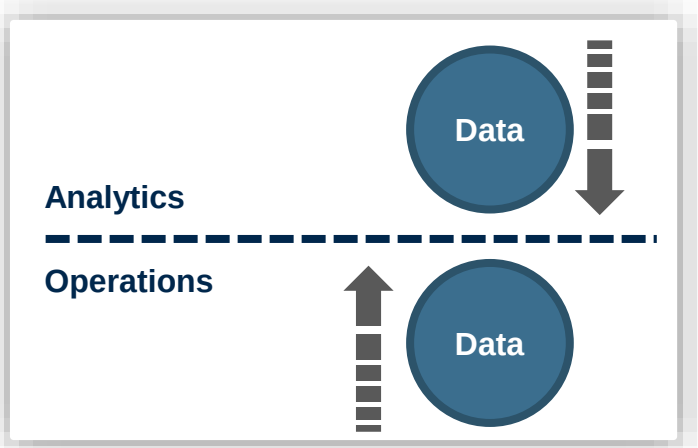
Technological Innovation

1. Data & Compute Convergence

The Secret is to Bring Analytics & Operations Together

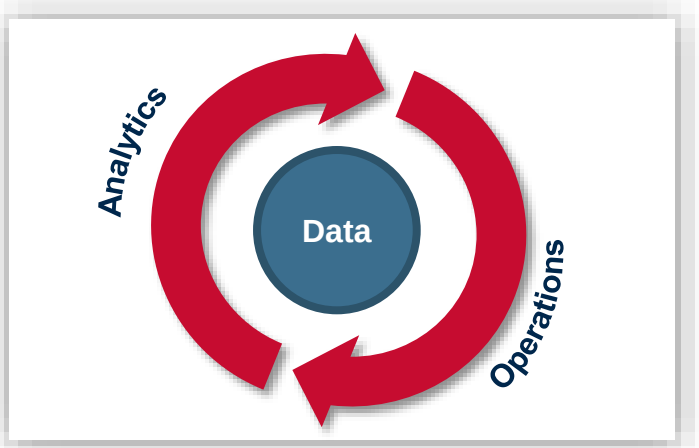


Before



- Complex and Slow
- Multiple Versions of the Truth
- Difficult Governance
- High TCO

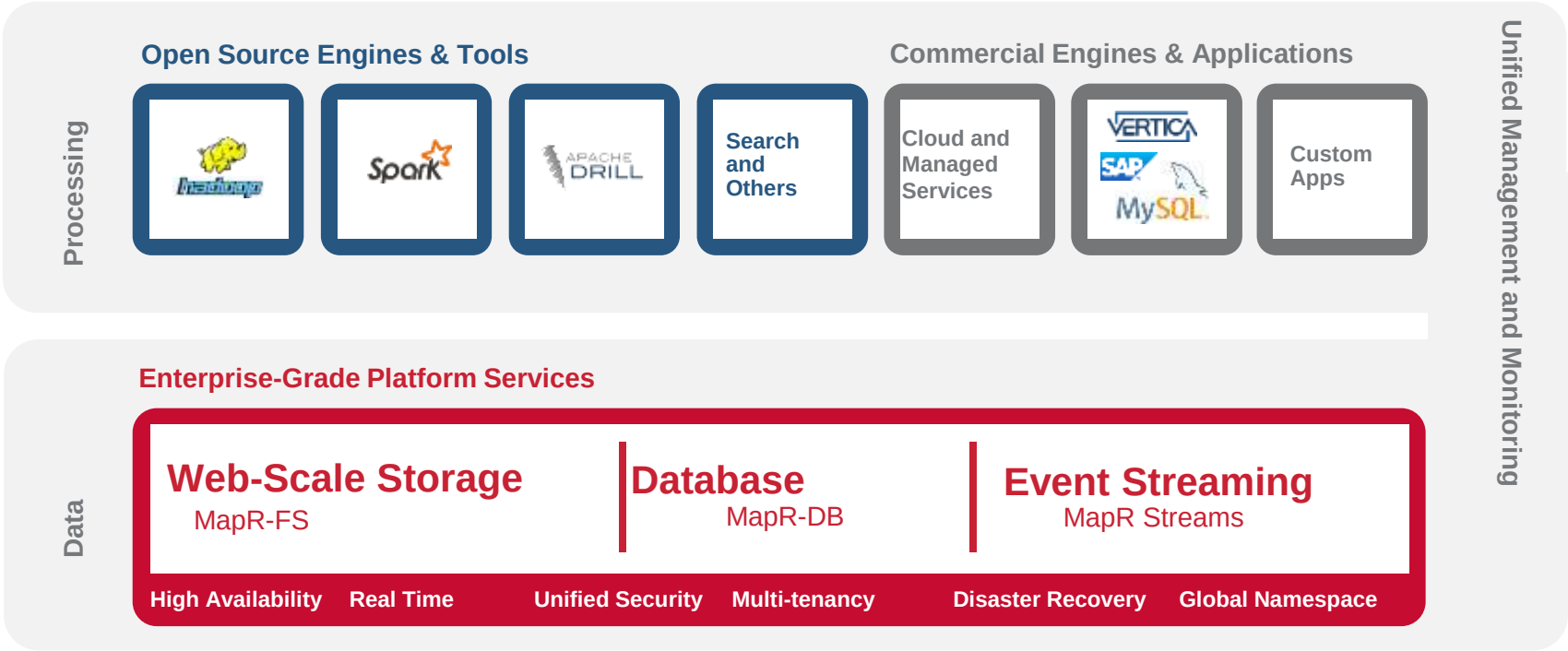
Converged



- Real-Time Data to Action
- Single Data Copy
- Easy Governance
- Low TCO – Scales Horizontally

HTAP (Hybrid Transaction/Analytical Processing) – Gartner Group 2015

The MapR Converged Platform



Example: Predictive Analytics at American Express

- *AmEx MyOffers* App
- Runs on MapR Converged Data Platform

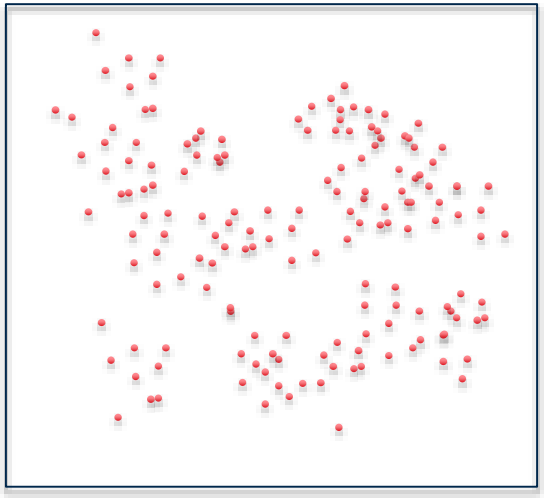
- Analytics: Mahout machine learning

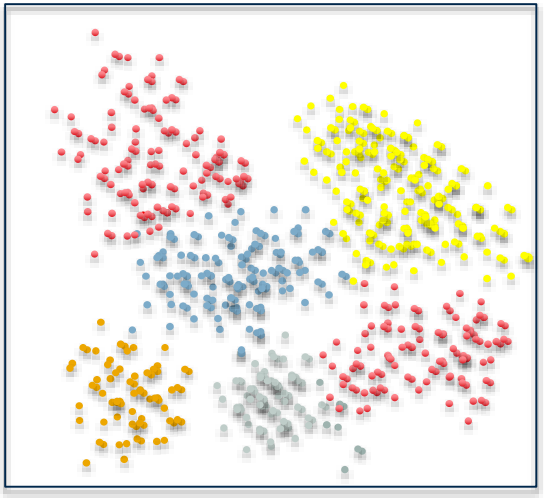
- Operations: Real-time recommendations based on location

Predictive Analytics

Co-occurrence network

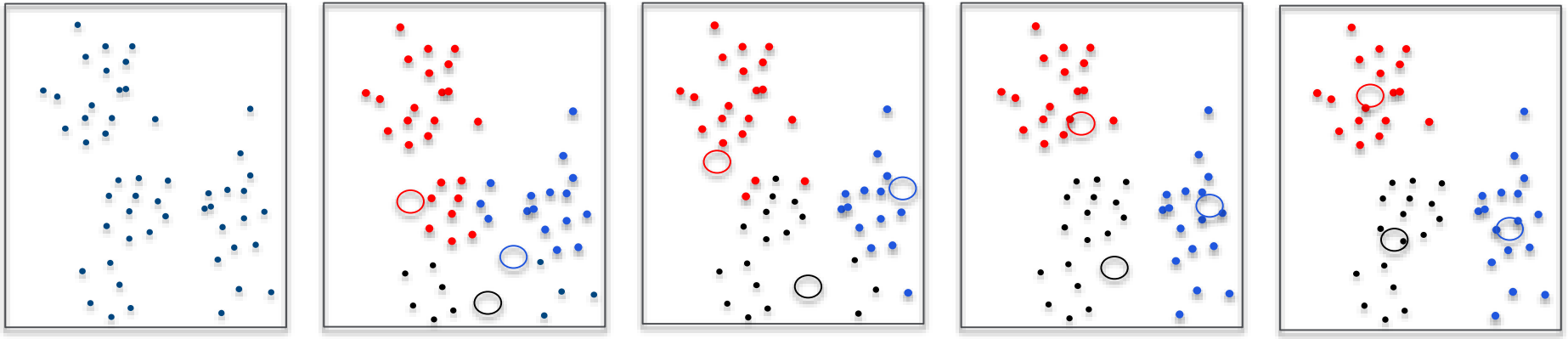
- Cardmember spend graph
- Merchant data
- Location
- Thumb up/down feedback





Type	Machine Learning	Time Series	Discrete Choice	Regression	Classification	Recommendations
Use Cases	Customer Retention	Customer Sentiment	Targeted Marketing	Risk Assessment	Fraud Detection	Diagnostics

Predictive Analytics



A. Input Data | B. Seed Point Selection | C. Iteration 2 | D. Iteration 3 | E. Final Clustering

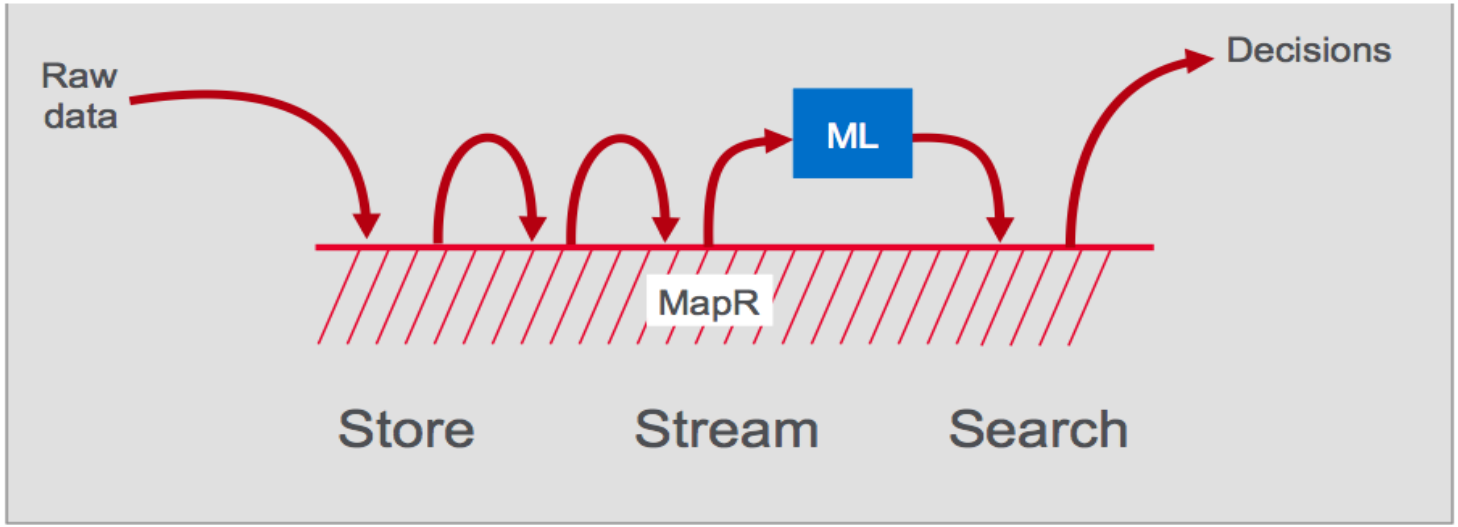
Iteration

Platform requirements

Large-scale, offline iterative ML training algorithm

+

Search engine plus location data stream for fast recommendations



American Express

Customer Retention

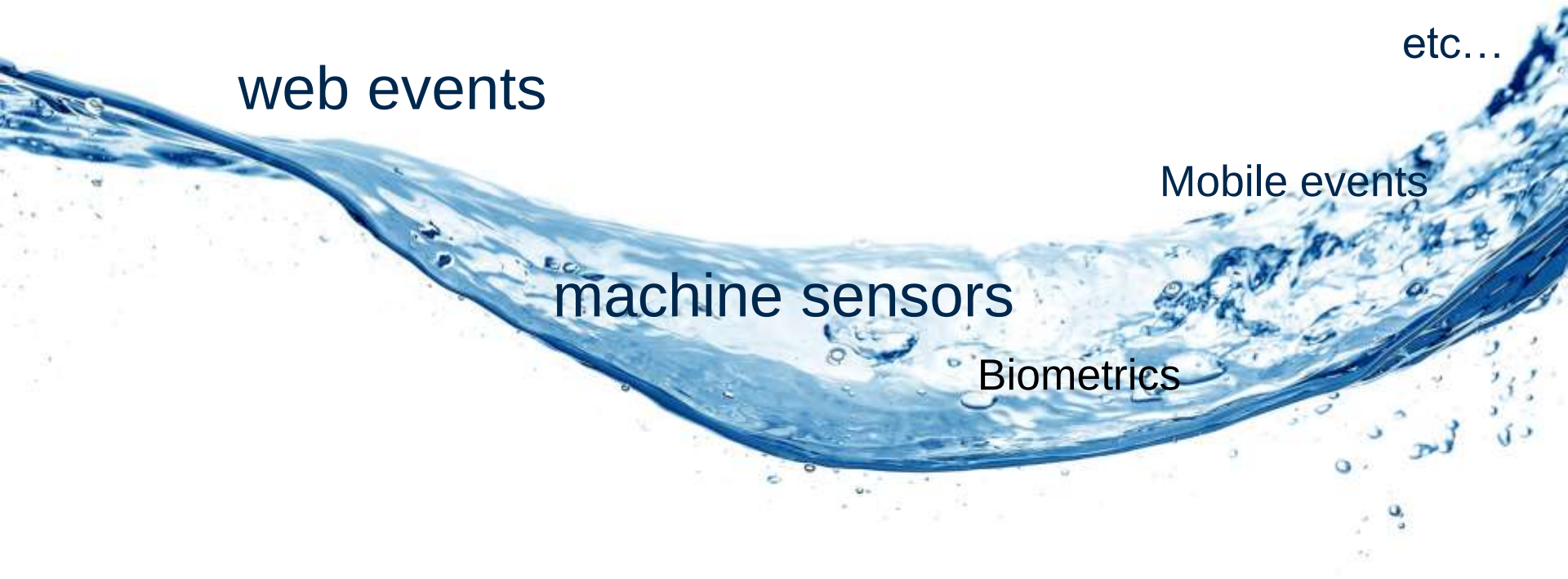
Offers based on real-time location data coupled with deep analytical insights

Saved customers over $190M

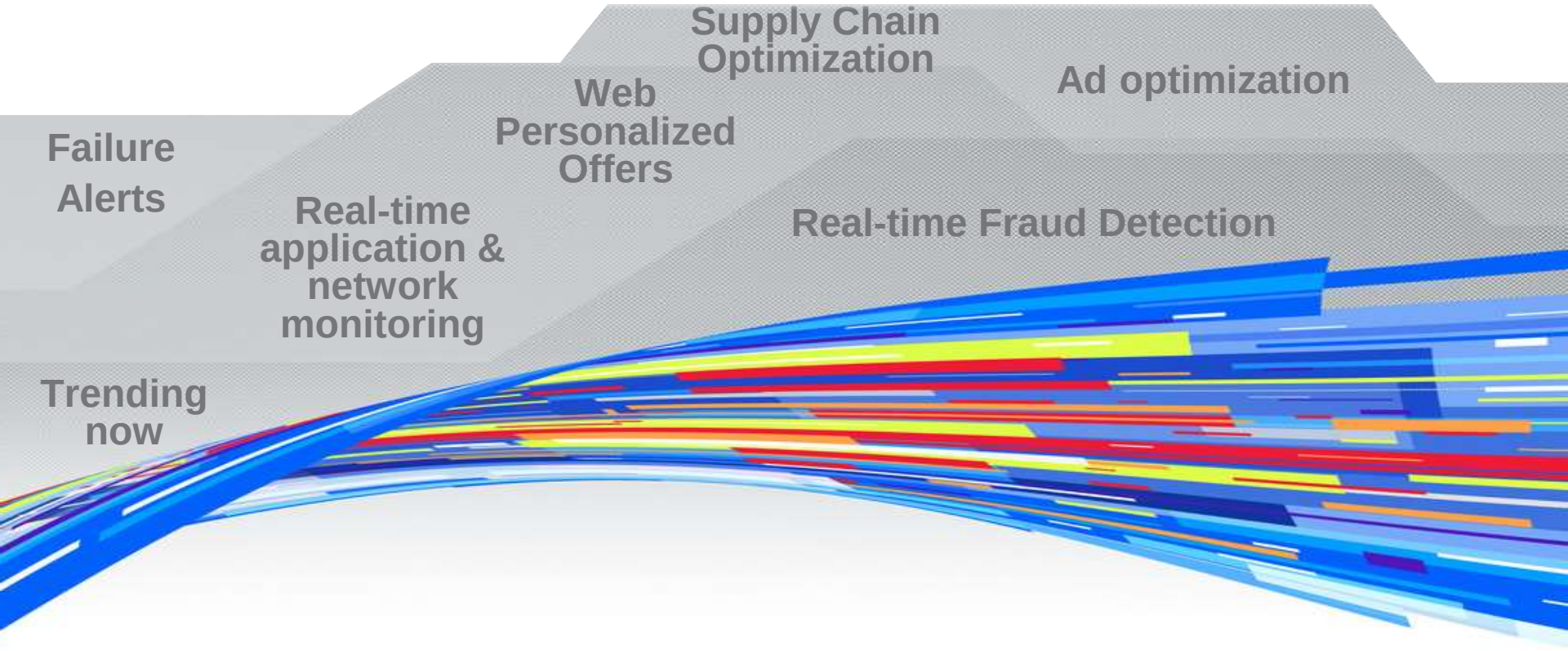
Technological Innovation

2. Stream Processing

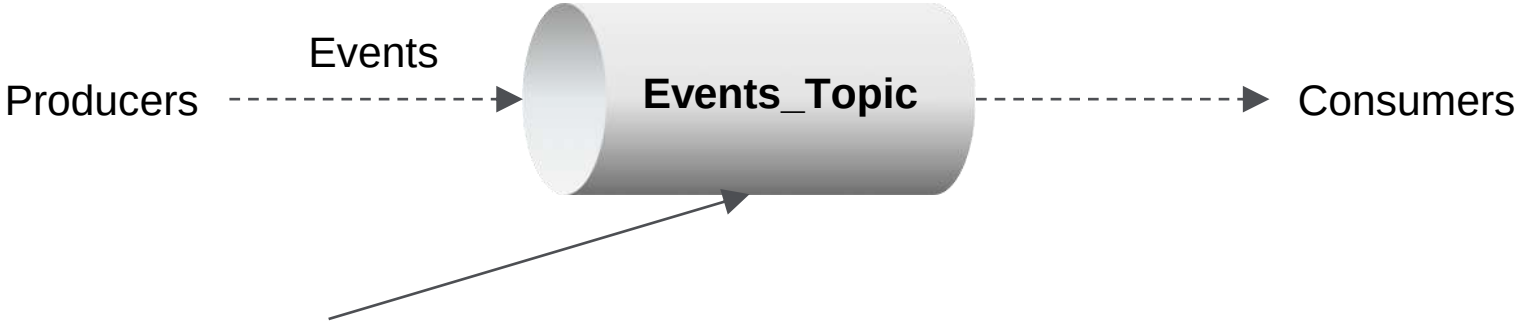
Event-based Data Flows



Streams Enable Real-Time Applications



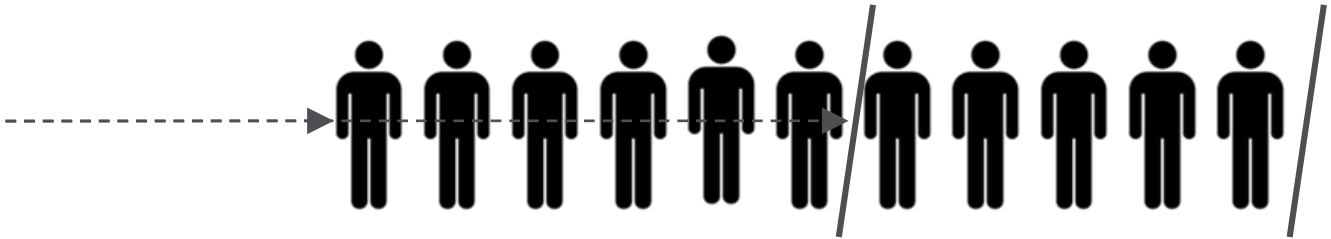
What's a Stream?



A **stream** is an unbounded sequence of events carried from a set of producers to a set of consumers.

Producers and consumers don't have to be aware of each other, instead they participate in shared **topics**. This is called publish/subscribe.

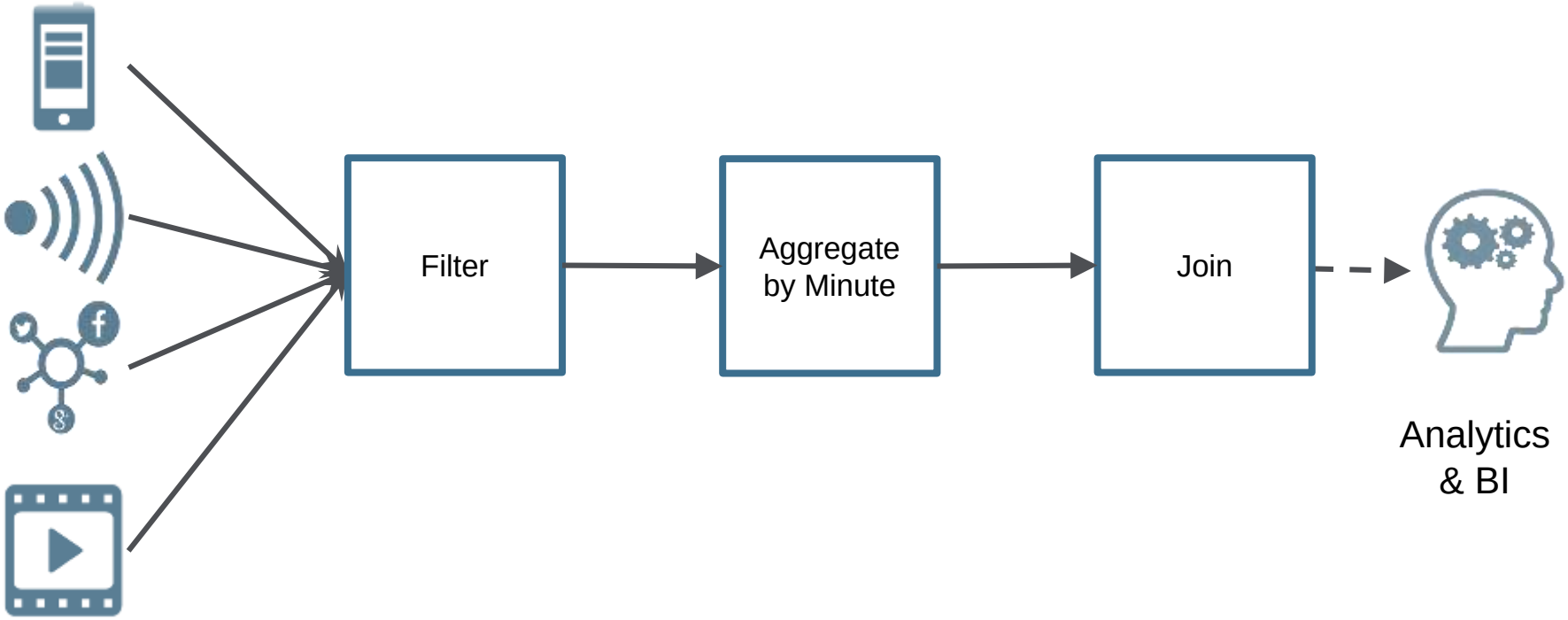
Streams vs Queues



Events are delivered in the order they are received, like a queue.

Unlike with a queue, events are persisted even after they're delivered.

Streams Simplify Data Integration Pipelines



Examples of Producers & Consumers

Social Platforms

Sensors

Servers
(Logs, Metrics)

Mobile Apps

Other Apps &
Microservices

Alerting Systems

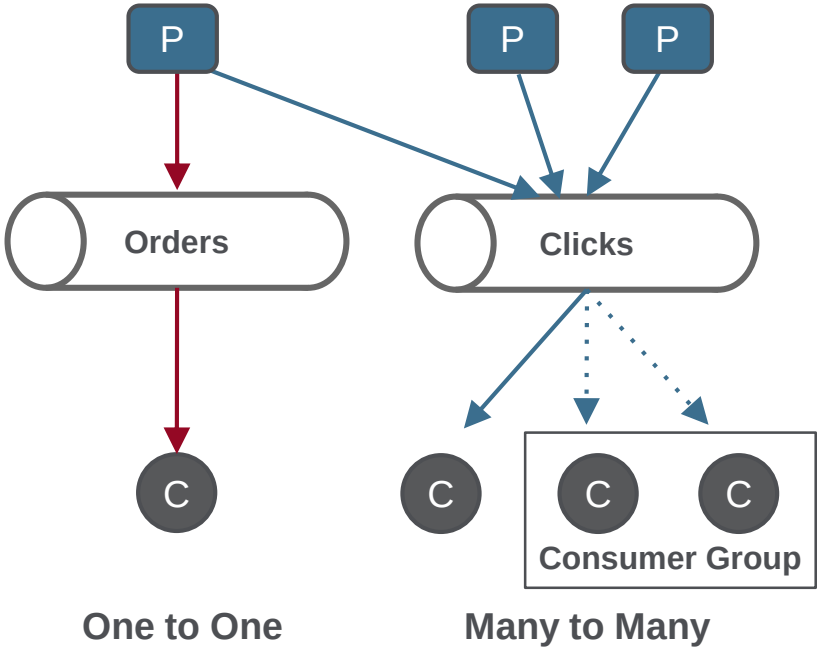
Databases &
Search Engines

Stream Processing
Frameworks

Dashboards

Other Apps &
Microservices

Flexible Communication Modes



Transforming the Health Care Ecosystem

"The Stream is the System of Record"

–Brad Anderson
VP Big Data Informatics

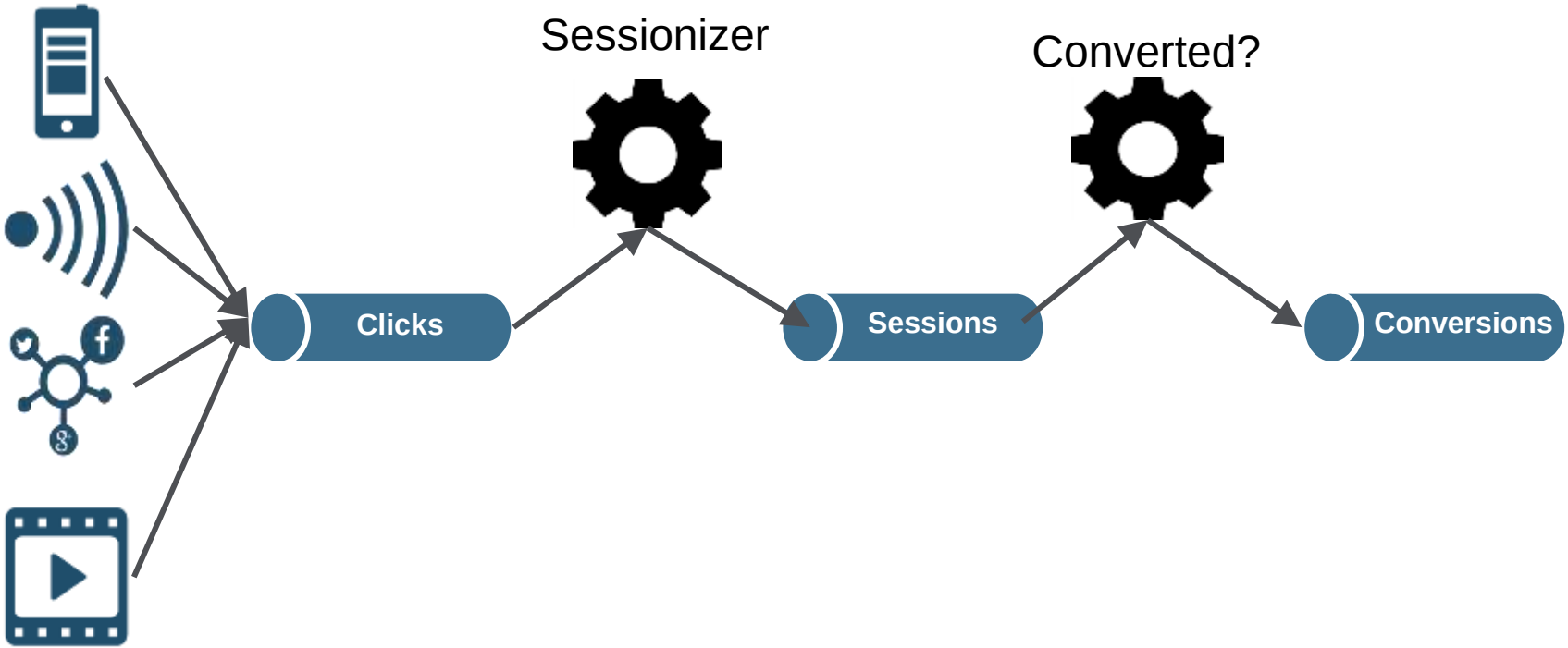
JSON DB (MapR-DB)

Graph DB (Titan on MapR-DB)

Search Engine (Elastic-Search)

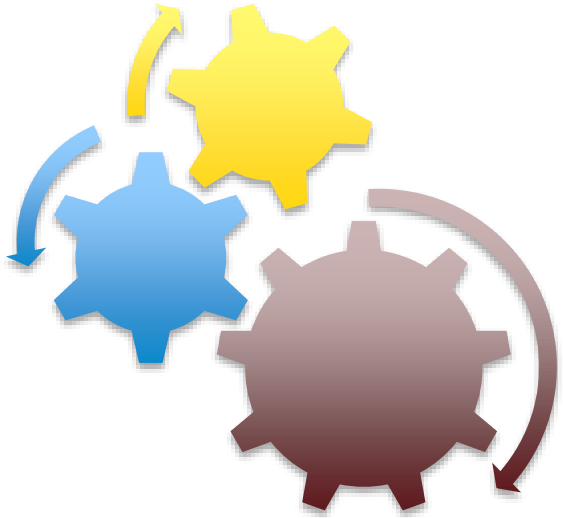
Electronic Medical Records

Streams Enable Event-based Microservices



Microservices Overview

- **Microservices** is an approach to application development in which a large application is built as a suite of modular services.
- Each module supports a specific business goal and uses a simple, well-**defined** interface to communicate with other modules.

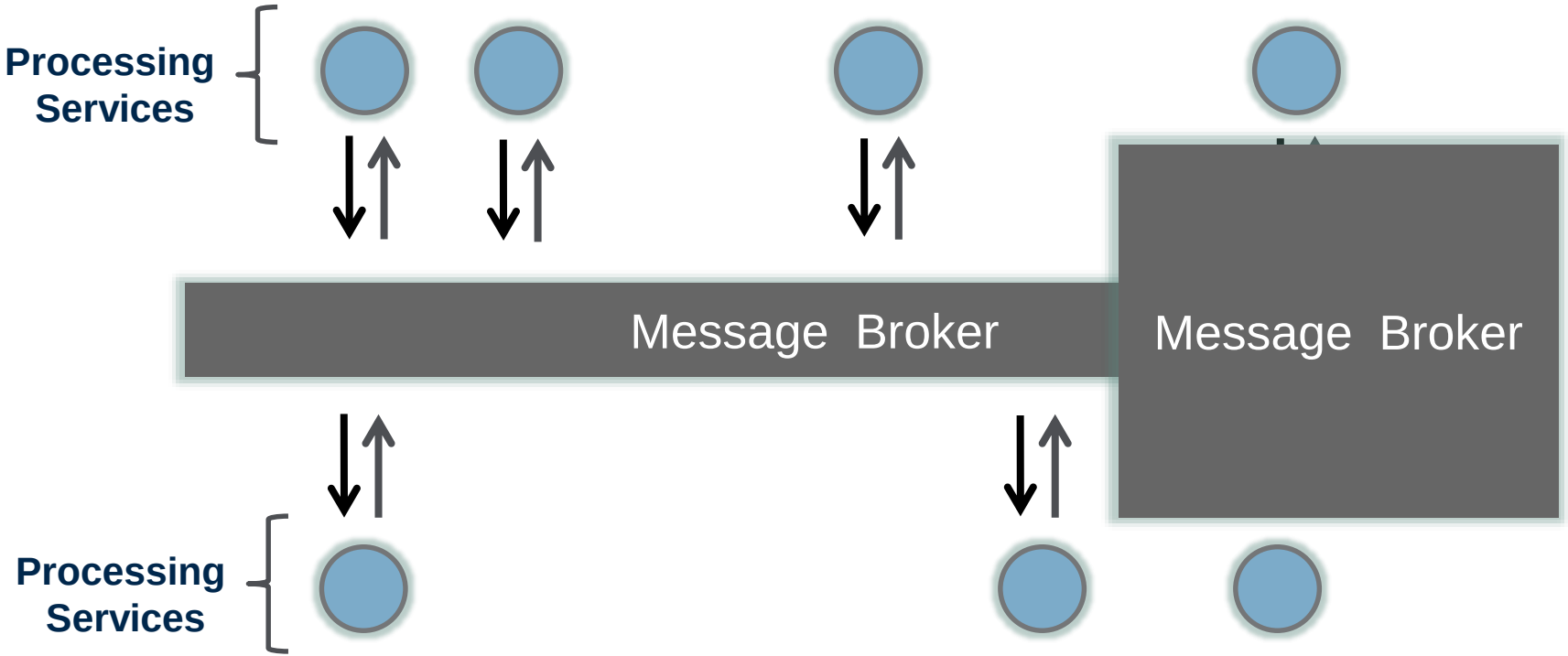


Event-Driven Microservices

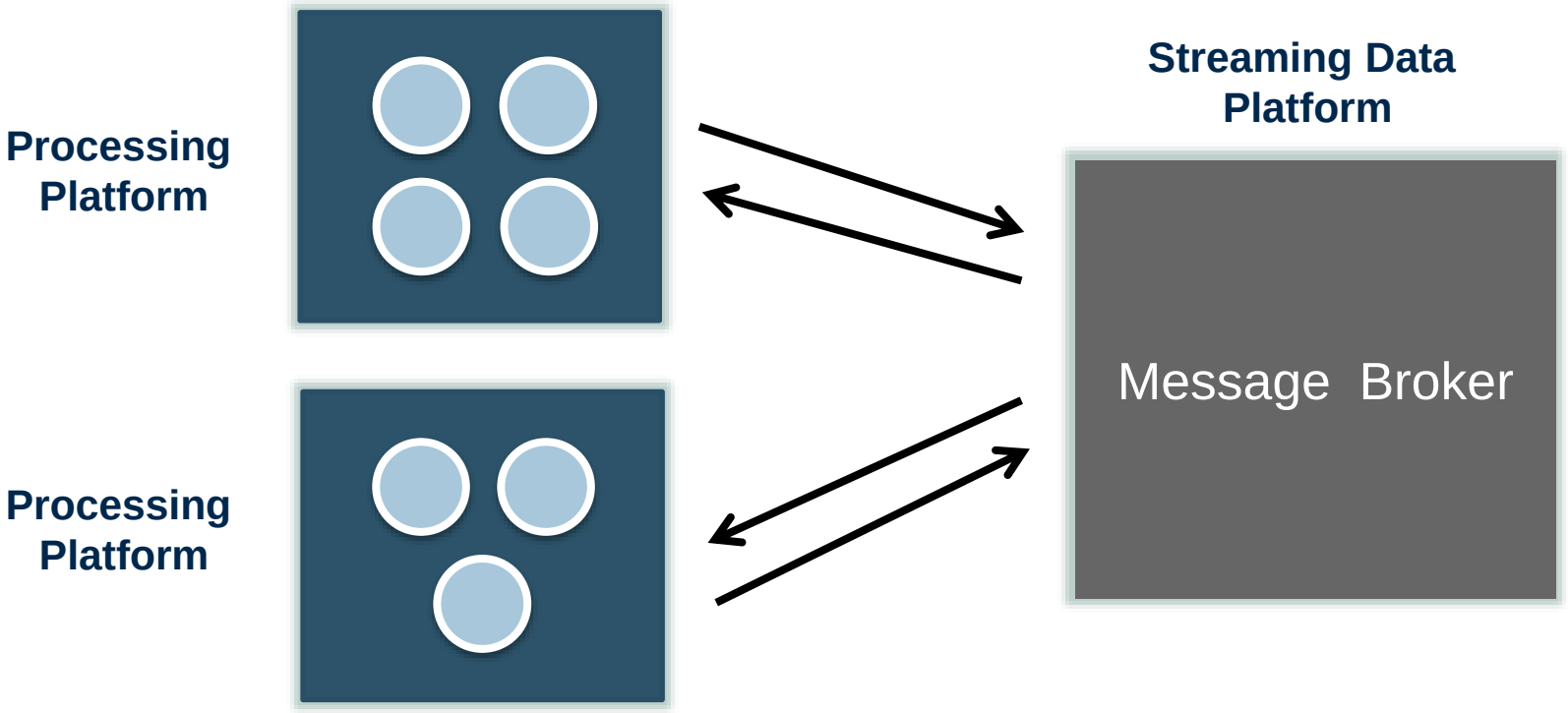
Microservices combined with Streams and a Converged Platform provides:

- Application development across file, database, document and streaming services
- Ultra scale, utility grade performance
- Greater efficiency and simplicity than alternative architectures
- Integrated data-in-motion and data-at-rest and continuous and low latency processing

Microservices Architecture

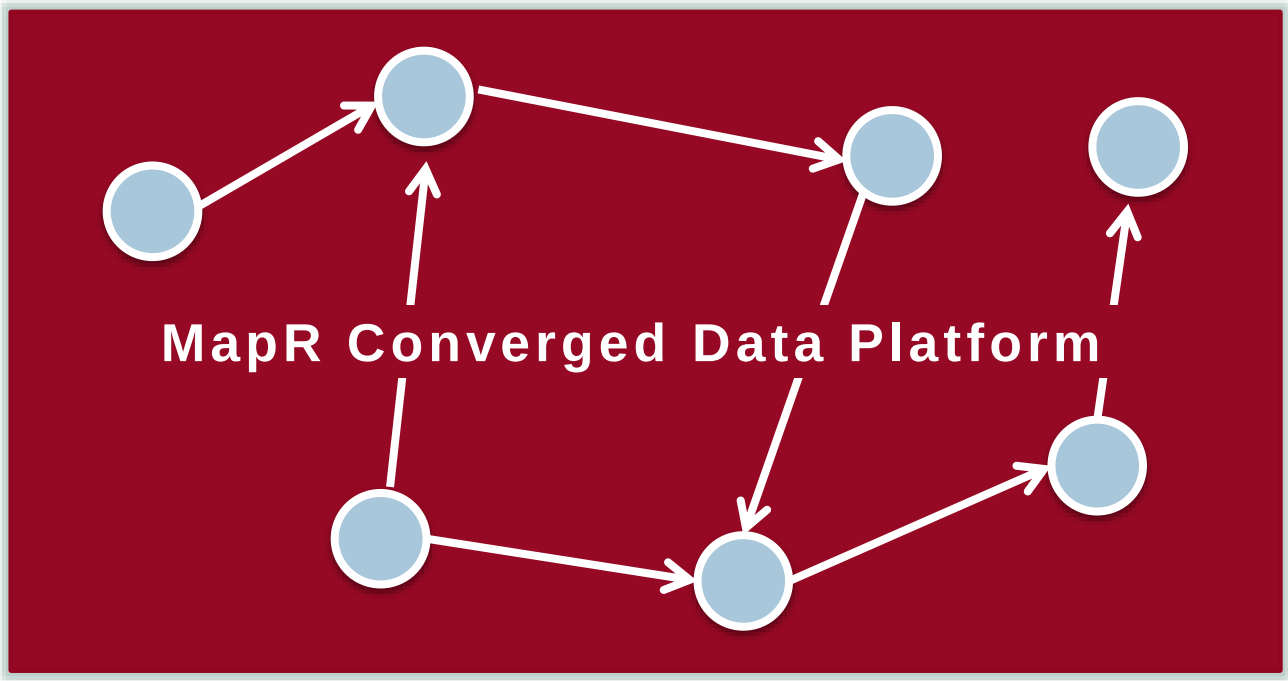


Microservices Architecture



Converged Microservices Architecture

Converged Message and Processing Services



Technological Innovation

3. Global Cloud Processing

Global Cloud Processing Platform

Distributed processing across clusters, data centers and public and private cloud environments

Single global namespace
Strong global consistency

Supports global apps with active/active replication

India

1.3 billion residents

640,000 villages

22 official languages

60% live on $2/day

75 million homeless

Largest illiterate population in the world

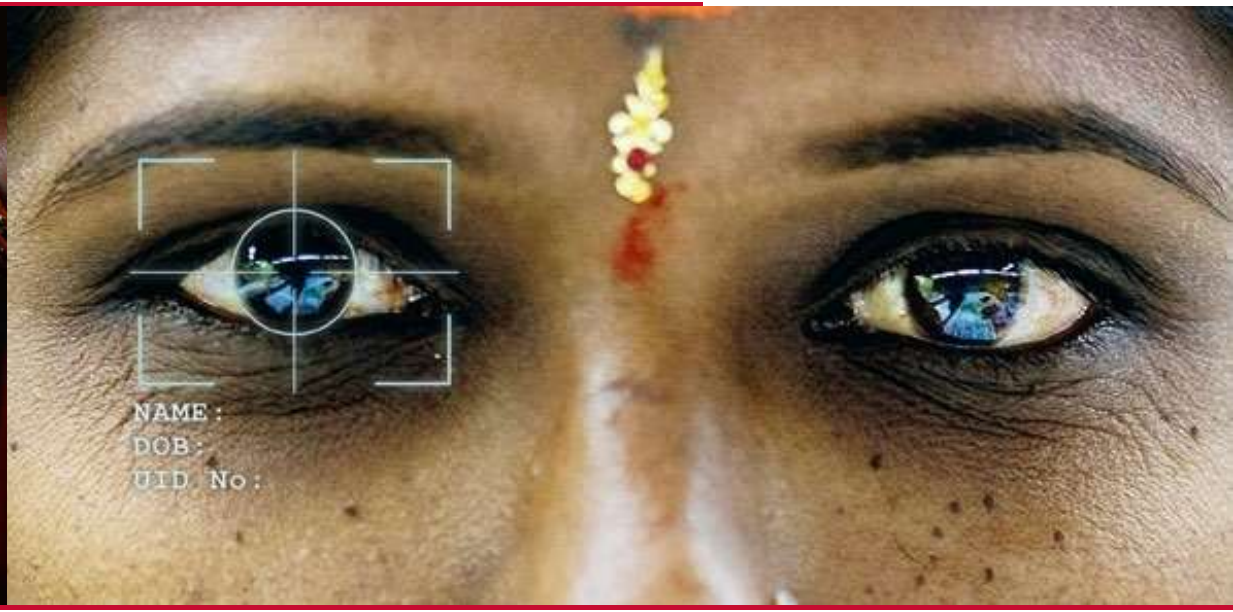



Indian Subsidy Program

$40 billion per year on subsidies

40% leakage due to fraud, impersonation

Largest Biometric Database in the World



20 BILLION

BIOMETRICS

Massive NoSQL Database Application

1M registrations (Trillions of matches) / day

5 TB incremental data / day

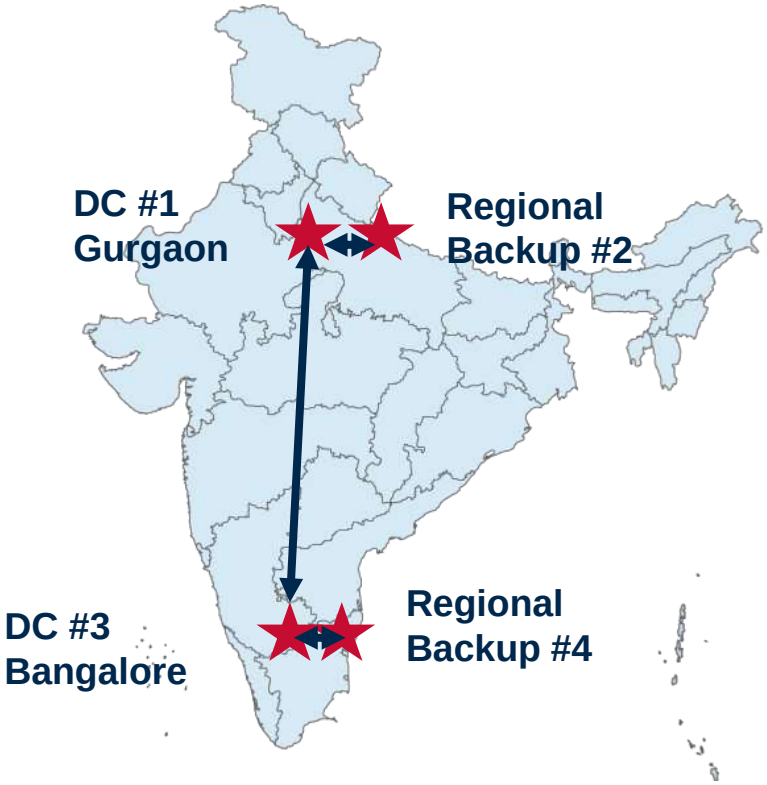
100M+ authentications / day

200 ms response

0.0001 false positive rate

Billion+ audit records per day

Always-on high availability



- Mirroring and backup across DCs
- Strong consistency for update and insert

Broad Application Support through Open APIs

Enrollment Services (40k enrollment stations)

Authentication Services

Fraud detection frameworks

Administrative Services

Analytics and Reporting Services

Logistics services

Contact Center

Impact

$1B saving per year from decreased fraud

>1 billion Aadhaars issued

95% of adults in India

500K-700K new enrollees everyday

240 million bank accounts

122 million LPG consumers

113 million ration cards

Key Technological Innovations

1. Data & Compute Convergence
2. Stream Processing
3. Global Cloud Processing

Thank You

CEE-SEC(R)

Software Engineering Conference in Russia

@DrCrystalV

maprtech

mapr-technologies

MapR

cvalentine@mapr.com

maprtech