Event Driven Architecture for Real-Time Analytics

Mike Spicer
Lead Architect - IBM Streams
IBM Watson and Cloud Platform

November 2017

IBM

Agenda

- The Move To Real-Time
  - Traditional Data At Rest Architectures
  - Why Change?
  - What Needs To Change?

- Event Driven Architecture
  - Components & Workloads
  - Different Demands & Concerns

- Use Cases & Examples

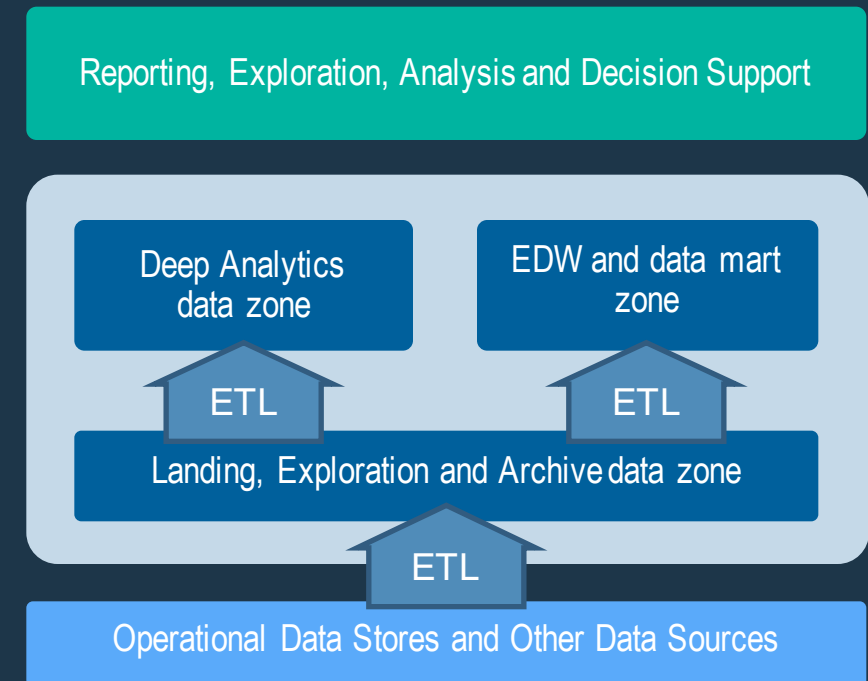- Real-Time Is Now

The Move To Real-Time

Traditional Data At Rest Architecture

Data ingested and stored in Zones

- Periodic updates
- ETL tasks
- Point in time view of the world

Query and Analysis of stored data

- Interactive queries & analytics
  - Request response
  - Set based
- Model building & reporting

Reporting, Exploration, Analysis and Decision Support

Deep Analytics data zone

EDW and data mart zone

ETL

ETL

Landing, Exploration and Archive data zone

ETL

Operational Data Stores and Other Data Sources

Why Change?

The world has changed

An instrumented & connected world can provide better data

- Data from many sources gives a fuller picture of the world
- Connectivity gives up to the moment view of the world

Better data can allow better decisions

Timely decisions can enable new opportunities

"There are two options: adapt or die"

Andy Grove

What Needs To Change?
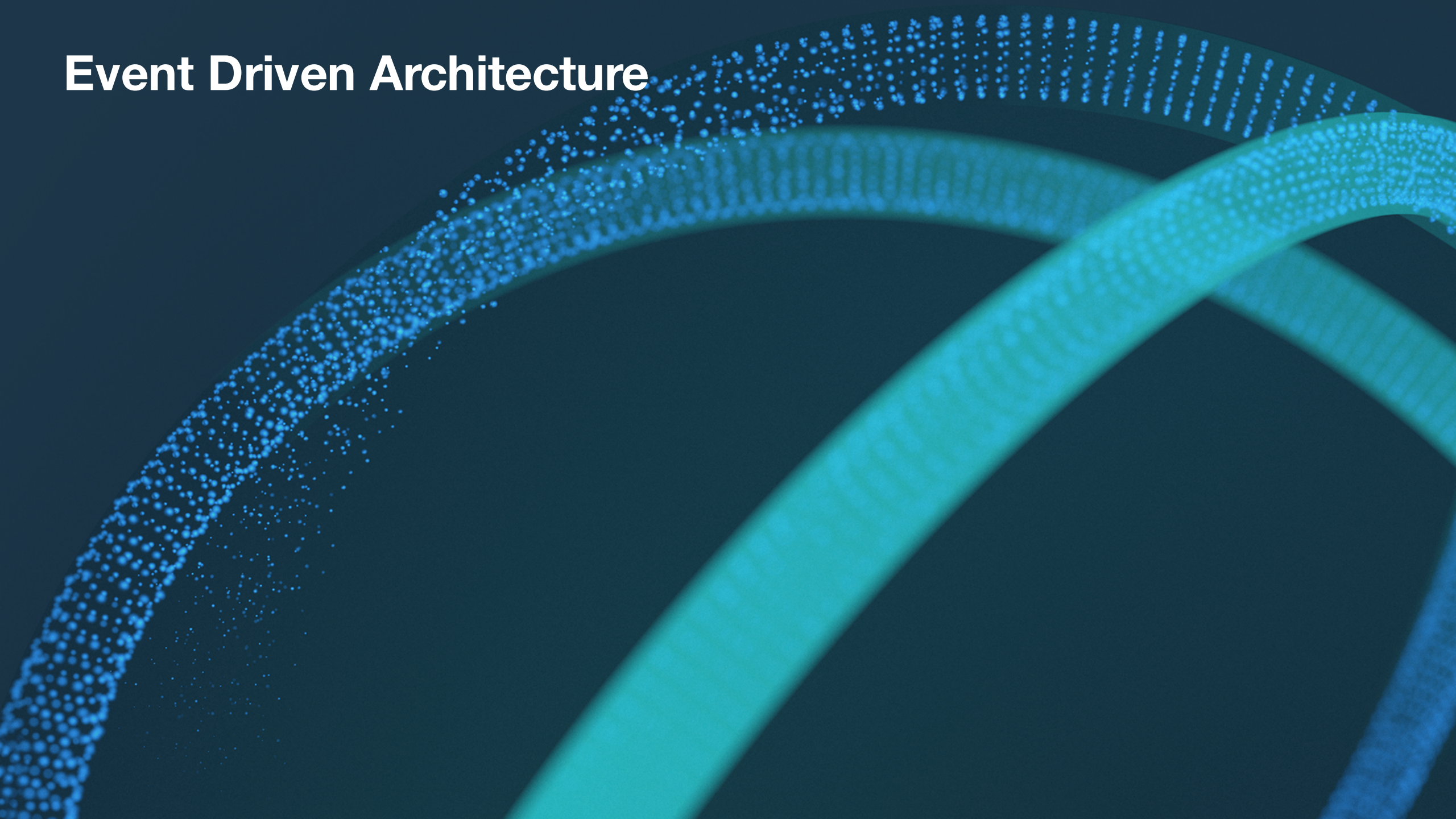
Current data not stale data

Personalized data not generic data

Full context not partial view

Analytics that detect or predict opportunities NOW not in the past

Actions that effect the outcome not just explaining why it happened

Event Driven Architecture

Event Driven Architecture

Data continuously ingested for processing
- Does not have to be stored

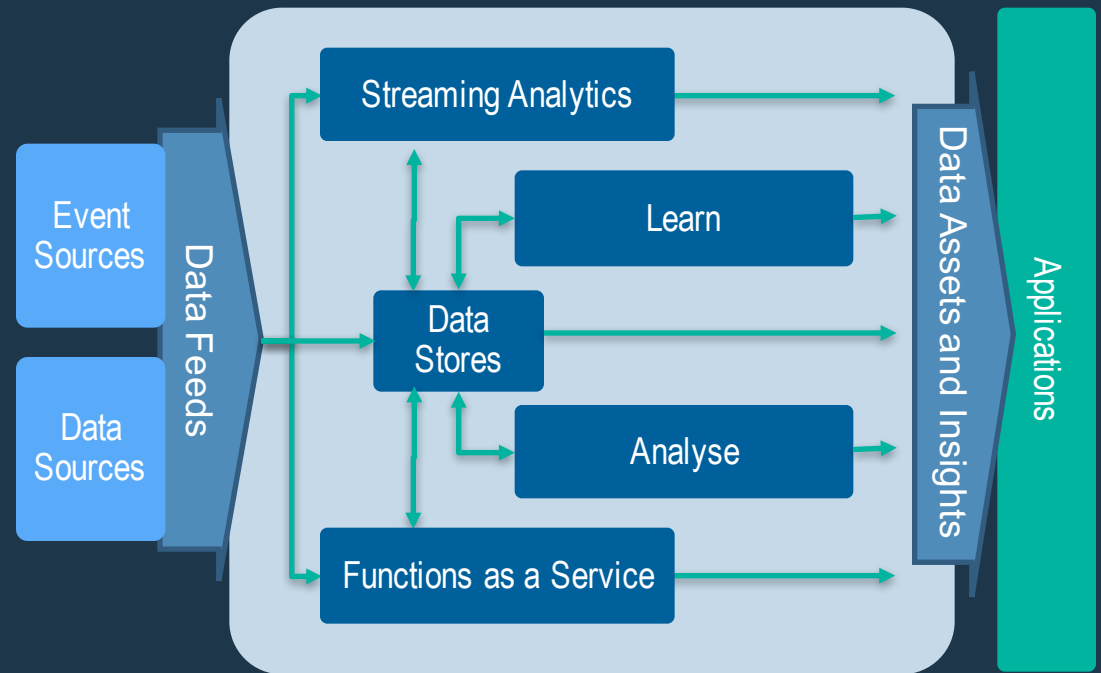**Continuous Query and Analysis of the feed**
- Data pipelines
- Analytics pipelines

Query and analysis of stored data
- Raw and derived data
- Different stores for different workloads

Produces data assets and insights
- Continuous update of the current context
- Feed of events of interest
- Deployed Pipelines and Models

Acquiring and Ingesting Events

Event and Data Sources
- Varied sources and data types
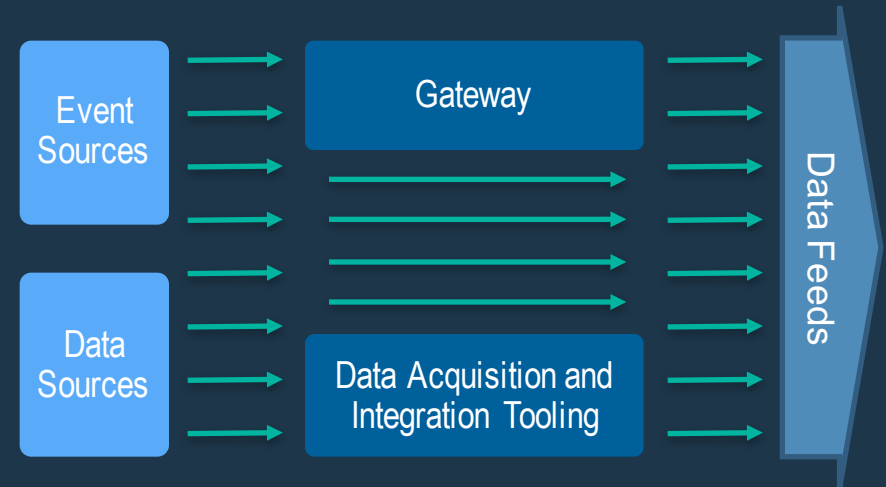- Sensors, apps, clickstream, logs, stores

Gateway
- Scalable Web Endpoint
- Protocol Handlers (http, mqtt, websockets etc.)

Data Acquisition and Integration Tooling
- Apache Edgent, NiFi, MiNiFi etc.
- Change Data Capture
- Application Integration (IBM App Connect)
- Connectors

Data Feeds
- Apache Kafka as ingestion firehose
- Message Queues for internal eventing/control

Processing Events

Streaming Analytics

- Continuous ingest and analytics
- Correlation across events and streams
- Windowing and stateful compute
- Consistent/high-volume streams of data

Functions As A Service (FaaS)

- Adhoc on demand invocation
- Discrete event or task processing
- Stateless compute
- Inconsistent and variable feeds
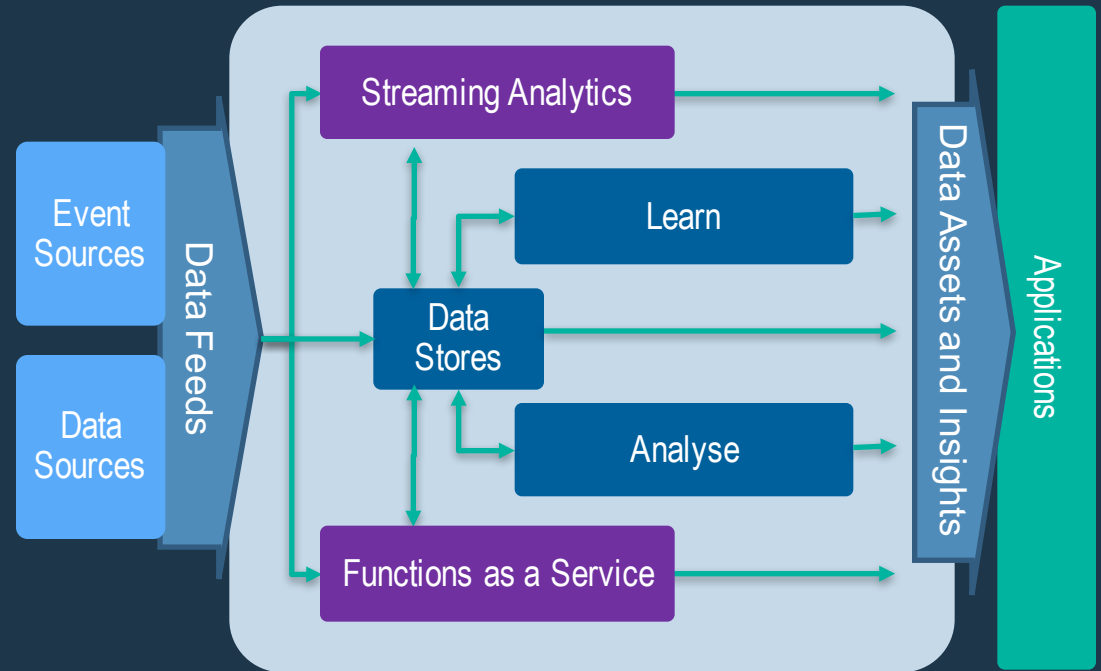
FaaS/Streams Processing Models and Use Cases

Characteristics common to FaaS:

- Stateless compute for **discrete event or task processing**
- Inconsistent or variable loads
- Ideal for out-of-band post processing

Pre & Post Processing

Use Cases:
- Interactive Tier (Mobile & Web)
- Data Events (Cloudant/Object Storage)
- Messaging (Kafka, Streams)
- State changes for individual events

Composition

Use Cases:
- Coordination of loosely coupled events across services

Characteristics common to Streams:

- Consistent/High-Volume streams of data
- Stateful compute optimized for continuous ingest and analytics across **multiple streams of data**

Data Flow Pipelines

Use Cases:
- Ingesting of multiple streams (Call Center, clickstream, account changes)
- Real-time dashboards

Ingest	Prepare	Detect	Decide	Act

Interest Events – Out of band processing

Streaming Analytics Application Pattern

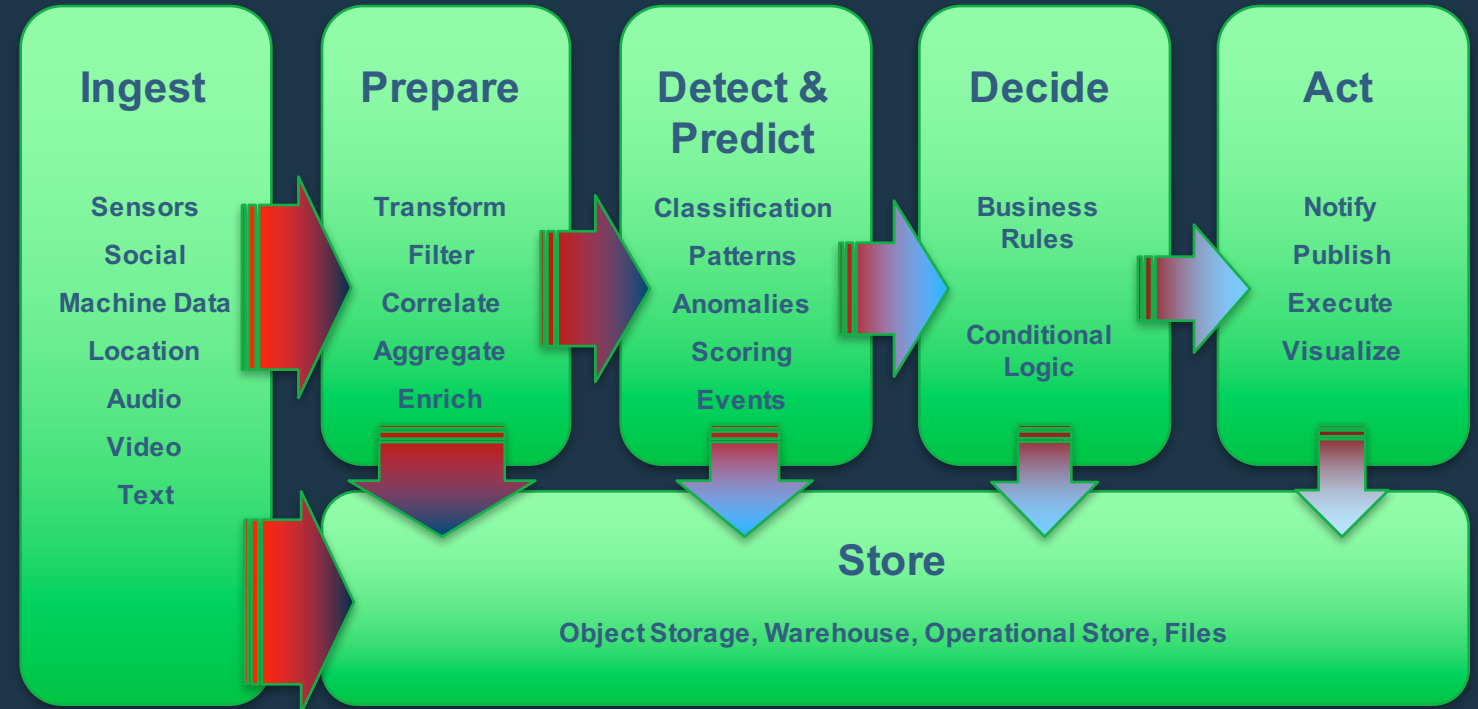**Ingest data from many sources**

Prepare data for analysis

Detect & predict events & patterns in the data

Decide how the results should be handled

Act on the results of interest

Store any data that is of longer term value

Ingest

Sensors
Social
Machine Data
Location
Audio
Video
Text

Prepare

Transform
Filter
Correlate
Aggregate
Enrich

Detect & Predict

Classification
Patterns
Anomalies
Scoring
Events

Decide

Business Rules

Conditional Logic

Act

Notify
Publish
Execute
Visualize

Store

Object Storage, Warehouse, Operational Store, Files

Storing Events

Raw Events
- Historical view for model building etc.

Derived Data
- Transformed, enriched & aggregated data

Profiles and Current Context
- Correlated data across events & sources
- Real-time and reference data

Different Stores For Different Workloads
- SQL & NoSQL
- Object Storage
  - Columnar format for timestamped data
- Warehouse & Operational Databases
- KV & Graph Stores
- Document Stores and Search Engines
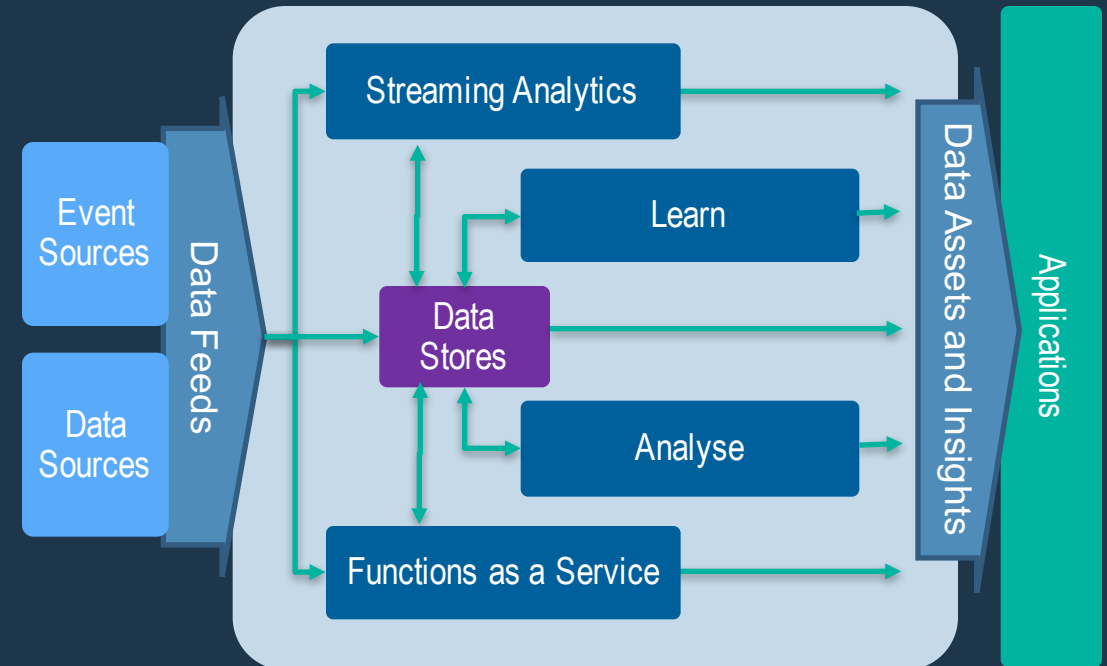
Analyzing and Learning

Detecting and Predicting Insights

- Opportunities & events of interest
- Classifications, patterns, anomalies

Analytical Frameworks & Runtimes

- Internal & external services & embedded capability

Exploratory Analysis

- Queries and analytics on data sets & feeds

Model Lifecycle

- Model building and learning
  - Offline against data sets, online against data feeds
- Published Models
- Deployment to Streaming Analytics, FaaS and API
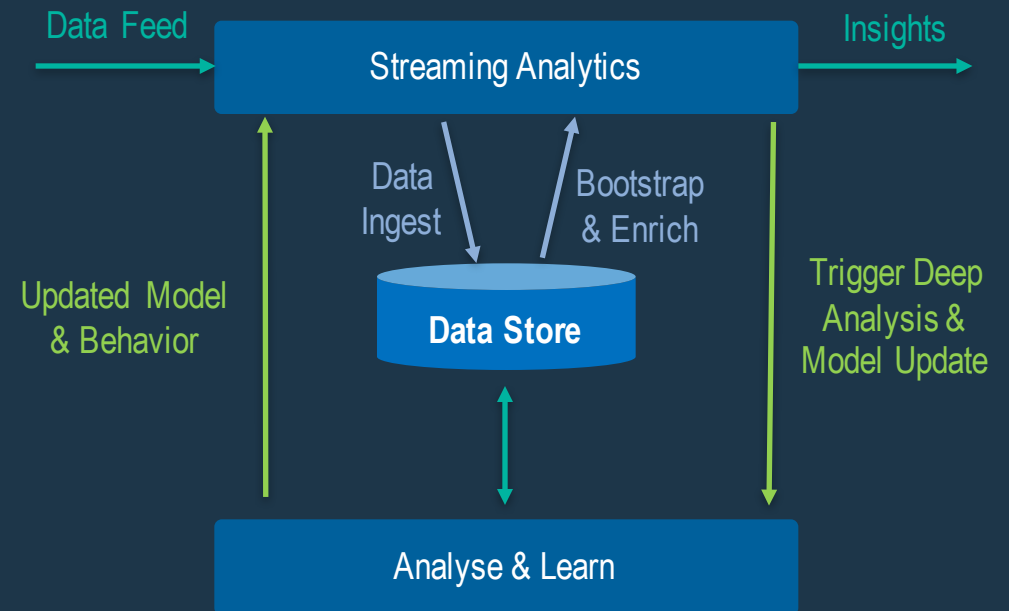- Dynamic refresh with on-line update
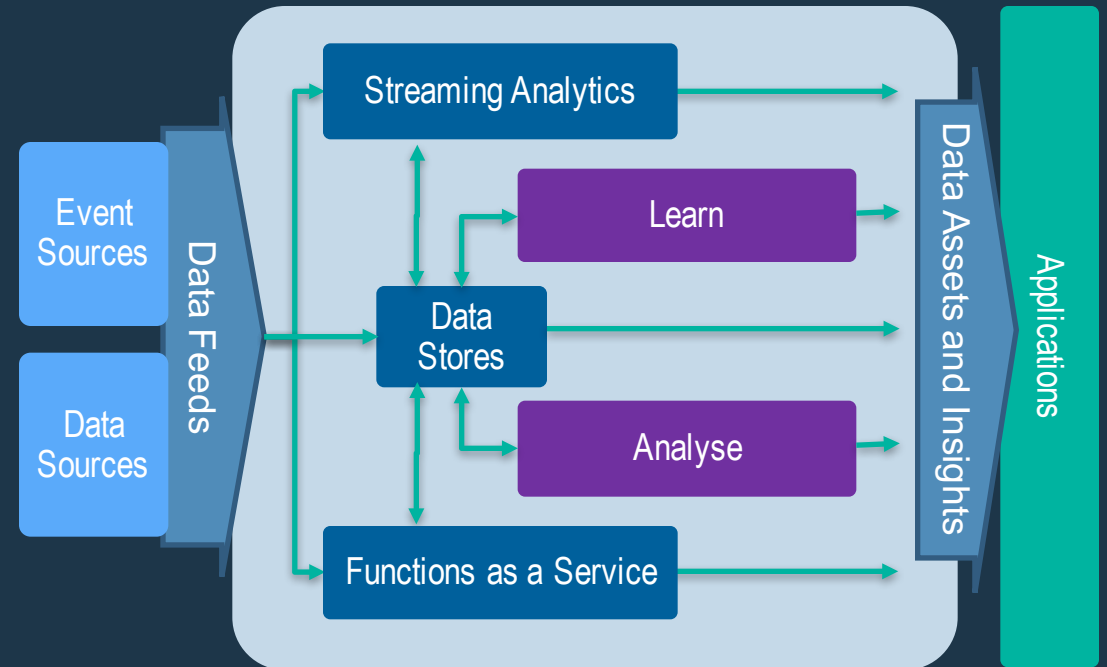
Adaptive Real-Time Analytics

A Comprehensive View of Past, Present and Future

- Current data from a variety of sources
- Real time analysis with low latency results
- Historical context for deep analysis
  - You may not know something is significant until later

Adapt to changes in the real world as they happen

- Trigger model update when unexpected behavior is observed
- Deep analysis when events of interest are detected
- Change behavior to detect and predict events we didn't know were of interest before

Data Feed → Streaming Analytics → Insights

Data Ingest

Bootstrap & Enrich

Updated Model & Behavior

Data Store

Trigger Deep Analysis & Model Update

Analyse & Learn

Analytics & Decision Making

GeoSpatial
- Location based analytics
- Geofencing & map matching
- Spatio-temporal hangout detection

TimeSeries
- Timestamped data analysis
- Anomaly detection & forecasting

Text Analytics
- NLP & NLU
- Sentiment analysis & entity extraction

Video and Audio
- Speech to Text
- Image recognition

Rules
- Decisions described as business logic

CEP
- Temporal pattern detection

Entity Analytics
- Relationships between entities
- Probabilistic matching

Data Assets and Insights

Raw, Derived and Context Data
- Constantly updated
- Exposed through the appropriate stores
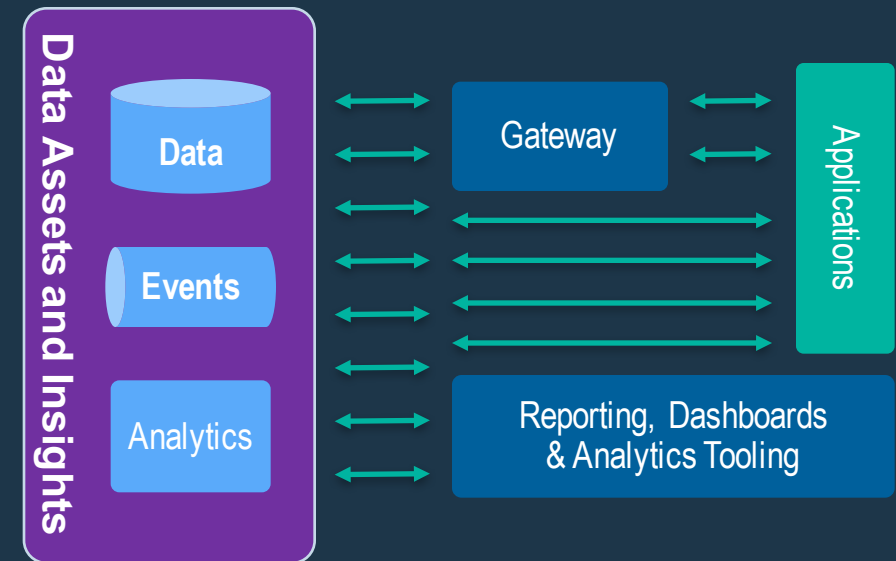
Event Streams
- Processed and derived events
- Opportunities and events of interest

Analytics
- Insights from ML Models
- Invocation of data and analytical pipelines

Gateway
- Scalable Web Endpoint
- Protocol Handlers (http, mqtt, websockets etc.)

Platform or Single Framework?

One Size does not fit all

- Different workloads have different characteristics
- Choose the right tool for the right job

One Size does fit some, sometimes

- Requirements change
- Technologies change

Platforms should help not hinder

- Choice not Confusion
- Simplicity not Complexity
- Integrated not Bundled

Clip Art Credit : http://www.supercoloring.com/silhouettes/orchestra

Platform Fabric

Collaboration and Sharing
- Team based development

Metadata and Asset Catalog
- Discoverable assets
- Automatic configuration
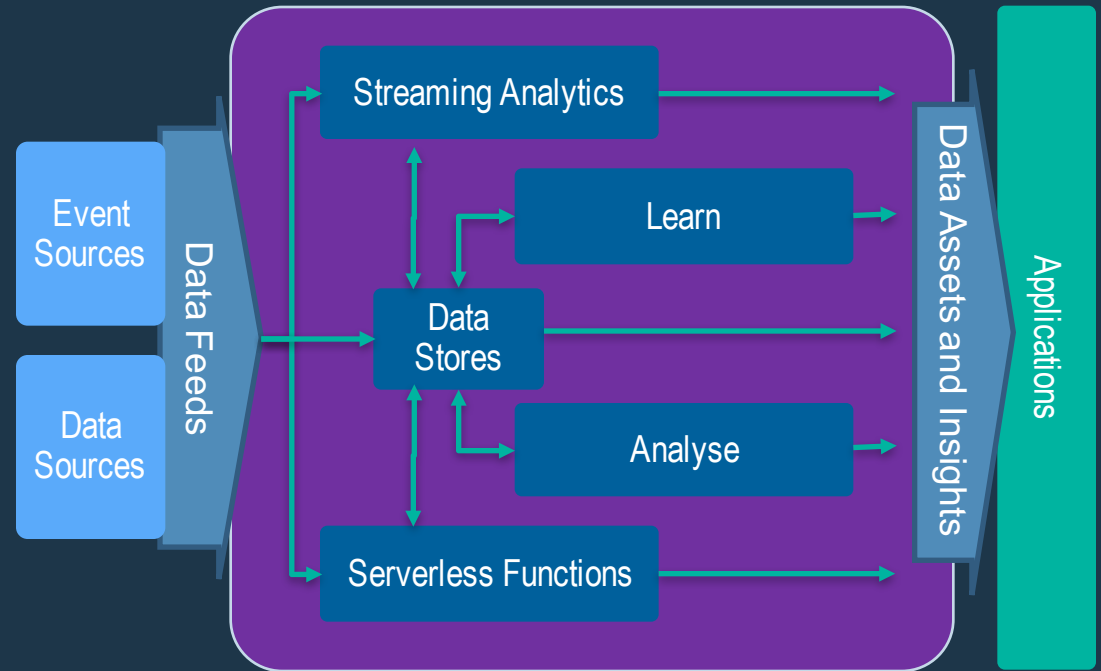
Lineage and Governance
- Trusted data and insights
- Traceable provenance & usage

Integration and Orchestration
- Composable micro services
- Dynamically constructed & coordinated across runtimes

Administration & Infrastructure
- Metrics and Monitoring
- Authentication and Authorization
- Logging and Audit

Event Sources

Data Sources

Data Feeds

Streaming Analytics

Learn

Data Stores

Analyse

Serverless Functions

Data Assets and Insights

Applications

Different Demands & Concerns

The Feed Never Ends

- The collection is unbounded
    - Not a request response set based model
- Windows can provide bounds
    - Sliding, Tumbling, Session
    - Size and trigger are separate concepts

The Firehose Doesn't Stop

- Keep drinking and keep up
    - Resilient and self-healing
    - Dynamic and adaptable
    - Processing rate >= Feed rate
        - Buffering doesn't save you
        - Increased latency and eventual overflow

Different Demands & Concerns

The Only Constant Is Change

- Changing requirements
  - Data sources & formats
  - Analytics, insights & opportunities
- Data feed rates
  - Periodic peaks and troughs
  - Prolonged change
- Adapt or die
  - Dynamic behavior
  - Dynamic configuration
  - Dynamically composed applications
  - Dynamic and elastic scaling

UNCERTAINTY AHEAD

Different Demands & Concerns

Guarantees Are Not Guaranteed

- Guaranteed Delivery != Guaranteed Processing
- Local Guarantees != Global Guarantees
  - External effects must be accounted for
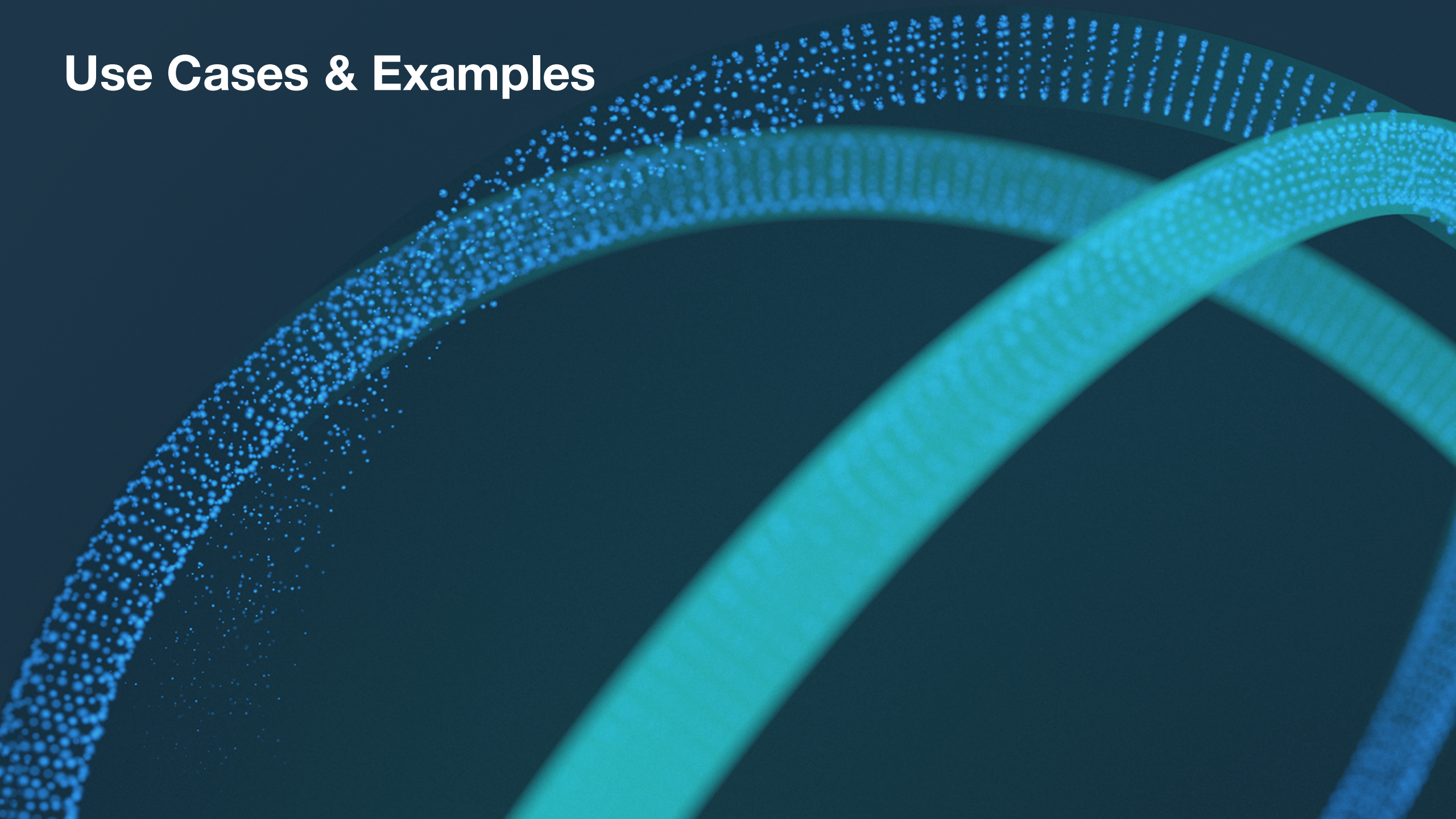 - XA Transactions are not the answer

The World Doesn't Wait

- Insights are perishable
- Opportunities are lost
- Latency matters

Photo Credit : https://www.flickr.com/photos/sepblog/4072462666/

Use Cases & Examples

Event Driven Architecture Is Broadly Applicable

Telecommunications
- Cognitive call center
- CDR processing
- Social analysis
- Churn prediction
- Geomapping

Transportation
- Intelligent traffic management
- Automotive telematics
- Trend Detection

Energy & Utilities
- Call Deflection and Optimization
- Transactive control
- Phasor Monitoring Unit
- Down hole sensor monitoring

Health & Life Sciences
- ICU monitoring
- Epidemic early warning system
- Remote healthcare monitoring
- HIPAA, PCI Compliance
- Transcription Services

Natural Systems
- Wildfire management
- Water management

Law Enforcement, Defense and Cyber Security
- Real-time multimodal surveillance
- Situational awareness
- Cyber security detection
- Data Leakage

Stock market and Banking
- Impact of weather on securities prices
- Analyze market data at ultra-low latencies
- Momentum Calculator
- Regulatory Monitoring and Audits

Insurance and Fraud prevention
- Detecting multi-party fraud
- Real time fraud prevention
- Know your customer (360)
- Next Best Offer

e-Science
- Space weather prediction
- Detection of transient events
- Synchrotron atomic research
- Genomic Research

Others

IoT Telematics

Large Engine Manufacturer
- Improved Customer Service Experience
  - Predict maintenance
  - Classify type and severity of issues
  - Minimize service times
- Enablers
  - Full context of all Customer history and current interactions
  - Sensors providing up to the moment telematics
  - ML models, GeoSpatial Analytics, and Business rules
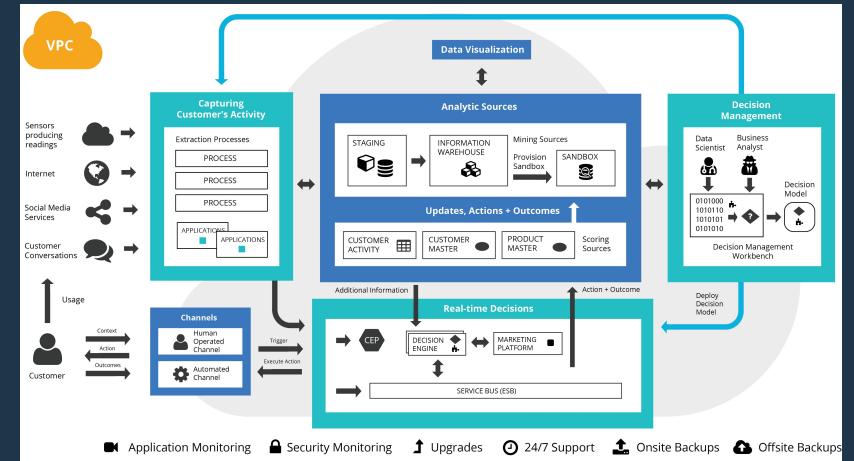
Tour De France Digital Experience
- Improved Fan and Viewer Experience
  - Complete and up to the moment view of all relevant data
  - More engaging and accurate analysis of current events
  - Multi-Channel Access to Insights
- Enablers
  - Bike & rider data, weather & reference data
  - GeoSpatial mapping
  - Window based stats, rankings and groupings
- Dimension Data Video
 https://www.youtube.com/watch?v=uxbMJjL7_u4
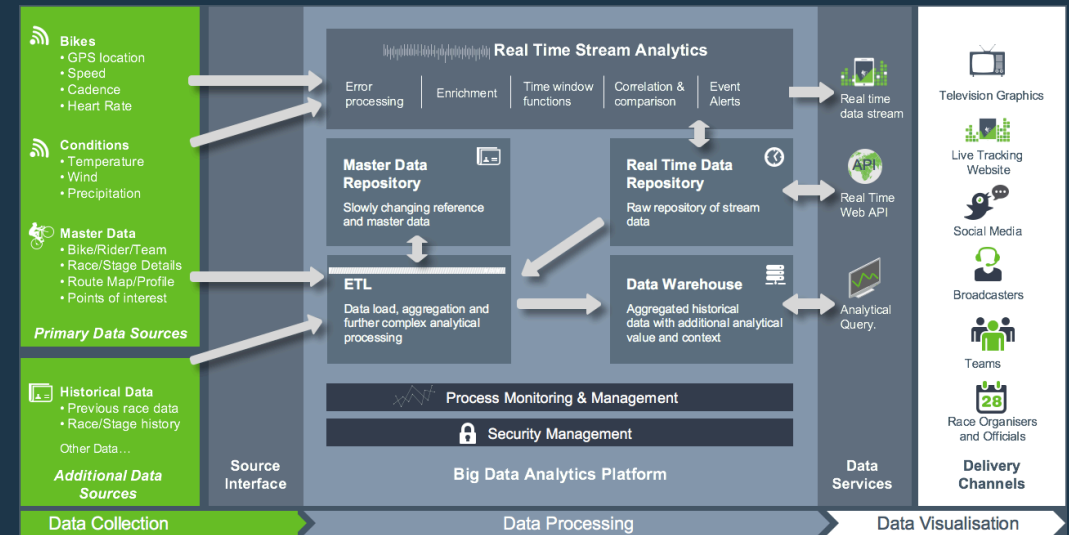
Healthcare Patient Monitoring

Neonatal ICU Monitoring

- Better care and earlier detection of dangerous infection
  - Constantly monitor physiological sensors
  - Detect patterns and predict infections
  - Provide personalized advice to prevent events
  - University of Ontario Institute of Technology
    https://youtu.be/WNccLBzR_l4

Predictive Care for Diabetes

- Personalized experience tailored to the individuals needs
  - Maintain current context and patient profiles
    - Patient history
    - Current behavior & sensor readings
  - Detect patterns and predict events
 - Provide personalized advice to prevent events
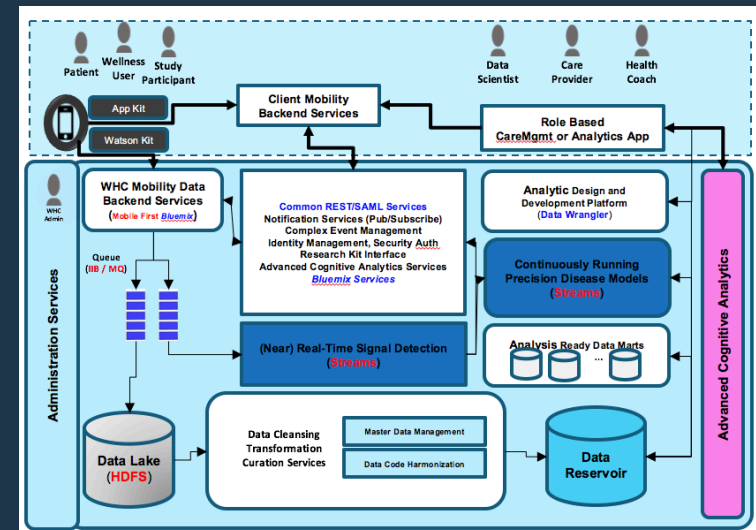
Semiconductor Manufacturing

Improve manufacturing efficiency

- Traditional approach could not keep up
  - 250 million records an hour and growing
- Real-time test result analysis
- Predictions, alerts & diagnostics
- Early detection to prevent wafers from being scrapped

Event Driven Architecture for automatic monitoring

- In-flight analytics comparing real-time and historical results
- Multivariate yield prediction models
- Exploration/experimentation of new analytic applications
- Reuse of data and analytics across applications
- New analysis and modeling tools
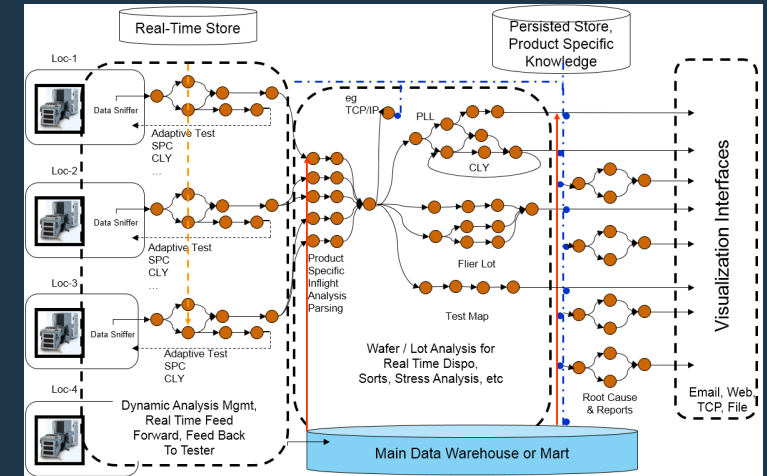
North American Telco Real-Time Advertising

Moved to real-time advertising based on individual profiles

- Clickthrough and revenue up 50% over prior segmentation based marketing/advertising
- In production without failures for over 3 years

Personalised up to date data and full context

- 30M customer profiles continually updated based on cross channel interactions
- CDRs, purchase, usage, TV watching and click stream data

Detecting opportunities and acting in real-time

- 10ms response time to look up customer profile, evaluate models, return ad
- 31B Customer data points processed daily via 500 analytical models
- 220M Customer behavioral events ingested daily
- 32M Customer profiles updated Daily
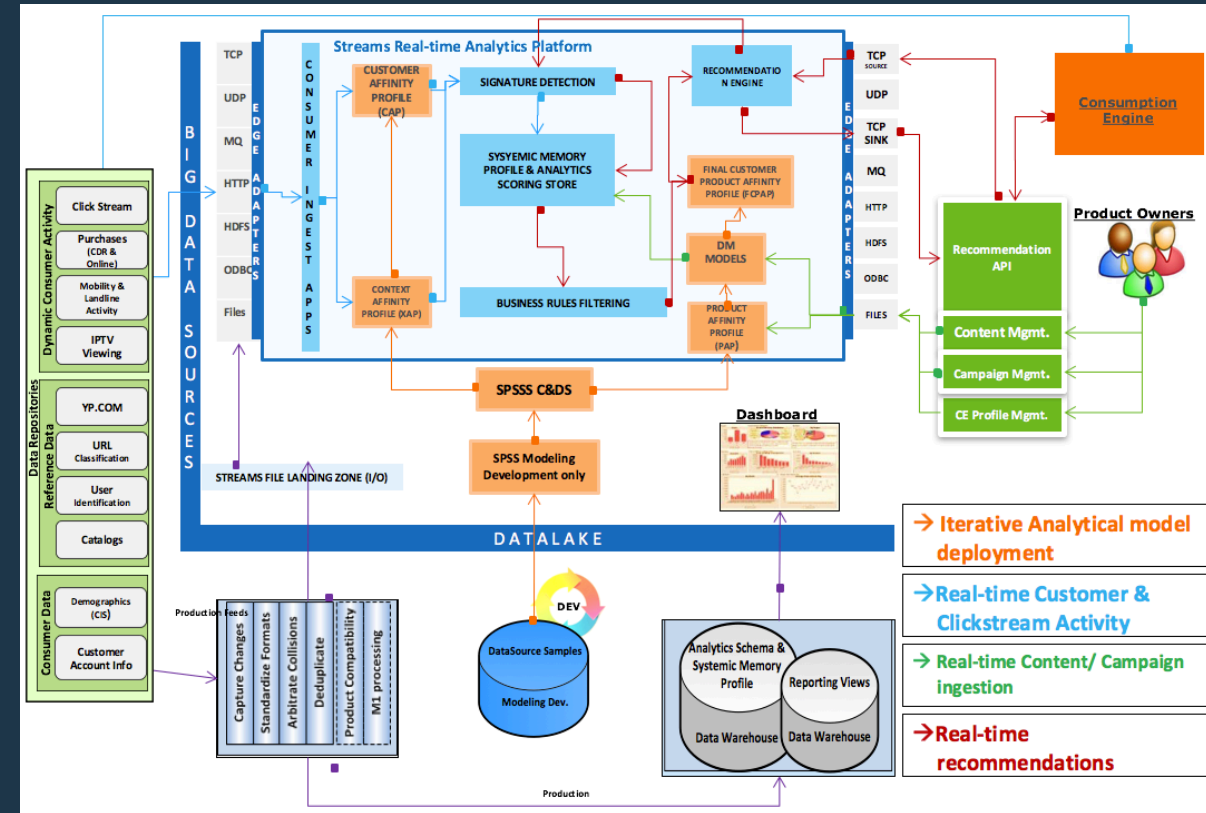
Cognitive Customer Experience

Enabling personalized customer experience

Traditional	Cognitive
• Channel-centric interactions	• Context shared across channels in real-time (cross channel experience)
• Batch analysis in the data warehouse	• Personalization
• Periodic and reactive updates	• Predictive & proactive
	• System self-learns using feedback from users

Know Me • Engage Me • Empower Me
Client

verizon✓

5/22/17 World of Watson 2016

Real-time cognitive customer experience platform

Customer Interactions

Cognitive Platform/Tools

Insights

WMQ — IBM SPSS — redis

IBM Streams

Solr

hadoop kafka Spark

- Call categorization
- Call propensity
- Repeat callers
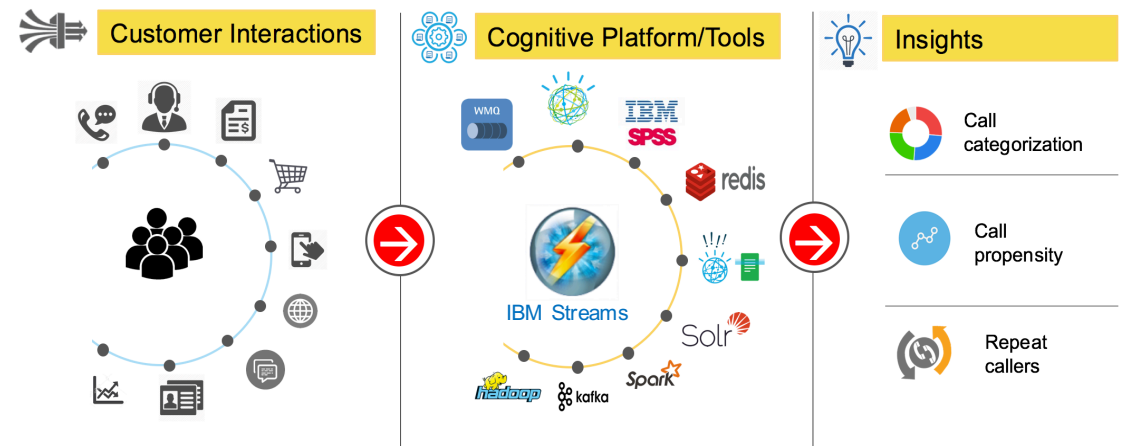
Fast integration with exiting systems and technologies using IBM Streams' pre-packaged connectors

verizon✓

5/22/17 World of Watson 2016 10
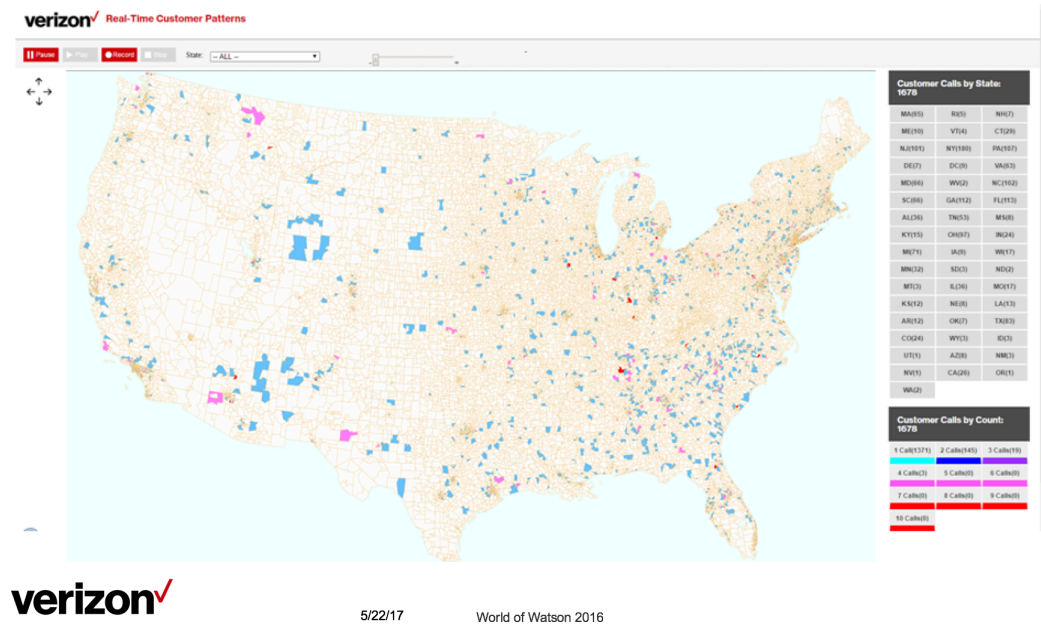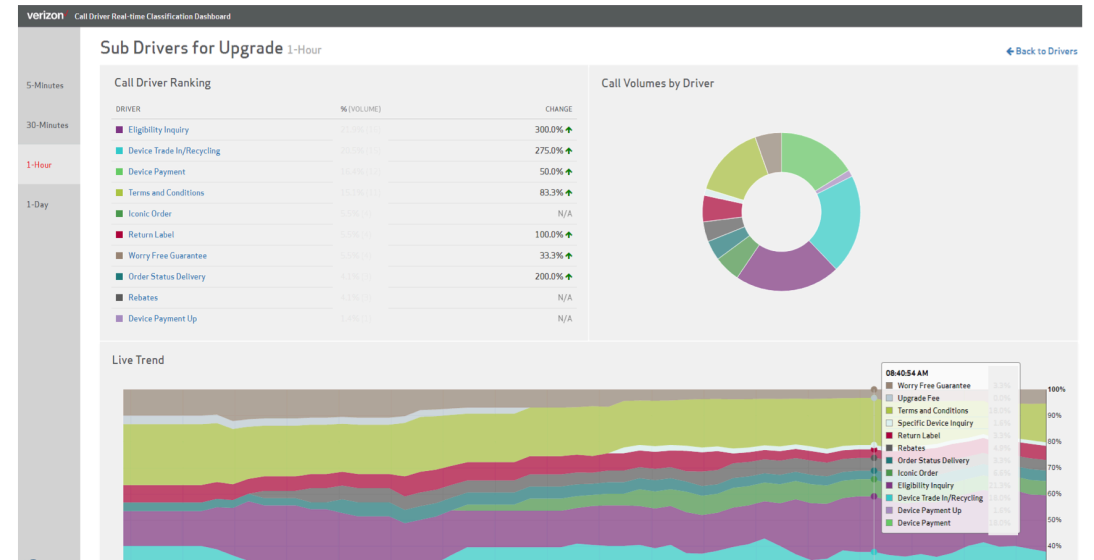
Cognitive Customer Experience

Real-time call patterns – heat map

Real-time call categorization

Cognitive Customer Experience

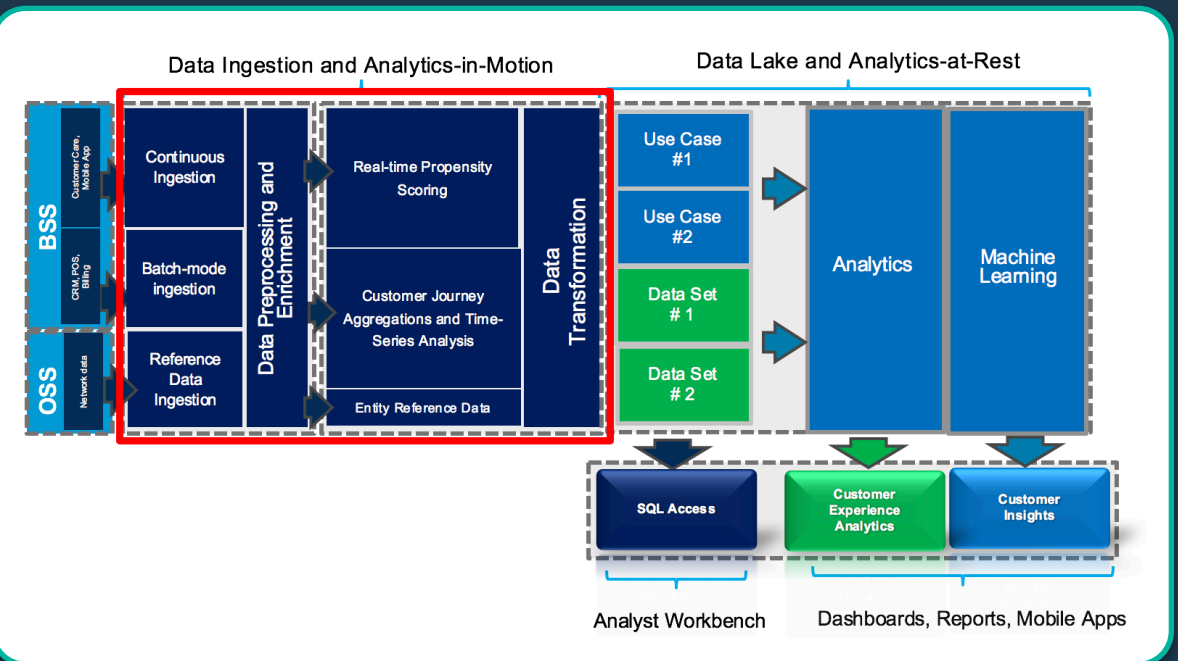## IBM Streams – analyze more, store less, and act NOW

Ingestion micro-applications	Analytics micro-applications	Target systems
▪ Real-time speech	▪ Machine learning pipelines	▪ Watson Explorer
▪ Log records from different system	▪ NLP pipelines	▪ Hadoop
▪ Clickstreams	▪ Aggregations	▪ Data warehouse
▪ Billing	▪ Feed correlations	▪ Relational databases
▪ Agent interactions	▪ Sorting and sequencing	▪ K-V stores

verizon✓ World of Watson 2016   5/22/17

Data Ingestion and Analytics-in-Motion

Data Lake and Analytics-at-Rest

Analyst Workbench

Dashboards, Reports, Mobile Apps

Real-Time is Now

Crossing The Chasm
From Early Adopters to Early Majority

The technology is ready

- Enabling capabilities have evolved and matured
  - Connectivity, messaging, streaming analytics & machine learning
  - Performance, resiliency, scaling & security
- Many large Enterprises have been in production for several years

The benefits are real

- Greatly improving existing solutions and experiences
- Enabling new business opportunities

Adoption is accelerating

- New projects & expanding usage
- More advanced business critical use cases

Why Wait?

Event Driven Architecture for Real-Time Analytics

Current data not stale data

Personalized data not generic data

Full context not partial view

Analytics that detect or predict opportunities NOW not in the past

Actions that effect the outcome not just explaining why it happened

Thank You

The **Watson** & **Cloud** Platform