



# Modernising Your Analytics Insights



# Modernising Your Analytic Insights

With Stream Processing & Learning Models

Creating a culture of agile data innovation:

- speeding up data ingestion, preparation & validation
- driving self-service insights into all the metrics that matter
- streaming & processing data in real-time as events occur
- empowering next-generation predictive & machine learning models

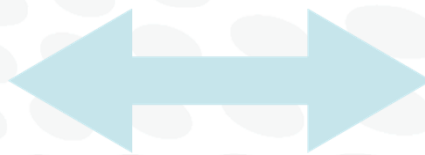
Together, we'll take a look at modern data engineering frameworks, and how our clients are effectively using the capabilities to transform and optimise their decision-making.



## Agenda

- 8:30** Registration
- 9:00** Welcome
- 9:10** **Data & Analytics in a modern world**  
CDP
- 9:40** Coffee break
- 10:00** **Why Streaming is dramatically changing our analytic world**  
Hortonworks
- 10:30** **Moving forward & discussion time**  
CDP and Hortonworks
- 10:45** Presentation end





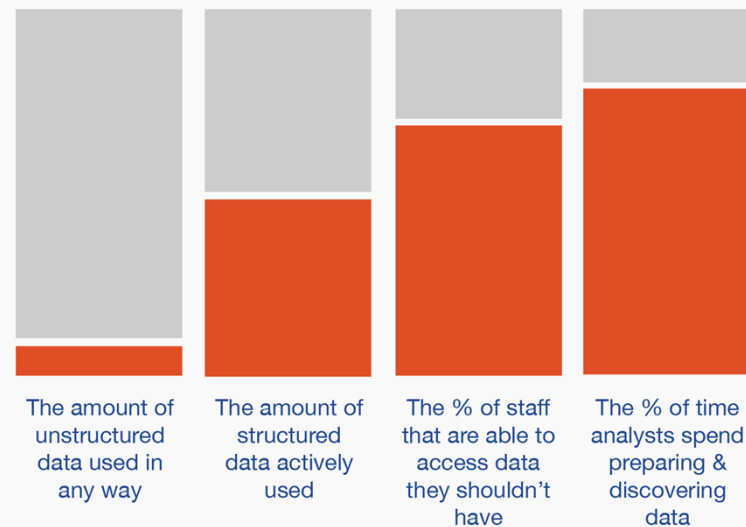
## THE PROBLEM:

- Struggle to utilise the value of data
- Business never stand still
- Limits to traditional Data Warehousing

## THE OPPORTUNITY:

- Use data more effectively
- Leverage data for maximum advantage

### Average utilisation of time and data cross-industry



Harvard Business Review - What's Your Data Strategy? June 2017



# ARCHITECTURE LANDSCAPE



## Why we should embrace the new architecture



Ability to store both structured & unstructured information



Enables realtime event management



Provides capability for realtime dashboards & real time operational reporting



Data Laboratory, supports prediction, machine learning & modelling



Can rapidly ingest new data sources



Data Factory, productionise real time analytics



Massively scalable with low storage costs



Supports agile development approaches



Services based architecture



Is a modern, cost effective & future-proofed framework



# TRADITIONAL vs NEW ARCHITECTURE

## Traditional

- Optimised for batch processed data processes
- Large development effort
- Model data on load
- Mostly structured data
- Monitoring and measuring the business



## New Architecture

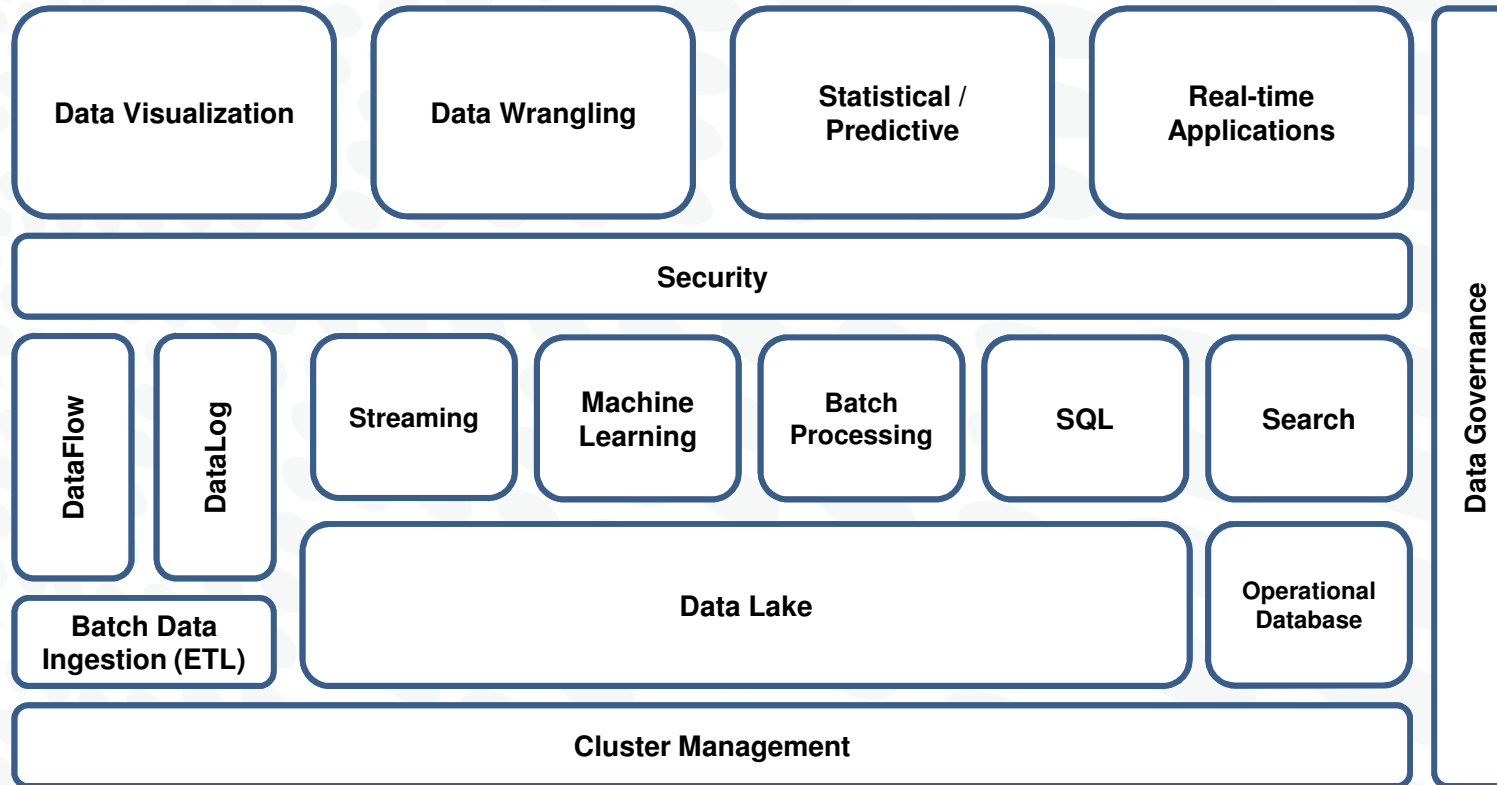
- Augmentation of existing environment
- Streamed data or Historical archiving
- Fast Capture of a variety information sources –
- Model data on use
- Supports a wide variety use cases not traditionally seen in DWH
- Driving the business forwards







# ARCHITECTURE LANDSCAPE





## DELIVERY FOCUS

### Foundation

- Data Lake
- Distributed processing
- Security
- Batch Ingestion
- SQL-like access Interface

### Streaming

- Real-time data ingestion
- Simple & Complex event process

### Data Discovery

- Data wrangling
- Data visualisation

### Advanced Analytics

- Machine Learning
- Predictive

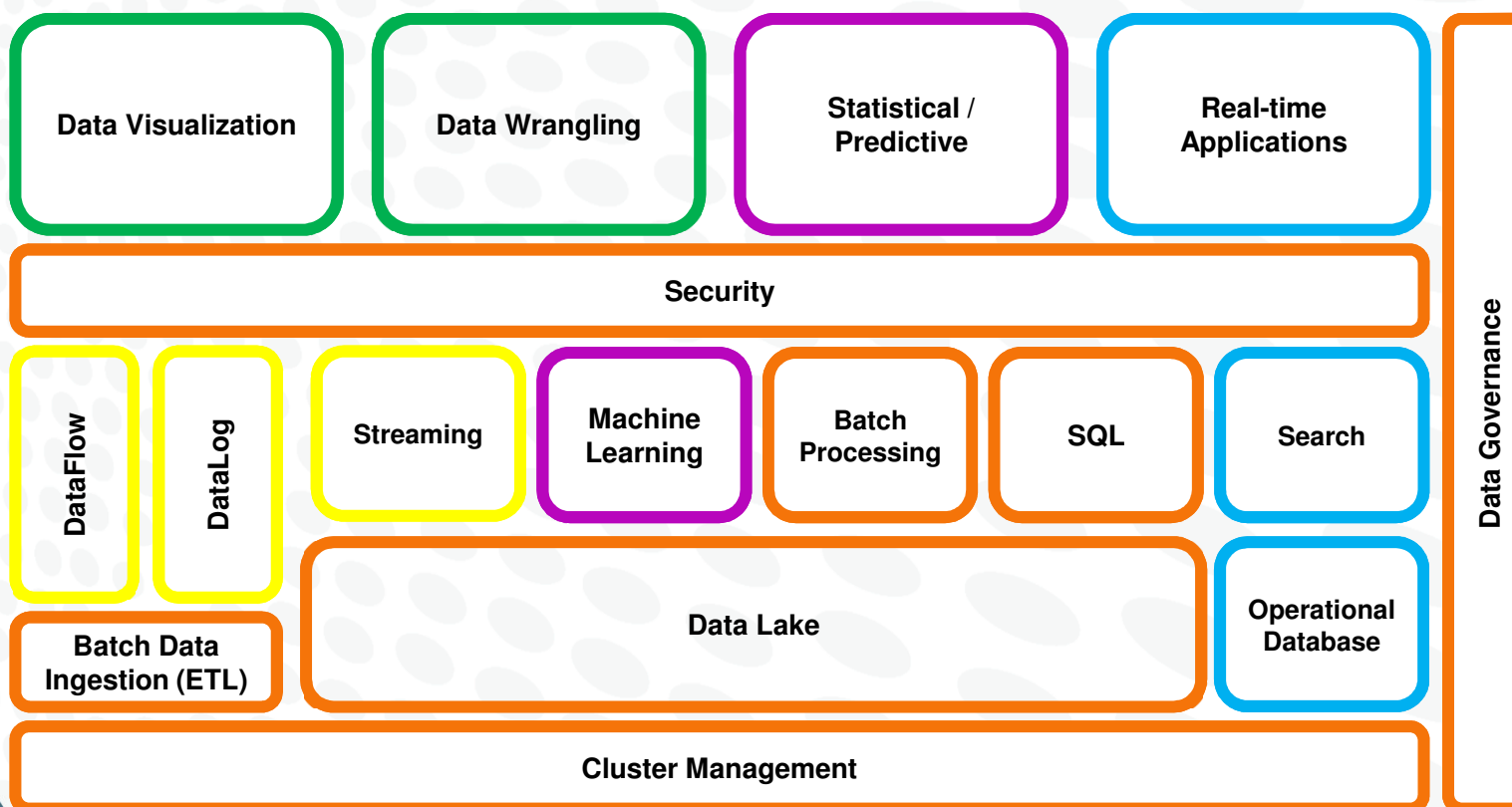
### Specialised Components

- Real-time application
- Search capability
- Operational database
- Time Series database





## SESSION FOCUS





## 1. Foundation

2. Data Discovery

3. Advanced Analytics

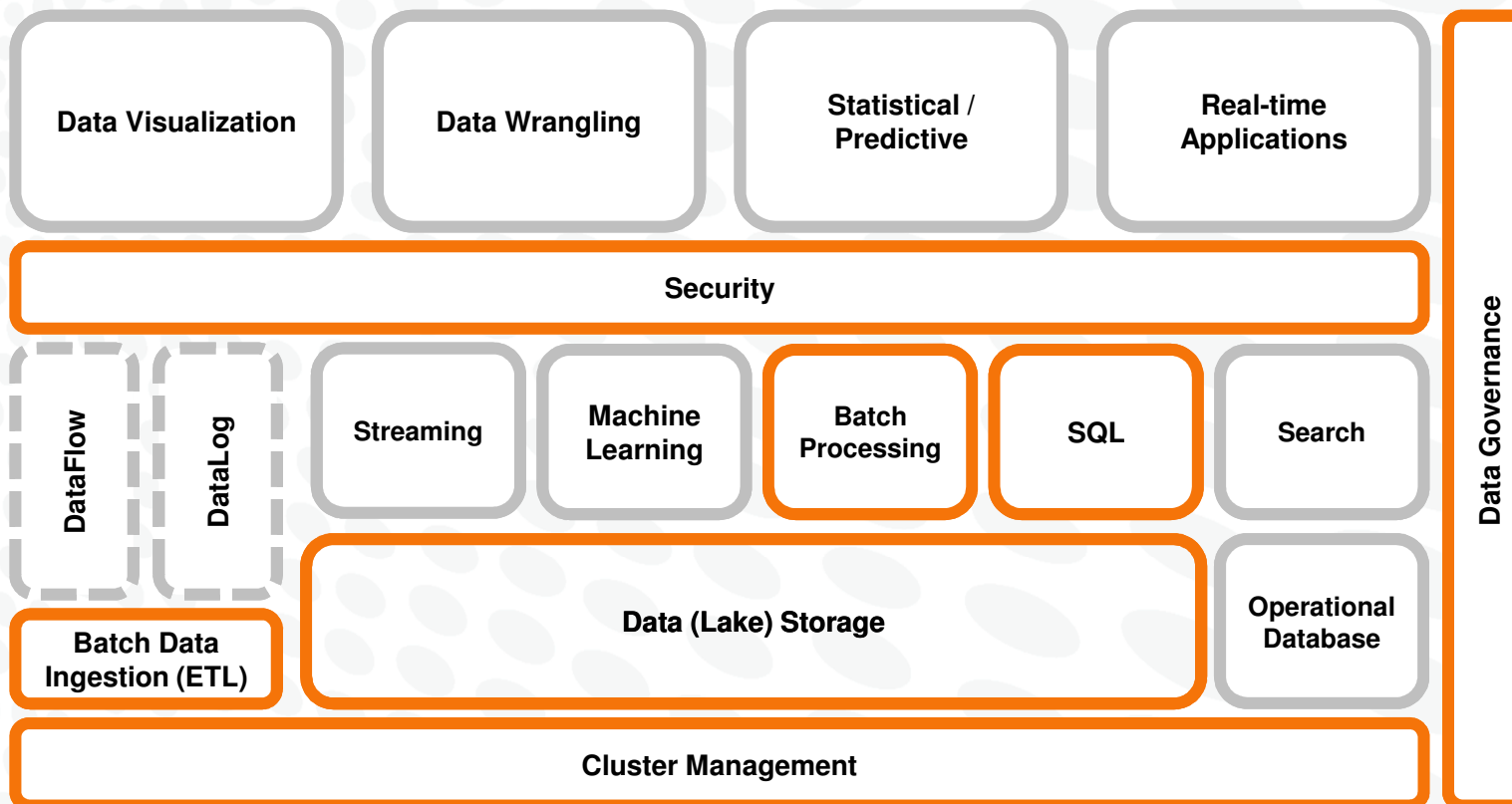
4. Streaming

5. Specialist Components





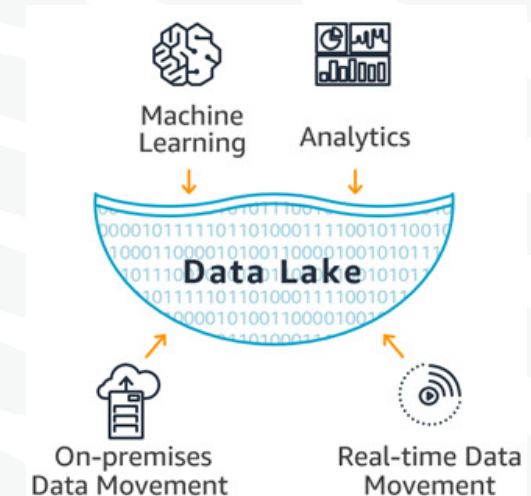
# FOUNDATION



## Data Lake

### Many use cases amongst others

- ✓ An environment that ensures optimal use of structured and unstructured data to deliver business value
- ✓ Central repository of critical and sensitive data
- ✓ Data maintained over long duration
- ✓ Users can access and analyze data in new and different ways
- ✓ Conceptual not technical environment







## CUSTOMER FOUNDATION USE CASES

- ✓ Analytical Data Lake
- ✓ Data ingestion Layer





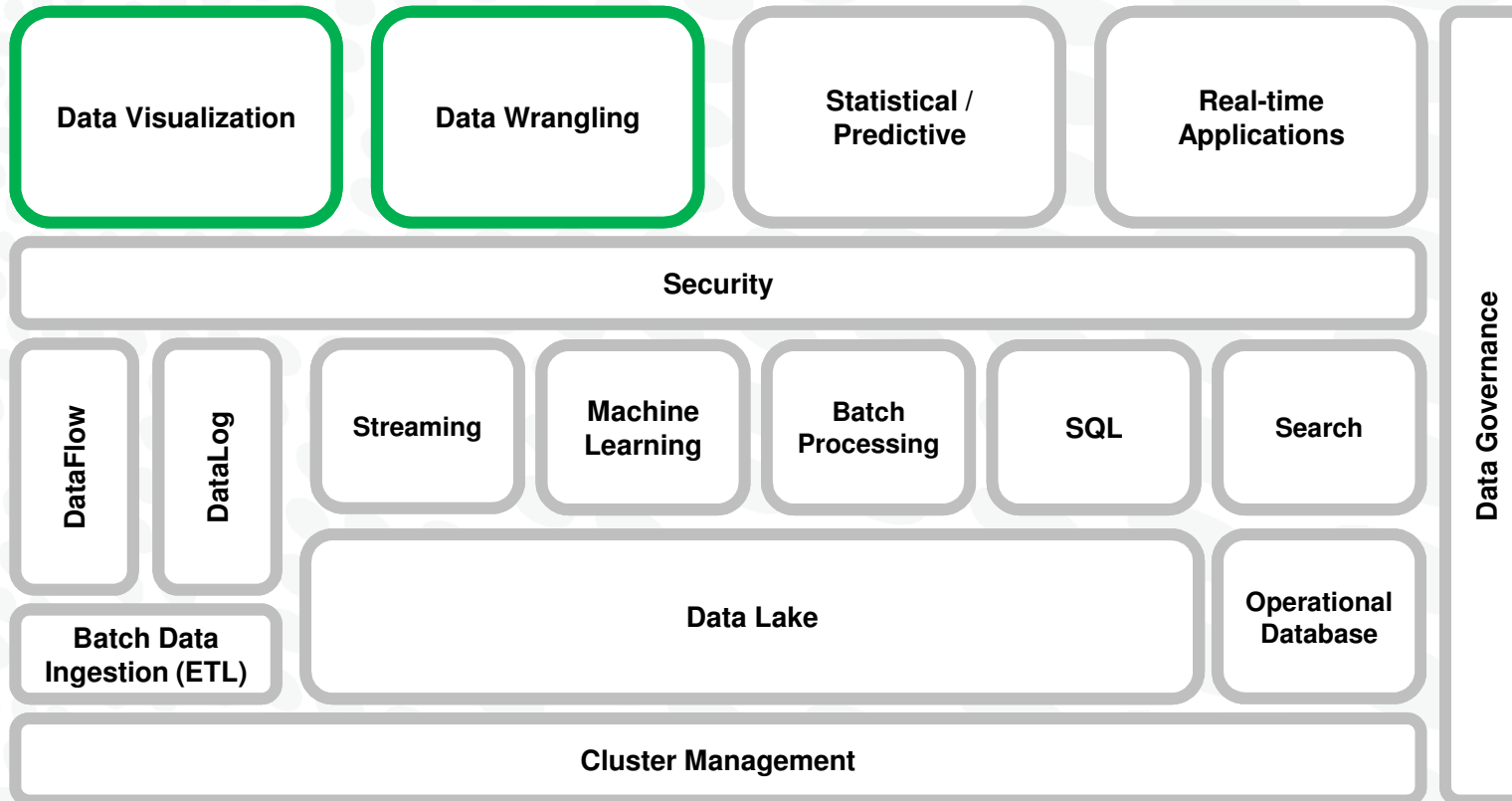


1. Foundation
2. Data Discovery
3. Advanced Analytics
4. Streaming
5. Specialist Components





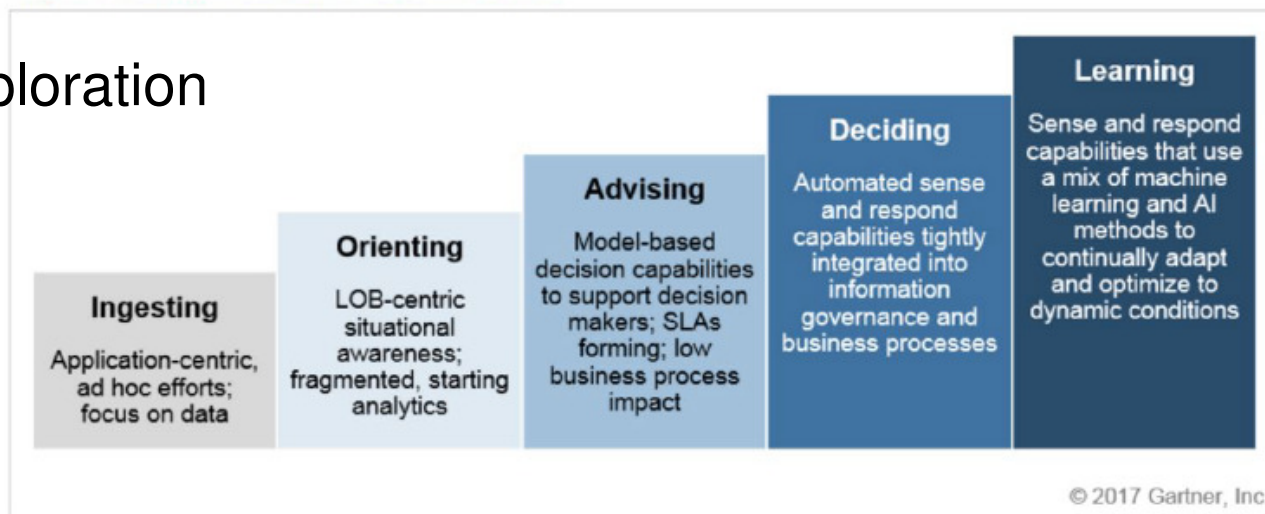
# DATA DISCOVERY



## Data Wrangling/ Data Exploration/ Data Understanding

“Data wrangling is an iterative agile process for **exploring, combining, cleaning** and **transforming** raw data into curated datasets for data science, data discovery, business intelligence and analytics.”

### Data Exploration



AI = artificial intelligence; LOB = line of business; SLA = service-level agreement

Source: Gartner (April 2017)



## DATA DISCOVERY USE CASES

- ✓ Data preparation and visualisation
- ✓ Data wrangling process, preparation for analytics



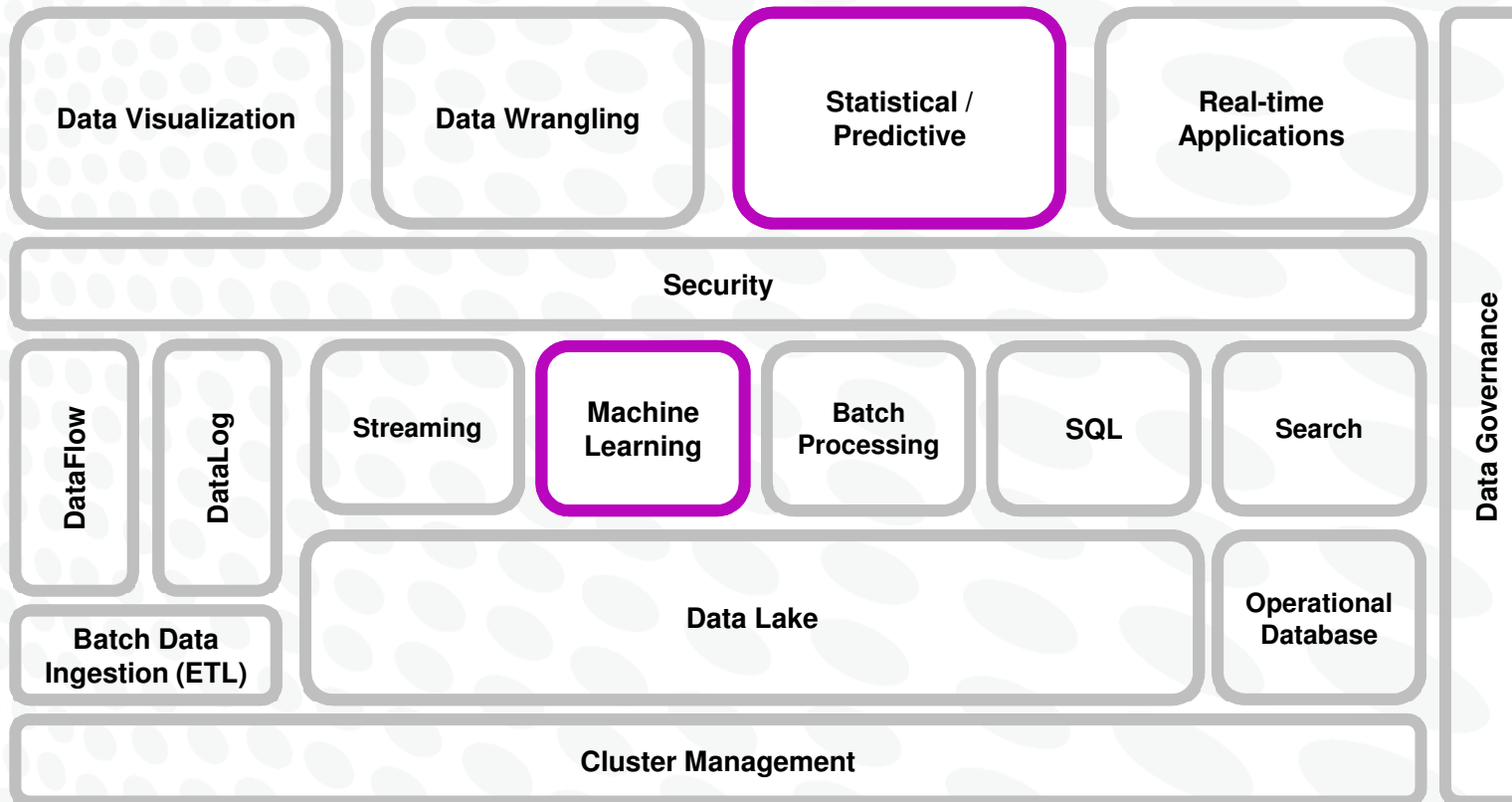


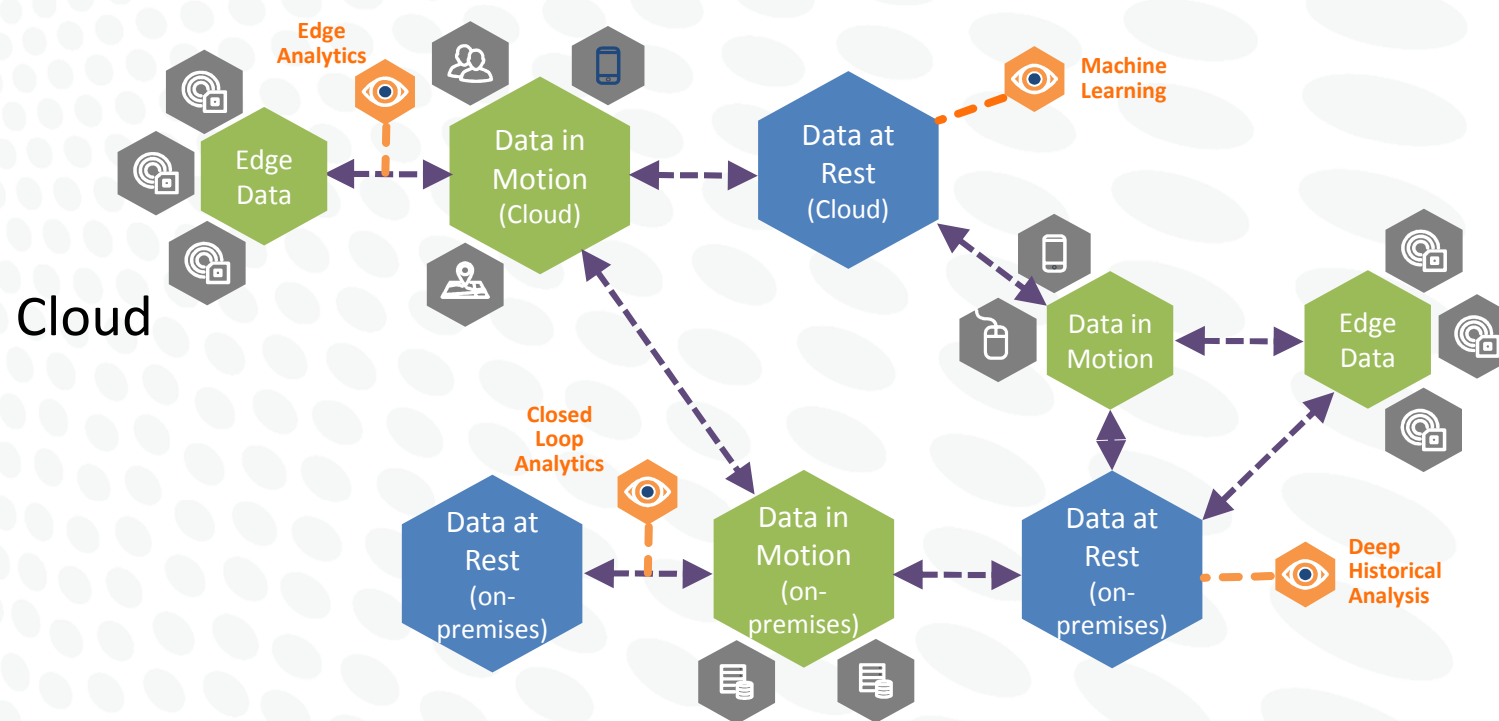
1. Foundation
2. Data Discovery
3. Advanced Analytics
4. Streaming
5. Specialist Components





# ADVANCED ANALYTICS





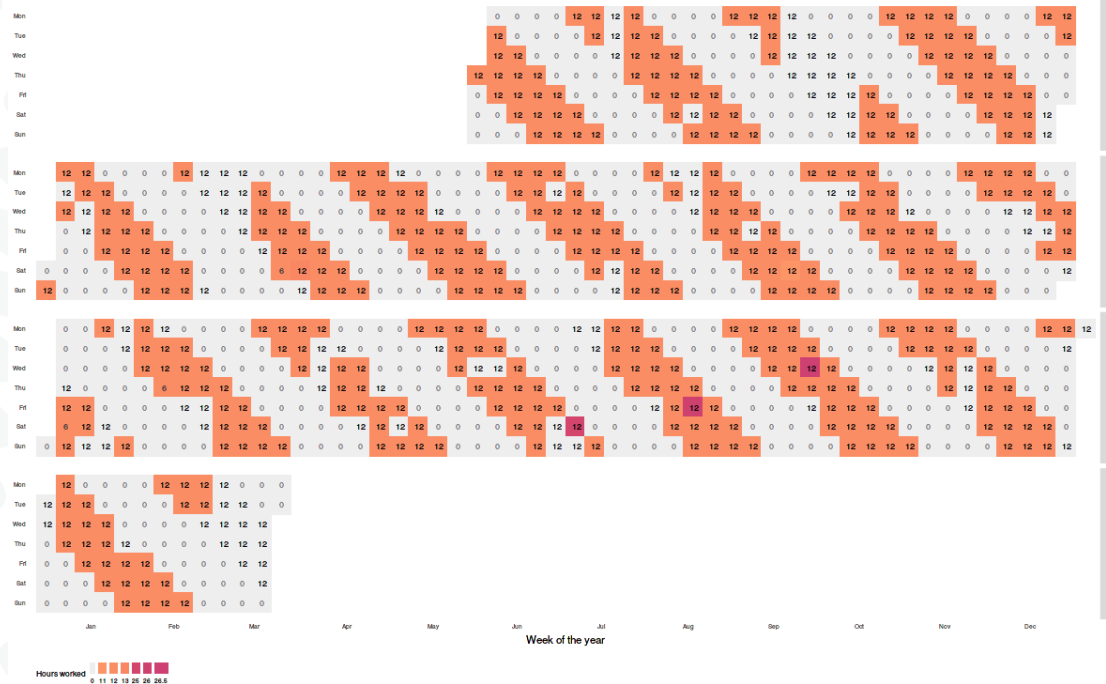


# ADVANCED ANALYTICS USE CASES

✓ Work pattern time series analysis

✓ Risk prediction

Employee 00021228 - Site 417 - Scheduled  
Scheduled working hours and expected hours given the work pattern





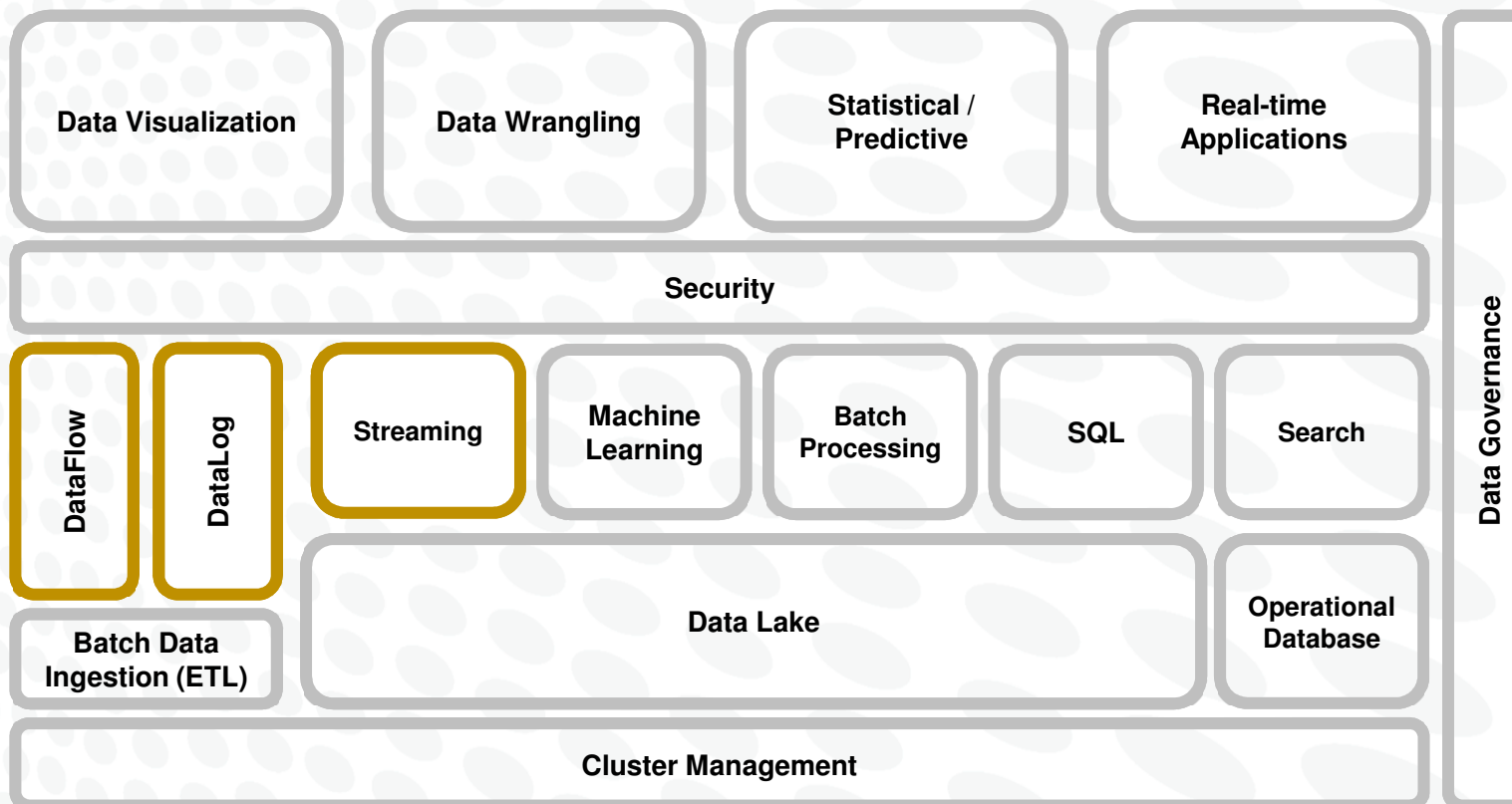


1. Foundation
2. Data Discovery
3. Advanced Analytics
4. Streaming
5. Specialist Components





# STREAMING

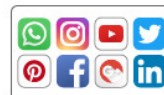




# STREAMING

Apply Lightweight Adaptable Streaming SQL

## COLLECT



Apps



Services



Devices

## ANALYZE

Streaming SQL

### Stream Processor

Stream Analytics  
Stream Data Integration

Filtering & Transformation  
Complex Event Processing  
Real-time ETL  
ML Model Serving  
Streaming ML  
Anomaly Detection  
Incremental Aggregation

## ACT

APIs



Actions



Alerts



Visualizations



## PERSIST

Analytics  
Stores



Event to Action within Millisecond

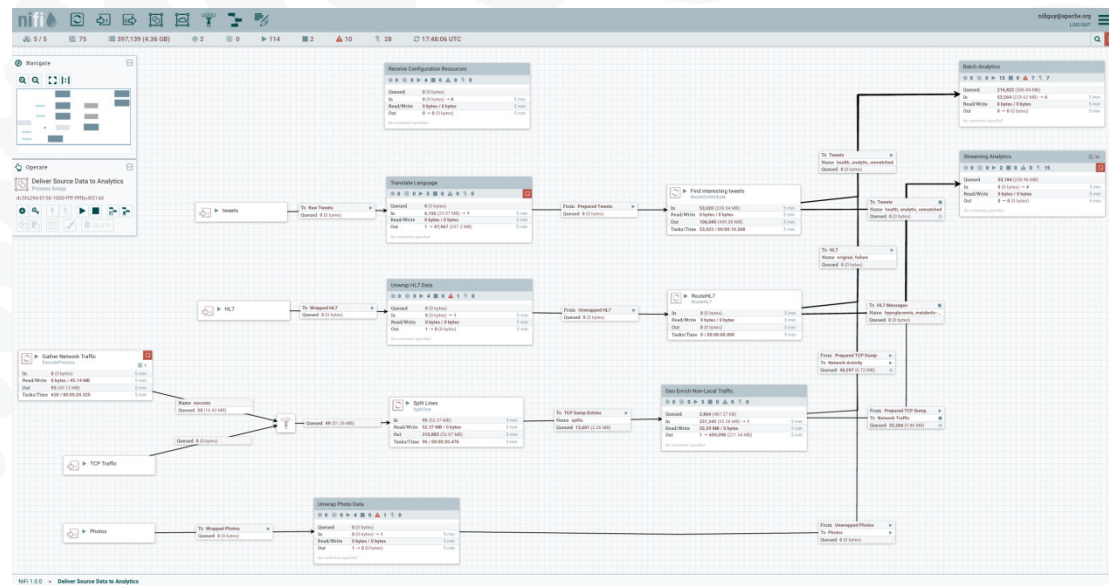
100K+ throughput with just two nodes, scale to millions with Kafka





# STREAMING USE CASES

- ✓ Retail real time cash slip processing to detect anomalies and patterns
- ✓ IOT device data collection or transport patterns (late busses)



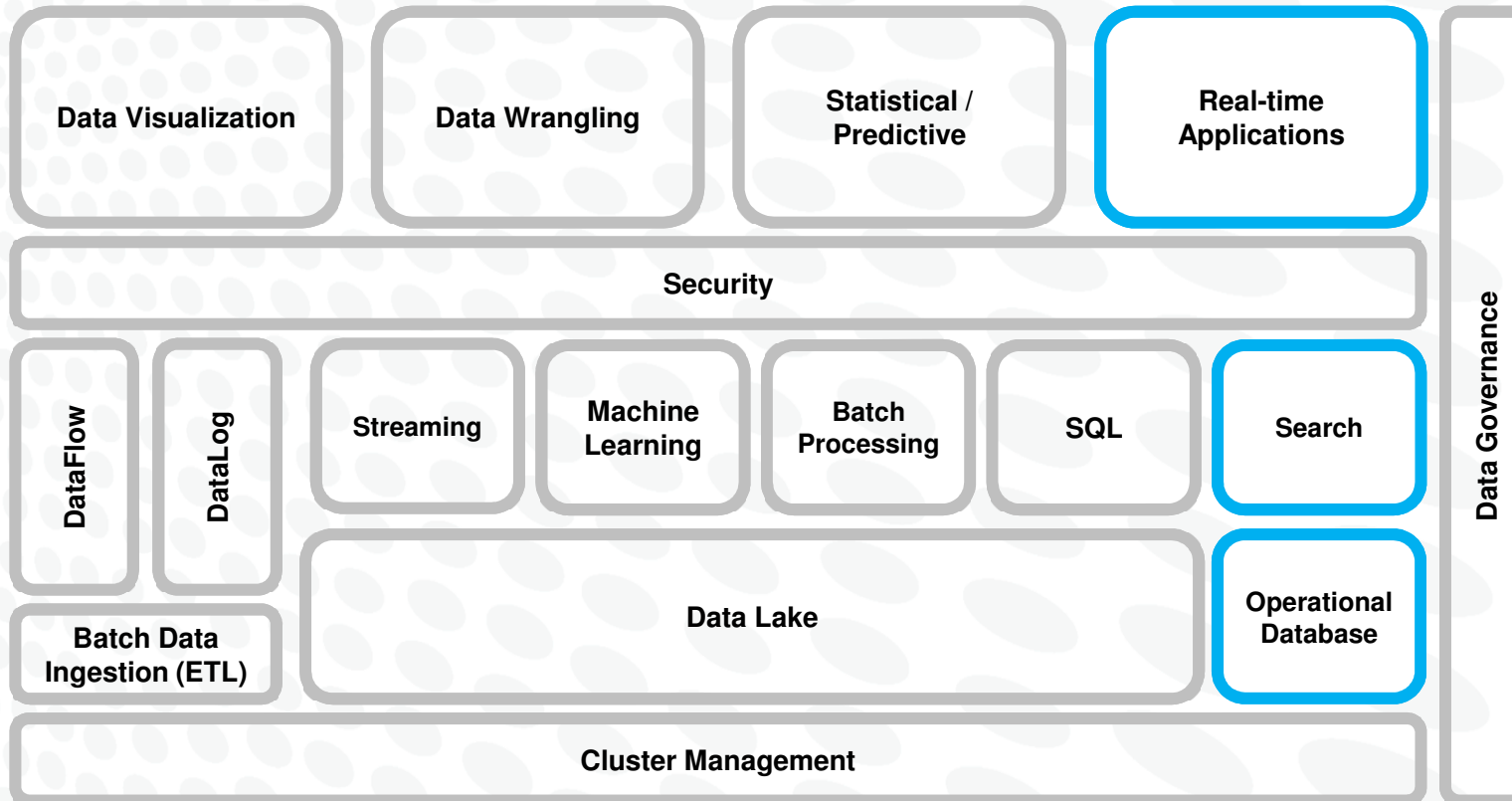


1. Foundation
2. Data Discovery
3. Advanced Analytics
4. Streaming
5. Specialist Components



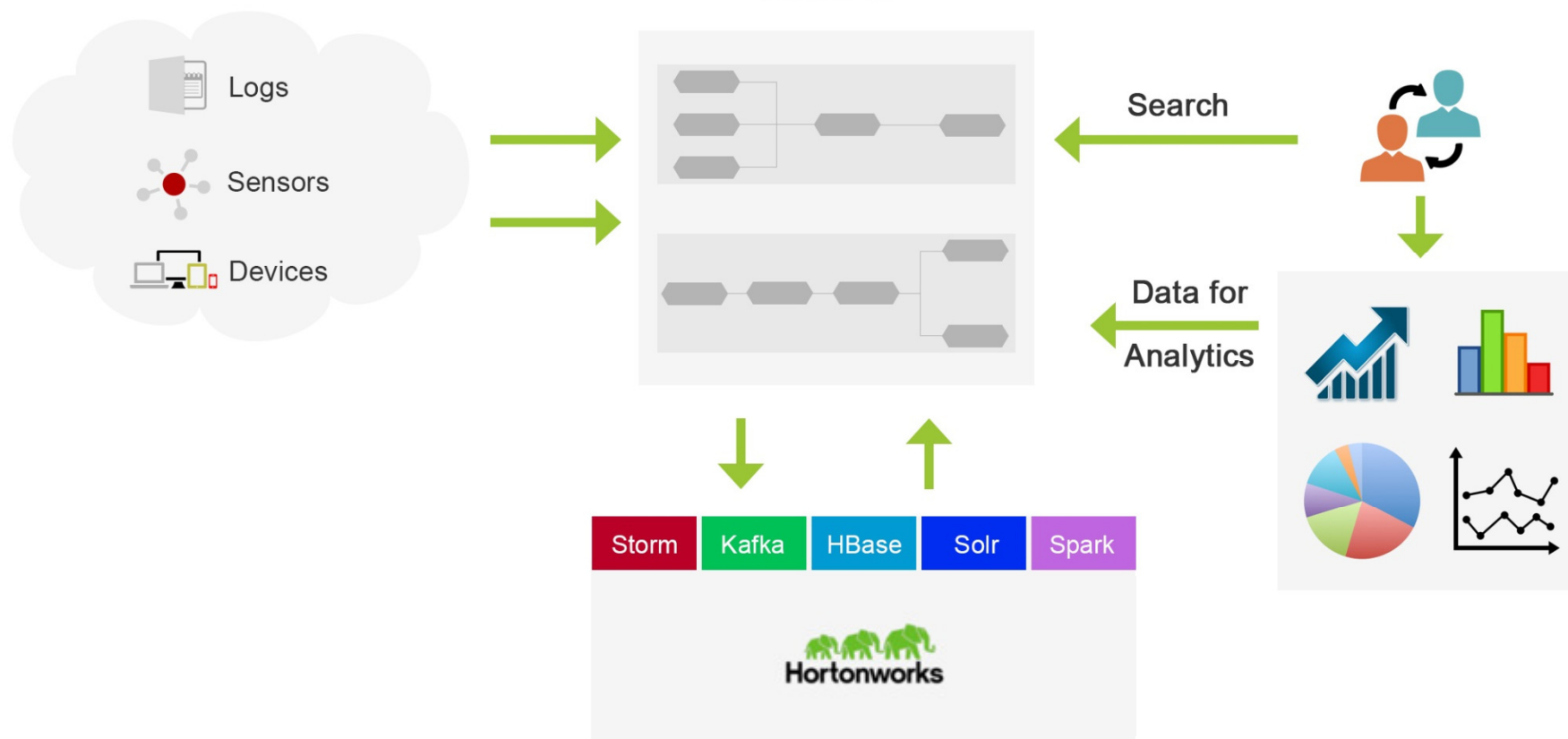


## SPECIALIST COMPONENTS



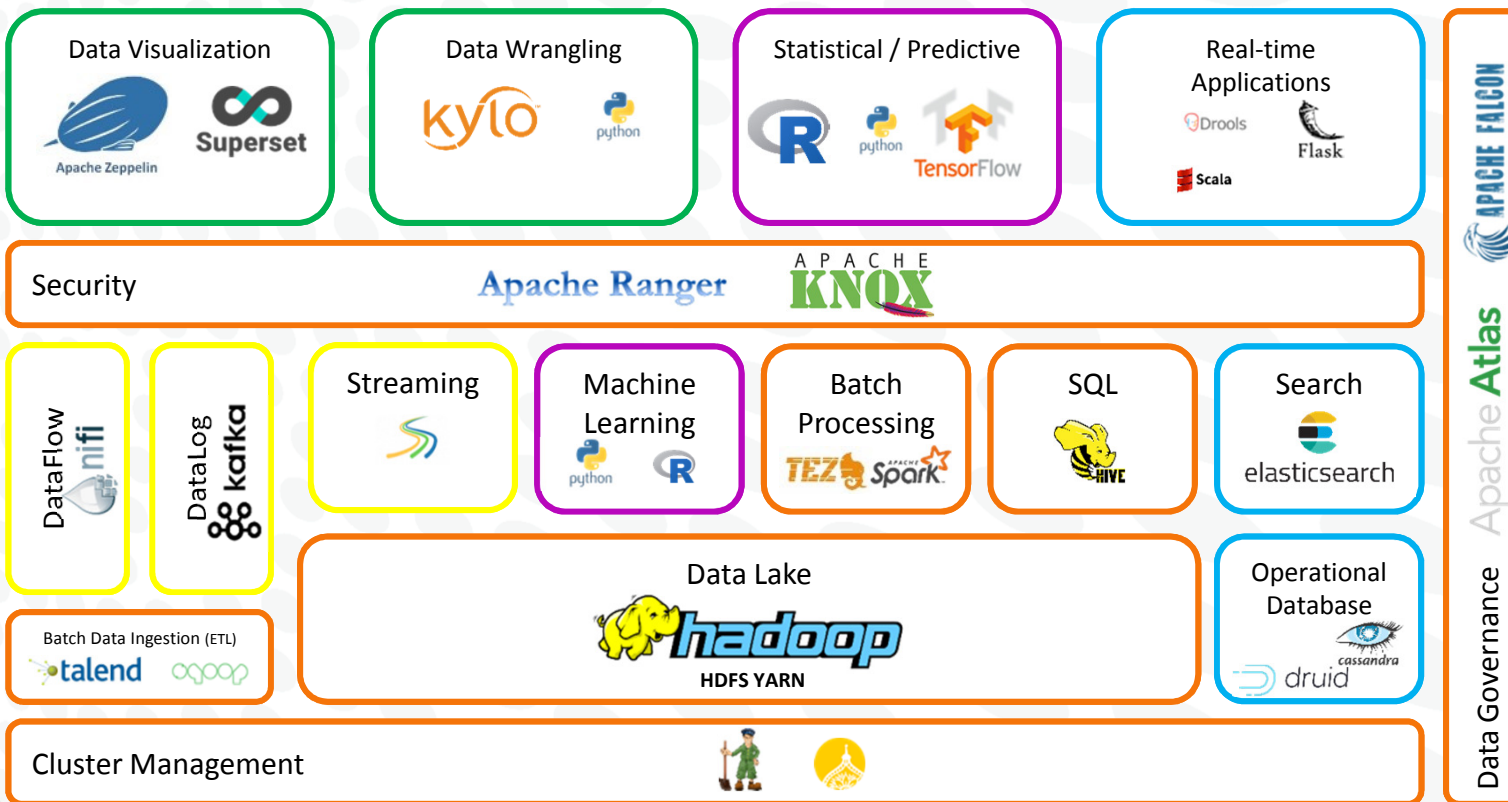


## SPECIALIST COMPONENTS





# OPEN TECHNOLOGY LANDSCAPE







# HYBRID TECHNOLOGY LANDSCAPE

Transition and Augmentation



Data Visualization



Data Wrangling



Statistical /



Real-time Applications



Security

Apache Ranger



DataFlow



Data in Motion



Streami



Machine Learning



Batch Processi



SQL



Search



Batch Data Ingestion



HDF



Data at Rest



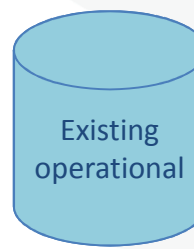
HDP



Cluster Management



APACHE FALCON  
Data Governance Apache Atlas



Existing ETL





Coffee break



## Agenda

- 8:30** Registration
- 9:00** Welcome
- 9:10** **Data & Analytics in a modern world**  
CDP
- 9:40** Coffee break
- 10:00** **Why Streaming is dramatically changing our analytic world**  
Hortonworks
- 10:30** **Moving forward & discussion time**  
CDP and Hortonworks
- 10:45** Presentation end





## CDP GROUP – WHO ARE WE

specialist consultants for analytics, planning & data integration



dashboards &  
visualisation



data integration



planning &  
forecasting

**GOAL:** Driving insight into your business

### OUR FOCUS:

- Design & develop analytics, planning & data engineering solutions
- Enhance your ability to make high-impact & factually informed decisions
- Reputation based on our deep level of expertise



1



✓ **Augmented Analytics with built-in AI**

- Contextual visualisation recommendations
- Automated blend & join data-preparation
- Smart natural language

2



✓ **Data Science Platform for collaboration**

- Incorporate all your R, Python, Spark & others
- Model & library management
- Version & deployment control

3



✓ **SQL Workload Enhancement over HIVE**

- Rich & complex SQL with outstanding performance
- Re-use existing DW queries against Hadoop
- Federate DW in place augmented with Hadoop



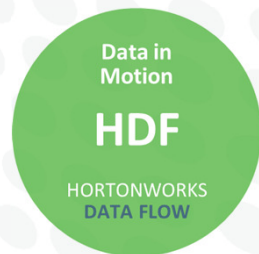
# HORTONWORKS OPEN-SOURCE

1



- ✓ Modern data architecture
- Scalable, secure, hybrid data lakes
  - Enables deep / machine learning

2



- ✓ Real-time streaming platform
- Data pipelines, ingestion & routing
  - High-volumes from sensors & devices





Hortonworks



# ● HOW DO YOU BEGIN THE MODERN ANALYTICS JOURNEY?





# PERFORMANCE PLAN

Our Performance Plan approach:

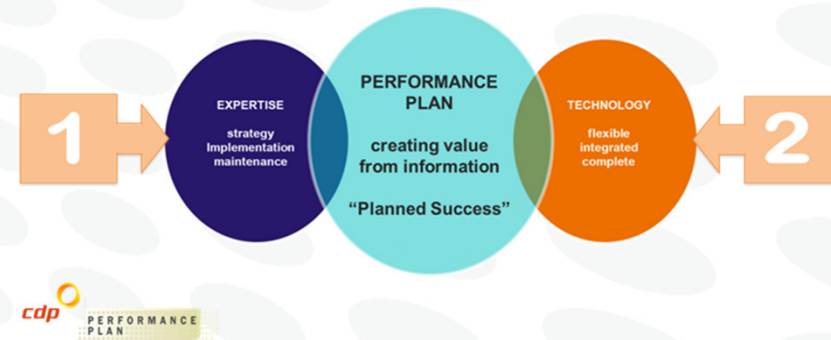
- identifies core business information requirements
- measures against current capability
- creates a plan to deliver.

Information, once captured:

- presented back as a templated plan
- road map of prioritised quick-win deliveries.

CDP is offering a first steps Performance Plan engagement from your attendance today:

- 5 Days - \$7.5k



Typical initial delivery  
20 days or \$30k





Thank you