Analyzing BigData with Machine Learning and Hadoop Clusters

Microsoft

Sudhir Rawat, Data Engineer, Microsoft (@rawatsudhir)

saltmarch
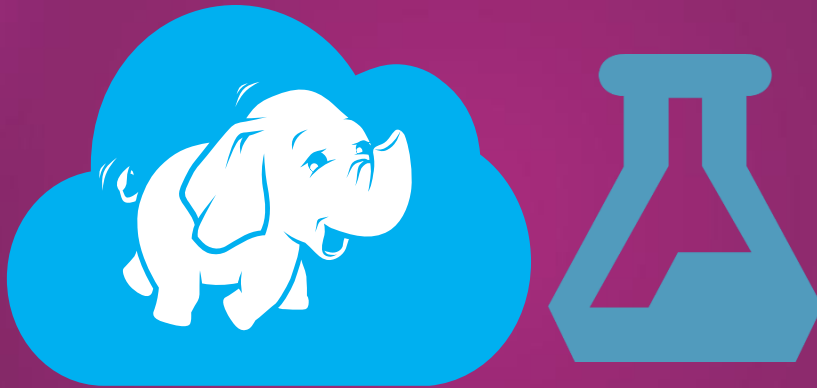GREAT INDIAN
DEVELOPER
SUMMIT

Agenda

- Capabilities of HDInsight and AML
- Build solution using HDInsight and AML
- Demos

BREAK APART....

Azure HDInsight

Managed 100% Apache Hadoop

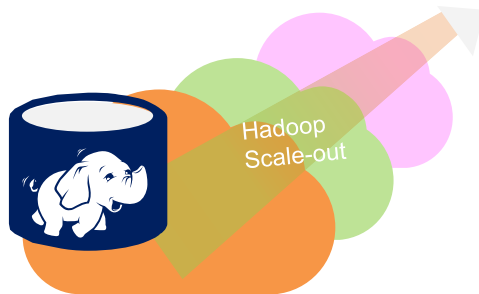Process structured, semi-structured, unstructured data

Scale elastically on demand

Powered by Azure

➡ Big Data Meets The Cloud

99.9% availability
Azure SLA

Hadoop Scale-out

Terabytes to

Petabytes
Scale-out

Deployed in

minutes
Within a few clicks

Azure HDInsight running Windows/Linux

- Choice of Windows or Linux clusters
  - Managed & supported by Microsoft
  - Re-use common tools, documentation, samples from Hadoop/Linux ecosystem
  - Add Hadoop projects that were authored on Linux to HDInsight
  - Easier transition from on-premises to cloud

ubuntu®

Windows Server
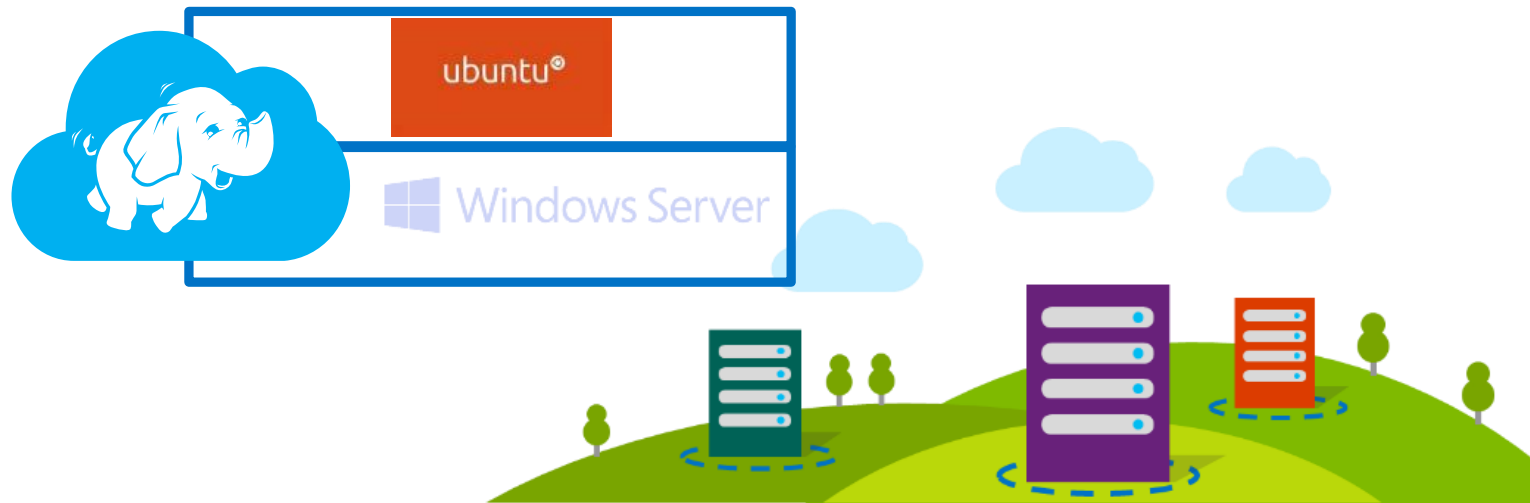
Big Data @ Microsoft - Options
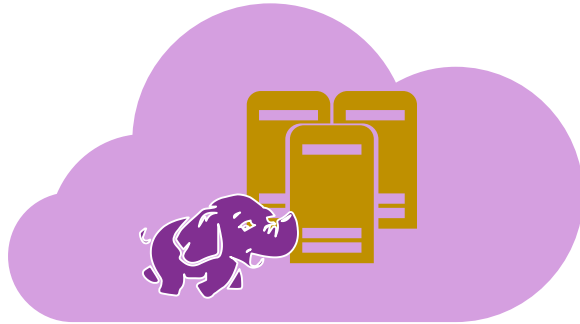
Hadoop in IaaS

Pros
- ❖ Complete Control
- ❖ On-Demand Cluster Sizing
- ❖ Storage - Local or Cloud

Cons
- ❖ Only VMs managed for HA
- ❖ Administration required
- ❖ Clusters need to stay active

Azure Data Lake Managed Clusters

Pros
- ❖ Fully managed – SLA bound
- ❖ Flexible resizing
- ❖ Pay-on-use
- ❖ Customization Options
- ❖ Deployed in minutes

Cons
- ❖ Forgo some control

Azure Data Lake Analytics

Pros
- ❖ Abstracted from clusters
- ❖ Automated resource alignment
- ❖ Easy to use interface and APIs
- ❖ Familiar languages
- ❖ Faster time to market

Cons
- ❖ None

Demo

Azure Data Lake Analytics

AZURE ML

Let's start with a game...

Delivering Advanced Analytics

Business users access results from anywhere, on any device

Data

Cloud
- HDInsight
- SQL Server VM
- SQL DB
- Blobs & Tables

Local
- Desktop files
- Excel spreadsheet
- Other data files on PC

Microsoft Azure Machine Learning

Data to model to web services in minutes

ML Studio

Integrated development environment for Machine Learning

Storage space

API

Web

Model is now a web service

Microsoft Azure Marketplace

Monetize this API

Clients

Devices   Applications   Dashboards

Business problem   Modeling   Deployment   Business value

GREAT INDIAN DEVELOPER SUMMIT

www.developersummit.com

How can Advanced Analytics help you?

EXAMPLE SOLUTIONS

| Finance and risk	Sales and marketing	Customer and channel	Operations and workforce
Revenue Forecasting	Sales forecasting	User segmentation	Agent allocation
Portfolio optimization	Demand forecasting	Personalized offers	Warehouse efficiency
Investment modelling	Sales lead scoring	Product recommendation	Smart buildings
Fraud detection	Marketing mix optimization		Predictive maintenance
Risk management			Supply chain optimization

Demo

BRING IT TOGETHER....

Conceptual
Cortana Analytics Suite - Layer Stack
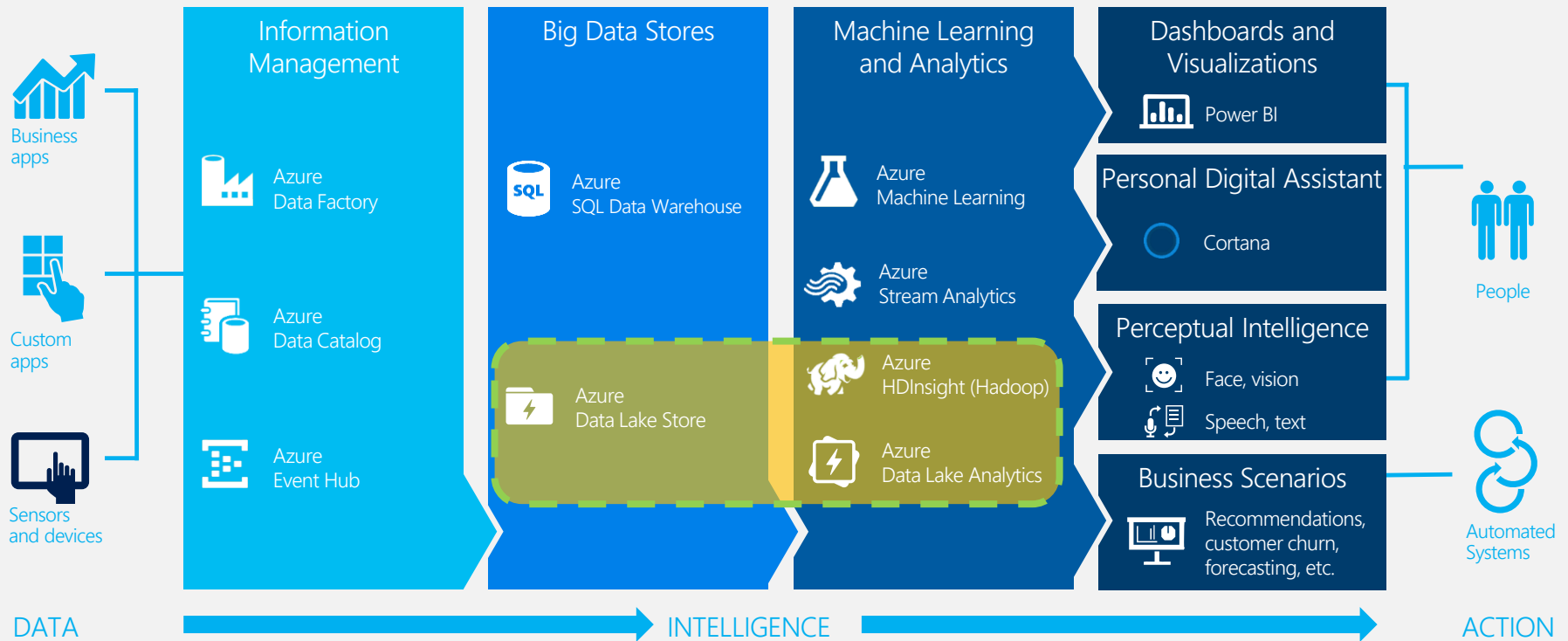
Logistics – IoT use case

Background
- One of the leading multinational automobile corporations that is one of the largest companies in the world by revenue. They manufacture over 10 million vehicles a year.
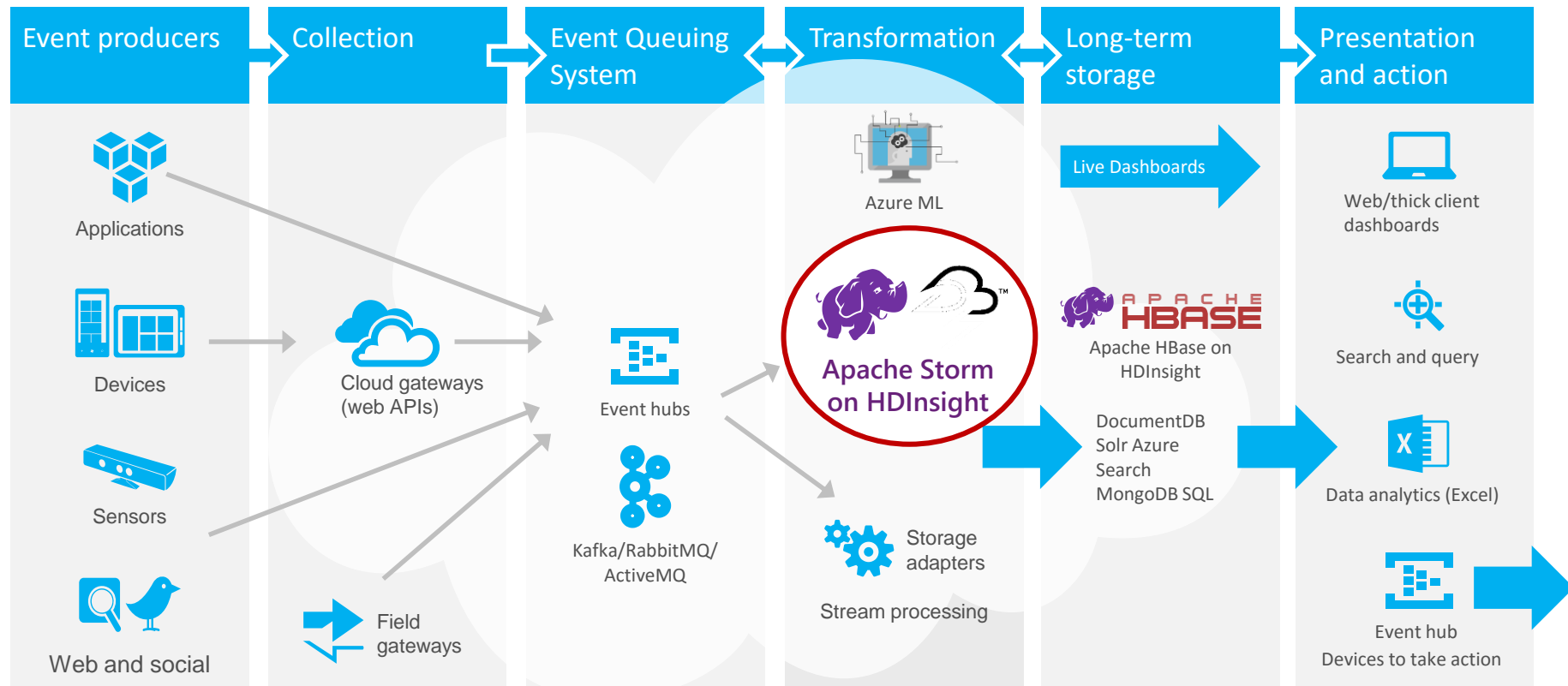
Problem
- Needed to analyze the telemetry.
- Wanted to build a scalable, reliable, and highly available solution that has the ability to receive and process a large volume of vehicle information and maintenance events

Solution
- Use Azure Data Lake, HDInsight, Storm in HDInsight, HBase in HDInsight, Event Hubs, Machine Learning, and Power BI
- Collect IoT data from automobiles:
  - Telemetry data comes in real-time
  - Able to process and generate insights around vehicle information and maintenance events

Fleet Management – Data Flow

Event producers	Collection	Event Queuing System	Transformation	Long-term storage	Presentation and action
Applications			Azure ML	Live Dashboards	Web/thick client dashboards
Devices	Cloud gateways (web APIs)	Event hubs	Apache Storm on HDInsight	Apache HBase on HDInsight	Search and query
Sensors		Kafka/RabbitMQ/ ActiveMQ	Storage adapters	DocumentDB Solr Azure Search MongoDB SQL	Data analytics (Excel)
Web and social	Field gateways		Stream processing		Event hub Devices to take action

Demo