

IBM



Cloud
Data
Services.

IBM Cloud Data Services

Put data to work
with advanced analytics on Cloud



IBM Cloud Data Services

The Sky's The Limit

Kevin McIntyre

WW Sales & Business Development, Cloud Data Services

IBM Analytics



Today

Topic	Speaker
The Sky's the limit: An overview of IBM's data solutions on cloud	Kevin McIntyre Sales & Business Development Leader Cloud Data Services IBM® World Wide
Enterprise-grade NoSQL for web and mobile data	Adam Kocoloski CTO Cloud Data Services, IBM World Wide
Cloud data warehousing for the next generation of Builders	
Tackling Big Data with Hadoop service via Cloud	Chris Howard Analytics Architect IBM Analytics, Asia Pacific
IBM and Spark for future fast analytics.	
The Big Knowledge Project Platform for Media Analytics leveraging IBM Cloud Data Services	Shiv Sehgal Solutions Architect RSG Media Systems



Agenda

- **Introduction to IBM Cloud Data Services**
- **Let's Bust Some Cloud Myths**
- **Why IBM for your Cloud Analytics Journey?**
- **Success Stories**



Let's Level Set

- **Hosting Services (IaaS – Infrastructure as a Service):** Bring your software licenses to be hosted. You are still responsible for the management and operations of the environment.
 - Examples: Hosting your software licenses on IBM SoftLayer, AWS, Microsoft Azure, etc.
- **Software as a Service:** Applications that are consumed over the internet and are typically not customizable and developed
 - Examples: TurboTax, Salesforce.com, Gmail, IBM Silverpop, IBM Cognos on Cloud
- **Platform as a Service (inclusive of Database as a Service):**
 - Examples: IBM CDS Offerings, IBM Bluemix, AWS database offerings, etc.



Why the Journey to Cloud-based Analytics?

MISSION

To provide the **best experience for developers and enterprises** with a **comprehensive set of rich, integrated cloud data services** covering content, data and analytics.

FASTER INNOVATION

Instant provisioning saves weeks of data center setup

BETTER IT ECONOMICS

Pay as you go with no big up-front capital investments

LOWER RISK OF FAILURE

Fully managed 24x7 so you can focus on new development



IBM Cloud Data Services



Cloudant	dashDB	BigInsights on Cloud	Spark as a Service	DB2 on Cloud
NoSQL DBaaS	Analytic Data Warehouse	Hadoop in the Cloud	Fully-managed Spark Service	Hosted Database in the Cloud
<ul style="list-style-type: none"> Global data distribution Massively scalable Eventually consistent data model Built for mobile, Systems of Engagement 	<ul style="list-style-type: none"> SQL interface Massively parallel ACID compliance Columnar, in-memory performance BLU augmented with NZ in-DB analytics Built for Systems of Insight 	<ul style="list-style-type: none"> Bare metal performance Build on reference architecture BigInsights enterprise features 	<ul style="list-style-type: none"> Optimized for extremely fast and large scale data processing Spark SQL, Streaming, MLlib, GraphX Build and run apps benefiting from operational, maintenance and hardware excellence 	<ul style="list-style-type: none"> Power of DB2 Fast Provisioning Flexible pricing No loss of DBA control Built for Systems of Record



MYTH #1:

**“IBM isn’t a cloud
company”**



Some Facts

- **IBM's overall cloud business is running at an \$8.7 billion annual run rate**
 - Growing 75% per year
 - Projected to be \$40 billion by 2018 and representing 44% of the corporation's revenue
 - Revenue and business from IaaS, PaaS, SaaS offerings
- **Strong partnerships, investments, and acquisition bring value to our clients**
 - Acquisitions of SoftLayer, Cloudant, Compose, BlueBox
 - Partnerships with Twitter, Facebook, Box, the Weather Company
- **IBM's leading IP leadership, investment in open source, integrated experience, and flexibility of deployment makes the IBM Cloud and offerings best suited for enterprises and developers alike**

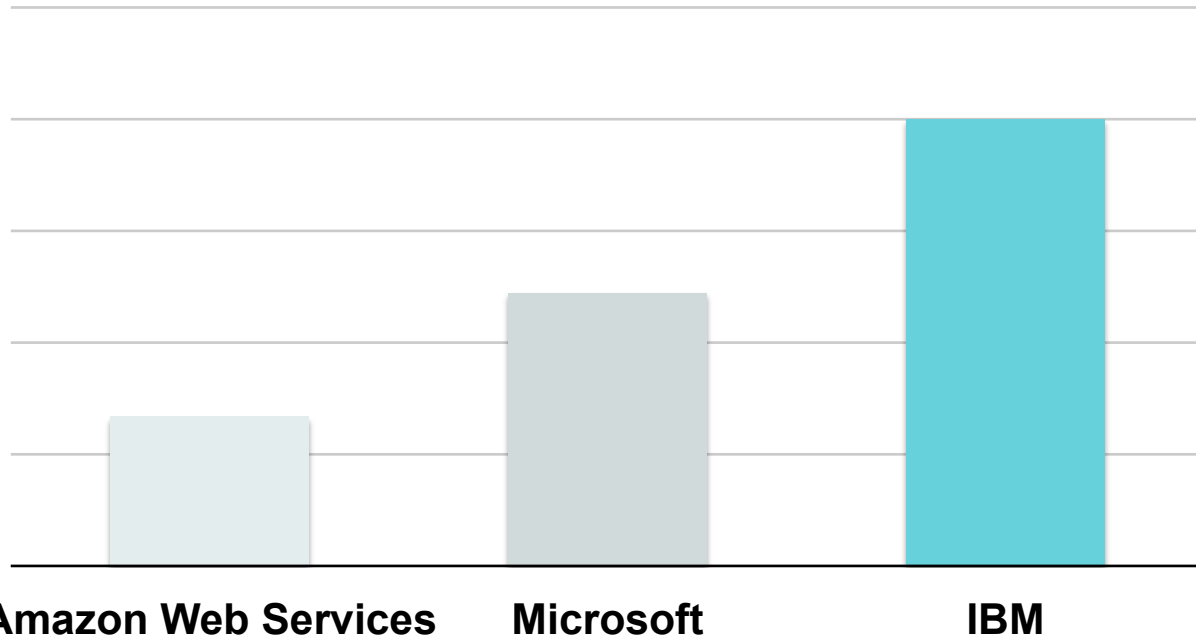


40,000+ cloud consultants and experts

- **Choice of deployment; public, private and hybrid**
- **Modular services to fit your need**
- **Envision, build and deploy, manage and transform with expert services**



IBM Voted #1 Cloud Provider by Enterprise Users



Based on an April 2015 451 Research survey of more than 2,000 enterprise cloud users - commissioned by Microsoft



The IBM Cloud

- “Bare-metal” outperforms virtualized
- Dedicated hardware
- 40 data centers worldwide

SOFTLAYER®
an IBM Company

Data Centers in Sydney & Melbourne





IBM Bluemix



Bluemix is an open-standard, cloud-based platform for building, managing, and running applications of all types (web, mobile, big data, new smart devices, and so on).



Go Live in Seconds

The developer can choose any language runtime or bring their own. Zero to production in one command.

Layered Security

IBM secures the platform and infrastructure and provides you with the tools to secure your apps.

On-Prem Integration

Build hybrid environments. Connect to on-premise assets plus other public and private clouds.

DevOps

Development, monitoring, deployment, and logging tools allow the developer to run the entire application.

Flexible Pricing

Sign up in minutes. Pay as you go and subscription models offer choice and flexibility.

APIs and Services

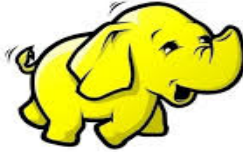
A catalog of IBM, third party, and open source API services allow the developer to stitch an application together in minutes.



IBM IP Leadership + Open Source



hadoop





A complete and growing Portfolio



Finance

- Cognos Controller
- Cognos Disclosure Management
- Concert
- Cognos TM1



Sales

- Incentive Compensation Management
- Territory Management
- Quota Management



Horizontals

- SPSS Modeler Gold
- Watson Analytics
- Watson Curator
- Business Intelligence
- Case Manager
- Content Manager on Demand
- Internet of Things Foundation
- Insight as a Service



Marketing

- Social Media Analytics
- SPSS Data Collection



Risk

- Risk Content & Data Services
- Algo Risk Content
- Algo Risk Service
- OpenPages GRC



Operations

- Maximo Asset Management
- Maximo Inventory Insights
- Intelligent Operations Center
- Intelligent Transportation
- Intelligent Water
- Intelligent City Planning & Operations
- Insights Foundation for Energy
- Facilities & Real Estate Management
- COPLINK
- Counter Fraud Management
- Decision Optimization



Platform / Development

- Cloudant
- Informix
- dashDB
- BigInsights
- DataWorks
- DB2 on Cloud
- SQL Database
- Analytics for Apache Spark
- Content Fabric



Engineering

- Managed Continuous Engineering
- Rational DOORS Next Generation
- Continuous Engineering
- Internet of Things Workbench



IBM
Data



BUSTED



MYTH #2:

“My company isn’t ready for the cloud”











IBM Cloud Data Services: Worldwide Success Stories



A **Global global BPO and IT services company operating in India and the US** is moving away from an on-premises Oracle & Informatica data warehouse environment to an IBM Cloud solution with **dashDB** and Cognos on Cloud.



A **US Insurance company** is using our **BigInsights on Cloud** platform to analyze internal emails for employee compliance to prevent risk & audit exposures.



A **large US retailer** is using **Cloudant** as a their scalable and flexible data management platform to support their mobile wallet offering that manages promotions, coupons, and customer loyalty for millions of users. From purchase to production in 8 weeks.



A **Spanish taxi company** required a database solution that scales with a growing user base and new territories for their corporate taxi application. **Cloudant** makes the application more scalable and available to users wherever they are.



BUSTED



MYTH #3:

**“Cloud offerings aren’t
secure enough for my
sensitive data”**



Unparalleled security

Don't avoid the security conversation, **START IT**

- **6,000+** IBM security experts worldwide
- **3,000+** IBM security patents
- **4,000+** IBM security clients worldwide
- **70+** new products/enhancements
- **27** leadership positions in analyst rankings
- **25** IBM Security labs worldwide





Use Case #1

A US \$30 billion insurance company is embracing a hybrid enterprise Hadoop strategy. They are transitioning some of their applications to run on IBM's BigInsights on Cloud offering so they can stay focused on the insight and results, not managing the environment.

Leveraging the Bluemix Platform and BigInsights on Cloud, this insurer is able to move this pricing application to cloud – using cloud runtimes for the application and the only Hadoop managed service with ANSI compliant SQL – all from IBM



Let's Look At One Offering's Security Features

▪ Security Features:

- *Encryption at rest*: Automatic with Advanced Encryption Standard (AES) in Cipher-Block Chaining (CBC) mode with a 256 bits key.
- *Encryption in transit*: Secure Socket Layer (SSL) is automatically configured when your dashDB database is provisioned. The dashDB console itself is automatically deployed with HTTPS so all your exchanges with the console are also protected with SSL.
- *Database activity monitoring* with Guardium to understand what sensitive data may be in your database and a connections report to see who is accessing it
- *Database access controls* including table level privileges and role based access control
- The database server employs a host firewall to protect listening services against port scans and other network security threats

• Security Certifications & Attestations:

- 3Q: US Safe Harbor, ISO 27001k, SOC2 Risk Assessment Report
- 4Q: HIPAA, PCI-DSS
- Early 2016: SOC2-type 2 certification



dashDB





BUSTED



Use Case Examples



**FASTER
INNOVATION**

Use Case #2

A global pharmaceutical company with more than 90 years of innovation and leadership in diabetes care is using IBM's Enterprise Hadoop as a Service (EHaaS) offering

They are using BigInsights on Cloud on sets of electronic medical records (EMR) data to analyze the relevance of a pharmacological treatment of obesity and obtain costs estimates to build an economic model for obesity treatments.

LIFE WITH REAL-TIME ANALYTICS AND REALLY GOOD SOURCES

BETTER IT
ECONOMICS

thered10

Use Case #3

The Red 10 is a data-driven marketing analytics firm in the UK. With dashDB, they are able to provide real-time analytics and updates to give an accurate view to the audiences

With dashDB, they can provide 1) a live view of the UK & Ireland markets, 2) new segmentation based on live contact views, 3) an instant view of all relevant information, and 4) the right message, at the right time, through the right medium.

This enables growth for less and increased conversion across the sales funnel for their clients.



**LOWER RISK
OF FAILURE**

Use Case #4

A US major grocery store retain chain needed a data management solution to support the development and release of a new web app in less than four months.

Cloudant's schema-less architecture, horizontal scalability, and guaranteed performance empowered the customer's developers to launch a new web application to engage their clients.



IBM Cloud-based analytics summary

- **Global Operations**
- **100s of dedicated enterprise clients**
- **50,000+ users**
- **300% year to year growth**
- **300+ Cloud Data Services team members**





Moving Forward....

- **On to Adam Kocoloski to cover Cloudbant and dashDB in more detail**
 - CTO of IBM Cloud Data Services
 - Co-founder of Cloudbant (acquired by IBM)
 - Ph.D. from Massachusetts Institute of Technology

IBM



Cloud
Data
Services.

IBM



IBM Cloudant

Enterprise-grade NoSQL for web and mobile data

Adam Kocoloski

Distinguished Engineer & CTO, Cloud Data Services

IBM Analytics

IBM Cloudant



JSON Document Store for Web & Mobile Apps
Elastically Scalable & Highly Available
Hybrid Cloud (Public | Private | Open Source | Client)
Key-Value, MapReduce, Search, & Geospatial

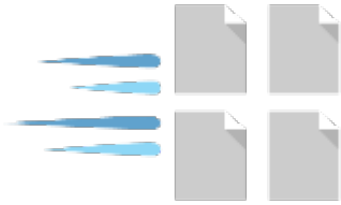
```
{
  "_id": "json",
  "_rev": "3-91769d0354c1c9ef70e07f79295175bf",
  "json": "JavaScript Object Notation",
  "data_types": {
    "number": 1,
    "string": "one",
    "boolean": true,
    "array": [
      "one",
      "two",
      3
    ],
    "object": {
      "key": "value"
    }
  },
  "applications": [
    "Web",
    "Mobile",
    "IoT"
  ],
  "likes": [
    "easy to read",
    "lightweight compared to XML",
    "Web Dev friendly"
  ]
}
```

Document Database

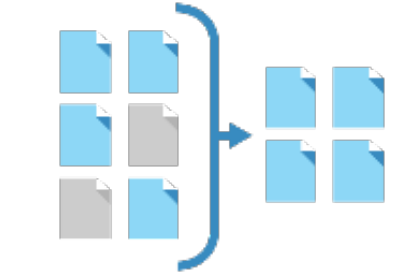
- Documents are stored in the popular [JSON](#) format with a flexible schema
- JSON is the standard data format for modern web, mobile, and IoT applications
- A database is a logical collection of documents, with single set of access permissions
- Cluster can hold any number of databases



Fully Featured Index & Query Options



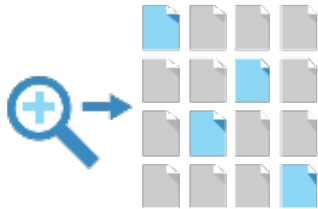
Primary
Doc Lookup



Secondary/Views
with MapReduce



Cloudant Query
Declarative Syntax



Search
Leveraging Lucene

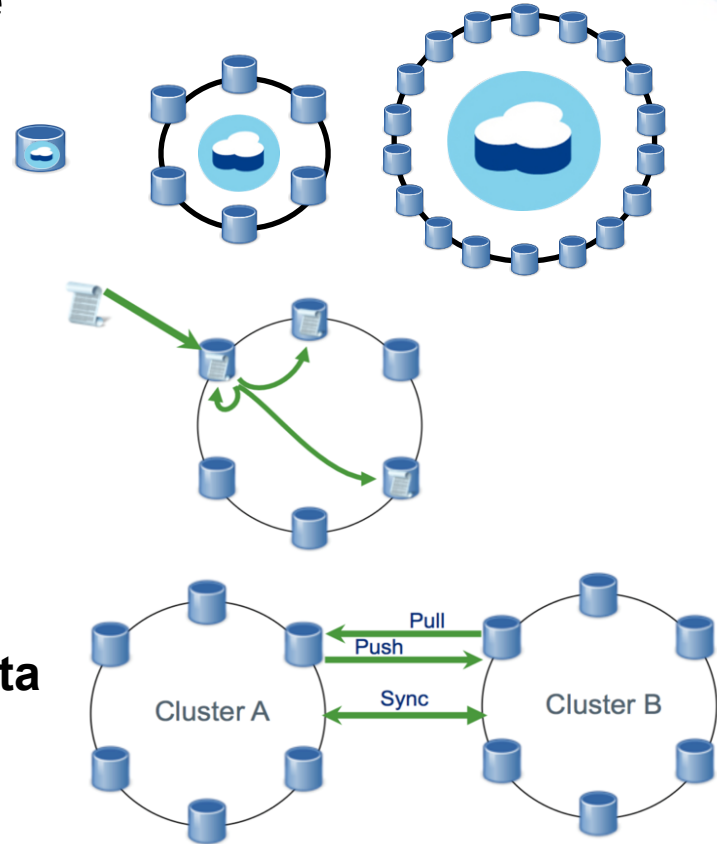


Geospatial
4D, polygon



Scalable, Durable, Available

- Seamlessly add more machines to scale
- Data is auto-sharded across cluster
- All data stored in triplicate
- Designed so no single point of failure for reads or writes
- Industry-leading flexibility in replication
- Geo-load balancing for user access to data closest to them

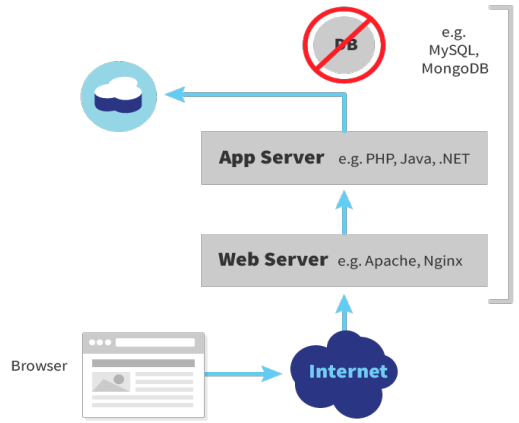
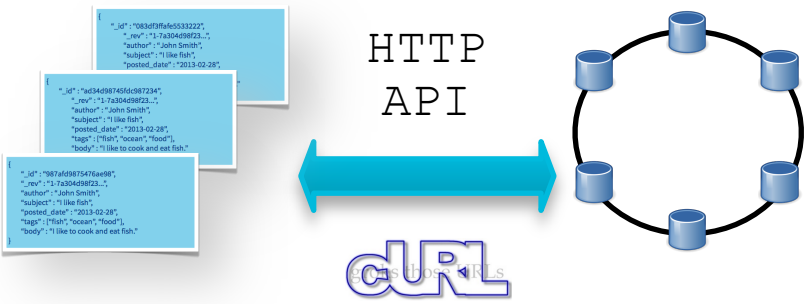






Developing with Cloudant

Databases				
Name	Size	# of Docs	Update Seq	Actions
._replicator	232.5 KB	117	1	[+][+][+]
._warehouse	1.7 KB	1	4	[+][+][+]
agents	76.8 KB	1002	1	[+][+][+]
animaldb	6.0 KB	12	4	[+][+][+]
banking	1.0 KB	6	2	[+][+][+]
bizcard	222 bytes	1	1	[+][+][+]
blog	5.6 KB	19	5	[+][+][+]
blog_answers	6.1 KB	17	4	[+][+][+]



python™

node.js

Microsoft .NET

Java™

php

The Ruby Toolbox

3-Tier Architecture





Cloudant Deployment Options

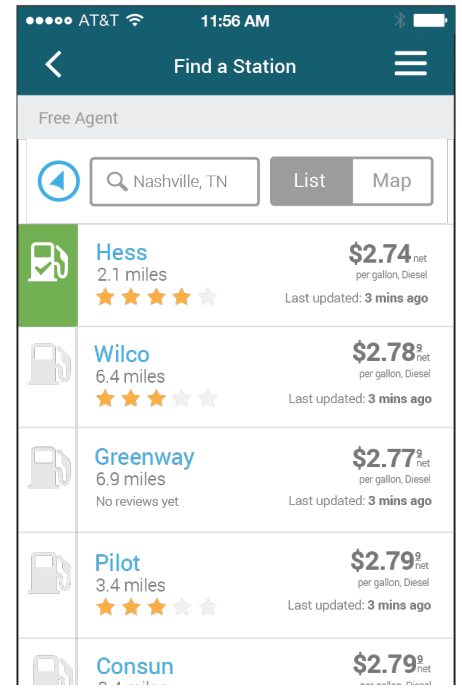
Cloudant Managed Guaranteed SLA with Public Cloud DBaaS		Cloudant Local The Power of DBaaS in the privacy of your data centers
Cloudant Multi-tenant	Cloudant Dedicated	
<i>Hosted & Managed by Cloudant</i>	<i>Hosted & Managed by Cloudant</i>	<i>Customer-hosted & managed with Cloudant DevOps tooling</i>
<i>Community Support</i>	<i>24x7 Premium Support</i>	<i>24x7 Premium Support</i>
<i>Use for development & prototyping</i>	<i>Use for production deployment, development</i>	<i>Use for production deployment, development</i>
<i>Monthly, metered usage fee</i>	<i>Monthly, per-node fee</i>	<i>Up-front perpetual license, or monthly, per-node fee</i>
<i>Available on SoftLayer, Rackspace, Amazon, Azure</i>	<i>Available @ 30+ SoftLayer, AWS, Rackspace, Azure data centers</i>	<i>On-premise or cloud platform of choice</i>
<i>Multi-tenant clusters</i>	<i>Single-tenant clusters</i>	<i>Single-tenant clusters</i>



Why do customers use Cloudant?

- **Scales massively & elastically**
 - Handle millions of daily active users
 - Cluster & Cross-Data Center
- **Managed for you 24x7**
 - Removes risk of project & SLA failure
 - Stay focused on core competencies
- **Agile development for web & mobile**
 - No rigid schemas to slow development
 - JSON, HTTP, and Sync
 - Analytics, Search, Geospatial
- **Many case studies on cloudant.com**





The combination of Cloudant's **advanced geospatial capabilities**, **security**, and **managed service** give Comdata a competitive advantage in terms of the experience they can deliver to their end users



“The Cloudant team were extremely friendly, highly professional and very solutions-minded.”

—Joe Bondi,
CTO and Co-founder, FitnessKeeper



Explosive growth (>20M daily active users) resulted in problems scaling their relational database. Cloudant allows their team to **focus on user experience** and front-end development.

IBM



Cloud
Data
Services.

IBM



IBM dashDB

Cloud data warehousing for the next generation of Builders

Adam Kocoloski

Distinguished Engineer & CTO, Cloud Data Services

IBM Analytics



IBM dashDB

Fast, fully managed, cloud data warehouse service that leverages integrated analytics to deliver answers as fast as you can think. dashDB's unique built-in analytics, R predictive modeling and business intelligence tools free you to analyze your data to get precise insights, quicker.

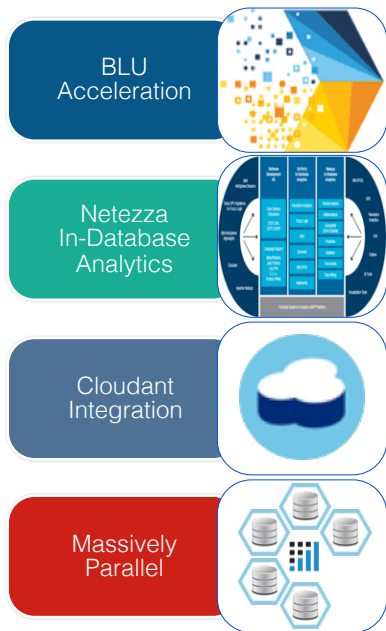
- Built-in performance with in-memory technology
- Predictive modeling built into the database (linear regression, k-means clustering, Esri compatible, et al.)
- Works with an ecosystem of apps and tools
- Integrated security and maintenance





dashDB – Analytics Warehouse as a Service

In-database analytics capabilities for best performance atop a fully-managed warehouse



- Fully-managed data warehouse on cloud
- **DB2 BLU** columnar technology + **Netezza** in-database analytics
 - BLU in-memory processing, data skipping, actionable compression, parallel vector processing, , “Load & Go” administration
 - Netezza predictive analytic algorithms, fully integrated RStudio & R language
- Oracle compatibility
- Massively Parallel Processing (**MPP**)



BLU's Breakthrough Technologies

- **Dynamic In-Memory**

In-memory columnar processing with dynamic movement of data from storage data



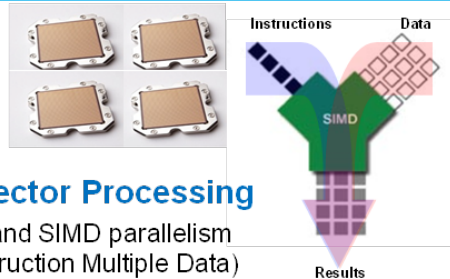
- **Actionable Compression**

Patented compression technique that preserves order so that the data can be used without decompressing



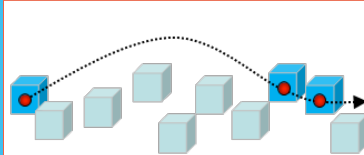
- **Parallel Vector Processing**

Multi-core and SIMD parallelism (Single Instruction Multiple Data)



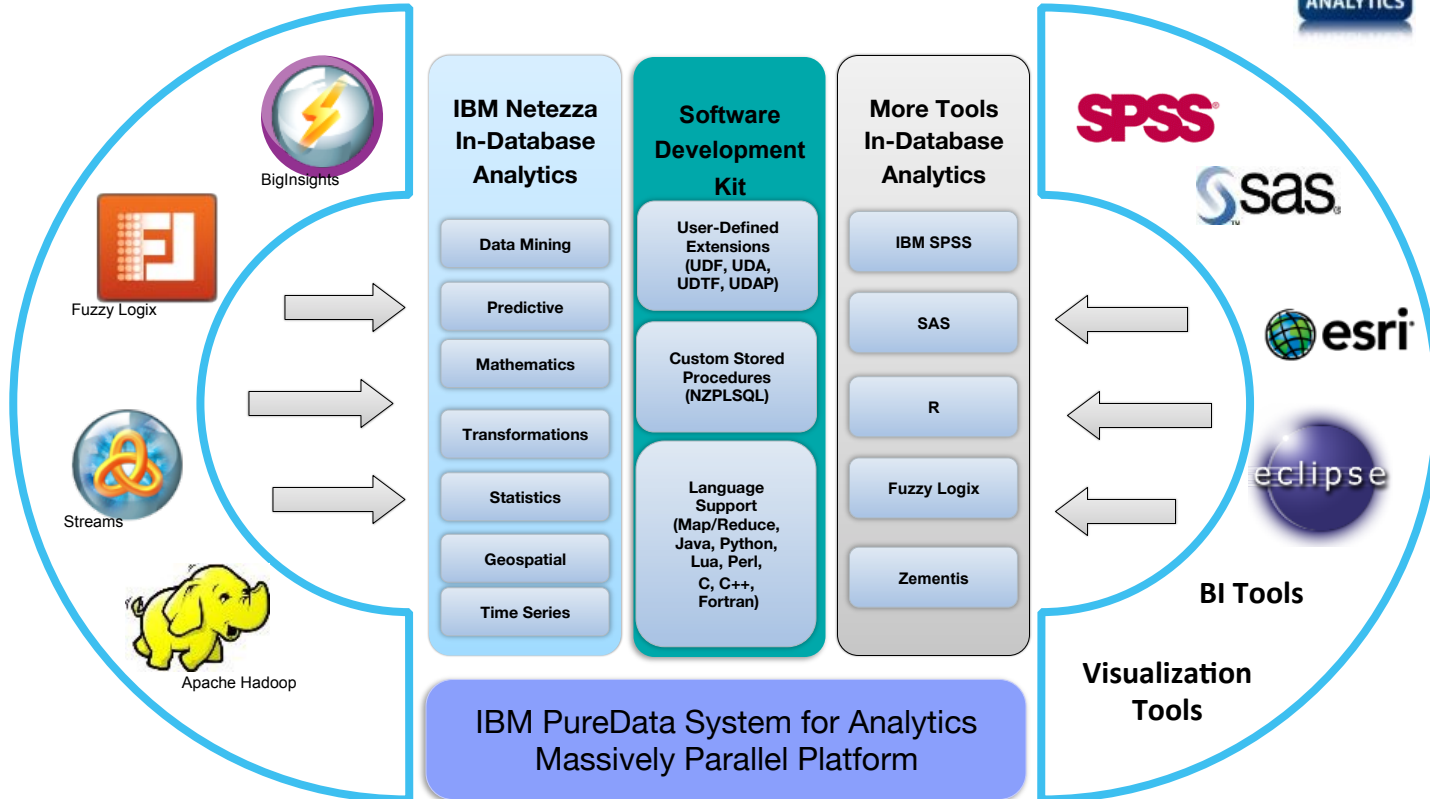
- **Data Skipping**

Skips unnecessary processing of irrelevant data





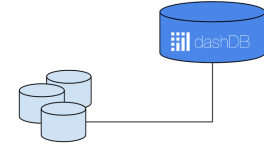
IBM Netezza Analytics at a Glance





dashDB Key Use Cases

Extend / Modernize on premises Environments



Insights from JSON Data



In-Database Analytics



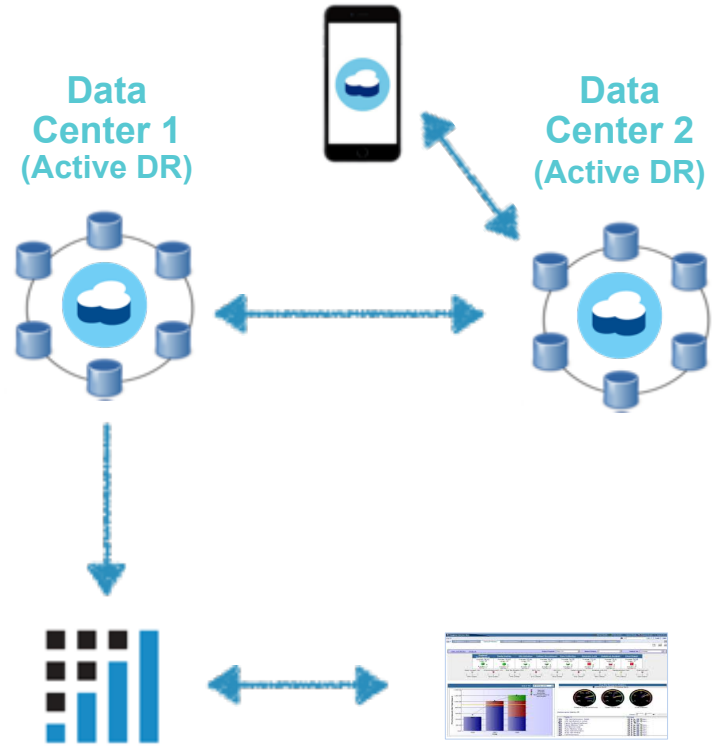
Data Warehouse & Analytics Service





Cloud Analytics for Web & Mobile

- Application powered by Cloudbant supports millions of users
- New user events are pulled incrementally and continuously into dashDB
- Deliver reports to your choice of BI tool
- Analyze data to better understand user experience and increase monetization





dashDB Client Success





Customer: Avnet Services

dashDB is an amazing fit for our business, and I can't wait to show our dashDB-based solution to customers who are used to running on a competing database.

*Here's the report that used to run in 30 minutes, and here's the one that runs in 30 seconds. **I'm not going to have to sell this. It's going to sell itself.***

- Ted Johnson, Technical Director & Senior Architect, Avnet Services





Key Takeaways

- **dashDB's is IBM's next generation data warehouse**
- **Enterprise-class performance and “load and go” simplicity deliver faster time-to-value**
- **Innovative analytics on JSON data enable insights from Systems of Engagement**
- **Rich Oracle and Netezza compatibility ensures viability of existing analytics**

IBM



Cloud
Data
Services.

IBM



IBM Cloud Data Services

BigInsights on Cloud

Chris Howard
Chief Analytics Architect, IBM Analytics, Asia Pacific



Agenda

- **Evolution of Big Data Analytics space**
- **Open Data Platform and BigInsights**
- **Hadoop as a service – BigInsights on Cloud**



What is changing in the realm of big data & analytics?

Data has emerging as the world's newest resource for competitive advantage

Over 2 billion people (25% of the world's population) are online

Gartner

Decision-making is moving from the elite few to the empowered many

Every driver generates 900 rows of data per 15 minute commute

Directline Insurance 2013

As the value of data continues to grow – current systems can't keep pace

Global data center traffic will grow at an annual rate of 25% reaching 7.7 zettabytes by end of 2017

Cisco Global Cloud Index: 2012 - 2017

Hadoop has become the way to store massive volumes of information and perform analytics on a wider set of data



The Hadoop Market is Evolving Rapidly

SQL

Machine Learning

Data access for the developer expanding to data insight for the scientist

Emerging Technology

Board room

Promising technology is now transforming business strategy

Open Source

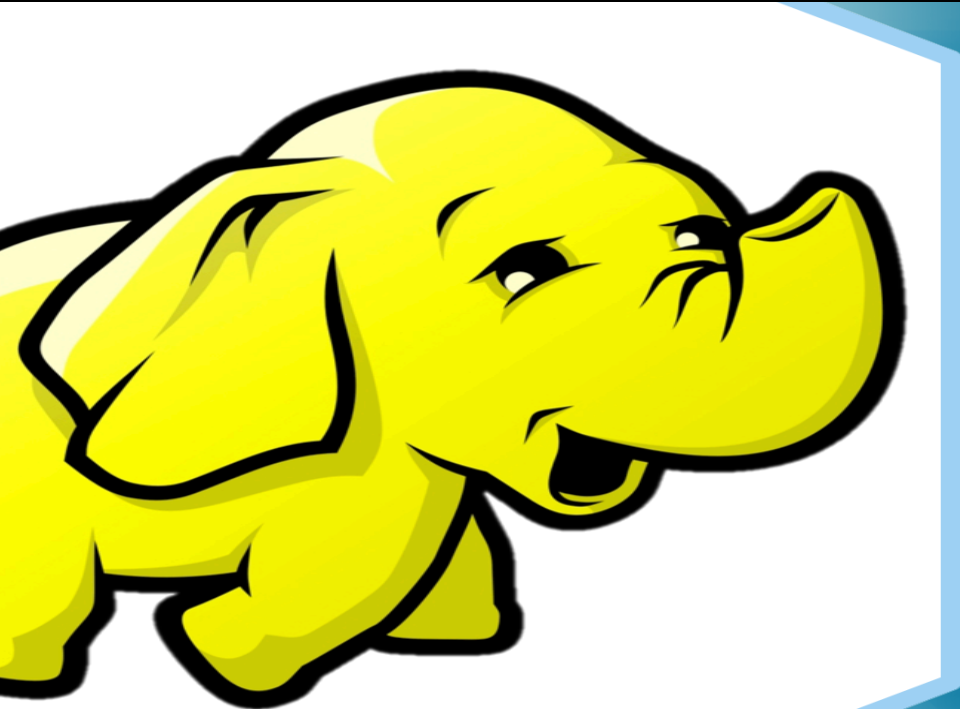
Consistent Platform

Industry shifting to an open and consistent platform to drive innovation for all



Open Data Platform (ODP)

Community-based effort to standardize Apache Hadoop for improved adoption

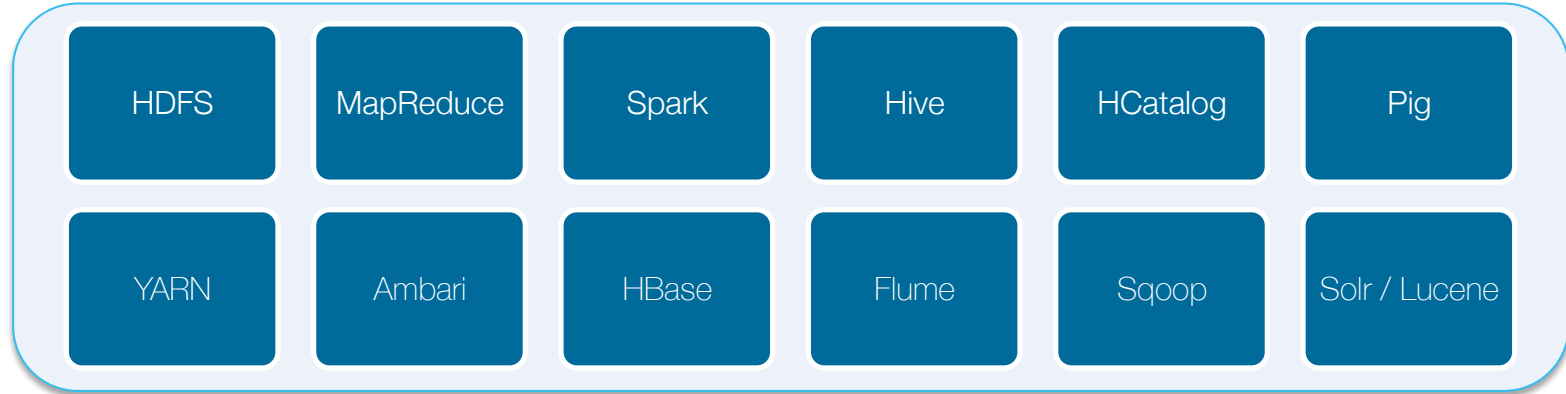


- **Certify** a standard “ODP Core” set of open source Hadoop family projects with specific versions and patch levels
- **Develop** tools and methods to help solution providers to test applications against the ODP Core
- **Contribute** changes and fixes in the ODP Core Hadoop family projects to the ASF using the ASF processes



IBM Open Platform (IOP)

IBM Open Platform (IOP)

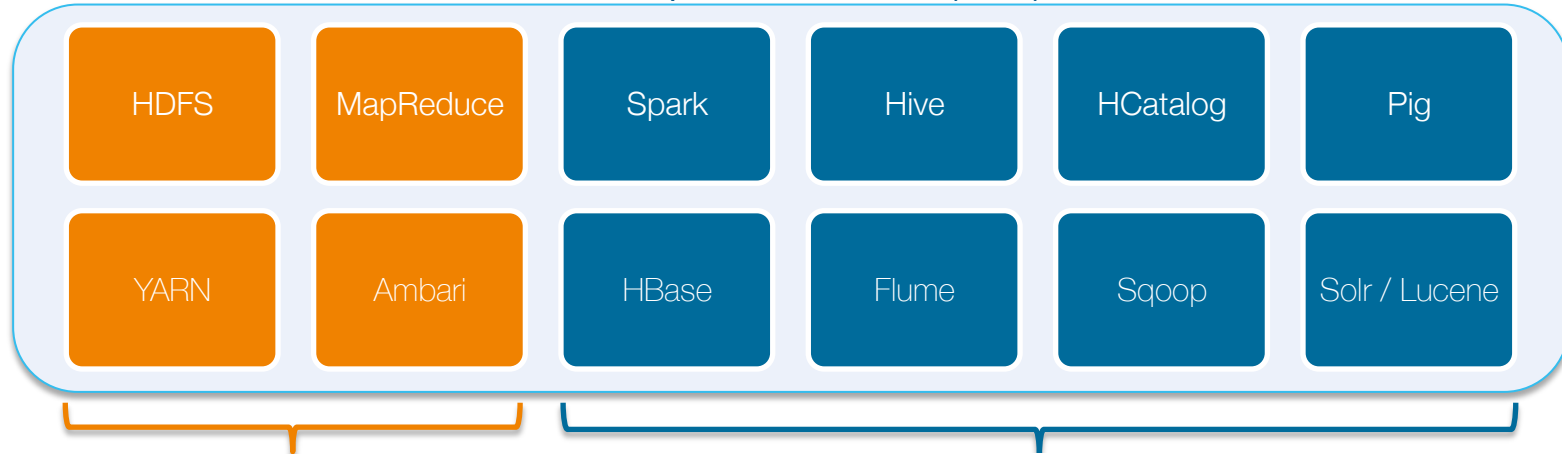


- **IOP package of 100% open source Hadoop distributions from IBM and the Apache Software Foundation**
 - Includes [Apache Spark](#) for in-memory MapReduce processing
 - Includes [Apache Ambari](#) for simplified Hadoop administration
 - **Free for production usage**
 - Support (paid) available for customers who desire it



IBM Open Platform (IOP) Adopts ODP Core Standards

IBM Open Platform (IOP)



Open Data Platform
(ODP)

Hadoop Open-source Components

- **ODP certification & standards will initially target core Hadoop packages, with plans for further coverage of the IOP stack in the future**
 - Enables IBM Hadoop capabilities to run on other ODP-certified Hadoop distributions
 - Better compatibility with minimal testing required against ecosystem software



Usual objections related to Hadoop & Big Data

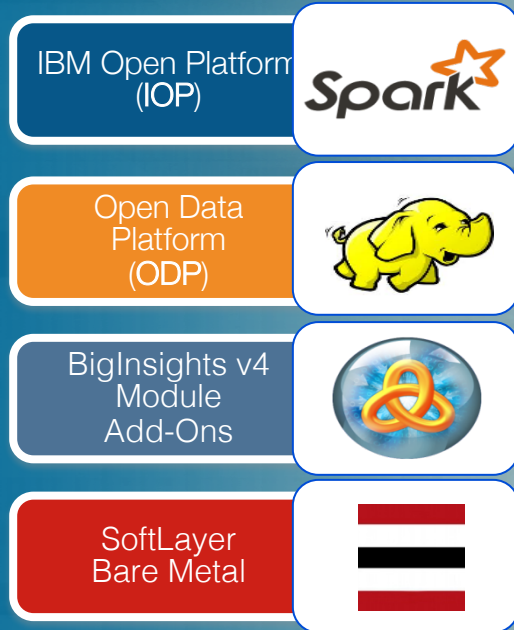
- Limited personnel, skills and data center capacity
- Many big data initiatives are early stage – fast moving ecosystem, requirements uncertain & evolving quickly
- *Spiky* or unknown infrastructure requirements
- Challenges meeting time-to-market expectations





IBM BigInsights on Cloud

Enterprise Hadoop as a Service (EHaaS)



- Simple IBM Cloud provisioning & scaling
- Performant bare metal deployments
- Managed solution
 - Monitoring for availability & security of critical platform components
 - Patching of OS, Hadoop, and BigInsights
- **IBM Open Platform (IOP) packages + BigInsights v4.0 module add-ons**
 - Latest open source packages (Hadoop 2.6, YARN, Spark, Ambari) available for no charge



IBM BigInsights on Cloud – Scope of Managed Operations

- **Managed operations:**

- Proactive monitoring for availability of critical platform components
- Ongoing patching for high severity fixes, security flaws, and new functionality
- 24 x 7 severity level-one support

Managed by Customers

Applications

Data

Hadoop
Operations

BigInsights
Operations

Managed by IBM

OS Patches

BigInsights
Patches

Operating
System

Security

Servers

Storage

Networking

IBM BigInsights on Cloud – v4.0 Paid Add-on Modules

IBM Open Platform (IOP) with Apache Hadoop

- 100% open source distribution, Open Data Platform (ODP) standards, free for production use



IBM BigInsights on Cloud – v4.0 Paid Add-on Modules

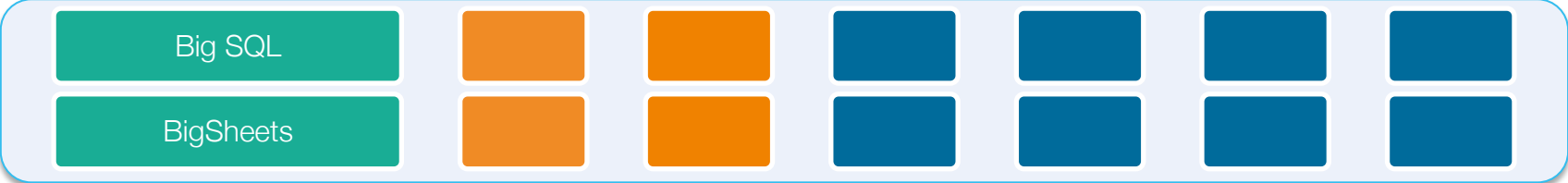
IBM Open Platform (IOP) with Apache Hadoop

- 100% open source distribution, Open Data Platform (ODP) standards, free for production use



BigInsights Analyst Module

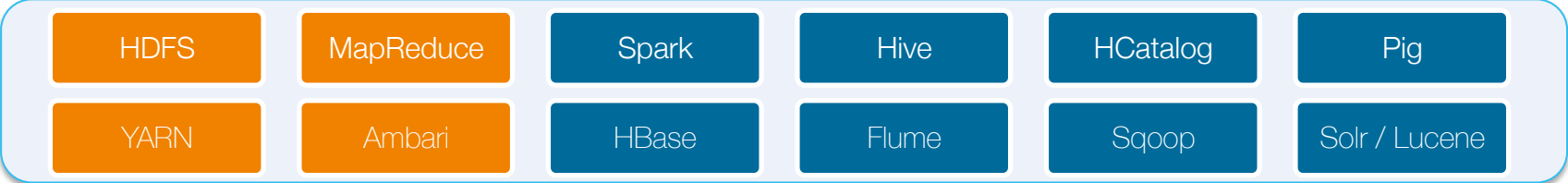
- Includes Big SQL, BigSheets, and the IOP stack



IBM BigInsights on Cloud – v4.0 Paid Add-on Modules

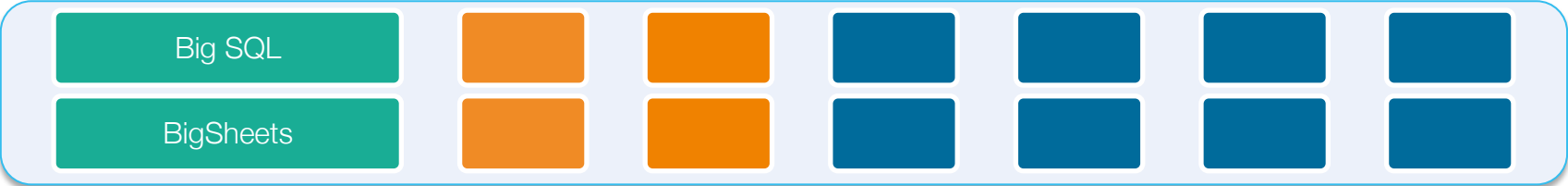
IBM Open Platform (IOP) with Apache Hadoop

- 100% open source distribution, Open Data Platform (ODP) standards, free for production use



BigInsights Analyst Module

- Includes Big SQL, BigSheets, and the IOP stack



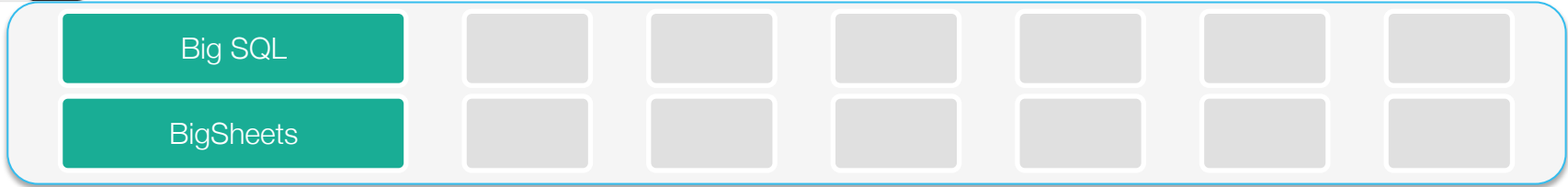
BigInsights Data Scientist Module

- Includes Big R, machine learning & text analytics libraries, as well as Analyst Module and IOP stacks





BigInsights Analyst Module – Detailed



- **Big SQL**
 - ANSI SQL 2011 compliant, built for native Hadoop data sources
 - Executes queries **3.6x** faster than Impala, **5.4x** faster than Hive
 - Supports IBM Cognos, SPSS, MicroStrategy
 - Runs **100%** of TPC-DS (RDBMS benchmark standard) queries at 30TB scale
- **BigSheets**
 - Spreadsheet-style analysis for business users
 - Scalable to massive datasets, multiple data sources
 - Built-in parsing for multiple (structured and semi-structured) data formats
 - Visualize results through spreadsheets, charts, and graphs
 - Entirely driven by graphical user interface (no programming skills required)



BigInsights Data Scientist Module – Detailed



- **Big R**
 - Explore, visualize, and transform BigInsights data using R language syntax
 - Partitioning of large data & parallel cluster execution of push-down R code
 - Connect against BigInsights using RStudio, work with native R environment
- **Text Analytics**
 - Extract information from unstructured data sources for business insight
 - Apply user-defined or pre-built rules for creation & extraction of key data
 - Users do not need to know AQL: driven by graphical user interface (GUI)



IBM Cloud Data Services

Spark as a Service

Chris Howard
Chief Analytics Architect, IBM Analytics, Asia Pacific



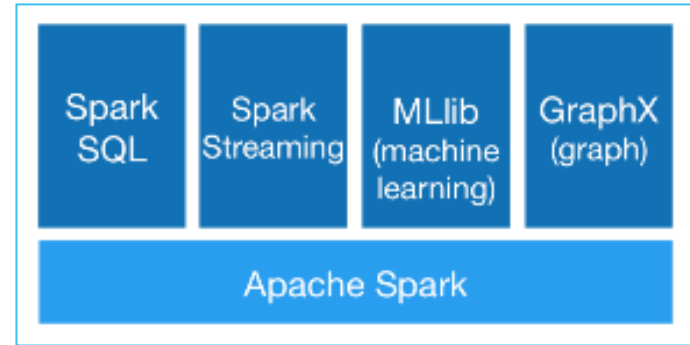
Agenda

- **What is Spark ?**
- **How is Spark Used ?**
- **IBM's Spark as a Service Offering**



Apache Spark

- An Apache Foundation **open source project**. Not a product.
- An **in-memory compute engine** that works with data. Not a data store.
- Enables **highly iterative analysis** on large volumes of data at scale
- **Unified environment for** data scientists, developers and data engineers
- Radically simplifies process of developing **intelligent apps** fueled by data.

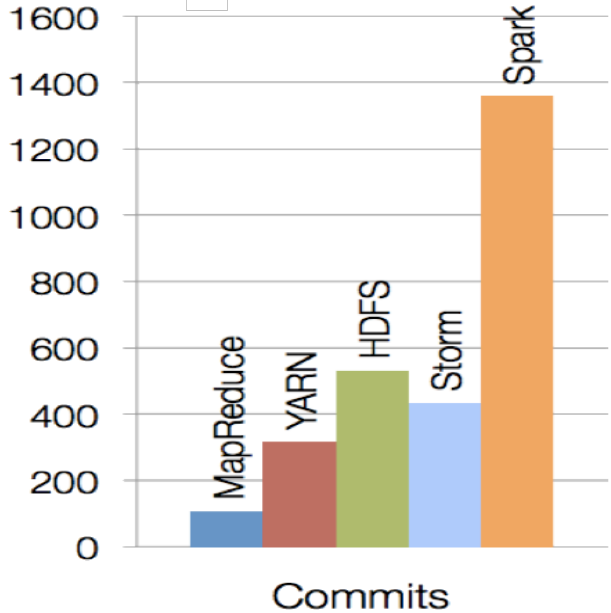


from <http://spark.apache.org>

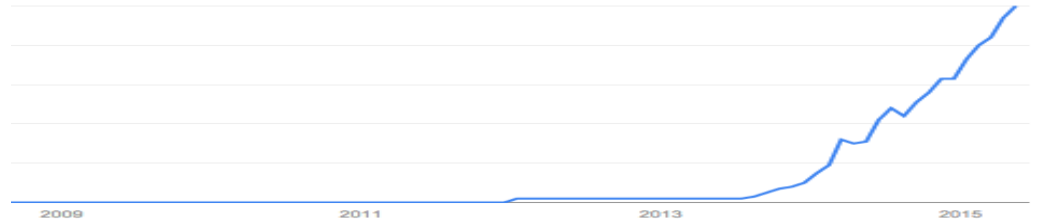


Evolving quickly, High Traction

Spark is one of the most active open source projects



Interest over time (Google Trends)



Job Trends (Indeed.com)



Source: <https://www.google.com/trends/explore?q=apache%20spark&cmpt=q&tz=>
<http://www.indeed.com/jobanalytics/jobtrends?q=apache+spark&l=>



Key reasons for interest in Spark

High Performance



- In-memory architecture greatly reduces disk I/O
- Anywhere from **20-100x faster** for common tasks

Productive



- **Concise and expressive syntax**, especially compared to prior approaches (up to 5x less code)
- **Single programming model** across a range of use cases and steps in data lifecycle
- **Integrated with common programming languages** – Java, Python, Scala
- **New tools** continually reduce skill barrier for access (e.g. SQL for analysts)

Leverages existing investments



- Works well within **existing Hadoop ecosystem**

Improves with age



- **Large and growing community** of contributors continuously improve full analytics stack and extend capabilities



Spark-as-a-Service targets 4 key personas

Data Scientist

- Access powerful tools to tease out the insights they're looking for, then make them actionable immediately

```
square <- function(x)
{
  return(x*x)
}
...
{
  "employees" : {
    "ID" : "1234"
    ...
  }
}
```

App Developer

- Add intelligence to their apps in a simple and straightforward no-hassle manner

Business Analyst

- Answer the questions that the organization needs quickly and easily, and without getting IT involved

```
SELECT
  FirstName, LastName
FROM
  employees
WHERE
  id = 1234;
```

Data Engineer

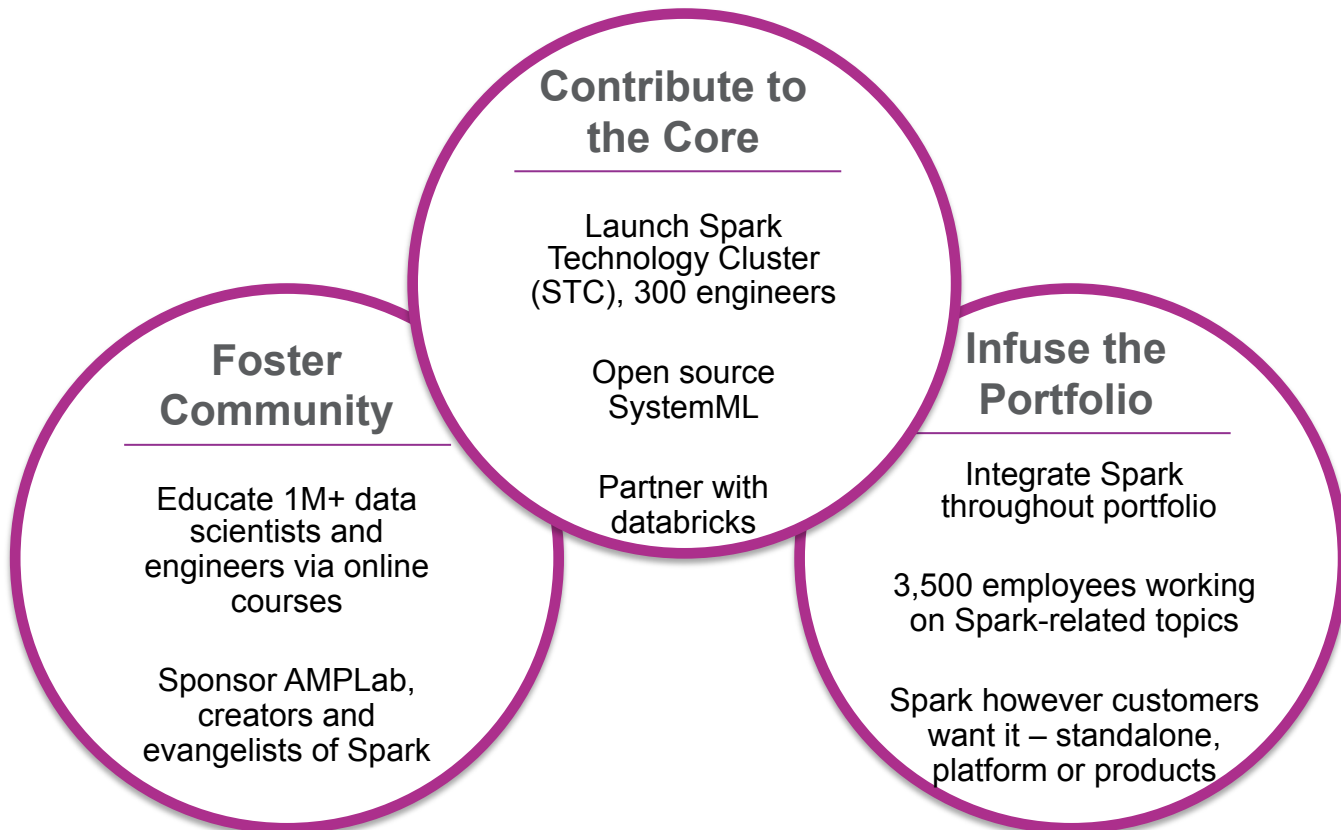
- Easily build data pipelines that power dashboards and data platforms while ensuring high quality

```
# mysqlimport -u root -
ptmpassword --local test
employee.txt
```





IBM is all-in on its commitment to Spark





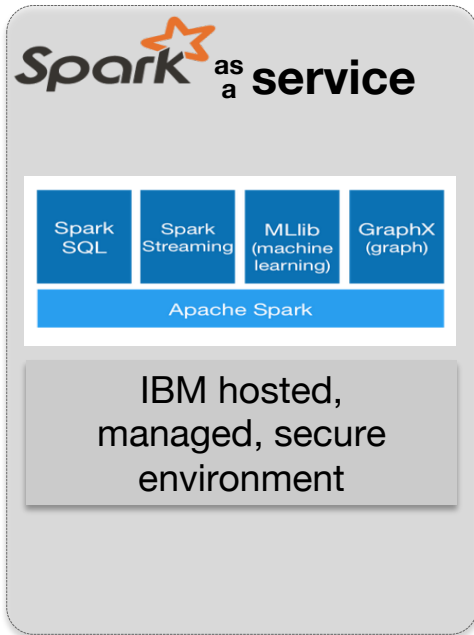
IBM's vision for IBM Analytics for Apache Spark

We make Spark
ACCESSIBLE and **USEFUL**

- Free Trial
- As-A-Service
- Pay as you go
- Managed
- Education
- Datasets
- Notebooks / Tools
- Templates / Boilerplate
- Autoscaling (elastic)
- Connectors



Apache Spark as a Service Offering



Fully-managed Spark environment accessible on-demand

- Access to Spark's next-generation performance and capabilities, including built-in machine learning and other libraries
- Pay only for what you use
- No lock-in – 100% standard Spark runs on any standard distribution
- Elastic scaling – start with experimentation, extend to development and scale to production, all within the same environment
- Quick start – service is immediately ready for analysis, skipping setup hurdles, hassles and time
- Peace of mind – fully managed and secured, no DBAs or other admins necessary



Our first job.....FIND ALIENS!

Leverage Spark capabilities to create a strong analytic foundation to enable new ways to analyze and collaborate on telescope data, initially for professional researchers but eventually to engage all "citizen scientists"



SETI INSTITUTE

Data Sources



Transactional



Social



Application



User Generated



Journal



Video and Audio



Machine / Sensor



Documents



Third Party

Cloud Data Services

Interactive and Real-time Analytics



Spark as a Service

NoSQL Doc Store



Cloudbant Managed DBaaS

Landing, Exploration, Archive



BigInsights EHaaS

Reporting, Analytics



dashDB Managed DWaaS



Hosted Data Services

DB2, Informix, BigInsights HPaaS, MDM, Streams, Sensemaking, Info Server, Guardium, Optim, ECM

Data Movement and Refinement *DataWorks*

On Premises



Real Time Analytics *Streams*

NoSQL Doc Store



Cloudbant Local

Data Warehousing



DB2 with BLU, PDA, PDOA

Deep Analytics, Modeling



PDA

Transactional Systems



DB2, Informix, FileNet

Landing, Exploration, Archive



BigInsights

Reporting, Analytics



DB2 with BLU, PDA

Information Governance *MDM, SenseMaking, Info Server, Guardium, Optim, ECM ILG*

External-Facing Applications



Web or Mobile Systems of Engagement

MobileFirst Platform

Internal Insight



Decision Management

SPSS, ODM



Predictive Analytics

SPSS



Visualization

Cognos, SPSS, Watson Analytics



Reporting

Cognos, Watson Analytics



Discovery Exploration

Watson Explorer



Enterprise Content

Content Foundation, FileNet, Watson Explorer, Case Manager



Getting Started

- **Discover** - Visit [IBM Big Data Hub](#) to read the latest news
- **Learn** - Start with the “Spark Fundamentals” at [Big Data University](#)
- **Try Spark** - Sign up for Apache Spark as a Service on IBM Bluemix at www.spark.tc/beta
- **Try Spark with Hadoop** - Download at IBM.com/Hadoop
- **Engage** - Join the IBM Spark Technology Center at www.spark.tc
- **Converse** - #SparkInsight

IBM



Cloud
Data
Services.

IBM