



IBM Cloud Data Services

Put data to work
with advanced analytics on Cloud



Today's Agenda

Topic	Speaker
An introduction to IBM's Cloud Data Services	Kieran Hagan Analytics Platform Technical Sales Manager IBM A/NZ
Enterprise-grade NoSQL for web and mobile data	Bharath Kadium Sales Lead Cloud Data Services, IBM A/NZ
Cloud data warehousing for the next generation of Builders	
Tackling Big Data with Hadoop service via Cloud	Suraj Pandey Technical Lead. Cloud Data Services, IBM A/NZ.
IBM and Spark for future fast analytics.	

IBM



Cloud
Data
Services.

IBM Cloud Data Services

Kieran Hagan

Analytics Platform Technical Manager A/NZ

IBM Analytics

khagan@au1.ibm.com



First... a disclaimer



CHRIS HOWARD
Chief Analytics Architect

Visual Analysis Results

Geolocation data

Physical data attributes



Kieran Hagan
Technical Sales

Possible match?

Possible match?

Possible match?



Cloud Data Services

- **IBM's growth as a Cloud Company**
- **Why is now the time to consider Cloud?**
- **How is IBM unique in its Cloud offerings?**



What's in a “cloud”? New terms to consider...

■ Infrastructure as a Service (IaaS)

Bring your software licenses to be hosted (hint: you still perform the management and operations of the environment).

- Examples: Hosting your software licenses on IBM SoftLayer, AWS, Microsoft Azure, etc.

■ Platform as a Service (PaaS)

Infrastructure and software provided by a vendor. You control the software deployed and configuration

- Examples: IBM Bluemix, AWS database offerings

■ Software as a Service (SaaS)

Applications that are used over the internet, typically pre-built applications

- Examples: Salesforce.com, Gmail, Watson Analytics, dashDB,



IBM is a cloud company

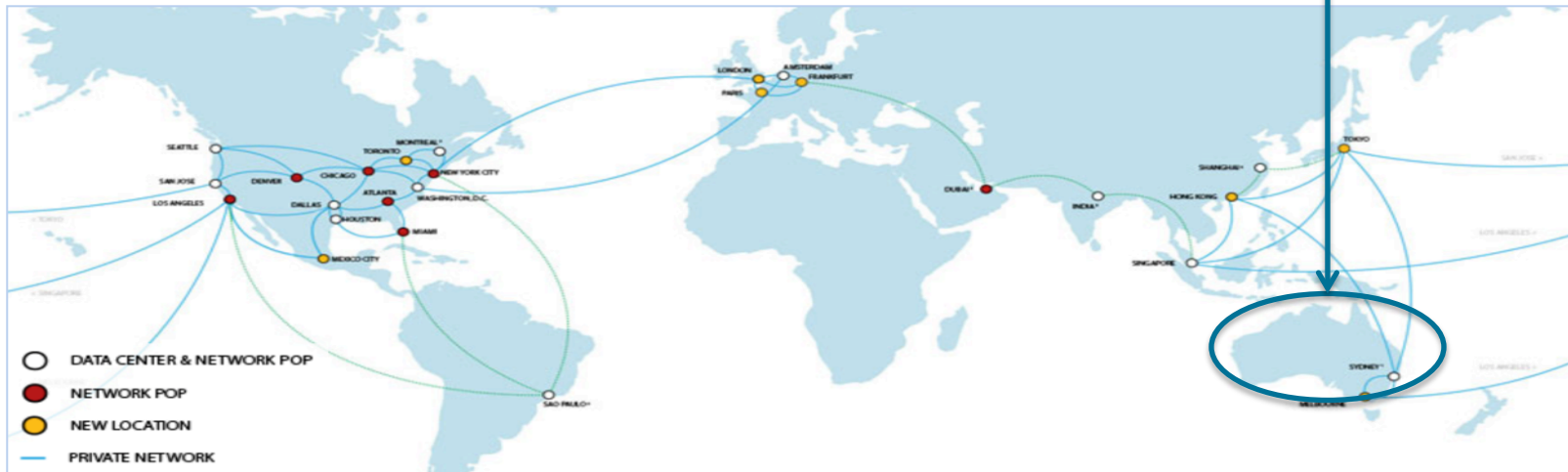


The IBM Cloud

- “Bare-metal” options at launch
- Dedicated hardware
- Over 40 data centers worldwide

SOFTLAYER®
an IBM Company

Data Centers in Sydney & Melbourne





IBM – Managing data in the cloud



DB2 on Cloud	dashDB	BigInsights on Cloud	Spark as a Service	Cloudant
<i>Hosted Database in the Cloud</i>	<i>Analytic Data Warehouse</i>	<i>Hadoop in the Cloud</i>	<i>Fully-managed Spark Service</i>	<i>NoSQL DBaaS</i>



A complete and growing Portfolio



Marketing

- Social Media Analytics
- SPSS Data Collection



Finance

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



Operations

- Maximo Asset Management
- Maximo Inventory Insights
- Intelligent Operations Center
- Intelligent Transportation
- Intelligent Water
- Intelligent City Planning & Operations
- Insights Foundation for Energy



Risk

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



CDS Platform

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



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



Engineering

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



Some Stats

- **IBM's 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 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 IP leadership, investment in open source, integrated experience, and flexibility of deployment makes the IBM Cloud best suited for enterprises and developers alike**



40,000+ cloud consultants and experts

- **Choice of deployment; public, private and hybrid**



Is now the time?







Are these areas of interest?

Implementing in the Cloud through a fully managed service yields the largest business benefits in the following categories

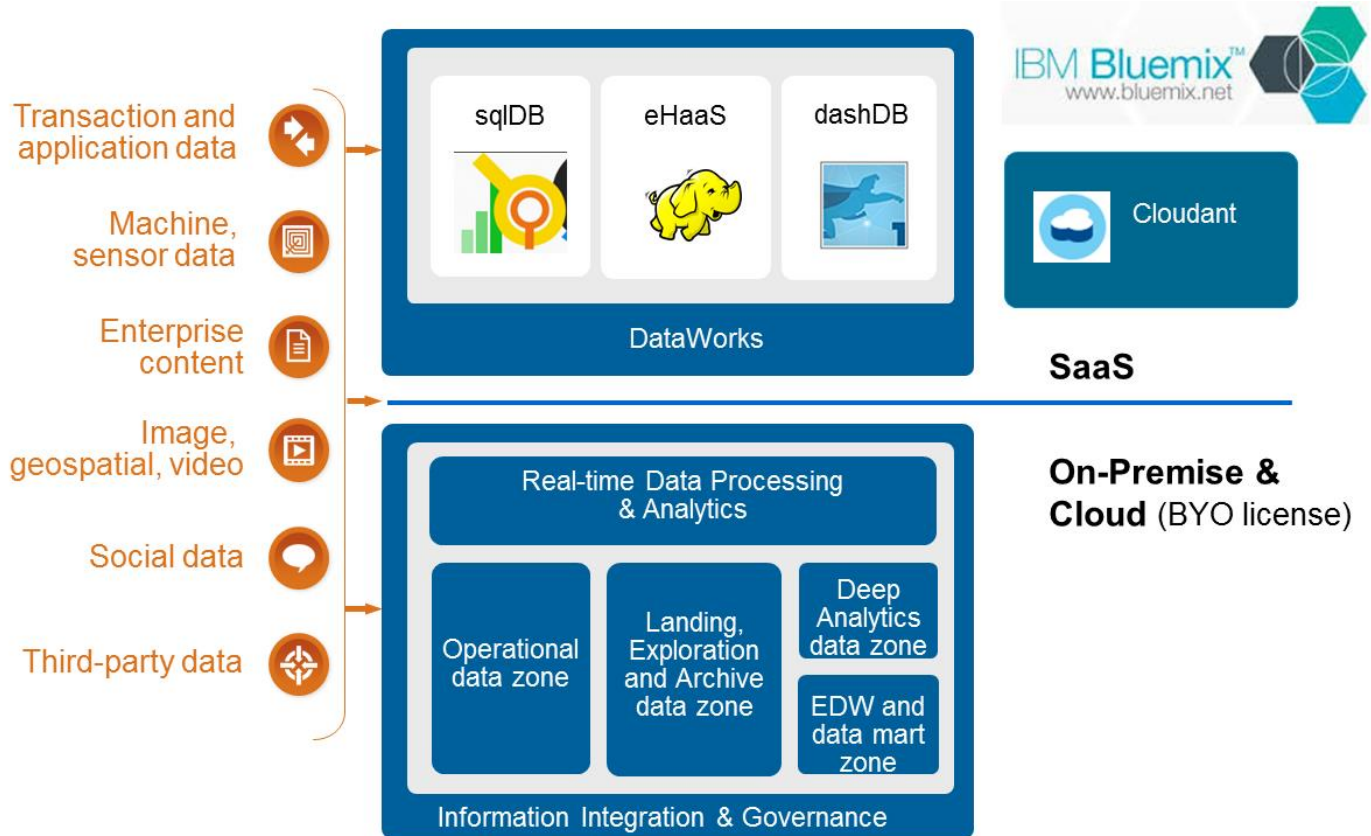
- **Hardware:** Reduced Cost of Hardware: Capex + Opex
- **Software:** Reduced Cost of Software
- **Design:** Reduce Cost to Implement
- **Administration:** Reduce Cost of Administration



Why IBM Cloud?



The IBM Difference – completeness and flexibility





IBM CDS MISSION

Provide the best and most **comprehensive set of rich, integrated cloud data services** covering content, data and analytics for building new systems of engagement or insight.

IBM



Cloud
Data
Services.

IBM



IBM Cloud Data Services

Cloudbant, Dash DB and more..

Data on Premises – Workload Optimised Products

- Offerings for structured and semi-structured data, transactional and analytics workloads.
- Knit together with the InfoSphere IIG technologies for data movement and integration.

ANALYTICAL TRANSACTIONAL

BigInsights

- Enterprise class Hadoop and real-time
- BigSQL for easier analytics
- IBM differentiators like GPFS



Cloudbant Local

- Massively scalable
- Eventual consistency model
- Built for Systems of Engagement



UNSTRUCTURED

STRUCTURED

InfoSphere.
software

DataStage

On-Prem to Cloud
ETL Connector

DB2 BLU

- SQL interface
- ACID compliance
- Columnar, in-memory performance
- DB2 Built for Systems of Insight



DB2

- SQL interface
- ACID compliance
- Flexible HA and DR options
- Built for Systems of Record



IBM Cloud Data Services

Mixed workloads and data types are knit together with **DataWorks** for true hybrid services

ANALYTICAL

TRANSACTIONAL



Enterprise Hadoop

- Bare metal performance
- Build on reference architecture
- BigInsights enterprise features



Cloudbant DBaaS

- Global data distribution
- Massively scalable
- Eventually consistent data model
- Built for mobile, Systems of Engagement

UNSTRUCTURED STRUCTURED

DataWorks

Data Refinery Services



dashDB

- SQL interface
- ACID compliance
- Columnar, in-memory performance
- BLU augmented with Netezza in-DB analytics
- Built for Systems of Insight



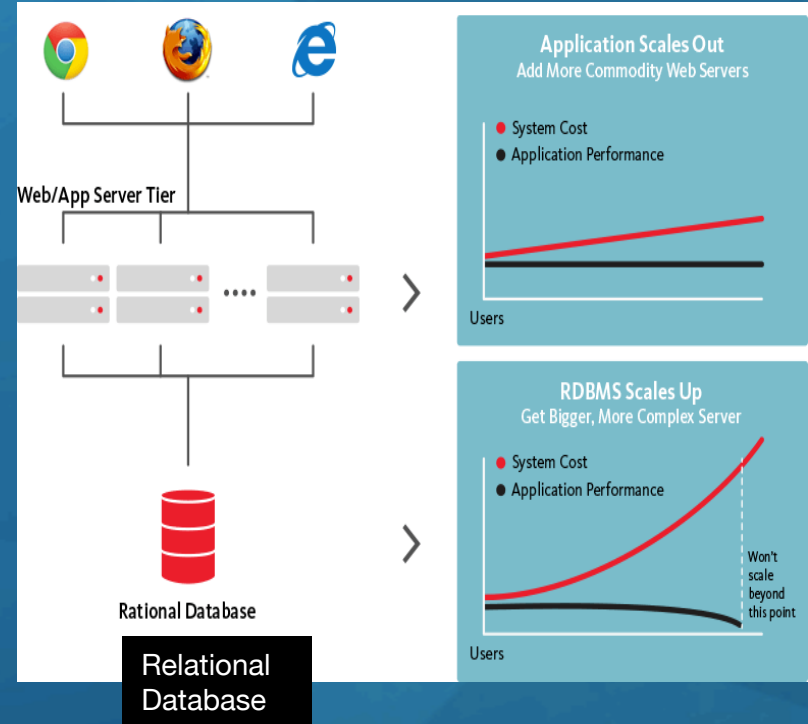
SQLDB

- DB2 for Bluemix
- SQL interface
- ACID compliance
- Same skillset and data formats as on-premises relational databases
- Built for Systems of Record



NoSQL vs Relational DBMS?

Requirement	Cloudant NoSQL	RDBMS
Elastic Scalability	Easy	Hard
Multi-structured data	Easy	Hard
Multi-data center	Easy	Hard
Data mobility	Easy	Hard



Cloudant – NoSQL Database as a Service

Truly hybrid – available as a fully-managed cloud service and an on-premises software solution



IBM Cloudant®

A fully-managed NoSQL database layer that can be **developed & deployed in days**



- Operational NoSQL JSON store
- Master-less architecture for maximum **scalability & availability**
- Advanced APIs
 - REST (HTTP) API
 - Replication & synchronization
 - Geo-load balancing
 - Incremental MapReduce indexes
 - Military-grade Geospatial indexes
 - Lucene full-text search
- Offline access to mobile apps & data



The Cloudant DBaaS Stack

Visualisation

Geo-Load Balancing

Lucene
Search

Chainable
MapReduce

GeoSpatial
Indexing

Management
Monitoring

IOQ

Horizontal Clustering Framework



Apache CouchDB



- Dashboards: Monitoring, Admin, Development
- Connects users to closest copy of data
- Developer APIs
- Prioritizing IO types; prevents “noisy neighbors” in multi-tenancy
- Clustering API, Sharding, Intra-cluster messaging
- JSON Doc Store, HTTP API, GET/PUT docs, Views, Replication...

 = Cloudant built & integrated

Cloudbant Deployment Options

Public Cloud, Private Cloud, or connect them together for Hybrid Cloud

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

Cloudant is used by web and mobile leaders



- 27,000 users
- Billions of daily transactions
- Millions of DBs
- Across 41+ DCs

eCommerce
Online Education
Gaming
Financial Services
Networking
Other Industries



IBM Cloudant®



IBM

SAMSUNG

Scales & remains available to **1 billion** users across Asia, North America, Europe

Transactional Throughput	300 million requests / day (3,500 / second)
Cluster Distribution	Global (mobile devices)
Media Types Ingested	Structured, semi-structured, unstructured (logs, audio)



fully managed services in support of massive concurrent user growth

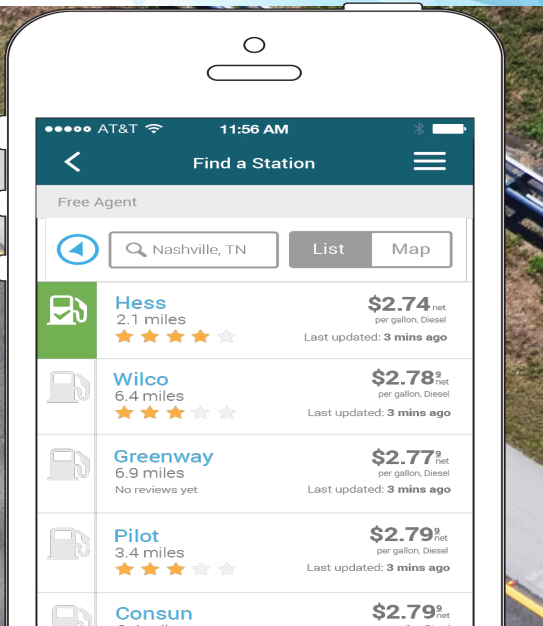
Transactional Throughput	2 billion requests / day (20 : 1 read-writes)
Data Volume	130 TB
Cluster Growth	From 6 to over 200 servers (in 12 months)



Fidelity Investments finds incredible value in database as-a-service

“**Cloudata** allows our architects and senior backend engineers to work on cranking out functionality versus being a slave to operations concerns and database tiers.”





The combination of Cloudata'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



Novartis successfully manages clinical trial data with high availability, stability, and security

“ *Everything is incremental in the **Cloudant** world, so we only have to [execute] operations on the small amount of data that’s changing each day. Processes that were occurring in the realm of ten to twelve hours, we have ”*
down to minutes.

John Walker, IT Director for Novartis



IBM Acquires Compose - Compose DBaaS Offerings



JSON document database.



An open-source, blazingly fast, low maintenance, key/value store, often used for data caching



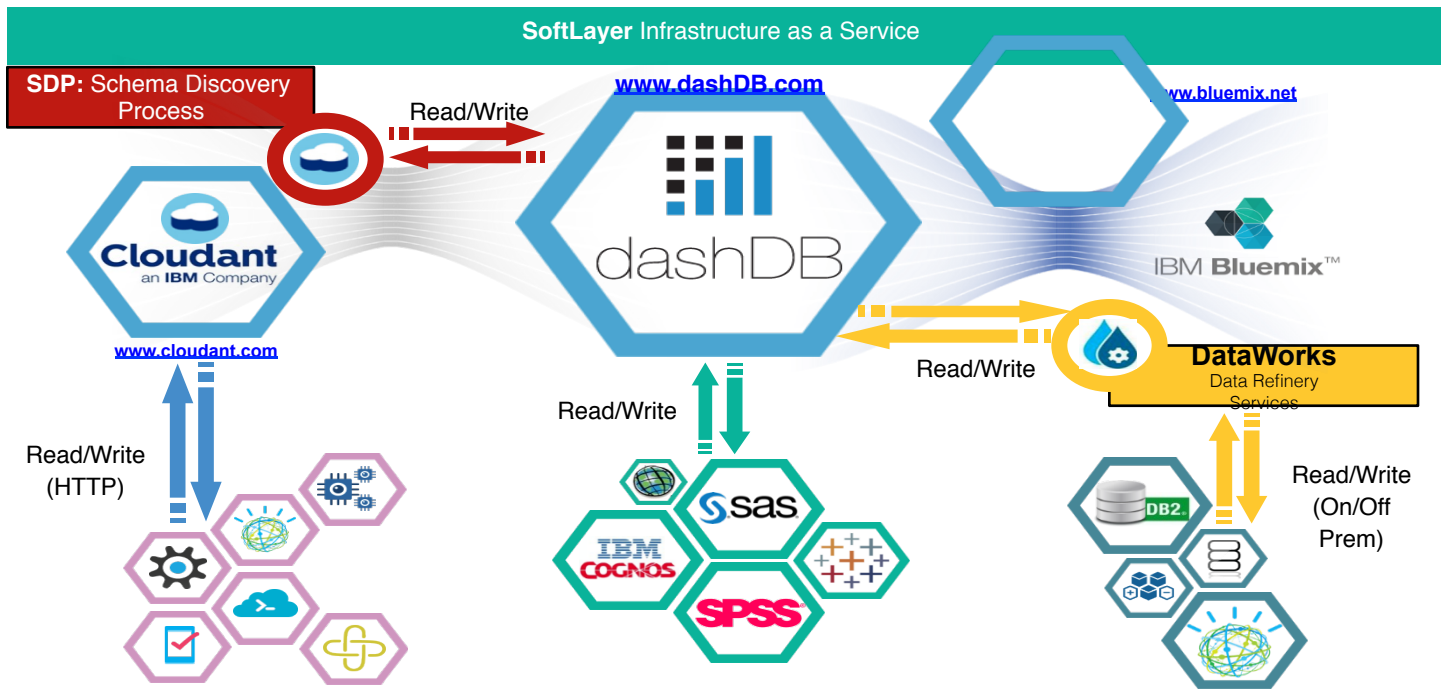
Combines the power of a full text search engine with the indexing strengths of a JSON document database



A powerful, open source object-relational database that is highly customizable.



A JSON document based, distributed database with an integrated administration and exploration console.



Cloud-Based Systems of **Engagement**

(NoSQL, Mobile Apps, Internet of Things, Social Media)

IBM & Third Party **Integrations**

(Cognos, SPSS, SAS, Tableau, ESRI ArcGIS)

Systems of **Record & Insight**

(Watson Analytics, DB2, HDP, flat files)

IBM



Cloud
Data
Services.

dashDB OVERVIEW



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.

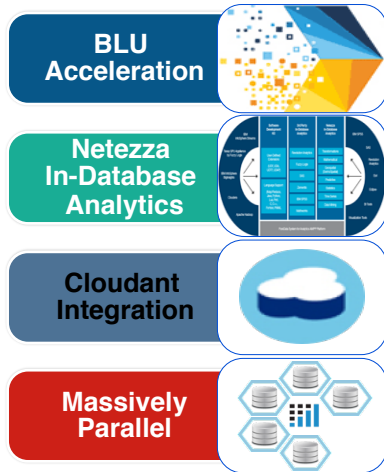
- Built-in performance with in-memory technology
- Predictive modeling built into the database (linear regression, k-means clustering, Esri compatible, et al.)
- Works with an ecosystem of apps and tools
- Integrated security and maintenance





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**)



IBM DB2/PDA



Oracle



Cloudbant



Other relational systems



IoT, Social Media



Datastake



Data cleansing & integration with numerous options (Object stores, ETL, Dataworks)



dashDB data warehousing and analytics services allows you to focus on the business, not the business of data warehousing

Systems of Record & Systems of Engagement as data sources for dashDB



Tableau



SAS



Cognos



esri



IBM SPSS



Aginity



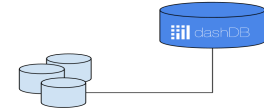
Watson Analytics
**In Roadmap*

IBM & Third Party Integration with BI, Data Visualization, and Analytic Apps

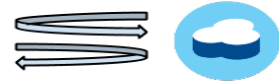


dashDB Key Use Cases

Extend / Modernize on premises Environments



Insights from JSON Data



In-Database Analytics



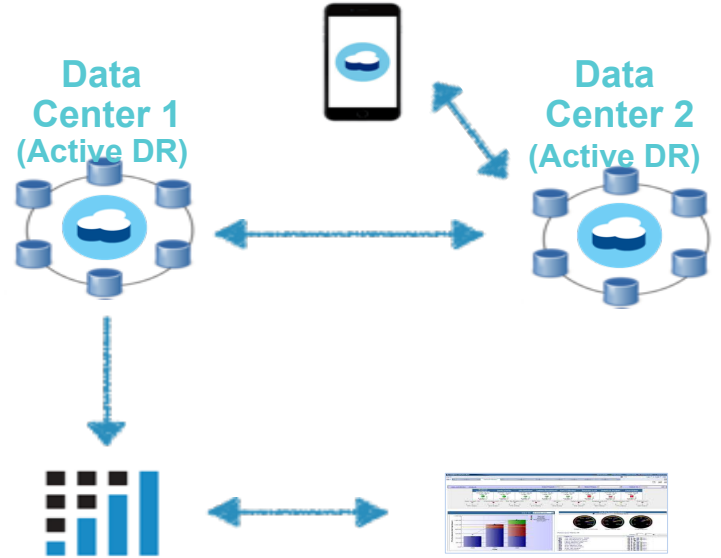
Data Warehouse & Analytics Service





Cloud Analytics for Web & Mobile

- Application powered by Cloudant 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





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 Cloud Data Services

BigInsights on Cloud

Suraj Pandey

Cloud Data Services Technical Lead, Australia and New Zealand



Agenda

- **IBM and Open Source – Hadoop**
- **Hadoop as a service – BigInsights on Cloud**



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

Data is the new Oil

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

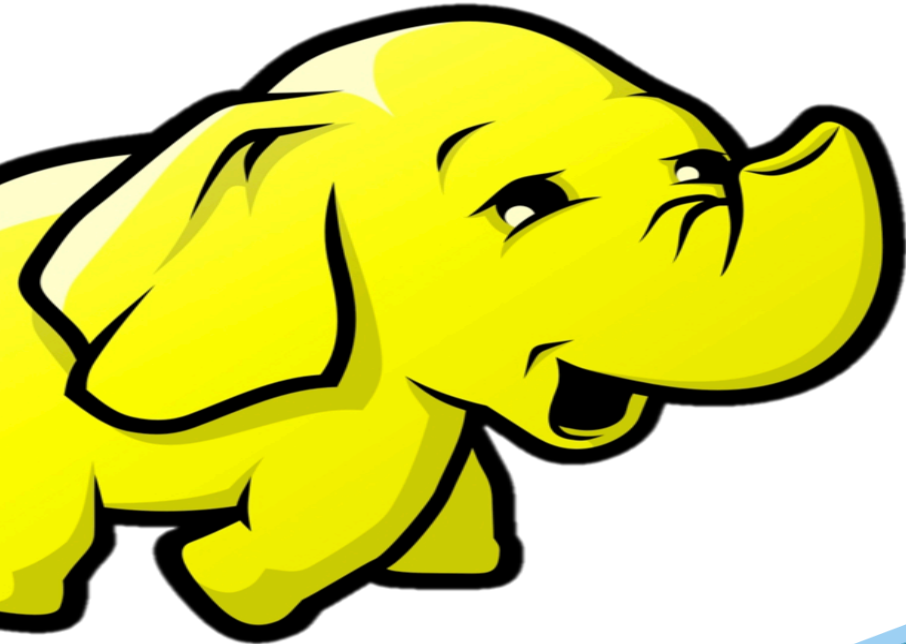
Consistent Platform

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



Open Data Platform Initiative (ODPi)

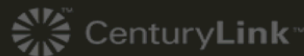
Community-based effort to standardize Apache Hadoop for improved adoption



- **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

ODPi Members

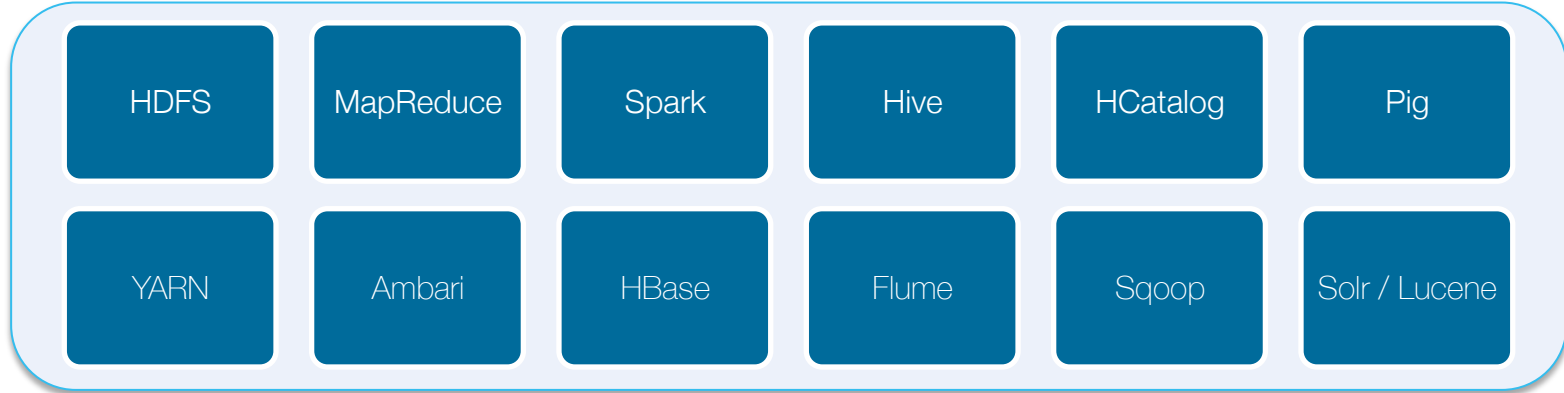
ODPi brings industry leaders together to accelerate the adoption of Apache Hadoop and related Big Data technologies and make it easier to rapidly develop applications. Members to date represent a diverse group of Big Data solution providers and end users including:





IBM Open Platform (IOP)

IBM Open Platform (IOP)

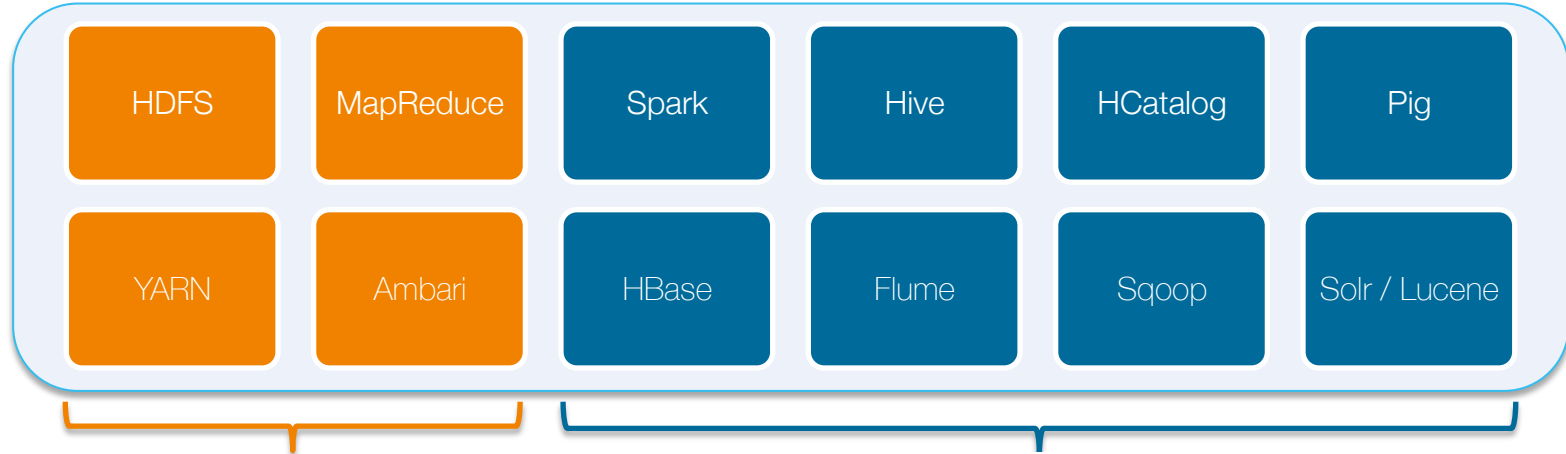


- **IOP package of 100% open source Hadoop distributions from IBM and the Apache Software Foundation**
 - Includes [Apache Spark](#) for in-memory MapReduce processing
 - Includes [Apache Ambari](#) for simplified Hadoop administration
 - **Free for production usage**
 - Support (paid) available for 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 any ODP-certified Hadoop distributions
 - Better compatibility with minimal testing required against ecosystem software



Usual story related to Hadoop & Big Data

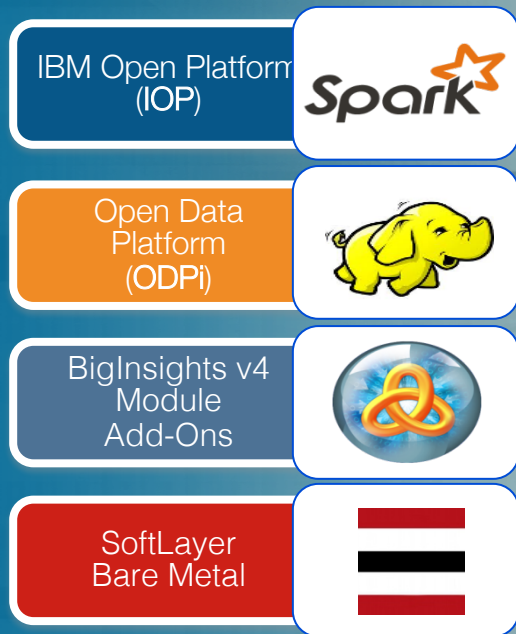
- Limited personnel, skills and data center capacity
- Many big data initiatives are early stage – fast moving ecosystem, requirements uncertain & evolving quickly
- *Spiky* or unknown infrastructure requirements
- Challenges meeting time-to-market expectations





IBM BigInsights on Cloud

Enterprise Hadoop as a Service (EHaaS)



IBM BigInsights on Cloud

- 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 v4.1 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



Managed by IBM



IBM BigInsights on Cloud – v4.1 Paid Add-on Modules

IBM Open Platform (IOP) with Apache Hadoop

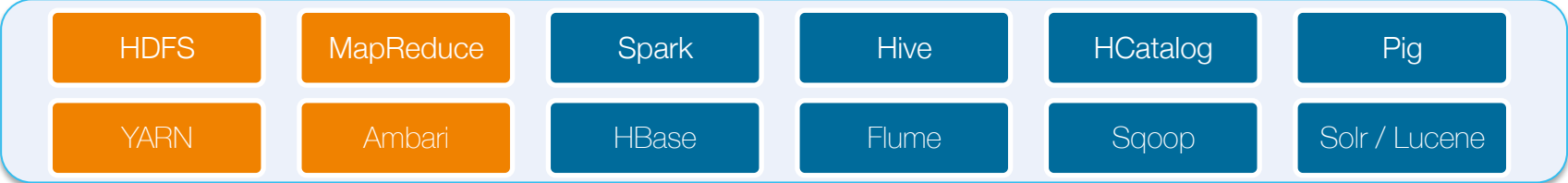
- 100% open source distribution, Open Data Platform Initiative (ODPi) standards, free for production use



IBM BigInsights on Cloud – v4.1 Paid Add-on Modules

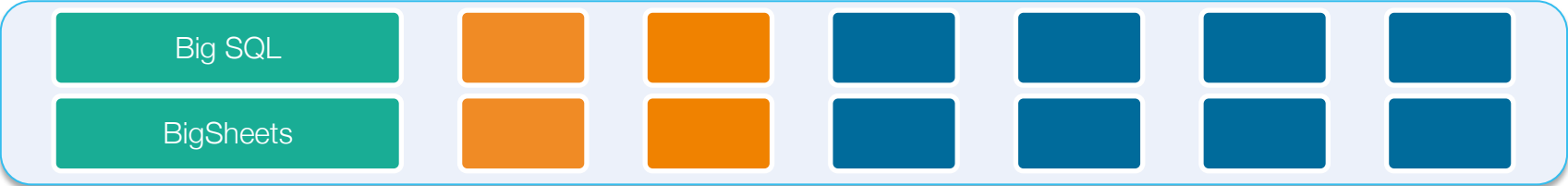
IBM Open Platform (IOP) with Apache Hadoop

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



BigInsights Analyst Module

- Includes Big SQL, BigSheets, and the IOP stack



IBM BigInsights on Cloud – v4.1 Paid Add-on Modules

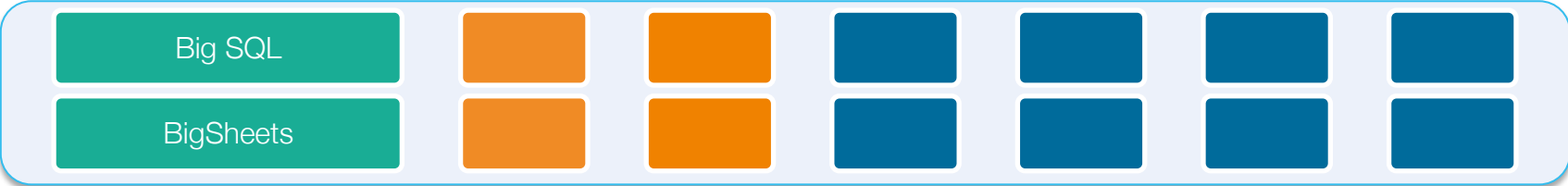
IBM Open Platform (IOP) with Apache Hadoop

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



BigInsights Analyst Module

- Includes Big SQL, BigSheets, and the IOP stack



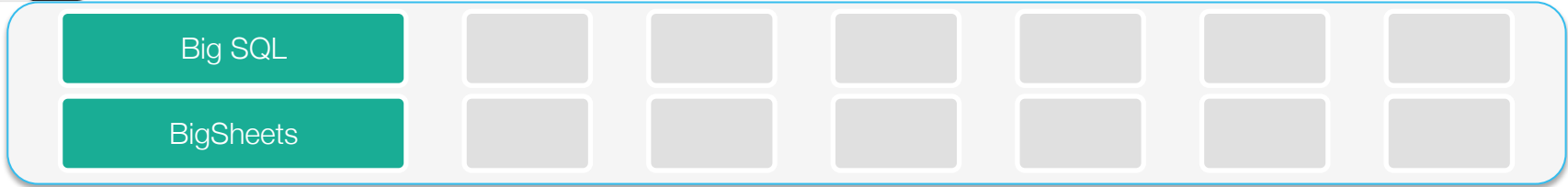
BigInsights Data Scientist Module

- 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)



Hadoop Advantages

Unlimited Scale

- Multiple data sources
- Multiple applications
- Multiple users

- Reliability
- Resiliency
- Security

Enterprise Platform

Wide Range of Data Formats

- Files
- Semi-structured
- Databases





Hadoop MapReduce Challenges

- Need deep Java skills
- Few abstractions available for analysts

Ease of Development



In-Memory Performance



- No in-memory framework
- Application tasks write to disk with each cycle

- Only suitable for batch workloads
- Rigid processing model

Combine Workflows





IBM Cloud Data Services

Spark as a Service

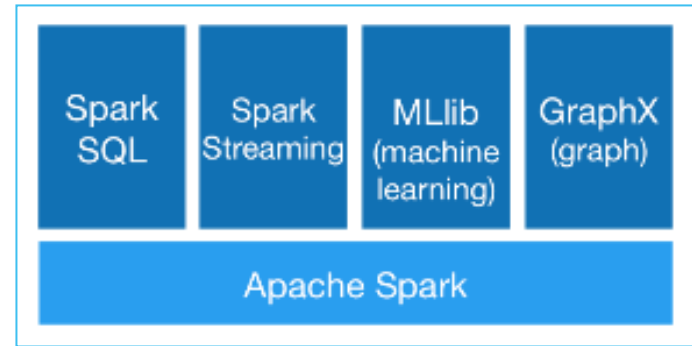
Suraj Pandey

Cloud Data Services Technical Lead, Australia and New Zealand



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.

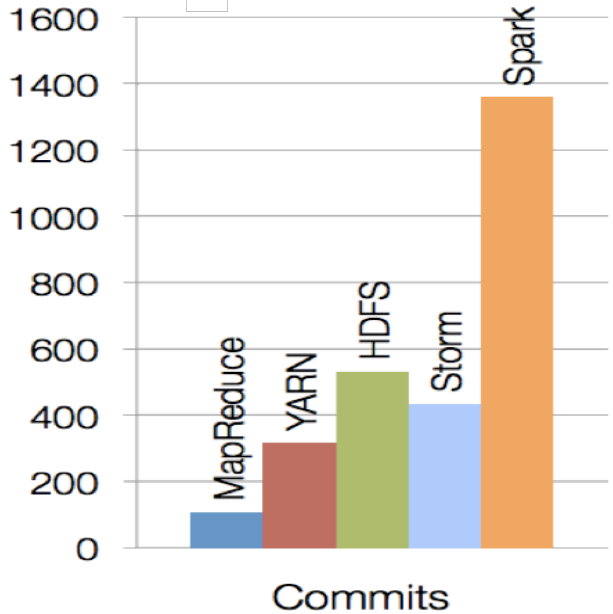


from <http://spark.apache.org>

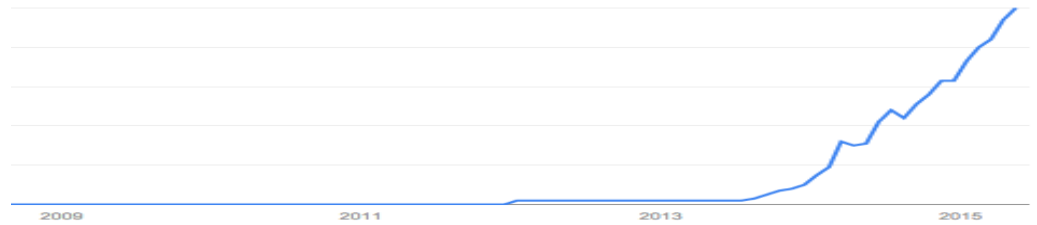


Evolving quickly, High Traction

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**

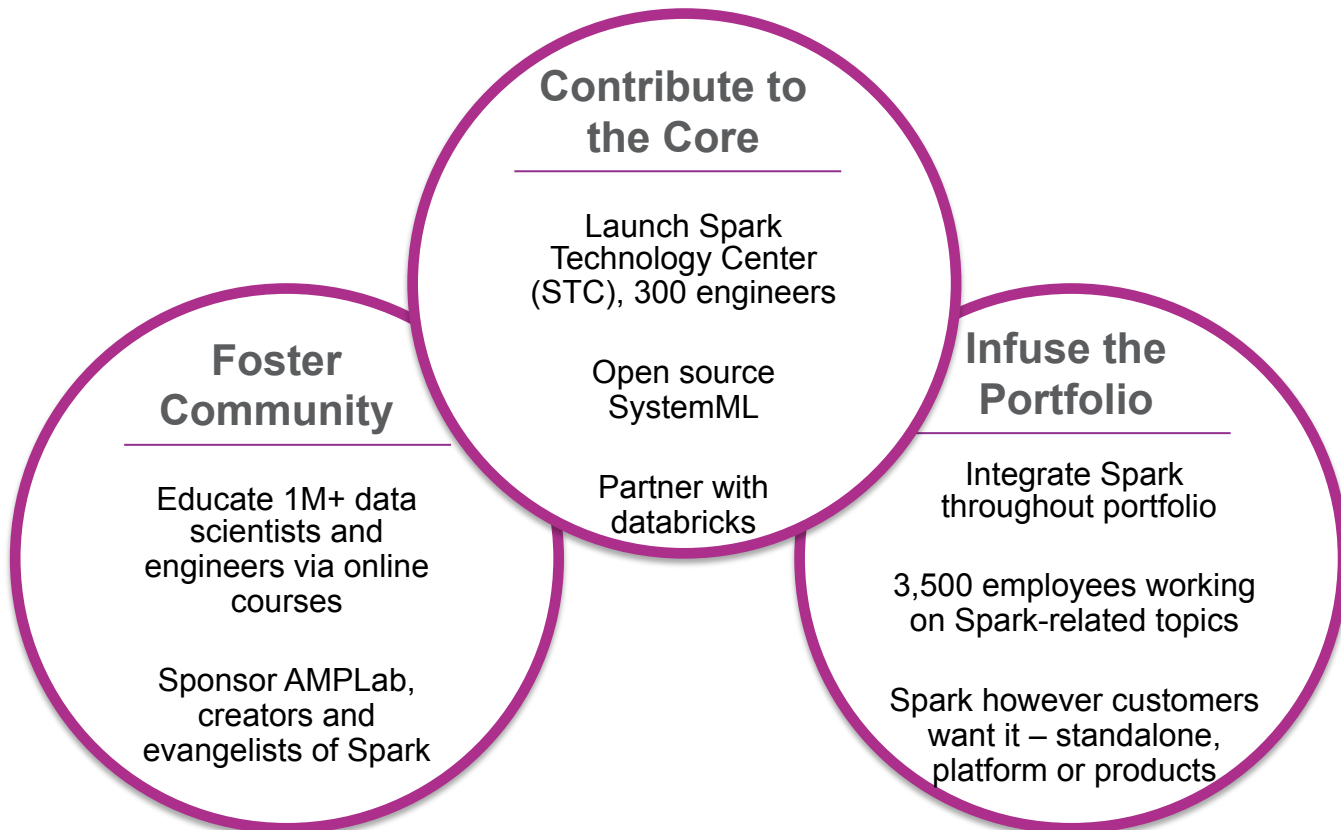
Improves with age



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



IBM is all-in on its commitment to Spark





IBM's vision for IBM Analytics for Apache Spark

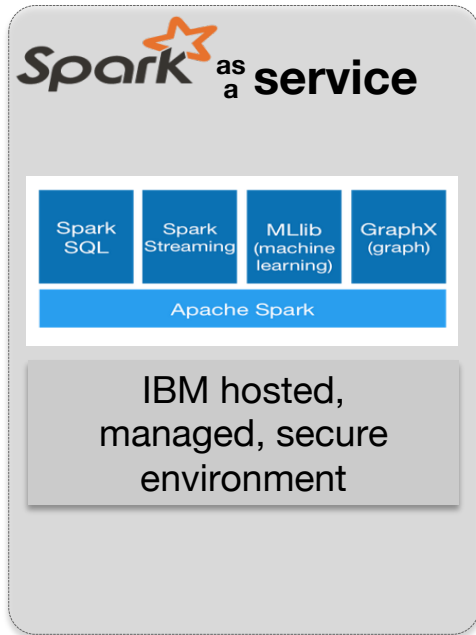
We make Spark
ACCESSIBLE and **USEFUL**

- 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

<http://ibm.com/spark>



Fully-managed Spark environment accessible on-demand

- Access to Spark as a Service for data processing at scale
- 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



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 <http://ibm.com/spark>
- **Try Spark with Hadoop** - Download at IBM.com/Hadoop
- **Engage** - Join the IBM Spark Technology Center at www.spark.tc

IBM



Cloud
Data
Services.

IBM