IBM Cloud Data Services
Overview
Agenda

- Benefits
- Cloud Data Services
  - Cloudant
  - dashDB
  - DB2 on Cloud
  - Compose.io
  - BigInsights on Cloud
  - Spark
  - DataWorks
Why Move to Cloud-based Data Services?

- Faster Innovation
- Better IT Economics
- Lower Risk of Failure
New Markets and New Buyers
Cloud Data Services Span a Variety of Needs

**Customer Data Center (On-Premises)**
- **Software**
  - Select, purchase, provision and operate hardware & software
  - Complete control of every operating parameter
- **Appliance**
  - Expertly engineered hardware & software
  - Customer responsible for operation
  - Fixed configuration = less control

**Cloud Data Center (Off Premise)**
- **Cloud Image**
  - Expertly engineered cloud configurations
  - IaaS manages hardware
  - Customer has full control over software & operation
- **Managed Service**
  - IBM operates service and controls everything
  - IBM delivers on SLA
  - Customer concerns only with data

---

Control — Simplicity
Data on Premises – Workload Optimized Products

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

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

**Cloudant Local**
- Massively scalable
- Eventual consistency model
- Built for Systems of Engagement

**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

**DataStage**
On-Premises to Cloud ETL Software
IBM Cloud Data Services

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

BigInsights for Apache Hadoop
IBM Analytics for Apache Spark
- Spark for in-memory processing
- Built on IBM Open Platform
- Bare metal performance
- Enterprise features

Cloudant
- Database as a Service (DBaaS)
- Massively scalable for global data distribution
- Eventually consistent data model
- Built for mobile, Systems of Engagement

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

DB2 on Cloud
- Hosted on SoftLayer virtual private nodes
- Administration by the customer DBA
- 4 high performance hardware configurations
- 2 database software software tiers
- DB2 10.5 native encryption
- Built for Systems of Record
Cloudant – Non-Relational Operational DBaaS

- **Distributed NoSQL “Data Layer”**
  - Powering Web, Mobile & IoT since 2009
  - Transactional JSON NoSQL Document database accessible using a RESTful API
  - Spreads data across data centers & devices for scalability & high availability

- **Available As**
  - A fully managed DBaaS
  - By you on-premises using Cloudant Local
  - Hybrid

- **Ideal for applications that require**
  - Massive, elastic scalability
  - High availability
  - Geo-location services
  - Full-text search
  - Occasionally connected users
  - Flexible database schemas that are fluid
## Cloudant Is Fully Hybrid Cloud Enabled

- **Public Cloud, Private Cloud or combine them into a Hybrid Cloud**

<table>
<thead>
<tr>
<th>Cloudant Managed</th>
<th>Cloudant Local</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guaranteed Performance with Public Cloud DBaaS</td>
<td>The Power of DBaaS in the privacy of your data centers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Cloudant Dedicated</th>
<th>Cloudant Multi-tenant</th>
<th>Cloudant Local</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hosted &amp; Managed by Cloudant</td>
<td>Hosted &amp; Managed by Cloudant</td>
<td>Customer-hosted &amp; managed with Cloudant DevOps tooling</td>
<td></td>
</tr>
<tr>
<td>24x7 Premium Support</td>
<td>Community Support</td>
<td>24x7 Premium Support</td>
<td></td>
</tr>
<tr>
<td>Use for <strong>production deployment</strong>, development</td>
<td>Use for <strong>development</strong> &amp; prototyping</td>
<td>Use for production deployment, development</td>
<td></td>
</tr>
<tr>
<td>Monthly, per-node fee</td>
<td>Monthly, metered usage fee</td>
<td>Up-front perpetual license, or monthly, per-node fee</td>
<td></td>
</tr>
<tr>
<td>Available @ 30+ SoftLayer, AWS, Rackspace, Azure data centers</td>
<td>Available on SoftLayer, Rackspace, Amazon, Azure</td>
<td>On-premise or cloud platform of choice</td>
<td></td>
</tr>
<tr>
<td>Single-tenant clusters</td>
<td>Multi-tenant clusters</td>
<td>Single-tenant clusters</td>
<td></td>
</tr>
</tbody>
</table>
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)
# dashDB – Key Use Cases

| Cloudant Analytics | Easy synchronization of JSON to structured data  
|                    | Allows analytics via standard BI tools  
|                    | In-database predictive algorithms allow greater insight for Cloudant users than ever before  
| Extend or Modernize | Extend on-premise data warehouses to the cloud  
|                    | Flexible, cost-effective growth  
|                    | Hybrid Cloud model to support ground to cloud  
| In-Database Analytics | Robust predictive analytic algorithms  
|                    | Integrated with R  
|                    | Watson Analytics Ready  
|                    | Analytics Ecosystem with Partners  
| Data Warehouse & Analytics Service | Data Warehousing and Analytics in the Cloud  
|                                    | Cloud Agility and Flexibility  
|                                    | Analytics for Cloud Data, Data Marts, and development and test environments  

© 2016 IBM Corporation
dashDB – Loading Data is Easy...

- From a File
  - Local on your Desktop
  - SoftLayer Swift Object Storage
  - Amazon S3

- DataWorks

- InfoSphere DataStage

- Replicating Data from a Cloudant Database

- Built-in Twitter Load

- SoftLayer Data Transfer Service

- Remote Load, Import, Ingest Utilities

- Data Conversion Workbench
dashDB – Aspera for High-Speed Data Transfer

- **Load from Desktop now includes Aspera as a file transfer option (currently in beta)**
  - Requires installation of Aspera Connect browser plugin

- **Greater than 100x transfer speeds** possible (compared to regular transfer method)
  - Very dependent on the network so you may only see moderate speed-ups in some cases

- **Data is not staged on SoftLayer Object Storage – direct transfer to dashDB server**

- **Beta includes limitations around size of each data transfer**
  - Entry plan: 400 MB
  - Enterprise plans: 2 GB
DB2 on Cloud

- Provides a Hosted DB2 Environment
  - Provisioned via Bluemix
  - Hosted on IBM SoftLayer
    - Virtual private nodes (not shared) or bare metal depending on configuration size
  - Administered by your organization's administrators
    - e.g. OS updates, DB2 updates, backups, HA configuration
  - Paid for on a month-to-month basis via a subscription model with support included

- Benefits
  - Convenience without the loss of control on cost effective infrastructure
  - Four high performance hardware configurations and two database software tiers
  - BLU Acceleration
  - Native encryption support included in all configurations ensuring data security in the cloud

- Five T-shirt sized configurations
  - Small, Medium, Large, X-Large and 2X-Large

- Two versions of DB2 available
  - IBM DB2 Workgroup Server Edition
  - IBM DB2 Advanced Enterprise Server Edition
**IBM Compose – Services & Deployments**

- **Compose is a cloud DBaaS platform that delivers IBM-Managed, Self-Hosted, and Public open source database services that can be adopted individually or as part of a fit-for-purpose production data strategy**
  - Compose databases are delivered as a complete, production-ready configuration
  - Daily backups and weekly / monthly snapshots
  - Intuitive dashboards with integrated monitoring and alerting
IBM Compose – Services & Deployments

Compose services are deployable to both SoftLayer and AWS. Available on SoftLayer as:
- IBM-Managed service
- Public Multi-Tenant
** Self-Hosted services coming soon

Available on AWS as:
- IBM-Managed service
- Self-Hosted service
- Public Multi-Tenant

IBM SoftLayer

Amazon AWS
IBM Compose – Services & Deployments

Compose services complement hosted and managed services from IBM Cloud Data Services (CDS)

<table>
<thead>
<tr>
<th>IBM Compose</th>
<th>Cloudant</th>
<th>Analytics for Apache Spark</th>
<th>BigInsights on Cloud</th>
<th>dashDB</th>
<th>DB2 on Cloud</th>
<th>Object Storage</th>
<th>DataWorks</th>
<th>Watson Analytics</th>
<th>IBM Graph</th>
</tr>
</thead>
</table>

**IBM SoftLayer**
Introducing IBM Compose Enterprise

- Offers benefits of a managed Compose platform on a dedicated, single-tenant data cluster
  - Delivered as a platform: customers have the flexibility to deploy any combination of Compose services that fit within the cluster’s capacity

- Deployable as a Self-Hosted (customer managed) Compose Enterprise cluster on a customer’s own IaaS or licensed as an IBM-Managed cluster
  - IBM-Managed available for both IBM SoftLayer and Amazon AWS
  - Self-Hosted available for AWS today; availability on SoftLayer coming soon
  - Deployments to a customer’s own data center (on-premises) coming in future
  - Once provisioned, the developer experience of building applications and working with the database platform is identical for Compose Enterprise data clusters (hosted or managed)

- Provides Enterprise-specific capabilities over Compose’s public and multi-tenant offerings:
  - Encryption at-rest
  - Additional plugins
  - Granular control over backups (coming soon)
When to Choose Compose Enterprise over Compose Public (Multi-Tenant)?

- **Data isolation**
  - Compose Enterprise’s dedicated cluster mitigates “noisy neighbors” and the potentially-negative performance implications of shared & multi-tenant environments
  - Addresses regulatory compliance and security concerns

- **Access to optimized configurations not supported by Compose Public**
  - Optimize performance characteristics of your clusters
  - Customize your environment with non-standard database configurations (e.g. Elasticsearch plugins)

- **Run Compose databases within non-public data centers via Self-hosted Enterprise clusters**
  - Customer’s self-managed IaaS account gives them control over networks, integration with other apps, etc.
  - Ability to deploy within a customer’s private (non-SoftLayer, non-AWS) data center coming in future
  - Alignment with IBM Hybrid Cloud Strategy of Public, Dedicated, and Local
Deployment Options for Compose Enterprise – Self-Hosted, IBM-Managed

- **Self-Hosted:** Customers install Compose data cluster into their own IaaS account
  - Customer must provision their own network, storage, and services
  - Compose software installed self-service via compose.io
  - Separate billing for IaaS and Compose software licensing
  - Use IaaS-provider VPC to connect apps and services
    - VPC: Virtual Private Cloud (SoftLayer, AWS)
  - Not sold by IBM today (ability to sell subscriptions is coming)

- **IBM-Managed:** Hosted and managed in the cloud by IBM
  - Customers select from pre-defined cluster sizes, with flexibility to grow
  - Similar experience to Cloudant Dedicated customers
  - Future support for Bluemix Dedicated
  - Single bill from IBM
  - Sold today via PPA / IBM Salesforce
### Compose Enterprise: IBM Managed is Available in 3 Sizes

<table>
<thead>
<tr>
<th>Plan</th>
<th>Cluster Description</th>
<th>Single Service Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Starter Cluster</strong></td>
<td>• Dedicated 3 node cluster</td>
<td>• MongoDB ➔ supports 160GB disk OR</td>
</tr>
<tr>
<td></td>
<td>• 16GB RAM per node</td>
<td>• Elasticsearch ➔ supports 160GB disk OR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• PostgreSQL ➔ supports 160GB/240GB** disk OR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Redis ➔ supports 24GB RAM</td>
</tr>
<tr>
<td><strong>Transactional Cluster</strong></td>
<td>• Dedicated 3 node cluster</td>
<td>• MongoDB ➔ supports 640GB disk OR</td>
</tr>
<tr>
<td></td>
<td>• 64GB RAM per node</td>
<td>• Elasticsearch ➔ supports 640GB disk OR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• PostgreSQL ➔ supports 640GB/960GB** disk OR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Redis ➔ supports 96GB RAM</td>
</tr>
<tr>
<td><strong>Large Transactional Cluster</strong></td>
<td>• Dedicated 3 node cluster</td>
<td>• MongoDB ➔ supports 2.4TB disk OR</td>
</tr>
<tr>
<td></td>
<td>• 256GB RAM per node</td>
<td>• Elasticsearch ➔ supports 2.4TB disk OR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• PostgreSQL ➔ supports 2.4TB/3.6TB** disk OR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Redis ➔ supports 366GB RAM</td>
</tr>
</tbody>
</table>

- Instance is a 3 node cluster
- Purchased only through an IBM seller/reseller
- Minimum term is 1 month (matches dashDB and Cloudant minimum license terms)
- Will add other plans such as a bulk data cluster, optimized for large storage and low transactions
- Will add incremental capacity parts to grow existing clusters
- Will offer a Bluemix Dedicated version in March at a similar premium to dashDB and Cloudant

** PostgreSQL capacity planning for a single Enterprise instance, followed by total capacity when spread across multiple smaller database instances**
IBM-Managed Enterprise Flexibility Supports Combinations of Services

- Transactional Cluster configured for a single service (one MongoDB database)

- Alternatively, a Transactional Cluster may be configured for a variety of services. New services can be added until 80% RAM or 100% disk is utilized.

- Simplifies trial & experimentation for customers using new database types
  - Start with MongoDB, add Elasticsearch for analytics, and layer on Redis for rapid data caching

- As use cases evolve and new services become available through Compose, customers of Compose Enterprise have flexibility of service capacity to adapt
## Compose Enterprise: Availability Roadmap

<table>
<thead>
<tr>
<th>Part</th>
<th>Self-Serve PAY-GO via Compose.io</th>
<th>IBM PPA</th>
<th>Bluemix Dedicated</th>
<th>Bluemix Local</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Hosted on AWS</td>
<td>Feb 2016</td>
<td>2Q16</td>
<td>TBD</td>
<td>N/A</td>
</tr>
<tr>
<td>Self-Hosted on SoftLayer</td>
<td>2Q16</td>
<td>2Q16</td>
<td>TBD</td>
<td>N/A</td>
</tr>
<tr>
<td>IBM-Managed on AWS (1 part per defined configuration)</td>
<td>TBD</td>
<td>Feb 2016</td>
<td>TBD</td>
<td>N/A</td>
</tr>
<tr>
<td>IBM-Managed on SoftLayer (1 part per defined configuration)</td>
<td>TBD</td>
<td>Feb 2016</td>
<td>March 2016</td>
<td>N/A</td>
</tr>
<tr>
<td>Self-Hosted Local</td>
<td>N/A</td>
<td>TBD</td>
<td>N/A</td>
<td>TBD</td>
</tr>
</tbody>
</table>

### Compose Enterprise Launch:
- Supports 3 cluster sizes
- Expand cluster capacity by purchasing a new cluster (Note: maximum size of a single database remains the same)

### Compose Enterprise Post-GA:
- Add incremental capacity to an existing cluster (upgrade or replace nodes) via vertical scaling
- Add new nodes into a cluster via horizontal scaling
- Support for new cluster types (optimized for bulk data or memory-heavy services)
IBM InfoSphere BigInsights – Cloud Deployment Options

Manage less, analyze more

**IBM Analytics for Hadoop**
- Prototype, create mash-ups in the cloud for non-production use
- Empowers developers to rapidly drive insight from all data
- Two-node Docker Instance
- Enterprise features – BigSheets, Big SQL, Text, and BigR
- Delivered via IBM Bluemix
- 50 GB – input data space
- Extendable, Free 14-day Trial

**BigInsights for Apache Hadoop**
- For Production deployments at scale in the cloud
- Delivers flexibility and efficiency with BYOL and PAYG pricing
- Scale to meet spikes in demand without on-premise infrastructure
- Perform enterprise-class, complex analytics on Big Data Available via the IBM Cloud Marketplace
- Web-based UI for Sizing/Pricing
BigInsights for Apache Hadoop – Options

Secure, Dedicated Bare-metal Infrastructure

IBM Open Platform

IBM BigInsights Data Scientist
- Text Analytics
- Machine Learning on Big R
- Big R
- Big SQL
- BigSheets

IBM BigInsights Analyst
- Big SQL
- BigSheets

Small Nodes
- Basic data extraction, transformation, file processing, search
- 16 cores, 64GB RAM, 20 TB data disk, 8 TB OS disk, 10 GB dual path network

Medium Nodes
- Data warehouse optimization – store new data or extend warehouse
- 16 cores, 128GB RAM, 28 TB data disk, 8 TB OS disk, 10 GB dual path network

Large Nodes
- Advanced Analytics – intensive data processing
- 16 cores, 192GB RAM, 32 TB data disk, 8 TB OS disk, 10 GB dual path network
Spark is an open source in-memory application framework for distributed data processing and iterative analysis on massive data volumes

- An Apache Foundation open source project; not a product
  - Spark is open, accelerating community innovation
- An in-memory compute engine that works with data; not a data store
  - Spark is fast—100x faster than Hadoop MapReduce
- Enables highly iterative analysis on large volumes of data at scale
  - Spark is about all data for large-scale data processing
- Unified environment for data scientists, developers and data engineers
  - Spark supports agile data science to iterate rapidly
- Radically simplifies the process of developing intelligent apps fueled by data
  - Spark can be easily integrated with IBM solutions
IBM Analytics for Apache Spark

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

IBM is adopting Spark throughout the IBM Analytics portfolio…

**Within Platforms**
- IBM Open Platform
  - Spark included with distribution
- BigInsights for Apache Hadoop
  - Spark included with IOP and BigInsights runs on IOP
- IBM Streams
  - Spark is a data-at-rest engine
- IBM DataWorks
  - Spark is the distributed data processing engine

**Within Solutions**
- IBM Commerce
  - DemandTec
    - Spark used for marketing automation
  - Unica
    - Spark used for pricing automation
- Watson Analytics
  - Spark used to analyze and refine data
- Watson Health
  - Spark used to analyze massive amounts of personal health data

**Future**
- IBM SPSS
- IBM PureData for Analytics
  - Enable Fluid Query to query Spark data
- IBM Research
  - Currently has over 30 active projects utilizing Spark
  - …and many more
A fully managed data preparation and movement service for IBM Cloud Data Services that enables business analysts, developers, data scientists and engineers to put data to work.

Access and publish data from on-premises and cloud sources
Fully managed by IBM with PAYG and Subscription options
Powered by Spark for a speedy and responsive experience
Available in Bluemix and Watson Analytics

**Analysts & Data Scientists**
Find and use the data they need to accelerate data based business decisions using timely accurate and trusted information

**Developers**
Quickly develop data-rich applications by embedding the DataWorks service REST API into new or existing applications

**IT Admins & Data Engineers**
Enable self-service data access to end users and deliver data faster while still maintaining data governance and security
DataWorks Application

A simple and intuitive application for end users to visually interact with DataWorks to combine, clean, shape and deliver trusted data with confidence

- Empowers users to find relevant and useful data
- Automatically profiles and classifies data on ingestion
- Delivers profiling and quality metrics to visualize, understand and trust the data
- Easily refine and enrich the quality of data by filtering, joining, sorting and more...
- Allows users to save and run activities and deliver the data to their desired targets
Stephen Holt – seholt@us.ibm.com

NY Area Cloudant PoT 4/6 – http://ibm.biz/MobileData