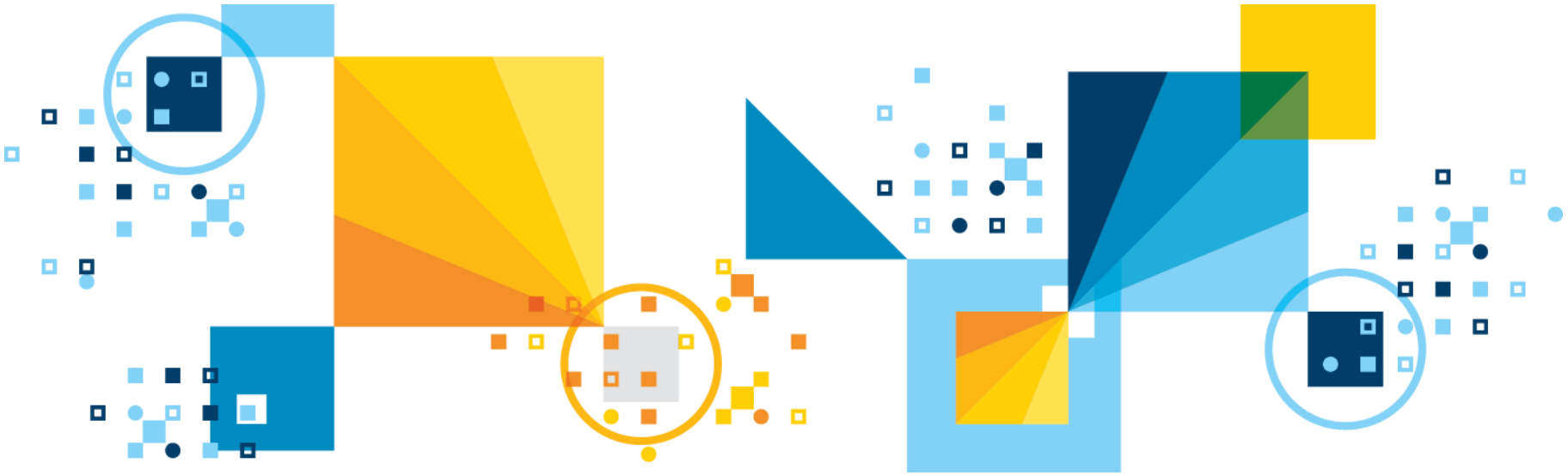
IBM Cloud Data Services (Public, Private & Hybrid Cloud)

Data Server Day - Tuesday, May 16th, 2017

Kelly Schlamb
IBM Canada Ltd.
Email: kschlamb@ca.ibm.com, Twitter: @KSchlamb



Legal Disclaimer

- © IBM Corporation 2017. All Rights Reserved.
- The information contained in this publication is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this publication, it is provided AS IS without warranty of any kind, express or implied. In addition, this information is based on IBM's current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this publication or any other materials. Nothing contained in this publication is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software.
- References in this presentation to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates. Product release dates and/or capabilities referenced in this presentation may change at any time at IBM's sole discretion based on market opportunities or other factors, and are not intended to be a commitment to future product or feature availability in any way. Nothing contained in these materials is intended to, nor shall have the effect of, stating or implying that any activities undertaken by you will result in any specific sales, revenue growth or other results.
- If the text contains performance statistics or references to benchmarks, insert the following language; otherwise delete:
  Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.
- If the text includes any customer examples, please confirm we have prior written approval from such customer and insert the following language; otherwise delete:
  All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.
- Please review text for proper trademark attribution of IBM products. At first use, each product name must be the full name and include appropriate trademark symbols (e.g., IBM Lotus® Sametime® Unyte™). Subsequent references can drop "IBM" but should include the proper branding (e.g., Lotus Sametime Gateway, or WebSphere Application Server). Please refer to http://www.ibm.com/legal/copytrade.shtml for guidance on which trademarks require the ® or ™ symbol. Do not use abbreviations for IBM product names in your presentation. All product names must be used as adjectives rather than nouns. Please list all of the trademarks that you use in your presentation as follows; delete any not included in your presentation. IBM, the IBM logo, Lotus, Lotus Notes, Notes, Domino, Quickr, Sametime, WebSphere, UC2, PartnerWorld and Lotusphere are trademarks of International Business Machines Corporation in the United States, other countries, or both. Unyte is a trademark of WebDialogs, Inc., in the United States, other countries, or both.
- If you reference Adobe® in the text, please mark the first use and include the following; otherwise delete:
  Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
- If you reference Java™ in the text, please mark the first use and include the following; otherwise delete:
  Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.
- If you reference Microsoft® and/or Windows® in the text, please mark the first use and include the following, as applicable; otherwise delete:
  Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both.
- If you reference Intel® and/or any of the following Intel products in the text, please mark the first use and include those that you use as follows; otherwise delete:
  Intel, Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
- If you reference UNIX® in the text, please mark the first use and include the following; otherwise delete:
  UNIX is a registered trademark of The Open Group in the United States and other countries.
- If you reference Linux® in your presentation, please mark the first use and include the following; otherwise delete:
  Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both. Other company, product, or service names may be trademarks or service marks of others.
- If the text/graphics include screenshots, no actual IBM employee names may be used (even your own), if your screenshots include fictitious company names (e.g., Renovations, Zeta Bank, Acme) please update and insert the following; otherwise delete: All references to [insert fictitious company name] refer to a fictitious company and are used for illustration purposes only.

International DB2 Users Group

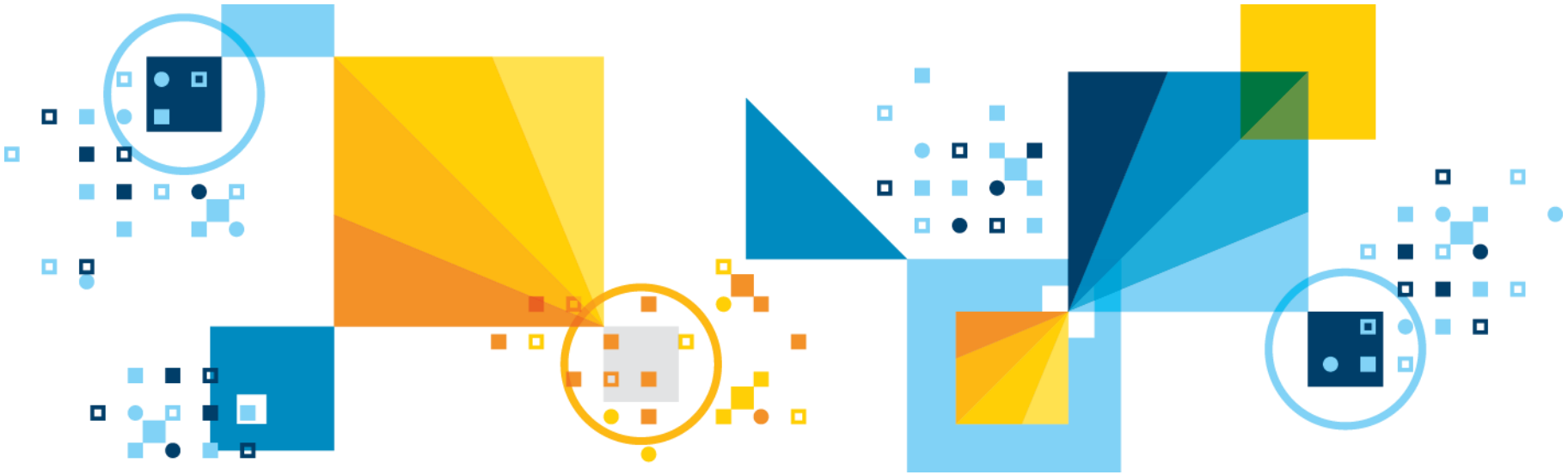
IDUG DB2 Tech Conference
Lisbon, Portugal
October 1-5, 2017

- The premier European DB2 Conference will take place on **October 1-5, 2017 at Lisbon, Portugal.** Make your plans to attend and experience the latest in DB2 technologies, networking opportunities and the technical content you need to be successful.
- The 2017 event offers new educational opportunities and more training than ever before! Attendees will experience: 5 days of education sessions
- Half and full-day **workshops**
- More than **100 one-hour technical sessions**
- Three expert panels on **z/OS, LUW & Application Development**
- And much more!
- Register at www.idug.org before **7 August 2017** to receive ***Early Bird Discount!***
- Conference hotel: **EPIC SANA Lisboa Hotel, Lisbon**

Agenda

- **The Cloud Journey**
  - Cloud deployment models and as-a-service
  - SoftLayer & Bluemix
- **IBM's Cloud Data Services**
  - DB2 on Cloud
  - Cloudant
  - Compose
  - BigInsights
  - Apache Spark
  - dashDB for Analytics
  - dashDB for Transactions
  - dashDB Local (for Analytics)
  - Bluemix Lift
  - Bluemix Data Connect
- **Hybrid Data Warehousing**

The Cloud Journey



Why the Journey to Cloud-based Data Services?

MISSION

To provide the **best experience for data professionals** to engage and build with a **comprehensive platform of trusted, flexible data services** covering content, data and analytics.

FASTER INNOVATION	BETTER IT ECONOMICS	LOWER RISK OF FAILURE
Instant provisioning saves weeks of data center setup	Pay as you go with minimal up-front capital investments	Fully managed 24x7 so you can focus on new development

Public and Private Cloud Drives Better Business Economics

Agility is the #1 reason to adopt the Cloud NOT cost savings. Cloud enables businesses to compete *faster*. And speed kills, a chess grandmaster would lose to a beginner that got four moves every turn.

1 Cost Agility

- Shifts fixed to variable cost
- Pay as and when needed
- CAPEX vs. OPEX

2 Business Scalability

- Provides limitless, cost-effective computing capacity to support growth

3 Market Adaptability

- Faster time to market
- Supports experimentation

4 Masked Complexity

- Expands product sophistication
- Simpler for customers and users

5 Context-Driven Variability

- User defined experiences
- Increases relevance

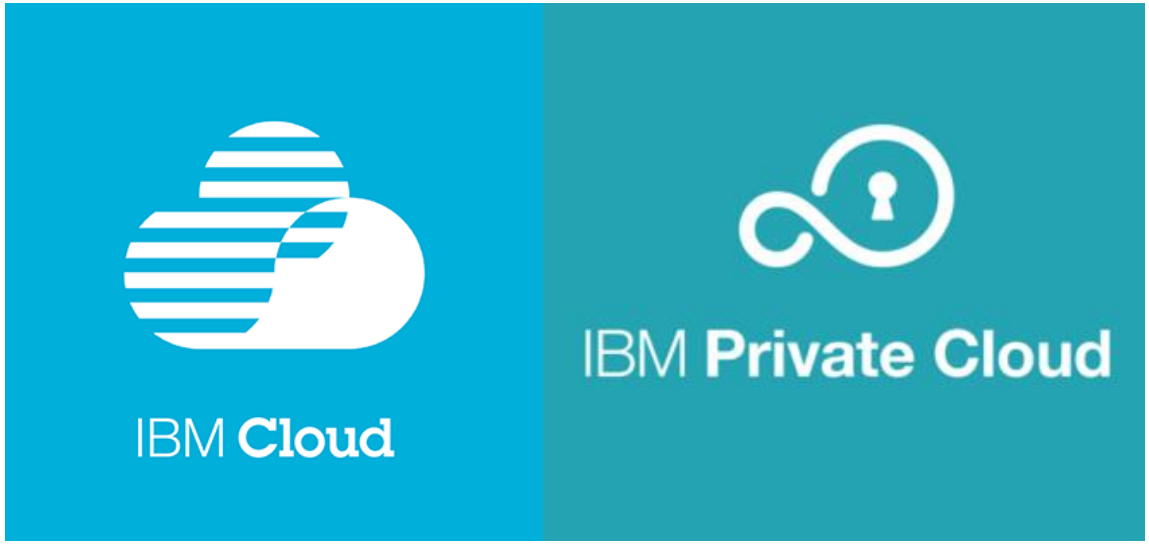
6 Ecosystem Connectivity

- New value nets potential new business

Source: IBV Analysis

Its not about Cloud or On-Premises its about Cloud AND On-Premises

It's About Hybrid



"Through 2020, the most common use of Cloud services will be a hybrid model combining on-premises and external Cloud services."

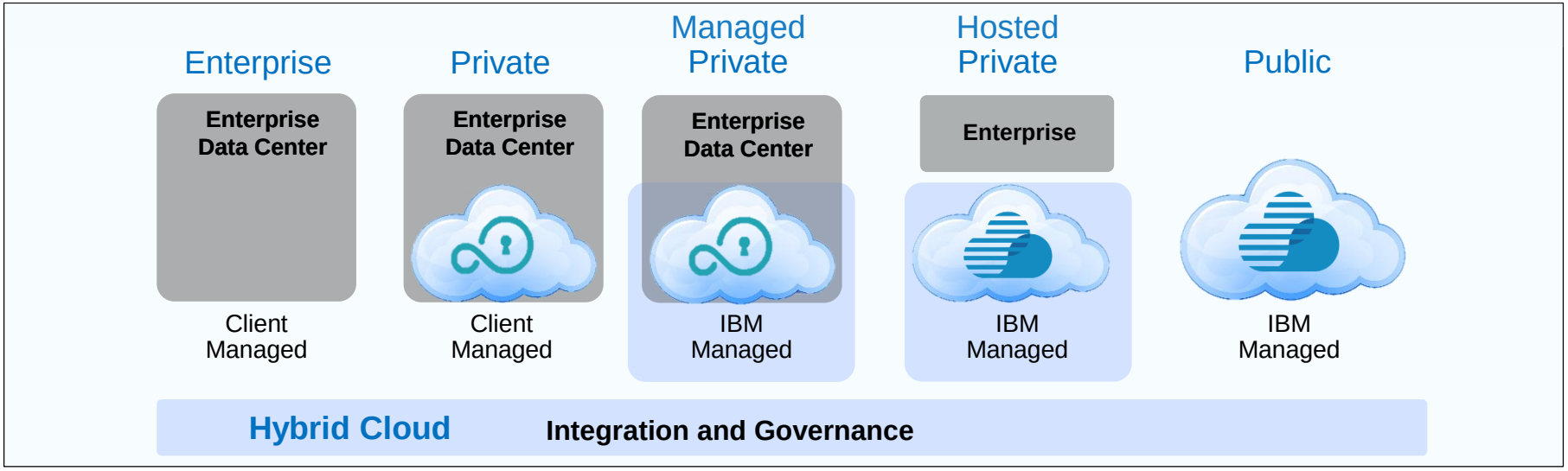
 Gartner, Cloud Computing Innovation Key Initiative Overview

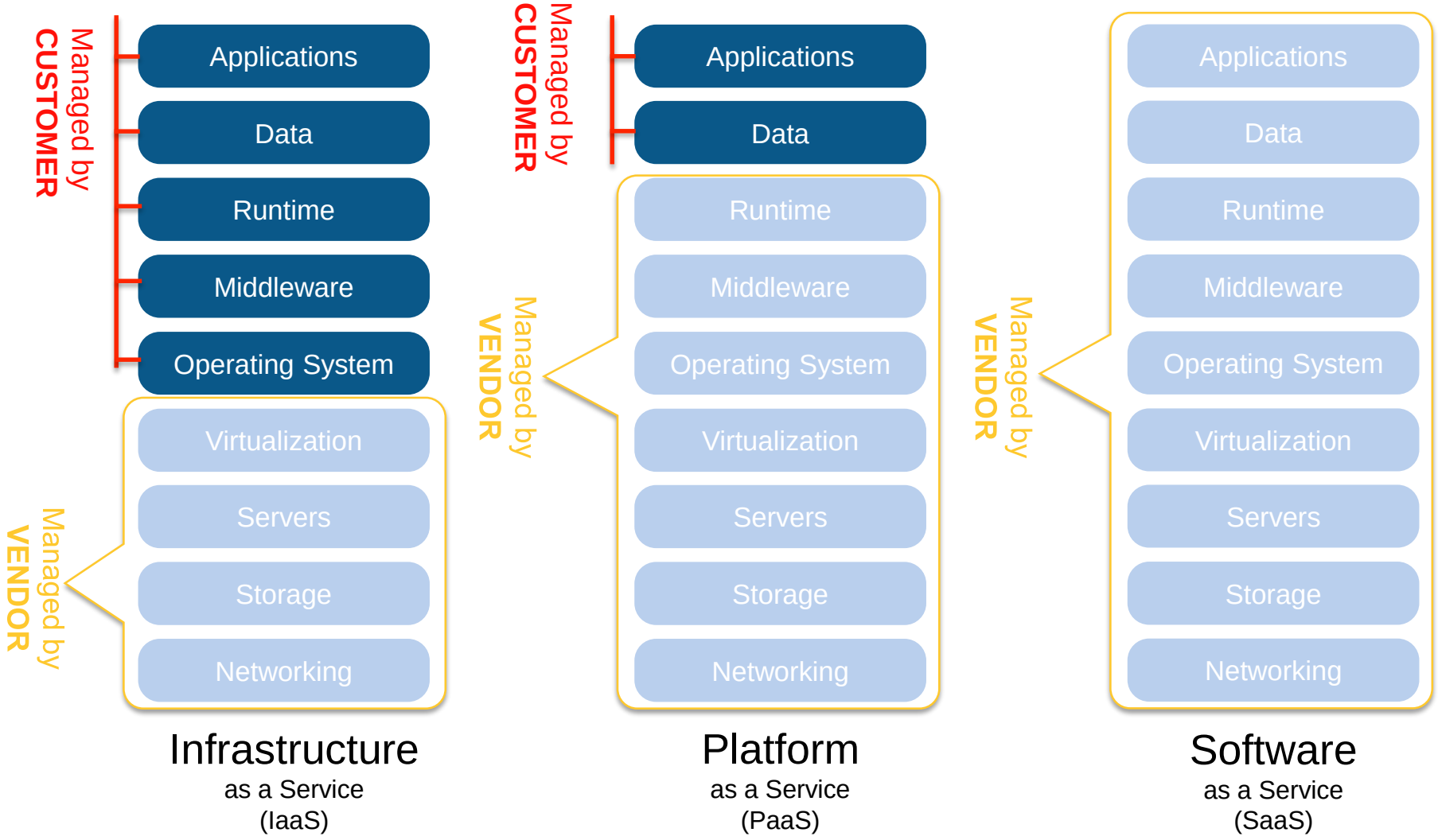
Cloud Deployment Models

1. **Private Cloud** – A Cloud dedicated for sole use and resides either on-premises or off-premises that is Client-managed or IBM managed.

2. **Public Cloud** – When the Cloud is off-premises and resources may be shared between clients. Hundreds to thousands of clients with controls to enforce isolation between client's data and resources. Most new ISVs are purely Public.

3. **Hybrid Cloud** – A combination of Public and Private Clouds. Most enterprise clients are Hybrid.




Managed by CUSTOMER
Applications
Data
Runtime
Middleware
Operating System
Managed by VENDOR
Virtualization
Servers
Storage
Networking
Infrastructure
as a Service
(IaaS)
Managed by CUSTOMER
Applications
Data
Managed by VENDOR
Runtime
Middleware
Operating System
Virtualization
Servers
Storage
Networking
Platform
as a Service
(PaaS)
Managed by VENDOR
Applications
Data
Runtime
Middleware
Operating System
Virtualization
Servers
Storage
Networking
Software
as a Service
(SaaS)

SOFTLAYER an IBM Company

IaaS – Infrastructure-as-a-Service

SoftLayer is IBM's Infrastructure-as-a-Service (IaaS)

- IaaS lets clients rent hardware, storage and networking without needing to know anything about the underlying data center facilities
- Servers come with operating systems to allow initial access but are not patched or managed from the default install by the IaaS
- IaaS lets you to pick where your server will be located

IBM Bluemix™

PaaS – Platform-as-a-Service

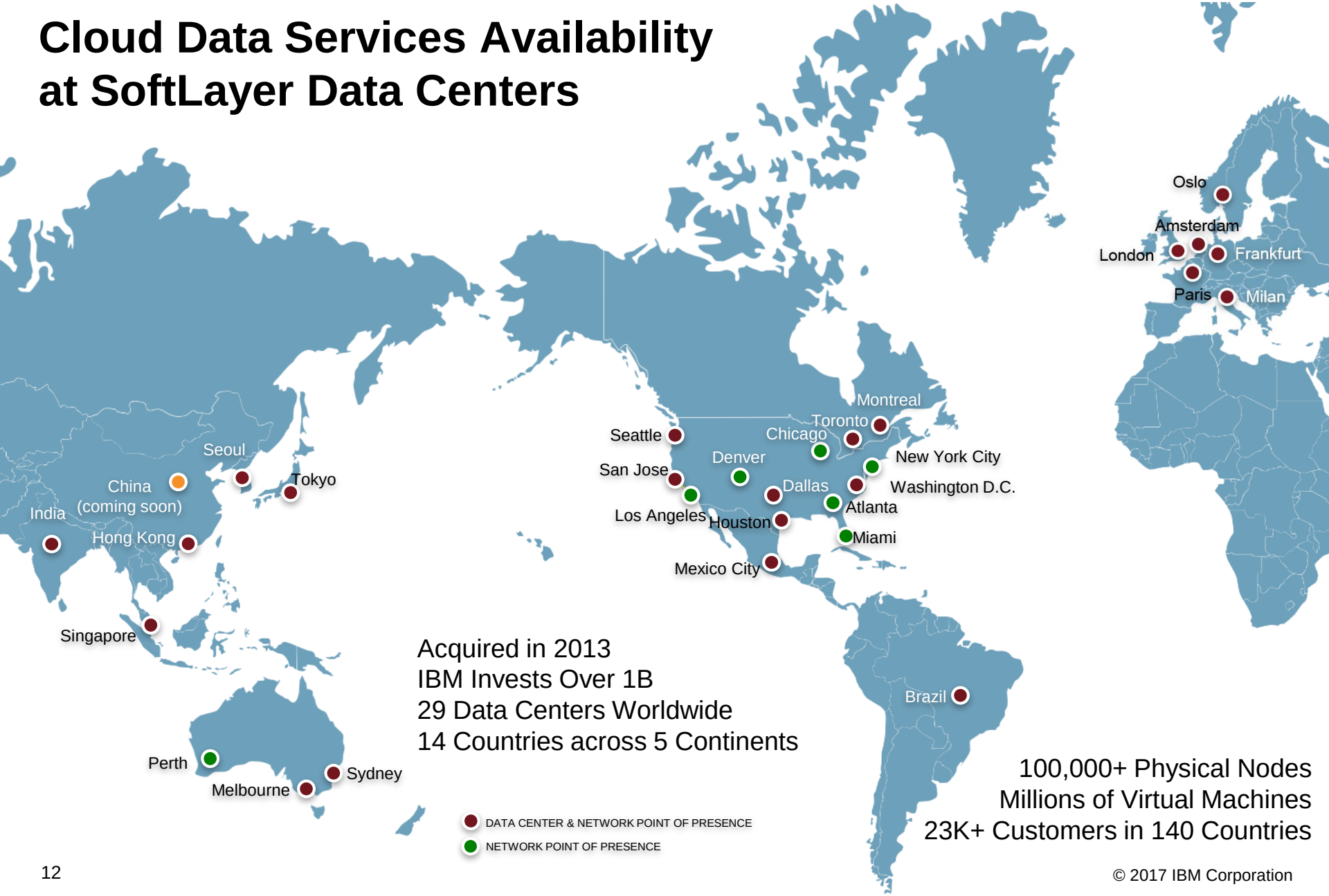
Bluemix is IBM's Platform-as-a-Service (PaaS)

- PaaS lets clients rent middleware services (databases, app servers, storage, Hadoop) without needing to know anything about the underlying hardware, storage and networking they are housed on
- PaaS still needs to know how much data you have to store
- Bluemix runs on top of SoftLayer
  - The services in Bluemix (the PaaS) are consumers of the storage, compute and network resources provided by SoftLayer (the IaaS)

Applications run on top of Bluemix

- The applications (SaaS) are consumers of the middleware services provided by the PaaS

Cloud Data Services Availability at SoftLayer Data Centers

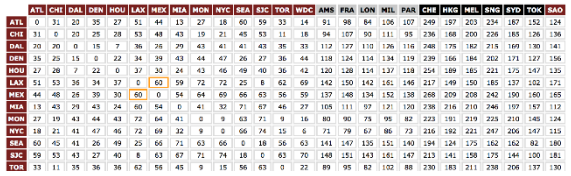


Acquired in 2013
IBM Invests Over 1B
29 Data Centers Worldwide
14 Countries across 5 Continents

100,000+ Physical Nodes
Millions of Virtual Machines
23K+ Customers in 140 Countries

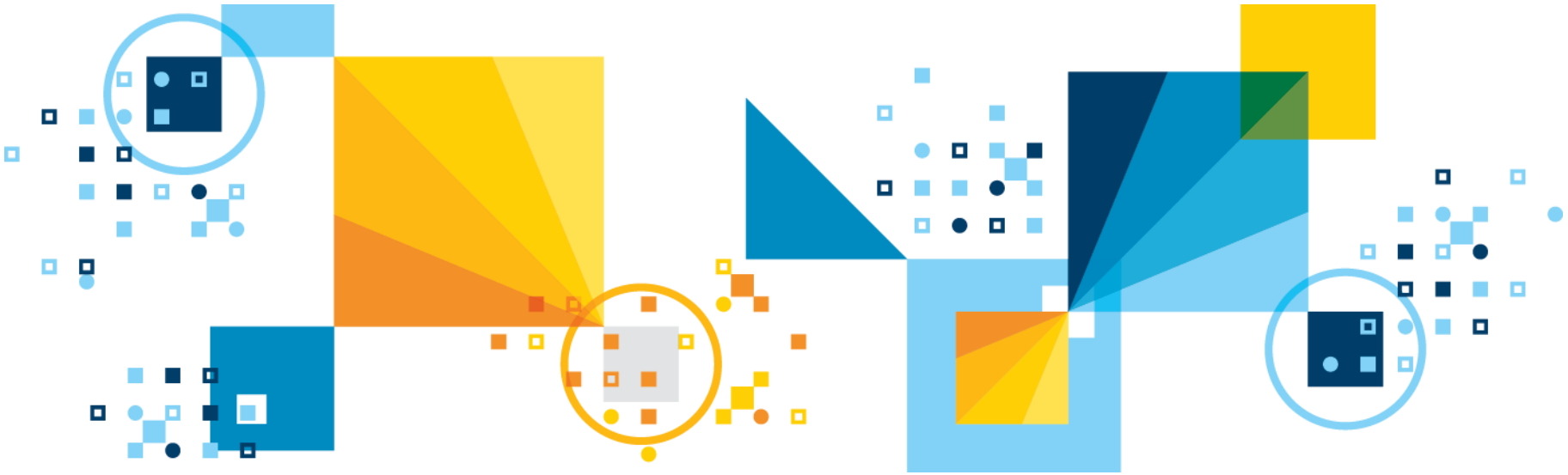
Network Latency to SoftLayer Data Centers

- Depending on distance…
  - Intra-continental: 4ms-100ms
  - Inter-continental: 70ms-250ms



Real-time latency tools:
http://www.softlayer.com/data-centers
http://lg.softlayer.com

IBM's Cloud Data Services



Data On-Premises or in the Private Cloud

ANALYTICAL

TRANSACTIONAL

UNSTRUCTURED/SEMI-STRUCTURED

STRUCTURED

InfoSphere. software

DataStage
On-Prem to Cloud ETL Connector

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

dashDB Local

- Containerized version of dashDB for Analytics
- Powered by DB2 BLU & Netezza in-database analytics

DB2

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

IBM Cloud Data Services Analytics and Transactional

IBM solutions combine workloads and data types for true managed hybrid services

ANALYTICAL

TRANSACTIONAL/OPERATIONAL

BigInsights on Cloud

- Spark for in-memory Hadoop
- Built on IBM Open Platform
- Bare metal performance
- BigInsights enterprise features

Cloudant

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

SCYLLA. mongoDB redis RethinkDB

UNSTRUCTURED

STRUCTURED

Bluemix Data Connect
Data Refinery Services

Watson Analytics
Analytics & Visualization Services

dashDB for Analytics

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

dashDB

dashDB for Transactions

- SQL interface, ACID compliant
- Optimized for transactional workloads
- Built for Systems of Record
- Oracle Compatibility

dashDB

PostgreSQL MySQL

DB2 on Cloud

- Hosted DB2 service
- Flexible deployments for analytics or OLTP

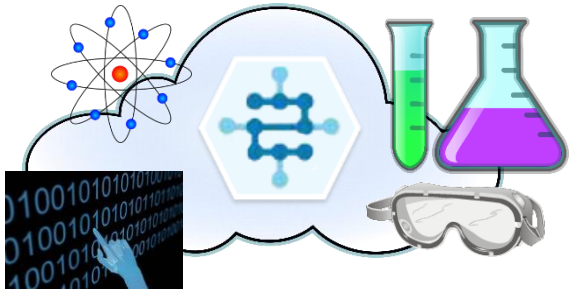
IBM DB2 on Cloud

- DB2 Relational SQL DBMS
- Hosted by IBM - on Bluemix/SoftLayer and Amazon Web Services (AWS) data centers worldwide
- Five high performance hardware configurations and two database software tiers (Standard / Advanced) to match capability and affordability needs
- Flexibility to configure the environment to meet your needs and workloads
- Managed by your DBAs
- Features include encryption, Oracle Compatibility, BLU Acceleration, etc.
- Monthly pay-as-you-go (credit card) and subscription pricing

Common DB2 on Cloud Use Cases

Cloud-based production environments

Test, development, prototyping, and proofs-of-concept



Disaster recovery

Consolidation of smaller, on-premises databases

18

DB2 on Cloud Plans and Prices (Bluemix Public / SoftLayer)

- **5 T-Shirt sized configurations: Small, Medium, Large, X-Large, 2X-Large**
- **2 versions of DB2: Workgroup (Standard), Advanced Enterprise (Advanced)**
- **Single part number for each size and version combination**
- **All configurations include native encryption**

Size / DB2 Plan	Small	Medium	Large	X-Large	2X-Large
Standard	$1,030	$1,750	$3,090	$6,180	-
Advanced	$1,290	$2,270	$4,120	$8,240	$18,000
Nodes	Virtual Private	Virtual Private	Virtual Private	**Bare Metal**	**Bare Metal**
Cores	2 x 2.0 GHz	4 x 2.0 GHz	8 x 2.0 GHz	12 x 2.4GHz	32 x 2.7 GHz
Memory	8 GB	16GB	32GB	128GB	1TB
Storage	100GB, 500GB SAN 100GB @ 500 IOPS	100GB & 1TB SAN 100GB @ 1,200 IOPS	100GB & 2TB SAN 100GB @ 1,600 IOPS	2 x 800GB SSD @ RAID 1 (~800GB) 6 x 1.2TB SSD @ RAID 10 (~3.5TB)	2 x 800GB SSD @ RAID 1 (~800GB) 16 x 1.2TB SSD @ RAID 10 (~9.5TB)
Network	1Gbps Public & Private Uplinks			10Gbps Redundant Public & Private Uplinks	
OS	RHEL 7.0				

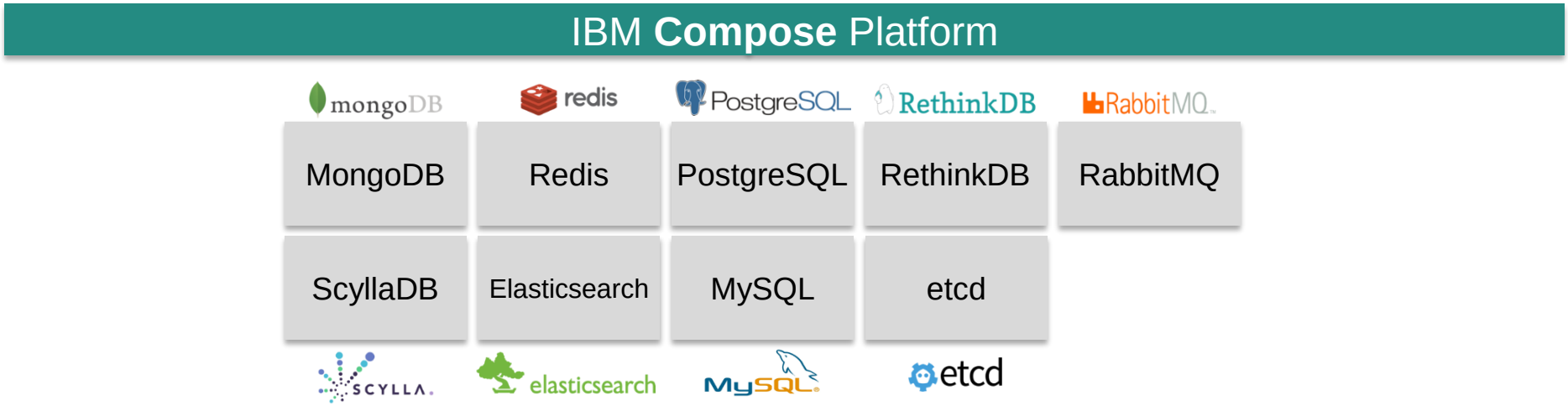
Cloudant – Non-Relational Operational DBaaS

- **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
  - 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

IBM Cloudant®

Build More ***Grow More*** ***Sleep More***

IBM Compose – Managed Platform of DBaaS Offerings



- **Compose is a managed platform for open-source DBaaS**
  - Services can be adopted individually via **Public** multi-tenant deployments
  - Entire catalogue can be licensed & deployed a la carte via **Compose Enterprise**

- **Best-practice delivery & configuration of open source technologies**
  - All services are production-ready and configured for HA out of the box
  - Automated (no-cost) backups, elastic scale-out, intuitive dashboards

IBM Compose – Services & Deployments

IBM **Compose** Platform

MongoDB	Redis	PostgreSQL	RethinkDB	RabbitMQ
ScyllaDB	Elasticsearch	MySQL	etcd	

Available on **SoftLayer** as:
- Enterprise
- Public Multi-Tenant

Compose services are deployable to **SoftLayer**, **Bluemix**, **AWS**, and **GCP**

Available on **AWS** as:
- Enterprise
- Public Multi-Tenant

IBM **SoftLayer** | Amazon **AWS**

IBM **Bluemix** | **Google** Cloud Platform

BigInsights on Cloud offering options

Enterprise

- Dedicated hardware
- Small/Med/Large dedicated hosts
- Monthly subscription
- IBM Open Platform
- IBM BigInsights
- VPN option
- Secure: ISO27K1, SOC2, HIPAA certified

Basic

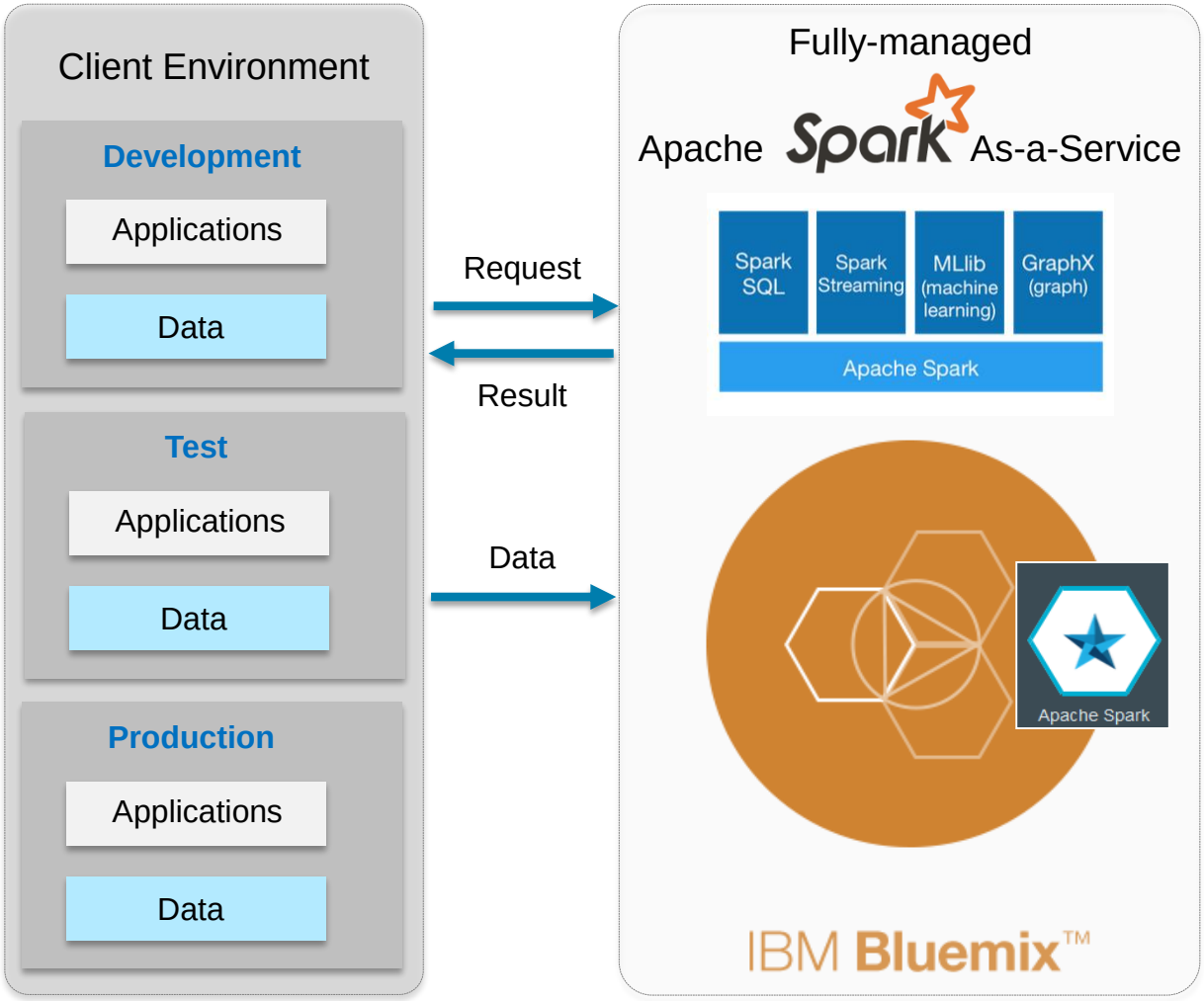
- Shared hardware
- Virtual nodes: 4 core, 24GB, 1TB disk
- Object Storage service integration
- Hourly pay-as-you-go
- Free usage for trial period
- IBM Open Platform only

The analytics operating system of the future…

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

Apache Spark in Bluemix



- Access to Spark's next-generation performance and capabilities, including built-in machine learning and other libraries
- **Personal (free) and reserved purchase options**
- 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

IBM dashDB – Analytics Warehouse as a Service

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

BLU Acceleration

Netezza In-Database Analytics

Cloudant NoSQL Integration

dashDB MPP

dashDB for Analytics

- Fully-managed data warehouse on cloud
  - Choice of SoftLayer or Amazon Web Services
- **BLU Acceleration** 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)
- On disk data encryption and secure connectivity

IBM dashDB – A Fully Managed Analytics Warehouse

Relieving the provisioning and management burden on your data warehouse DBA staff

IBM sets up, optimizes, and manages all aspects of the dashDB environment

- OS and database software installation
- Optimized configuration for analytics workloads
- BLU Acceleration "load and go" simplicity
  - No need for creating auxiliary structures like indexes or aggregates
  - Automatic memory management
  - Automatic statistics gathering
  - Automatic space reclamation
  - Pre-configured workload management
- Ongoing OS and database software maintenance
- Automated daily backups
- 24/7 monitoring and restart after hardware or software failure
- Ongoing risk assessment and security monitoring
- Upgrade support moving to a larger dashDB size

DBA only has to focus on data delivery

- Schema definition
- Data loading
- User management and database access control
- Application, ETL connectivity

dashDB for Analytics Plans

Entry (Multi-tenant)

- Shared, multi-tenant environment
- **20 GB** SAN storage capacity
- $50 / month (USD) flat rate
- FREE if < 1 GB of data used

SMP Small

- Dedicated, single tenant environment
- Virtual environment
- 16 vCPUs
- 64 GB memory
- **1 TB*** of data storage and temp space (SAN)
- $1170 / month (USD)

SMP Medium

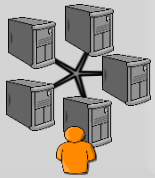
- Dedicated, single tenant environment
- Bare metal
- 32 cores
- 256 GB memory
- **4 TB*** of data storage and temp space (SAN)
- $4700 / month (USD)

SMP Large

- Dedicated, single tenant environment
- Bare metal
- 32 cores
- 256 GB memory
- **12 TB*** of data storage and temp space (SAN)
- $7370 / month (USD)

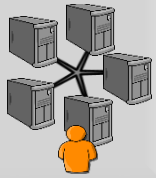
MPP Small

- Dedicated, single tenant environment
- SoftLayer Bare metal
- 3 node clusters and up
- 24 cores per node
- 256 GB memory per node
- **5 TB*** of data storage and temp space (SSD) per node
- $5410 / month (USD) per node



MPP Small for AWS

- Dedicated, single tenant environment
- AWS Dedicated Virtual environment
- 3 node clusters and up
- 32 vCPUs per node
- 244 GB memory per node
- **8 TB*** of data storage and temp space (SSD) per node
- $6250 / month (USD) per node



* Storage split between user data and temporary data. Sizes shown indicate amount of uncompressed, pre-load user data and uncompressed temporary data that can be stored (assuming typical compression rates).

dashDB Security

Functional

- ✓ Encrypted data at-rest using AES 256
- ✓ Encrypted data in-motion - via SSL/TLS
- ✓ Application authentication using embedded LDAP server
- ✓ Authorization – Grant permissions to specific tables, rows, roles
- ✓ Optionally restrict access to specify client host names

Infrastructure

- ✓ Automated backups (encrypted) on a daily basis for recovery purposes
- ✓ Dedicated bare metal nodes for dedicated local storage and compute (MPP)
- ✓ Host firewall to protect from port scans/network intrusion
- ✓ Private VLAN and VPN options available with ICIAE network service

Governance & Compliance

- ✓ ISO 27001 certified
- ✓ SOC 2 Type 1 & SOC 2 Type 2 certified, SOC 3
- ✓ HIPAA-ready
- ✓ Privacy Shield certified
- ✓ More coming!

Operational

- ✓ Monitored 24x7
- ✓ Fully managed for setup, configuration, tuning and DR operations
- ✓ InfoSphere Guardium-based audit reports and monitoring
- ✓ Sensitive data monitoring through Guardium Database Activity monitoring to discover sensitive data, connections and SQL statements

SoftLayer: physical security compliance

IBM dashDB for Transactions

Transactional database capabilities for best performance atop a fully-managed instance

Excellent Transactional Performance

Oracle Compatibility

Robust Security

dashDB **for Transactions**

- Fully-managed transactional database as a service
- **High performance database engine optimized for transactional workloads**
- **Row-organized tables**
- **Oracle compatibility**
- On disk data encryption and secure connectivity
- 3 plans plus HA versions
  - 2 cores, 8 GB memory, 500 GB HDD
  - 12 cores, 128 GB memory, 1.4 TB SSD
  - New 48 cores, 1 TB memory, 11 TB SSD

dashDB for Transactions Plans

Precise Performance 500
(2 cores - 8 GB RAM)

- Dedicated, single tenant environment
- Virtual environment
- 2 vCPUs
- 8 GB memory
- 500 GB HDD storage for client data
- $250 / month (USD) or $500 / month (USD) for HA configuration

Precise Performance 1400
(12 cores - 128 GB RAM)

- Dedicated, single tenant environment
- Dedicated bare metal
- 12 physical cores
- 128 GB memory
- 1.4 TB SSD storage for client data
- $4,000 / month (USD) or $8,000 / month (USD) for HA configuration

Precise Performance 10000
(48 cores - 1 TB RAM)

New

- Dedicated, single tenant environment
- Dedicated bare metal
- 48 physical cores
- 1 TB memory
- 11 TB SSD storage for client data
- $18,000 / month (USD) or $36,000 / month (USD) for HA configuration

Each plan available in a High Availability (HA) configuration

- Primary/standby pair with synchronous replication
- Single IP address for simplified client access to cluster
- Automatic failover

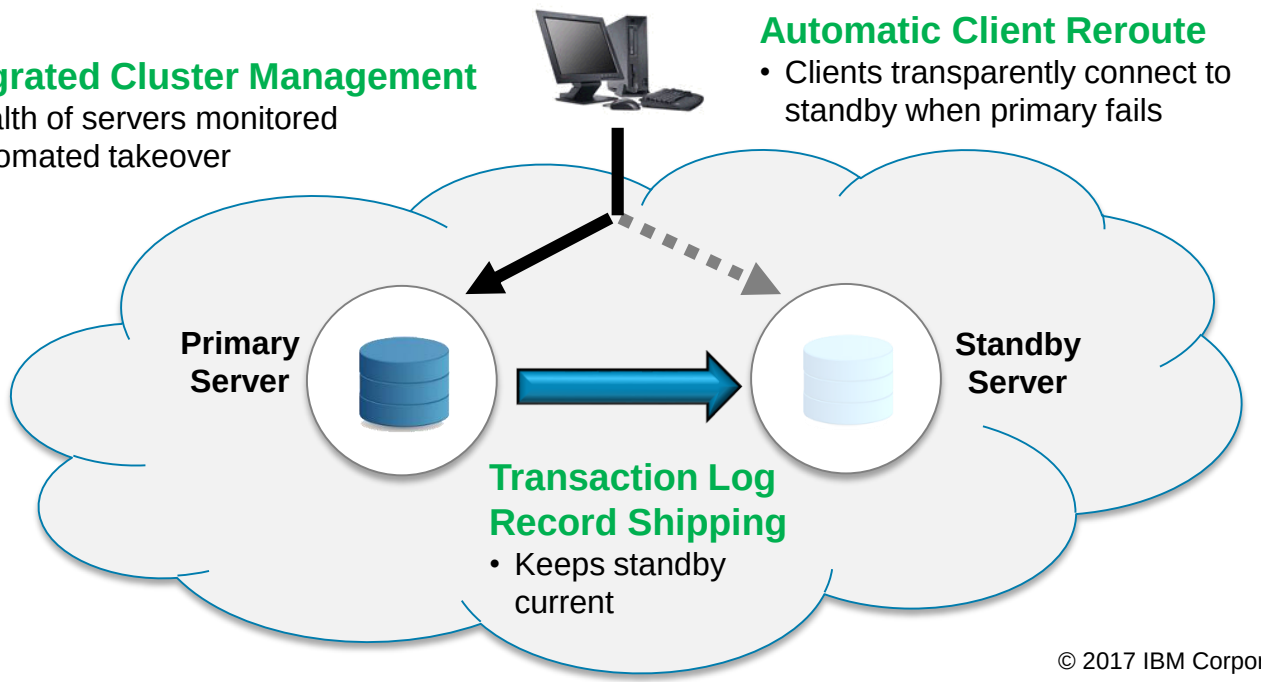
- **Fully managed**
- **Built-in Oracle compatibility**
- **Optimized for transactional workloads**
- **Data encryption with robust security**

dashDB for Transactions – High Availability Configuration

- **Reduced outage times for unplanned outages and maintenance activities**
- **Configuration includes both a primary server plus a warm standby server**
- **Log shipping keeps standby up-to-date for fast failovers**
- **Simplified client configuration – floating IP address always points to current primary**
- **Not logged operations that compromise data recoverability are blocked**

dashDB
for Transactions

High Availability Configuration



IBM Data Server Manager – Natively Supports dashDB

IBM Data Server Manager is a critical integrated part of the hybrid IBM Data Store value proposition. It's the up and running, management, and the problem determination console for DB2, **dashDB**, and BigInsights BigSQL

IBM Data Server Manager is more than a tool for monitoring DB2. It is a set of common services and capabilities for Enterprise and Application DBAs, Data Engineers and Analysts and Developers

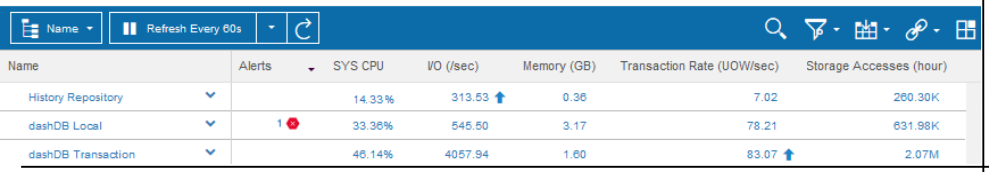
It's the center of the Hybrid Enterprise

New with Data Server Manager 2.1.2: By establishing a connection to a dashDB Local, dashDB for Transactions, or dashDB for Analytics database, Data Server Manager automatically reconfigures its feature set to compliment the selected database type

Data Server Manager 2.1.2 (released in December 2016)

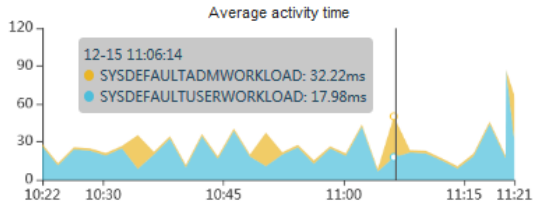
Manage your IBM data Enterprise

- Enterprise view of DB2 and dashDB KPIs and alerts from one dashboard
- Data Server Manager Login authentication through LDAP
- New enterprise grid view shows hundreds of DB2 and dashDB servers at a glance



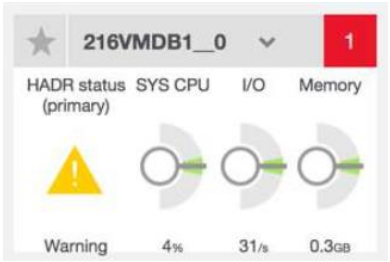
Monitor In Depth

- New powerful database performance overview page for rapid problem determination
- Database time breakdown
- 18 KPIs in one screen for realtime or historical analysis
- Breakdown by workload
- For DB2 and dashDB



Ensure DB2 High Availability

- See the availability of all of your HADR and pureScale clusters at a glance
- Dashboard, SNMP and Email alerts that highlight potential availability problems before they happen

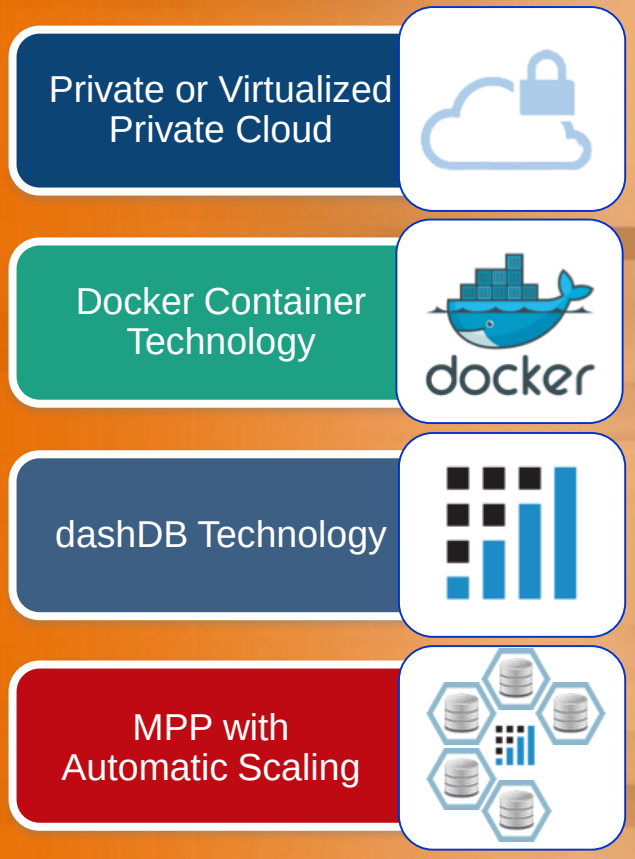


Make remote data act like local data

- Make data in remote DB2, BigSQL or dashDB databases look and act like local tables with 11.1.1.1 and Fluid Query
- Explain and monitor queries with calls to remote data sources with 11.1.1.1

Introducing IBM dashDB Local (for Analytics)

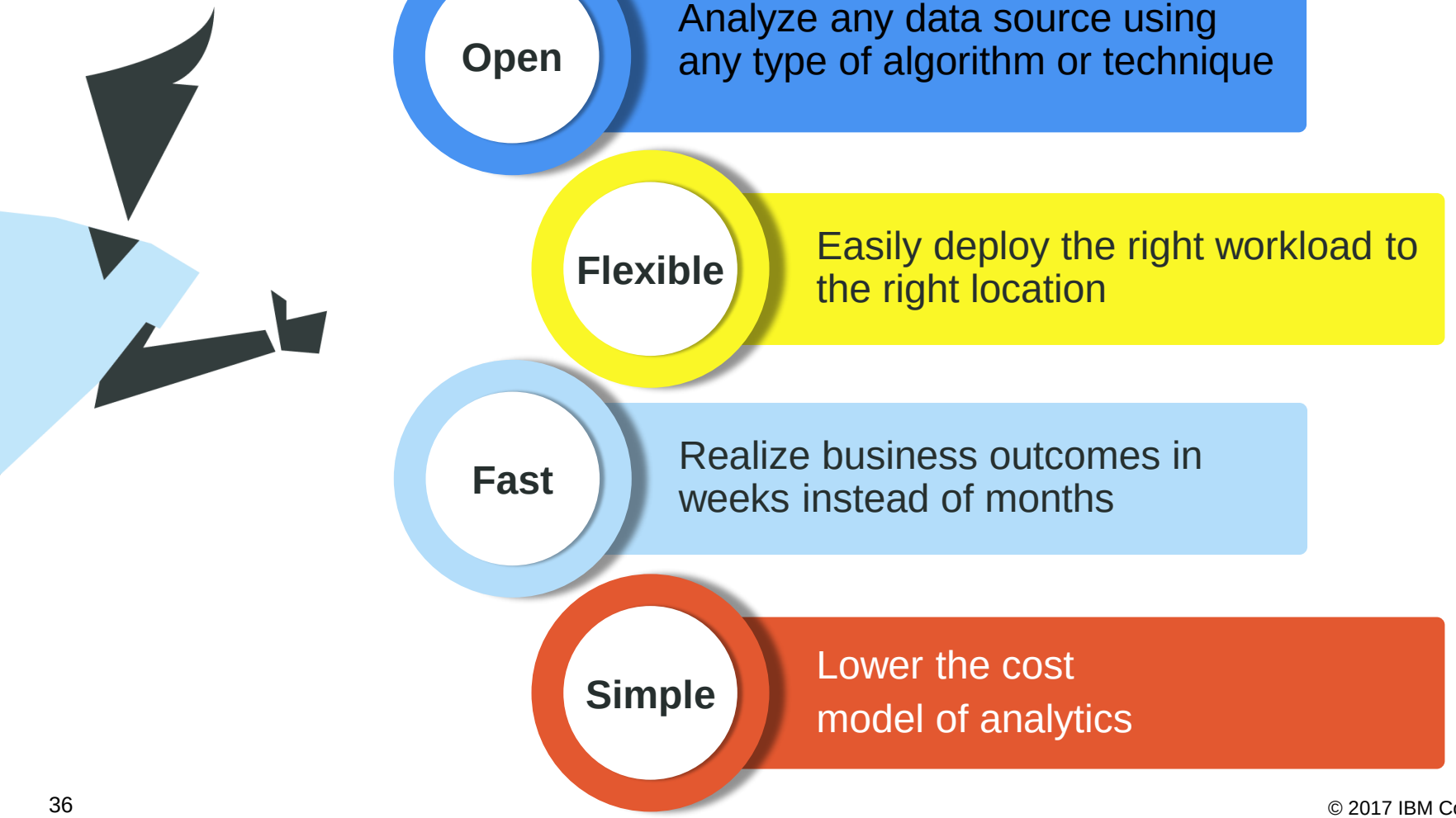
Benefits of dashDB Technology with Fast Deployment into Private Cloud Environment



- Highly flexible data warehouse
- Optimized for fast and flexible deployment into **private or virtual private clouds**
- Uses **Docker** container technology
- Built on top of **dashDB technology**, it shares the benefits of
  - BLU Acceleration in-memory columnar technology
  - Netezza In-database Analytics
  - Oracle Compatibility
- **Massively Parallel Processing** (MPP) with automated scaling capabilities to increase infrastructure efficiency

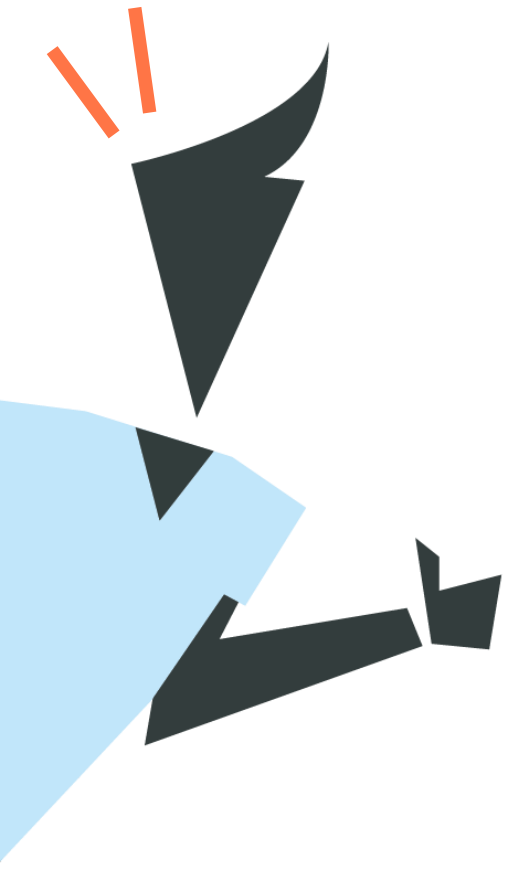
dashDB Local

Key Differentiators



dashDB Local

Addressing many needs



"**I need** more data warehouse capacity. I like the advantage of using SDE to achieve elasticity to continually meet service levels and maximize use of existing resources such as commodity hardware."

"**I need** to automate the provisioning of the right combinations of data and analytic services or solutions by user."

"**I need** a warehouse/data mart warehouse/data mart that is easy to deploy and requires little tuning or management."

"**I need** a cloud strategy, but need data to stay more directly under business control or on-premises due to internal requirements and other mandates."

"**I need** a cost-effective, high-performance processing engine to gain deeper insights from massive amounts of data being generated from mobile, web, and IoT applications."

"**I don't want** to spend millions re-writing applications to work with Hadoop, especially when working with structured data, and commodity hardware."

Use Case Examples

Build new or replace current data warehouse production environment

- Market-leading performance and flexible resource allocation in private cloud environments offer attractive price/performance and TCO
- Simple compute-oriented pricing model facilitates hosted or on-premises IaaS and bare metal server deployments
- Clients who want to start small and grow on-demand
- As a stepping stone in the journey towards IBM-managed cloud services

Augment your warehouse footprint

- Dynamic, cost-effective development and test environments
- Client-managed backup and DR strategy
- Dynamic analytic side car for self-service analytics and independent analytics
- Enable self-service analytic services for data scientists powered by Spark

Store your older / colder structured SQL data

- Use for a logical data warehouse solution instead of Hadoop
- Benefit from a native SQL data warehouse engine (performance, in-database analytics)
- Unified working behavior, look & feel, apply same skills
- Fluid Query enables integration and query routing in both directions

dashDB Local with an Unmatched Value Proposition!

Simple

Deploy in minutes
Load & go
No expensive tuning

Fast

World-class in-memory
columnar engine
Operates on compressed data

Flexible

Automatic orchestrated configuration
and adaption to infrastructure
Elastic Scale up/down
Scale out/in based on business needs

Agile

Runs wherever docker runs.
On-premise or hosted IaaS
Move applications as needed

Smart

Powered by Apache Spark
Analyze data in-database
Run Spark MPP apps

Very Secure

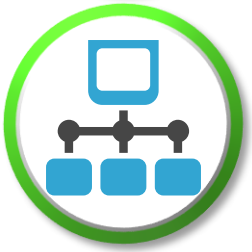
Embedded authentication
Closed container for restricted access
UI console

Other vendors offer NO or ONLY partial software defined environment (SDE) capabilities!

IBM dashDB Local and Fluid Query: Self Service Integration

Discovery & Exploration

- Land data in Hadoop for discovery, exploration & “day 0” archive
- Queries can access data across dashDB Local, Hadoop and other database sources in your logical data warehouse



Build bridges to RDBMS islands

- Combine data from different enterprise divisions currently trapped in separate database implementations
- Access structured data from familiar sources like Oracle, DB2, SQL Server and dashDB Local

Queryable Archive

- Query historical data on Hadoop with Big SQL or from dashDB Local
- Combine Hadoop data in IBM Big SQL, Hive, Impala or Spark SQL with other data sources

Data Warehouse Capacity Relief and Disaster Recovery

- Offload cold data to relieve resources on the primary data warehouse
- Copy data to Hadoop as a disaster recovery solution (immutable backup or compressed read) or Active-Passive dashDB Local

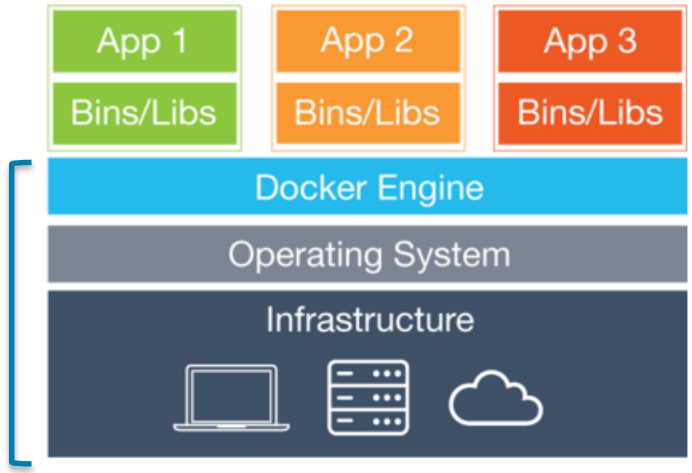
What is Docker?

- **Docker Inc. is the original author and primary sponsor of the Docker open source project**
  - Started as an internal project within dotCloud
  - Released as open source project in March 2013
  - Quickly gained traction from open source community, as well as attention from venture capitalists and technology partners (including IBM)
- **Docker is an open platform for developing, shipping, and running applications**
  - Solves dependency chaos in application development and reduces development life cycle
- **Docker brings container technology to the masses**
  - Many companies have been using container technology
  - Docker provides tooling and platform to facilitate adoption

How does Docker Work?

Applications encapsulated in containers

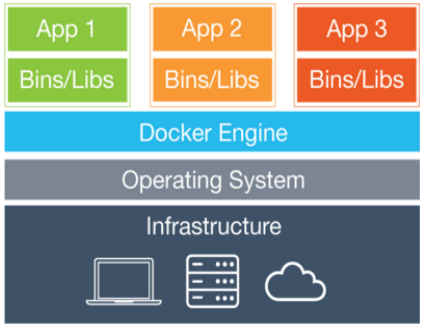
Host operating system + Infrastructure



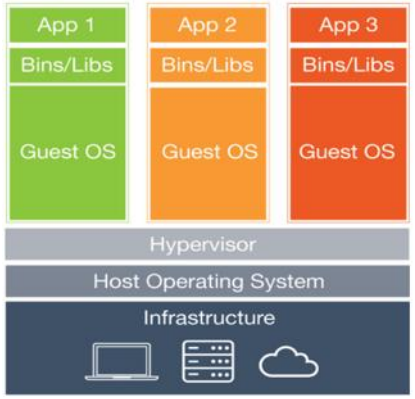
- Docker separates applications from infrastructure using “containers”
- Containers = software + complete filesystem associated with software
  - everything software needs to run: code, runtime, system tools, system libraries
- 1 Host runs multiple guest instances
  - Each guest is a container with its own root file system, processes, memory, devices and network ports

Docker vs VMware (Virtual Machine)

Docker



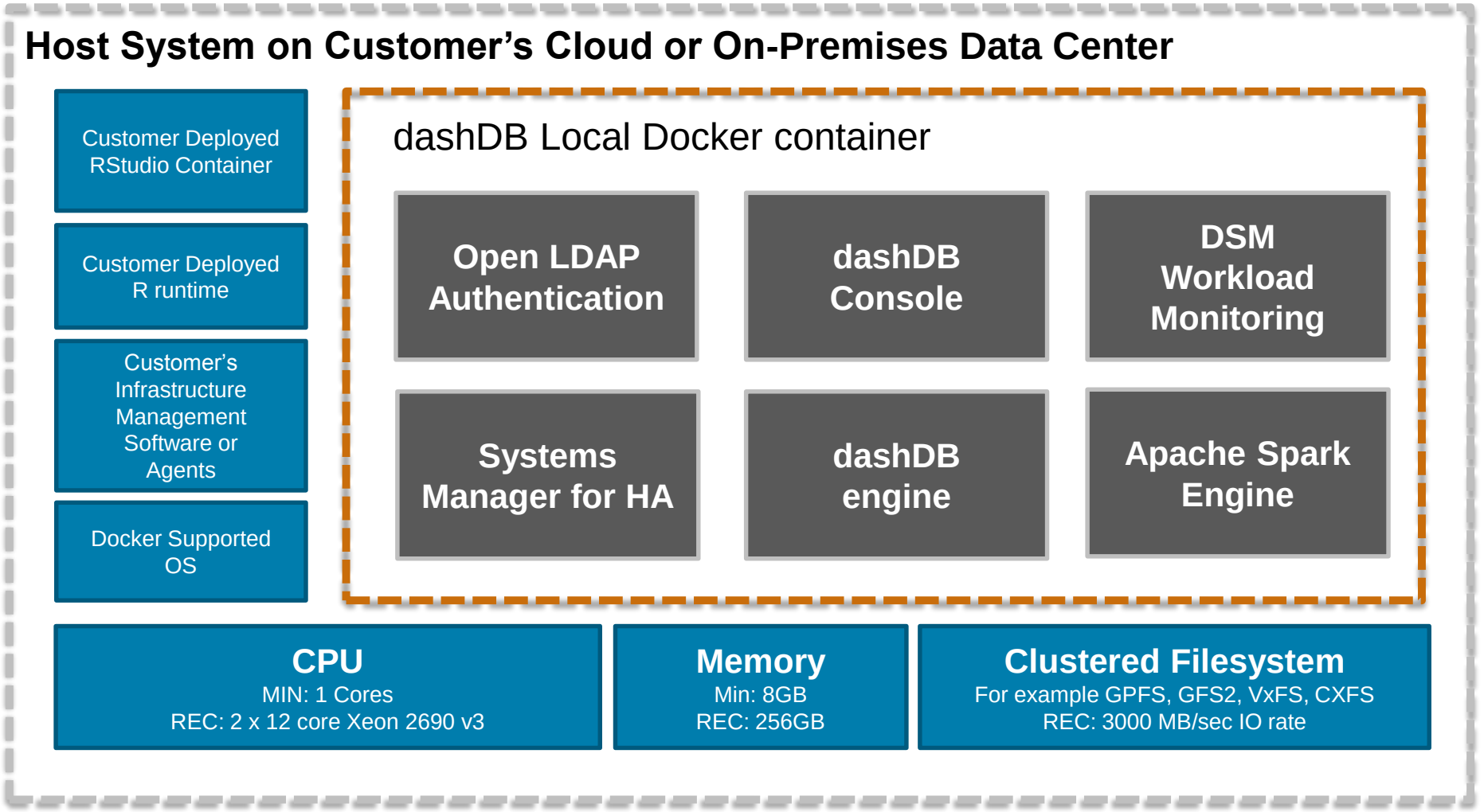
Virtual Machines



Docker	Virtual Machines
Based on Container technology	Virtualization Technology (hypervisor)
Provides OS-Level process isolation	Provides hardware virtualization
Containers include the application and all of its dependencies • Host kernel is shared amongst Docker containers • Run as an isolated process in userspace on the host operating system	Each virtual machine includes the application, the necessary binaries and libraries and an entire guest operating system
Advantageous for packaging and shipping • Easier and more lightweight to deploy and faster to start up than virtual machines	Best for native execution to share and manage hardware, allowing multiple different environments, isolated from each other, to be executed on the same physical machine

Both technologies can be used in conjunction

dashDB Local: Architecture



Scale Up CPU, Memory, IO according to your infrastructure availability

dashDB Local Container Stack

Application container is consistent and "**stateless**" from any changes

Container Stack

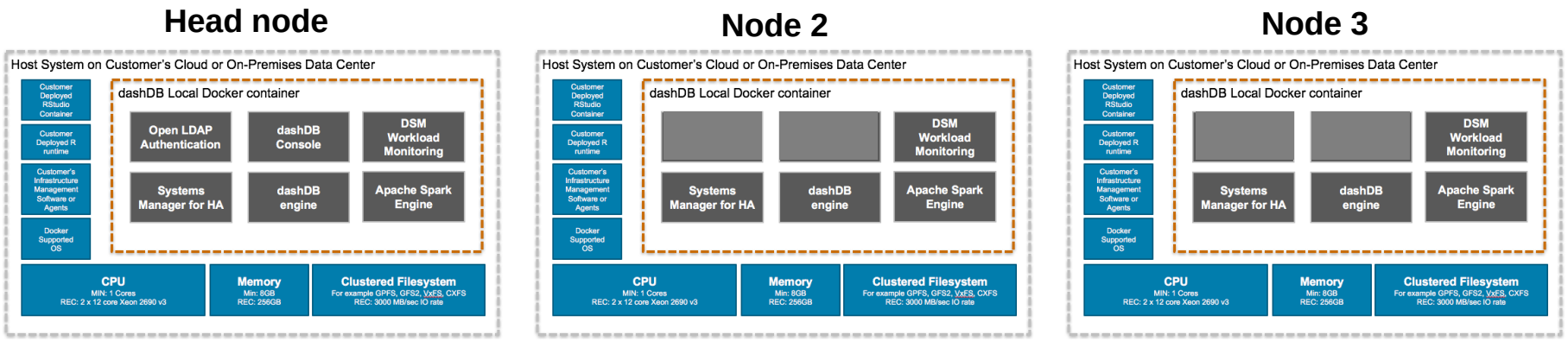
- **CentOS 7.2 - Consistent with current dashDB host OS level**
- **dashDB engine 11**
- **LDAP authentication**
- **DS Server (Web server) and dashDB Console**
- **System Manager for HA**
- **Workload Management/Monitoring via enterprise DSM**

Not included (different from dashDB managed services):

- R Studio: user install instructions provided
- Guardium

dashDB Local Scale Out on Multiple Host Systems According to Your Infrastructure Availability

- One container per host
- Host can be bare metal or virtual
- 1 head node
- Minimum 3 nodes, maximum 24 nodes
- LDAP server and Console is active on one node at any given time



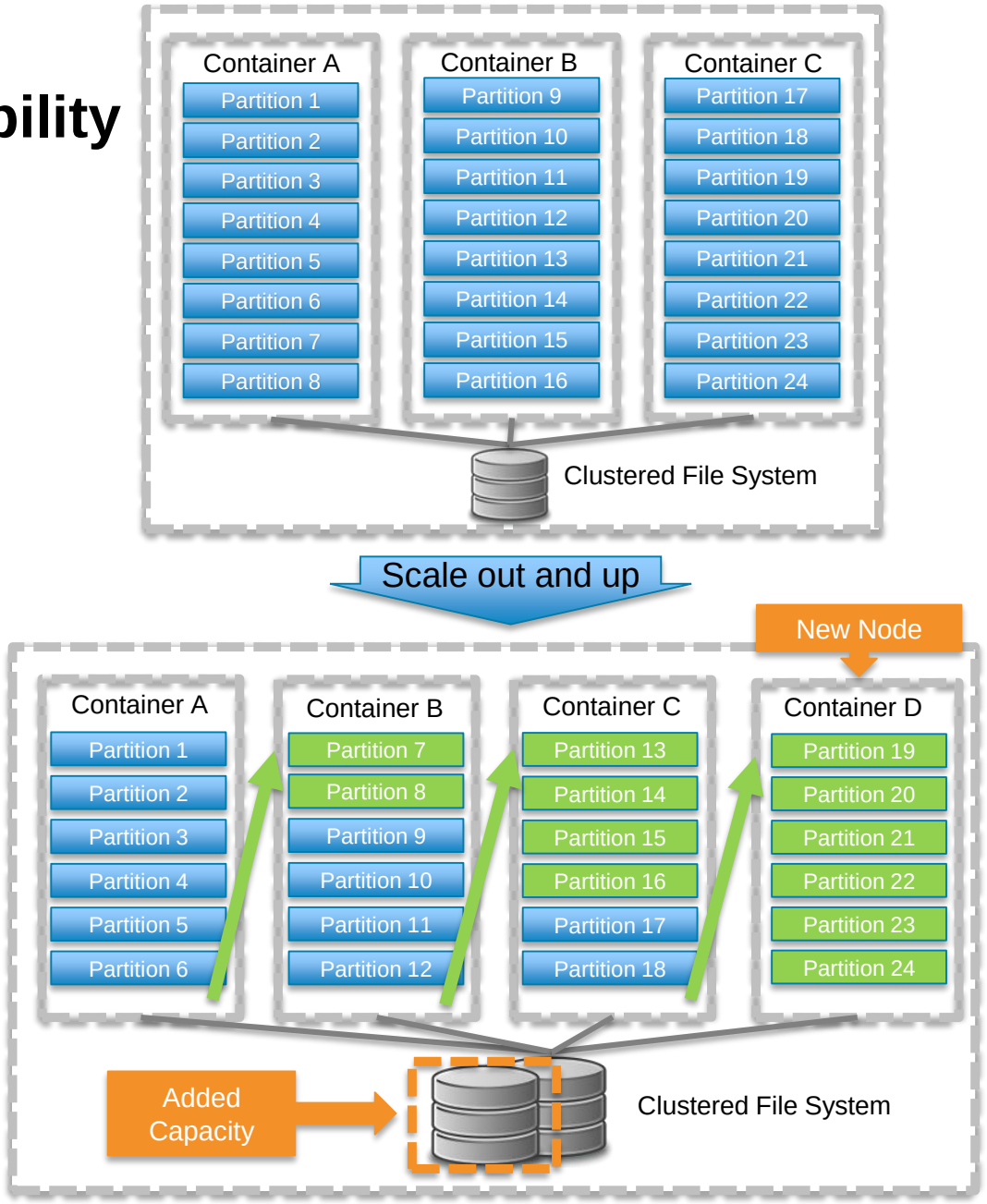
dashDB Local – MPP Elastic Scalability

- **Automatic detection of hardware resources for scaling operations**
  - Docker container automatically detects hardware resources available
  - Minimal interruption to your database operation - requires temporarily stopping dashDB Local services while you implement the changes

- **Scale in or scale out: removing existing or adding new nodes**
  - Partitions will be automatically re-balanced across all containers
- **Scale up or scale down: change the resources available to a node**
  - Easy because independence between application container and named volume storage container
- **Portability: Replacing the node and clustered file system**
  - Stop container on old server, copy data to new file system, start container on new server

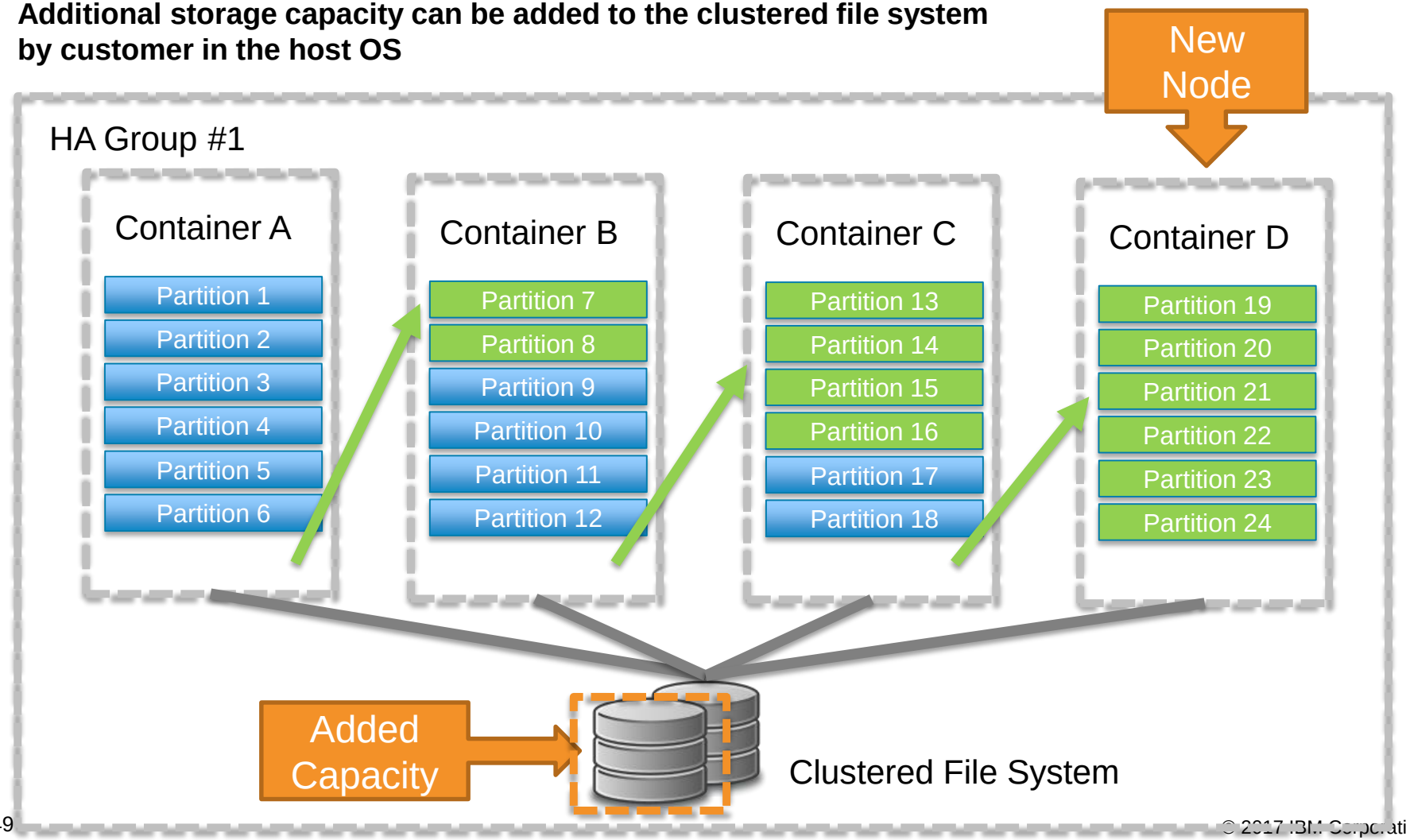
dashDB Local:
SMP + MPP Elastic Scalability

- **Scale up or scale down:**
  - Adjust CPU + Memory resources available to a node
- **Scale in or scale out:**
  - Add or remove nodes as needed
  - Partitions automatically re-balanced across containers
  - No data redistribution required



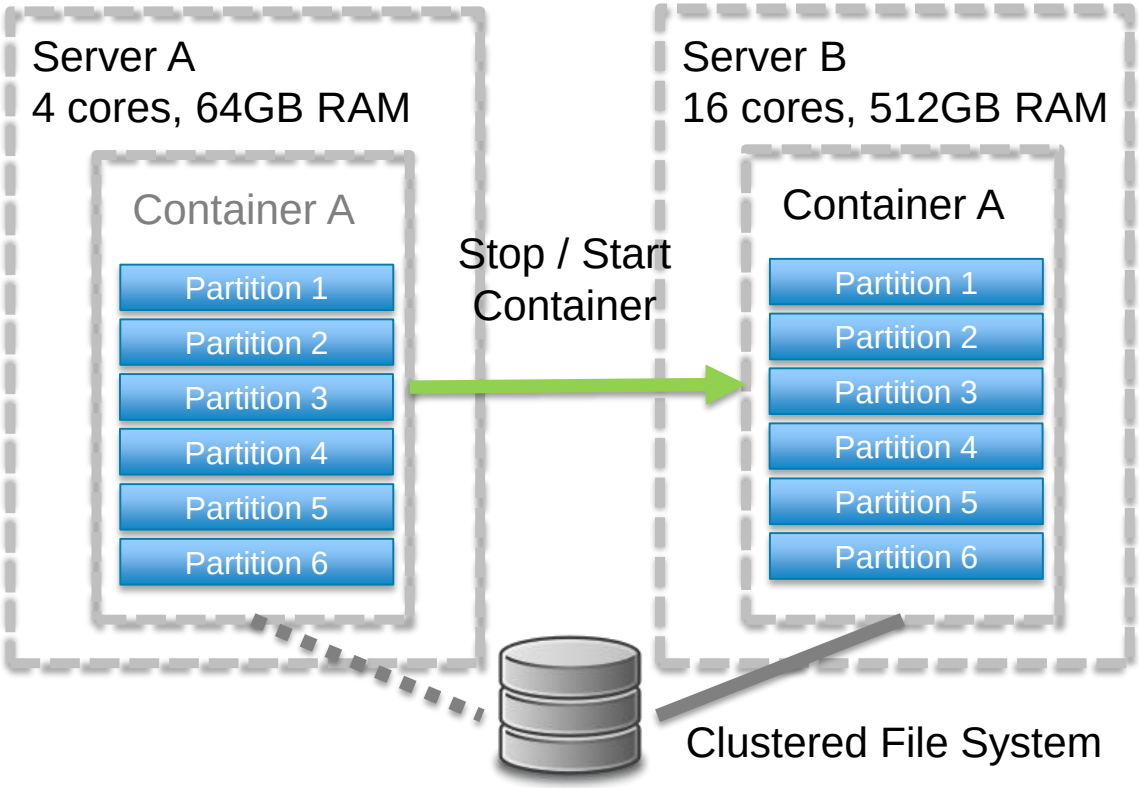
Elasticity: Scale Out

- **For expansion, a new node will be added to an HA group and the partitions will be re-balanced across all the containers**
- **Additional storage capacity can be added to the clustered file system by customer in the host OS**



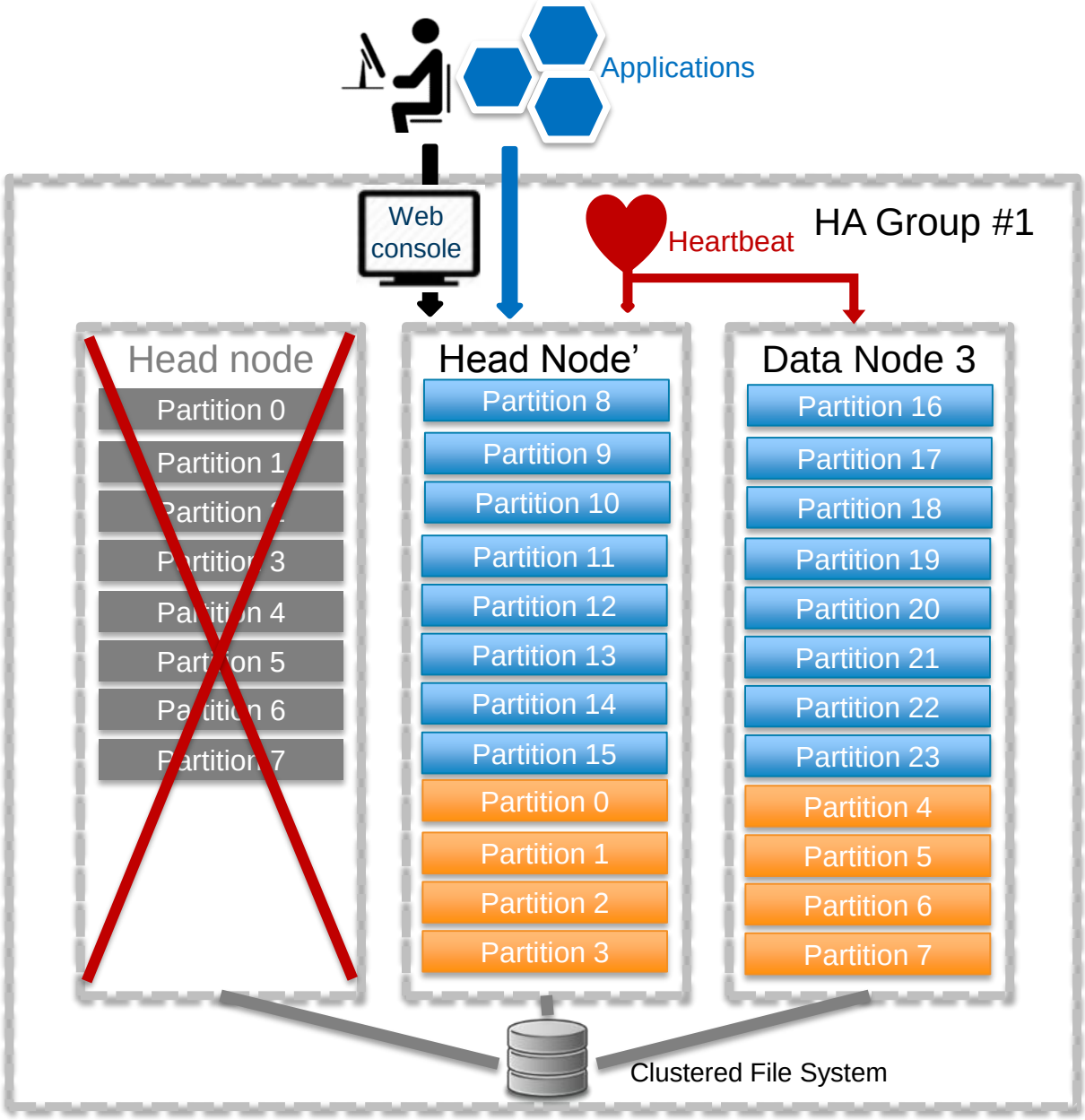
Elasticity: Scale Up is as Easy as a Container Restart

- **Scale-up is easy due to independence between container and named volume storage**
- **Scenario 1: Replacing the server completely (Server A → Server B)**
  - Stop container on old server
  - Re-create dashDB Local container on new server (we will transparently re-link the stateless container to stateful data on the named volume)
- **Scenario 2: Increasing the resources on the same server**
  - Simply restart of the container



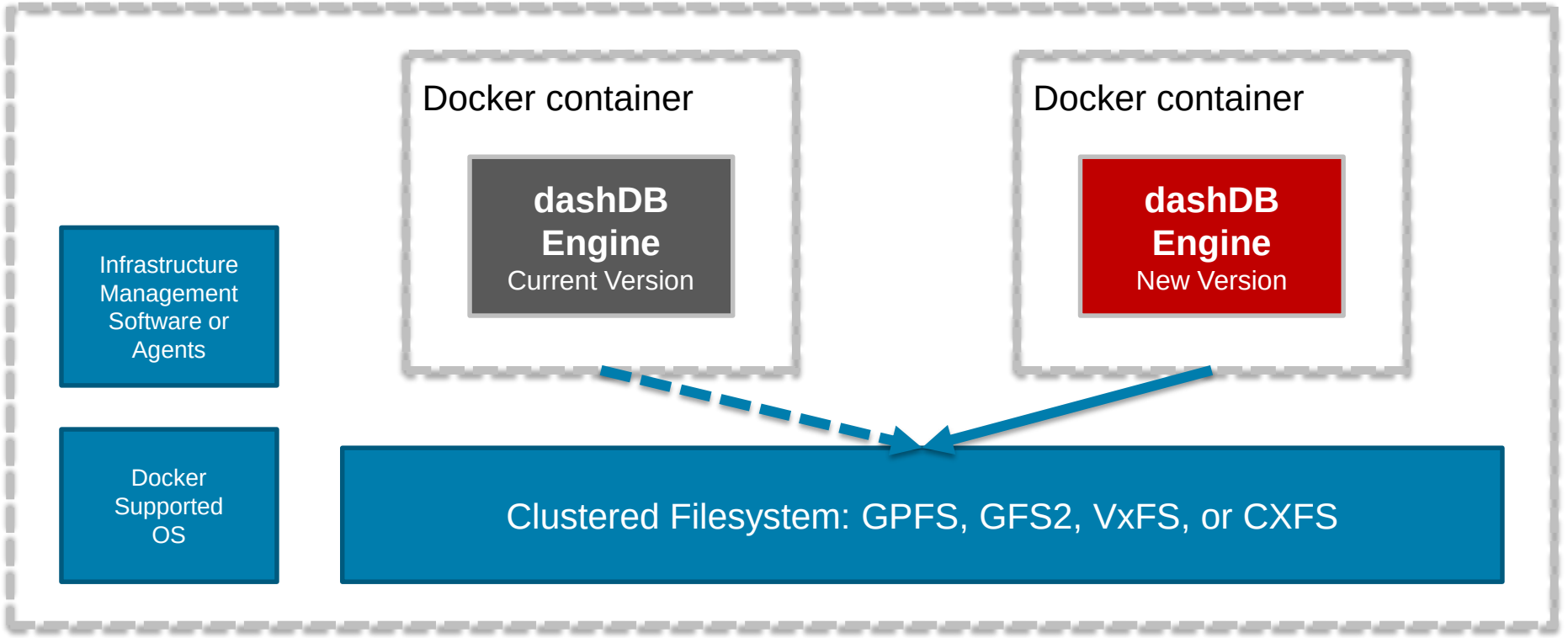
Node Failover

- **If a container in an HA group has failed, the partitions will be re-balanced among the remaining running containers**
- **If a head node failure occurs, a new head node will be selected from the active data nodes**
  - Web console is restarted
  - Application needs to be re-routed



dashDB Local: Easier Upgrade

- **Runtime and Storage are Separate – makes software upgrades simpler**
- **Container includes updates needed**



dashDB Local – Access to Docker Container

Console UI and database are to be accessed through given endpoints

- **Web Console UI access through URL via HTTPS**
  - Typically at https://<IP_of_dashDBLocal>:8443
- **Database accessed through JDBC / ODBC via database client**
  - Can do most normal database administrative operations
    - Example: create table, load data, run SQL, monitor workload
  - CLPPlus remote client access available
    - Host side CLPPlus immediately available via `docker exec –it dashDB clpplus`
- **Wrappers for command line nz_* tools available through `docker exec`**
  - Generically branded as db_* tools

Support for dashDB Local

- **IBM supports “dashDB Local container” only**
  - IBM Support will help triage dashDB Local issues
  - If the issue is the Docker engine or the host OS, they would be redirected

- **Customer is responsible for acquiring the level of support they need for Docker and host OS**
  - Customers can leverage Docker CS engine (from Docker) or Open source docker RPMs that come with the Linux distros (such as Red Hat)
  - Docker will provide commercial support only for Docker CS engine and not for Docker RPMS and it’s Customer choice to embrace either path.

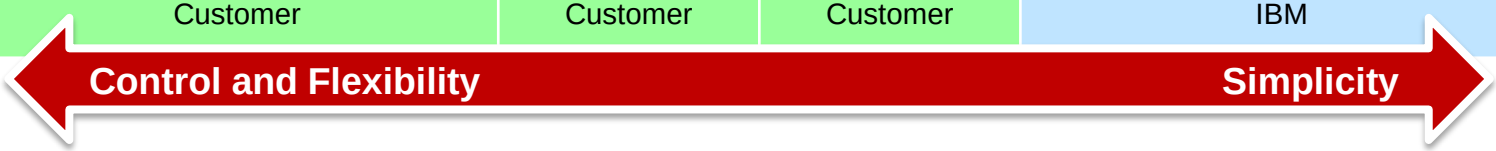
- **dashDB Local Support follow the traditional call-in, PMR support process that we have for our on-premises software**

dashDB Local
Minimum and Recommended System Requirements

SYSTEM REQUIREMENTS	MINIMUM Single Host	MINIMUM Multi-Host (MPP)	RECOMMENDED 4TB raw data per Host
Host Operating System	Docker supported OS (Kernel 3.10+)	Docker supported OS (Kernel 3.10+)	CentOS 7 / RHEL 7
Processor	1x 2.0 GHz core	8x 2.0 GHz core	2 x 2.6 GHz 12-core Xeon 2690v3 CPU
RAM	8GB	64GB	256GB
Storage mounted on /mnt/bludata0	20GB	200GB	Network attached (SAN) 2000GB SSD @ 3000MB/sec IO
Docker Version	1.11.1	1.11.1	1.11.1

DB2 and dashDB – Who Does What?

Activity	DB2 Software (on-premises)	DB2 Software (BYOL on IaaS)	DB2 on Cloud	dashDB Local	dashDB for Analytics	dashDB for Transactions
Provision hardware infrastructure	Customer	IaaS Vendor	IBM	Customer or IaaS Vendor	IBM	
Manage hardware infrastructure	Customer	IaaS Vendor	IBM	Customer or IaaS Vendor	IBM	
Database manger software installation	Customer		IBM	IBM *	IBM	
Database manager instance creation	Customer		IBM	IBM *	IBM	
Database creation	Customer		Customer	IBM *	IBM	
Database configuration	Customer		Customer	IBM *	IBM	
Manage database environment	Customer		Customer	Customer	IBM	
Database manager software maintenance	Customer		Customer	IBM *	IBM	
OS maintenance	Customer		Customer	Customer	IBM	
Setup encryption	Customer		Customer	IBM	IBM	
Database backup and restore	Customer		Customer	Customer	IBM	



* For initial setup or subsequent updates of dashDB Local, customer simply deploys an IBM-provided Docker image with all software (including the database environment) pre-configured

IBM Bluemix Data Connect

A fully managed **data preparation** and **integration** service

Access data wherever it resides
Combine data from multiple sources
Transform data and make it fit for use

Self-Service Access from Anywhere, Anytime
Making Data Simple for Everyone

Business Analysts

Can find and use the data they need to accelerate data based business decisions using timely accurate and trusted information

Data Engineers

Can easily build data pipelines and workflows to transform and move data without the need for deep ETL development skills

Developers

Can easily develop data-rich applications by embedding the Data Connect REST API into new or existing applications

IBM Bluemix Data Connect: Data Access

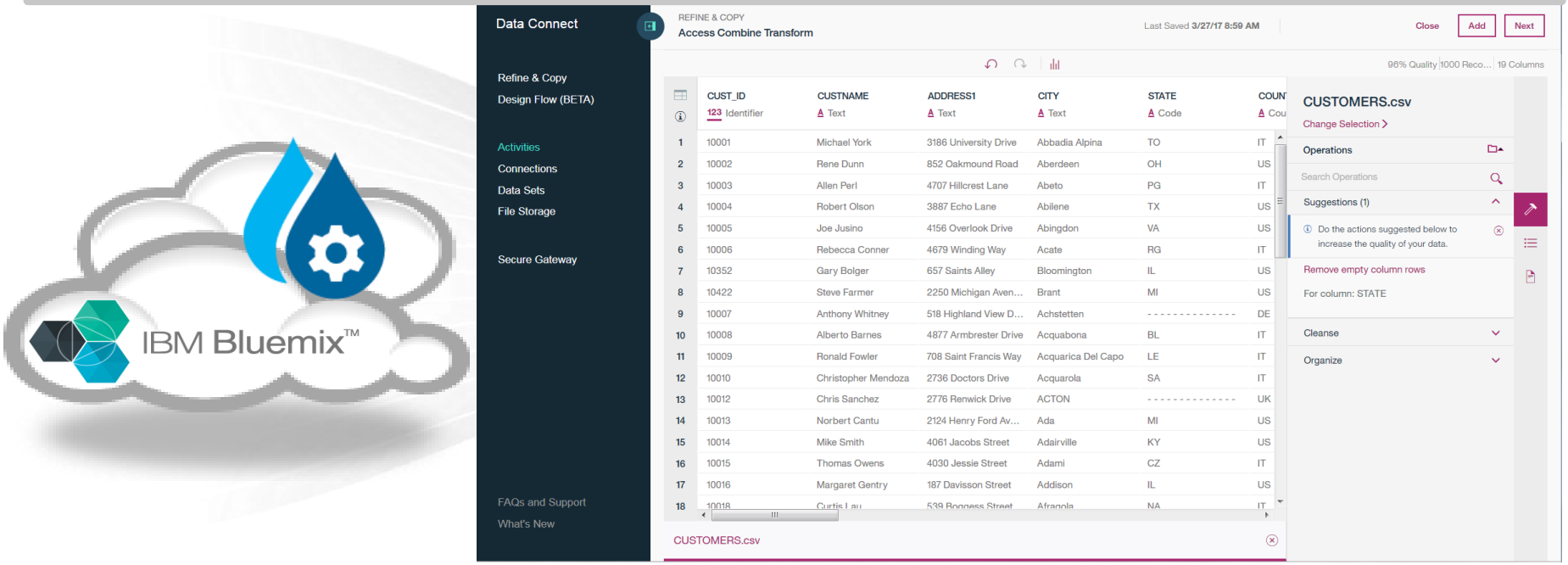
Cloud Source	On-Premises Source	Cloud Target	On-Premises Target
Amazon Redshift	Apache Hive	Amazon S3	Files [Desktop]
Amazon S3	Cloudera Impala	Bluemix Object Storage	Hortonworks HDFS
Apache Hive	Files [Desktop]	Drop Box	IBM BigInsights™
BigSQL	Hortonworks HDFS	IBM Cloudant™ [2]	IBM DB2® LUW, z/OS, iSeries
Bluemix Object Storage	IBM BigInsights™	IBM dashDB	IBM dashDB Local
Drop Box	IBM DB2® LUW, z/OS, iSeries, BigSQL	IBM BigInsights™ on Cloud	IBM Pure Data for Analytics®
IBM BigInsights™ on Cloud	IBM dashDB Local	IBM DB2® on Cloud	Microsoft SQL Server
IBM Cloudant™ [1]	IBM Informix®	IBM Watson™ Analytics	MySQL Enterprise & Community
IBM dashDB	IBM Pure Data for Analytics®	Microsoft Azure	Oracle
IBM DB2® on Cloud	Microsoft SQL Server	PostgreSQL on Compose	Teradata
Microsoft Azure	MySQL Enterprise & Community	SoftLayer Object Storage	
PostgreSQL on Compose	Oracle		
Salesforce	Pivotal Greenplum		
SoftLayer Object Storage	PostgreSQL		
	Sybase & Sybase IQ		
	Teradata		

Data Source Connectivity Matrix

All of the supported targets are compatible with each source – Some version restrictions apply

IBM Bluemix Data Connect

Simple and **intuitive** self-service **data preparation** for business users to access, combine, cleanse and transform data with ease and confidence

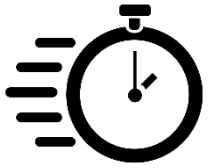


- Access data from multiple relevant cloud or on-premises data sources
- Automatically analyzes, profiles and classifies the data upon ingestion
- Delivers quality scores and metrics to understand and trust the content
- Easily refine, enrich and cleanse the data using a robust catalog of transformation operations
- Save, edit, run and schedule activities to deliver the data to a diverse set of targets

IBM Bluemix Lift

A fully managed **ground-to-cloud** data migration service

IBM Bluemix Lift makes it easy to **quickly, securely and reliably** migrate your database from private on-premises data centers to an IBM cloud data property



Fast

Migrate quickly to the cloud

Embedded with IBM Aspera, Lift makes over-the-wire data transfer blazing fast. It allows you to begin migrating your applications quickly to the cloud sooner than before.

Nonintrusive

Zero downtime migrations

Lift captures changes to your source database and replays them to your target database so applications using the source DB can run uninterrupted.

Secure

Secure data movement

Lift protects your sensitive and proprietary data when you are moving it across the Internet by establishing a secure, encrypted connection, end to end.

IBM Bluemix Lift – Capabilities

- **One-time & incremental migration for Netezza or CSV file set to dashDB**
  - **Ultra high speed data movement into dashDB using IBM Aspera** with option to slow down for bandwidth-constrained networks
  - **Zero-downtime for source database** during migration with built-in change data capture
  - **Schema compatibility** assessment and migration
  - Support for dashDB Entry and Enterprise plans
- **One-time migration for DB2 LUW or CSV file set to DB2 on Cloud**
- **Test with sampling of data:** Use a sample of your data set to test target schema
- **Security from top to bottom:** Lift encrypts your data as it travels over-the-wire
- **Self healing:** Lift gracefully handles services upgrades and unexpected outages
- **Data load API:** Manage data migration activities via a published API
- **Blazing fast end-to-end data migration speeds**
  - Observed cross-country Lift migration rate with **Netezza to dashDB: 171 GB/hour**
  - Observed cross-country Lift migration rate with **DB2 to DB2 on Cloud: 27 GB/hour**
  - Typical end-to-end data migration rates seen with **Competitor: 5-25 GB/hour**

Current Lift Migration Sources and Targets

Where we migrate *to*

dashDB
data warehousing
and analytics

DB2 on Cloud
relational database

Where we migrate *from*

IBM Netezza

IBM DB2

CSV/Flat Files

Oracle, SQL Server
and other sources use this
option to get data in

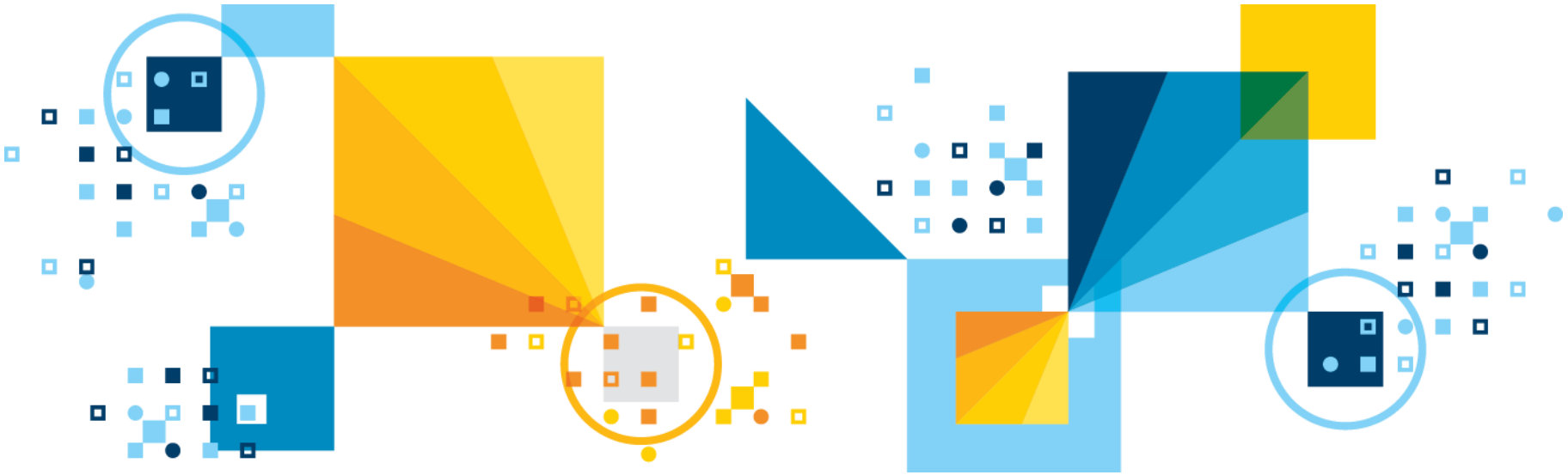
IBM Bluemix Lift

Fits neatly into our ground-to-cloud, data movement portfolio of offerings

25 TB and smaller	25 TB to 100 TB	100 TB and beyond
IBM Bluemix Lift	**IBM Data Transfer Service**	**IBM Mass Data Migration Service**
Available today	Available today	1H 2017

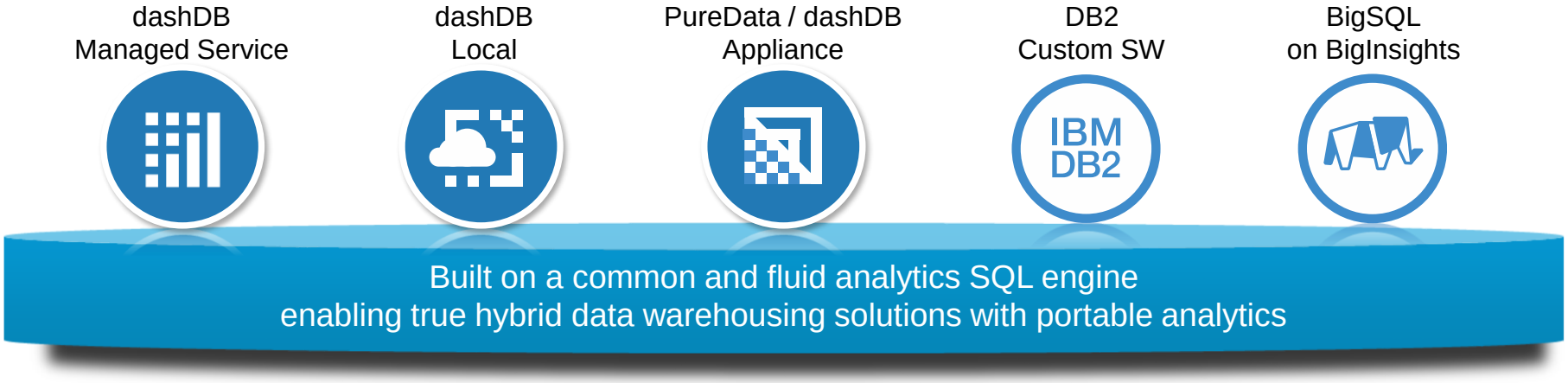
IBM Lab Services technical consultants available to help with data on boarding

Hybrid Data Warehousing



IBM Hybrid Data Warehouse Strategy

Offer clients **choice** in selecting the best (combination of) data stores to satisfy hybrid data warehouse solution needs



Data Virtualization

Common Fluid Query capabilities for query federation and data movement

Operational compatibility

Reuse operational and housekeeping procedures

Application compatibility

Write once, run anywhere

Licensing

Flexible entitlements for business agility & cost-optimization

Next-gen analytics

Common programming model for in-DB analytics

Ecosystem

One ISV product certification for all platforms

Cloud Data Warehouse Use Cases

Standalone, high performance data warehouse in the cloud

USE CASE	SCENARIO
Quickly build data driven 'born in the cloud' applications	Mobile, web, IoT data captured in Cloudant JSON documents are immediately transformed into structured data managed in dashDB for subsequent analysis.
Consolidate and integrate data silos and data marts	Move on-premises data warehouse/data marts and multi vendor cloud databases into a cost-effective, high performing cloud data warehouse.
Quickly build self service or SaaS applications	Deliver new, self service analytic solutions and applications that take advantage of massive amounts of data being generated on the Internet.

Hybrid Cloud Data Warehouse Use Cases

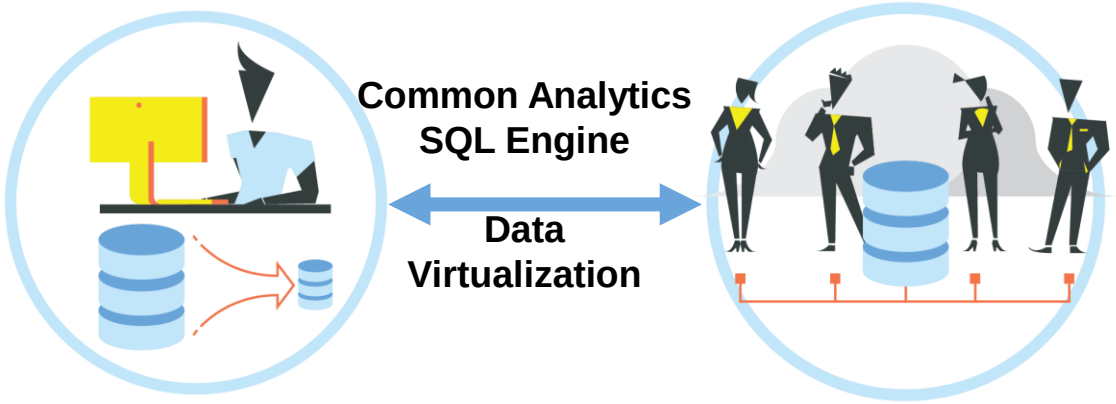
Extend on-premises data warehouse environment with cloud-like flexibility

USE CASE	SCENARIO
Dev/Test Ecosystem	Quickly and easily test new applications and data sources before production implementation
Accelerate LOB Analytic Projects	Quickly stand up a integrated data mart service that can meet breadth of data source and analytic techniques such as IoT data sources and Spark Analytics
Free up on-premises data warehouse capacity	Migrate a subset of applications/data running on-premises DW to the cloud with the option to remain hybrid or fully migrate, that is, dashboards/reporting, queryable archive

A Hybrid Data Warehousing Example

IBM PureData System for Analytics (PDA) integrates with dashDB for hybrid data warehousing

PDA on-premises data warehouse
(structured, defined set of data, traditional BI)

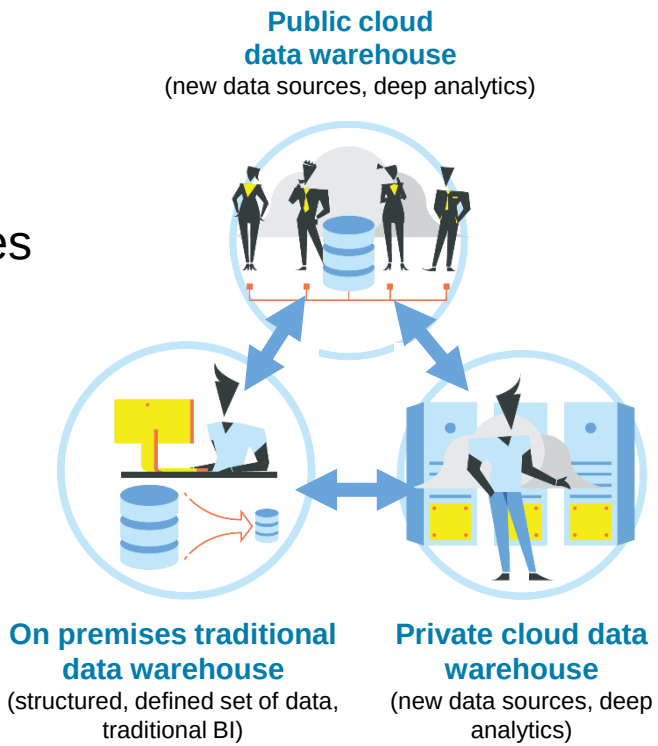


dashDB for Analytics public cloud data warehouse
(new data sources, new analytics)

- IBM Fluid Query provides data virtualization through federated queries
- PDA users now have visibility to new IoT data sets captured in dashDB
- IT maintains governance and control of which data they expose to which PDA users

Hybrid Data Warehouse Benefits

- Enables a controlled journey to the cloud and empowers users with self service access to combinations of data sets and analytic tools
- Offers flexibility in choosing the data warehouse form factor (combinations) that best suit the business needs
- Accelerates time to value:
  - Start small and move fast
  - Enable course corrections without penalties
- Bridges data stores for seamless data integration or application movement
- Leverage existing investments through a write once – run anywhere application paradigm



Try dashDB out for free!

dashDB for Analytics
http://ibm.biz/dashDB_Analytics

Free Entry Plan

dashDB for Transactions
http://ibm.biz/dashDB_TX

7 Days Free

dashDB Local
http://ibm.biz/dashDBLocal

Free Trial

IBM Bluemix

https://console.ng.bluemix.net/

70