



# HRG Assessment:

## IBM dashDB

IBM dashDB is a family of cloud based data warehouse and database solutions for analytics and transaction processing. IBM's dashDB produces data-based insights using in-memory and columnar processing as well as in-database analytics technologies that include the R statistical language, the RStudio development environment, the Apache Spark analytics operating environment, and Netezza's in-database analytics and analytics library. Customers can run "What-if" scenarios and ad hoc queries to examine trends, provide analysis, and gain competitive advantage. There are several different editions of dashDB: dashDB for Analytics, dashDB Local, and dashDB for Transactions. IBM dashDB for Analytics is a fully managed data warehousing public cloud service, managed and maintained by IBM, ensuring that customers are always at the most current release levels. IBM dashDB Local is a customer-managed data warehousing private cloud service which runs on a customer's infrastructure, or on a hosted service like SoftLayer, AWS, or Azure. IBM dashDB Local gives customers data warehouse deployment control and flexibility. IBM dashDB for Transactions is a fully managed operational database public cloud service that is not covered in this Assessment.

### The Challenge

The world is increasingly instrumented with chips and sensors that track and measure transactions and interactions resulting in continually increasing volumes of data. The business environment supported by IT is evolving so rapidly that it is a requirement to deliver an infrastructure that is fluid, highly scalable, and able to support applications that are rapidly developed and assimilated. The challenge is to deliver systems and services to support end to end supply chain interactions, business intelligence, and analytics at a reduced cost, with better customer satisfaction, while growing market share and stakeholder value.

The ability to customize offers as a transaction or customer engagement is underway is becoming a business requirement. Next generation customer experience applications require near real-time integration and analysis of distributed and heterogeneous data sources (including structured, unstructured, poorly structured, high volume and high velocity data). The ability to manage and analyze this data and transform it into actionable information is the key to achieving the true benefits of Cloud Computing, Big Data and the Internet of Everything.

## Transforming the Organization

Big Data, Social, and sentiment analysis are transforming how customers use data (structured and unstructured). The key to successful business evolution is having choice and flexibility in terms of identifying what data is useful and then harvesting that data without the cost and complexity that previous generation “so called” simplified methods have imposed. Using data to influence the outcomes of transactional engagements between customers and businesses as they are occurring is evolutionary. Data made useful through real time analytics during a customer engagement facilitates upselling through the introduction of highly relevant add-on offers. Most data used to enhance the customer experience on mobile and web based platforms is unstructured or poorly structured data from sources such as Facebook, Twitter, LinkedIn, Pinterest and Instagram.

Today, no data, regardless of type or location, is beyond reach and integration. Users are no longer limited by the physical or virtual location of data, which application it comes from, or even if the data is structured or unstructured. Businesses require operational analysis and decision-making based on near real time data flows from across the organization and beyond, collected in a wide range of formats. The ability to integrate and analyze active real-time operational data from transactional systems, historical data from traditional data warehouses and data marts, and unstructured high velocity data is a requirement for organizations to meet constantly evolving challenges and opportunities.

Applications (initially, the automation of business logic) must undergo continual change for businesses to evolve. HRG believes that over time applications and analytic tools will merge and become the next generation of applications or insight systems that will be able to change business processes through their discoveries in real-time.

### dashDB

IBM dashDB consists of a family of pre-configured data warehouse solutions available as either a fully managed service on a public cloud or as a client-managed service in a private cloud, for use on a customer’s existing IT infrastructure. By using both dashDB managed service and dashDB Local, customers can easily move data and workloads between them. Flexibility in hosting and the ability to dynamically move data and workloads provides value for those dashDB Local customers with remote field offices or remote researchers, as well as for those departmental users with well understood functional requirements such as preproduction engineering prototyping and concept testing.

The dashDB managed service and dashDB Local share a common look and feel and a common code base that incorporates key Netezza in-database analytics capabilities and analytic libraries which are being migrated to the new dashDB code base. IBM dashDB and dashDB Local provide on-demand data warehousing as well as in-memory and columnar processing derived from IBM DB2 with BLU Acceleration.

Both dashDB and dashDB Local provide in-memory column store technology and an MPP-aware query engine, query planning, query optimization, dynamic in-memory processing on compressed columnar data, parallel vector processing, and data skipping (derived from IBM DB2 with BLU Acceleration) and are optimized for highly parallelized multi-core systems. The MPP feature for large data sets lets

customers rapidly scale from a single node system to an enterprise-class MPP data warehouse cluster, and then scale back to a single SMP node once a period of peak demand has passed. The dashDB analytics engine offers SQL compatibility with DB2, Netezza, and Oracle for ease of migration, along with generic programming support using ODBC, JDBC, SQL PL and PL/SQL.

## dashDB for Analytics (managed service)

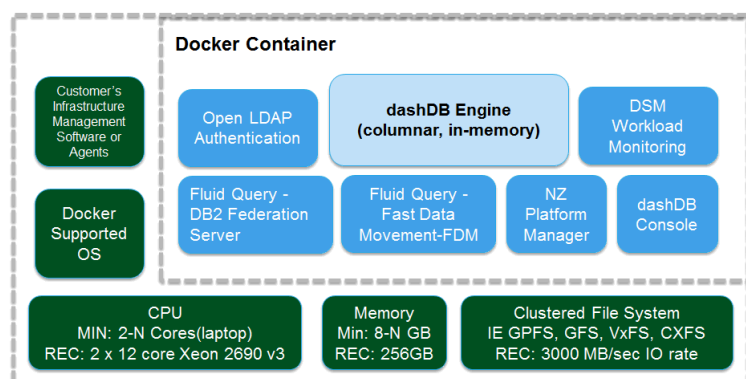
The dashDB for Analytics managed service is an on demand data warehouse service hosted on a public cloud, available via IBM Bluemix running on IBM SoftLayer, also available on Amazon Web Services (AWS). The dashDB managed cloud service integrates with IBM Cloudant and other IBM cloud services available through Bluemix, enabling the quick stitching together of cloud based solutions. For example, deep data analysis is enabled using dashDB with integrated tools such as Watson, Cognos and SPSS. Integration with Aspera helps clients to rapidly load data over a WAN. Last, IBM DataWorks Lift quickly moves data up to the cloud from traditional data center IT sources for analysis using dashDB's in-database analytics. The IBM dashDB managed service for Analytics can be used as a data source for federated queries using Fluid Query which is available with IBM PureData System for Analytics and Netezza, its predecessor. *(For more on Fluid Query, please see the section below on dashDB Local.)*

Customers using the dashDB managed service can access JSON data, typically web, mobile or Internet of Things high velocity data, through IBM Cloudant by applying automatic schema discovery and data mapping, and then loading and synchronizing JSON data imported from Cloudant directly into dashDB. In this manner data from sensors (IoT and Point of Sale devices) or social media platforms can be used to populate a dashDB table through synchronized replication. Many next generation customer-facing applications will benefit by using this data for upselling and for making customer recommendations while an engagement is taking place. Recently, IBM and Apple announced that Cloudant will act as a data layer for new next generation mobile applications. Now, JSON data captured and stored in Cloudant is available to dashDB to run analytics such as reporting, data visualization or geospatial analytics for these new next generation mobile applications.

## dashDB Local

dashDB Local is a client managed data warehouse service for software defined environments (SDEs) such as private clouds, virtual private clouds and other infrastructures that support Docker container technology. Rapid deployment using Docker provides customers immediate access to in-memory data processing, in-database analytics, and a scalable MPP architecture combined with Oracle and Netezza compatibility. dashDB Local is packaged in a Docker container (see graphic) which runs on Linux, Mac OS, or Microsoft Windows

Architecture: Host System on Customer's Cloud or On-Premises Data Center



Scale up CPU, Memory, IO according to infrastructure availability

operating environments, or on a private hosted service like SoftLayer, AWS, or Microsoft Azure. According to IBM, customers have reported using dashDB Local to get a data warehouse up and running in less than 20 minutes for an MPP cluster and in less than 10 minutes for a single node SMP system.

dashDB Local includes Apache Spark as an embedded analytics run-time environment, allowing users to run SQL and Spark queries leveraging dashDB Local in-database analytics. dashDB Local enables the set-up of MPP Spark cluster members co-located with data warehouse cluster members, providing fast exchange of data between members through shared memory. Data scientists can use Scala and Python to develop and deploy applications on Spark with structured data in dashDB serving as the storage layer. IBM says that Spark integration is coming soon for dashDB managed service.

The Fluid Query federation and data movement capability derived from Netezza and embedded in dashDB Local enables the routing and movement of data across physical and virtual data stores. In the near term HRG expects to see IBM deliver dynamic predicate push-down as a capability for use with Spark to dynamically push filtering down to the data source engine to support dashDB Local to dashDB Local federated queries, dashDB to dashDB Local federated queries, and dashDB to Hadoop query federation. All this means faster analysis of data spread across multiple data sources. HRG expects to see support for third party query federation for Oracle, Teradata, Microsoft and generic JDBC data sources in the future.

## Security

The IBM dashDB managed service uses IBM Security Directory Server LDAP for user authentication and IBM Security Guardium Data Activity Monitor for automatic encryption, database activity monitoring and access control. Guardium which is supported for some editions of the dashDB managed service is not yet available for use with dashDB Local. dashDB managed service stores data in the database and automatically saves encrypted backup images using industry standard AES 256 symmetric cryptography. dashDB automatically encrypts data at rest, and the encrypted data can be analyzed in much the same way that BLU compression allows analytics to be performed on compressed data. Encrypting data in transit can be accomplished using SSL port 50001 as long as the unencrypted port 50000 which transmits in clear text is disabled.

dashDB has been developed and tested using best practices including risk assessment, threat modeling, and static and dynamic code analysis. dashDB database activity is monitored, including who has access, who is connected, as well as what actions are performed and what changes are made. This information is made available to customers through the dashDB console. The console supports HTTPS encrypted communication with dashDB and facilitates control over who and what can access which data.

## MPP

Using the dashDB MPP feature for large data sets, customers can easily scale from a single node to an enterprise class MPP data warehouse cluster on demand and then scale back to a single SMP node once the period of peak load has passed. Within the MPP architecture data can be exchanged between servers (nodes) within the query engine in the native BLU columnar vector format. Currently the maximum limit

for dashDB Local MPP scale out is 24 nodes (August 2016). HRG expects IBM to easily address this limitation when customer demand warrants it.

## Highly Available

In the event of a node or server failure, dashDB Local provides high availability functionality so the data warehouse can continue its activities when failures occur. This functionality is based on a heartbeat mechanism with automatic restart and failover. The heartbeat detects when a node, a partition, the web console or the IP address is down, and the cluster manager takes appropriate action. The customer may also provision redundant cluster file system storage for more reliable storage management.

## Scale Out and Scale In

dashDB provides simple scalability in several ways. CPU and memory resources are added to a given cluster to scale up. New nodes are added to a cluster to scale out. Once the peak workload is completed, resources (CPU, memory, nodes) can be returned to the pool to scale in. These actions require a customer to stop and then restart the service on the system with the additional resource added. This feature is not yet fully automatic, but HRG expects IBM to make a fully automated feature available in the future. The workload compute resource requirements of a business can vary widely, and dashDB Local's flexible scaling capabilities ensure that critical SLA metrics are cost effectively met when business workload demands expand and contract based on seasonal demand changes. dashDB uses dynamic in-memory columnar data store technology to minimize I/O, achieving significant speed when contrasted to conventional row-store databases.

## In-Database Analytics

dashDB in-database analytics using R, RStudio and Netezza in-database analytics bring the query to the data right where the data lives. Additional analytics tools are available from IBM and from partners including Cognos, ESRI, and Aginity. dashDB is designed to work with Watson Analytics for cognitive intelligence. As noted in the dashDB Local section above, dashDB's integration with Spark enables data scientists to use dashDB as a data layer for their applications written in Scala or Python, with tight integration between Spark cluster members and dashDB cluster members.

## Use Case: Healthcare and Personalized Medicine

The use of remote monitoring devices to collect and transmit patient data to healthcare professionals drives growth in healthcare Big Data. Telemetric data lets caregivers monitor and track individuals enabling early intervention for conditions as they worsen. Using telemetric healthcare devices improves timely access to care for elderly and rural patients. The use of these devices reduces the cost of care by reducing hospital readmissions and related transport costs. Improved access to, and responsiveness of, caregivers is a significant benefit. Today many patients communicate with doctors, nurses, and caregivers using on-line chat, a web portal, or email. This on-line communication further increases the volume of EHR (Electronic Health Records) and other patient-centric data.

Ultimately the dashDB Local platform could do the analytics to match a patient's health records across the complete gamut of settings in which care and services are provided, allowing clinicians, members of a care team, or even patients to view the complete picture of an individual patient's care requirements. Healthcare management and delivery mechanisms will continuously evolve with the increased use of mobile smart phone and remote monitoring technologies to manage and report real-time health information and data.

Healthcare providers generate increasing volumes of data from their operations, patients, websites, and other sources. Keeping track of that data and analyzing it to provide meaningful and actionable information provides significant benefits to patients, providers, and payers. With in-memory processing and in-database analytics capabilities, IBM dashDB is well positioned to help in the effective diagnosis and treatment of illnesses when symptoms are first presented in a clinical or remote setting. Through real-time analytics, dashDB could help identify global syndromic health threats and developing pandemics as they emerge. Additionally, this solution can be used to help inform first response teams for most effective intervention, containment, and treatment, thus avoiding untold suffering and unnecessary deaths.

Using dashDB and dashDB Local, researchers can use in-memory processing and in-database analytics to compare and analyze anomalies or differences of one genome compared to the base level mapping of the entire human genome or an equivalent proxy. This will enable researchers to identify and decipher disease specific genetic markers resulting in individual/personalized preventive and therapeutic medical applications. Security, privacy, disaster recovery, availability, and other emerging concerns will have to be addressed in order to pave the way for fulfilling the promise that personalized medicine holds. However, dashDB's encryption of data and strong access control go a long way toward addressing these concerns.

## **Use Case: Oil & Gas**

Oil and Gas information technology, to be effective, must ingest, integrate, analyze, and administer data from thousands of sensors installed in wells and surface facilities to continuously monitor, in real-time, all equipment, wells, and environmental conditions. Using the dashDB Local in-memory processing and in-database analytical capabilities, Oil and Gas professionals can now compare current conditions and data against all known and similar situations, enabling them to rapidly make correct decisions and produce actionable insights into logistics, supply chain, marketing, and trading to improve and enable management of the entire energy production and delivery supply chain from first oil to delivery. Predictive in-database analytics using R, and combining current and real-time data with historical data, will help oil and gas suppliers predict and eliminate potential equipment and well failures and eliminate human error.

## **Use Case: Financial Services**

Financial services firms challenged with accelerated and shorter product and services innovation cycles have to deploy and administer core infrastructure systems while working against limited time, money and skilled resources. Good information based on the best available data is central to analytics, planning, and decision-making, and dashDB has an important role to play here. The ability to rapidly respond in real-time to competitive and market place changes is becoming a requirement throughout the financial services sector.

Executives, directors, and managers in financial services firms understand the benefit of having access to the complete range of corporate and market data without having to be exposed to its underlying complexity. Today, no data, regardless of type, location, or format, is beyond integration. It is the nature of businesses to grow and to accumulate data, and those businesses that interpret historical business and market data accurately and apply that knowledge to the analysis of current business and market data in the shortest elapsed time will tend to be market leaders. As companies collect increasing volumes of data from their business operations, websites, marketing initiatives, sales and other sources, keeping track of that information and making it useful to the business through data analytics using the in-memory processing and in-database analytics capabilities that dashDB brings will provide significant distinctive competitive advantage.

## **Conclusions and Recommendation**

Solutions that organize and make sense of the increasing flows of data that are available today are essential to realizing the benefits of Cloud Computing, Big Data and the Internet of Everything. Data has been combined for many years for traditional data warehouses and business intelligence. New applications that combine real-time information from multiple sources are becoming critical for business success. Transitioning to more “digital and social” business models and evolving into digital, social, and cloud based entities, businesses need the right data and data management tools to support today’s customer facing and B-to-B application evolution.

Organizations are flattening and simplifying the IT infrastructure by implementing converged and hyper-converged systems to gain in energy efficiency and reduce IT’s physical infrastructure requirements. These organizations are moving to hybrid cloud implementations to make information and applications more accessible to customers and employees while increasing productivity and reducing cost. Many organizations will, over time, move almost completely to the Cloud. IBM dashDB can help organizations transform themselves to better take advantage of Cloud, Big Data, and IoT, and prepare for a future where change is the only constant. The dashDB cloud-based managed service delivers simplicity, letting customers leave implementation on the cloud, as well as application and system maintenance, to IBM. IBM dashDB and dashDB Local deserve serious consideration if you want to fully exploit and benefit from the increasingly competitive and continually changing global business environment.

*Harvard Research Group is an information technology market research and consulting company. The company provides highly focused market research and consulting services to vendors and users of computer hardware, software, and services. For more information please contact Harvard Research Group:*

## **Harvard Research Group**

Harvard, MA 01451 USA

Tel. (978) 456-3939

e-mail: [hrg@hrgresearch.com](mailto:hrg@hrgresearch.com)

<http://www.hrgresearch.com>