IBM dashDB

*Cloud-based data warehousing as-a-service,
built for analytics*

Highlights

- Gain instant access to critical business insights

- Load, analyze and visualize your data at rapid speeds

- Upload data from multiple sources and integrate with R

- Achieve greater insights with in-database predictive analytic algorithms

- Extend on-premises data warehouse environments to the cloud

- Analyze JSON data through native integration with IBM Cloudant

IBM® dashDB™ is a fast, fully managed, cloud data warehouse that utilizes integrated analytics to rapidly deliver answers. dashDB's unique in-database analytics, R predictive modeling and business intelligence tools free you to analyze your data and get precise insights, quicker.

dashDB is simple to get up and running with rapid provisioning in IBM Bluemix™. You can test the solution or start using dashDB for no charge, for up to one gigabyte of data and then just $50 US per month for 20 gigabytes of data storage. Larger instance sizes with multi-terabyte capacity are available as you grow your data, and as your users require a dedicated environment. Massively Parallel Processing (MPP) enables even faster query speeds as well as larger scale data sets.

IBM dashDB provides the simplicity of a data warehouse appliance as a service, with the agility and scalability of the cloud for any size organization. You can rapidly compose analytic applications using the rich set of developer and complementary services in IBM Bluemix or with your favorite on-premises tools.

Data warehousing before dashDB

Historically, building a data warehouse was a painstaking endeavor. You had to decide on specific data warehousing software and then determine and secure the proper balance of hardware and storage to allocate for it. Once you decided on the physical makeup of the data warehouse, you would then be tasked on both building the physical system as well as the logical data models that would support your initiative. When you needed to expand the data warehouse (and you definitely would, as the data that you are collecting is always growing and new applications are built on top), you would then need to purchase new allocations of processing power, storage and software.

The entire process introduced risks each and every time you would alter the data warehouse. Did you keep the proper balance of processing to data? Did you procure the proper hardware? Was the data allocated across the system correctly? Was the hardware out of date and incompatible with the newest technology on the market? Was the software in need of updates? At any point in the growth process, you were opening yourself up to a host of issues that could occur. Add to this the fact that you had to pay for capacity, even if you were not yet using it. You and your organization were taking all of the risk.

Enter the data warehouse appliance

Data warehouse appliances helped alleviate much of the pain of building your own data warehouse. These systems came pre-configured and integrated the data warehouse for analytical performance. All you really had to do was pick out your model and size, and once plugged in and turned on, you loaded your data into the warehouse.

The allocations of hardware and software were already done for you, yet the appliance still required you to buy a new system when you hit your resource limits. With the amount of growing data, new applications and new users, implementing a new warehouse could get expensive as you dynamically scale. Add to the situation upgrades, patches, maintenance and overall hardware depreciation, and you are still left with much of the maintenance of a traditional data warehouse.

The data warehouse appliance is still an essential part of an organization's decision management system, but is there some sort of complementary technology that would alleviate some of these growing pains? Enter IBM dashDB.

Benefits of the cloud

Through the scalable and on demand nature of cloud computing, you can get the data warehouse up and running very quickly with rapid cloud provisioning. Since there isn't any infrastructure investment, you are empowered with true business agility. You buy what you want, when you want. You are in charge.

As a fully managed cloud service, the day-to-day backend of maintaining dashDB is taken care of for you. This maintenance is not only limited to whatever fixpacks may or may not be applied (these actually become irrelevant because you never know that they have occurred until you read the corresponding documentation) but also the versioning. With dashDB, our dedicated engineers and developers are continuously building new functionality, compatibility and integrations into the product—and for you, this happens automatically.

Protect your data with dashDB

From design to deployment, dashDB is optimized to provide thorough security coverage of your data. dashDB includes multiple layers of security such as automatic encryption for data at rest and in transit, database activity monitoring with IBM InfoSphere® Guardium®, advanced database access control, and deployment hardening.

Security starts with design and development best practices. dashDB is developed using practices such as risk assessment, threat modeling and static and dynamic code analysis using IBM AppScan®.

Grow faster by focusing on your business, not the business of data warehousing

The simplicity of cloud is key—dashDB is a fully managed service and you have just heard a few of the cloud's inherent benefits, such as automatic fixpacks and versions, rapid cloud provisioning and business agility. By utilizing dashDB as a service, you pay for the product as you go, based on what you use. This plan differs greatly from buying a complete data warehouse. With dashDB, as your data grows, you pay for added capacity—simple and controlled.

In concert with simple pricing, the fact that this solution is a service and not a hardware cluster or appliance helps you to grow as your business demands. There is no purchase, installation and testing of new software and hardware—you simply grow as you need in the cloud.

Built-in performance with in-memory technology delivers fast answers

What makes dashDB different from other cloud-based data warehouses? A lot. At the core of dashDB is IBM's BLU Acceleration technology.

IBM BLU Acceleration is an in-memory database technology that provides cutting-edge warehousing performance without the typical constraints of in-memory solutions. Since dashDB is built on BLU, it maintains all of its advantages:

- **Advanced processing:** dashDB does not require the entire dataset to fit in-memory while still processing at lightning-fast speeds—it uses a series of patented algorithms that nimbly handle in-memory data processing.
- **Prefetching of data:** dashDB is designed to anticipate and "prefetch" data just before it's needed, and to automatically adapt to keep necessary data in or close to the CPU.
- **No decompression required:** dashDB preserves the order of data and performs a broad range of operations—including joins and predicate evaluations—on compressed data, without the need for decompression, drastically speeding the processing of data.
- **Data skipping:** When dealing with big data, there is a good chance that you don't need *everything* in the data warehouse to answer a particular query. dashDB's BLU Acceleration is designed to automatically determine which data would not qualify for analysis within a particular query. Large quantities of irrelevant data can then be skipped over during a query, saving you time and resources.

MPP capabilities enable faster queries and massive data sets

dashDB's MPP builds upon the benefits of the standard dashDB service with even more speed and scale, so you can handle much larger data sets. The MPP architecture is a networked cluster of servers working in parallel to speed up query fulfillment. In the dashDB MPP cluster, multiple servers work on the same query simultaneously, and the processing of a query at each server is further parallelized across all the processors.

In a standard architecture, parallelization occurs only at the processor level. With an MPP architecture, a query is broken up into pieces so that multiple servers, each with their own local storage and compute capacity, are working on separate pieces of the data. This team effort drastically speeds up the querying process, and reduces I/O requirements. Each individual server working on a query in MPP leverages BLU dynamic in-memory columnar store technology, which minimizes I/O even further and achieves an order of magnitude in speed when compared to conventional row-store databases.

With MPP, performance improvements are increased with each new server added to the network cluster. For example, if a query takes one hour in a standard architecture using a single server, it would take approximately 15 minutes with an MPP cluster utilizing just four servers. Adding one more server, for a total of five, reduces the query time to 12 minutes; six servers reduces the query time further to 10 minutes; and so on. Therefore, with dashDB MPP, scaling out is as simple as adding additional servers to your cluster.

Built for analytics to help you understand your data and business

The evolution of in-memory processing is moving faster than ever due to the substantial growth in data volumes that organizations are tackling. As hardware and memory are commoditized, we are able to push more of the data to the memory and process it there.

Organizations used to have to wait to receive analytic reports from the data warehouse; yet with new advances with in-memory computing, this is no longer the case. Results are now available for decision making in real-time. This enhanced speed also makes further analysis of the results feasible for the analyst who may wish to dig deeper into results.

The data warehouse is now being used as the main data store for analytics. Therefore, it makes sense to bring analytics to the data warehouse rather than move data out to analyze it elsewhere.

In-database analytics for greater efficiency and performance

Building upon the performance of BLU Acceleration, dashDB also integrates IBM Netezza® Analytics for fully integrated in-database advanced analytics. The same technology has been used by IBM Netezza appliances and IBM PureData™ for Analytics systems.

What this means is that with dashDB, you get a myriad of predictive modeling algorithms *built directly into the database*. These algorithms are available whenever you want to use them.

An example of the algorithms included with dashDB are:

• Linear regression
• Decision tree clustering
• K-means clustering
• Esri-compatible geospatial extensions

By running the analytics natively in the database, where the data resides, your organization can gain huge efficiencies. Rather than having to extract the data, send it somewhere else, stage it and then process it, you are leaving the data where it resides (in the data warehouse) and then applying the analytics directly to it.

Compatibility with advanced tooling like R and IBM Watson Analytics

R is an open source programming language that was developed for advanced data analysis and graphical visualization. It can be used to analyze data from many different data sources including external files or databases. dashDB integrates R for predictive modeling through an R runtime alongside the data. A web console can be used to load data and perform analytics within minutes. Data analysis could include SQL, BI tools, or R scripts and models. With dashDB and open source R, your analytical options are broad and varied.

IBM dashDB includes RStudio — a fully integrated R development environment designed to provide quick, R-based predictive analytics. RStudio provides an R language code completion feature, integrated help for R packages, file management capabilities and much more. If you need to install additional R packages, you can do so easily from within RStudio.

dashDB was developed with the larger business intelligence ecosystem in mind. The dashDB service works natively with core IBM technologies like IBM Watson™ Analytics, IBM Cognos® BI, IBM DataWorks and others, yet it definitely does not stop there. dashDB was built to work with IBM's myriad of business partners and BI tool sets including Looker, Aginity Workbench, Tableau and many others.

The IBM partner ecosystem is growing rapidly, and you can connect multiple third-party tools to the dashDB service. For example:

• Connect IBM InfoSphere Data Architect to design and deploy your database schema
• Connect Esri ArcGIS to perform geospatial analytics and map publishing with your data
• Connect an IBM Cognos server to run Cognos reports against your data
• Connect SQL-based tools such as Tableau, Microstrategy, or Microsoft Excel to manipulate or analyze your data
• Connect your Bluemix™ applications that require an analytics database
• Connect Aginity Workbench to migrate Netezza data models and data to dashDB

Use cases

There are many ways that organizations are taking advantage of dashDB. Listed below are four main use cases that we are seeing dashDB customers active in today.

1. **Augmenting the existing data warehouse – Hybrid**
   It is stated that over 90 percent of clients plan to augment their existing data warehouse[1]. dashDB is well structured for this usage. Through dashDB, organizations are able to extend on-premise data warehouse environments to the cloud. Since you pay for the capacity you need today, the platform is elastic and is there when you need it. With dashDB, you are not forced into buying utility before you need it.

2. **Analysis of NoSQL data**
   The second key use case comes from core IBM integration between Cloudant® and dashDB. You can easily synchronize your JSON data within Cloudant to structured data within dashDB. This process then allows for traditional BI and analytics common in data warehouses. Since dashDB has in-database predictive algorithms built in, those clients that are using Cloudant can analyze their JSON document stores, hassle-free.

3. **Data science data store**
   Specific to this idea of in-database analytics, the third key use case is detailed here. For the statistically inclined as well as for data scientists, dashDB maintains a robust set of predictive analytic algorithms. As stated, dashDB has R runtime and RStudio built in. R is widely used among statisticians and data miners for developing statistical software and data analysis.

4. **Standalone data warehouse as a service**
   dashDB is heavily used as a standalone data warehouse in the cloud. Whether it be small start-up datamarts that you spawn off of some Cloudant data, a basic test or development environment, or even full enterprise data warehousing on the cloud, dashDB is available 24x7 for you and your organization.

Getting started with dashDB

IBM dashDB is a fully managed data warehousing service, so there is no hardware or software to purchase and install. Simply sign up for an account at www.dashdb.com and start using it for free. If you need assistance learning dashDB or determining whether it is a good fit for your organization's requirements, feel free to contact us online at dashdb.com, and we will answer your questions or set up an in depth discussion with our technical team.
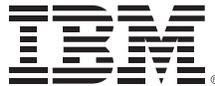
About IBM dashDB solutions

IBM provides the most comprehensive portfolio of data warehousing, information management and business analytic software, hardware and solutions to help clients maximize the value of their information assets and discover new insights to make better and faster decisions and optimize their business outcomes.

For more information

TTo learn more about dashDB, please contact your IBM representative or IBM Business Partner, or visit the following website: www.dashdb.com

Additionally, IBM Global Financing can help you acquire the IT solutions that your business needs in the most cost-effective and strategic way possible. We'll partner with credit-qualified clients to customize an IT financing solution to suit your business goals, enable effective cash management, and improve your total cost of ownership. IBM Global Financing is your smartest choice to fund critical IT investments and propel your business forward.

For more information, visit **ibm.com**/financing

IBM

dashDB