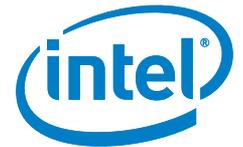


## SOLUTION BRIEF

### Cloud-Based Data Warehousing

Intel® Xeon® Processor E5-2600 v3 Product Family



# Powerful, Agile, Cloud-Based Data Warehousing with IBM dashDB\* MPP on Intel® Architecture

**Delivering performance, scalability, and agility, IBM dashDB\* MPP servers based on Intel® architecture building blocks provide a fully managed, cloud data warehouse service with integrated, in-database analytics.**

---

“The massive growth of data is redefining today’s competitive landscape by demanding more performance than ever before. With IBM dashDB\* and other cloud data services, supported by Intel® architecture, we get an advanced suite of analytical offerings, with the massive performance capability to accelerate our delivery of products and services for media companies.”

– Mukesh Sehgal,  
President and CEO, RSG Media

Many organizations still face significant challenges in realizing the full opportunity of data analytics. They need data warehouse infrastructure to be simple and agile enough to add new services quickly. Services must also reach into all corporate data stores, while giving users the power to control the environment, independent of IT.

Those challenges become even more pronounced as data stores continue to grow, making it critical for data warehousing solutions to incorporate optimizations that take advantage of each successive generation of Intel’s leadership in processors, storage, and network controllers. Data warehouse infrastructures must also evolve to strategic technology shifts such as moving to the cloud.

To help keep pace with these requirements, IBM recently introduced dashDB\* MPP, a new high-performance, scalable data warehouse service that is fully managed in the cloud. Built to deliver next-generation analytics and take advantage of innovations in Intel® architecture, dashDB\* MPP changes the game by offering:

- **Self-service analytics for business users.** Design queries and do predictive modeling, without being a data scientist or worrying about the infrastructure that is working behind the scenes.
- **Simplicity for IT departments.** Eliminate the overhead of constant infrastructure maintenance to serve emerging needs from line-of-business users.

Innovations throughout the Intel architecture-based hardware stack help generate notable improvements in database performance from dashDB MPP. Intel® processors rapidly execute queries against massive data sets, while Intel® Solid-State Drives (Intel® SSDs) enable high read-write performance while providing advanced data protection. Intel® Ethernet Controllers deliver efficient traffic handling that helps drive performance. As a result, businesses can handle larger data sets, complete more queries faster, and make more informed decisions.

### Next-Generation Performance Advances

Lab testing by IBM demonstrates performance gains of approximately 1.2x on a dashDB workload between generations of the Intel® Xeon® processor E5-2600 product family, as illustrated in Figure 1.<sup>1</sup> The workload behind these test results is designed to simulate real-world business analytics, using BD-Insights\*, an IBM-developed benchmark. A total of 30 deep-analytics queries are used, in two broad categories:

- **Intermediate queries** represent typical analytic queries used to generate sales reports.
- **Complex queries** represent hand-crafted, deep-dive analytic queries created by sales analysts.

Testing used 60 work streams (representing 60 users), each running analytics queries in pseudo-random order. Two sets of throughput data are presented in the figure:

- **Full-workload** results are based on the use of all 30 queries, which vary significantly in complexity.
- **Processor-intensive queries** results are based on a subset of the 10 most compute-intensive queries.

### Server-Hardware Innovations That Help Deliver the Performance Boosts

The cloud infrastructure that hosts dashDB MPP is powered by the ongoing design advances of the Intel Xeon processor E5-2600 product family. As the heart of the high-performance, agile data center, this platform delivers significant business benefits. When paired with Intel® technologies for local storage and networking, the advantages are even more striking.

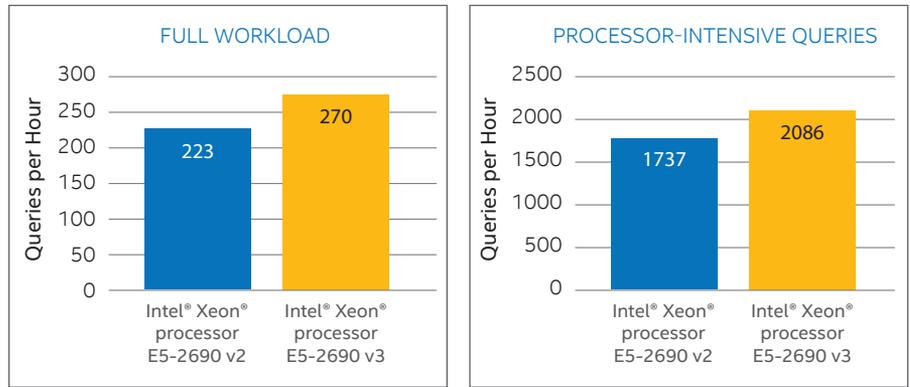


Figure 1. Performance of dashDB\* MPP increases up to 1.2x between processor generations.<sup>1</sup>

### Compute: The Intel® Xeon® Processor E5-2600 v3 Product Family

The balanced platform for dashDB MPP analytics begins with the Intel Xeon processor E5-2600 v3 product family. Key architectural advances over the previous generation that contribute to dashDB performance include the following:

- **High number of cores** allow more simultaneous operations, for faster results from large numbers of simultaneous queries.
- **Large cache** keeps more data available for fast access by the processor cores, for faster responsiveness.
- **Fast Intel® QuickPath Interconnect links** move data through the platform rapidly, accelerating analytics workloads.
- **DDR4 memory** provides high memory bandwidth to support data-intensive analytics.

Data movement within the platform is accelerated by Intel® Data Direct I/O (Intel® DDIO), which lets the Ethernet controller communicate directly with processor cache, bypassing slower system memory. Code in dashDB MPP also takes advantage of Intel® Advanced Vector Extensions (Intel® AVX) instructions to speed up columnar reads. These capabilities enable outstanding responsiveness in generating fast answers to sophisticated business questions.

### Storage and Networking: Intel® Solid-State Drives and Intel® Ethernet Controller X540

Intel® SSD Data Center S3700 Series accelerates data movement for table storage in dashDB MPP. Built for tight, platform-level integration with the rest of the Intel® hardware stack, these solid-state drives deliver high performance and protect against data loss and corruption.

**Note:** IBM SoftLayer\* dynamically assigns storage resources on demand, potentially including non-Intel solid-state drives.

Optimal transfer of data among hosts is enabled by network adapters based on the Intel® Ethernet Controller X540. Offering 10-gigabit port speed and built for robust integration with the rest of the hardware and software stack, the controller offers efficient load balancing across processor cores, supporting high network I/O.

For dashDB and other demanding business workloads, this ability to optimize the efficiency of data handling from multiple streams can make data warehousing infrastructures more responsive to business needs that require rapid transformation of massive data sets into actionable insight.

## Data Warehousing Born in the Cloud: IBM dashDB

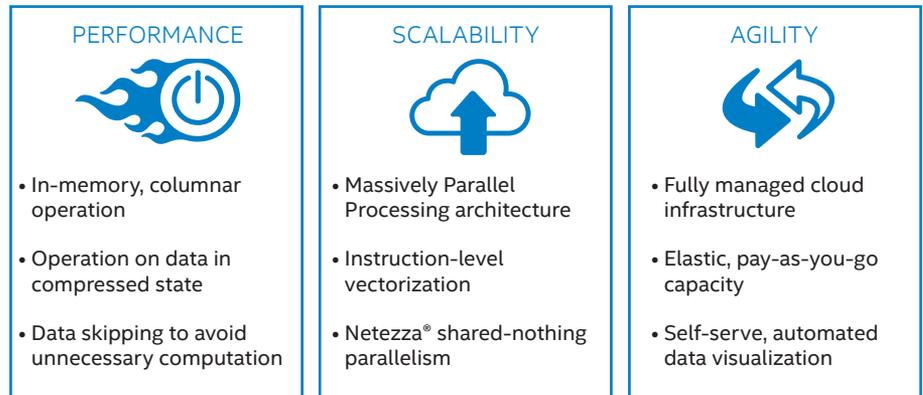
Combining mature technologies for data warehousing and analytics with cutting-edge cloud architecture, dashDB MPP provides businesses with performance, scalability, and agility, as illustrated in Figure 2.

dashDB MPP employs the massively parallel processing (MPP) architecture. MPP allows the dashDB service to scale out across large numbers of servers, taking advantage of the local compute and storage resources throughout the cluster. Scaling to more servers also reduces the I/O requirements on each node. A query that takes 20 minutes on a standard, single-server architecture, for example, could theoretically be accomplished in 10 minutes using two servers with MPP architecture, in four minutes using five servers, and so on.

This solution architecture directly benefits from the compute, storage, and networking benefits of the Intel architecture hardware stack, providing timely insights into business questions such as, “How are we doing?”, “What’s responsible for those results?”, and “What should we do to improve?”

**Note:** The dashDB\* product is offered at multiple plan levels. The results and discussion in this paper refer to the Enterprise MPP.4 release, which is required to obtain massively parallel processing (MPP) functionality. For details on all dashDB offerings, please visit <https://console.ng.bluemix.net/catalog/dashdb/>.

In concert with Intel’s commitment to the advancement of cloud technology, dashDB is a fully managed cloud service that streamlines operations related to data warehousing and analytics. Provisioning is simple, allowing implementation in just minutes, with zero capital expenditures. Elastic capacity enables businesses to spin resources up and down at will, paying only for what they use. And all back-end maintenance is handled by IBM, transparently.



**Figure 2.** IBM dashDB\* offers advanced performance, scalability, and agility.

Business users can simply access dashDB from a browser, effortlessly transform JSON data into a relational form, synchronize with the data warehouse, and access it as a table. They can rapidly explore the data, run queries against it, and view multiple visualizations, automatically.

The IBM Bluemix\* platform enables businesses to create a new dashDB service easily and free of charge from the service catalog on the Bluemix portal. That simple dashDB instance can then be upgraded and be seamlessly connected to other IBM Cloud Data services through Bluemix, to accelerate time to market, improve uptime, and streamline development of web and mobile applications. It also takes advantage of IBM SoftLayer\* bare-metal infrastructure to deliver a high-performance, flexible environment.

### High-Performance In-Memory Computing with IBM BLU Acceleration\*

To enable rapid results on complex queries and massive data sets, dashDB MPP takes advantage of IBM BLU Acceleration, a patented, in-memory columnar data-store technology that accelerates processing, even when the data set is too big to fit into memory. Innovations in the memory subsystem of the Intel Xeon processor E5-2600 v3 product family, such as DDR4 memory and Intel DDIO, can have a direct positive impact on this in-memory operation.

BLU also enables dashDB to take optimal advantage of server-platform parallelism. It is highly optimized to benefit from both multi-core hardware architecture and software vectorization, increasing performance through intelligent use of single-input, multiple-data (SIMD) instructions. A further innovation that increases efficiency and throughput is dashDB’s data-skipping functionality, which enables it to identify ranges of column data that do not need to be included in a given query, thus reducing processing requirements for the analysis at hand.

### Infrastructure as a Service for dashDB on IBM SoftLayer

IBM SoftLayer is a premier option for deployment of dashDB that takes advantage of Intel architecture through IBM’s status as an Intel® Cloud Partner. The seamless, automated cloud environment provides global infrastructure that moves businesses one step closer to decoupling their compute operations from specific geographies, with elasticity and scalability on demand as new business needs arise. Capacity for SoftLayer includes nearly 200,000 servers worldwide.

Intel® Trusted Execution Technology (Intel® TXT) is enabled on some IBM SoftLayer servers, helping establish a trusted launch environment for applications. At system startup, Intel TXT verifies critical attributes of system firmware and software components against a known-good system image to ensure that they have not been tampered with. This hardware-based protection helps protect system integrity below the software level, with mechanisms that are out of reach from software compromise. For more information, see [www.intelcloudfinder.com/softlayer](http://www.intelcloudfinder.com/softlayer).

## Conclusion

With its massively parallel, in-memory columnar design, dashDB MPP is a powerful engine for business analytics, built to take excellent advantage of Intel architecture at the processing, local storage, and network-connectivity levels. As a cloud-native set of technologies, dashDB offers a fully managed data warehousing service, with integrated, in-database analytics. For further information, contact your IBM Cloud Data Services sales rep or send email to [dashDB\\_Info@wwpd1.vnet.ibm.com](mailto:dashDB_Info@wwpd1.vnet.ibm.com).

Co-engineering by IBM and Intel extends the performance and scalability of dashDB MPP, offering operational advantages from the architectural innovations of the Intel Xeon processor E5-2600 v3 product family.

Line-of-business users now have simplified, independent means to unlock their data's potential through reports, visualizations, and dashboards. Data analysts have the power to perform advanced multi-dimensional analysis and data mining to produce rapid insights. IT organizations can focus on high-value activities, rather than maintaining the analytics infrastructure. Business as a whole becomes more intelligent and capable, paving the path to success.

For more information, visit [intel.com/xeon](http://intel.com/xeon) and [ibm.com/dashdb](http://ibm.com/dashdb)



<sup>1</sup> Testing performed by IBM.

Hardware configurations:

- Server based on two Intel® Xeon® processors E5-2690 v2 @ 3.00 GHz, 256 GB DDR3 RAM; table storage: six Intel® Solid-State Drive Data Center S3700 Series (800 GB each); staging path and load copy: six Seagate Constellation\* hard disk drives (1 TB each); network interfaces: SuperMicro AOC-2UR6-I4XT 2U Ultra Riser\*, Intel® Ethernet Controller X540 (10 Gbps).
- Server based on two Intel® Xeon® processors E5-2690 v3 @ 2.6 GHz, 256 GB DDR4 RAM; table storage: six Intel® Solid-State Drive Data Center S3700 Series (800 GB each); staging path and load copy: six Seagate Constellation\* hard disk drives (1 TB each); network interfaces: SuperMicro AOC-2UR6-I4XT 2U Ultra Riser\*, Intel® Ethernet Controller X540 (10 Gbps).

Software configuration (all test systems): CentOS® 6.6 (2.6.32-504.12.2.el6.x86\_64); dashDB\* MPP code base db2\_main\_linuxamd64\_s150619; GPFS gpfs.base-4.1.0-6.x86\_64.gpfs.crypto-4.1.0-6.x86\_64.gpfs.docs-4.1.0-6.noarch, gpfs.ext-4.1.0-6.x86\_64.gpfs.gpl-4.1.0-6.noarch, gpfs.gplbin-2.6.32-504.12.2.el6.x86\_64-4.1.0-6.x86\_64.gpfs.gskit-8.0.50-32.x86\_64.gpfs.msg.en\_US-4.1.0-6.noarch.

Intel® Compiler Version 13.0 Build 20130327.

Performance and compression data is based on measurements and projections using IBM benchmarks in a controlled environment. The actual throughput, performance or compression 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.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.

Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Notice Revision #20110804

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document. You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Intel products described herein. You agree to grant Intel a non-exclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at [intel.com](http://intel.com).

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

Copyright © 2015 Intel Corporation. All rights reserved. Intel, the Intel logo, and Xeon are trademarks of Intel Corporation in the U.S. and other countries.

\*Other names may be trademarks of their respective owners. Printed in USA 0815/RA/MESH/PDF ♻️ Please Recycle 332852-001US