Bloor

InDetail

**InDetail** Paper by Bloor
Author **Philip Howard**
Publish date **September 2016**

# IBM DB2 version 11.1 (and the hybrid cloud)

**...this is not entirely a new product release in the conventional sense of that term – some of the major new features in version 11 have previously been available within IBM's cloud-based dashDB product. What this means is that these capabilities have already been tried and tested in operational environments.**

Author **Philip Howard**

# Executive summary

This is, theoretically, a review of the latest release of IBM's flagship database DB2. It specifically refers to the version of DB2 that runs in distributed environments and not to the mainframe version. This release (version 11 – technically 11.1) was announced in April 2016 for general availability in June of this year. However, this is not entirely a new product release in the conventional sense of that term. Briefly – we will discuss this in detail later – some of the major new features in version 11 have previously been available within IBM's cloud-based dashDB product. What this means is that these capabilities have already been tried and tested in operational environments and we would therefore expect the new release to be more robust and stable than one might normally expect.

More generally, this is a somewhat unusual product evaluation. Aside from the fact that major features of this release have been previously available elsewhere there are two further reasons for this. The first is that this paper does not focus on DB2 generally but specifically on the features in the latest release of DB2. In other words, it assumes a reasonable degree of familiarity with the DB2 database. However, having said that, we will have to explore some of the capabilities that were previously available within the product – or, specifically, within DB2 with BLU Acceleration – in order to understand the significance of the latest functionality. Secondly, the focus of this paper is not so much on the features of the product as on the impact of those features, though of course we have to describe the former in order to explain the latter. In particular, we will be discussing the use of DB2 within hybrid cloud environments.

## Fast facts

IBM DB2 needs little introduction: it has been one of the market's leading database products for upwards of a quarter of a century. Originally it was a single database that aimed to do everything but over the last few years IBM has recognised that transactional environments require different architectures from data warehousing and it has delivered options appropriately so that there have been two primary deployment options, both from a technical and marketing perspective. However, in the latest release of DB2 all the features that support any particular environment are included within the software and treated simply as installation options that you select or not. The same is true with respect to the various editions of the product that you can license.

## Key findings

In the opinion of Bloor Research, the following represent the key features of the latest release. The list here is relatively comprehensive because not all of these will be discussed in detail within the body of this paper.

- Version 11 supports BLU Acceleration for massively parallel processing (MPP) based systems. It was previously only available for symmetric multi-processing (SMP) environments.
- The latest release incorporates new features taken from the Netezza environments, notably that product's in-database analytics (INZA) capabilities.
- IBM Federation Server, which was originally a part of DB2 and then split out and marketed and sold separately, has been returned to its original home as a part of the database. This is specifically to support logical data warehousing and hybrid cloud environments, which are major foci in this release.
- On this same theme, though neither of these is new in this release, DB2 for Big Data enables DB2 to work alongside IBM BigInsights and, of course, DB2 is available in the cloud (as a hosted service).

**“The latest release incorporates new features taken from the Netezza environments, notably that product's in-database analytics (INZA) capabilities.”**

> **IBM wants to emphasise its support for hybrid cloud environments.**

- You can upgrade directly to version 11 from version 9.7 and later, without having to go through any intermediate steps.
- Previously IBM has expended considerable effort to ensure SQL compatibility with Oracle. In this release that effort has been extended to Netezza and PostgreSQL (the latter because the former is based on it).
- HP-UX and Solaris are no longer supported platforms as clients transition to Linux or AIX, running on either Intel or Power processors.
- For Linux environments running VMWare IBM has introduced support for virtualised RDMA (remote data memory access) adapters. RDMA is faster than Sockets (which is also supported) for low latency networking but normally you need multiple such adapters. Virtual support for RDMA means that you only need one physical adapter in Linux/VMWare environments.
- Synchronous and near-synchronous options are now available for HADR (high availability data replication) with DB2 pureScale, in addition to the asynchronous capability that was previously available. Note that you can have only one synchronous copy but any number of asynchronous ones.
- Encryption support has been enhanced with central key management. Previously you had to have local key stores. This release is also previewing hardware-based encryption.
- There are a number of enhancements supporting transactional environments based on IBM PureScale (the clustered version of DB2). These include performance enhancements and throughput optimisation, as well as some particular features for supporting geographically distributed installations. There are also new bundled failover options.
- There are new licensing options with this release that simplify processor and sub-capacity licensing. New license metrics have been designed to facilitate hybrid cloud deployments and to ensure compatibility between on-premises and cloud-based installations.

## The bottom line

IBM does not describe version 11 as a major release. The only reason it has a new number (eleven versus ten something) is that IBM wants to emphasise its support for hybrid cloud environments. We accept that this is important. More generally, it's a question of what's not to like? You certainly couldn't say that there is anything obvious that you would like to see in DB2 that isn't there and it is not hard to see that all of the new features are potentially beneficial. However, we do not altogether subscribe to the idea that this is not a major release. In our opinion, the extension of BLU acceleration to MPP-based systems is very significant. We are also pleased to see Federation Server as, once again, a part of the database. In our view it should never have left. Of course, this capability is fundamental to supporting logical data warehouses and, more broadly, to supporting hybrid cloud environments. So, we are inclined to think that actually this is a major release, at least in its implications if not in its actual features.

# Blu acceleration

Before discussing the role of DB2 within hybrid cloud environments, which is the main focus of this paper, we will turn our attention to what used to be known as BLU Acceleration. We say "used to" because it was historically a set of capabilities for improving analytic performance that only applied to SMP-based solutions. With this release, BLU Acceleration is available for all DB2 implementations regardless of whether they are SMP or MPP-based. While "BLU" will, no doubt, continue to be referred to for some time to come, in practice its capabilities are now just options that you choose to implement or not. It is no longer a separate product.

BLU Acceleration, as it used to exist, had five major sets of capabilities that it provided:

- "Actionable" compression. Not only does IBM offer advanced compression capabilities but for many query processes you do not need to de-compress the data in order to query it. This is because the techniques now used for compression (known as frequency-based dictionary compression) preserve order. This means that predicate evaluation, joins, grouping and count queries can all be performed without de-compressing the data.
- Data may be held in columnar format. This is not an all or nothing choice. Where it makes sense to use columns you can use columns and where it makes sense to use rows you can use rows combined with conventional indexes. We do not believe that we need to rehearse the value of using a columnar approach for appropriate query types. Note that column-organised tables do not have secondary structures, such as materialised query tables, thereby eliminating any need for synchronisation.
- It provides in-memory processing through what it calls "dynamic in-memory" capabilities that allow columns of data to reside in memory for query purposes. Unlike some other databases that have been designed purely to provide in-memory capability BLU Acceleration was developed on the premise that your data warehouse will always exceed the size of the memory that is available. Of course, you can always pay a fortune to have perhaps hundreds of terabytes of memory but that is an impractical proposition for most organisations.
- It leverages the ZoneMap technology gained when IBM acquired Netezza. The company calls this data skipping and the way that it works is that the database holds metadata about where data is stored and this allows you to "skip" those areas of the database where the data that the query requires is not stored. This can very significantly speed up relevant queries.
- It uses parallel vector processing. This is cross-core parallelism: you can parallelise across the cores within a single CPU as well as across sockets – and parallelises Multiple Data elements per Single Instruction cycle in each core (SIMD). IBM claims one of its strengths is the efficiency of the parallelism made possible by such deep engineering to reduce "memory access latency", in other words the time it takes to get data from RAM into the CPU where it can be processed. Most systems have a fair bit of memory access latency, but IBM leverages CPU cache to minimise this.

As we have mentioned, this only ran on SMP-based systems. However, last year (2015) IBM released dashDB. This is a managed data warehousing service runing in the cloud. Note that DB2 in the cloud is hosted but not managed. dashDB is, in effect, a combination of DB2 with BLU Acceleration and Netezza (PureData for Analytics) technologies. Moreover, dashDB is a "cloud first" environment. In other words, it represents an environment where new capabilities can be implemented in a very agile manner and with short development cycles. In any case, the key

> **"With this release, BLU Acceleration is available for all DB2 implementations regardless of whether they are SMP or MPP-based."**

point here is that this is the first place where BLU Accelearation was developed to run on MPP platforms, which is why we stated previously that DB2 version 11 doesn't entirely reflect what we are used to thinking of as a new release.
The same also applies to SQL compatibility with Netezza (and PostgreSQL). And this capability in turn means that the sort of in-database analytics that Netezza has historically supported are now also vailable for DB2. In-database analytics significantly improves performance for appropriate analytics, whether via user defined functions, third party libraries from vendors such as Fuzzy Logix, or directly from IBM environments such as SPSS.

Finally, with respect to BLU Acceleration we should say that the availability on MPP-based platforms is not simply a question of providing improved analytic performance. For instance, you now no longer have to make a choice between query performance and the ability to parallelise back-ups. Now you can do both. This will be of particular benefit, for example, to very large SAP customers running SAP BW workloads on top of DB2 with DPF (data partitioning feature). We would expect these users to migrate to BLU-based MPP systems while medium to large SAP customers have the option to simplify their environment by implementing BLU and removing DPF.

**“In-database analytics significantly improves performance for appropriate analytics, whether via user defined functions, third party libraries from vendors such as Fuzzy Logix, or directly from IBM environments such as SPSS.”**

# Logical data warehousing

The concept behind logical data warehousing (LDW) is that different types of data are stored in different types of repositories, that queries can span these repositories, and that it is the entire environment that comprises the enterprise data warehouse rather than a single monolithic entity, which has historically been the case. There are several pre-requisites for supporting an LDW. In particular, these include understanding the metadata relevant to each data store, being able to push down queries to each repository, and having an optimiser that understands both the constituent databases and the characteristics of the network. The technology that provides this set of capabilities is known variously as either data federation or data virtualisation. In IBM's terms this is provided by Federation Server.

What is now Federation Server started life as DataJoiner, which was a feature of DB2 more than a decade ago. It then became an independent product known as DB2 Information Integrator, then WebSphere Information Integrator and then it was spun back out again as Federation Server within the InfoSphere family and, finally, it is back with DB2 where it belongs. The reason this travelogue is important to understand is because DataJoiner was the first product of its type in the market and it was the leading product in the market. However, in moving it away from the database – the product leverages the DB2 optimiser, amongst other things – focus was lost. In particular, moving it into the WebSphere group was a mistake and capabilities fell behind those of its competition. Fortunately, IBM saw the error of its ways and significant efforts have been made to rectify its market position over the last few years, which we expect to continue now that it is back where it belongs.

The second aspect of LDW support is DB2 for Big Data. This was released initially for deployment with DB2 version 10.5 and is similarly available for version 11. Effectively this is a way of licensing DB2, IBM BigInsights (which is IBM's Hadoop implementation) along with Federation Server to link these two environments together, along with SQL compatibility (see next) between DB2 and BigInsights. This is significant not merely for its own sake but also because it helps in integrating the semi-structured "schema after" data platform of Hadoop (BigInsights) with on-premises "schema before" SQL platforms such as DB2.

It is worth commenting on IBM's strategy here. Without going into too many details of products that have not actually been released yet, the intention is for all IBM databases to be able to share common SQL capabilities so that queries can span any and all of the company's databases, including DB2, Informix, dashDB, Netezza, BigInsights and so on. What this means in practice is that IBM is developing what it calls the "Common Analytics Engine", of which the most important component is a common SQL engine. This is what has already been implemented in dashDB and which will be implemented in other products over time. This will mean that queries will be portable across all of IBM's environment so that, in principle, you can write a query once but deploy it anywhere. This will require SQL compatibility across all of these product lines and will not be a simple task. Big SQL 3.0, for example, which is used with BigInsights is based on the SQL 2011 standard, so all the other products will have to be brought up to this level and then common extensions supported. However, we have already started to see this convergence and we expect significant progress to be made in this area over the coming months.

Finally, from a strategic perspective, IBM will leverage not just Federation Server for this purpose but also other integration technologies such as BigIntegrate (an existing version of DataStage) as well as open source technologies like Apache Spark. In this last case it is worth noting that DB2 and Informix Spark adapters are already available.

**"IBM is developing what it calls the "Common Analytics Engine" – when fully implemented this will mean that queries will be portable across all of IBM's environment."**

# Hybrid cloud

As a general rule hybrid cloud environments are analytic. This might be considered a contentious statement, depending on what you understand by the word "hybrid". We define hybrid cloud as *"an environment that supports the continuous interworking of applications and functions, some of which are based in on-premises environments, and some of which are deployed in public or managed cloud environments"*. We tend to refer to environments where there are no such continuous interactions as either siloed or heterogeneous, depending on the degree of integration across the ecosystem. There are, for example, lots of non-analytic applications running in the cloud, and many of these run alongside and interact with non-analytic applications that run on-premises. However, we would argue that the extent of these interactions is limited, is often batch-based and is sporadic rather than continuous. Conversely, consider an MDM application wherein customer data is augmented with data from sentiment analysis. Now, that sentiment analysis might or not be done in the cloud, but it is certainly hybrid in the true sense of that word. Going forward we expect the Internet of Things (IoT) to be a major driving force behind the hybrid cloud, with a lot of collected data being stored and analysed in the cloud, and with the results of those analyses being operationalised within on-premises environments and applications.

What this means in practical terms is that a lot of LDW implementations are going to leverage hybrid cloud environments with existing investments in on-premises systems being extended and augmented by data stored and analysed in the cloud.

From IBM's perspective all the comments made in in the previous section about logical data warehousing apply here. We should also mention Cloudant. This is a cloud-based NoSQL database provided by IBM that is primarily oriented towards the storage and processing of JSON documents. It integrates with dashDB. Cloudant is important in this context because a lot of IoT devices generate JSON documents and many IoT environments involve operational processing as well as providing machine generated data for analytic purposes. One can easily, therefore, imagine a scenario in which Cloudant and dashDB are operating together in the cloud and both are operating in concert with DB2. In some other environments it will be a question of dashDB and DB2 acting in synergy, while in others you may have DB2 in the cloud. Or, of course, you may be using BigInsights rather than Cloudant. The point is that IBM wants to give users the option of deploying whatever combination is appropriate for their application, in a way that is consistent across all of their platforms.

Finally, it is worth commenting that IBM has introduced a number of DB2 licensing packages specifically directed at cloud environments, sold with a monthly charge and with predictable maintenance term.

**"Going forward we expect the Internet of Things (IoT) to be a major driving force behind the hybrid cloud, with a lot of collected data being stored and analysed in the cloud, and with the results of those analyses being operationalised within on-premises environments and applications."**

# Summary

Although there are a number of improvements for DB2 in transactional environments in this release, the bulk of new capabilities are focused on analytic environments. In this context, it is worth commenting that IBM has been developing parallel database capabilities for a long time. Of course, that does not necessarily prove anything (Yahoo! has been building search engines for longer than Google) but in this particular case we are inclined to think that this is significant. It should mean, for example, that not only does IBM understand how to make SAP BW work well with DB2 DPF (via parallel query, data distribution, workload optimisation and so forth) but, more importantly for this release, how DB2 BLU on an MPP platform can leverage the same expertise.

That is a specific point, as is the introduction of BLU Acceleration for MPP-based systems, and as is the introduction of the Common Analytic Engine, which will (and already does to some extent) provide SQL compatibility across the IBM portfolio of databases, regardless of whether these are deployed on-premises or in the cloud. More generally, a significant part of this paper has been strategic and theoretical and it has been broader than just considering DB2. We have been discussing IBM's approach to logical data warehousing and the hybrid cloud and, in order to do that, it is necessary to understand the company's strategy. This has meant that we have diverged somewhat from a pure consideration of DB2. However, we make no apology: as a general rule, databases such as DB2 no longer live in glorious isolation. Regardless, this release clearly provides some significant additional features, but it is the direction of travel that is most significant.

**"...as a general rule, databases such as DB2 no longer live in glorious isolation. Regardless, this release clearly provides some significant additional features, but it is the direction of travel that is most significant."**

**FURTHER INFORMATION**

Further information about this subject is available from
*www.BloorResearch.com/update/2298*

## About the author

**PHILIP HOWARD**
**Research Director / Information Management**

Philip started in the computer industry way back in 1973 and has variously worked as a systems analyst, programmer and salesperson, as well as in marketing and product management, for a variety of companies including GEC Marconi, GPT, Philips Data Systems, Raytheon and NCR.

After a quarter of a century of not being his own boss Philip set up his own company in 1992 and his first client was Bloor Research (then ButlerBloor), with Philip working for the company as an associate analyst. His relationship with Bloor Research has continued since that time and he is now Research Director, focused on Information Management.

Information management includes anything that refers to the management, movement, governance and storage of data, as well as access to and analysis of that data. It involves diverse technologies that include (but are not limited to) databases and data warehousing, data integration, data quality, master data management, data governance, data migration, metadata management, and data preparation and analytics.

In addition to the numerous reports Philip has written on behalf of Bloor Research, Philip also contributes regularly to ***IT-Director.com*** and ***IT-Analysis.com*** and was previously editor of both ***Application Development News*** and ***Operating System News*** on behalf of Cambridge Market Intelligence (CMI). He has also contributed to various magazines and written a number of reports published by companies such as CMI and The Financial Times. Philip speaks regularly at conferences and other events throughout Europe and North America.

Away from work, Philip's primary leisure activities are canal boats, skiing, playing Bridge (at which he is a Life Master), and dining out.

## Bloor overview

Bloor Research is one of Europe's leading IT research, analysis and consultancy organisations, and in 2014 celebrated its 25th anniversary. We explain how to bring greater Agility to corporate IT systems through the effective governance, management and leverage of Information. We have built a reputation for 'telling the right story' with independent, intelligent, well-articulated communications content and publications on all aspects of the ICT industry. We believe the objective of telling the right story is to:

- Describe the technology in context to its business value and the other systems and processes it interacts with.
- Understand how new and innovative technologies fit in with existing ICT investments.
- Look at the whole market and explain all the solutions available and how they can be more effectively evaluated.
- Filter 'noise' and make it easier to find the additional information or news that supports both investment and implementation.
- Ensure all our content is available through the most appropriate channel.

Founded in 1989, we have spent 25 years distributing research and analysis to IT user and vendor organisations throughout the world via online subscriptions, tailored research services, events and consultancy projects. We are committed to turning our knowledge into business value for you.

## Copyright and disclaimer

Bloor Research International Ltd
20–22 Wenlock Road
LONDON N1 7GU
United Kingdom

Tel: **+44 (0)20 7043 9750**
Web: **www.Bloor.eu**
email: **info@Bloor.eu**