Bloor

White Paper

White Paper by Bloor
Author **Philip Howard**
Publish date **January 2017**

IBM Informix on Cloud

Siloed and heterogeneous cloud deployments are essentially no different from on-premises deployments: they just happen to be instantiated somewhere else.

Author **Philip Howard**

Introduction

Informix, one of the databases offered by IBM, has a long and illustrious history. It is well known as a platform for supporting applications developed by third parties and ISVs, primarily for use in transactional environments. This is because Informix is a robust offering that combines high performance and functionality with minimal administrative requirements. However, this paper does not discuss these features of the product in any detail. Instead, we are here concerned with the latest release of IBM Informix, which offers the ability to host Informix on the IBM Cloud (SoftLayer and Bluemix).

There are two environments where it might be useful to host Informix in the Cloud. Firstly, because it represents a cost-effective alternative to implementing the database on-premises and, secondly, within hybrid cloud environments. While the first of these is simple enough we need to be clear about what we mean when we talk about hybrid cloud environments before we can sensibly discuss how Informix might fit into this environment.

Informix is a robust offering that combines high performance and functionality with minimal administrative requirements.

Hybrid cloud

A good starting point will be the definition of a hybrid cloud environment. Most definitions you find on the Internet are either misleading or incomplete. For instance, *"a hybrid cloud is an integrated cloud service utilising both private and public clouds to perform distinct functions within the same organisation."* This fails to recognise that many organisations have on-premises environments that are not any sort of private cloud. Another example is *"a hybrid cloud is a cloud computing environment which uses a mix of on-premises, private cloud and third-party, public cloud services with orchestration between the two platforms."* This is accurate enough but it is too broad and says nothing very useful from a data or application perspective. Moreover, the distinction between private cloud and on-premises environments is not a particularly useful one (the former is just a deployment option) and in the remainder of this paper "cloud" is taken to mean third-party cloud environments, whether public or managed. Going beyond this, it is necessary to differentiate between the following:

> ***...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.***

1. Siloed cloud environment.
There are applications, with associated data, running in public or managed cloud environments. There are also applications, with associated data that run on-premises (whether in a private cloud or not – this is not a useful distinction in the current context). These are siloed. This is not a hybrid environment: there is neither integration nor orchestration. The fact that you have some applications running in the cloud and some on-premises does not mean that you have a hybrid cloud environment. Indeed, this is a siloed cloud environment regardless of where things are deployed.

2. Heterogeneous cloud environment.
The on-premises and cloud-based applications exchange information on a periodic basis. For example, at month end. These are still siloed applications but there is a bridge (some form of data integration) between the two. For example, you might be using Workday for your payroll and transferring data from that environment into your on-premises financial application at the month end, but these remain essentially siloed applications that have been integrated. This is a heterogeneous cloud environment.

3. True hybrid cloud environment.
As a counter-example consider master data management (MDM) with a single customer view extended to include the storage and analysis of social media data on each customer, logs that record how that customer has browsed your web site, text details based on phone records and emails of your interactions with that customer, and so on. This extraneous information can be used to extend and enrich the single customer view and it will often – it doesn't have to be – stored and managed in the cloud, while your MDM solution remains on-premises. This sort of environment, in our opinion, represents the true value of hybrid cloud, where the on-premises and cloud data and applications are effectively extensions of one another rather than things that are essentially separate and distinct. This, in our view, is what hybrid cloud is all about.

A hybrid cloud can therefore be defined 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".*

Having thus defined different classes of cloud-based application it should be clear that siloed and heterogeneous cloud deployments are essentially no different from on-premises deployments: they just happen to be instantiated somewhere else. Hybrid use cases, on the other hand, are rather different from traditional on-premises deployments and it is worth considering these differences in some detail. We can do this by discussing use cases.

Hybrid cloud use cases

There are a variety of environments that lend themselves to hybrid cloud deployments. These include:

- Extensions to existing on-premises deployments. Perhaps the most popular of these is extended master data management, as discussed above, but this would also apply to extended CRM (customer relationships management) and other applications that can be augmented in a similar fashion.
- Hybrid analytic environments. While extensions of the type related to MDM and CRM may simply consist of additional detail, it is also common – perhaps more common – to include analytic capabilities against this additional detail. This may range from relatively simple business intelligence for, say, sentiment analysis, to quite sophisticated analytics that may effectively create a hybrid transactional and analytic environment where operational processing is taking place on-premises and analytics in the cloud. Yet another extension to this principle may occur within a logical data warehouse (LDW) where traditional data warehousing is implemented on-premises but additional analytic processing for unstructured data is being carried out in the cloud.
- The Internet of Things (IoT). Many IoT implementations will adopt a hybrid cloud environment, often of the hybrid analytic type just described. However, IoT deployments may be complicated by the fact that sensors, devices and gateways may include their own intelligence. Firstly, this means that IoT environments may consist of a multi-tier architecture rather than just the two-tier hybrid cloud environment we have been discussing. Secondly, sensors and devices will not usually be deployed in the cloud while gateways may or not be cloud-based. Further, even when we talk about edge processing being on-premises that may not literally be the case: it is for this reason that some vendors refer to edge-based equipment being *"in the fog"*.
- Elasticity. Here the idea is to cater for peaks of activity that cannot be anticipated as opposed to month-end and year-end peaks that can be planned for (and which may represent a good case for non-hybrid cloud deployments). For example, Tennis Australia – the governing body for tennis in Australia – uses a hybrid cloud infrastructure to cater for social media-based fan interaction. The problem here is that you can't predict in advance when a match or remark (by a player or commentator) or event will trigger a potential avalanche of comments and tweets. On the other hand, you don't want to impair the quality of the fan experience. The elasticity of a cloud-based solution allows provisioning of extra compute capacity as required, so a consistent experience is provided to fans.
- DevTest. Development and testing processes have very variable infrastructure demands, not only across development projects but also within them. Features such as elastic scaling that support this, are much easier and cheaper to provide in the cloud than in on-premises environments. Similarly, hybrid cloud is proving popular where large media or other files need to be shared across multiple sites or customers. For example, one major game developer uses a hybrid cloud model for sharing software during the testing process. In another example, the same approach is being used by a media management company: internal data is held on-premises for security reasons while data being shared is posted in the cloud.

"The elasticity of a cloud-based solution allows provisioning of extra compute capacity as required, so a consistent experience is provided."

IBM Informix and the cloud

> **IBM Informix has a long history of being embedded in a variety of applications and this suggests that the product is easy to implement and deploy.**

Informix is IBM's preferred database offering for IoT and Bloor Research has previously published a paper discussing IBM Informix and its role in supporting IoT environments. You can read this at: ***http://www.bloorresearch.com/research/white-paper/ibm-informix-and-the-internet-of-things/*** and we do not, therefore, intend to spend much time repeating the comments we made in that paper. However, there are characteristics of Informix which, while suiting it to IoT, also suit it for cloud – and especially hybrid cloud – environments.

Embedding

IBM Informix has a long history of being embedded in a variety of applications and this suggests that the product is easy to implement and deploy. This has implications both for deployment in the cloud and in the fog. In the latter case, footprint is a major consideration and IBM Informix requires just 100 Mb to install on disk. More generally, the main requirements for embedding a database are that you can fire it up and forget about it, that it performs adequately, that it is robust and does not fall over, that it is flexible enough to support different deployment models, and that it is easily updated (without appreciable downtime) when requirements change. In so far as IBM Informix is concerned most of these can be taken as given. However, there are a couple of points worth discussing.

Firstly, it is essential that any embedded database is invisible and remains that way. This is true regardless of whether you are simply collecting data and passing it on (in IoT implementations) or in transactional environments or you are performing some analytics on the data. Particularly with respect to analytics, to get good performance, you need, at least in most environments, to create indexes, materialised views and other such database constructs to achieve the necessary performance. While this is feasible it is not flexible enough when additional requirements need to be supported. Every time you add functionality you need to change the supported indexes. Worse, different workloads may mean that different indexes, materialised tables and so forth will be differently suitable for different customers. Moreover, these workloads may change over time. What this will mean is that the database will need to be tuned on an ongoing basis to maintain performance, which is impractical in IoT environments and undesirable for cloud deployments. For these reasons a traditional relational database will not be suitable for embedding at any level within your infrastructure, precisely because these all require exactly this sort of tuning. Fortunately, this is not the case with IBM Informix because the product not only requires no tuning but also includes self-healing and self-tuning autonomics that handle these requirements automatically. As we discuss in the next section, this fire and forget functionality applies to analytics as well as operational environments.

Warehouse Accelerator

IBM offers various Warehouse Editions of Informix. These all include the Informix Warehouse Accelerator (IWA) to support analytics. This is an extension to the normal database used for transactional purposes and will typically be required in hybrid environments where you need to support both transactional and analytic processing within the same scenario. IWA enables query processing in-memory and provides a column-based approach to avoid any requirement for indexes, temporary space, summary tables or partitions. In other words, it is entirely suitable for supporting analytic applications because the lack of these features means that administration is both minimised and consistent across transactional and analytic environments. The Warehouse Accelerator can be implemented on the same system as the relevant transactional system. For hybrid environments you use Smart Analytics Studio, which is a graphical development tool, to define the data (and its schema) that you want to query and the Warehouse Accelerator will automatically offload this data, which is now stored separately from the OLTP environment.

It is processed in its own memory space so that there is no conflict with the operational aspects of the environment and transactional performance will not be impacted. Note that there is no need to change your existing business intelligence tool(s).

There are several other features worth mentioning:

- The Accelerator uses vector processing. This is a form of processing that takes advantage of modern day CPU characteristics, which is orders of magnitude faster for computationally intensive tasks, which analytics frequently are.
- The database optimiser has been specifically optimised to support both transactional and analytic workloads where a hybrid environment is being deployed. It is also worth noting that the optimiser knows where the data is and where it is not. The optimiser determines whether the query can be satisfied by the Accelerator and, if so, it routes the query there. If not, it will choose to execute the query within Informix, itself. Now, if a query saves the result into a temporary table as part of the Select statement, as is often done by certain BI tools, then the Accelerator can speed up that portion of the query.
- In-database analytic capabilities are available from Fuzzy Logix, which has ported its library of analytic and statistical capabilities to run within the IBM Informix database.
- Informix itself uses what IBM calls "deep compression technology" which provides benefits both in terms of storage capacity and performance. A proprietary encoding method (approximate Huffman encoding) allows predicate evaluation without having to decompress the data. This is a substantial advantage where memory is limited, especially at device or sensor level.

Finally, there are a couple of other points to make. The first is that the environment just described supports both transaction and analytics running on the same system, but also hybrid cloud environments where transactions are hosted on-premises and analytics in the cloud. Or, if you are processing unstructured data (see next) then it could well be that you have both a hybrid transactional and analytic environment in the cloud linked to (structured) transaction processing on-premises. Secondly, in addition to support for unstructured data, IBM Informix also supports (analytic) processing of both time series and geo-spatial data, which we also discuss below.

Unstructured data

There are many environments where support for relational data alone is not enough. Moreover, the sort of data you are collecting and processing may change over time. For this reason, a database that supports a flexible schema will be preferable. IBM Informix provides exactly that by supporting JSON (where each data object has its own schema) as first class objects within the database. This is important not just from a flexibility perspective but also because JSON is commonly used as a data exchange medium within IoT environments, as well as more widely. Moreover, IBM has recognised that there is substantial existing use of MongoDB within both the IoT and cloud communities, so it has implemented API compatibility with MongoDB, a wire listener that acts as a mid-tier gateway between MongoDB and IBM Informix, and the database supports BSON (binary JSON), which is the form of JSON supported by MongoDB. Again, BSON is treated as first-class.

Time series and geo-spatial data

There are many applications that require an understanding of time. In smart metering and other sensor-based environments you will be taking measurements on a regular basis and you need to know what value was recorded and at what time. You also need to be able to handle data that arrives out of time sequence. Bearing in mind that measurements are taken every 15 minutes, say, it would be wasteful to

“Informix itself uses what IBM calls “deep compression technology” which provides benefits both in terms of storage capacity and performance.”

record a time stamp for every single measurement: it is more efficient to simply store the start date and time, and record what the time interval between measurements is. This is what is known as a regular time series. It saves disk space and will provide better performance characteristics when querying the data. On the other hand, in capital markets you need to know when a trade was initiated and when it was completed. Consequently, transactions occur at irregular intervals resulting in an irregular time series, where individual time stamps need to be recorded. IBM Informix has native support for both regular and irregular time series and this is a key feature as such support is not commonly offered by competitive products.

Geo-spatial support is also important in both IoT-based and other environments. For example, in logistics, where you are tracking (and re-routing) delivery vehicles or in insurance (a non-IoT use case) for determining whether houses have been built on a flood plain. Geo-spatial capabilities are more widely provided by vendors than time series support, but are still limited, with many supporting vendors being in the data warehousing space as opposed to those also supporting transactions and operations.

Flexible Grid

Finally, a major consideration for moving to the cloud, regardless of the type of cloud implementation, is for cost savings. Clearly, features such as elasticity (paying only for the capacity you use) and fire-and-forget (thereby minimising administration costs) are important ancillary considerations in addition to the cost savings involved in cloud deployments per se. However, at the same time, users still want to have enterprise-grade software. Of course, "enterprise-grade" is a misnomer: all organisations, of all sizes, want the scalability, performance and reliability that are associated with so-called enterprise-grade software. And this particularly applies when the deployment is mission critical. To support such environments, IBM Informix provides several options, including High Availability Data Replication (HADR), Remote Secondary Standby Database Servers and Shared Disk Secondary Servers. An option that combines the last two of these is known as Flexible Grid. The Flexible Grid supports the definition of a multi-node heterogeneous cluster that makes it possible to run an application on any node within the grid. This means that you can have a geographically dispersed environment with different commodity hardware (and operating systems) implemented in different locations, depending on need, and yet have the whole environment centrally controlled. Not only are DML operations replicated, but so are DDL operations. This means that when a table is created on one node, it is automatically created on all nodes within the Flexible Grid. In addition, it is possible to automatically start replication of the table as soon as it is created, which greatly simplifies the deployment of a replication solution. A major feature of the Flexible Grid is that it supports continuous availability. That is, operations can continue regardless of whether downtime is planned or unplanned. For example, Game Show Network uses the Informix Flexible Grid and – at the point that we spoke with them – had had no unplanned (or planned) downtime for two years. According to Susan Marciano, Vice President of Technical Operations: *"the Flexible Grid feature of Informix enables us to perform rolling upgrades without any outages, so players can go on playing with no interruption and no impact on our revenue. That's worth its weight in gold."* Other use cases occur wherever 24x7 operations are critical and, especially, where costs are a major factor, since the Flexible Grid runs on commodity hardware and is therefore well suited to deployment in the cloud. Minimal administration is a further major benefit along the same theme.

"the Flexible Grid feature of Informix enables us to perform rolling upgrades without any outages, so players can go on playing with no interruption and no impact on our revenue. That's worth its weight in gold."

Conclusion

There are both specific and generic reasons for considering IBM Informix for cloud and hybrid cloud deployments. Specifically, the time series and geo-spatial capabilities of the database lend themselves to IoT-based applications, which will often involve hybrid cloud. The ability to replicate time series and other data from the edge into the cloud will support (predictive) analytics for these environments.

More generically, the argument is that in an ideal world you would like to have the same type of database both on-premises and in the cloud. This means having features that directly support these environments. The most obvious of these are that you need to have support for all types of data (times series, geo-spatial, relational, non-relational) and this in turn means simplified deployment. You will also want the ability to run the same applications across all the various topographical options (on-premises, full cloud and hybrid) that are available. Which in turn means supporting both transactional and analytic processing on the various types of data we have already discussed. In addition, you want the sort of enterprise class software that is suitable for mission critical environments (including enterprise-class security), while also offering features that are complementary to the cloud to help in reducing costs, such as subscription-based pricing where you pay only for what you use. Once you put all this together you realise that there may not be very many database products that can offer all of this. IBM Informix is one that does.

> **...the argument is that in an ideal world you would like to have the same type of database both on-premises and in the cloud.**

FURTHER INFORMATION
Further information is available from
www.BloorResearch.com/update/2312

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 channels.

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**