Bloor

InPerspective

IBM Cloud Pak for Data Deployment Options

"

IBM Cloud Pak for Data is a robust data and AI platform. It is well-suited for implementing a comprehensive data management solution that allows you to take advantage of IBM's broad catalogue of data management services from a single, central location.

"

Introduction

IBM Cloud Pak for Data is a robust data and AI platform. It is well-suited for implementing a comprehensive data management solution that allows you to take advantage of IBM's broad catalogue of data management services from a single, central location. Given the depth of IBM's AI capabilities – many of which are either available explicitly as part of the platform or baked into other services present within it – it also makes for an attractive way to operationalise AI across your environment, up to and including the entirety of data management.

There are a number of reasons this might be appealing. The breadth of capability enabled by the aforementioned range of services is one; the quality of IBM's individual solutions is another; and the interoperability offered by the platform as a whole is a third. However, it is safe to say that one of the key selling points of IBM Cloud Pak for Data (and in truth this goes for IBM's other Cloud Paks as well) is that the options it offers for implementation and deployment are highly versatile and customisable.

In terms of implementation, the platform itself is structured around a selection of core services that IBM considers essential for the product's use. This selection can then be expanded by adding on various extended services, allowing you to customise IBM Cloud Pak for Data to the use cases relevant to you while only paying for what you actually need past the core package. At the same time, the core services by themselves offer an effective, if lean, "all-in-one" solution for data and AI management.

Moreover, this flexibility carries over to the platform's deployment options. As you might imagine given the product's name, it is fully capable of supporting the cloud, including public, private, hybrid and multi cloud environments, and it is *"cloud-native"*, meaning that it has been designed to allow you to take full advantage of the cloud's technological and economic benefits. These capabilities are most apparent in the base product and its deployment options, which include all the aforementioned support for cloud as well as on-premises. Two additional IBM Cloud Pak for Data offerings are also available: IBM Cloud Pak for Data System, which deploys IBM Cloud Pak for Data as an on-premise, hyper-converged infrastructure, and IBM Cloud Pak for Data as a Service, which does so as a managed service on the IBM Cloud, albeit with a constrained selection of specific IBM data and AI services. Additional services for this last option are rolling out on a progressive basis.

In this report, we seek to describe each of the aforementioned deployment options and discuss when they might best suit your business needs. To begin with, we will lay out a summary of IBM Cloud Pak for Data and some of its latest additions, in order to establish what we are dealing with, then proceed to highlight each deployment option and discuss when and why they're most worth using.

> "
> **...it is safe to say that one of the key selling points of IBM Cloud Pak for Data (and in truth this goes for IBM's other Cloud Paks as well) is that the options it offers for implementation and deployment are highly versatile and customisable.**
> "

Product summary

Platform structure

IBM Cloud Pak for Data is an end-to-end data and AI platform that leverages a range of IBM services to help you build a data and AI management solution. Services are provided and deployed as containers within the context of a unified platform, which provides all the usual benefits of both. They are separated into two categories: core services, which are included in the platform without additional cost, and expanded services, which are priced separately. Each of these categories is further subdivided into sections, such as AI, analytics, data governance, data sources, dashboards, developer tools, industry solutions, and storage, and in almost all cases these subcategories are present in both the core and expanded services.

The core services alone form a competent data management solution that includes such capabilities as data cataloguing and governance (provided via Watson Knowledge Catalog); AI and machine learning including model management (Watson Machine Learning, Watson OpenScale, and Watson Studio); streaming analytics (Streams); self-service data preparation (Data Refinery); dashboarding and data visualisation (Cognos Dashboards); data virtualisation (Data Virtualization); and access to or integration with several Db2 products. The expanded services add significantly to this, including a variety of more specific AI and analytics solutions, data integration via DataStage, and more general integration with several third-party products.

Note that the above is very much not a complete list, and that IBM Cloud Pak for Data also provides access to a range of industry accelerators, each of which offers a prefabricated set of configurations and assets designed to provide an out-of-the-box solution for a particular industry. These solutions will not necessarily be complete, but at the very least should let you hit the ground running. The platform also offers third-party support for more than 50 independent software vendors.

We should also reiterate that even aside from the AI services we've mentioned explicitly, AI and machine learning are present in services throughout the platform. In fact, it is clear that enabling and accelerating AI adoption is one of IBM's

core aims with this suite, not just in terms of providing that AI directly but also in helping you to build a foundation of AI-ready data that allows you to implement it effectively.

IBM Cloud Pak for Data is available for purchase through IBM directly or from various online marketplaces, including the Red Hat Marketplace and Azure Marketplace.

Recent developments

IBM has adopted what is essentially a two-pronged value proposition for IBM Cloud Pak for Data: cost reduction on the one hand, and innovation enablement leading to *digital transformation* on the other (a *"survive and thrive"* approach). Not coincidentally, these two propositions are well-suited to, and clearly informed by, current events, which for those of you reading in the distant future refers to the 2020 COVID-19 pandemic. These themes are particular apparent in the 3.5 release, which offers enhanced automation capabilities, simplified platform management and an improved user experience in order to reduce costs, and greater AI, governance and security capabilities, as well as compatibility with additional IBM and third-party services, to reduce risk, address compliance issues, and aid in digital transformations.

Particularly notable new features include:

- Automated query optimisation for data virtualisation
- Data connections that are definable, manageable, and reusable at the platform level
- Centralised *"deployment spaces"* for deploying and managing all of your runtime jobs
- A *"single pane of glass"* administration view for all of your operations from a single location
- A management console for IBM databases running either on or off the platform
- Support for Red Hat OpenShift Operators, enabling automated installs, upgrades and lifecycle management within the platform
- Federated learning that enables you to train your AI models on data sets in disparate locations in the same manner as federated querying (currently in technical preview and due for full release in early 2021)

> **...it is clear that enabling and accelerating AI adoption is one of IBM's core aims with this suite.**

Deployment options

T here are three deployment options available for IBM Cloud Pak for Data: the core IBM Cloud Pak for Data, IBM Cloud Pak for Data System, and IBM Cloud Pak for Data as a Service. In short, the former is the most flexible, offering deployment on public, private, hybrid and multi cloud as well as on-premises; IBM Cloud Pak for Data System deploys Cloud Pak on a high-performance, on-premises, hyper-converged infrastructure; and IBM Cloud Pak for Data as a Service deploys on the IBM Cloud as a fully IBM-managed service, with all the attendant advantages that brings.

IBM Cloud Pak for Data

The core IBM Cloud Pak for Data offering is as a self-managed software platform deployable on-premises or in the cloud, with support for AWS (Amazon Web Services), Microsoft Azure, Google Cloud Platform and IBM Cloud, as well as private cloud, multi cloud, and hybrid cloud environments. It is essentially a sort of 'default' option for IBM Cloud Pak for Data, by far the most flexible and least specialised of the three deployment options available.

That said, it does not offer the performance of IBM Cloud Pak for Data System or the hands-off deployment of IBM Cloud Pak for Data as a Service. Accordingly, you will want to leverage this deployment option when you want to deploy to an environment unsupported by the other provided options, which includes private clouds, public clouds other than IBM, and hybrid and multi cloud environments. You might also opt for it if the trade-offs made by the other options don't appeal to you. For example, you might decide that you prefer the flexibility of the cloud to the raw performance of an on-premises, hyper-converged infrastructure, but at the same time want to leverage functionality not yet available as part of IBM Cloud Pak for Data as a Service. Finally, this option may appeal if you want to maximise your options in the future. For instance, perhaps you want to adopt an on-premises deployment

initially (perhaps even as an intentional stopgap or provisional solution) before a cloud transformation at a later date. Although you could accomplish this with an IBM Cloud Pak for Data System deployment, it's going to be much easier with an option that supports both cloud and on-prem to begin with (and it's also worth noting that IBM takes pains to make this sort of cloud migration relatively easy).

In cases like the above, this core offering should meet your needs. Its strength is its flexibility, and its comparative weakness is only that it doesn't offer the more specialised benefits that the other deployment options do. It is, in effect, the yardstick against which they should be compared against.

IBM Cloud Pak for Data System

IBM Cloud Pak for Data System involves deploying the IBM Cloud Pak for Data software platform to pre-configured and optimised (by IBM) hardware leveraging a hyper-converged infrastructure (sometimes referred to as an HCI). In turn, an HCI is a software platform that is designed to simplify and empower your data centre by unifying your storage, compute, networking and virtualisation resources under a single roof. This has a number of advantages, not least of which is that it allows you to manage all of those resources centrally, and that it enables you to distribute and thus optimise your processes and applications across all of those different areas. HCIs are relatively little-known (hence why we felt the need to give a brief explanation here) but we are told that they are becoming increasingly in mainstream.

The biggest advantage of this methodology is that it results in extremely high performance, suitable for use cases when the flexibility and other advantages of the cloud are less relevant than millisecond-latency and sub-second response times. This is partly enabled by the use of on-premises hardware, but is aided significantly by the fact that the particular software and hardware offered as part of this deployment have been optimised specifically to enable speed

> "
> **The core IBM Cloud Pak for Data offering is as a self-managed software platform deployable on-premises or in the cloud, with support for AWS, Microsoft Azure, Google Cloud Platform and IBM Cloud, as well as private cloud, multi cloud, and hybrid cloud environments.**
> "

> **...the platform comes pre-optimised for handling data and AI workloads with the greatest possible performance while ensuring security, data protection and compliance.**

and low-latency while still offering an out-of-the-box analytics experience at scale, Price-performance and cost effectiveness are also considerable: IBM estimates that the TCO for IBM Cloud Pak for Data System will break even with deployment to a leading public cloud at about 9 months, with it improving in relative terms after that.

In particular, the platform comes pre-optimised for handling data and AI workloads with the greatest possible performance while ensuring security, data protection and compliance. This dovetails nicely with the inherent security advantages of on-prem, and makes IBM Cloud Pak for Data System a good solution if you have data that cannot or should not be moved to public clouds. In fact, the platform provides a trio of predefined adoption patterns tailored to different uses cases: data modernisation, DataOps, and the AI lifecycle. These offer common packages of services that come pre-configured and can help you get started and improve time to value.

Another major benefit of IBM Cloud Pak for Data System is that it allows you to address and modernise your data management infrastructure and applications simultaneously. It also provides a single point of contact for customer support for both software and hardware (and while we're on the subject, IBM provides installation and automatic upgrades as a standard feature, as well as support for business-critical services and built-in redundancy). This puts it squarely in between the other two IBM Cloud Pak for Data deployment options when it comes to utilisation of IBM support.

IBM Cloud Pak for Data as a Service

IBM Cloud Pak for Data as a Service deploys an integrated selection of IBM Cloud Pak for Data services (as of writing, approximately 30 of the approximately 50 – so about 60% – of the services on offer) as part of a fully managed service on the IBM Cloud with subscription and pay-as-you-go licensing options. The most notable missing services are Data Virtualization and a number of Watson Knowledge Catalog features: the core data management solution remains intact (and we are told IBM is planning to implement these services at a later date). On the other hand, it is easy to see why a managed solution might be appealing: managed services are simple, adaptable, hands-off, and offer great time-to-value. This is a potent brew. Moreover, all this makes managed solutions a natural fit for the cloud environment.

To elaborate, the stated reason organisations move to the cloud is often a combination of elasticity and cost-effectiveness. One of the most significant advantages of the cloud is that it allows your deployments to be highly flexible and dynamically scalable, and the aforementioned benefits follow on from these properties. If your system can scale to precisely what you need, when you need it, not only does that benefit your organisation in terms of agility, it also ensures that you're only ever paying for the resources that you're actually using.

Moreover, these benefits are most effectively realised through the use of managed services. By deploying IBM Cloud Pak for Data as a managed service, you enable all of the flexibility and scalability outlined above and thus take full advantage of the cloud and what most organisations want out of it. What's more, by externalising the management of IBM Cloud Pak for Data to IBM itself, you allow the product experts at IBM to handle your deployment instead of doing it yourself, reducing your own organisation's workload and likely your costs. Furthermore, with IBM Cloud Pak for Data as a Service's recent integration with IBM Cloud Satellite – a distributed cloud solution that makes it possible to launch services and workloads across distributed environments – customers can eliminate the need to move or copy across cloud-based deployments.

Popular use cases for IBM Cloud Pak for Data as a Service include

providing a fast and easy way to enable AI throughout your data management lifecycle, as well as using governance and DataOps to break down data silos with relatively little effort. In both cases, its chief benefit boils down to very fast time to value. You could also deploy as a Service as a temporary measure in order to get a solution off the ground quickly, before switching to another deployment style at a later date. However, if IBM Cloud Pak for Data as a Service is a viable deployment option for you to begin with – because its limited services cover your needs, because you don't need the performance of dedicated and optimised hardware, because you don't want to deploy on a private cloud, or a public cloud other than IBM – it is probably also the most appealing in general, in which case we see little incentive to move away from it.

Conclusion

In the course of this report, we have described and discussed IBM Cloud Pak for Data and the three major deployment options it offers. Each option has its place, and you will need to consider the needs of your organisation carefully before deciding which is the best fit. That said, regardless of what your requirements are, IBM Cloud Pak for Data almost certainly provides at least one method of deployment that will work for you.

If you are interested in IBM Cloud Pak for Data, a free trial is available at *https://dataplatform.cloud. ibm.com/registration/ stepone?context=cpdaas&apps=all*

"
Popular use cases for IBM Cloud Pak for Data as a Service include providing a fast and easy way to enable AI throughout your data management lifecycle, as well as using governance and DataOps to break down data silos with relatively little effort.
"

FURTHER INFORMATION
Further information about this subject is available from
www.bloorresearch.com/update/2637

About the authors
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 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.

DANIEL HOWARD
Senior Analyst:
Information Management and DevOps

Daniel started in the IT industry relatively recently, in only 2014. Following the completion of his Masters in Mathematics at the University of Bath, he started working as a developer and tester at IPL (now part of Civica Group). His work there included all manner of software and web development and testing, usually in an Agile environment and usually to a high standard, including a stint working at an 'innovation lab' at Nationwide.

In the summer of 2016, Daniel's father, Philip Howard, approached him with a piece of work that he thought would be enriched by the development and testing experience that Daniel could bring to the table. Shortly afterward, Daniel left IPL to work for Bloor Research as a researcher and the rest (so far, at least) is history.

Daniel primarily (although by no means exclusively) works alongside his father, providing technical expertise, insight and the 'on-the-ground' perspective of a (former) developer, in the form of both verbal explanation and written articles. His area of research is principally DevOps, where his previous experience can be put to the most use, but he is increasingly branching into related areas.

Outside of work, Daniel enjoys latin and ballroom dancing, skiing, cooking and playing the guitar.

Bloor overview

Technology is enabling rapid business evolution. The opportunities are immense but if you do not adapt then you will not survive. So in the age of *Mutable* business Evolution is Essential to your success.

We'll show you the future and help you deliver it.

Bloor brings fresh technological thinking to help you navigate complex business situations, converting challenges into new opportunities for real growth, profitability and impact.

We provide actionable strategic insight through our innovative independent technology research, advisory and consulting services. We assist companies throughout their transformation journeys to stay relevant, bringing fresh thinking to complex business situations and turning challenges into new opportunities for real growth and profitability.

For over 25 years, Bloor has assisted companies to intelligently evolve: by embracing technology to adjust their strategies and achieve the best possible outcomes. At Bloor, we will help you challenge assumptions to consistently improve and succeed.

Copyright and disclaimer