IBM Cloud Pak for Data System with Netezza Performance Server

Highlights

- A data platform with built-in governance
- Simple data management and analysis
- Highly available services for adaptability
- Instant pre-configured software provisioning with hardware
- 100 percent compatibility with Netezza's IBM PureData System for Analytics

Quickly scale and deploy a complete private cloud for your enterprise data and AI architecture

For years, companies have been accumulating data at an impressive pace. Some of this data is being duplicated across various clouds and repositories at astounding levels. This duplication leads to data silos, high costs, delayed projects and increased security risks.

Data users assigned to an analytics initiative spend as much as 60 percent of their overall time trying to locate the data they need. Organizations have existing processes to store data in multiple repositories, clouds and back-end systems and have invested resources to move data to a central location where it can be managed, controlled and made accessible. However, most centralized big data projects still fail to deliver on the promise of easy and controlled data access. Turning that failure to success would allow data scientists to access data across their organization wherever it resides. This greatly reduces the amount of time spent searching for data.

As such, IBM developed Cloud Pak for Data System, a hyper-converged system that combines software, storage, compute and networking. Built on a governed data and AI platform, Cloud Pak for Data System simplifies and unifies the management, governance and analysis of data. It allows you to provision and deploy data services flexibly and rapidly, tailored to specific needs. Users can automatically add discovered hardware nodes and add software from an app-like experience to expand their System.

The Cloud Pak for Data System installation is designed to be as simple as using the software itself.

Hyper-converged technology makes it possible for users to scale and evolve their infrastructures simply and economically as application loads change. Cloud Pak for Data System includes a flexible microservices software architecture and field-programmable gate arrays (FPGAs).

Accelerate time to value with Cloud Pak for Data System

Customers are aware of the value and speed cloud provides to their applications and services and want the same efficiency in their own data centers. Cloud Pak for Data System offers accelerated time to value, allowing an entire private cloud system to be stood up for a data and AI architecture in hours. That's why we also call it a "cloud in a box."

Hyper-convergence enables a software-defined data center to provide cost-effective agility, scalability and security. Customers can quickly develop applications and analyze their data in a security-rich environment within their own firewalls and expand as they need rather than resort to expensive capital outlays. Most importantly, developers don't need to change how they deploy applications, whether they're working in the cloud or their own data center.

When your needs outgrow your existing installation, Cloud Pak for Data System simplifies expansion. Buy small increments of computing or storage capacity as needed, rather than spending excessively on capital. Based on application demand and delivery, budgeting can be completed in quarterly expansions on an ad hoc basis. When the hardware arrives, it's pre-loaded and configured with software and licenses. Once installed in the data center, the existing cluster will automatically discover new hardware and licenses, inherit all user configurations, and be ready to use in a few hours. The platform evolves based on each enterprise's needs and scalability, and it's all managed and highly available to help ensure maximum performance and productivity.

Benefits of bringing data & AI workloads to a "cloud in a box" using Red Hat OpenShift

Natively built with Red Hat OpenShift Container Platform, Cloud Pak for Data System can support seamless migration of containers.

With OpenShift, users can easily write and deploy applications knowing that they'll run on a platform optimized for Red Hat OpenShift. When choosing to deploy a private cloud on premises, Cloud Pak for Data System provides optimized hardware to increase the container performance of the Red Hat cluster while speeding the time to value of data workloads.

Cloud Pak for Data use cases

The Cloud Pak for Data platform delivers on six primary uses cases:

- **Automated AI lifecycle** - Data science teams are looking for integrated systems to manage assets across the AI lifecycle while enterprise CDOs want to ensure the governance of AI models and data associated with them. Cloud Pak for Data enables the end-to-end AI lifecycle all the way from data preparation to building and deploying models to managing them at scale while ensuring governance of data and AI models, including bias detection, explainability, model fairness and drift detection.

- **Data Modernization** - Enterprises are grappling with the proliferation of data, with data distributed across multiple silos, databases, and clouds. Cloud Pak for Data addresses this through governed data virtualization, enabling self-service access to data in real time and including a comprehensive set of capabilities to discover, prep, transform, govern, catalog and access data at scale across all enterprise data sources.

- **Data Ops** - Organizations want to enforce policy controls to allow access to ALL relevant data sets. Cloud Pak for Data enables enterprises to scale their data operations and enforce policies on individual columns and rows, so that assets with sensitive data can still be used. Among other things, this use case includes capabilities around data preparation, data quality, lineage and regulatory compliance.

- **AI for Financial Operations** - Automate and integrate planning across your organization, from financial planning & analysis to workforce planning, sales forecasting and supply chain planning. Enable your organization to deliver more agile, reliable plans and forecasts to drive better business performance.

- **AI for Customer care** - Automating customer care is one of the key enterprise use cases of AI. Among other things Watson helps reduce time to resolution, decrease call volume and increase customer satisfaction. Watson Assistant (WA) can provide AI-powered automated assistance to customers or employees through web/mobile or voice channels and enable human agents to better handle customer inquiries. Watson Discovery (WD) complements Watson Assistant and can help unlock insights from complex business content–such as manuals, contracts and scanned pdfs.

- **Self-service Analytics** - Data is the world's most valuable asset and is a competitive differentiator. Data warehouse & BI is one of the foundational use cases which entails collecting relevant data and building reports or dashboards to derive business insights. Cloud Pak for Data includes everything you need to develop, visualize and analyze data at scale.

The platform also offers *Magento* -- IBM's unique system monitoring and management framework. This framework includes resource managers and a policy engine that, together, provide autonomous monitoring, reporting, and repair of the software and hardware resources of the System.

The additional value of Cloud Pak for Data System

Cloud Pak for Data System delivers on the use cases above with accelerated deployment, optimized performance, and reduced footprints—all behind a firewall on premise in your data center.

Beyond the benefits of the platform, the hyper-converged infrastructure offers proven performance, scalability, intelligence and simplicity for business needs. It helps companies with:

- **Total Cost of Ownership**: At low levels of usage, a public cloud may be cost-effective, but as companies ramp up usage, an on-premises implementation through Cloud Pak for Data System is much more cost-effective.

- **Optimized hardware:** Based on Intel Xeon Scalable processors with optimized software extensions for improved performance and security, Cloud Pak for Data System outperforms commodity hardware, providing the same benefits with less infrastructure.

- **Proximity**: Cloud Pak for Data System brings compute to your data, which results in lower network/data transfer costs and better performance especially as data volumes grow. Many public cloud providers charge each time the data stored in the cloud is accessed or moved. Over time, it can add up to a large sum. This is in addition to the data storage costs in the cloud.

- **Fully Dedicated:** With the addition of Cloud Pak for Data System, you can gain the benefits of the data and AI platform without sacrificing other workloads. This limits the impact that other workloads can have on the platform performance.

- **Continuous Availability/Business Continuity:** Redundancy is built into the System at every level to ensure there is no single point of failure across most hardware and software components.

- **Privacy/Security:** The platform operates behind your firewall and adheres to strict security compliance guidelines, helping organizations keep up with ever-changing data governance regulations.

- **Installation:** IBM provides full installation of the Cloud Pak for Data System at your data center, ensuring that the System is up and running once set-up is complete.

- **Serviceability:** IBM fixes hardware failure at no additional cost. Users can choose to create tickets for hardware replacements, use call-home features and start software upgrades without downtime.

- **Extensions:** Extended services on Cloud Pak for Data allow users to expand the data and AI capabilities of the platform through services such as Data Virtualization, Watson Assistant, Watson Studio and Watson OpenScale.

Netezza Performance Server

On top of the benefits outlined above, another key value-add for Cloud Pak for Data is Netezza Performance Server.

Netezza Performance Server for Cloud Pak for Data allows customers to augment the market leading in-database capabilities of legacy IBM PureData Systems with the full spectrum of analytics and AI solutions modern businesses require. The upgrade to Netezza Performance Server on Cloud Pak for Data and Cloud Pak for Data System is 100% compatible with existing Netezza appliances. Upgrading to the new version is a seamless lift and shift, not a migration. It's as simple as *"nz_migrate."* This saves time, effort and ultimately cost that would have gone toward migrations.

Designed for deep analysis of complex and diverse data volumes scaling into the petabytes, Netezza Performance Server delivers actionable business insight with industry-leading speed and cost of ownership. Customers benefit from in-database analytics and hardware-accelerated machine learning while leveraging Cloud Pak for Data System's powerful AI capabilities such as IBM InfoSphere DataStage, IBM Cognos Analytics and IBM Watson Knowledge Studio. This combination allows customers to see hidden trends in customer behavior, uncover subtle changes in the market for competitive advantage and model the impact of changes made based on this intelligence.

Netezza Performance Server requires minimal administration and tuning both for initial deployment as well as ongoing maintenance. Netezza Performance Server allows data scientists to spend a majority of their time analyzing results and providing insight rather than preparing data.

Netezza Performance Server delivers a distinct performance advantage over other analytic options. Asymmetric massively parallel processing (AMPP)™ architecture combines Red Hat OpenShift and containers running on blade servers and NVMe disk storage with hardware-accelerated data filtering using field programmable gate arrays (FPGAs). This combination delivers fast query performance on complex analytic workloads, providing sophisticated analytics to drive business insight.

Data scientists can build their models using all enterprise data. Netezza Performance Server consolidates all analytics into a single platform where the data resides. Linear scalability with little to no impact on performance allows data scientists and quantitative teams to operate inside the appliance without having to off-load massive data sets to separate infrastructure.

Hardware acceleration allows data scientists to iterate and fine-tune analytical models faster to arrive at the best solution. Once the model is developed, it is seamlessly executed against the relevant data. Prediction and scoring can also be done where the data resides. Users get their predictive scores in near real-time, operationalizing advanced analytics by making it available throughout the enterprise.

Netezza Performance Server supports an extensive array of built-in analytical tools. It is delivered with a library of more than 200 pre-built, scalable, in-database analytic functions that execute analytics in parallel while abstracting away the complexity of parallel programming from developers, users and DBAs. Netezza Performance Server leverages native Cloud Pak for Data System support for Jupyter or Zeppelin notebooks, RStudio and a wide variety of other machine learning environments including SPSS and the Watson Catalog.

The analytics functionality extends to in-database geospatial analytics that are compatible with industry-standard ESRI GIS formats. Integration with existing geospatial analytic environments is simplified as a result. Because data lives in a variety of places and formats, IBM's Fluid Query allows users to query data in multiple data stores: third-party, open source, Hadoop, cloud and remote data sources. Users familiar with Fluid Query will find 100% compatibility in Performance Server.

Companies have accumulated and analyzed data in PureData System for Analytics for years – realizing the value of data-driven business decisions. With Cloud Pak for Data System and Netezza Performance Server, users benefit from a hyper-converged, containerized, microservices "cloud in a box" architecture – bringing the scalability and economics of the cloud to their own data centers. Users can enjoy flexibility with even faster and easier to manage analytics capabilities – a key component to every customer's journey to AI.

With Netezza Performance Server on Cloud, users can now deploy Netezza Performance Server on private clouds and public clouds—such as IBM Cloud, Amazon Web Services (AWS) or Azure—through Cloud Pak for Data.

Why IBM?

- **Simplify deployment** - rapidly stand up a new private cloud data and AI platform, behind the firewall, in a matter of hours.

- **Scale to fit your needs** - get pay-as-you-grow capacity with plug-and-play hardware nodes and an intuitive software app store.

- **Unify IT asset management** - simplify software and system management with a unified and intuitive dashboard.

- **Overcome data silos** - seamlessly connect and integrate data across all clouds.

- **Leverage optimized hardware** - get integrated FPGA hardware acceleration for AI inference that supports common data and AI use cases and workloads.

- **Red Hat in a box** - running natively on Red Hat OpenShift, IBM Cloud Pak for Data System helps enable you to set up a Red Hat-based private cloud.

IBM

Next steps

→ Get started today with a free trial

→ Schedule a free consultation with an IBM expert

For more information

To learn more about Cloud for Data Pak System, visit ibm.biz/CloudPak4DataSystem.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.