CABOT PARTNERS
OPTIMIZING BUSINESS VALUE

Netezza Performance Server: The Leading Enterprise Data Warehouse

Ravi Shankar, Ph.D., M.B.A
Srini Chari, Ph.D., MBA

July 2020

Cabot Partners Group, Inc.
100 Woodcrest Lane, Danbury, CT 06810.
www.cabotpartners.com
info@cabotpartners.com

Sponsored by IBM

Table of Contents

Executive Summary

Advanced analytics and artificial intelligence (AI) are rapidly transforming the economy and society. Applications of these fast-growing technologies enable organizations to predict and shape future outcomes, empower people to do higher-value work, automate decisions and operations, and reimagine new business models.

However, deploying and scaling AI across the enterprise is not easy especially as the volume, velocity, and variety of data continue to explode. What's needed is a well-designed, agile, scalable, high-performance, modern, and cloud-native data and AI platform that allows clients to efficiently traverse the "AI Ladder" with trust and transparency. An enterprise data warehouse (EDW) is a critical component of this platform.

EDWs are central repositories of integrated data from many sources. They store current and historical data used extensively by organizations for analysis, reporting, and better insights and decision-making. Historically, data warehouse appliances (DWAs) have delivered high query performance and scalability, but are now struggling to transform data into timely, actionable insights with the data explosion.

What's needed is a modern EDW with a common, collaborative, cloud-like containerized environment to run all analytic processes from data ingest to insights with choice, flexibility, business continuity, and high-performance. IBM's Netezza Performance Server (NPS) which incorporates IBM Cloud Pak for Data, a leading hybrid cloud data platform, provides all these capabilities and is a key component of a client's system of insights.

With the same code base on-premises or on-the-cloud, NPS radically modernizes IBM Netezza – the industry's first and leading DWA. It preserves Netezza's hallmark of simplicity and protects current investments in tools and training by retaining the same ecosystem of business intelligence (BI), Data Science, and EDW tools. Current Netezza clients can now upgrade to NPS with a single *nz_migrate* command.

We present a comprehensive comparative analysis of the key services at each layer of a modern EDW: Infrastructure, Enterprise Ready, Data Management, Analytics and Modernize. Each service for NPS, Snowflake, Amazon Redshift, Microsoft Azure Synapse, and Teradata is scored and compared based on its capability/strength and customer/community acceptance.

Based on this analysis, we believe NPS which includes IBM Cloud Pak for Data System is outperforming competitors and fast-tracking the delivery of client value in their advanced analytics and AI journey. This is because Netezza Performance Server delivers a secure, hybrid multi-cloud experience with higher performance/scalability, superior enterprise-grade capabilities, smarter analytics, and a better ROI.

The economic impact of AI is $13 trillion over the next decade

Clients must traverse the AI Ladder

EDWs are the bedrock of analytics/AI

DWAs have delivered high query performance and scalability

Data explosion is driving new requirements for EDWs

Innovative technologies are driving EDW modernization and 12% annual growth

Hybrid cloud data platforms are ideal for EDWs

Hybrid cloud data services market is growing 20.53% annually

Hybrid Cloud Crucial for Enterprise Data Warehouse (EDW) Modernization

Analytics and artificial intelligence (AI) solutions are profoundly transforming how businesses and governments engage with consumers and citizens. Across many industries, high-value transformative use cases in personalized medicine, predictive maintenance, fraud detection, cybersecurity, logistics, customer engagement, geospatial analytics, and more are rapidly emerging. The economic impact of just AI is huge: an estimated $13 trillion is expected to be added to the global economy in the next decade.[1]

However, for analytics and AI to become an integral part of an organization, a large percentage of clients indicate that numerous deployment challenges with data and infrastructure must be overcome – data volumes (50%), data quality and management (47%), and skills (44%).[2] Moreover, clients must efficiently traverse the "AI Ladder" with trust and transparency and deploy enterprise data warehouses (EDWs) which are key for analytics/AI.

EDWs are central repositories of integrated data from many different sources. They store current and historical data that is used extensively by organizations for analysis, reporting, deeper insights, and better decision-making. For the last 15 years, a data warehouse appliance (DWA) – a combination of hardware and software designed specifically for analytics with massively parallel processing (MPP) – has delivered high query performance and scalability. The IBM Netezza, the first DWA in the industry, has a large customer base.

Today, DWAs have become expert integrated systems with built-in analytic functions, simplified user experience, and are easy to deploy with no tuning and minimal maintenance. They use modern database capabilities such as MPP architectures, zone mapping, data compression, and more.

But even DWAs are now struggling to transform data into timely, actionable insights to enhance business value especially as the volume and variety of data continue to explode – globally data is expected to grow from 33 zettabytes (ZB or 10^{21} bytes) in 2018 to an estimated 175 ZB by 2025[3]. This data explosion and increasing demands from end-users are driving new EDW requirements to improve agility, balance self-service initiatives, manage complexity, and ensure data quality, security, governance, and compliance[4].

Fortunately, the relentless rate and pace of technological innovations in cloud/edge computing, high-performance systems, and data management are enabling EDW and DWA modernization and expanding their role and importance. Consequently, the worldwide data warehousing market is expected to grow 12% annually from $13 billion (2018) to $ 30 billion (2025)[5].

A modern EDW requires a common, collaborative, cloud-like environment to run all analytic processes from data ingest to insights with choice, flexibility, business continuity, and high-performance. Hybrid cloud data platforms provide this. They combine the benefits of a public cloud with that of an on-premises infrastructure to offer a better EDW solution.

Consequently, the worldwide data services for the hybrid cloud market is expected to grow at a healthy CAGR of 20.53% from 2016 to 2021 as enterprises prioritize a balance of public and private clouds. Only 31% of enterprises see public cloud as their top priority, while a combined 45% of enterprises see hybrid cloud as the future state[6].

---

[1] Tim Fountaine, Brian McCarthy and Tamim Saleh, "Building the AI-Powered Organization", Harvard Business Review, July-August 2019.

[2] Ritu Jyoti, "Accelerate and Operationalize AI Deployments Using AI - Optimized Infrastructure", IDC Technology Spotlight, June 2018.

[3] Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

[4] Melissa Coates, "Designing a Modern Data Warehouse + Data Lake", Blue Granite

[5] https://www.gminsights.com/industry-analysis/data-warehousing-market

[6] https://www.ibm.com/downloads/cas/V93QE3QG

84% of organizations plan to use multi-clouds

Hybrid multi-cloud platforms combine the benefits of a public cloud and on-premises infrastructure to deliver a better EDW solution

EDWs with cloud-native architectures are the strategic direction

Today, large organizations leverage almost five clouds on average. About 84% of organizations have a strategy to use multiple clouds[7] and 56% plan to increase the use of containers[8]. Netezza Performance Server (NPS) which incorporates IBM Cloud Pak for Data is the only major EDW to currently support containers, giving it unique multi-cloud capabilities. Hybrid multi-cloud data platforms have significant advantages in modernizing DWAs and EDWs.

Advantages of Hybrid Multi-Cloud Platforms for EDW Modernization

A hybrid, open, multi-cloud platform allows organizations to take advantage of their data and applications wherever they reside, on-premises, and across many clouds. Here are some key pros and cons of deploying EDWs over on-premises, hybrid, or public clouds (Figure 1):

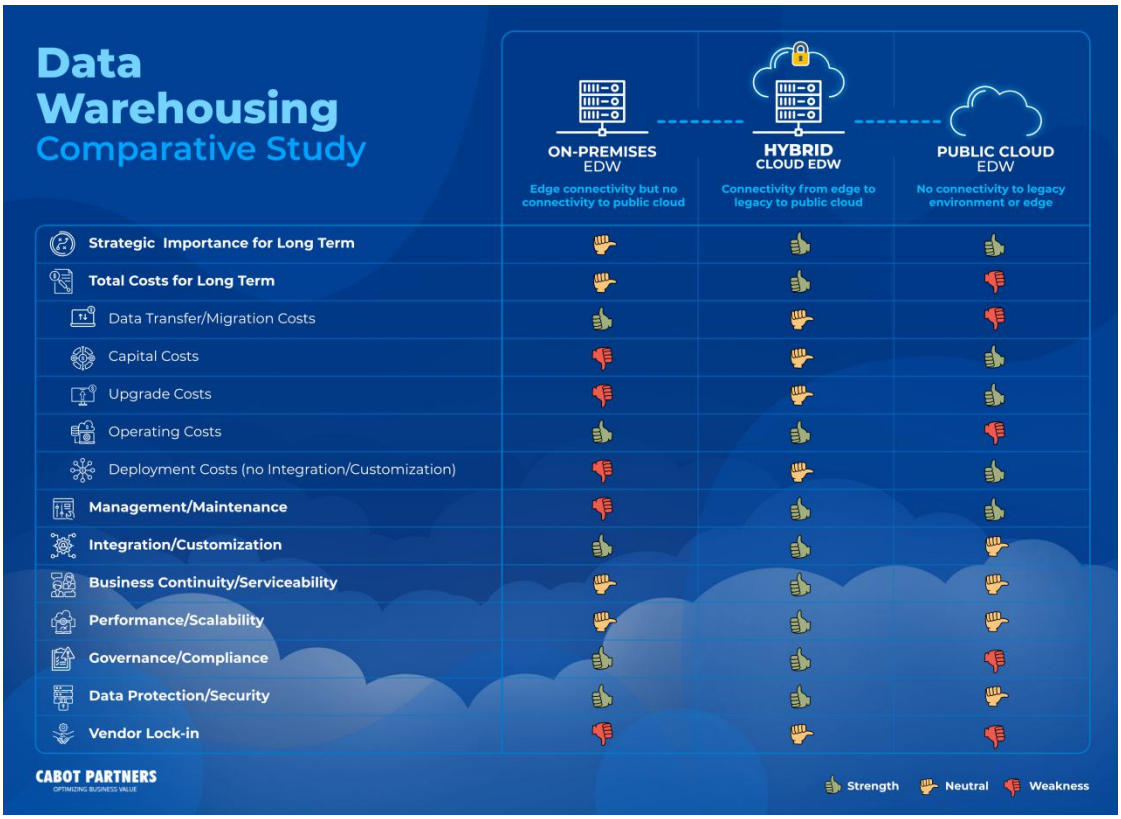

Figure 1: Comparing Enterprise Data Warehouses on On-Premises, Public and Hybrid Cloud

- **Strategic for the long-term:** About 80% of enterprise workloads are still on-premises[9] and still strategic, the public/hybrid cloud is even more strategic driving most of the innovation, growth, and investment in analytics.
- **Total long-term costs:** On-premises costs are predictable and become more favorable with greater utilization. Public cloud costs are unpredictable and good for short, infrequent spiky workloads and consumption-based pricing produces greater accountability of the user population. However, these costs grow steeply with higher utilization typical for most EDWs today. In addition, there are many other hidden costs such as long-term contracts, incremental, supplementary licensing fees,and more.

[7] RightScale® STATE OF THE CLOUD REPORT 2019 from Flexera™
[8] https://www.redhat.com/cms/managed-files/rh-enterprise-open-source-report-detail-f21756-202002-en.pdf
[9] Nagendra Bommadevara, Andrea Del Miglio, and Steve Jansen, "Cloud adoption to accelerate IT modernization", McKinsey & Company, 2018

Hybrid clouds ideal to optimize total costs amidst growing data volume and variety

Data transfer/migration costs grow considerably on public clouds

Capital and upgrade costs are significant for on-premises

Operating costs are higher on public clouds as usage grows

Management /Maintenance are easier on public and hybrid clouds

Integration/ Customization are easier on on-premises and hybrid clouds

With hybrid cloud EDWs, customers can prudently optimize costs using on-premises assets for predictable workloads and offload spiky workloads to the public cloud. This is very effective for the long-term as a smaller on-premises hardware footprint can meet immediate requirements, and incremental needs for resources during peaks can be satisfied by the public cloud. Key components of the total costs include:

- Data Transfer/Migration Costs: For on-premises infrastructure, these are negligible since most of the data for the entire analytics workflow typically reside on-premises. Significant for public clouds since many analytics workflows require substantial movement of data to and from the public cloud. Often enterprises are limited in their ability to move datasets from the cloud back to their on-premises equipment or to another cloud. Moreover, cloud providers charge fees for transferring data out their cloud environment which dramatically increases costs – particularly as datasets continue to grow. Also migrating on-premises workloads to the public cloud is hard and time-consuming.

  In hybrid clouds, there is limited movement of data throughout the analytics workflow to and from the public cloud, and so these costs are low to medium. With consistent cloud-native architectures, migrating workloads from on-premises to public clouds is also relatively easy and less expensive.
- Capital Costs: Significant capital investment for on-premises IT infrastructure is needed to handle peak loads and may result in lower and sub-optimal utilization under normal operations. For public clouds, customer capital costs are negligible. For hybrid clouds, some capital investment for IT infrastructure is needed for certain critical analytics workloads to run on-premises with the rest offloaded to the public cloud. This may result in better utilization and lower capital costs compared to the all on-premises alternative.
- Upgrade Costs: Significant capital expense for hardware upgrades over time are needed to modernize on-premises IT infrastructure to drive innovation. For public clouds, the customer incurs a negligible capital expense for hardware upgrades over time since the provider is responsible for the infrastructure. For hybrid clouds, a modest capital expense for hardware upgrades over time is needed to modernize infrastructure.
- Operating Costs: Since the customer typically owns and operates on-premises assets, costs are predictable and high utilization environments provide better economics than public clouds which are better for short spiky workloads. With a hybrid cloud, the customer can prudently minimize costs by largely using on-premises assets for predictable workloads and offloading spiky workloads to the public cloud.
- Deployment Costs (no Integration/Customization): Significant for on-premises infrastructure, since provisioning and deploying resources and analytics workflows take more time and effort. Whereas costs are low on public clouds with faster provisioning and deployment as the process is automated. On hybrid clouds, costs are significant since connectivity between on-premises and public cloud and maintaining two environments could add another layer of complexity. However, this could be alleviated with a consistent cloud-native containerized architecture.

- **Management/Maintenance:** Moderately hard for on-premises since customers must invest in scarce skills and resources to maintain and operate these environments. Much easier with public clouds since customers typically can use a centralized portal with process automation. For hybrid clouds, it is relatively straightforward for customers to maintain and operate with the right pre-determined operating policies and procedures for workload placement on-premises or on-the-cloud.
- **Integration/Customization:** Easier for on-premises customers to customize and integrate newer solutions with their legacy solutions. This is harder to do on public clouds. On hybrid clouds, it is easier to integrate legacy systems with newer custom solutions from the edge to multiple clouds seamlessly.

Hybrid clouds deliver the best performance and scalability

Governance / Compliance operations can be tailored end to end on hybrid clouds

Hybrid clouds protect sensitive data with minimal vendor lock-in

Hybrid multi-clouds provide considerable choice and flexibility for EDW modernization

Netezza Performance Server integrates database, server, storage and advanced analytic capabilities to accelerate a client's analytics/AI journey

- **Business Continuity/Serviceability:** It can be tailored to provide higher service level agreements (SLAs) for on-premises customers. It is harder to do for public clouds, but they can deliver excellent business continuity. Hybrid clouds can provide high SLAs and excellent business continuity even with disasters.
- **Performance/Scalability:** Excellent performance on-premises with hardware accelerators, faster storage, and proximity to data, but harder to scale to address new business requirements. Lower performance for large-scale analytics on public clouds since maintaining data proximity is hard and optimized storage and computing infrastructure are typically not available. But public clouds can easily scale to meet new business requirements for smaller data sizes. However, as data sets continue to grow exponentially, beyond a few 100s of terabytes, these environments have limited elasticity. Hybrid clouds have excellent performance with hardware accelerators, faster storage, and proximity to data either on-premises or on-the-cloud and can also easily scale to meet new business requirements.
- **Governance/Compliance:** Excellent for on-premises since these operations can be tailored to meet individual enterprise and regulatory requirements. Public clouds have limited ability to tailor these operations for individual customers since they are set broadly by the cloud provider. Hybrid clouds are excellent since these operations can be tailored to meet individual enterprise and regulatory requirements consistently end-to-end.
- **Data Protection/Security:** On-premises and hybrid clouds are excellent since sensitive data can be stored and managed for individual customer requirements and protocols. Public clouds are somewhat vulnerable since their infrastructure is shared and many enterprises are reluctant to part with their mission-critical data.
- **Vendor Lock-in:** Strong for on-premises and public clouds especially with the underlying software infrastructure. Also, data migration to an alternate solution is complex and expensive.

A hybrid multi-cloud environment empowers customers to experiment with and choose the tools, programming languages, algorithms, and infrastructure to build data pipelines, train and make analytics/AI models ready for production in a governed way for the enterprise, and share insights throughout the workflow.

IBM Netezza Performance Server combines DWAs with hybrid multi-clouds to modernize EDWs

Netezza Performance Server – a Hybrid Cloud Modernized EDW

Netezza Performance Server is a purpose-built, standards-based, data, and AI system that integrates database, server, storage, and advanced analytic capabilities into a single, easy-to-manage hybrid cloud platform. Powered by IBM Cloud Pak for Data, it can scale (Figure 2) from small to multi-petabyte capacity to accelerate a client's analytics/AI journey: Collect, Organize, Analyze, Infuse and Modernize[10].

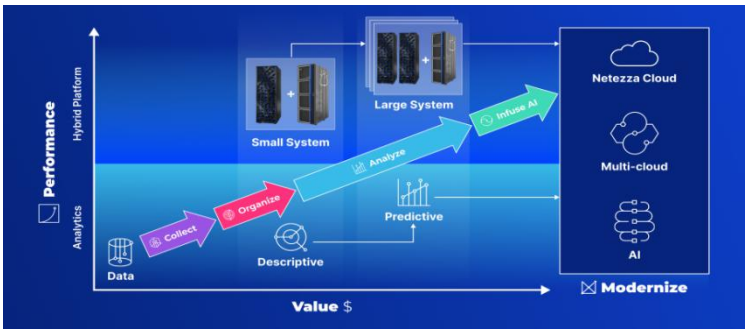

Figure 2: IBM Netezza Performance Server Accelerates the Analytics/AI Journey

[10] https://www.ibm.com/analytics#your_journey

Netezza Performance Server is powered by IBM Cloud Pak for Data and supports multiple clouds

Simplifies analytics by consolidating all analytic activity to where the data resides with unique data virtualization

Single, modernized and unified GUI and extensive library of analytic tools and programming languages

More performance and scalability

Architecture: Netezza Performance Server's cloud-native MPP platform architecture (Figure 3) is built on IBM Red Hat OpenShift and provides a significant boost in performance and scalability without the need to copy the data into a separate analytics server. It supports multiple clouds – currently Amazon Web Services (AWS) and IBM Cloud with support for other clouds are expected in the future.

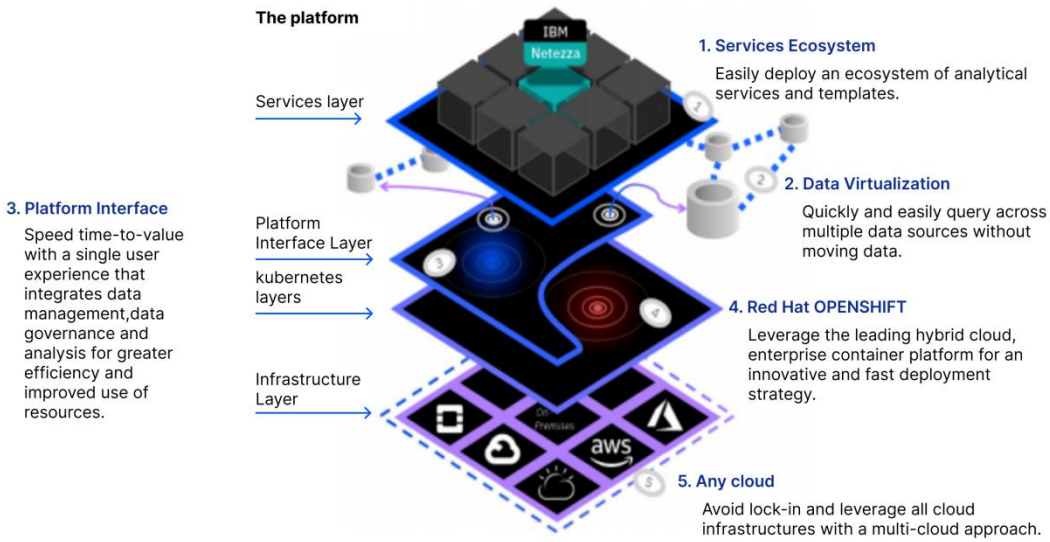

Figure 3: Netezza Performance Server Platform Architecture

Key benefits of the Netezza Performance Server include:

- **Simple/Smart:** It integrates easily with leading applications through standard ODBC, JDBC, and OLE DB interfaces and is 100% compatible with a customer's existing data integration and BI tools. It profoundly simplifies analytics by consolidating all analytic activity to one place: where the data resides. In addition, the data virtualization capability that comes with IBM Cloud Pak for Data enables users to query any data source, anywhere, without moving any data—letting users treat many databases as one.

  A single, modernized, and unified GUI helps administrators monitor and manage system resources, administer database objects, configure workload management, view active sessions, and monitor system resource utilization for capacity planning. For data scientists, it offers a built-in analytical infrastructure and an extensive library of statistical and mathematical functions, supporting a breadth of analytic tools and programming languages, including Open Source R. It is delivered with a library of more than 200 prebuilt, scalable, in-database analytic functions that execute analytics in parallel while abstracting away the complexity of parallel programming from the developers, users, and DBAs. One noteworthy key differentiator is its in-database geospatial analytics that is compatible with industry-standard ESRI GIS formats. This enables easy integration with existing geospatial analytic environments.
- **Better Performance/Scalability:** Its unique asymmetric MPP, hybrid columnar acceleration assist, and workload-optimized storage provide significant advantages on-premises or on public clouds while minimizing data movement. State-of-the-art hardware such as faster cores and advanced NVMe flash drives enable faster analytics/AI at scale. To enhance performance on-premises, it leverages a field-programmable gate array (FPGA). On the cloud, it uses additional workload-optimized compute cores.

High availability and resilience with unique backup and restore

- **Enterprise Ready:** The built-in automation and auto-recovery provide a highly available and fault-tolerant deployment with minimal human touch, ensuring the EDW is available 24x7x365. It can access a customer's cloud data lake, whether on IBM Cloud Object Store or others (e.g. AWS S3). Backups can also be scheduled and kept in object stores as well as replicated to multiple availability zones (Figure 4). Architected for high availability and resiliency, failure of a processing node or any element in the system causes no significant performance degradation ensuring a robust, production-ready environment throughout the lifecycle. It is also fully supported by IBM.

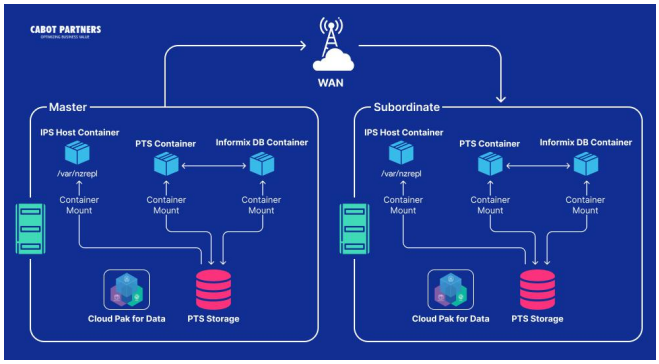

Figure 4: Netezza Performance Server Replication Architecture for Backup and Restore

NPS always includes IBM Cloud Pak for Data and this helps improve the ROI and value for clients.

Improve ROI with IBM Netezza Performance Server

TVO framework caused by Technical/ Business and Direct/ Derived cost and value drivers

The Total Value of Ownership (TVO) framework (Figure 5) illustrates the value of NPS. It goes beyond just the Total Cost of Ownership (TCO) and categorizes interrelated cost/value drivers (circles) for analytics/AI by each quadrant: Costs, Productivity, Revenue/Profits, and Risks. Along the horizontal axis, the drivers are arranged based on whether they are primarily Technology or Business drivers. Along the vertical axis, drivers are arranged based on ease of measurability: Direct or Derived.

Key cost/value drivers identified and estimated holistically

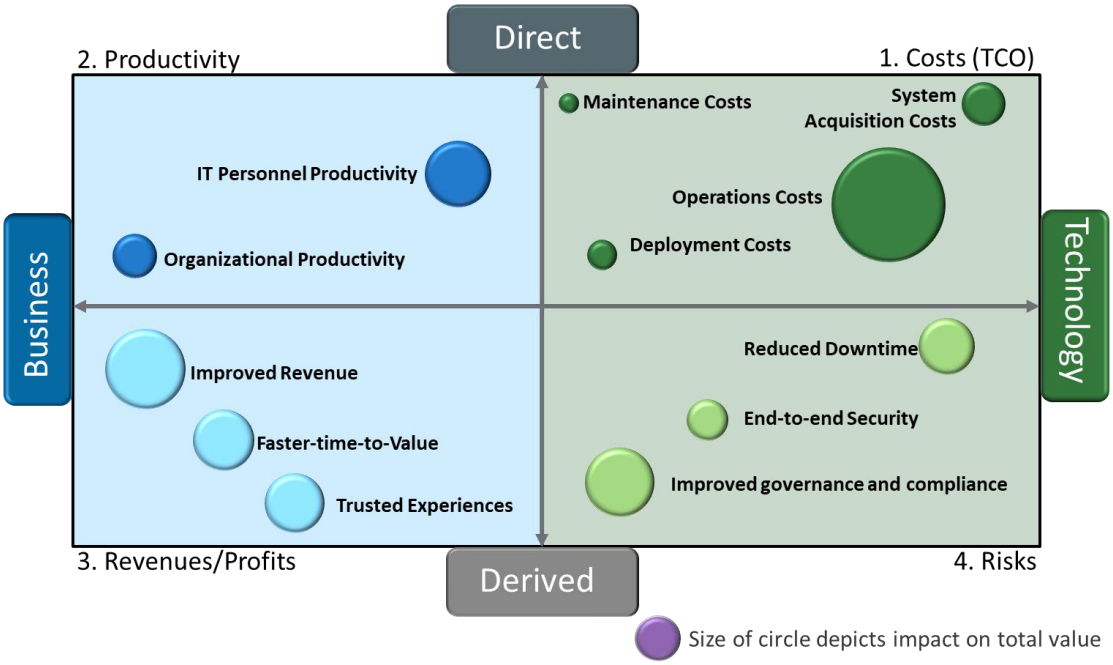

Figure 5: TVO Framework for Analytics/AI with Cost/Value Drivers

EDW operations costs typically are very high

NPS minimizes financial penalties separation in public clouds

NPS allows customers to prudently handle spikes, minimize usage costs and avoid pausing

Public clouds typically incur significant data transfer/migration costs as data volumes grow

NPS enhances productivity with a modern architecture/ technology, a unified GUI, greater flexibility and less data movement

The cost/value drivers for analytics/AI are depicted as circles whose size is proportional to the potential impact on a client's Total Value (Benefits – Cost) of Ownership or TVO as follows:

1. **Total Costs of Ownership (TCO):** Typical costs include one-time acquisition costs for the hardware and deployment, and annual costs for software, maintenance, and operations. For highly used analytics/AI environments, **operations costs** can be very high. All the operating cost benefits of hybrid clouds described earlier apply to NPS. In addition, NPS offers big advantages over competitors to address:
   a. Workload separation: Traditionally, EDW implementations suffer cost, performance, SLA, and predictability issues because of resource (I/O and CPU) contention and collision among competing workload elements. Public cloud EDWs such as Snowflake achieve workload separation by assigning separate compute and disk clusters of various sizes (Virtual Data Warehouses – VDWs) to specific workload elements i.e. BI Reporting, ETL, Data Science, etc. However, this can become inefficient and expensive as these workload elements grow in variety and size. NPS supports software features optimized to fully exploit modern hardware stacks to manage contention irrespective of the workload environment and don't carry the financial penalties associated with workload separation.
   b. Usage costs: Today's EDW's find themselves as anchor tenants in a much larger analytics ecosystem where the loading, processing, and enrichment of data puts a higher processing burden and increases usage. For cloud-based EDW users, this can be costly. Better workload management on NPS allows customers to prudently minimize costs by combining an on-premises system for predictable workloads and offloading spiky workloads to the public cloud.
   c. Spikes and compute pause: Sporadic and unplanned demand spikes challenge EDW operations and can adversely impact end-user satisfaction. Many cloud EDW solutions support concurrent spikes by spinning up additional processing clusters. This additional compute cost is billed to the customer and can be very large. So, some cloud-based EDW providers i.e. Snowflake offer customers the ability to pause compute when there is no activity. However, EDW workloads offer little opportunity to pause. In addition, pausing deletes the cache and significantly deteriorates performance when restarted. The hybrid cloud NPS with workload management can address this without pausing.
   d. Data transfer/Migration costs: Many cloud-based analytics/AI workloads require substantial data movement to and from the cloud. Cloud providers typically charge fees to move data out of their cloud. This quickly adds up as data volumes grow. Also, migrating on-premises/legacy data to the public cloud is hard and time-consuming. Since NPS has the same code base on-premises or on-the-cloud, all analytic activity can be consolidated to where the data resides, minimizing data transfer/migration costs. Further, there is no expensive extract, transform, load (ETL), or migration when the data is needed elsewhere within a similar system.

2. **Improved Productivity:** The TVO model quantifies the productivity gains of data scientists, data engineers, applications developers, and the organization. All productivity benefits of hybrid clouds described earlier apply to NPS. In addition, NPS offers big advantages over competitors to enhance productivity with:
   a. Architecture/Technology: The NPS cloud native MPP architecture is built on IBM Red Hat OpenShift and provides a significant boost in performance and scalability without the need to copy the data into a separate analytics server. State-of-the-art hardware such as faster cores and advanced NVMe flash drives enable faster analytics/AI at scale. To enhance performance, FPGAs are used on-premises and additional workload-optimized cores are used on-the-cloud.

b. GUI: A single, modernized, and unified GUI helps administrators monitor and manage system resources, administer database objects, configure workload management, view active sessions, and monitor system resource utilization for capacity planning.
c. Data gravity: As datasets grow larger, they become harder, more laborious, and expensive to move. So, it makes sense to minimize data movement and move applications and processing to the data where it resides at the edge, on-premises, or on the cloud.
d. Flexibility/Agility: NPS operates with equivalent functionality on-premises and on-the-cloud; avoiding vendor lock-in and accelerating value. Clients can run on virtualized and elastic compute infrastructure which is primarily Kubernetes based. Compute can be separated from storage with a solid network backbone.

3. **Revenue/Profits:** Faster time to value with better performance-optimized parallel and in-memory processing and a single open, flexible, and agile platform. Greater innovation and better decision-making capabilities spur growth, revenues, and improve profits. NPS offers big advantages over competitors to improve revenues/profits with:
   a. Scalability/Insights: NPS can be deployed initially for the right-sized environments for the data volumes and workloads and as data volumes grow, the architecture can scale quickly and easily as the growth needs change and deeper insights with trusted data are needed.
   b. Build once and deploy anywhere: NPS can be deployed faster – on-premises, on-the-cloud, or as a hybrid solution. A portable, container-based platform (Red Hat OpenShift) speeds up the iterative process to transform the business and move to the cloud and then deploy across multiple clouds.
   c. Innovative features: For data scientists, NPS offers a built-in analytical infrastructure and an extensive library of statistical and mathematical functions, supporting a breadth of analytic tools and programming languages, including Open Source R. It is delivered with a library of more than 200 prebuilt, scalable, in-database analytic functions that execute analytics in parallel while abstracting away the complexity of parallel programming from the developers, users, and DBAs.
   d. Management services: A consistent set of services helps streamline operations and manage costs as workloads scale. Secure and up-to-date software from a trusted source gives the business the ability to deploy production workloads with both confidence and speed.

NPS increases revenues/profits with deeper insights, better deployment, greater innovation and consistent management services

4. **Risk Mitigation:** Lower risk of project failure (even well-planned Analytics projects have up to 60% failure rate[11] with a streamlined workflow with reusable components, better ecosystem, and business/IT collaboration and enhanced security/ privacy. It is improved governance with better data cleansing/quality and process consistency. NPS offers big advantages over competitors to mitigate risks with the following additional features:
   a. Availability: The built-in automation and auto-recovery provide a highly available and fault-tolerant deployment with minimal human touch, ensuring the EDW is available 24x7x365. Failure of a processing node or any element in the system causes no significant performance degradation ensuring a robust, production-ready environment throughout the lifecycle.
   b. Backups: Can be scheduled and kept in object stores and replicated to multiple availability zones.

NPS mitigates downtime, security and governance risks

The simplicity, ease of development and deployment, and out-of-the-box performance with no indexing and tuning and minimal administration through the lifecycle lower TCO. The ecosystem of BI, Data Science, EDW tools remain the same and can coexist with existing Netezza customers. This lowers the TCO and maximizes ROI. NPS has all the key features needed for a well-architected modern EDW.

NPS delivers better ROI with simplicity and the same familiar ecosystem of analytic tools

[11] Why big data projects fail and how to make 2017 different, Expansion of Gartner's prediction that 60% of big data projects fail; By Sameet Agarwal, Network World Feb 16, 2017.

The High-Level Architecture of a Modern EDW

Figure 6 depicts a high-level stack architecture of the key layers needed for EDW modernization.

The five key layers of a modern EDW

Figure 6: A High-Level Stack Architecture of a Modern Enterprise Data Warehouse

Key functional layers of a modern EDW: Modernize, Infrastructure, Enterprise Ready, Data management and Analytics

Key layers in the stack are briefly described – starting with Modernize:

1. **Modernize:** EDWs must offer a modern, flexible and extensible platform that is efficient, agile, and support features such as data virtualization, containerization, multi-cloud management, self-service, and enable faster and easier migration of data.
2. **Infrastructure:** In a modern EDW environment, data moves from platform to platform (i.e. on-premise, hybrid and public cloud) frequently as users repurpose it for multiple use cases, making it difficult to govern data and track its lineage. To maintain such an environment, it is important to have an infrastructure that is robust, scalable, and flexible.
3. **Enterprise Ready:** EDWs ensure collaboration across the enterprise by guaranteeing that data is provided with relevant business context. For this, it must comply with relevant security, compliance, service level, and governance standards.
4. **Data management:** A critical EDW component is the ETL process which consolidates data from multiple sources and transforms it into a useful, modeled, and consistent format. Other key services in this layer include governance and support for big data, mixed workloads, and more.
5. **Analytics:** This leverages large volumes of disparate data in the data warehouse for discovery, self-service analytics, dashboards/visualization, machine learning, in-database analytics, and more.

Figure 6 provides the framework for comparing the Netezza Performance Server with similar offerings.

Key Enterprise Data Warehouse Products Used for Comparative Analysis

The focus of this analysis is on major EDW solution providers with the breadth of services/offerings spanning both on-premises and cloud deployments. In addition to IBM Netezza Performance Server, for this comparative analysis, we chose the following four EDW solutions (vendor's summary descriptions):

- **Snowflake** is an analytic data warehouse provided as Software-as-a-Service (SaaS). Snowflake's data warehouse is not built on an existing database or "big data" software platform such as Hadoop.

The Snowflake data warehouse uses a new SQL database engine with a unique architecture designed for the cloud. Snowflake is not available as an on-premise appliance.

- **Amazon Redshift** is a fully managed petabyte-scale cloud-based data warehouse product designed for large scale dataset storage and analysis. It is also used to perform large scale database migrations. Each Redshift cluster runs its own Redshift engine and contains at least one database. Amazon Redshift is available only on the AWS cloud and is not available as an on-premises appliance.
- **Microsoft Azure Synapse** is an analytics service that brings together enterprise data warehousing and big data analytics. It gives users the freedom to query data using either serverless on-demand or provisioned resources—at scale. Azure Synapse is available only on the Microsoft Azure cloud and is not available as an on-premises appliance.
- **Teradata Vantage** is an analytics platform that combines open source and commercial analytic technologies to enable descriptive, predictive, and prescriptive analytics. As a unified analytic and data framework, it contains a cross-engine orchestration layer that pipelines the right data and analytic request to the right analytic engine across a high-speed data fabric. It is available as an on-premises appliance and on major public cloud providers.

Key competitors considered are: Snowflake, Amazon Redshift, Microsoft Azure Synapse and Teradata Vantage

Comparative Analysis Methodology/Results by Each EDW Framework Layer

We used the EDW framework (Figure 4) for the comparative analysis and evaluated key services in each layer: Infrastructure, Enterprise Ready, Data Management, Analytics and Modernize. The information used was obtained from recently published articles (see Appendix for more references). It was also validated through a rigorous process of interviewing subject matter experts and practitioners with hands-on experience in building and deploying data warehouse solutions on each of the five providers.

This analysis evaluated in detail and graded services at each layer of the modern EDW with extensive research and interviews

For every alternative, each service listed in the following tables for every layer was graded on a sliding scale of 0 – 5 (0 is the least and 5 is the best) based on the capability and the strength of services as well as customer/community acceptance of that service. The axis of the spider chart for each layer and alternative depicts this score for each service. These scores for each service starting with Modernize are then aggregated to derive the final score for that layer of the EDW alternative.

Modernize: The services and comparison features for the Modernize layer are shown in Table 1

The Modernize layer – building a modern EDW architecture that provides choice and flexibility across the enterprise

Services	Comparison Features
Data Virtualization	Out-of-the-box or native solution, variety of sources supported, security and governance, Data Federation
Containerization	CI/CD Pipelines, Docker Images for drivers and extensions, Database engine and services decoupled from hardware
Cloud Management	Centralized portal for monitoring and maintenance activities
Multi-Cloud	Support of on-premises and major public clouds including IBM, AWS, Azure, Google and others
Self Service Capability (Data Preparation and Visualization)	Self-service analytics, Self-service Visualization, Data Preparation, Data Store
Migration	Backup and restore, Migration tools and support, Compatibility, Complexity and effort involved, Ability to migrate from on-premises DW to Cloud
Data Lake	Integrate diverse data, Governance, Security, Self-service queries

Table 1: Services and Comparison Features for the Modernize Layer

The results of the analysis for the Modernize category are shown in Figure 7.

NPS offers the best flexibility and choice with a modern solution that is efficient, agile and supports features such as: data virtualization, containerization, multi-cloud management, self service and faster and easier migration of data

IBM provides innovative, cloud-native, out-of-the-box services with multiple deployment choices for the Modernize layer

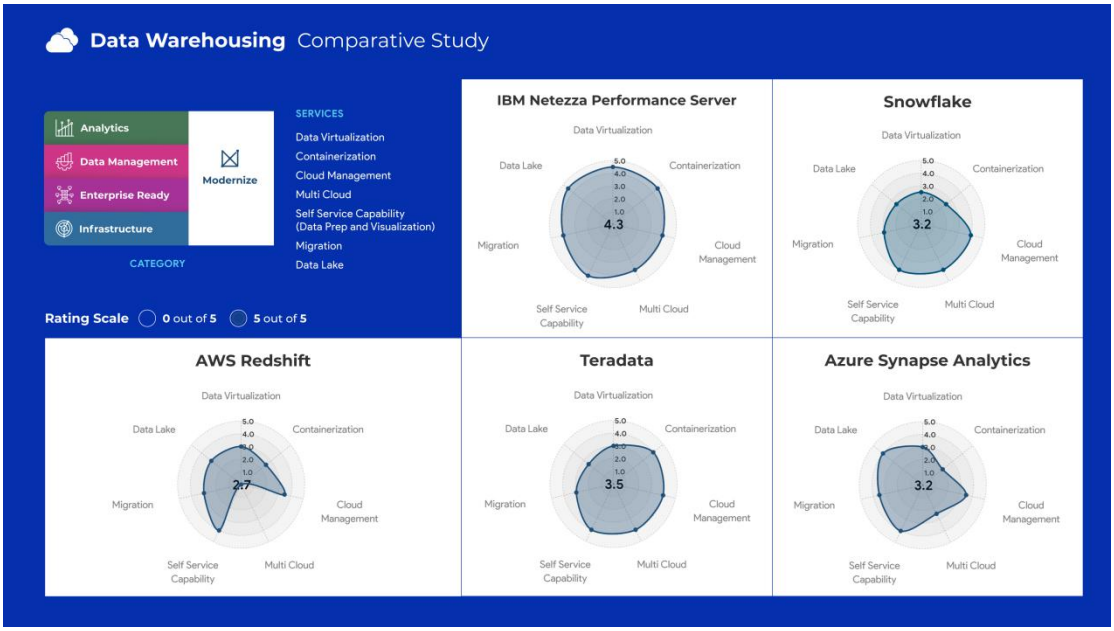

Figure 7: Comparative Analysis for the Modernize Layer

Modernize – Synopsis: Netezza Performance Server (NPS) provides a feature-rich true cloud-ready software-defined environment enabling customers to build composable microservices-based insights. Since it is cloud-native, users can deploy NPS anywhere – on-premises and on hybrid or public clouds such as IBM Cloud and AWS. NPS with IBM Cloud Pak for Data brings virtualization, self-service (data preparation and visualization), and data lake tools required for a modern EDW.

Snowflake has many innovative features, but its use is confined to a public cloud. A significant shortcoming is that, unlike NPS, it lacks containerization and out-of-the-box data virtualization capabilities. AWS Redshift is popular, but its use is primarily limited to the AWS environment. Azure Synapse Analytics has good self-service and data lake integration capabilities and is again limited to the Azure environment. Teradata is feature-rich and it can function as an appliance in an on-premises environment and in the cloud. However, it doesn't have out-of-the-box virtualization capability.

"I've done more POCs than I can count, but moving Netezza on-prem to Netezza on Cloud was the most seamless, free of hiccups, and in exceeding expectations, than any POC I've ever done, or even heard of. Zero query breakage, reports balance to the penny, and blazing boost in speed. Add to this IBM's aggressive, proactive support, and the overall experience was outstanding."

David Birmingham
Principal Solution Architect
Sirius

Infrastructure: The services and comparison features for the Infrastructure layer are in Table 2.

The Infrastructure layer – delivering diverse platforms and tools with reliable high-performance systems and software

Services	Comparison Features
Software	Client tools, Drivers and Plugins, Ability to install software in customer machines
Data Warehouse Appliance	Built in compute (multi-core processors) and ready to run software, Networking and Storage, Optimized for data warehouse and analytics, MPP Architecture, Fully functional data warehouse (DBMS, Connectivity Plugins, Compression, Data Availability & Scalability, Security), FPGA
Storage	Object Storage, Block Storage, Shared File Storage, Archival Storage, Storage Gateway, Data Transfer Service, Geo Spatial Objects and Functions, NVMe solid state drives, flash storage
MPP Architecture (Compute and Storage)	Share vs. Shared Nothing, Compute, Storage
Load Balancing, Multi-clusters	Load Balancing: Replication vs Multi-cluster Modern Solutions:Provisioning time, Centralized Storage, Multi-cluster, Metered usage

Table 2: Services and Comparison Features for the Infrastructure Layer

The results of the analysis for the Infrastructure layer are shown in Figure 8.

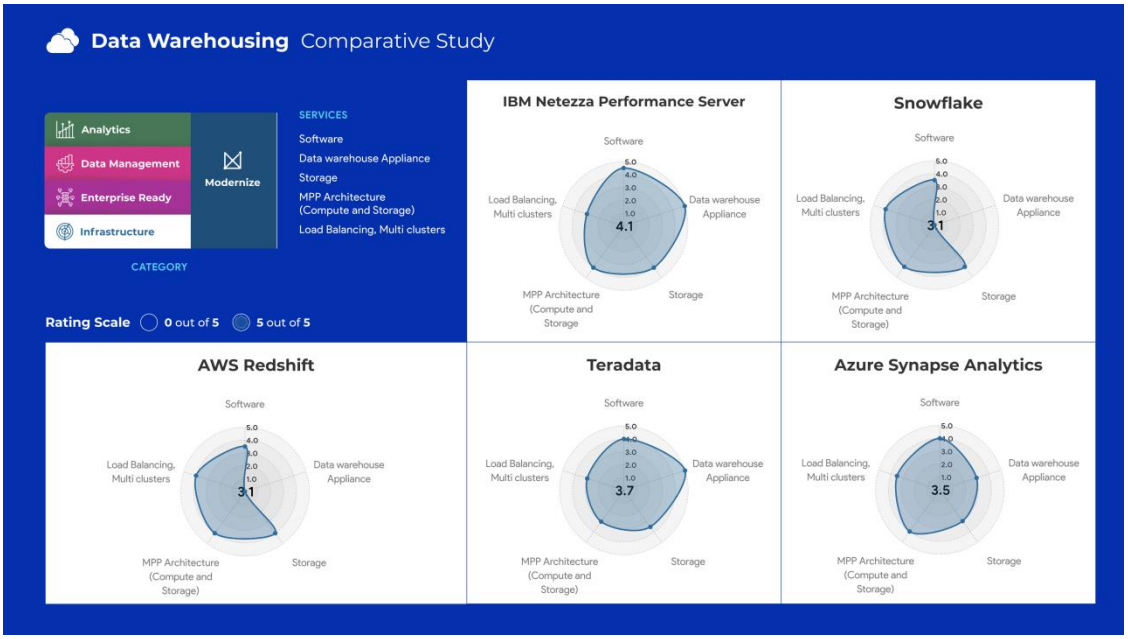

Figure 8: Comparative Analysis for the Infrastructure Layer

NPS offers a common code base on-premises or on the cloud with workload optimized storage for the Infrastructure layer

Infrastructure – Synopsis: NPS and Teradata lead in this layer and provide appliance and cloud solutions. However, Teradata's last major upgrade was in 2017 and its price/performance is weaker compared to NPS. A major advantage of NPS is its common code base across the appliance and cloud platforms. In the appliance mode, additional performance is gained by the FPGA processors resident in the appliance. NPS also has a workload-optimized storage solution that improves load balancing and hence performance.

"Our longest queries returned in a fraction of their original time, and saw no concurrency drag running hundreds at once."

Customer from Healthcare Industry

Snowflake, Redshift, and Azure Synapse analytics mainly function in a cloud environment and do not have appliance-based on-premises solutions. This is a major shortcoming for large enterprises that have legacy data warehouses and require a highly available and scalable on-premises EDW.

The Enterprise Ready layer – addressing governance, security, quality, availability and compliance needs

NPS, being part of the IBM Cloud Pak for Data ecosystem, has enterprise grade containerization capabilities and can be integrated with enterprise management and lifecycle operations

Enterprise Ready: The services and comparison features for the Enterprise Ready layer are in Table 3.

Services	Comparison Features
Security	Integrated security across all layers - Firewall, Policies and rules, SSH, SSL, Authentication & Authorization, Data security, Infrastructure security
Governance	Centralized Portal, Policy Management, Monitoring and control
Compliance	Support of standard compliance - example: SOC1, SOC2, HIPAA
Integration	APIs, Authentication, Integration with other sources, Plugins and extensions
Support and Maintenance	Upgrades, Monitoring, Workload management, Support services
Availability	Availability performance (Downtime in a year), Availability Zones
Scalability	Horizontal Scalability, Provisioning time

Table 3: Services and Comparison Features for the Enterprise Ready layer

The results of the analysis for the Enterprise Ready layer are shown in Figure 9.

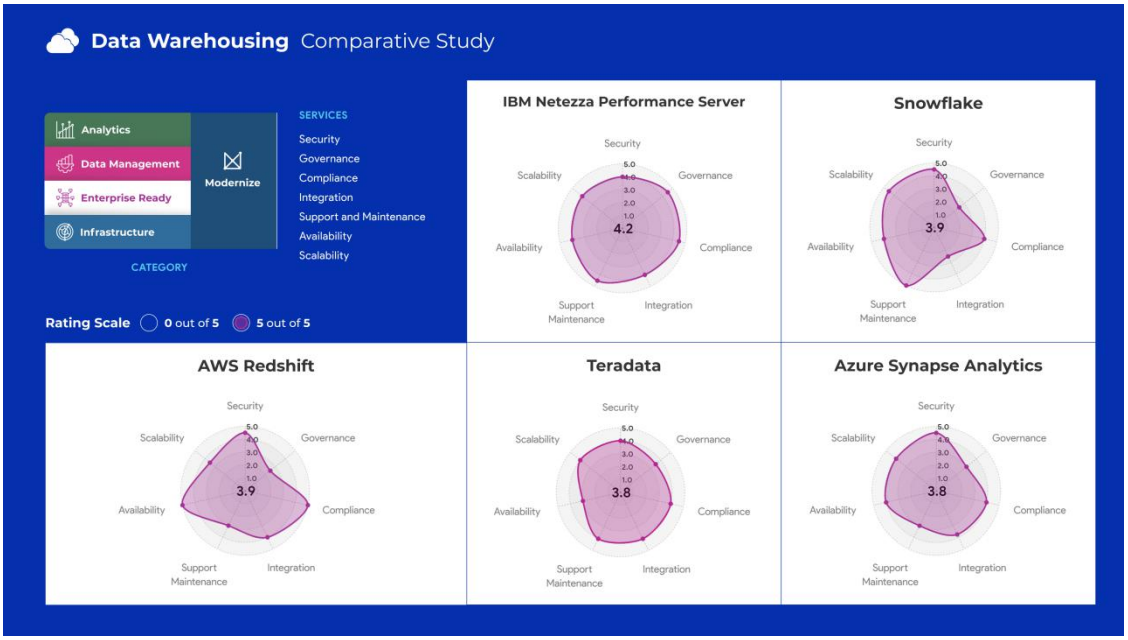

Figure 9: Comparative Analysis for the Enterprise Ready Layer

Enterprise Ready – Synopsis: All solutions are quite competitive in this layer. Netezza Performance Server, being part of the IBM Cloud Pak for Data ecosystem has enterprise-grade containerization capabilities and can be integrated with enterprise management and lifecycle operations. Snowflake and Redshift, being cloud-only solutions, are limited since integration with on-premises infrastructure will result in additional cost and effort. Teradata is a strong contender, However, it lacks an out-of-the-box solution. Synapse as part of the Azure ecosystem is strong, however, it is new to the market and needs additional validation with customers.

Data Management: The services and comparison features for the Data Management layer are in Table 4.

The Data Management layer – consolidating data from multiple sources and transforming it into a useful, modeled and consistent format

Services	Comparison Features
Data Warehouse	Optimized for reads and OLAP, Storage Limits, Concurrency, Compression, Scripting Languages - Procedures and Functions, Performance, Scalability and Integration features, Connectivity with BI, ETL & Visualization Tools, Workload Management, Loading and unloading, In-Memory Analytics, Backup and Restore, SQL Compatibility
ETL/Data Integration	Connectivity plugins and extensions, Tools supported
MDM / Data Quality	Extract data from variety of sources, Data Cleansing & Enrichment, Data Standardization, Define Transformations and Rules, Real time processing and batch processing, Define rules and work flows, Impact Analysis
Data Governance	Clearly defined roles for a Data Governance for frame work (Stewards, Analysts, Data Scientists, Engineers, etc.), Collaboration, Ability to create and maintain Data Catalogue with clear classification and definition of business terms, Ability to define policies and rules, Collaboration
Batch Processing	Bulk Load Utilities, ETL Tools Supported
Real Time Processing	Streaming Utilities, Built in solutions and plugins
Big Data Solutions	Ability to offload processing to Big Data (Hadoop and other services)
Mixed Workloads	Suitable for OLTP & OLAP, Ability to handle transactions (Huge and small volume) and queries (simple to complex) concurrently, Direct Query from BI and Analytics Applications which requires high performance and lower response time

Table 4: Services and Comparison Features for the Data Management Layer

The results of the analysis for the Data Management layer are shown in Figure 10.

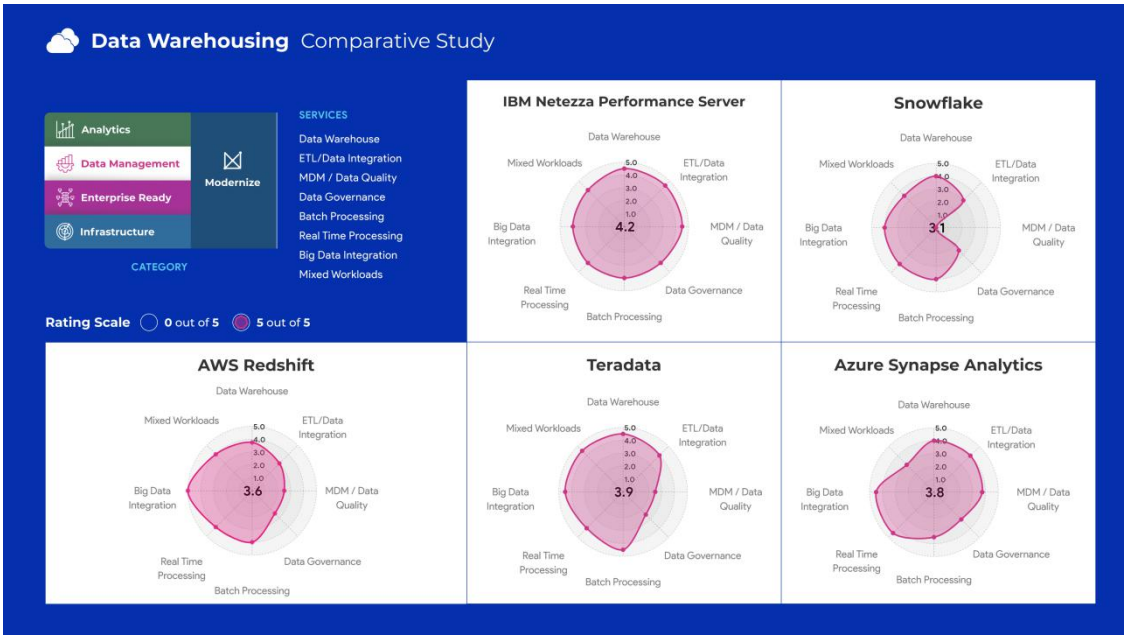

Figure 10: Comparative Analysis for the Data Management Layer

Netezza is a leading EDW for decades with excellent performance and unique ETL, MDM and governance tools

Data Management – Synopsis: NPS is a strong and leading database for EDW implementations for decades. It is supported by most of the leading ETL tools. Netezza was the first DWA in the industry. Its MPP architecture with FPGA provides increased throughput and faster execution time. NPS has utilities to load and unload huge volumes of data. As part of the IBM Cloud Pak for Data framework, it incorporates ETL, MDM, and data governance tools under a single umbrella.

"We ported a six-year-old custom-code application running thousands of ELT operations. Ran first time without a hiccup and finished so fast we thought it had failed."

Leading National Concierge company

The Snowflake data warehouse uses a SQL database engine with a unique architecture designed for the cloud and is offered as Software-as-a-Service (SaaS). It is not available as an appliance. Snowflake processes queries using MPP compute clusters where each node in the cluster stores a portion of the entire data set locally. This approach offers the data management simplicity of a shared-disk architecture, but with the performance and scale-out benefits of a shared-nothing architecture. Virtual data warehouses and zero movement data sharing are some of the notable and unique features of Snowflake. Though it is supported by many other ETL and data governance tools, it lacks out-of-the-box solutions for MDM. Lack of a robust workload management solution particularly for very large datasets presents scalability and governance challenges for Snowflake users. Also, Snowflake uses local NVMe for cache, but its persistent data resides on 100x slower S3/cloud object storage. There are multiple functions that can't be cached, and this limits performance as well as scalability/elasticity.

Amazon Redshift is a data warehouse product that is part of AWS and not available as an appliance. Redshift delivers fast query performance by using columnar storage technology to improve I/O efficiency and parallelizing queries across multiple nodes. It is compatible with PostgresSQL. Redshift Spectrum based on S3 storage provides the multi-cluster capability and supports querying nested data in Parquet, ORC, JSON, and Ion file formats. Redshift has a good workload management solution but needs partner solutions for MDM and data governance. ETL is a challenge with Redshift as AWS Glue ETL is not yet a mature product.

Teradata Vantage is a strong contender in data management. Like Netezza, Teradata is one of the leading big data and EDW solutions for decades. Teradata's MPP and shared-nothing architecture allows petabyte scaling and massive query processing. Teradata also has utilities to load and unload huge volumes of data. Teradata is well known for workload management and its performance – although its price/performance is much weaker than NPS and the last hardware upgrade was in 2017. Teradata's Intelligent Memory architecture gives extended memory for faster retrieval of hot data. It doesn't have strong out-of-the-box ETL, MDM, and data governance tools.

Microsoft Azure Synapse Analytics is new to the industry and evolving as a strong contender in the cloud. It is an MPP data warehouse with shared and distributed storage. It is not available as an on-premises appliance. For an on-premises warehouse solution, customers can use Microsoft PDW (Parallel Data Warehouse) if needed. As part of the Azure cloud, Synapse leverages services and tools of the ecosystem. The COPY command with Azure Synapse Analytics allows loading files from BLOB storage. It also allows CSV, PARQUET, and ORC file formats. Synapse Analytics supports SQL, .Net, Python, Scala, R, and Java. Apache Spark in Azure Synapse Analytics can be used for in-memory processing.

Analytics: The services and comparison features for the Analytics layer are shown in Table 5.

Services	Comparison Features
Dashboards and Visualization	BI Tools Supported, Built in visualization and analytics plugins
Explore/Discover	Automatic Discovery, Search and View Assets, Query and View Data
Data Preparation Tools	Data Sources Supported - Relational, Big Data, Data Lake, Explore/Discover, Transform, Blend, Refine & Enrich, Visualize
Data Mining	Data Preparation, Modeling, Pre-processing and Processing, Mining tasks (Anomaly detection, Dependency Modeling, clustering, classification, regressing, summarization), Validations
In-Database Analytics	Out-of-the-box analytics and mining, SQL Functions and Libraries for various algorithms including but not limited to data exploration, model diagnostic, classification, regression, clustering, time series, etc
Machine Learning	IDEs, Note Books, Languages, Modeling, Big Data Integration, Spark Engine, Data Cleansing

Table 5: Services and Comparison Features for the Analytics Layer

The results of the analysis for the Analytics layer are shown in Figure 11.

Figure 11: Comparative Analysis for the Analytics Layer

IBM has a large and mature portfolio of BI, AI and ML solutions for the Analytics layer optimized for NPS

Analytics – Synopsis: Netezza Performance Server (NPS) is part of the IBM Cloud Pak for Data ecosystem. Moving analytics to the Netezza Performance Server is straightforward with IBM's embedded analytic platform. In addition, the data virtualization capability that comes with IBM Cloud Pak for Data enables users to query any data source, anywhere, without moving any data. IBM Watson Studio—which comes with Netezza Performance Server for IBM Cloud Pak for Data helps data scientists and analysts prepare data and build models at scale across any cloud. The Netezza analytics functionality also includes the in-database geospatial analytics that is compatible with industry standards. The presence of FPGA in the Netezza appliance improves performance considerably.

"Netezza Performance Server enables us to make the most out of our data. By leveraging the solutions capabilities, we are able to reduce the time to generate key reports and make better, more informed decisions in line with our customer needs."

Leading European Bank

Snowflake has a rich portfolio of analytical features and services including from partners. The Snowflake Connector for Spark ("Spark connector") brings Snowflake into the Apache Spark ecosystem, enabling Spark to read data from and write data to, Snowflake. From Spark's perspective, Snowflake looks like other Spark data sources (PostgreSQL, HDFS, S3, etc.). However, this imposes additional infrastructure and network capacity requirements.

Amazon Redshift is part of AWS and as such it leverages analytics capability of SageMaker and other AWS AI services and tools. Visualization capability within AWS is less mature. Redshift does not have in-database capability. Spark on Amazon EMR is a viable option for advanced analytics. However, this necessitates moving data from S3 to HDFS which can be complex.

Teradata Vantage is a strong contender in this category. Teradata Vantage Analyst and Teradata Aster Analytics provide a rich set of analytic tools and services. Teradata Vantage provides in-database analytics using SQL and supports R Studio and Jupyter. Teradata Intelligent Memory speeds data warehouse query performance by ensuring that the most frequently used data is kept in memory.

Synapse studio, a new solution, combines all the analytic services into a single ecosystem. Features such as visualization, machine learning are available within Azure and integrated with Synapse studio. Many features are still in the early stages of release.

Consolidated Comparative Analysis – Comprehensive Summary View

Figure 12 depicts a comparative analysis dashboard followed by a brief summary of each solution

Figure 12: Dashboard of Comparative Analysis

IBM Netezza Performance Server leads in all of a modern EDW

All the EDW solutions considered in this comparative analysis use the latest innovations such as MPP architecture, columnar storage, pipelined execution, just-in-time compilation to get the best results. Naturally, there are differences in a few features and functions. The leaders are those that can combine the benefits of a public cloud with that of an on-premises infrastructure to offer a better EDW solution especially in environments that are legacy data-rich. Data migration is expensive and time-consuming, and an EDW solution that facilitates moving compute closer to the data is ideal.

- **Netezza Performance Server** is a purpose-built, standards-based, data, and AI system that integrates database, server, storage, and advanced analytic capabilities into a single, easy-to-manage platform on-premises as an appliance or on IBM Cloud and Amazon Web Services (AWS). Netezza Performance Server appliance and Netezza on the cloud has the same code base and is part of the IBM Cloud Pak for Data ecosystem. For current Netezza users, the ecosystem of BI, Data Science, EDW tools remains the same both for the appliance and on-the-cloud, protecting the investment in tools and training which leads to a lower TCO. It dramatically simplifies analytics by consolidating all analytic activity to one place: where the data resides. Moving analytics to the Netezza Performance Server is straightforward with IBM's embedded analytic platform. In addition, the data virtualization capability that comes with IBM Cloud Pak for Data enables users to query any data source.
- **Snowflake** data warehouse is an analytic data warehouse provided as Software-as-a-Service (SaaS). Snowflake's data warehouse is not built on an existing database or "big data" software platform such as Hadoop. The Snowflake data warehouse uses a SQL database engine with a unique architecture designed for the cloud. Snowflake is not available as an on-premises appliance. Snowflake defines a virtual warehouse as a cluster of compute resources. This warehouse provides all the required resources, such as CPU, memory, and temporary storage, to perform operations in a Snowflake session. Virtual data warehouses and zero movement data sharing are some of the notable and unique features of Snowflake. Lack of a robust workload management solution presents scalability and governance challenges for Snowflake users. Most of the other services such as machine learning, MDM, ETL require solutions outside the ecosystem which often requires additional network resources and data movement

  Amazon Redshift is a fully managed petabyte-scale cloud-based data warehouse which is part of the larger cloud computing platform AWS and not available as an appliance. Redshift delivers fast query performance by using columnar storage technology to improve I/O efficiency and parallelizing queries across multiple nodes. It is compatible with PostgreSQL. Redshift Spectrum based on S3 storage provides the multi-cluster capability and the ability to query in a multi-cluster environment. Though it has good scalability and availability, for larger implementations, the cost can be high. Redshift lacks mature out-of-the-box ETL and BI tools. Since it is available on AWS, services such as MDM and data virtualization require integration with third-party solutions.
- **Teradata Vantage** is an analytics platform that combines open source and commercial analytic technologies to enable descriptive, predictive, and prescriptive analytics. It is available as an on-premises appliance and supports major public clouds. It brings all the services needed for data warehousing and analytics in a single platform. Teradata Aster is another product in the portfolio which covers services required for an entire analytics lifecycle around data warehousing (data collection/data preparation, analytics - machine learning and mining, query processing, and visualization). Teradata is not very strong in a few out-of-the-box solutions such as ETL, MDM, and BI tools and will require third-party solutions to satisfy these needs.
- **Microsoft Azure Synapse** is an analytics service that brings together enterprise data warehousing and big data analytics. It gives users the freedom to query data using either serverless on-demand or provisioned resources—at scale. Azure Synapse is available only on the Azure cloud and is not available as an on-premises appliance. For on-premises warehouse solutions, customers can use Microsoft PDW (Parallel Data Warehouse. Since Synapse is a part of the Azure ecosystem, it brings all the services needed for the analytics lifecycle under one umbrella. Azure Synapse has a mature set of services for ETL, BI, and Machine learning. The product is quite new and may take some more time to mature.

Compared to alternatives, IBM Netezza Performance Server provides a modern EDW for deeper insights and to accelerate the analytics/AI journey

Conclusions and Recommendations

IBM Netezza Performance Server is the only major EDW that provides a hybrid, multi cloud containerized solution with data virtualization, simplicity and excellent price performance

To maximize value from analytics/AI, customers must collaborate to collect data and make it simple and accessible and organize data to create a business-ready analytics base to analyze this data. The traditional EDW and DWA were designed to do this. However, as the volume, velocity, and variety of data continue to explode, EDWs must be modernized.

Modern EDWs must provide a common, collaborative, cloud-like environment to run analytic processes from data ingest to insights with choice, flexibility, business continuity, and high-performance.

In addition to IBM Netezza Performance Server (NPS), there are several notable modern EDW solutions in development: Snowflake, Amazon Redshift, Microsoft Azure Synapse, and Teradata. But only NPS provides a hybrid, multi-cloud containerized solution with data virtualization, simplicity, and excellent price-performance.

Based on a comprehensive comparative analysis of the services at each layer of a modern EDW, we believe IBM Netezza Performance Server has several advantages over its competitors. NPS is fast-tracking the delivery of client value and insights in their advanced analytics and AI journey by:

- Providing the same code base and consolidating all analytic activity to where the data resides – on-premises or on multiple clouds.
- Monitoring, managing, and administering system resources, database objects, and applications with a single common, unified, modern GUI.
- Supporting an extensive analytic and mathematical library and a breadth of programming tools.
- Offering better performance/scalability with unique asymmetric MPP, hybrid columnar acceleration assist, faster cores, NVMe flash drives, and advanced workload management.
- Delivering business continuity, highly available, and fault-tolerant deployment with built-in automation, auto-recovery, scheduled back-ups, and replication across multiple availability zones.
- Lowering the TCO with simplicity, ease of development and deployment, and out-of-the-box performance with no indexing and tuning and minimal administration throughout the lifecycle.
- Protecting an existing Netezza client's investment since the ecosystem of BI, Data Science, EDW tools remain the same, and upgrading to NPS only requires a single *nz_migrate* command.

IBM Netezza Performance Server lowers TCO and protects an existing Netezza customer's investment with a single command to upgrade

For more information, visit IBM Netezza Performance Server

"We trust IBM's position as a leader in Data Management for Analytics and we really like the performance and the value we saw in Netezza Performance Server. With just one simple command, nz_migrate, we moved our tables, data and stored procedures from our existing Netezza into the new Netezza Performance Server, and we did this all by ourselves. It's that simple."

VP, Information Technology
Leading Global Market Research Company

Cabot Partners is a collaborative consultancy and an independent IT analyst firm. We specialize in advising technology companies and their clients on how to build and grow a customer base, how to achieve desired revenue and profitability results, and how to make effective use of emerging technologies including HPC, Cloud Computing, Analytics and Artificial Intelligence/Machine Learning. To find out more, please go to www.cabotpartners.com.

Appendix

Interviews

We conducted over ten interviews with subject matter experts and practitioners listed in the table below. In addition, we also interviewed several IBM employees.

Name	Role
Parsu Nurani	Vice President, Software Development
Ajay Asthana	Managing Principal, Analytics
M.R. Pamidi	Editor, Technical newswire
Gord Sissons	Managing Principal, IT Technical Analysis
John Wesley	Owner and CEO, IT Services
Thiruvengadarajan Chandrasekaran	Senior Architect, Business Intelligence
Deepak Ravikumar	Data Scientist
Vrishod Sathiyanarayanan	Data Scientist
Naveen Raju	Data Engineer
Veeramanikandan R	Senior Consultant, BI & Data Warehousing
Rajamani Ramalingam	Lead, DevOps

Additional References

In addition to the footnotes in the main body, here are additional references by topic and solution provider.

1. Modernize

1.1 IBM Netezza Performance Server

- https://developer.ibm.com/clouddataservices/docs/ibm-cloud-pak-for-data/data-virtualization/
- https://improvado.io/blog/data-preparation-tools
- https://www.datameer.com/news/datameer-partners-ibm-new-data-science-machine-learning-platform/
- https://www.g2.com/products/ibm-cloud-pak-for-data/features
- https://www.ibm.com/analytics/data-lake
- https://www.ibm.com/cloud/cloud-pak-for-management
- https://www.ibm.com/downloads/cas/3BY1PNGG
- https://www.ibm.com/downloads/cas/AGDXB6PM
- https://www.ibm.com/downloads/cas/LJWBXDPO
- https://www.ibm.com/in-en/analytics/netezza
- https://www.ibm.com/in-en/products/cloud-pak-for-data/deployment-model-options
- https://www.ibm.com/support/knowledgecenter/SS5FPD_1.0.0/com.ibm.ips.doc/postgresql/admin/c_adm_ips_webconsole.html
- https://www.ibm.com/support/knowledgecenter/SS5FPD_1.0.0/com.ibm.ips.doc/postgresql/admin/c_sysadm_netezza_cli_overview.html
- https://www.ibm.com/support/knowledgecenter/SS5FPD_1.0.0/com.ibm.ips.doc/postgresql/ips_compat.html
- https://www.ibm.com/support/knowledgecenter/SSEP7J_11.1.0/com.ibm.swg.ba.cognos.ug_cog_rlp.doc/c_cubingconcepts.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/wsj/analyze-data/analytics-dashboard.html
- https://www.openshift.com/
- https://www.trifacta.com/partners/ibm/

1.2 Snowflake

- https://www.snowflake.com/product/
- https://docs.snowflake.com/en/user-guide/intro-editions.html
- https://www.snowflake.com/blog/beyond-modern-data-architecture/
- https://www.snowflake.com/blog/data-lake-or-data-swamp/
- https://www.snowflake.com/blog/snowflake-as-your-data-lake-or-even-data-ocean/
- https://www.snowflake.com/blog/want-ubiquitous-cloud-computing-with-your-multi-cloud-strategy-better-get-a-cross-cloud-platform/
- https://www.snowflake.com/data-warehousing-glossary/query-manager/
- https://www.snowflake.com/data-warehousing-glossary/what-are-ad-hoc-queries/
- https://www.snowflake.com/news/snowflake-announces-new-self-service-data-warehouse/
- https://www.snowflake.com/trending/what-self-service-analytics
- https://www.snowflake.com/use-cases/snowflake-as-your-data-lake/

1.3 AWS Redshift

- https://aws.amazon.com/blogs/apn/driving-hybrid-cloud-analytics-with-amazon-redshift-and-denodo-data-virtualization
- https://aws.amazon.com/blogs/apn/tag/data-virtualization.
- https://aws.amazon.com/blogs/big-data/agile-analytics-with-amazon-redshift/
- https://aws.amazon.com/blogs/big-data/how-to-migrate-from-ibm-netezza-to-amazon-redshift-with-no-downtime/
- https://aws.amazon.com/blogs/database/integrating-teradata-with-amazon-redshift-using-the-aws-schema-conversion-tool/
- https://aws.amazon.com/marketplace/pp/Talend-Talend-Data-Preparation-for-AWS/B01KAPUAY0
- https://aws.amazon.com/solutions/implementations/data-lake-solution/
- https://docs.aws.amazon.com/prescriptive-guidance/latest/patterns/migrate-a-teradata-database-to-amazon-redshift-using-aws-sct-data-extraction-agents.html
- https://docs.aws.amazon.com/redshift/latest/dg/c_redshift-sql-implementated-differently.html
- https://docs.aws.amazon.com/redshift/latest/dg/c_unsupported-postgresql-features.html
- https://docs.aws.amazon.com/redshift/latest/dg/c_unsupported-postgresql-functions.html
- https://docs.aws.amazon.com/redshift/latest/dg/c-using-spectrum.html
- https://docs.aws.amazon.com/redshift/latest/dg/federated-overview.html
- https://docs.aws.amazon.com/SchemaConversionTool/latest/userguide/agents.dw.html
- https://www.denodo.com/en/data-virtualization/data-federation

1.4 Teradata

- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/cpd/svc/dv/dv-teradata.html
- http://assets.teradata.com/resourceCenter/downloads/Brochures/EB4403.pdf
- http://assets.teradata.com/resourceCenter/downloads/Brochures/EB6293.pdf
- http://assets.teradata.com/resourceCenter/downloads/Brochures/Teradata_Aster_Discovery_Platform_EB7573.pdf
- http://assets.teradata.com/resourceCenter/downloads/Brochures/Teradata_Aster_Discovery_Portfolio_EB7509.pdf
- http://assets.teradata.com/resourceCenter/downloads/WhitePapers/EB3154.pdf
- https://assets.teradata.com/resourceCenter/downloads/Datasheets/Hybrid-Cloud-Options-for-Teradata-Vantage-Datasheet.pdf
- https://cloud.google.com/blog/products/data-analytics/bringing-teradata-vantage-to-google-cloud
- https://datometry.com/data-virtualization-technology/
- https://docs.teradata.com/reader/N06ooHwOGcycpbHnjyA_rw/NW_tWI0tw6cWHYF5ANN9Sg
- https://docs.teradata.com/reader/NC0jOtZwcMAeYkU69mLwyA/RCZYvbHoMRqRv19n5VO1fw
- https://en.wikipedia.org/wiki/Teradata
- https://github.com/Teradata/docker-images
- https://hub.docker.com/search?q=teradata&type=image
- https://kylo.io/

- https://www.teradata.com/Blogs/Introducing-Teradata-IntelliCloud-Our-Next-Generation-Managed-Cloud
- https://www.teradata.com/Cloud
- https://www.teradata.com/Cloud/Data-Lake
- https://www.teradata.com/Products/Ecosystem-Management/IntelliSphere/Unity
- https://www.teradata.com/Products/Ecosystem-Management/IntelliSphere/Viewpoint
- https://www.teradata.com/Products/Software/Integrated-Data-Warehouses
- https://www.teradata.com/Products/Software/Vantage/Analyst

1.5 Microsoft Azure Synapse Analytics

- https://docs.microsoft.com/en-us/sql/azure-data-studio/data-virtualization-extension?view=sql-server-ver15
- https://www.denodo.com/en/denodo-platform/denodo-platform-for-azure
- https://docs.microsoft.com/en-us/sql/linux/quickstart-install-connect-docker?view=sql-server-ver15&pivots=cs1-bash
- https://azure.microsoft.com/en-in/services/synapse-analytics/#features
- https://azure.microsoft.com/en-in/services/synapse-analytics/
- https://azure.microsoft.com/en-us/services/private-link/
- https://azure.microsoft.com/en-in/services/synapse-analytics/#features
- https://azure.microsoft.com/en-in/blog/azure-sql-data-warehouse-is-now-azure-synapse-analytics/
- https://azure.microsoft.com/en-us/services/data-share/

2. Infrastructure

2.1 IBM Netezza Performance Server

- http://www.praxiumgroup.com/adug-www/prev_pres/ADUG%202013-04-05%20-%20PureData-TechDeepDive.pdf
- https://github.com/IBM/nzgo
- https://www.ibm.com/support/knowledgecenter/SS5FPD_1.0.0/com.ibm.ips.doc/postgresql/ips_overview.html
- https://www.ibm.com/support/pages/ibm-puredata-system-analytics-creating-network-bond-redhat-enterprise-linux-rhel-5-n1001-series-appliances
- https://www.ibmbigdatahub.com/sites/default/files/document/redguide_2011.pdf
- https://www-01.ibm.com/common/ssi/ShowDoc.wss?docURL=/common/ssi/rep_ca/8/760/ENUSJP20-0038/index.html&lang=en&request_locale=en

2.2 Snowflake

- https://community.snowflake.com/s/article/Caching-in-Snowflake-Data-Warehouse
- https://docs.snowflake.com/en/user-guide/intro-key-concepts.html#snowflake-architecture
- https://docs.snowflake.com/en/user-guide/warehouses.html
- https://docs.snowflake.com/en/user-guide/warehouses-multicluster.html
- https://www.snowflake.com/product/architecture/

2.3 AWS Redshift

- https://docs.aws.amazon.com/redshift/latest/dg/c_challenges_achieving_high_performance_queries.html
- https://docs.aws.amazon.com/redshift/latest/gsg/getting-started.html
- https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html

2.4 Teradata

- https://www.teradata.com/Products/Hardware/IntelliFlex
- http://assets.teradata.com/resourceCenter/downloads/Brochures/EB7865.pdf
- http://assets.teradata.com/resourceCenter/downloads/WhitePapers/EB9836.pdf
- https://assets.teradata.com/resourceCenter/downloads/Brochures/EB9388.pdf
- https://docs.teradata.com/reader/8va3hHldqttqINim95eDKQ/djgioihNSM~~53k4c8YuZw
- https://docs.teradata.com/reader/jRMubtQoiqTllrJvfVcfxQ/XfApZLOzy1pofaRVuxuHhA

- https://docs.teradata.com/reader/N06ooHwOGcycpbHnjyA_rw/GHWXgJUwpkmJ4L_bObOQSA
- https://docs.teradata.com/reader/Repwkrf9vULBIMFEsQL6nA/iIaxUyTOu29oXzi4ugXZEQ
- https://kr.cloudera.com/solutions/gallery/teradata-appliance-for-hadoop-with-cloudera.html
- https://www.teradata.com/Products/Hardware/IntelliBase

2.5 Microsoft Azure Synapse Analytics

- https://docs.microsoft.com/en-us/sql/analytics-platform-system/parallel-data-warehouse-overview?view=aps-pdw-2016-au7
- https://azure.microsoft.com/en-in/services/storage/
- https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/massively-parallel-processing-mpp-architecture#synapse-sql-mpp-architecture-components
- https://docs.microsoft.com/en-us/sql/analytics-platform-system/hardware-configurations?view=aps-pdw-2016-au7
- https://techcommunity.microsoft.com/t5/azure-synapse-analytics/performance-benchmark-azure-synapse-analytics-data-warehouse/ba-p/1381302
- https://www.slideshare.net/jamserra/scaling-sql-server-to-hundreds-of-terabytes

3. Enterprise Ready

3.1 IBM Netezza Performance Server

- http://www.praxiumgroup.com/adug-www/prev_pres/ADUG%202013-04-05%20-%20PureData-TechDeepDive.pdf
- https://www.ibm.com/cloud/cloud-pak-for-management
- https://www.ibm.com/downloads/cas/ZB43B1YJ
- https://www.ibm.com/in-en/cloud/multicloud-manager
- https://www.ibm.com/mysupport/s/topic/0TO50000000IYkUGAW/cloud-pak-for-data?language=en_US
- https://www.ibm.com/support/knowledgecenter/en/SSQNUZ_2.1.0/com.ibm.icpdata.doc/iira/iira-create-proj.html
- https://www.ibm.com/support/knowledgecenter/en/SSQNUZ_2.1.0/com.ibm.icpdata.doc/zen/overview/security.html
- https://www.ibm.com/support/knowledgecenter/SS5FPD_1.0.0/com.ibm.icpds.doc/admin/vulnerability_scan.html
- https://www.ibm.com/support/knowledgecenter/SS5FPD_1.0.0/com.ibm.ips.doc/postgresql/odbc/c_datacon_plg_overview.html
- https://www.ibm.com/support/knowledgecenter/SSQNUZ
- https://www.ibm.com/support/knowledgecenter/SSQNUZ_2.1.0/com.ibm.icpdata.doc/zen/install/upgrade-overview.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/cpd/troubleshoot/troubleshooting.html

3.2 Snowflake

- https://docs.snowflake.com/en/user-guide/admin-user-management.html
- https://docs.snowflake.com/en/user-guide/data-load-snowpipe-rest-apis.html
- https://docs.snowflake.com/en/user-guide/ecosystem-security.html
- https://docs.snowflake.com/en/user-guide/intro-editions.html
- https://docs.snowflake.com/en/user-guide-data-share.html
- https://www.snowflake.com/developers/
- https://www.snowflake.com/product/security-and-trust-center/
- https://www.snowflake.com/resources/data-applications/?utm_CTA=website-PPC
- https://www.snowflake.com/snowflakes-security-compliance-reports/

3.3 AWS Redshift

- https://aws.amazon.com/blogs/big-data/scale-your-amazon-redshift-clusters-up-and-down-in-minutes-to-get-the-performance-you-need-when-you-need-it/
- https://docs.aws.amazon.com/redshift/latest/dg/c_workload_mngmt_classification.html
- https://docs.aws.amazon.com/redshift/latest/dg/concurrency-scaling.html
- https://docs.aws.amazon.com/redshift/latest/mgmt/configuring-connections.html
- https://docs.aws.amazon.com/redshift/latest/mgmt/iam-redshift-user-mgmt.html
- https://docs.aws.amazon.com/redshift/latest/mgmt/metrics.html
- https://docs.aws.amazon.com/redshift/latest/mgmt/security-compliance.html
- https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html

3.4 Teradata

- https://assets.teradata.com/resourceCenter/downloads/Datasheets/EB7351_Premier_Support.pdf
- https://assets.teradata.com/resourceCenter/downloads/Datasheets/Hybrid-Cloud-Options-for-Teradata-Vantage-Datasheet.pdf
- https://assets.teradata.com/resourceCenter/downloads/Product-Support-Policies.pdf
- https://cloud.google.com/blog/products/data-analytics/bringing-teradata-vantage-to-google-cloud
- https://docs.teradata.com/reader/YGWiBKmUdWChYIxyxZOK7g/rT4f_iYRDaAkMLJfUKFB2A
- https://en.wikipedia.org/wiki/Teradata
- https://www.teradata.com/Blogs/Introducing-Teradata-IntelliCloud-Our-Next-Generation-Managed-Cloud
- https://www.teradata.com/Products/Ecosystem-Management/IntelliSphere/Viewpoint

3.5 Microsoft Azure Synapse Analytics

- https://azure.microsoft.com/en-in/support/legal/preview-supplemental-terms/
- https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-overview-manage-security
- https://www.slideshare.net/jamserra/azure-synapse-analytics-overview
- https://www.stitchdata.com/resources/bigquery-vs-azure-synapse-analytics/

4. Data Management

4.1 IBM Netezza Performance Server

- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/wsj/catalog/get-started-op.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/cpd/svc/datastage/c_supported_conn_stgs.html
- https://www.ibm.com/support/pages/connecting-microsoft-windows-azure-sql-database-information-server-using-datadirect-jdbcodbc-drivers
- https://www.ibm.com/support/pages/connecting-datastage-amazon-redshift-odbc
- https://www.ibm.com/support/knowledgecenter/SSZJPZ_11.7.0/com.ibm.swg.im.iis.ds.nav.doc/containers/cont_iisinfsrv_ds_and_qs.html
- https://www.ibm.com/support/knowledgecenter/SSQNUZ_2.1.0/com.ibm.icpdata.doc/zen/install/hadoop.html
- https://www.ibm.com/support/knowledgecenter/SSQNUZ_2.1.0/com.ibm.icpdata.doc/zen/install/cdh.html
- https://www.ibm.com/support/knowledgecenter/SS5FPD_1.0.0/com.ibm.ips.doc/postgresql/udfdev/c_udf_create_shlibs.html
- https://www.ibm.com/support/knowledgecenter/en/SSZJPZ_9.1.0/com.ibm.swg.im.iis.ds.parjob.adref.doc/topics/c_deeadvrf_Accessing_SQL_Server_from_DataStage.html
- https://www.ibm.com/support/knowledgecenter/en/SSULQD_7.1.0/com.ibm.nz.adm.doc/r_sysadm_wlm_feature_summary.html
- https://www.ibm.com/support/knowledgecenter/en/SSULQD_7.1.0/com.ibm.nz.adm.doc/c_sysadm_wlm_nz_appliance.html

4.2 Snowflake

- https://community.snowflake.com/s/article/Tech-in-5-Real-Time-Architecture
- https://community.snowflake.com/s/article/Using-DBT-to-Execute-ELT-Pipelines-in-Snowflake
- https://docs.snowflake.com/en/sql-reference/data-types-text.html
- https://docs.snowflake.com/en/sql-reference/stored-procedures-api.html
- https://docs.snowflake.com/en/user-guide/data-load-snowpipe.html
- https://docs.snowflake.com/en/user-guide/data-pipelines-intro.html
- https://docs.snowflake.com/en/user-guide/ecosystem-all.html
- https://docs.snowflake.com/en/user-guide/ecosystem-etl.html
- https://docs.snowflake.com/en/user-guide/query-size-limits.html
- https://docs.snowflake.com/en/user-guide/spark-connector.html
- https://docs.snowflake.com/en/user-guide-data-load.html
- https://docs.snowflake.com/en/user-guide-data-unload.html
- https://stackoverflow.com/questions/59295563/maximum-javascript-string-size-in-snowflake-stored-procedures

- https://support.snowflake.net/s/question/0D50Z00007CZcqmSAD/what-is-limit-on-number-of-columns-how-to-do-a-sparse-table
- https://support.snowflake.net/s/question/0D50Z00008BDIPTSA5/snowflake-oltp
- https://www.snowflake.com/trending/big-data-tools
- https://www.snowflake.com/trending/data-governance-framework
- https://www.snowflake.com/use-cases/data-engineering/
- https://www.xplenty.com/blog/redshift-vs-snowflake/

4.3 AWS Redshift

- https://aws.amazon.com/blogs/big-data/metadata-classification-lineage-and-discovery-using-apache-atlas-on-amazon-emr/
- https://aws.amazon.com/blogs/big-data/run-mixed-workloads-with-amazon-redshift-workload-management/
- https://aws.amazon.com/glue/
- https://aws.amazon.com/kinesis/
- https://aws.amazon.com/marketplace/search/?page=1&category=07b485fb-a5a4-4d61-aa53-06a55497a64b
- https://aws.amazon.com/marketplace/search/?page=1&filters=VendorId&VendorId=d13da660-f20f-4a0b-9f73-77a73706ce96&category=07b485fb-a5a4-4d61-aa53-06a55497a64b
- https://aws.amazon.com/mp/scenarios/bi/mdm/
- https://aws.amazon.com/redshift/partners/?partner-solutions-cards.sort-by=item.additionalFields.partnerName&partner-solutions-cards.sort-order=asc&awsf.partner-solutions-filter-partner-type=use-case%23data-integration
- https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/what-is-datapipeline.html
- https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-what-is-emr.html
- https://docs.aws.amazon.com/redshift/latest/dg/c_redshift_system_overview.html
- https://docs.aws.amazon.com/redshift/latest/dg/cm-c-implementing-workload-management.html
- https://docs.aws.amazon.com/redshift/latest/dg/getting-started-federated.html
- https://docs.aws.amazon.com/redshift/latest/dg/materialized-view-overview.html
- https://docs.aws.amazon.com/redshift/latest/dg/stored-procedure-create.html

4.4 Teradata

- http://assets.teradata.com/resourceCenter/downloads/Brochures/EB6664.pdf
- http://assets.teradata.com/resourceCenter/downloads/WhitePapers/EB7805.pdf
- http://www.sourcemediaconferences.com/CDIFL07/pdf/ShainmanM_Enterprise_MDM_Complementing-and-Extending-Active.pdf
- http://www.tensupport.com/library/materialview.pdf
- https://assets.teradata.com/resourceCenter/downloads/Datasheets/EB8387_Master_Data_Management_Datasheet_FINAL.pdf
- https://assets.teradata.com/resourceCenter/downloads/WhitePapers/EB7614.pdf.pdf
- https://aws.amazon.com/marketplace/pp/Teradata-Teradata-Vantage-Advanced-DIY/B06Y4ZM9KY
- https://docs.teradata.com/reader/C8cVEJ54PO4~YXWXeXGvsA/qA9DefitTUVFMUjuX2puVA
- https://docs.teradata.com/reader/Eay_qOFnZprmAcMX6RlaEQ/A144z0FeaA23lBqLlSBqVg
- https://docs.teradata.com/reader/JtLhZxnZVIJAs8pZG1VVfg/root
- https://docs.teradata.com/reader/kmuOwjp1zEYg98JsB8fu_A/Njtxp90eGc06tsr~V1MZyg
- https://docs.teradata.com/reader/N06ooHwOGcycpbHnjyA_rw/GHWXgJUwpkmJ4L_bObOQSA
- https://docs.teradata.com/reader/N06ooHwOGcycpbHnjyA_rw/IYGgzFEUZ4scJTSYEt6Wsg
- https://docs.teradata.com/reader/qs3BwCv08lKpQd7ePrYOvA/DUnM7nEmLaiRfTfpgiMiJg
- https://docs.teradata.com/reader/utmttCBNdC6kDW8061XWDQ/lTdTBjXrWnjlN2sAEgn5Og
- https://docs.teradata.com/reader/vNIQtW0hqhEdO0IUGs1D3Q/2tJRZOXzPwbj2GrhsC351g
- https://docs.teradata.com/reader/XE1o9QrvcquDZ7akeQqyag/HsKbz8Dj8TpWiyZfB2FJ7Q
- https://docs.teradata.com/reader/yxBNqiK9WZh15WoRzMnBrw/HFIqM4Ne2bPfzi1RLIPx_g
- https://docs.teradata.com/reader/zzfV8dn~lAaKSORpulwFMg/4yYWcJ1P1RdqxRQ6258EPQ
- https://downloads.teradata.com/tools/articles/teradata-parallel-transporter/teradata-parallel-transporter-1-basics
- https://www.predictiveanalyticstoday.com/teradata-bigdata-analytics/
- https://www.teradata.com/Blogs/How-to-Enjoy-Hybrid-Partitioning-with-Teradata-Columnar
- https://www.teradata.com/Partners/Informatica
- https://www.teradata.com/Products/Applications/Master-Data-Management

4.5 Microsoft Azure Synapse Analytics

- https://www.slideshare.net/jamserra/azure-synapse-analytics-overview
- https://azure.microsoft.com/en-in/services/stream-analytics/
- https://docs.microsoft.com/en-in/azure/synapse-analytics/spark/apache-spark-overview
- https://docs.microsoft.com/en-us/azure/hdinsight/
- https://docs.microsoft.com/en-us/azure/hdinsight/hadoop/apache-hadoop-etl-at-scale
- https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables?tabs=sql-pool
- https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-workload-management#:~:text=Synapse%20SQL%20pool%20workload%20management,your%20workload%20utilizes%20system%20resources.
- https://docs.microsoft.com/en-us/sql/data-quality-services/introduction-to-data-quality-services?view=sql-server-ver15
- https://docs.microsoft.com/en-us/sql/master-data-services/master-data-services-installation-and-configuration?view=sql-server-ver15#-on-an-azure-virtual-machine
- https://profisee.com/solutions/cloud/

5. Analytics

5.1 IBM Netezza Performance Server

- https://developer.ibm.com/technologies/analytics/articles/ba-data-mining-techniques/
- https://en.wikipedia.org/wiki/SPSS_Modeler
- https://improvado.io/blog/data-preparation-tools
- https://www.datameer.com/news/datameer-partners-ibm-new-data-science-machine-learning-platform/
- https://www.ibm.com/downloads/cas/ZB43B1YJ
- https://www.ibm.com/support/knowledgecenter/SSEP7J_11.1.0/com.ibm.swg.ba.cognos.ug_cog_rlp.doc/c_cubingconcepts.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/wsj/analyze-data/analytics-dashboard.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/wsj/analyze-data/data-science.html
- https://www.trifacta.com/partners/ibm/

5.2 Snowflake

- https://hunters.ai/hunters-snowflake-rsa/
- https://towardsdatascience.com/machine-learning-in-snowflake-fdcff3bdc1a7
- https://www.snowflake.com/blog/boost-your-analytics-with-machine-learning-and-advanced-data-preparation/
- https://www.snowflake.com/blog/creating-a-security-data-platform-with-snowflake-and-hunters-ai/
- https://www.snowflake.com/blog/machine-learning-and-data-warehousing/
- https://www.snowflake.com/blog/machine-learning-in-minutes-announcing-zepl-in-partner-connect/
- https://www.snowflake.com/news/qubole-snowflake-bring-machine-learning-cloud-data-warehouse/
- https://www.snowflake.com/use-cases/enabling-data-driven-application-developers/ai-machine-learning-data-science/
- https://www.thoughtspot.com/snowflake
- https://www.thoughtspot.com/thoughtspot-and-snowflake-announce-partnership-deliver-search-ai-driven-analytics-database-cloud
- https://www.trifacta.com/blog/trifacta-for-snowflake-part-2/

5.3 AWS Redshift

- https://aws.amazon.com/marketplace/pp/prodview-qm4jjwykj4yxy?qid=1580398223144&sr=0-12&ref_=srh_res_product_title
- https://aws.amazon.com/quicksight/
- https://docs.aws.amazon.com/machine-learning/latest/dg/binary-model-insights.html

5.4 Teradata

- https://assets.teradata.com/resourceCenter/downloads/Datasheets/Hybrid-Cloud-Options-for-Teradata-Vantage-Datasheet.pdf
- http://assets.teradata.com/resourceCenter/downloads/Brochures/Teradata_Aster_Discovery_Platform_EB7573.pdf
- https://docs.teradata.com/reader/N06ooHwOGcycpbHnjyA_rw/17LbqaEWrntdYuCmcyR1Zg
- https://docs.teradata.com/reader/N06ooHwOGcycpbHnjyA_rw/d4LKYX4~1jG3cCIeRLWkXw
- https://docs.teradata.com/reader/N06ooHwOGcycpbHnjyA_rw/IYGgzFEUZ4scJTSYEt6Wsg
- https://docs.teradata.com/reader/Tk6W9zbo2NHGakhfyvWmmw/IB57v7SHpsdbA1Jv89XCcg
- https://www.teradata.com/Products/Software/Database
- https://www.teradata.com/Products/Software/Vantage/Analyst

5.5 Microsoft Azure Synapse Analytics

- https://azure.microsoft.com/en-in/blog/azure-sql-data-warehouse-is-now-azure-synapse-analytics/
- https://azure.microsoft.com/en-in/services/synapse-analytics/#features
- https://azure.microsoft.com/en-in/services/synapse-analytics/#overview

CABOT PARTNERS

OPTIMIZING BUSINESS VALUE

Cabot Partners Group, Inc.
100 Woodcrest Lane, Danbury, CT 06810
www.cabotpartners.com | info@cabotpartners.com