

IBM Cloud Pak for Data System: The Leading Hyper-Converged Platform for Data and AI

Sponsored by IBM

Ravi Shankar, Ph.D., MBA and Srinu Chari, Ph.D., MBA
April 2020

<mailto:info@cabotpartners.com>

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.

The "AI Ladder" is an intricate, multistep workflow where clients must collaborate at each step to **collect** data to make it simple and accessible, **organize** data to create a business-ready analytics foundation such as a data warehouse for business intelligence (BI), **analyze** data to build analytics and AI with trust and transparency and **infuse** AI across the business.

However, before climbing the AI Ladder, clients must **modernize** their information architecture to have choice and flexibility. Several hybrid cloud data platforms provide some of these capabilities to operationalize AI at scale with business continuity even during disasters. But many organizations are struggling with "build vs. buy decisions." Often, they may lack the time, money and/or expertise to build, manage and operate these environments from scratch.

So, companies are now choosing to **buy pre-integrated solutions** that are optimized to drive more innovation, improve productivity and shorten time to market, while reducing costs, complexity and deployment risks. IBM Cloud Pak for Data System delivers these benefits.

IBM Cloud Pak for Data System is a hyper-converged inter-operable platform that provides software defined network, storage and compute to reduce complexity, increase scalability, accelerate time to value and maximize ROI. Though not an appliance, the System delivers these benefits through an appliance-like experience. It is not limited to installations with a few racks but rather provides a fully automated way to expand network, storage and compute to build an entire data center as a center for gravity for relational and non-relational data. The system hardware comes pre-installed and optimized with the most current version of IBM Cloud Pak for Data. Users of the System access a data lake that's scalable to large petabytes, a hybrid cloud data management software platform that pre-integrates data and AI capabilities as modular services and an information architecture that supports governance.

Starting with **modernize**, we present a comprehensive comparative analysis of the services at each step of the AI Ladder offered by major providers of hybrid cloud data platforms. Based on this analysis, we believe IBM Cloud Pak for Data System is outperforming competitors and fast-tracking the delivery of client value in their advanced analytics and AI journey by delivering a full end-to-end platform in one system.

IBM Cloud Pak for Data System is the ideal platform for clients who want to quickly get started on their AI journey. Clients can also easily add on cartridges that extend data and AI services as needed and use Netezza Performance Server – the industry leading high-performance data warehouse solution built with open source and hardware acceleration.

Copyright © 2020. Cabot Partners Group, Inc. All rights reserved. Other companies' product names, trademarks, or service marks are used herein for identification only and belong to their respective owner. All images and supporting data were obtained from IBM or from public sources. The information and product recommendations made by the Cabot Partners Group are based upon public information and sources and may also include personal opinions of both Cabot Partners Group and others, all of which we believe to be accurate and reliable. However, as market conditions change and not within our control, the information and recommendations are made without warranty of any kind. The Cabot Partners Group, Inc. assumes no responsibility or liability for any damages whatsoever (including incidental, consequential or otherwise), caused by your or your client's use of, or reliance upon, the information and recommendations presented herein, nor for any inadvertent errors which may appear in this document. This paper was developed with IBM and Intel funding. Although the paper may utilize publicly available material from various vendors, including IBM, it does not necessarily reflect the positions of such vendors on the issues addressed in this document.

Hybrid Cloud Data Platforms are Crucial for Analytics and AI

Analytics and AI 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 and more are rapidly emerging. In fact, AI adoption alone has grown an astounding 270% in the last four years and 40% of organizations expect it to be the leading game changer in business.¹ The economic impact of AI is immense with an estimated \$13 trillion added to the global economy over the next decade.²

However, for analytics and AI to become an integral part of an organization, numerous deployment challenges with data and infrastructure must be overcome – data volumes (50%), data quality and management (47%) and skills (44%).³

To overcome these challenges and best implement analytics/AI workflows across the enterprise, a common, collaborative, cloud-like containerized environment to run all analytic processes together in an inter-operable way is needed. An open, governed information architecture (IA) and data management solutions that promote choice, flexibility, business continuity and high-performance are also critical. 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 solution for analytics/AI (Table 1).

Public Cloud	On-Premises
<ul style="list-style-type: none"> • High scalability and flexibility for unpredictable workload demands with varying peaks/valleys • Rapid software development, test and proof of concept pilot environments • No capital investments required to deploy and maintain infrastructure • Faster provisioning time and reduced requirements on IT expertise as this is managed by the cloud vendor 	<ul style="list-style-type: none"> • Hard to move existing data lakes into the cloud because of large data volumes – bring compute to where data resides • Supports edge analytics for making immediate decisions • Provides dedicated and secure environments for compliance to stringent regulations and/or unique workload requirements • High/custom SLA performance and efficiency

Table 1: Benefits of Public Cloud and On-Premises Infrastructures for Analytics/AI

Consequently, the worldwide data services for the hybrid cloud market is expected to grow at a healthy CAGR of 20.53% from 2016 to 2021⁴ 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.⁴

Today, large organizations leverage almost five clouds on average. The percentage of enterprises with a strategy to use multiple clouds is 84%⁵ and 56% of the organizations plan to increase the use of containers.⁶

Hybrid cloud data platforms that support a multi-cloud architecture will have a significant advantage.

¹ <https://futureiot.tech/gartner-ai-adoption-growing-despite-skills-shortage/>

² Tim Fountaine, Brian McCarthy and Tamim Saleh, “Building the AI-Powered Organization”, Harvard Business Review, July-August 2019.

³ Ritu Jyoti, “Accelerate and Operationalize AI Deployments Using AI - Optimized Infrastructure”, IDC Technology Spotlight, June 2018.

⁴ <https://www.ibm.com/downloads/cas/V93QE3QG>

⁵ RightScale® STATE OF THE CLOUD REPORT 2019 from Flexera™

⁶ <https://www.redhat.com/cms/managed-files/rh-enterprise-open-source-report-detail-f21756-202002-en.pdf>

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

Hybrid cloud data platforms help overcome AI deployment challenges

The hybrid cloud data services market is growing 20.53% annually

A Hybrid Multi-Cloud Platform Operationalizes Analytics/AI

A hybrid, open, multi-cloud platform allows organizations to take advantage of their data and applications wherever it resides, on-premises and across many clouds. It provides clients with:⁷

- **Flexibility:** Freedom to choose and change environments, data and services as needed – making sure that there is no “Vendor Lock-in.”⁸
- **Integration:** Reduce data proliferation across multiple silos, databases and clouds. Integrate data stored and managed in a hybrid environment on-premises and across many private and public clouds.
- **Workload and resource optimization and portability:** Optimize cost, performance and agility, with the flexibility to move data and analytics/AI workloads between private environments and public clouds using containers to meet changing business demands.
- **Compliance:** Deploy cloud environments that are already compliant with regulatory requirements and provide for ease of update across platforms as inevitable change occurs.
- **Pre-integration:** Provide an integrated platform that pre-integrates data management, analytics/AI and development tooling needed to accelerate time-to-value.
- **Global delivery and deployment:** Distribute applications and data for better management of data sovereignty and compliance. Easily implement a growing array of microservices.
- **High performance:** Build, deploy and manage analytics/AI models at scale as needed.

An open hybrid multi-cloud environment empowers customers to experiment with and choose the programming languages, tools, algorithms and infrastructure to build data pipelines, train and productionize analytics/AI models in a governed way for the enterprise, and share insights throughout the workflow. In other words, efficiently traverse the AI Ladder and operationalize AI at scale in a managed way across the organization.

The AI Ladder Needs a Modern Cloud-Native Platform

The AI Ladder (Figure 1), developed by IBM, provides organizations with an understanding of where they are in their AI journey and a framework for determining where they need to focus. It is a guiding principle for organizations to transform their business by providing four areas to consider: how they collect data, organize data, analyze data and then ultimately infuse AI into their organization. It also shows how a modernized end-to-end platform eliminates the silos of inefficiency that exist today in traditional architectures and tooling and provides business continuity during disasters.

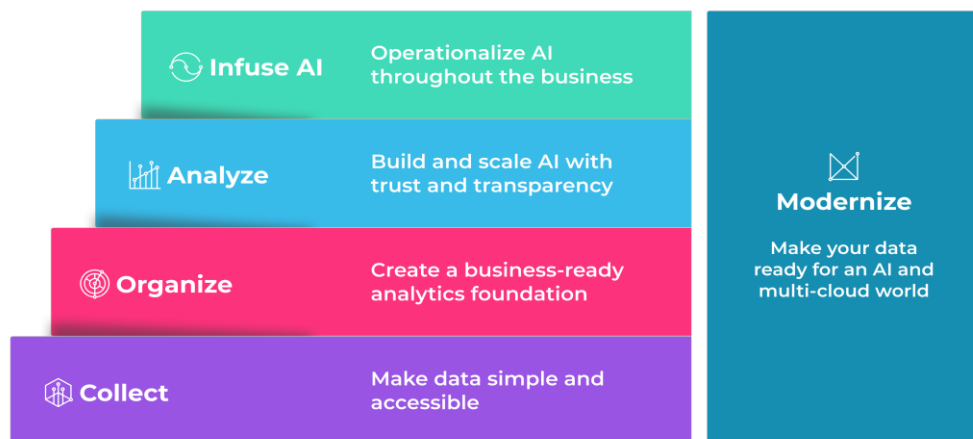


Figure 1: AI Ladder – a Guiding Principle for Business Transformation with Data and AI

⁷ <https://www.ibmdatahub.com/blog/vision-hybrid-cloud-big-data-and-analytics>

⁸ https://en.wikipedia.org/wiki/Vendor_lock-in

Hybrid multi-cloud platforms provide flexibility, integration, portability, compliance, and more

Hybrid multi-cloud solutions make it easy to traverse the AI Ladder to operationalize AI at scale

The AI Ladder provides actionable guidance to transform business with AI

Modernize to collect, organize and analyze data and infuse AI

IBM Cloud Pak for Data System provides a platform with pre-integrated data and AI services

IBM Cloud Pak for Data System is a hyper-converged system with pre-installed software for data and AI

Relative to a public cloud, IBM Cloud Pak for Data System cost savings grow with greater usage

The four steps that make up the AI ladder are:

1. **Collect:** Make data simple and accessible – collect data of every type regardless of where it resides, enabling flexibility in the face of ever-changing data sources while reducing costs, inefficiency and quality associated with traditional data management.
2. **Organize:** Create a business-ready analytics foundation – Organize all data into a trusted business-ready foundation with built-in governance, protection, and compliance regardless of where the data resides.
3. **Analyze:** Build and scale AI with trust and transparency – Analyze data in faster and smarter ways and benefit from AI models that empower organizations to gain new insights and make better, smarter decisions without the traditional challenges and silos.
4. **Infuse:** Operationalize AI throughout the business – Apply AI across the enterprise in multiple departments and within various processes, drawing on predictions, automation and optimization that a unified platform system supports.

Before climbing the AI Ladder, customers must **modernize** and build an information architecture that is efficient, agile and able to offer choice and flexibility across the enterprise. This expands the impact of AI across their organization and requires less assembly.

The AI Ladder must stand atop a modern cloud-native architecture with each step providing a robust foundation to deploy the corresponding critical data and AI services. This ensures customers get a consistent deployment experience with common services to manage, monitor and scale across all steps of the ladder – including AI applications. IBM Cloud Pak for Data System provides this platform with pre-integrated data and AI services and can significantly reduce the Total Cost or Ownership (TCO) for clients in their analytics/AI journey.

Modernize and Improve ROI with IBM Cloud Pak for Data System

[IBM Cloud Pak for Data System](#) is a hyper-converged system, based on Intel Xeon Scalable processors, that comes pre-installed with [IBM Cloud Pak for Data](#) platform software. The cloud-native platform uses a modular and scalable

approach to provide storage with continuous availability, automatic network/storage/compute expansions to reduce complexity and special skills to manage the infrastructure with seamless procedures to extend to multiple clouds.

Clients can quickly get started with instant pre-assembled provisioning and save on the ongoing costs typical with public clouds. Figure 2

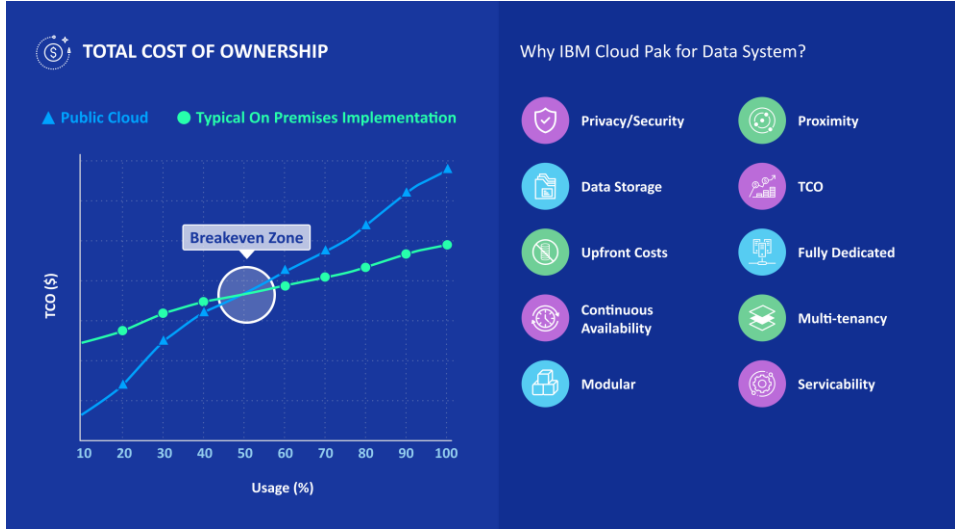


Figure 2: IBM Cloud Pak for Data System Lowers TCO Over Time Compared to the Public Cloud

depicts these cost savings that grow with increasing usage and the advantages of IBM Cloud Pak for Data System for analytics and AI. According to SwiftStack, the five-year on-premises TCO for IBM Cloud Pak for Data System storage could be 1/4th the TCO of public cloud storage.

While many public cloud providers proclaim huge savings and benefits, actual costs can quickly grow. Key factors that help improve the ROI of IBM Cloud Pak for Data System relative to a public cloud and traditional on-premises implementation are:

- **Usage:** At low levels of usage, a public cloud may be cost-effective, but as the usage starts ramping up, there is an inflection zone where an on-premises implementation is much more cost-effective.
- **Proximity:** IBM Cloud Pak for Data System keeps the data close to compute 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.
- **Upfront Costs:** Depending on the implementation, an on-premises solution will not incur many costs associated with transformation, migration and integration with existing systems.
- **Fully Dedicated:** There is no need to share on-premise infrastructure resources. This mitigates jitter and unpredictable performance because of noisy neighbors.
- **Multi-tenancy:** Many applications such as development, testing and production can be consolidated.
- **Continuous Availability/Business Continuity:** Global dedicated storage, multiple availability zones and Red Hat OpenShift provide seamless building of local S3 cache within a cluster and data replication/distribution to dedicated/off-premises object storages. This also helps provide business continuity during disasters.
- **Privacy/Security:** The platform is FIPS compliant and meets strict security compliance for GDPR.
- **Modularity:** Infrastructure can be purchased as needed and there is no need to invest capital up-front. This means users can start out small and buy expansion units to grow the System depending on their usage and storage needs.
- **Optimized hardware:** Hardware is selected and designed specifically for IBM Cloud Pak for Data System, based on Intel Xeon Scalable processors with optimized software extensions for improved performance and security.
- **Installation:** IBM provides full installation so that IBM Cloud Pak for Data System is up and running once set-up is complete.
- **Serviceability:** IBM fixes hardware failure (such as disc or memory breakdowns) at no additional cost. Users can access automatic ticket creation for hardware component replacements, use call-home features and start seamless software upgrades without downtime.

[**Netezza Performance Server**](#) is an optional add-on data warehousing solution to IBM Cloud Pak for Data System. It is built with open-source and optimized for high performance analytics with built-in hardware acceleration. Netezza Performance Server data is stored, managed, and processed on high performant NVMe drives giving 3000 MB/sec read and 700 MB/s write performance. The NVMe drives are self-encrypting drives (SEDs) that can be turned on-off based upon requirements.

Cartridges on IBM Cloud Pak for Data System allow users to extend data and AI capabilities through services such as Data Virtualization, Watson Assistant, Watson Studio and Watson OpenScale.

Intel powers IBM Cloud Pak for Data System with Intel Xeon Scalable processors and Intel SSDs. The hardware with workload-optimized software add-ons can help accelerate AI results and time to actions. These technologies can help speed development and insights with optimized frameworks and help collect and organize data faster, with a modernized, high-performance and secure platform that protects data.

Many unique design features lower the TCO of IBM Cloud Pak for Data System

Netezza Performance Server is optimized for data warehouses with built-in hardware acceleration

Intel provides a comprehensive value-added portfolio for AI

Key Data and AI Platforms Considered for Comparative Analysis

Cabot Partners considered many categories of leading data and AI solution providers:

- **Hyperscalers:** IBM Cloud Pak for Data System, Amazon AWS, Microsoft Azure, and Google Cloud – these offer data and AI services and position themselves as integrated platforms spanning data management, analytics, AI and development tooling.
- **Traditional providers:** Oracle, Teradata, SAS, SAP, Informatica, Cloudera – these offer data and AI products spanning only a subset of the data and AI lifecycle and position themselves as platforms.
- **Point products:** Dataiku, Alteryx, Domino Data Labs, Collibra, Alation, Denodo, Databricks, Data Robot, Snowflake, etc. – these address only a subset/niche of the data and AI Ladder lifecycle such as warehousing, data visualization or data catalog.
- **Hyper-converged systems:** Nutanix, Dell EMC, etc. – these are not tailormade for data and AI.

The focus of this analysis is on major solution providers with the breadth of services/offerings for all steps of the AI Ladder. So, we focused on the following providers in addition to **IBM Cloud Pak for Data System:**

- **Amazon Web Services (AWS) Outposts:** A fully managed service that extends AWS infrastructure, AWS services, APIs, and tools to the data center, co-location space, or on-premises facility for a consistent hybrid experience.
- **Microsoft Azure Stack/Hybrid Cloud:** Enables deployment of Azure services anywhere and extends Azure management to any infrastructure.
- **Google Anthos:** Open hybrid and multi-cloud application platform that enables clients to modernize their existing applications, build new ones and run them anywhere in a secure manner. Anthos enables consistency between on-premises and cloud environments and helps accelerate application development.
- **Oracle Cloud at Customer:** Platform as a Service (PaaS) used to build, deploy, integrate and extend applications in the cloud. Oracle Cloud Platform offerings include data management, application development, integration, content and experience, business analytics, systems management and security.

Comparative Analysis Methodology and Results for Each Step of the AI Ladder

We used the AI Ladder framework for the comparative analysis and evaluated key services in each step: Collect, Organize, Analyze, Infuse and Modernize. The information used was obtained from recently published articles. It was also validated through a rigorous process of interviewing subject matter experts and practitioners with hands-on experience in building and deploying data and AI solutions on each of the considered hybrid cloud data platforms (see Appendix for an exhaustive list of references).

For every alternative, each service listed in the following tables for every step 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 in a hybrid cloud environment as well as customer/community acceptance of that service. The axis of the spider chart for each step 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 step of the AI Ladder and alternative:

Modernize: The services and comparison of features for the Modernize step are shown in Table 2 (next page).

The competitive scope included hyperscalers, traditional providers, point products and hyper-converged systems

Competitors considered were AWS, Microsoft Azure, Google Anthos and Oracle

This analysis evaluated in detail and graded services at each step of the AI Ladder with extensive research and interviews

The Modernize step – building an information architecture for AI that provides choice and flexibility across the enterprise

Services	Comparison Features
Storage Services	Object Storage, Block Storage, Shared File Storage, Archival Storage, Storage Gateway, Data Transfer Service
Deployment Options (Private Cloud, Hyper-converged Systems, Public Cloud)	Ability to deploy in Private Cloud / Data Centers, Hyper-converged Systems and Public Cloud Vendors (IBM, AWS, Google, Azure, Oracle and others)
Multi-cloud Management	Unified portal to view and manage all resources across a multi-cloud environment, Security and Governance, Migration Options, Auto detect infrastructures
DevOps	Cloud-native (Containerized Solutions and Services), CI/CD Pipelines, Automation, Multi-language, Collaboration
Systems Hardware	Rack Solutions (Dimensions, Options with minimal features and ability to extend or add more features and devices), Cores, Storage, Power
Networking/Switches	Highly available Hyper-converged Modular Building Blocks, Compute/Storage with Hardware Acceleration, Management, Power, Rack Solutions, CPU (Socket Type, Clock Speed, Turbo Speed, No of Cores), Memory, Storage, Unlimited expansion, Start small Grow bigger, Number of switches, Network Speed/Bandwidth, Redundancy Switching, Brand Value
Continuous Availability	Availability performance (Downtime), Availability Zones, Kubernetes Container Platform
Long Term ROI	<p>Comparison of business value for 3+ years.</p> <p>Costs: Implementation (Hardware and Software) and ongoing cost (Labor, License and Support) Benefits: Productivity, Operational Efficiency, Cost of Security, Customer Retention, Increased Revenue, Reduced Cost (Infrastructure, Personnel), Mitigated risk of vendor lock in</p> <p>Licensing (Flat Fee, Pay Per Use, Monthly) - Flat fee and Monthly would be better for long term Managed Services, Customer Owned Infrastructure</p>
Day 2 Operations	Operational Support, Backup & Disaster Recovery, Quick Repairing, Upgrades, Resizing
Provisioning	Ease of use (Simple Provisioning), Time independent access, Mobility

Table 2: Services and Comparison Features for the Modernize Step of the AI Ladder

The results of the analysis for the Modernize step are shown in Figure 3.

IBM offers the best flexibility and choice: software only, hyper-converged or rack-based solutions with hybrid multi-cloud support

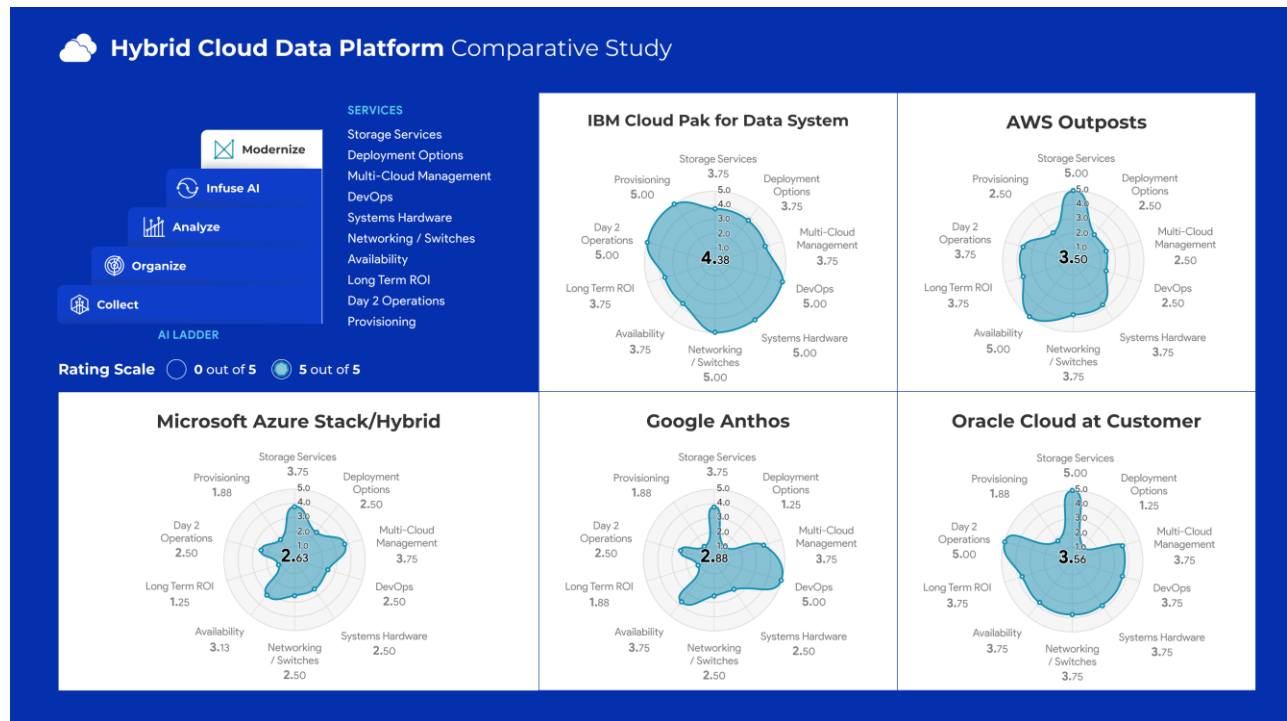


Figure 3: Comparative Analysis for the Modernize Step of the AI Ladder

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

The Collect step – the ability to collect all relevant data and make it accessible to the business no matter where the data resides

IBM Cloud Pak for Data System has a mature set of comprehensive, out-of-the-box services for the Collect step

Modernize – Synopsis: IBM is a strong provider with innovative, cloud-native and out-of-the-box services and solutions. IBM offers flexibility by providing multiple choices: software only, hyper-converged or rack-based solutions. AWS’s main shortcoming is that its hybrid solution is restricted to rack-based solutions and there are no options for using third-party or customer owned hardware. In addition, multi-cloud management is not their strong suit. Microsoft Azure Stack/Hybrid is mainly a software-only solution and offers multi-cloud management services. Google Anthos is a software-only solution. However, it has attractive and promising partner solutions, Intel, Lenovo and Dell being the leading partners. Oracle has weak multi-cloud management and integration support with other public cloud providers.

Collect: The services and comparison of features for Collect are shown in Table 3.

Services	Comparison Features
Databases	SQL - OLTP RDBMS, SQL - OLAP RDBMS, SQL - Mixed Workload RDBMS, NoSQL - Key-Value Database, NoSQL - Document Database, NoSQL - Graph Database, NoSQL - Time Series Database, Cluster - High Availability, Sharding - Horizontal Partitioning - Distributed Database, Data Replication - Change Data Capture, Data Masking, Data Encryption
ETL Tools	Visual Flow - An Easy-to-Use IDE and Code-Less Tools, Connections - Variety of source and target systems supported, Big Data - Data Integration, Analytics - Real-time analytics engines (Mainly Spark, Impala, etc.), Data Management - Data, cleansing, Data Profiling, Cloud - Support for cloud data sources - Data warehouse (e.g. Snowflake), Storage solutions (e.g. AWS S3), Execution - Ability execute or trigger remote processes in a hybrid environment, Management - Operation Management (Monitoring, Scheduling etc.)
Streaming Solutions	Point to Point Messaging System, Publish-Subscribe Messaging System, Reliability - distributed, partitioned, replicated and fault tolerance, Scalability - messaging system scales easily without down time, Durability, Performance, Extensibility
Big data Solutions	Data Processing, Predictive Applications - Identity Management, Fraud Analytics, Analytics - Risk Analytics, Statistical Analysis, Text Analytics, Content Analytics, Reporting Features - Real-Time Reporting, Dashboards, Security Features - Single Sign-on, Data Encryption, Technologies Support

Table 3: Services and Comparison Features for the Collect Step of the AI Ladder

The results of the analysis are shown in Figure 4.

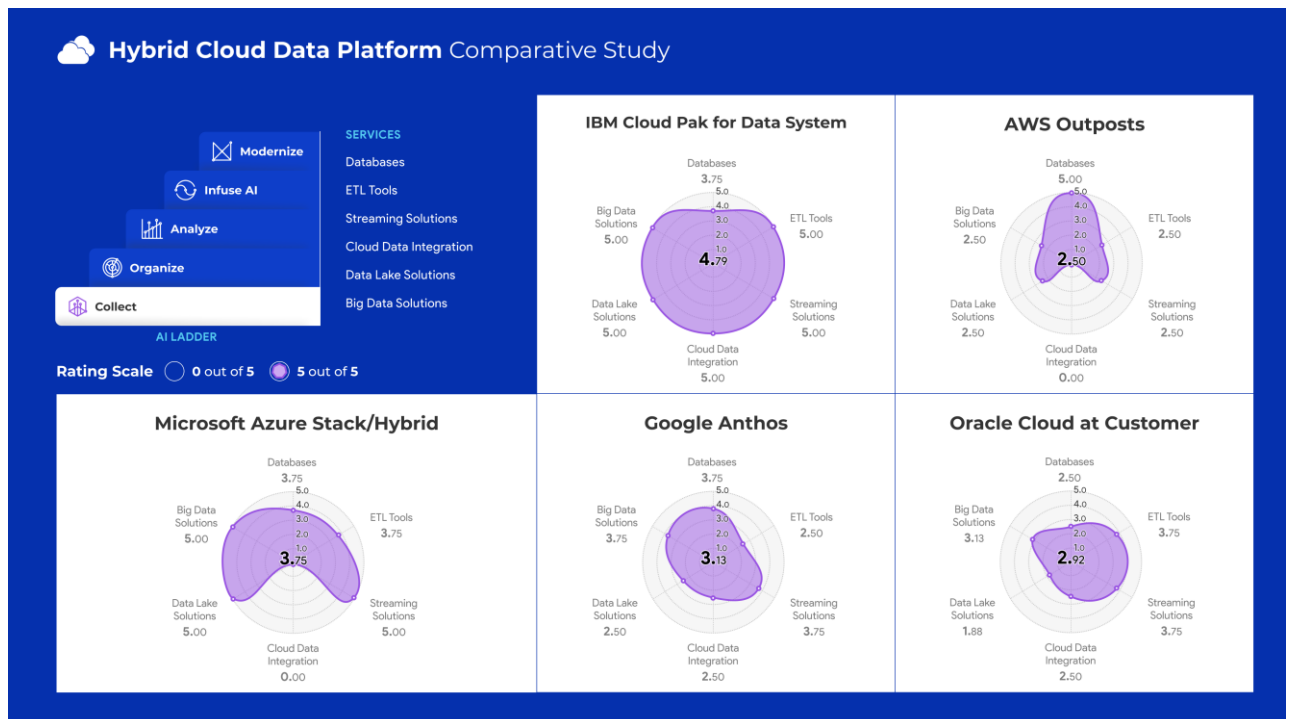


Figure 4: Comparative Analysis for the Collect Step of the AI Ladder

Collect – Synopsis: IBM Cloud Pak for Data System has a mature set of out-of-the-box services for a hybrid cloud. AWS Outposts has limited databases services. Microsoft Azure Stack/Hybrid is still in the early stages of deployment and acceptance, and it has limited database services for a hybrid cloud. Azure’s ETL and BI tools are strong. Oracle Cloud is strong in database and storage services but weaker in cloud data integration and big data services.

Organize: The services and comparison of features for the Organize step are shown in Table 4.

Services	Comparison Features
Data Virtualization	Source Systems supported, Schema or model creation, Security, Performance, Ease of use, Support of SQL and Other Scripting Languages
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
Security	Ability to define policies and rules, Integrated security across all layers - Firewall, policies and rules, SSH, SSL, Authentication & Authorization
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 workflows, Impact Analysis
Integration with Other Clouds	Data Integration, Application Integration, Integrated Security
Data Warehouse Solutions	Data modeling, Databases, OLAP Cubes, Connectivity with BI & Visualization Tools, Scripting Languages, Performance, Scalability and integration features
Cataloging	Create and maintain data dictionaries, Industry standard models with terms and classification
Data Lineage	Data and business lineage, Report generation, Cover variety of assets in the life cycle (databases, files, ETL, BI Tools, MDM)

Table 4: Services and Comparison Features for the Organize Step of the AI ladder

The results of the analysis for the Organize step are shown in Figure 5.

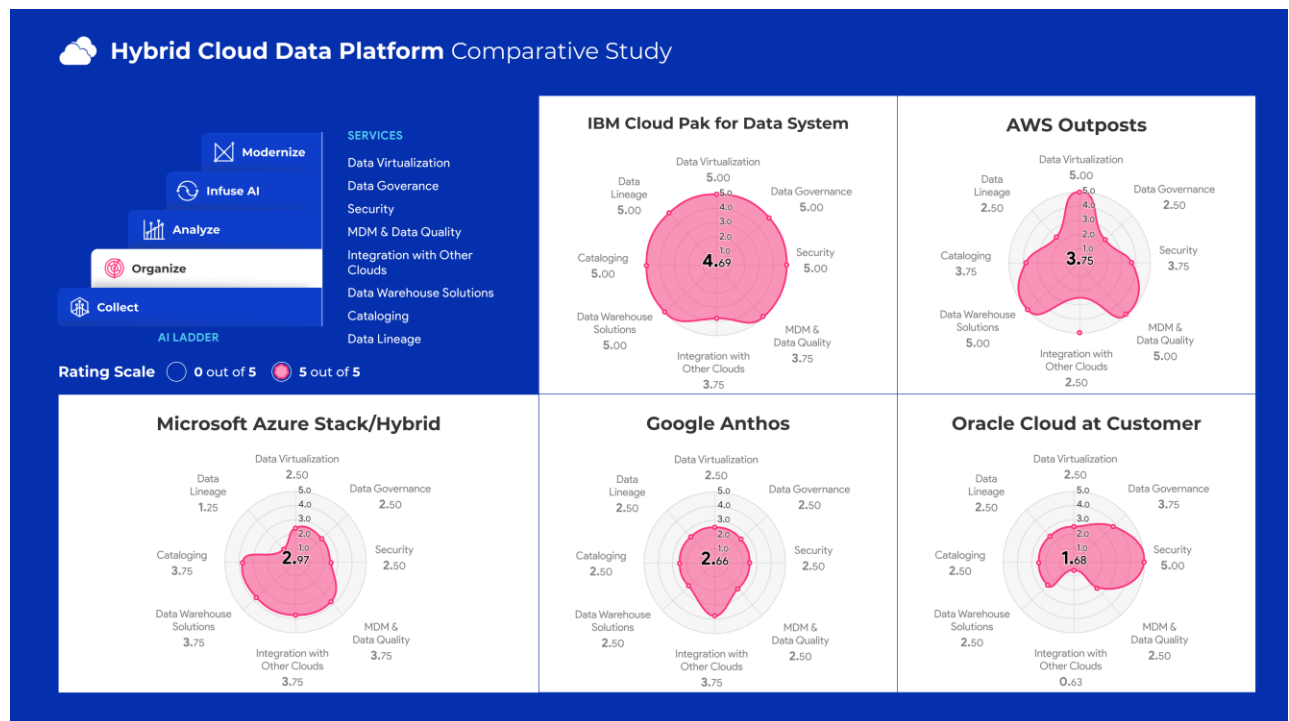


Figure 5: Comparative Analysis for the Organize Step of the AI Ladder

The Organize step – make high-quality data available to the right people, processes, and applications to make better business decisions, improve operations and innovate

IBM Cloud Pak for Data System has powerful out-of-the-box data quality, MDM and other services for the Organize step

Organize – Synopsis: IBM Cloud Pak for Data System leads with a powerful set of mature out-of-the-box services for most of the areas including data quality, MDM, etc. Data Virtualization has clearly defined workflows for personas. AWS Redshift is another popular data warehousing solution that clients could purchase. However, it is not yet available in Amazon Outposts. Other key services are available in the AWS marketplace from partners: Denodo, Profisee, Informatica and Snowflake. Microsoft Azure Stack/Hybrid is feature rich but has limited services. Conduit and Denodo are third-party Data Virtualization solutions available in Azure. Google Anthos services have not yet matured and found wider acceptance. Oracle has weak cloud/Data Virtualization capabilities and limited integration with other clouds.

Analyze: The services and comparison of features for the Analyze step are shown in Table 5.

Services	Comparison Features
Dashboards and Visualization	Visualization options, Data Source, Data extract or Dataset, Semantic Layer, In memory analytics engine, Dashboards and Reports, Scheduled reports and alerts, Schedule refreshes, Stream sources, Big data integration, APIs for integration, Paginated Reports
Explore/Discover	Automatic Discovery, Search and View Assets, Query and View Data
Data Profiling	Data sources, Analysis report with max length, uniqueness, null values or blanks, most frequently used values string patterns, min & max values, Automation & integration options
Data Preparation Tools	Data Sources Supported - Relational, Big Data, Data Lake, Explore/Discover, Transform, Blend, Refine & Enrich, Visualize
Text Analytics and data Mining	Data Preparation, Modeling, Pre-processing and Processing, Mining tasks (Anomaly detection, Dependency Modeling, clustering, classification, regressing, summarization), Validations
Machine Learning	IDEs, Note Books, Languages, Modeling, Big Data Integration, Spark Engine, Data Cleansing
Deep Learning	Deep Neural Networks, Supported Use Cases (Computer Vision, Speech Recognition, Natural Language Processing, Recommendation Engines), GPU Enabled Libraries & Deep Learning Frameworks, Compute with NVIDIA GPU or equivalent
Embedded Analytics	APIs and SDK for application integration, Authorization and Authentication

Table 5: Services and Comparison Features for the Analyze Step of the AI Ladder

The results of the analysis for the Analyze step are shown in Figure 6.

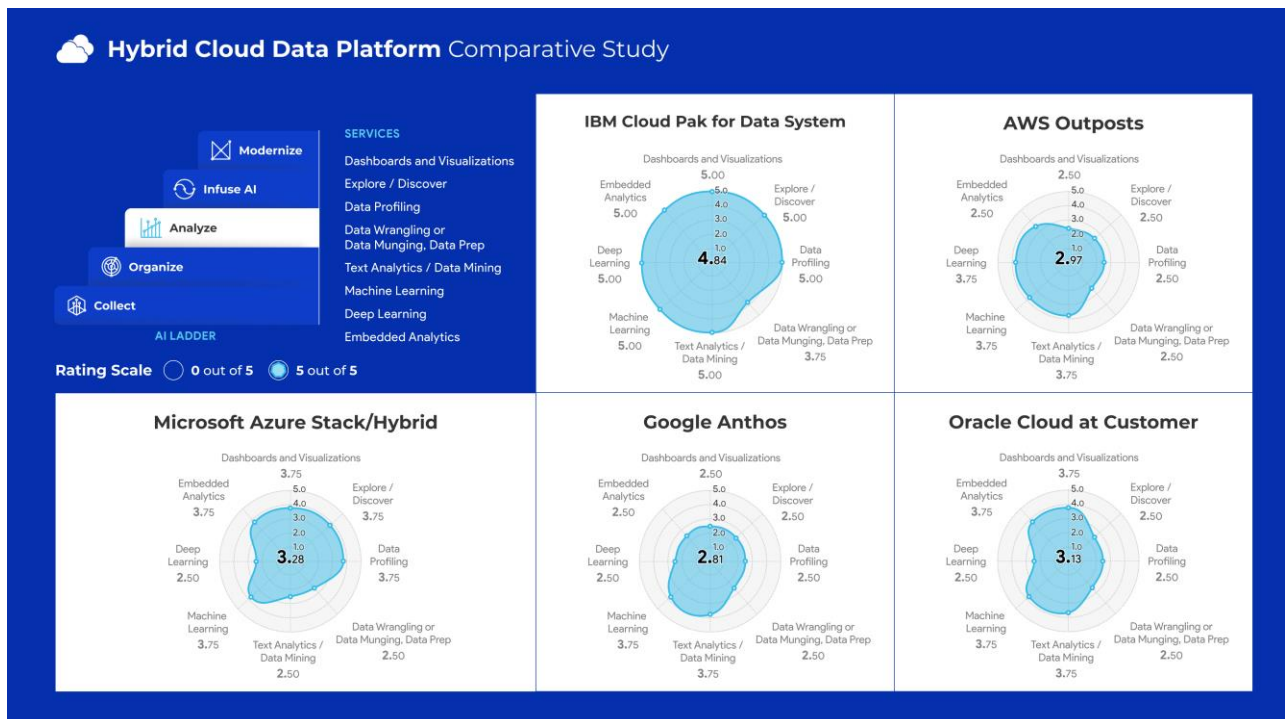


Figure 6: Comparative Analysis for the Analyze Step of the AI Ladder

The Analyze step – create the foundation to understand client data and glean insights and information needed to build and scale AI models with trust and transparency

IBM has a mature portfolio of BI, AI and ML solutions for the Analyze step

Analyze – Synopsis: IBM has out-of-the-box solutions for most of the required set of services. The IBM product portfolio for BI, AI and Machine Learning (ML) is mature and well-known in the industry. AWS does not have a mature set of tools for data visualization, wrangling and discovery. Nonetheless, AWS has popular third-party tools that are available in its marketplace. Microsoft Azure Stack/Hybrid has a leading set of services and tools in this category for the public cloud – but they are not yet available for the hybrid environment. Google Anthos has a good set of tools for AI and data mining. However, most of the services including visualization and discovery are not mature or have not found popular acceptance in the hybrid cloud environment. Oracle is a strong contender in most of these areas. However, its cloud capabilities including AI and BI Visualization tools have yet to get widespread customer traction.

Infuse: The services and comparison of features for the Infuse step are shown in Table 6.

Services	Comparison Features
Deployment and Automation	Ability to deploy or promote model to production, End to end life cycle management (Build / Train, Test, Deploy and integrate)
Operationalize AI	Application Integration, Visualization and Dashboard, Real-time Alerts and dashboards
Industry Accelerators	Voice to Text and Text to Voice, Document Processing, Language Processing, Domain Specific Solutions: Financial Crime / Fraud Detection, Call Center Solutions, etc.

Table 6: Services and Comparison Features for the Infuse Step of the AI Ladder

The results of the analysis for the Infuse step are shown in Figure 7.

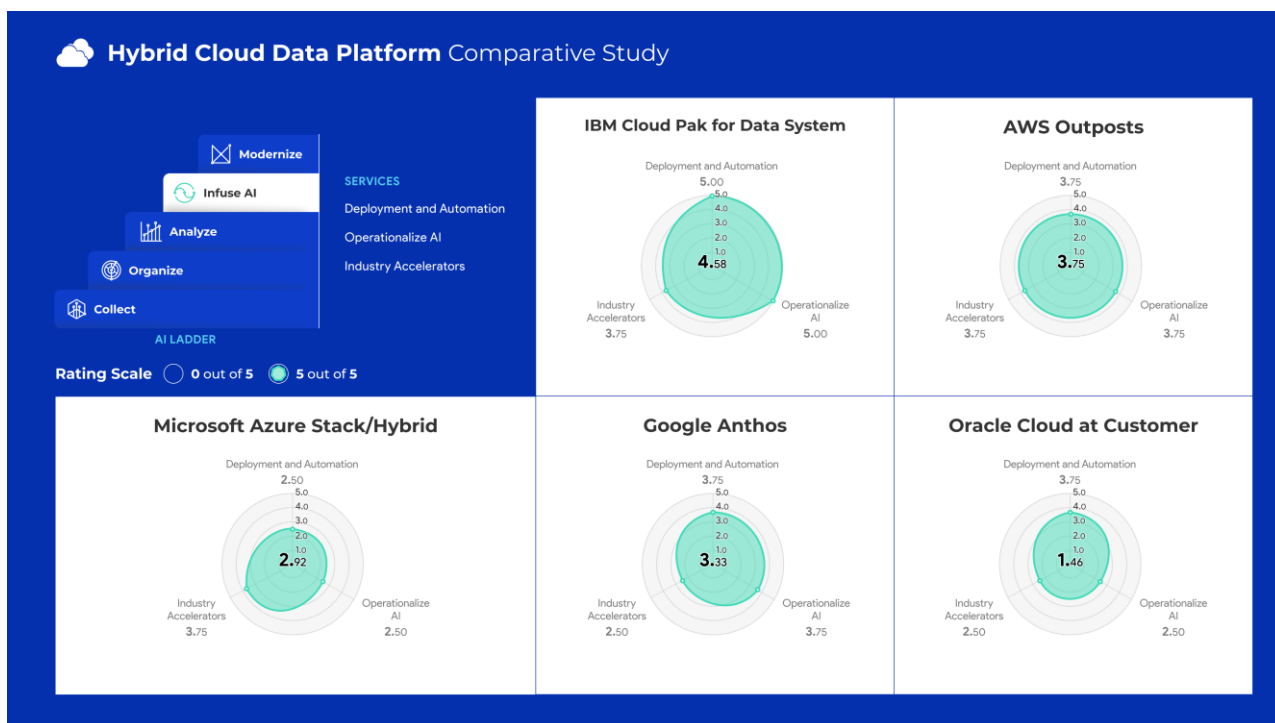


Figure 7: Comparative Analysis for the Infuse Step of the AI Ladder

Infuse – Synopsis: IBM services here are visionary and can rapidly bring value to clients using AI and ML. AWS and Google are strong contenders but have limited hybrid cloud solutions. Oracle and Microsoft Azure Stack/Hybrid also have a good set of features to operationalize AI. Azure services in this space are attractive and are finding greater acceptance in the marketplace. However, Oracle and Azure have limited hybrid multi-cloud capabilities.

The Infuse step – Operationalize AI throughout the business

IBM has visionary services for the Infuse step and can rapidly deliver value for AI/ML

Consolidated Comparative Analysis – Comprehensive Summary View

Figure 8 depicts a comparative analysis dashboard followed by a brief summary of each solution:

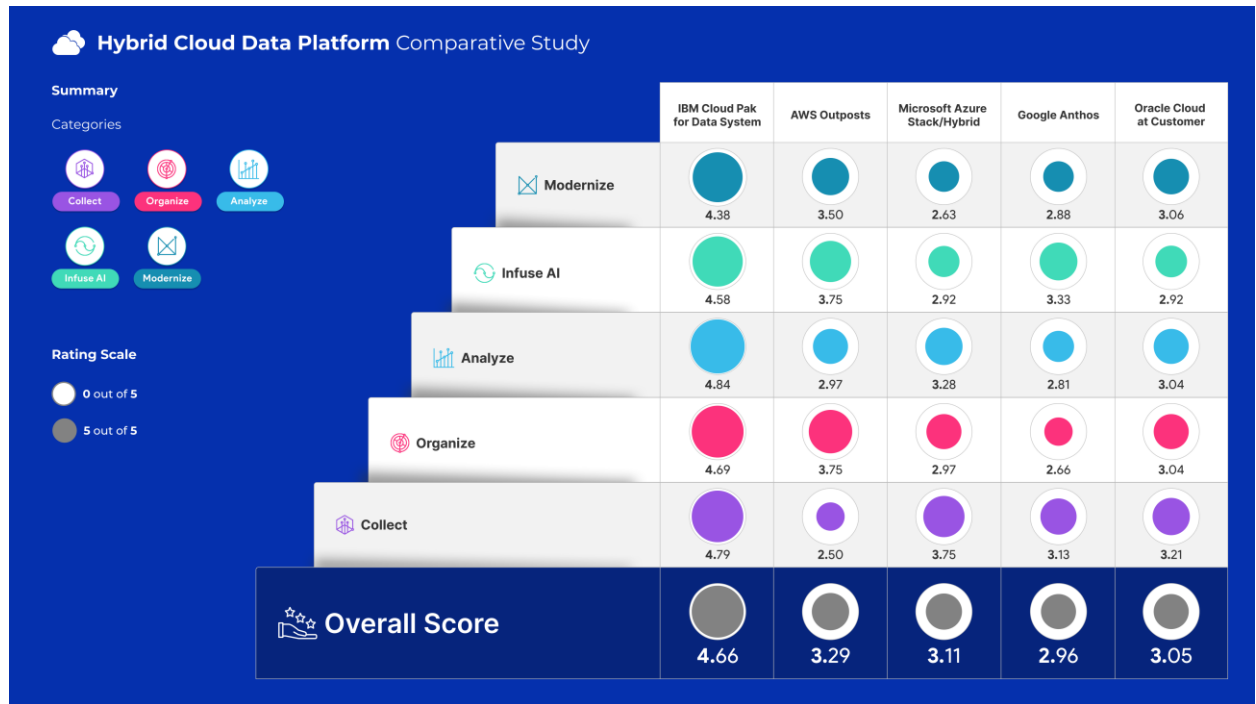


Figure 8: Dashboard of Comparative Analysis

- IBM Cloud Pak for Data System:** A fully integrated data and AI platform that modernizes how businesses collect, organize and analyze data and infuse AI throughout their organizations—and is a clear leader of the pack. IBM Cloud Pak for Data System integrates AI technology into a hybrid data management platform, that includes DataOps, and governance and business-ready analytics technologies. It can be implemented as a software-only solution; implemented on IBM, partner and the customer’s hardware; and on the public cloud. It has out-of-the-box solutions and third-party solutions that are readily available for hybrid cloud implementations.
- AWS Outposts:** A fully managed service that extends AWS infrastructure, AWS services, APIs, and tools to datacenter, co-location space, or on-premises facility for a truly consistent hybrid experience. It is a strong contender to IBM. However, most of the required services for the AI Ladder are not currently available for hybrid implementation. In addition, AWS has only a rack-based solution. AWS also lags in the ability to extend hybrid cloud management to other public cloud services.
- Microsoft Azure Stack/Hybrid:** A software-only hybrid solution which focuses mainly on cloud hybrid management, data services, Kubernetes apps, governance and security. Microsoft Azure Stack/Hybrid supports several market-leading services and out-of-the-box features.
- Google Anthos:** A software-only solution that is an open hybrid and multi-cloud application platform. Anthos solutions are available on Lenovo, Dell and Intel platforms. Anthos has a strong cloud-native foundation for a hybrid solution but currently has limited hybrid implementation services.
- Oracle Cloud at Customer:** A strong contender after AWS. Oracle has limited multi-cloud management and integration support with other cloud solutions. Oracle’s portfolio of services is strong for the public cloud but is limited for hybrid implementations. Oracle supports Private cloud as a rack offering but has limited support/coexistence with other public clouds (Microsoft Azure Stack/Hybrid is an exception) and Software as a Service.

IBM Cloud Pak for Data System leads in all steps of the AI Ladder

Compared to alternatives, IBM Cloud Pak for Data System provides a comprehensive modern platform to accelerate the journey to AI

Conclusions and Recommendations

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. Then, they must *analyze* this data to build analytics and AI models with trust and transparency, and finally operationalize and *infuse* these AI models everywhere across the business.

This intricate, multistep AI Ladder is hard to implement unless clients **modernize** their data and AI tools into a unified platform experience that leverages cloud-native architectures. In addition to IBM, Amazon, Microsoft, Google and Oracle are developing this hybrid cloud data platform vision, but only IBM provides the hyper-converged multi-cloud solution required.

Based on a comprehensive comparative analysis of the services at each step of the AI Ladder, we believe IBM Cloud Pak for Data System has key tactical advantages over the competitors and is fast-tracking the delivery of client value in their advanced analytics and AI journey by:

- Pre-integrating and pre-installing unique modular data and AI services across the AI lifecycle
- Offering a hybrid multi-cloud unified experience data platform in a single hyper-converged system that can scale up or down to meet business needs quickly
- Providing one inter-operable platform, with one unified user experience for all persona users covering all steps of the ladder to AI
- Lowering the TCO with modularity, data proximity, continuous availability, multi-tenancy, and more
- Enabling clients to modernize their infrastructure with high-performance servers, storage and networking all pre-integrated and pre-tested while protecting their current investments with business continuity even during disasters.

For more information, visit [IBM Cloud Pak for Data System](#)

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
Geetha Ramanan	Senior Architect, Cloud Practices
Sankarganesh Shanmugavel	Vice President, Infrastructure & DevOps
Thiruvengadarajan Chandrasekaran	Senior Architect, Business Intelligence
Vrishod Sathiyarayanan	Data Scientist
Rajamani Ramalingam	Lead, Dev ops
Gowtham Premkumar	Engineer, Dev ops

Additional References

In addition to the footnotes in the main body, here are additional references by topic/step of the AI Ladder and solution provider.

1. ROI References

- <https://blogs.cisco.com/datacenter/cisco-hyperflex-vs-public-cloud>
- <https://blogs.cisco.com/datacenter/aws-or-private-cloud-or-both-whats-your-strategy>
- <https://blogs.vmware.com/virtualreality/2014/06/identifying-questionable-assumptions-amazon-tco-comparison-calculator.html>
- <https://www.slideshare.net/DellWorld/cloud-to-build-or-buy-can-you-justify-onpremises-it>
- SwiftStack, “Data Analytics Hub – Total Cost of Ownership (TCO) Analysis”, provided by IBM.

2. Collect

2.1 IBM Cloud Pak for Data System

- <https://www.ibm.com/analytics/data-lake>
- <https://www.g2.com/products/ibm-cloud-pak-for-data/features>
- https://www.ibm.com/support/knowledgecenter/SSQNUZ_2.1.0/com.ibm.icpdata.doc/dfd/c_supported_conn_stgs.html
- https://www.ibm.com/support/knowledgecenter/SSQNUZ_2.1.0/com.ibm.icpdata.doc/zen/install/cdh.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/cpd/overview/whats-new.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/cpd/svc/data-source-svc.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/cpd/svc/datastage/c_multi-cld-data-int.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-datastage-amazon-redshift-odbc>
- https://www.ibm.com/support/knowledgecenter/en/SSZJPZ_9.1.0/com.ibm.svg.im.iis.ds.parjob.adref.doc/topics/c_deeadvrf_Accessing_SQL_Server_from_DataStage.html
- <https://www.ibm.com/support/pages/connecting-microsoft-windows-azure-sql-database-information-server-using-datadirect-jdbcodbc-drivers>

2.2 AWS Outposts

- <https://aws.amazon.com/rds/outposts/>
- <https://aws.amazon.com/rds/>
- <https://aws.amazon.com/blogs/database/creating-highly-available-ibm-db2-databases-in-aws/>
- <https://aws.amazon.com/outposts/features/>
- <https://aws.amazon.com/blogs/big-data/orchestrating-an-etl-process-using-aws-step-functions-for-amazon-redshift/>
- <https://aws.amazon.com/marketplace/search/?page=1&category=8ea84ab7-e3db-4532-8d97-1474f6d7137a>
- <https://aws.amazon.com/marketplace/search/?page=1&category=8ea84ab7-e3db-4532-8d97-1474f6d7137a>
- https://aws.amazon.com/marketplace/search/results?page=1&filters=vendor_id&vendor_id=d13da660-f20f-4a0b-9f73-77a73706ce96
- <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/emr/features/outposts/>

2.3 Microsoft Azure Stack/Hybrid

- <https://azure.microsoft.com/en-in/services/azure-arc/hybrid-data-services/>
- <https://azure.microsoft.com/en-in/services/stream-analytics/>
- <https://azure.microsoft.com/en-in/solutions/data-lake/>
- <https://databricks.com/session/etl-made-easy-with-azure-data-factory-and-azure-databricks>
- <https://docs.microsoft.com/en-us/azure/architecture/example-scenario/data/hybrid-etl-with-adf>
- <https://docs.microsoft.com/en-us/azure/data-factory/connector-amazon-marketplace-web-service>
- <https://docs.microsoft.com/en-us/azure/data-factory/connector-amazon-redshift>
- <https://docs.microsoft.com/en-us/azure/data-factory/connector-amazon-simple-storage-service>
- <https://docs.microsoft.com/en-us/azure/data-factory/connector-google-bigquery>
- <https://docs.microsoft.com/en-us/azure/data-factory/connector-google-cloud-storage>
- <https://docs.microsoft.com/en-us/azure/data-factory/connector-overview>
- <https://docs.microsoft.com/en-us/azure/data-factory/connector-spark>
- <https://docs.microsoft.com/en-us/azure/hdinsight/hadoop/apache-hadoop-etl-at-scale>
- <https://docs.microsoft.com/en-us/sql/data-quality-services/data-cleansing?view=sql-server-ver15>
- <https://docs.microsoft.com/en-us/sql/integration-services/connection-manager/hadoop-connection-manager?view=sql-server-ver15>
- <https://docs.microsoft.com/en-us/sql/integration-services/connection-manager/integration-services-ssis-connections?view=sql-server-ver15>

2.4 Google Anthos

- <https://cloud.google.com/solutions/open-source-databases/#products--services>
- <https://airflow.incubator.apache.org/>
- <https://beam.apache.org/documentation/pipelines/design-your-pipeline/>
- <https://beam.apache.org/documentation/runners/capability-matrix/>
- <https://beam.apache.org/documentation/runners/spark/>
- <https://cloud.google.com/blog/products/gcp/designing-etl-architecture-for-a-cloud-native-data-warehouse-on-google-cloud-platform>
- <https://cloud.google.com/dataflow/docs/guides/sql/data-sources-destinations>
- <https://cloud.google.com/dataflow/docs/quickstarts/quickstart-python>
- <https://cloud.google.com/data-fusion/>
- <https://cloud.google.com/data-fusion/docs/concepts/overview>
- <https://cloud.google.com/data-fusion/docs/quickstart/>
- <https://cloud.google.com/dataproc/>
- <https://cloud.google.com/manage-oracle-workloads/>
- <https://cloud.google.com/products/databases/>
- <https://cloud.google.com/products/data-transfer>
- <https://cloud.google.com/solutions/automated-network-deployment-multicloud>

- <https://cloud.google.com/solutions/build-a-data-lake-on-gcp>
- <https://cloud.google.com/solutions/data-lake/>
- <https://cloud.google.com/solutions/stream-analytics>
- <https://cloud.google.com/solutions/using-gcp-apis-from-an-external-network>
- <https://cloud.google.com/spanner/>
- <https://cloud.google.com/sql/?tab=tab2>
- <https://cloud.google.com/sql-server/>
- <https://cloudplatform.googleblog.com/2015/03/announcing-Google-Cloud-Dataflow-runner-for-Apache-Flink.html>
- <https://stackoverflow.com/questions/50942053/how-to-run-cloud-dataflow-pipelines-using-spark-runner>

2.5 Oracle Cloud at Customer

- <http://www.oracle.com/us/solutions/cloud/oracle-cloud-machine-ds-2949541.pdf>
- <https://blogs.oracle.com/bigdata/data-lake-solution-patterns-use-cases>
- <https://blogs.oracle.com/bigdata/what-is-a-data-lab>
- https://cloudmarketplace.oracle.com/marketplace/en_US/listing/62627436
- <https://docs.oracle.com/en/cloud/cloud-at-customer/>
- <https://docs.oracle.com/en/cloud/paas/integration-cloud-service/index.html>
- <https://docs.oracle.com/en/middleware/fusion-middleware/data-integrator/12.2.1.3/odikm/introduction3.html#GUID-1EDB1383-C6AB-454F-981B-C000F7925191>
- <https://docs.oracle.com/en/solutions/big-data-advanced-analytics/index.html#GUID-AD32BF4D-582A-4CC7-A42A-251C781081C2>
- <https://www.oracle.com/big-data/>
- <https://www.oracle.com/cloud/integration/>
- <https://www.oracle.com/cloud/integration/data-integration-platform-cloud/>
- <https://www.oracle.com/technetwork/database/database-technologies/bdc/big-data-connectors/overview/bigdata-connectors-datasheet-1883359.pdf>

3. Organize

3.1 IBM Cloud Pak for Data

- <https://developer.ibm.com/clouddataservices/docs/ibm-cloud-pak-for-data/data-virtualization/>
- <https://www.ibm.com/in-en/marketplace/infosphere-data-architect>
- https://www.ibm.com/support/knowledgecenter/en/SSEPEK_12.0.0/apsg/src/tpc/db2z_createsp.html
- https://www.ibm.com/support/knowledgecenter/SS5FPD_1.0.0/com.ibm.ips.doc/postgresql/sproc/c_sproc_store_d_procs.html
- https://www.ibm.com/support/knowledgecenter/SS5FPD_1.0.0/com.ibm.ips.doc/postgresql/sqltk/c_sqlcxt_install_and_setup.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/SSLVMB_sub/statistics_mainhelp_ddita/spss/base/idh_repo.html
- https://www.ibm.com/support/knowledgecenter/SSEP7J_11.1.0/com.ibm.swg.ba.cognos.ug_cog_rlp.doc/c_cubin_gconcepts.html
- https://www.ibm.com/support/knowledgecenter/SSQNUZ_2.1.0/com.ibm.icpdata.doc/igc/catalog-govern-overview.html
- https://www.ibm.com/support/knowledgecenter/SSZJPZ_11.7.0/com.ibm.swg.im.iis.ds.nav.doc/containers/containers_and_qs.html
- https://www.ibm.com/support/knowledgecenter/SSZJPZ_11.7.0/com.ibm.swg.im.iis.igcug.doc/topics/r_igcug_lineage.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/cpd/svc/data-source-svc.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/cpd/svc/dv/virtualizing_data.html
- <https://www.tableau.com/solutions/ibm>

- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/wsj/catalog/get-started-op.html

3.2 AWS Outposts

- <https://aws.amazon.com/es/blogs/apn/driving-hybrid-cloud-analytics-with-amazon-redshift-and-denodo-data-virtualization/>
- <https://aws.amazon.com/redshift/partners-detail/denodo/>
- <https://aws.amazon.com/solutions/case-studies/getsmarter-marketplace-case-study/>
- <https://aws.amazon.com/marketplace/search/?page=1&category=07b485fb-a5a4-4d61-aa53-06a55497a64b>
- <https://aws.amazon.com/blogs/big-data/metadata-classification-lineage-and-discovery-using-apache-atlas-on-amazon-emr/>
- <https://aws.amazon.com/marketplace/search/?page=1&category=07b485fb-a5a4-4d61-aa53-06a55497a64b>
- <https://aws.amazon.com/mp/scenarios/bi/mdm/>
- <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/data-warehouse/>
- <https://docs.aws.amazon.com/quicksight/latest/user/spice-in-an-analysis.html>
- <https://docs.aws.amazon.com/quicksight/latest/user/importing-data-to-spice.html>
- <https://docs.aws.amazon.com/redshift/latest/mgmt/configure-jdbc-connection.html>
- <https://docs.aws.amazon.com/redshift/latest/mgmt/configure-odbc-connection.html>
- <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%23business-intelligence&awsml.page-partner-solutions-cards>
- <https://docs.aws.amazon.com/redshift/latest/dg/stored-procedure-overview.html>
- <https://docs.aws.amazon.com/redshift/latest/dg/user-defined-functions.html>
- <https://docs.aws.amazon.com/glue/latest/dg/populate-data-catalog.html>
- <https://docs.aws.amazon.com/glue/latest/dg/aws-glue-api-catalog.html>
- https://aws.amazon.com/marketplace/pp/prodview-qm4jjwykj4xyy?qid=1580398223144&sr=0-12&ref=srh_res_product_title
- <https://atlas.apache.org/#/>
- <https://aws.amazon.com/blogs/big-data/metadata-classification-lineage-and-discovery-using-apache-atlas-on-amazon-emr/>
- <https://aws.amazon.com/blogs/security/how-to-automate-saml-federation-to-multiple-aws-accounts-from-microsoft-azure-active-directory/>
- <https://aws.amazon.com/blogs/security/how-to-set-up-federated-single-sign-on-to-aws-using-google-apps/>
- https://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_providers_saml_3rd-party.html

3.3 Microsoft Azure Stack/Hybrid

- <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://azure.microsoft.com/en-us/services/information-protection/>
- <https://azure.microsoft.com/en-in/services/security-center/>
- <https://azure.microsoft.com/en-in/blog/enhance-hdinsight-security-with-service-endpoints/>
- <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/>
- https://azure.microsoft.com/en-us/product-categories/integration/?ocid=AID754288&wt.mc_id=CFID0424
- <https://docs.microsoft.com/en-us/azure/azure-arc/servers/azcmagent-reference>
- <https://docs.microsoft.com/en-us/azure/azure-arc/servers/overview>
- <https://azure.microsoft.com/en-us/services/synapse-analytics/>
- <https://azure.microsoft.com/en-in/solutions/data-lake/>
- <https://azure.microsoft.com/en-us/services/analysis-services/>

- <https://docs.microsoft.com/en-us/azure/data-catalog/data-catalog-how-to-data-profile>
- <https://feedback.azure.com/forums/906052-data-catalog/suggestions/31919476-relationship-lineage>

3.4 Google Anthos

- <https://cloud.google.com/beyondcorp/>
- <https://cloud.google.com/bigquery>
- <https://cloud.google.com/bigquery#all-features>
- <https://cloud.google.com/bigquery#migrate-teradata>
- <https://cloud.google.com/blog/products/gcp/improving-data-quality-for-machine-learning-and-analytics-with-cloud-dataprep>
- <https://cloud.google.com/data-catalog/>
- <https://cloud.google.com/data-fusion/>
- <https://cloud.google.com/dataprep/>
- <https://cloud.google.com/identity/>
- <https://cloud.google.com/migrate/anthos/>
- <https://cloud.google.com/security/>
- https://cloud.google.com/solutions/migration/dw2bq/dw-bq-data-governance#security_framework
- <https://console.cloud.google.com/marketplace/details/informatica/iics-prod>
- <https://www.denodo.com/en/denodo-platform/test-drives/data-services-gcp>
- <https://www.informatica.com/in/solutions/explore-ecosystems/google-cloud-platform.html>

3.5 Oracle Cloud at Customer

- https://docs.oracle.com/middleware/1213/odsi/OSDID/intro_data_services.htm#OSDID103
- <https://www.oracle.com/a/ocom/docs/oci-data-catalog-data-sheet.pdf>
- <https://www.oracle.com/big-data/guide/what-is-a-data-catalog/>
- <https://www.oracle.com/cloud/integration/data-integration-platform-cloud/>
- <https://www.oracle.com/cloud/security/cloud-services/identity-cloud.html>
- <https://www.oracle.com/database/adw-cloud.html>
- <https://www.oracle.com/middleware/technologies/enterprise-data-quality.html>

4. Analyze

4.1 IBM Cloud Pak for Data

- https://datapatform.cloud.ibm.com/docs/content/wsj/refinery/refining_data.html
- <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/in-en/marketplace/infosphere-information-analyzer>
- https://www.ibm.com/support/knowledgecenter/SSEP7J_11.1.0/com.ibm.swg.ba.cognos.ug_cog_rlp.doc/c_cubin_gconcepts.html
- https://www.ibm.com/support/knowledgecenter/SSQNUZ_2.1.0/com.ibm.icpdata.doc/igc/catalog-govern-overview.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/cpd/svc/watson/knowledge-studio.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.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/wsj/analyze-data/deep-learning1.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/wsj/analyze-data/hadoop-analyze-clusters.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/wsj/analyze-data/ml-experimnt-builder_local.html

- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/wsj/local/hadoop-utilities.html
- <https://www.trifacta.com/partners/ibm/>
- <https://www.ibm.com/in-en/marketplace/cognos-dashboard-embedded>

4.2 AWS Outposts

- <https://aws.amazon.com/about-aws/whats-new/2019/12/introducing-amazon-sagemaker-studio-the-first-integrated-development-environment-ide-for-machine-learning/>
- <https://aws.amazon.com/augmented-ai/>
- <https://aws.amazon.com/blogs/aws/amazon-sagemaker-model-monitor-fully-managed-automatic-monitoring-for-your-machine-learning-models/>
- https://aws.amazon.com/blogs/awsmarketplace/accelerate-amazon-emr-spark-presto-and-hive-with-the-alluxio-ami/?nc1=b_rp
- <https://aws.amazon.com/blogs/machine-learning/build-amazon-sagemaker-notebooks-backed-by-spark-in-amazon-emr/>
- <https://aws.amazon.com/blogs/machine-learning/using-r-with-amazon-sagemaker/>
- <https://aws.amazon.com/comprehend/>
- <https://aws.amazon.com/deeplens/>
- <https://aws.amazon.com/deepracer/>
- <https://aws.amazon.com/machine-learning/amis/>
- <https://aws.amazon.com/macie/>
- https://aws.amazon.com/marketplace/pp/B07KRRSWLY?qid=1580397045844&sr=0-2&ref_=srh_res_product_title
- https://aws.amazon.com/marketplace/pp/prodview-qm4jjwykj4yxy?qid=1580398223144&sr=0-12&ref_=srh_res_product_title
- <https://aws.amazon.com/marketplace/search/results?x=0&y=0&searchTerms=data+prep>
- <https://aws.amazon.com/marketplace/solutions/machine-learning/data-prep>
- <https://aws.amazon.com/quicksight/>
- <https://aws.amazon.com/sagemaker/autopilot/>
- <https://aws.amazon.com/solutionspace/solutions/wipro-data-discovery-platform-on-aws/>
- <https://aws.amazon.com/textract/>
- <https://docs.aws.amazon.com/glue/latest/dg/aws-glue-programming.html>
- <https://docs.aws.amazon.com/sagemaker/latest/dg/gs-studio-ui.html>
- <https://towardsdatascience.com/launching-emr-backed-sagemaker-notebook-via-infrastructure-as-code-boto3-cloudformation-4d2ced33c417>

4.3 Microsoft Azure Stack/Hybrid

- <https://powerbi.microsoft.com/en-us/>
- <https://azure.microsoft.com/en-us/services/data-explorer/>
- <https://docs.microsoft.com/en-us/azure/data-catalog/data-catalog-how-to-data-profile>
- <https://azure.microsoft.com/en-in/updates/self-service-big-data-prep-dataflows-now-available-in-power-bi-embedded/>
- <https://powerbi.microsoft.com/en-us/blog/introducing-power-bi-data-prep-wtih-dataflows/>
- <https://azure.microsoft.com/en-in/services/cognitive-services/text-analytics/>
- <https://azure.microsoft.com/en-us/services/machine-learning/>
- <https://azure.microsoft.com/en-in/services/machine-learning/>
- <https://docs.microsoft.com/en-in/azure/hdinsight/r-server/r-server-overview>
- <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-access-data#supported-data-storage-service-types>
- <https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-machine-learning-mllib-ipython>
- <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/prepare-data>
- <https://azuremarketplace.microsoft.com/en-ca/marketplace/apps/microsoft-ads.dsvm-deep-learning>
- <https://azure.microsoft.com/en-in/services/virtual-machines/data-science-virtual-machines/>
- <https://powerbi.microsoft.com/en-us/power-bi-embedded/>

4.4 Google Anthos

- <https://cloud.google.com/ai-platform-notebooks/>
- <https://cloud.google.com/automl>
- <https://cloud.google.com/bigquery-ml/docs/bigqueryml-intro>
- <https://cloud.google.com/blog/products/gcp/improving-data-quality-for-machine-learning-and-analytics-with-cloud-dataprep>
- <https://cloud.google.com/data-fusion/docs/tutorials/targeting-campaign-pipeline>
- <https://cloud.google.com/datalab/>
- <https://cloud.google.com/dataprep/>
- https://cloud.google.com/dataprep/docs/html/Profile-Your-Source-Data_57344603.html
- <https://cloud.google.com/dataproc/docs/tutorials/bigquery-connector-spark-example>
- <https://cloud.google.com/dataproc/docs/tutorials/bigquery-sparkml>
- <https://cloud.google.com/dataproc/docs/tutorials/gcs-connector-spark-tutorial>
- <https://cloud.google.com/dataproc/docs/tutorials/spark-scala>
- <https://cloud.google.com/deep-learning-vm/>
- <https://cloud.google.com/natural-language>
- <https://cloud.google.com/natural-language/docs/sentiment-tutorial>
- <https://cloud.google.com/solutions/business-intelligence/>
- <https://support.google.com/datastudio/answer/7450249?hl=en>
- <https://oembed.com/#section3>
- <https://stackoverflow.com/questions/52612087/how-to-securely-embed-a-private-google-data-studio-report-without-a-user-needing>
- <https://developers.google.com/datastudio/solution/row-level-filtering>

4.5 Oracle Cloud at Customer

- <https://blogs.oracle.com/bigdata/what-is-oracle-cloud-infrastructure-data-science>
- <https://docs.cloud.oracle.com/en-us/iaas/Content/Compute/References/ngcimage.htm>
- <https://docs.oracle.com/en/solutions/big-data-advanced-analytics/index.html#GUID-CBD6C4CA-5967-400E-AFC9-17EE80CC02B2>
- <https://www.oracle.com/artificial-intelligence/>
- <https://www.oracle.com/cloud/compute/virtual-machines-data-science.html>
- <https://www.oracle.com/database/technologies/datawarehouse-bigdata/oml-notebooks.html>
- <https://www.oracle.com/in/business-analytics/products.html>
- <https://www.oracle.com/in/business-analytics/analytics-cloud.html>
- <https://www.oracle.com/middleware/technologies/bi-foundation/provider-services.html>

5. Infuse

5.1 IBM Cloud Pak for Data

- <https://developer.ibm.com/patterns/infuse-ai-into-your-application/>
- <https://github.com/IBM/icp4d-customer-churn-classifier>
- https://www.ibm.com/support/knowledgecenter/en/SSFHA8_1.2.1/wmla_kc_welcome.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/cpd/svc/fci/fci-get-started.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/cpd/svc/watson/wavi.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/wsj/analyze-data/ml-experimnt-builder_local.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/wsj/analyze-data/ml-spaces_local.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/wsj/analyze-data/visualization-parent.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/wsj/wmls/wmls-deploy-overview.html

- https://www.ibm.com/support/producthub/icpdata/docs/view/overview/SSQNUZ_current/cpd/overview/whats-new.html?t=Overview%20and%20architecture&p=overview

5.2 AWS Outposts

- <https://aws.amazon.com/blogs/big-data/analyze-data-in-amazon-dynamodb-using-amazon-sagemaker-for-real-time-prediction/>
- <https://aws.amazon.com/blogs/big-data/create-real-time-clickstream-sessions-and-run-analytics-with-amazon-kinesis-data-analytics-aws-glue-and-amazon-athena/>
- <https://aws.amazon.com/blogs/machine-learning/use-amazon-cloudwatch-custom-metrics-for-real-time-monitoring-of-amazon-sagemaker-model-performance/>
- <https://aws.amazon.com/kinesis/data-analytics/>
- <https://aws.amazon.com/machine-learning/amis/>
- <https://aws.amazon.com/machine-learning/containers/>
- <https://aws.amazon.com/quicksight/features-ml/>
- <https://aws.amazon.com/solutionspace/solutions/machine-learning-amazon-sagemaker-cloudwick/>
- <https://docs.aws.amazon.com/machine-learning/latest/dg/binary-model-insights.html>
- <https://www.slideshare.net/AmazonWebServices/operationalizing-machine-learning-to-deliver-content-at-scale-mae306-aws-reinvent-2018>

5.3 Microsoft Azure Stack/Hybrid

- <https://azure.microsoft.com/en-in/industries/>
- <https://azure.microsoft.com/en-us/resources/gigaom-delivering-on-the-vision-of-mlops/>
- <https://azure.microsoft.com/en-us/services/event-hubs/>
- <https://docs.microsoft.com/en-us/azure/architecture/reference-architectures/ai/realtime-scoring-python>
- <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-consume-web-service>
- <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-consume-web-service#consume-the-service-from-power-bi>
- <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-and-where>
- <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-app-service>
- <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-functions>
- <https://docs.microsoft.com/en-us/common-data-model/industry-accelerators>
- <https://docs.microsoft.com/en-us/power-bi/service-machine-learning-automated>
- <https://docs.microsoft.com/en-us/power-bi/service-machine-learning-integration>
- <https://docs.microsoft.com/en-us/power-bi/service-real-time-streaming>

5.4 Google Anthos

- <https://cloud.google.com/ai-hub/>
- <https://cloud.google.com/automl>
- <https://cloud.google.com/datalab/>
- <https://cloud.google.com/solutions/ai#solutions>
- <https://cloud.google.com/solutions/stream-analytics>

5.5 Oracle Cloud at Customer

- <https://blogs.oracle.com/bigdata/what-is-oracle-cloud-infrastructure-data-science>
- <https://www.oracle.com/data-science/cloud-infrastructure-data-science.html>

6. Modernize

6.1 IBM Cloud Pak for Data

- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/cpd/overview/whats-new.html
- https://www.ibm.com/support/knowledgecenter/SSQNUZ_2.1.0/com.ibm.icpdata.doc/zen/install/hadoop.html
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/cpd/access/data-sources.html
- http://www-01.ibm.com/common/ssi/ShowDoc.wss?docURL=/common/ssi/rep_ca/8/760/ENUSJP20-0038/index.html&lang=en&request_locale=en
- <https://cloud.ibm.com/catalog/services/ibm-cognos-dashboard-embedded>

- https://cloud.ibm.com/docs/services/cognos-dashboard-embedded?topic=cognos-dashboard-embedded-overview-of-cde&cm_sp=dw-bluemix--data--devcenter&cm_mc_uid=46190736986315287252796&cm_mc_sid_50200000=69576721528750235630&_ga=2.148575287.1493997892.158019164
- https://cloud.ibm.com/docs/services/cognos-dashboard-embedded?topic=cognos-dashboard-embedded-providingapplicationusersowndashboardassets&cm_sp=dw-bluemix--data--devcenter&cm_mc_uid=46190736986315287252796&cm_mc_sid_50200000=69576721528750235630&_ga=2.148575287.1493997892.158019164
- <https://community.ibm.com/community/user/security/blogs/adam-case/2019/11/25/ibm-cloud-identity-idaas-one-of-first-to-support-aws-session-tags>
- <https://www.ibm.com/cloud/cloud-pak-for-management>
- <https://www.ibm.com/in-en/cloud/multicloud-manager>
- <https://www.ibm.com/in-en/products/cloud-pak-for-data/deployment-model-options>
- <https://www.ibm.com/security/partners/aws>
- https://www.ibm.com/support/knowledgecenter/SSMKHH_9.0.0/com.ibm.etools.mft.doc/bp28350_.htm
- <https://www.openshift.com/>
- https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/cpd/troubleshoot/troubleshooting.html
- https://www.ibm.com/mysupport/s/topic/OTO5000000IYkUGAW/cloud-pak-for-data?language=en_US
- <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

6.2 AWS Outposts

- <https://aws.amazon.com/outposts/features/>
- <https://aws.amazon.com/blogs/big-data/embed-interactive-dashboards-in-your-application-with-amazon-quicksight/>
- <https://aws.amazon.com/eks/>
- <https://aws.amazon.com/hybrid/>
- <https://aws.amazon.com/marketplace/pp/HTBASE-JUKE-Multi-Cloud-Data-and-Control-Plane/B07KCZNRTL>
- <https://aws.amazon.com/outposts/features/>
- <https://aws.amazon.com/outposts/pricing/>
- <https://www.youtube.com/watch?v=Xnsrx3IU0Fc>
- https://aws.amazon.com/outposts/faqs/#Support_26_maintenance

6.3 Microsoft Azure Stack/Hybrid

- <https://docs.microsoft.com/en-us/azure/data-factory/data-flow-aggregate>
- <https://docs.microsoft.com/en-us/azure/hdinsight/>
- <https://azure.microsoft.com/en-in/overview/cloudnative/>
- <https://azure.microsoft.com/en-in/services/expressroute/>
- <https://azure.microsoft.com/en-in/services/virtual-wan/>
- <https://docs.microsoft.com/en-us/azure/databox-online/data-box-edge-overview>
- <https://docs.microsoft.com/en-us/azure-stack/operator/azure-stack-hci-overview?view=azs-1910>
- <https://docs.microsoft.com/en-us/azure-stack/operator/azure-stack-overview?view=azs-1910>

6.4 Google Anthos

- <https://cloud.google.com/storage/>
- <https://cloud.google.com/anthos/gke/>
- <https://cloud.google.com/anthos/run/>
- <https://cloud.google.com/anthos/service-mesh>
- <https://cloud.google.com/blog/topics/partners/google-cloud-partners-with-lenovo-on-hybrid-cloud-next19>
- <https://cloud.google.com/dell/>
- <https://cloud.google.com/kubernetes-applications/>
- <https://cloud.google.com/solutions/hybrid-and-multi-cloud-monitoring-and-logging-patterns>
- <https://cloud.google.com/solutions/manage-hybrid-cloud>

- <https://cloud.google.com/stackdriver/>
- <https://lenovopress.com/lp1215.pdf>
- <https://stackoverflow.com/questions/52612087/how-to-securely-embed-a-private-google-data-studio-report-without-a-user-needing>
- <https://www.cisco.com/c/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/at-a-glance-c45-742545.html>
- <https://www.dellemc.com/en-us/collaterals/unauth/white-papers/products/converged-infrastructure/vxflex-integrated-rack-for-anthos-solution.pdf>
- <https://www.intel.in/content/www/in/en/products/solutions/select-solutions/cloud/google-cloud-anthos.html>
- <https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/products/vsan/vmw-google-anthos-vsan-solution-overview.pdf>
- <https://cloud.google.com/products/operations>

6.5 Oracle Cloud at Customer

- <https://www.oracle.com/cloud/systems-management/cloud-services.html>
- <https://docs.oracle.com/en/cloud/cloud-at-customer/private-cloud-at-customer/start/start-concept-ib-fabric-architecture.html#>
- <https://www.oracle.com/a/ocom/docs/corporate/analystrelations/oracle-database-on-premises-cloud-options.pdf>
- <http://www.oracle.com/us/solutions/cloud/oracle-cloud-machine-ds-2949541.pdf>

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.