Greater Choice and Value for Advanced Analytics and AI

How IBM and Cloudera deliver better data access, analytics and decisions throughout your enterprise

Sponsored by IBM and Cloudera

Ravi Shankar, Ph.D., MBA and Srini Chari, Ph.D., MBA   **mailto:info@cabotpartners.com**

September 2019

Executive Summary

Advanced Analytics and Artificial Intelligence (AI) are poised to rapidly transform 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, processes and experiences, and reimagine new business models.

However, most organizations are stuck in experimentation in silos. Industrializing AI throughout the enterprise is not easy. There are many deployment challenges associated with data, talent and trust especially as data volume, velocity and variety continue to explode.

To amplify the value of AI and make it pervasive, it is imperative that clients consider best practices and solutions that address these challenges holistically across several dimensions: Business, Process, Applications, Data and Infrastructure. Doing so provides clients extensive choice and flexibility to maximize the Total Value (Benefits – Costs) of Ownership (TVO) from their investments. This is the goal of the IBM + Cloudera strategic alliance.

By maximizing the TVO, organizations can reduce costs, improve productivity, increase revenues/profits and mitigate risks while industrializing Analytics/AI deployments. This requires an open Information Architecture (IA) and data management solutions with choice and flexibility to operationalize, sustain and scale the intricate, multistep, ladder-like Analytics/AI workflows. Both Cloudera and IBM (especially with the Red Hat acquisition) are deeply committed to open source and hybrid multi-cloud technologies to provide this IA.

Without being prescriptive, this paper provides an overview of the rich and extensive portfolio of IBM and Cloudera products and services. Clients have complete flexibility to choose and customize their specific Analytics/AI solutions including selecting individual components. Anchored on an open framework that supports on-premises and multi-cloud deployments, this portfolio provides clients a valuable array of business, process, applications, data and infrastructure capabilities with unprecedented flexibility and choice to maximize the TVO of their Analytics/AI investments.

Clients deploying Analytics/AI solutions should seriously consider the rich array of products and services from IBM and Cloudera and make their own individual choices on selecting specific components based on their unique needs. Compared to public cloud and other niche solution alternatives that promote vendor lock-in, IBM and Cloudera solutions offer clients an industry-leading, open platform with an enterprise-grade Hadoop distribution plus an ecosystem of integrated products and services – all designed to help organizations industrialize Analytics/AI with greater choice and value.

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

Cabot Partners

Optimizing Business Value

Huge Value of Analytics, AI and Machine Learning (ML)

Analytics, Artificial Intelligence (AI) and Machine Learning (ML) 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 and more (Figure 1) are rapidly emerging. In fact, AI/ML adoption alone has grown an astounding 270% in the last four years and 40% of organizations expect it to be a game changer.[1] The economic impact of AI/ML is immense.
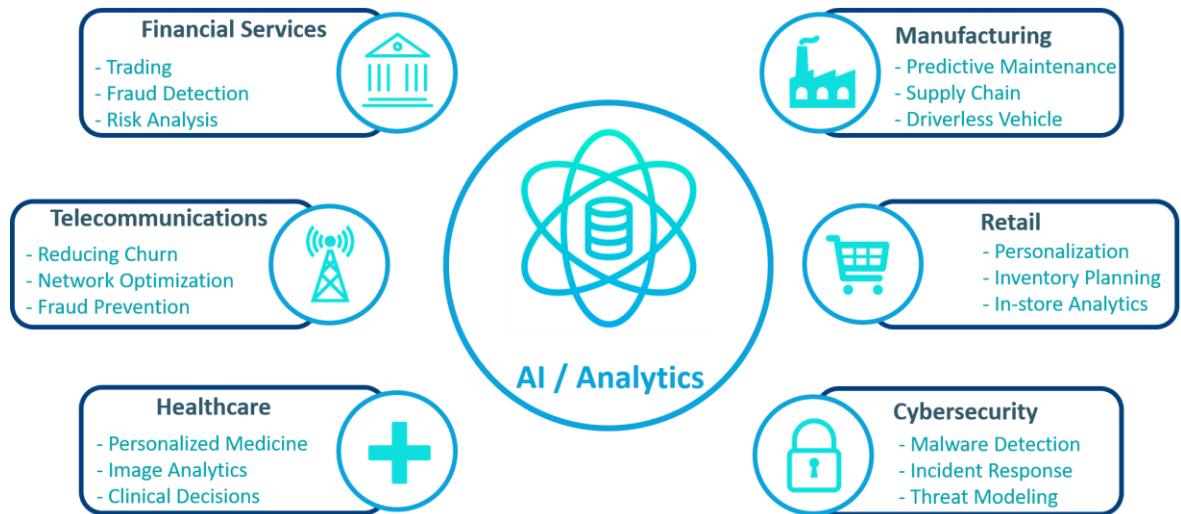

Figure 1: High Value Use Cases of Analytics and AI

Another recent survey[2] indicates that over the next five years AI is expected to have a positive impact on growth (90%), productivity (86%), innovation (84%) and job creation (69%). 77% of respondents expect AI to improve the sustainability of economic growth.

However, for Analytics, AI and ML to become a crucial integral part of an organization, numerous challenges must be overcome. In fact, 77% of respondents in another recent survey[3] say that "business adoption" of big data and AI initiatives continues to be a challenge and only 31% have a data-driven organization, fewer (28%) have a data culture.

To amplify the value of AI and make it pervasive, it is imperative that clients consider best practices and solutions that address these challenges holistically across several dimensions: Business, Process, Applications, Data and Infrastructure. Doing so will enable clients to maximize their Total Value of Ownership (TVO) from their investments. This is the goal of the IBM + Cloudera strategic alliance.

Best Practices to Maximize TVO of Analytics and AI Investments

AI is rapidly shaping the future of work by enabling organizations to predict and shape future outcomes, empower people to do higher value work, automate decisions, processes

---

[1] https://futureiot.tech/gartner-ai-adoption-growing-despite-skills-shortage/
[2] https://eiuperspectives.economist.com/sites/default/files/EIU_Microsoft%20-%20Intelligent%20Economies_AI%27s%20transformation%20of%20industries%20and%20society.pdf
[3] New Vantage Partners, "Big Data and AI Executive Survey 2019 Executive Summary of Findings", 2019.

and experiences, and reimagine new business models. In fact, AI pioneers see more value in the form of higher revenues (72%) and then secondarily in cost savings (28%).[4] Which is why organizations must carefully assess the total value of their AI /Analytics investments.

The TVO framework (Figure 2) goes beyond just the Total Cost of Ownership (TCO). It categorizes interrelated cost/value drivers (circles) for Analytics and AI by each quadrant: Costs, Productivity, Revenue/Profits and Risks. Along the horizontal axis, the drivers are arranged based on whether they are primarily Technical or Business drivers. Along the vertical axis, drivers are arranged based on ease of measurability: Direct or Derived.
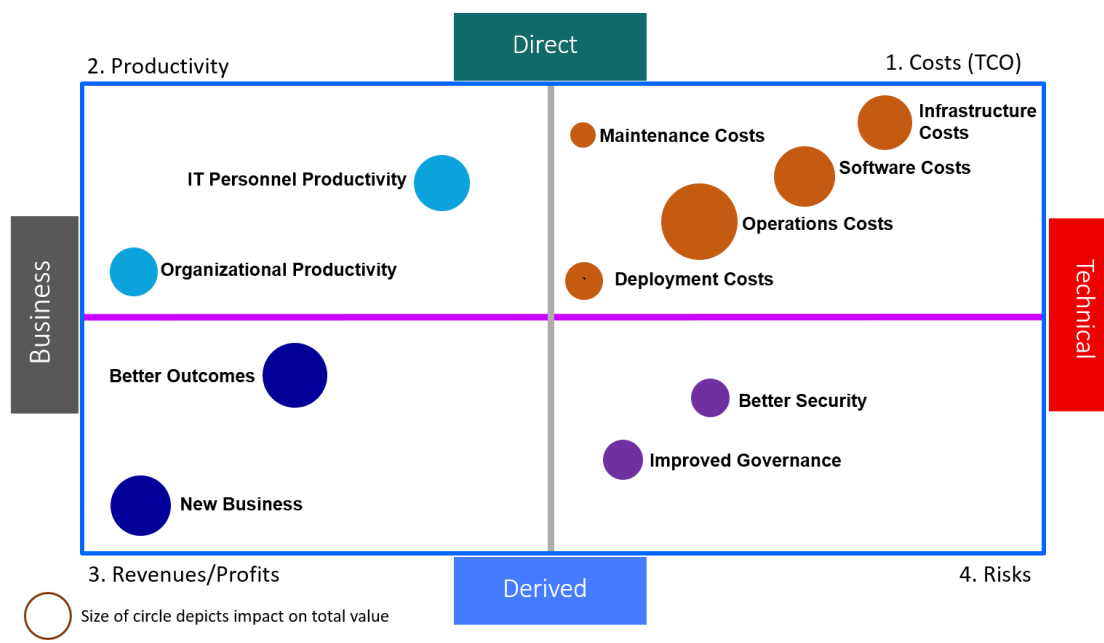

Figure 2: TVO Framework with Cost and Value Drivers for Analytics and AI

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 as follows:

1. **TCO:** Costs for infrastructure, software, deployment, maintenance, operations, etc.
2. **Enhanced Productivity:** Productivity gains of data scientists, data engineers, developers, analysts and the organization because of automation and shift to higher value work.
3. **Higher Revenue/Profits**: Better able to predict and shape future outcomes and reimagine new business models to spur growth, revenues and improve profits.
4. **Risk Mitigation**: Lower risk of project failure (even well-planned Analytics projects have up to 60% failure rate[5]) with better governance, security, privacy and compliance.

To maximize the TVO, organizations must operationalize, sustain and scale Analytics/AI. However, today, about 51% of organizations are stuck in experimentation because over 60% of organizations face challenges associated with Data, Talent and Trust.[6] To quickly identify and implement high value Analytics and AI use-cases, organizations need to overcome these

---

[4] Sam Ransbotham, David Kiron, Philipp Gerbert, and Martin Reeves, "Reshaping Business with Artificial Intelligence", MIT Sloan Management Review, 2017.
[5] 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.
[6] Forrester, "Challenges that hold firms back from achieving AI aspirations", 2019.

challenges and leverage corresponding emerging best practices[7][8][9] (Figure 3) in a consistent and repeatable manner at scale, across the business to maximize Analytics/AI value.
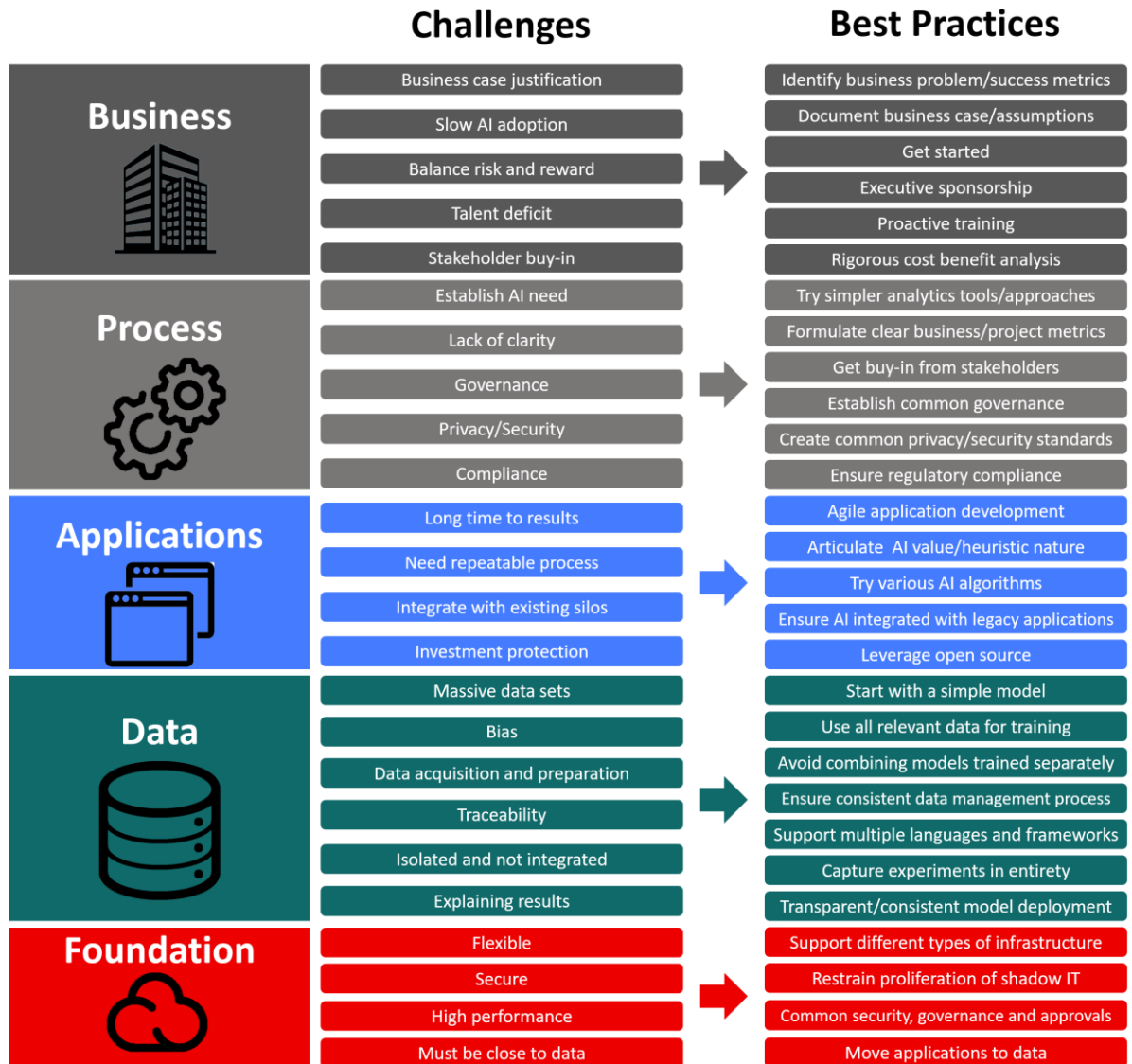
Figure 3: Challenges and Corresponding Best Practices for Analytics and AI

By implementing these best practices, customers can collaborate to *gather* data and make it simple and accessible, *arrange* to create a business-ready analytics foundation such as a data warehouse for Business Intelligence (BI), *analyze* to build Analytics and AI with trust and transparency and to operationalize and *industrialize* AI across the business.

This intricate, multistep, ladder-like journey and workflow is crucial to industrialize AI. However, this requires an open Information Architecture (IA) and data management solutions with choice and flexibility. "There is no AI without a robust and open IA". IBM and Cloudera deliver this open IA to clients on their journey to operationalize Analytics/AI.

---

[7]https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx

[8] https://blogs.oracle.com/bigdata/7-machine-learning-best-practices

[9] https://www.newgenapps.com/blog/best-practices-machine-learning-models-applications

The Analytics and AI Workflow Requires an Open Platform

Compared to the much-hyped focus on the compute-intensive training and inference tasks, there is little appreciation of the complexities and importance of data management. One of the most vexing challenges in deploying AI is how to manage all the data used throughout the workflow (Figure 4). The AI/Analytics workflow has more green data management blocks and fewer blue and red compute-intensive applications boxes.

AI algorithms become more accurate and efficient the more they get trained on large volumes of data from many sources, including valuable enterprise data. However, company-specific data is often fragmented – 80% of this data is locked in siloes and not easily accessible. So, it is important to have robust processes to regularly gather and arrange this diverse data from numerous sources and integrate it into the many constantly improving training models to deliver better insights and business value.
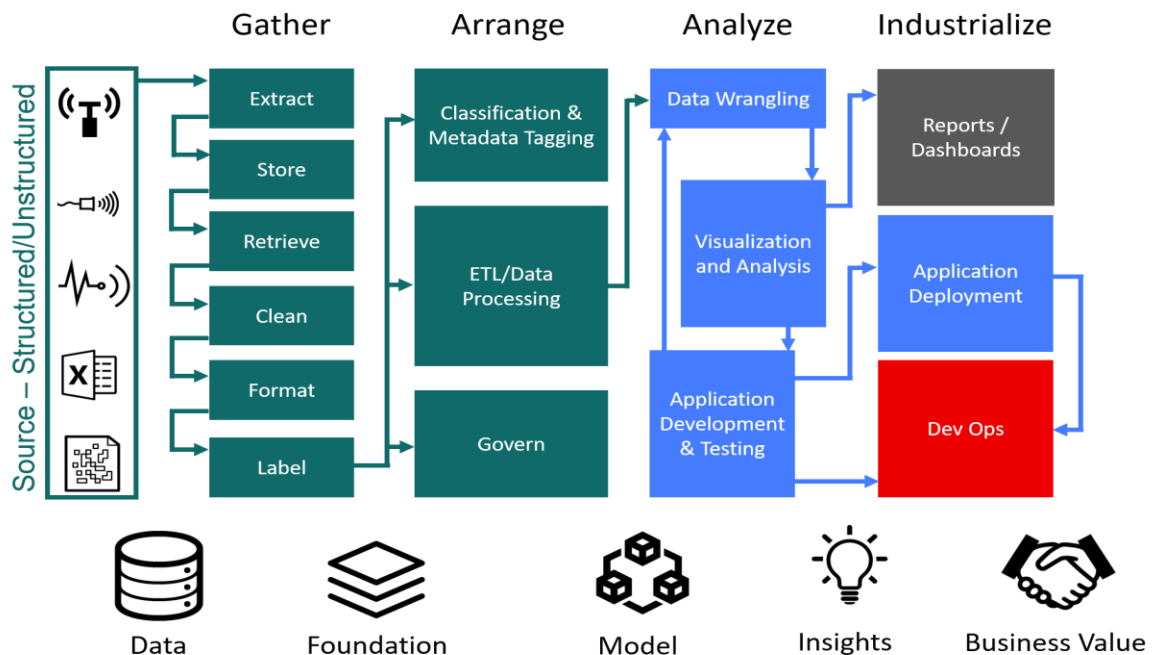
Figure 4: A Typical Intricate Iterative Analytics/AI Workflow

The various phases of a typical, intricate and iterative Analytics/AI workflow (Data to Foundation to Models to Insights to Business Value) are shown in Figure 4.

1. **Gather**: The first step in any Analytics/AI workflow is to acquire the data which can be structured or unstructured. It is important to accurately track data *provenance*, i.e., where each piece of data comes from and whether it is still up to date, since data often needs to be re-acquired in the future to run updated experiments.

   With data streaming in from hundreds of sensors, a single source (vehicle, plant equipment, building, gene sequencing machine, etc.) can produce terabytes of data each day. However, Data Scientists typically do not look at just one source. They may have to look at numerous sources and as time goes on, might have multiple Analytics/AI models, with multiple versions and hundreds of different data subsets. So, the data/storage management challenges compound exponentially.

Dealing with raw data is also not very convenient since it was generated and formatted without considering analysis requirements. Raw data often contains semantic errors, missing entries, or inconsistent formatting, so it needs to be "cleaned" prior to analysis.

This is a big challenge as data collection and preparation are very time-consuming activities – Data Scientists spend about 80 percent[10] of their time simply finding, cleansing and arranging data. So, solutions that efficiently ensure data is extracted, stored/retrieved, cleaned, formatted, labeled and readied for analysis are very crucial.

2. **Arrange**: For a robust analytics foundation, it is vital to classify, process and govern data:
   - <u>Classification and Metadata Tagging</u>: Standard metadata provides information that is adequate for a cursory understanding of data. Custom tags enable deeper analysis from high-volume, heterogeneous data by resolving data management problems and allowing users to easily search, discover, and understand data relationships to get the most value.
   - <u>ETL/Data Processing</u>: Extract, transform, load (ETL) is the process to copy data from one or more sources to a target system which holds the data in another way or context from the source(s). ETL makes it possible to extract data from multiple siloed source systems, enforce data quality and consistency standards, integrate and deliver data so that developers can build applications and end users can analyze and get insights.
   - <u>Governance:</u> As Analytics and AI are used more to answer important questions, the need to govern all IT assets (data, applications, processes, infrastructure, etc.) and ensure security, privacy and regulatory compliance increases. In fact, recent research indicates that by 2022, 65% of enterprises will task CIOs to transform and modernize governance policies to seize the opportunities and confront new risks posed by AI, Machine Learning, and data privacy and ethics.[11] Every organization must constantly check for good data quality from source systems and prevent unethical uses, inaccurate outcomes and prejudices perpetuated by faulty data and biased algorithms.

3. **Analyze:** Data wrangling, data visualization and application development and testing.
   - <u>Data Wrangling:</u> In addition to the tasks detailed in the Gather and Arrange phases, additional data handling may be required before Data Scientists can dive deep into analysis to build the application or training model. These additional tasks could include discovering, structuring, cleaning, enriching, validating and publishing data, rules and other documentation to help operationalize Analytics/AI downstream.[12]
   - <u>Visualization and Analysis:</u> Helps identify significant patterns and trends in the data and analysis results to gain better insights through simple charts like line, bar or other charts. This also aids the Data Scientist to analyze and investigate patterns and trends deeper with other algorithms and approaches to get more detailed insights from the data.
   - <u>Application Development and Testing:</u> Developing Analytics/AI applications requires a heuristic and iterative approach. Data scientists conduct many experiments with several algorithms to arrive at a practical training model which expresses the relationships between various types of data. Developers use these models to build new applications or to feed them the appropriate data from the AI application. This training model is tested and updated regularly even when in production and especially when new data is gathered to ensure that business objectives and prediction accuracy goals are being met.

---

[10]CrowdFlower, "2016 Data Science Report".
[11] https://www.idc.com/getdoc.jsp?containerId=prUS44420918
[12] https://www.trifacta.com/data-wrangling/

4. **Industrialize:** Deploy and run Analytics/AI applications to produce insights/value:
   - <u>Reports/Dashboards:</u> Provide the managers and organizations with actionable business insights to reduce costs, improve productivity, drive revenues/profits and lower risks.
   - <u>Application Deployment:</u> There are several ways to deploy applications and can be tailored to the needs of the customer/end-user. It can be as simple as generating a report or as complex as implementing a repeatable data science process. In addition, customers can deploy on-premises, on the cloud or on a hybrid cloud with pros and cons for each choice.
   - <u>Dev Ops:</u> Widely used to develop cloud services, Dev Ops enables IT operations staff, data scientists and developers to collaborate and choose the best infrastructure, automation tools and release process to deploy, support, update and scale Analytics/AI applications.

To implement Analytics/AI best practices and workflows across the enterprise, a common open collaborative cloud-like containerized environment to run all analytic processes is needed. In fact, in a recent survey[13], 69% of the enterprises indicated open source software plays a strategic role and 42% are already using it for analytics. In addition to containers, they are also using open source for cloud management, security and data bases.

Worldwide, from 2018 to 2022, the public cloud services market is expected to grow annually at 17.6%.[14] In the same period, the data services for hybrid cloud market is expected to grow a little faster at 18.4% annually.[15] Clearly, analytics is moving to hybrid clouds. A hybrid cloud combines or augments the benefits of an on-premises infrastructure with that of a public cloud and can offer a better solution for Analytics/AI workloads (Table 1).

Public Cloud	On-Premises
- 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.	- Hard to move existing data lakes into the cloud because of large data volumes - bring compute to where data resides - Inconvenient to use cloud for Edge analytics for immediate decisions - Need dedicated and secure environments for compliance to stringent regulations and/or unique workload requirements - High/custom SLA performance and efficiency.

Table 1: Making the Case for Public Cloud and On-Premises Infrastructure for Analytics/AI

An open hybrid 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, and share insights throughout the workflow.

Together, IBM and Cloudera are building solutions that recognize that "Data fuels the digital transformation, AI unlocks the value of Data and a Hybrid cloud democratizes Data".

---

[13] https://www.redhat.com/en/enterprise-open-source-report/2019
[14] https://www.gartner.com/en/newsroom/press-releases/2018-09-12-gartner-forecasts-worldwide-public-cloud-revenue-to-grow-17-percent-in-2019
[15] https://www.idc.com/getdoc.jsp?containerId=US44159818

IBM + Cloudera: Partnership to Maximize the TVO of Analytics/AI

The IBM + Cloudera strategic partnership reinforces a combined commitment to open source and cloud for Analytics/AI initiatives. It offers clients an industry-leading, enterprise-grade Hadoop distribution plus an ecosystem of integrated products and services – all designed to help organizations industrialize Analytics/AI. As a part of this partnership:

- IBM resells and supports several Cloudera products under a multi-year contract
- IBM and Cloudera provide migration assistance to several Cloudera unity products
- Cloudera resells key IBM products starting with IBM Watson Studio and IBM Big SQL.

Built on an open hybrid multi-cloud foundation (Red Hat), IBM and Cloudera provide a rich and extensive portfolio of products and services (Figure 5 – key products and services) to help clients overcome the many challenges and deploy best practices from the edge to AI.

These solutions provide clients a holistic combination of business, process, applications, data and infrastructure capabilities (Figure 5) with unprecedented flexibility and choice to reduce costs, improve productivity, drive revenues/profits and mitigate risks *i.e.* to maximize their Total Value of Ownership (TVO) of their Analytics/AI investments.
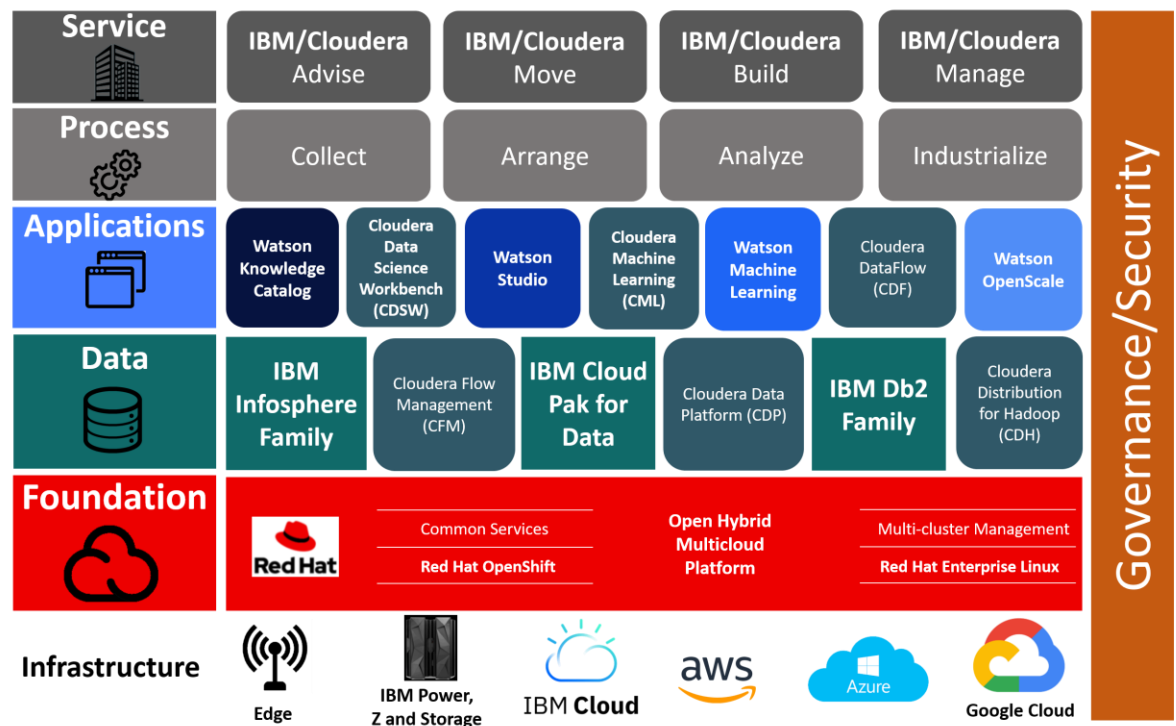
Figure 5: Non-prescriptive Conceptual Landscape of Key IBM and Cloudera Products/Services for Analytics/AI Provide Greater Choice and Value to Clients

Customers have the choice and flexibility to select some of the products/services in Figure 5 and other complementary offerings to fully deploy, customize and industrialize Analytics/AI in own their unique environments. A brief non-prescriptive description of some key IBM and Cloudera products/services is provided by layer:

Business/Process/Services: In their Analytics/AI journey (Collect, Arrange, Analyze and Industrialize), clients can fill their capability gaps from a business, process, people, technical and organizational perspective. IBM and Cloudera help organizations through the planning, feasibility, prototyping, and deployment of real projects with governance and security.

- Advise: *Cloudera* provides a white glove concierge service including AI expert and platform advising plus access to Cloudera Fast Forward Labs (CFFL) research reports, prototypes and experts onsite with client teams for 1 week/month. Cloudera also offers **s**trategy prescription for use cases, organizational and process design.

  IBM has a Data Science Elite team with deep Analytics/AI skills and expertise in industry-specific use cases: e.g. fraud detection, recommender systems, resource optimization, up-sell/cross sell, churn risk, etc. IBM offers collaborative use case discovery and design workshops with the customer, and mentors and enables client teams through hands-on learning.

- Move and Build: The *IBM* AI Learning program helps data scientists and developers work together to accelerate the process of moving to deployment and integrate AI into their applications. This program empowers a client's cross-functional team to deploy, monitor and optimize models quickly and easily. APIs are generated automatically to help developers infuse AI into their applications in minutes. Intuitive dashboards make it simple for client teams to manage models in production, and seamless workflows enable continuous retraining to maintain and improve model accuracy.

  Cloudera offers strategic guidance on its own foundational cloud and data tools and enablement services to deploy its data sciences development products in a client's environment following best practices. Clients can easily identify, build and document an application in their environment.

- Manage: *Cloudera* helps clients deploy applications in their Cloudera production environment, scale by gathering additional data and/or increasing performance of the production application, and operate this production environment by applying operational tooling, automated processes and best practices to maintain production models.

  IBM experts are available to consult and advise on multi-cloud approaches to modernize for AI/Analytics. They can provide insights on how hybrid data management can make business data simple and accessible, help clients arrange data for governance and integration and analyze data in smarter ways. They can also guide clients on how to improve deployments and explain outcomes.

  IBM also provides multi-vendor IT support services to help simplify IT support management by streamlining multiple OEM and vendor contracts to a single vendor with the expertise to care for all technology-support needs. Clients can improve availability, resolve issues quickly and reduce outages with IBM's proactive, reactive, onsite and remote support for their multi-vendor data center and across their IT environment.

Applications: Key products for AI/Analytics applications developers include:

- IBM Watson Knowledge Catalog provides a secure enterprise catalog management platform that is supported by a data policy framework. A catalog connects data and

Advisory services to help clients ramp up AI capabilities

Collaborate with clients to build, deploy and integrate AI workflows

Manage and scale AI workflows on-premises or on hybrid clouds with multi-vendor IT support

9

knowledge with the people who need to use it. The data protection policy framework ensures that data access is compliant with the business rules.

- Cloudera Data Science Workbench (CDSW) is a secure, self-service enterprise data science platform that lets data scientists manage their own analytics pipelines, thus accelerating machine learning projects from exploration to production. It enables data science teams to use their preferred data science packages to run experiments with on-demand access to compute resources. CDSW supports collaborative development and can run both in the public cloud and on-premises.

- IBM Watson Studio provides businesses with the environment and tools to solve their business problems by collaboratively working with data. They can choose the tools they need to analyze and visualize data, to cleanse and shape data, to ingest streaming data, or to create and train machine learning models.

- Cloudera Machine Learning (CML) enables clients to build, train, and deploy scalable AI /ML applications and is tightly coupled with CDSW and other Cloudera products.

- IBM Watson Machine Learning helps data scientists and developers work together to accelerate and simplify the process of moving to deployment and integrating AI/ML. It works with Watson Studio and empowers cross-functional teams to deploy, monitor and optimize models quickly and easily.

- Cloudera DataFlow (CDF) is a scalable, real-time streaming analytics platform that ingests, curates, and analyzes data for key insights and immediate actionable intelligence.

- IBM Watson OpenScale is the open platform for businesses to operationalize trusted AI and extend their deployments enterprise-wide. It provides insights into AI health, recommends next steps to improve outcomes, and orchestrates tasks to remediate issues around performance, accuracy, and fairness. It allows enterprises to automate and operate AI at scale—wherever it resides, across its entire lifecycle—with transparent, explainable outcomes. Watson OpenScale provides confidence in AI outcomes and bridges the gap between the teams that operate AI and the business units that use these applications.

Data: Key products to gather and arrange data for a business-ready analytics foundation are:

- IBM InfoSphere Family provides an advanced data integration platform that enables customers to cleanse, monitor, transform and deliver data. The scalable solution family provides massively parallel processing capabilities to help customers manage small to very large data volumes and delivers trusted information for Analytics/AI initiatives.

- Cloudera Flow Management (CFM) is a no-code data ingestion and management solution powered by Apache NiFi. With NiFi's intuitive graphical interface and 300+ processors, CFM delivers highly scalable data movement, transformation and management capabilities to the enterprise.

- IBM Cloud Pak for Data is a data and analytics platform with built-in governance. It simplifies and unifies how businesses collect, organize and analyze data to accelerate the value of data science and AI. This multi-cloud platform delivers a broad range of core data microservices, with the option to add more from a growing services catalog. It offers greater flexibility, security and control, and the benefits of the cloud without having to move a customer's data.

- Cloudera Data Platform (CDP) is an enterprise data cloud and offers a full complement of open-source data management and multi-function analytics, with the agility, elasticity, and ease of use of a public cloud-like experience. It provides a single control plane to manage infrastructure, data, and analytic workloads across hybrid and multi-cloud

environments with shared services to safeguard data privacy, ensure regulatory compliance, and protect against cybersecurity threats across all cloud environments.

- <u>IBM Db2 Family</u> delivers advanced data management and analytics capabilities for transactional and warehousing workloads and is supported across Linux, Unix and Windows operating systems. It includes advanced features such as in-memory technology, advanced management and development tools, storage optimization, workload management, actionable compression and continuous data availability.
- <u>Cloudera Distribution for Hadoop (CDH)</u> delivers the core elements of an Apache-licensed open source Hadoop – scalable storage and distributed computing – along with a web-based user interface and vital enterprise capabilities. CDH offers unified batch processing, interactive SQL and interactive search and role-based access controls.

Foundation: To avoid the pitfalls of single-vendor reliance, organizations can spread workloads across multiple cloud vendors with flexibility to use a cloud whenever they want. Deploying containers in clouds helps an enterprise use the right cloud for the right project. It also helps them unify the entire infrastructure from the edge to the data center to the cloud.

- <u>Red Hat</u> provides open source technologies to bring a consistent foundation on-premises or to any cloud deployment: public, private, hybrid, or multi.
- <u>Red Hat OpenShift</u> is a complete container application platform built on Kubernetes – an open source platform that automates Linux container operations and management.
- <u>Red Hat Enterprise Linux</u> and Red Hat OpenShift bring more security to every container and better consistency across environments.
- <u>Red Hat Cloud Suite</u> combines a container-based development platform, private cloud infrastructure, public cloud interoperability, and a common management framework into a single, easily deployed solution for clients who need a cloud and a container platform.

IBM and Cloudera also collaborate to optimize their products and services across the entire stack. For instance, CDP is optimized on IBM Power Systems and IBM Spectrum Scale Storage; offering clients even better TVO for their Analytics/AI initiatives.

Cloudera + IBM Maximize Total Value of Ownership of Analytics/AI

Together IBM and Cloudera provide unprecedented flexibility, choice and value for clients in their journey to industrialize Analytics/AI. The TVO model outlined here not only considers the TCO in the first quadrant (Figure 2) but also the benefits from the value drivers in the remaining three quadrants: Enhanced Productivity, Higher Revenues/Profits and Reduced Risks. Clients can maximize the TVO using hybrid data cloud solutions from IBM + Cloudera compared to pure public or pure private cloud alternatives by:

Lowering TCO: With a single data/analytics platform on hybrid/multi-cloud environments, the IBM and Cloudera portfolio of solutions enable applications to run on their preferred cloud platform based on economics and functionality; lowering costs. In addition, with no vendor lock-in, clients can keep options open to control costs and easily adopt new capabilities.

Migration tools allow easy movement of data between on-premises and clouds; providing customers the flexibility to choose the right infrastructure for the job. For example, customers can use public cloud infrastructure for development/test as initial provisioning is

faster and easier. To scale production applications, customers can use an on-premises data center and hybrid cloud for bursting.

IBM Cloud Pak provides full software support and helps protect the entire stack – from hardware to applications. Clients can rapidly migrate, integrate and modernize Analytics/AI applications. They are easily deployed, delivered as packages tailored for specific client use cases and offer significant reduction of operational and deployment costs. For example, IBM Cloud Pak for Automation helps transform business processes, decisions and content. A banking client was able to reduce manual processes by 80 percent[16]. IBM Cloud Pak for Multicloud Management provides multi-cloud visibility, governance and automation. It was able to help clients reduce the operational expenses of supporting large-scale cloud-native environments by 75 percent[17].

Enhancing Productivity: IBM + Cloudera solutions improve the productivity of all IT staff and the organization through automation and enable them to do higher value work with the following unique features provided by the open, hybrid multi-cloud architecture:

a. 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.

b. Flexibility and agility: Operate with equivalent functionality on and off premises and on all major public clouds; 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.

c. Self-service: Reduce time and effort to deploy common application types with self-serve public cloud-like experiences such as Data Flow and streaming, Data Engineering, Data Warehousing, Operational Database, Machine Learning and others.

d. Build once and deploy anywhere: A portable, container-based platform (Red Hat OpenShift) facilitates the iterative process to transform the business and move to the cloud and then deploy across multiple clouds in a simple and flexible way. A consistent set of management 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.

Increasing Revenues/Profits: Optimizing and integrating the solution stack on the foundation infrastructure delivers faster time to value with better scale, performance, ease of deployment and integration. In addition, support for in-memory, parallel and vector processing improve data ingestion and query performance. Other key features such as visual and SQL interfaces that are fully ANSI compliant with schema flexibility and reusable, portable Analytics/AI workflows help shape and produce better and faster outcomes.

Greater innovation and better decision-making capabilities with trusted data help clients improve customer service and support, reduce time to market, build better pricing models, identify new customers/markets/products/services, and more.

Mitigating Risks: Maintaining strict enterprise data privacy, governance, data migration, and metadata management across all environments lowers risks. CDP and IBM Cloud Pak for Data provide a single control plane to manage infrastructure, data, and analytic

---

[16] https://www.ibm.com/downloads/cas/O5A0BD4R
[17] http://ibm.biz/Ovum-WP

workloads across hybrid and multi-cloud environments. They also drive automation to dozens of application patterns and lifecycles and provide common security, governance, compliance standards and operations across the board. IBM Cloud Paks have gone through stringent vulnerability and best-practices scans, so customers can be assured that they are getting up-to-date software that secures the entire stack.

Cloudera has 1000s of existing clients getting substantial value from their Hadoop platform optimized to manage the sheer volume and variety of data that will only grow in the future. Fortunately, this platform can scale easily on affordable clusters – on-premises or on the cloud – with optional accelerators such as Graphics Processing Units (GPUs). Hadoop can also easily integrate with many other Big Data frameworks. This gives clients lifetime access and availability to their data – one of their most strategic assets that they can monetize.

Conclusions and Recommendations

Analytics and AI are profoundly transforming how businesses and governments engage with consumers and citizens. To amplify the value of AI and make it pervasive, it is imperative that clients consider best practices and solutions that address these challenges holistically across several dimensions: Business, Process, Applications, Data and Foundation. This intricate, multistep, ladder-like journey and workflow is crucial to industrialize AI.

IBM and Cloudera provide unprecedented flexibility, choice and value for clients in their journey to industrialize Analytics/AI and maximize their Total Value of Ownership (TVO). Compared to pure public cloud alternatives, these hybrid multi-cloud solutions lower TCO, enhance productivity, increase revenues/profits and lower risks for a broad spectrum of Analytics/AI workloads. Further, client investments in data lakes are protected lifetime.

Clients deploying Analytics/AI workflows should seriously consider solutions from the Cloudera and IBM strategic partnership for investment protection and the following reasons:

1. Reinforces a combined commitment to open source and hybrid cloud – both growing in use for Analytics/AI initiatives
2. Offers an industry-leading, enterprise-grade open platform plus an ecosystem of integrated products and services – all designed to help industrialize Analytics/AI
3. Provides a cloud-native architecture including containerized workloads, microservices and multi-cloud provisioning
4. Avoids vendor lock-in that is typical of many public cloud service providers
5. Supports cloud open source platforms/applications such as Kubernetes, open source frameworks and databases such as Hadoop, MongoDB and many open source programming languages, deep learning frameworks, and open source interfaces
6. Amplifies choice and value for clients on their journey to *gather* data and make it simple and accessible, *arrange* to create a business-ready analytics foundation such as a data warehouse for Business Intelligence (BI), *analyze* to build Analytics and AI with trust and transparency and to operationalize and *industrialize* AI across the business.