

IBM Data and Analytics Strategy

# Field Guide

IBM

Download the current version of the IBM  
Data and Analytics Strategy Field Guide



<https://www.ibm.com/cloud/garage/content/field-guide/data-analytics-field-guide>

# Transform your business with data, analytics, and cloud

Businesses need to move rapidly. Data and the related analytics are key to differentiation, but traditional approaches are often difficult and brittle. This can result in delays, business challenges, lost opportunities, and the rise of unauthorized projects.

You need a strategy to connect your business plan and outcomes to your data and analytics requirements. We can help you develop the roadmap to quickly achieve success.

## UNWIND THE DATA HAIRBALL

**Make data simple and accessible.** Collect your data, no matter where it lives, achieving freedom from ever-changing data sources.

**Build a trusted analytics foundation.** Organize your data into a trusted, secure, business-aligned source of truth.

**Scale insights on demand.** Analyze data in smarter ways to make decisions informed by evidence-based insights.

---

## What's inside?

This field guide provides a high-level overview of IBM's data and analytics approach.

### LEARN IT

A summary of the concepts.

### GET STARTED

Tips to modernize your data and analytics strategy.

# IBM's unique approach

Regardless of where you are on your data and analytics journey, IBM® Garage™ Methodology provides the strategy and expertise to accelerate your business outcomes.

Working with your teams and your data, in an iterative and agile approach, we define, design, and prove a solution with a continuously refined architecture and a roadmap to your success.

## DEVELOP A ROADMAP FOR A SMARTER BUSINESS

**Fast.** Accelerates transformation and alignment of your data and analytics initiatives to new and emerging business requirements.

**Focused.** Delivers early results in support of business outcomes with IBM and Partner expertise.

**Flexible.** Adapts to the desired speed of the business with lessons learned throughout and a continuously updated roadmap and action plan.

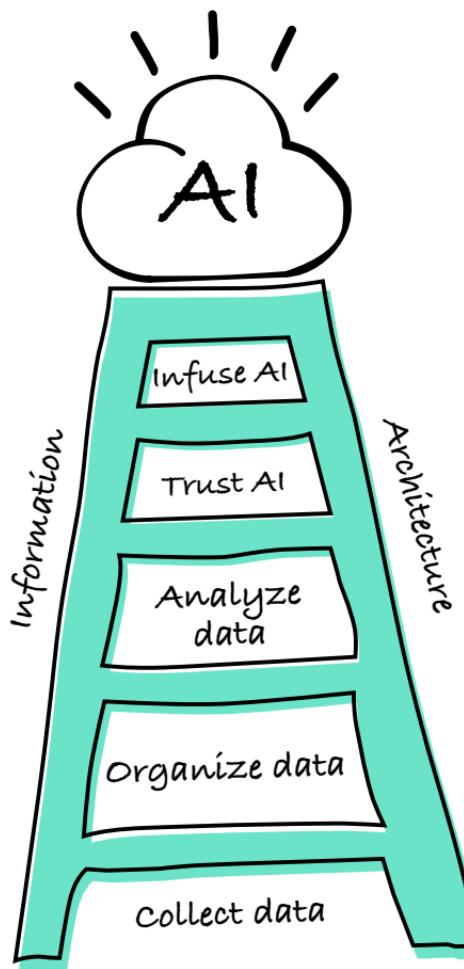


Learn more

Check out the IBM Analytics Services.

<https://www.ibm.com/analytics/services>

Deliver early results!



**The Garage Methodology** is your way to deliver accelerated results for your business objectives in your data and analytics journey.

# Data and analytics journey

No matter where you are on your data and analytics journey, you need to start with your business needs and objectives. With the proven practices of the Garage Methodology, our experts can help you assess, plan and prioritize, and build an architecture roadmap and action plan to deliver accelerated results for those objectives.

Our experts help you look holistically at your needs to deliver accelerated business outcomes with a roadmap for continued progression and success.

## IT'S YOUR JOURNEY, WE'LL MEET YOU WHERE YOU ARE

**Business initiative.** Start with your business outcomes, not technology.

**Focus and plan.** Together we define the focus and build the roadmap and action plan.

**Engage and prove.** We deliver early results working with your users, your data, and our combined expertise.

**Adapt, expand, iterate.** Leveraging lessons learned throughout, we adapt the architecture, roadmap, and actionable next steps.

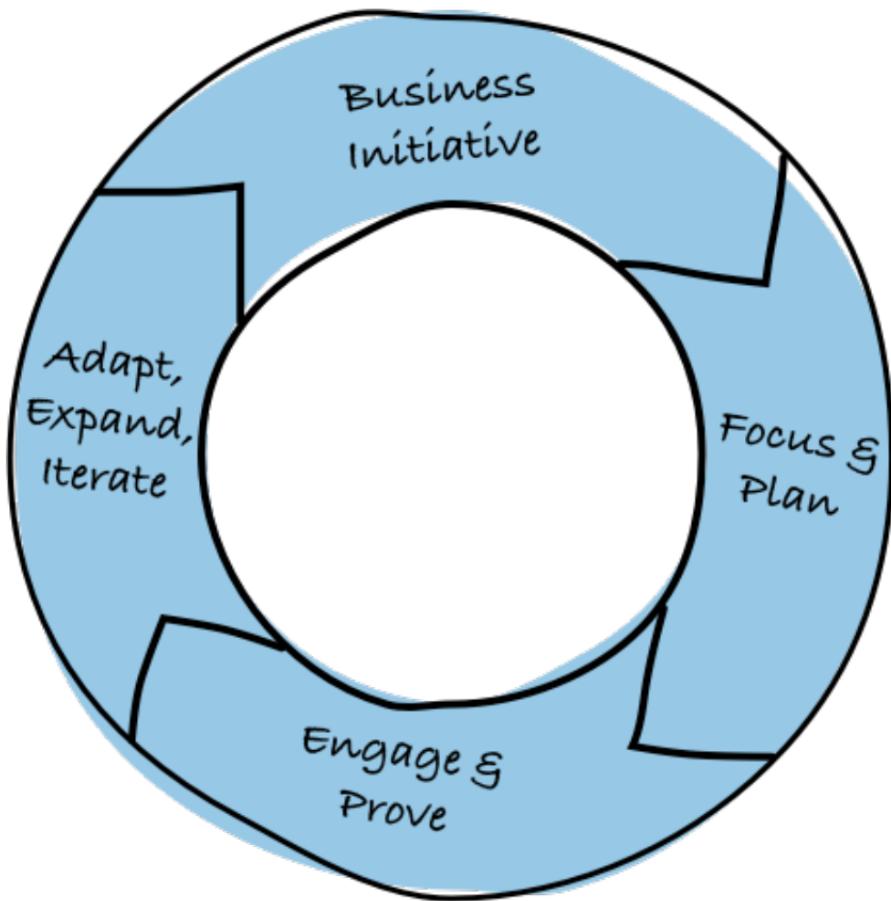


Learn more

Check out IBM Analytics Services.

<https://www.ibm.com/analytics/services>

Start where you are!



Deliver accelerated results and iterate on your roadmap  
with the method

# The three pillars of IBM Analytics

To improve efficiencies and lower cost, data needs to be easily accessible. To build trust and confidence in data, you need a way to better manage the data lifecycle and empower teams to become more self-reliant and data-driven.

## DATA FOR A SMARTER BUSINESS

**Make data simple and accessible.** Collect all of your data, no matter where it lives. Access your data freely without the constraints of fragmented and ever-changing data sources.

**Build a trusted analytics foundation.** Discover all types of data with confidence and build trust by integrating, cataloging and securing your data with built-in compliance and protection controls to deliver a single version of the truth.

**Scale insights on demand.** Scale the discovery of new insights using advanced analytics and data science to deliver better answers and smarter business processes.



Learn more

Check out IBM Analytics.

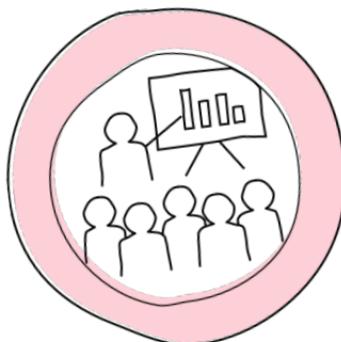
<https://www.ibm.com/analytics>



Hybrid data  
management



unified governance  
and integration



Data science and  
business analytics

The capabilities you need to modernize your data strategy.

# Make data simple and accessible

In today's digital enterprise, fast and actionable insights are the goal. Hybrid data management enables enterprises to drive innovation by seamlessly accessing, sharing and analyzing data.

To achieve this level of data management, all types of data—structured, semi-structured or unstructured—should be easily accessed and analyzed whether it's stored on-premises, in a public cloud, in a private cloud, or in any combination.

## GAIN VALUE FROM YOUR DATA WITHOUT LIMITS

**Adapt current data management to changing business needs.** Drive innovation across your business by integrating new cloud-based technologies with existing infrastructure.

**Leverage data of all types and sources with hybrid data management.** Yield new insights from a range of new data types and sources—including social, mobile, web and Internet of Things (IoT)—alongside your historical and operational data.

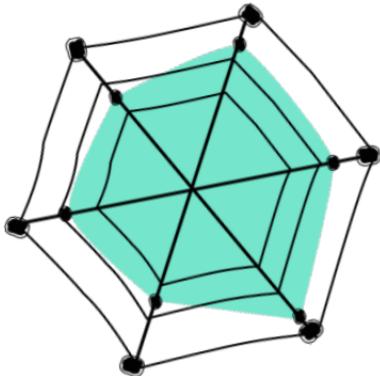
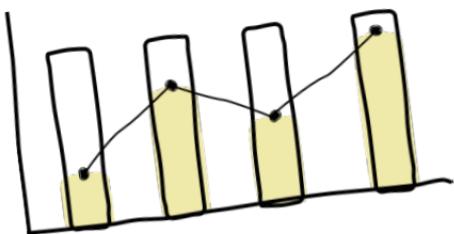
**Embrace open source technologies and data more securely.** Leverage open source innovation in a secure manner using IBM's technology and expertise.



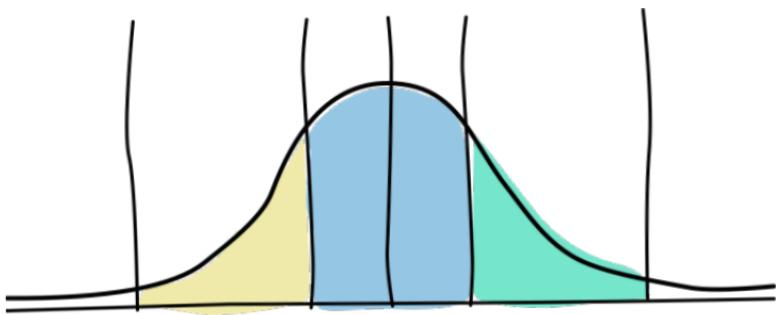
Learn more

Check out IBM's Hybrid Data Management.

<https://www.ibm.com/analytics/data-management>



001010  
11010110  
1100111



Achieve freedom from ever-changing data sources.

Leverage all data, structured or unstructured,  
wherever it resides.

# Build a trusted analytics foundation

Discover, integrate, and govern data to drive better business insights while mitigating compliance risks. Unified governance and integration empowers you to locate and trust your private, public, and partner data.

Governance and integration delivers the capabilities your organization needs to discover, profile, integrate, cleanse, protect, and catalog your critical business information assets throughout the data lifecycle.

## **KNOW YOUR DATA. TRUST YOUR DATA. USE YOUR DATA.**

**Trust your data.** Build confidence in your data with a holistic approach to managing, improving, and leveraging information.

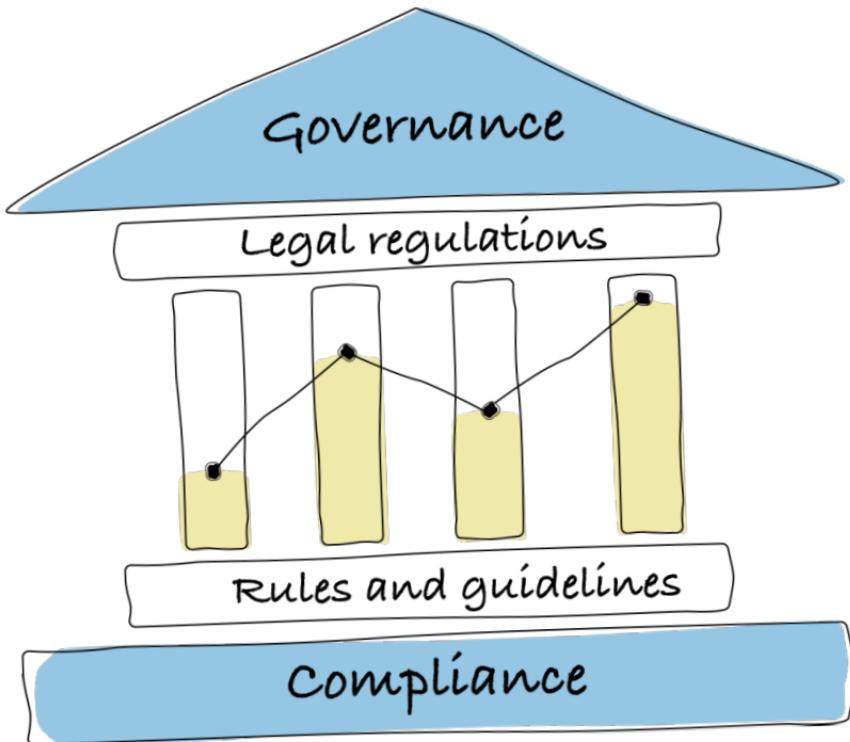
**Data integration, replication & quality.** Deliver, transform, and cleanse data from disparate sources with varying formats.

**Enable information driven insights.** Empower every data user and line of business leader with a high-value, 360-degree view of trusted data to drive business insights and intelligence.



Learn more

Check out IBM's view on Unified Governance and Integration.  
<https://www.ibm.com/analytics/unified-governance-integration>



Build a trusted analytics foundation to drive better  
insights and compliance.

# Cognitive classification

Data is more than its physical representation. Identifying the actual meaning of data in your overall information landscape provides a common understanding and allows you to protect, find, and reuse your data.

Manually assigning meaning to all technical data elements is a time-consuming and expensive operation. Automating this process with new innovations in data classification, natural language understanding, and machine learning, enable data stewards to focus on review, training, and curation.

## UNDERSTAND YOUR DATA

**Discover & ingest.** Find data sources that help you drive business transformation and innovation.

**Data classification.** Automate the classification of data based on machine learning.

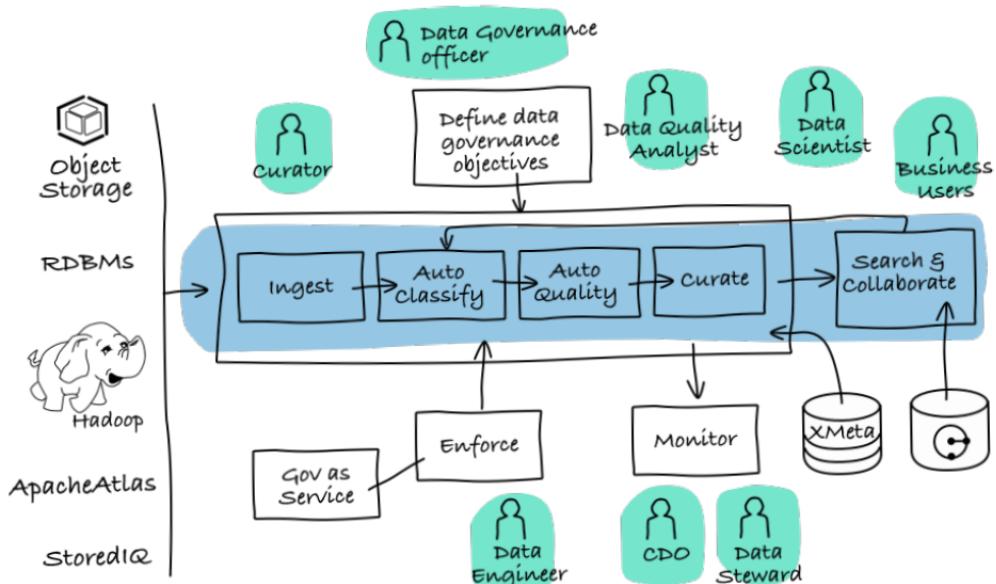
**Data quality.** Use tools to automatically assess the format and quality of data.



Learn more

Check out IBM's view on Unified Governance and Integration.

<https://www.ibm.com/analytics/unified-governance-integration>



Use machine learning to curate your data throughout the data lifecycle.

# Protect data while you put it to work

Data privacy controls how information is collected, used, shared, and disposed of, in accordance with policies or external laws and regulations. Data privacy is driven by business needs (corporate IP, loyalty obligations, protecting customers and their personal data) and regulatory compliance requirements (Healthcare, Personally-Identifiable, GDPR, Financial—formal laws & guidance, processes and reporting). Data topology provides the context for implementation with new ways of thinking, instead of traditional “one size fits all” approaches.

## PROTECT DATA; PREVENT BREACHES

**Meeting regulatory and compliance obligations.** Enact and assess the measures you are taking to ensure you are meeting regulatory requirements for protecting data for every use case within the data topology landscape.

**Privatize personal and business-sensitive information.** Implement direct and in-direct data privatization approaches for customer and employee personal data based on your use cases and data zones.

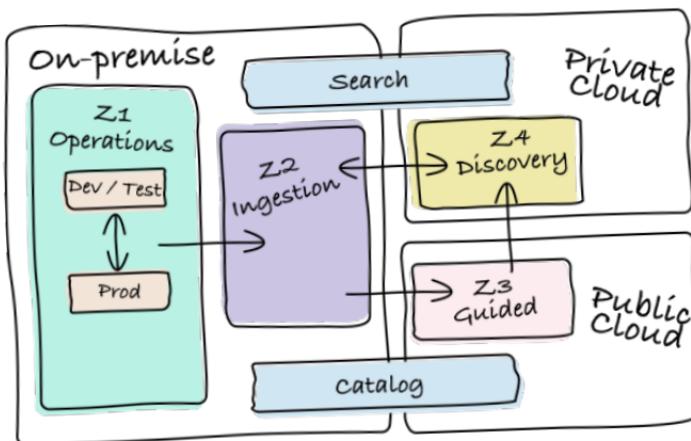
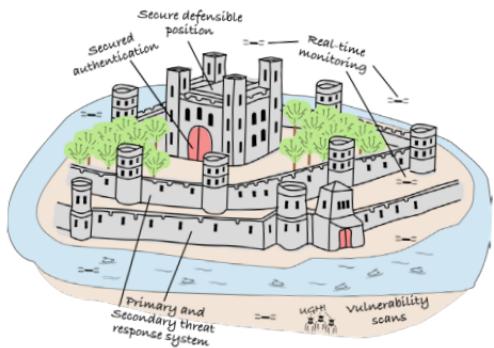
**Secure your data.** Physical and logical access controls along with activity and behavior monitoring provide the hardening approaches that wrap around and support privatizing and regulatory behaviors.



Learn more

Check out IBM's view on data governance.

<https://www.ibm.com/analytics/data-governance>



The right data protection approach starts with building a data topology to map the users, use, and flow of data in support of business objectives

# Meeting regulatory and compliance obligations

Business' need to build a trusted analytics foundation to drive better insights and compliance. Data privacy controls how information is collected, used, shared, and disposed of, in accordance with policies or external laws and regulations. Privacy laws differ by industry and geography, place strict requirements on the management of data, and must be accounted for in applications across the globe.

## YOU CAN'T BE COMPLIANT WITHOUT...

**Identity resolution.** Understanding and locating individuals, organizations, and their relationships (both obvious and hidden) is key to providing an understanding of your data across large, sparse, and disparate collections of data.

**Data lineage.** You have to trust your data. Trace your data's lineage to understand where it originated and how it was transformed everywhere it was used.

**Audit trails.** Maintain detailed records of changes and links to the data over time.



Learn more

Check out IBM's view on data governance.

<https://www.ibm.com/analytics/data-governance>



It's all about protecting personal data.

# Privatizing your data

The right data protection approach starts with building a data topology to map the users, use cases, and flow of data in support of your business objectives. Data protection strategies and technical approaches must be in sync with the data topology, applying the right techniques for the respective zones and use case(s).

Data topology zones span the spectrum from development to test to production. As data use straddles the balance of data privatization and statistical data relevance, this spectrum has become less black and white and more ‘continuous’.

## DATA PROTECTION AND CONTINUOUS DELIVERY

**Testing and development.** In the test and development zone, there is more tolerance for some statistical integrity loss. Lower, narrower environment monitoring is required. Employ traditional masking and anonymization techniques.

**Analytics and data science.** Sandbox model development requires more advanced de-identification techniques, including assessing re-identification risks. There is a lower tolerance for the loss of statistical integrity and reliability, and a need for lifecycle of data management.

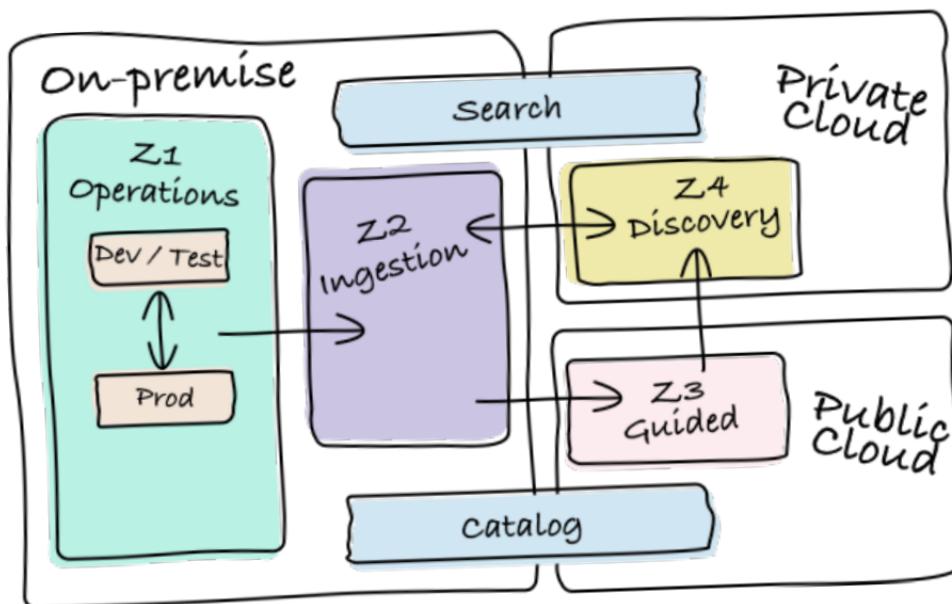
**Production.** In production, there is no tolerance for any loss of statistical integrity or reliability. Controlled, reversible pseudonymization along with the highest levels of monitoring and access control are required.



Learn more

Check out IBM's view on data governance.

<https://www.ibm.com/analytics/data-governance>



Data protection must take into account the blurring of lines between testing and production use cases.

# Personal data protection

New regulations around personal data use are concerned with breaches of personal data. Violations can be dealt with severely. Advanced approaches and techniques such as data pseudonymization and anonymization, in addition to traditional data masking and fabrication techniques, must be added to your data protection ‘tool kit’ to address the right protection for the right uses, while still supporting business and regulatory compliance requirements.

## WHAT'S IN YOUR DATA PROTECTION TOOLKIT?

**Direct identifiers.** Name, social security number, and driver's license number can directly identify a person. Apply masking techniques and pseudonymization de-identification procedures to address direct identifiers.

**Indirect identifiers.** Age, location, health condition, and purchases are all examples of publicly observable data that can indirectly identify an individual. Indirect identifiers can be linked together in original or supplemental data stores and pose the risk enabling the identification of an individual person.

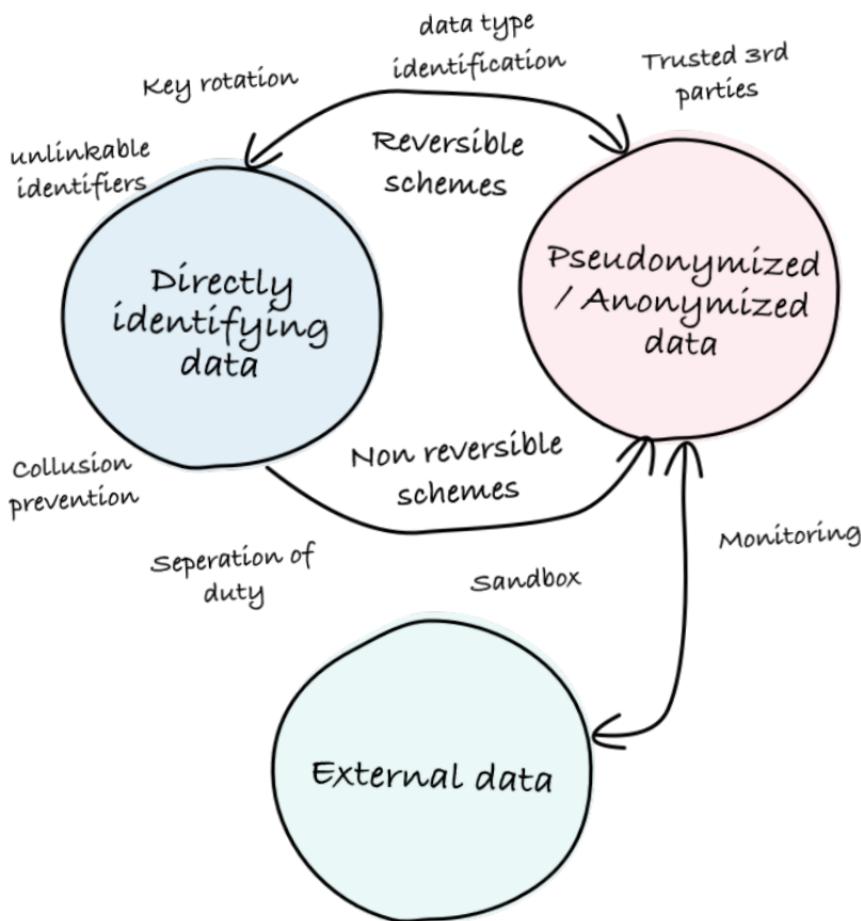
**Re-identification risks.** Applying the right amount of change to direct and indirect attributes while still preserving analytical value requires the on-going measurement of the risks of re-identification and the application of proper technical and organizational measures of data anonymization and protection.



Learn more

Check out IBM's view on data governance.

<https://www.ibm.com/analytics/data-governance>



Use data de-identification techniques to mitigate the risk of a data breach, unwanted re-identification of data, and leakage.

# Secure your data

Data is a critical resource in any organization. If it's lost, compromised, or stolen, the effects on your business can be devastating. You need to design security from the start, and have processes and systems in place to protect data - whether at rest or in transit - from unauthorized access, misuse, or disclosure.

Your defense must include rings of security where each layer ensures data confidentiality, availability and integrity by enforcing access controls, allowing time for reactionary strategies, creating a robust SIEM (Security Information Event Management) and enabling flexible response capabilities.

## SECURITY CANNOT BE AN AFTERTHOUGHT

**Enforce data protection.** Implement data encryption, data access control, key management, network security, security monitoring, and certification management as critical elements of your data protection solution.

**Identify vulnerabilities.** Routinely assess the environment for vulnerabilities so that you can identify and remediate issues fast.

**Implement security by design.** Design security at all levels of the architecture. Focus on security capabilities at the start of the process, not the end. Security must partner with the business teams that are driving innovation. If a breach occurs, you need to be able to quickly identify the nature and impact with security forensics.



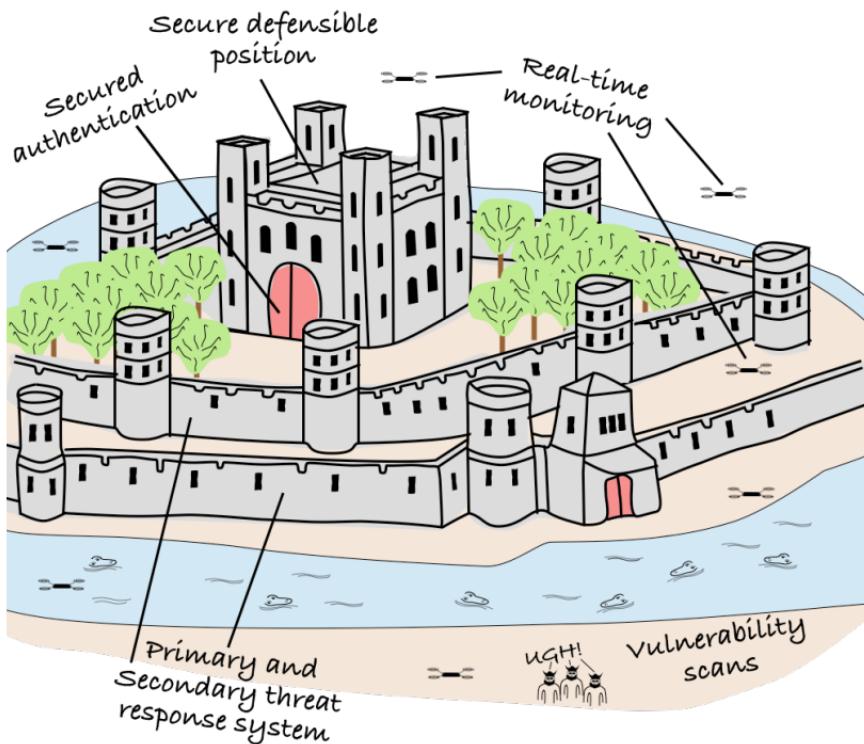
Learn more

Check out the IBM data and analytics architecture.

<https://www.ibm.com/cloud/garage/architectures/dataAnalyticsArchitecture/security-for-data>

Restricted public access

Rings of security strategy



A secure data strategy is built upon rings of defense.

# Crossing the chasm from data management to AI for the enterprise

The path to AI is a combination of specific roles, new processes, and enabling technologies. AI is a collaborative, team-based effort. The technology sandboxes deployed must support data in many forms, ensure data protection, maintain data privacy and security.

## START YOUR AI JOURNEY

**Build your team.** Include the data scientist, business analyst, data engineer roles on your cross-functional team. It is key to have someone who can understand, present, and interpret results to the end-users and business stakeholders.

**Define the use case in data science terms.** Align the practical application of data science with the goals of the business. How accurate are the predictions? How helpful are the created clusters in support of a digital engagement campaign? How effective are the measurements?

**Don't start before you have all relevant data.** Gather the internal data, data from other sources, and external data. For example, weather and social need to support your data science use case. Leverage automation to assess data quality and completeness.

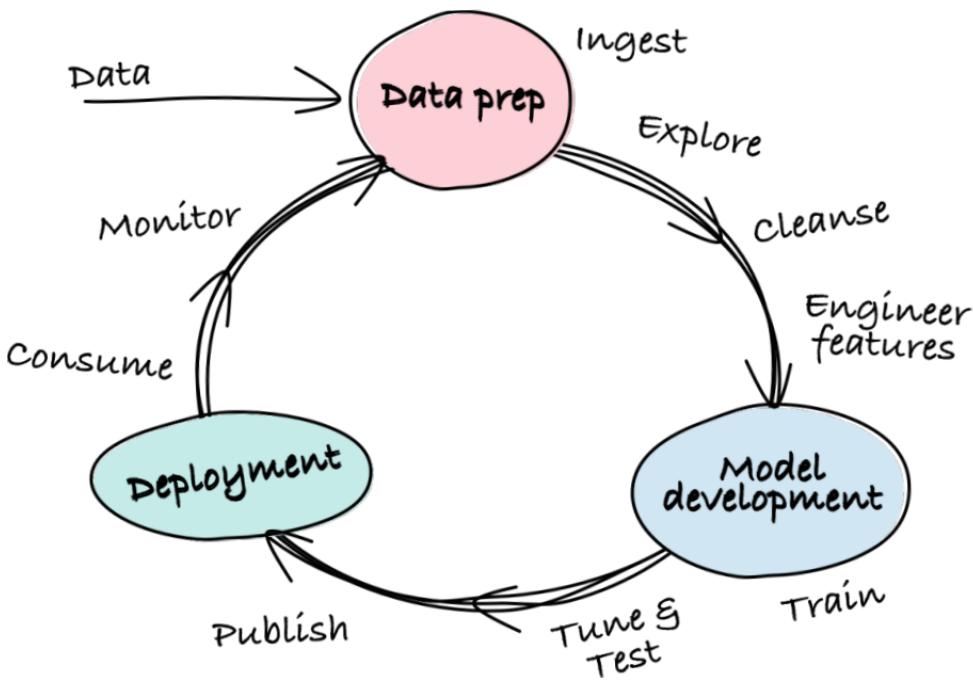


Learn more

Check out IBM's data science and advanced analytics.

<https://www.ibm.com/analytics/data-science>

Iterate, iterate, iterate!



Data science and AI is an ongoing, iterative process across data preparation, model development, and model deployment.

# Putting AI to work

After you've assembled the team, defined the business use case, and gathered the data, how do you put AI to work?

Discover patterns in data and anticipate what is likely to happen next; operationalize with machine learning to deliver faster, better business insights at scale.

## TRANSFORM DATA INTO BUSINESS OUTCOMES

**Prepare data.** Combine, transform, and prepare new features to build and train the data science models.

**Gain insights.** Build descriptive, prescriptive, and predictive models across all types of data to optimize complex business decisions. Apply descriptive knowledge to uncover relationships and create market segments, and predict future events using data mining and data science technology.

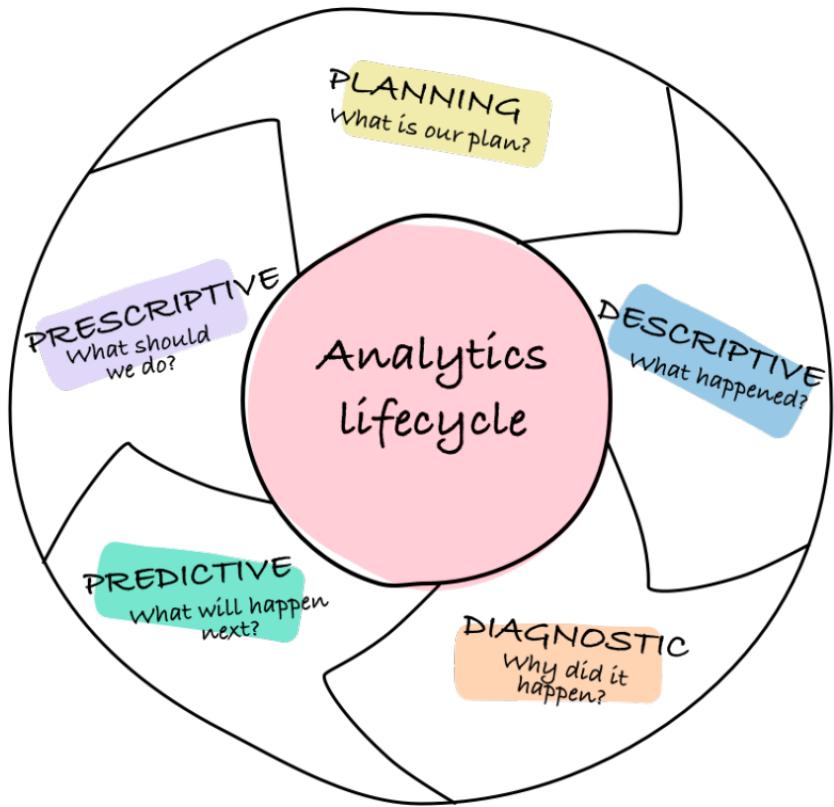
**Drive innovation with analytics.** Operationalize insights, build new data-driven applications, and act on real-time analytics. Feedback, model management, and retraining are essential to maintaining accuracy with ever-changing data.



Learn more

Check out IBM's data science and advanced analytics.

<https://www.ibm.com/analytics/data-science>



Leverage data science and analytics to understand the present, predict the future, and affect business outcomes.

# Data and analytics architecture

A complete information architecture for artificial intelligence should encompass data, analytics, and applications. Hybrid data management, data science, and visualization are core platform components in support of this architecture. The intersections of each domain are enabled through unified governance and integration.

## YOU CAN'T HAVE AI WITHOUT AN INFORMATION ARCHITECTURE

**Focus and plan.** While focus is the key to deliver early results, you need to consider data in a holistic context, and build a roadmap that ensures that what you develop today can grow and expand without significant future rework. Your action plan should allow business to “keep the lights on” while augmenting and transitioning to the future-state information architecture.

**Build a data strategy.** Connect your business plan and outcomes to your data and analytics requirements. Data strategy, data topology (classification, organization and flow), governance and common data design are the keys to becoming a data-driven organization that can quickly adapt to new business requirements and platform technologies.

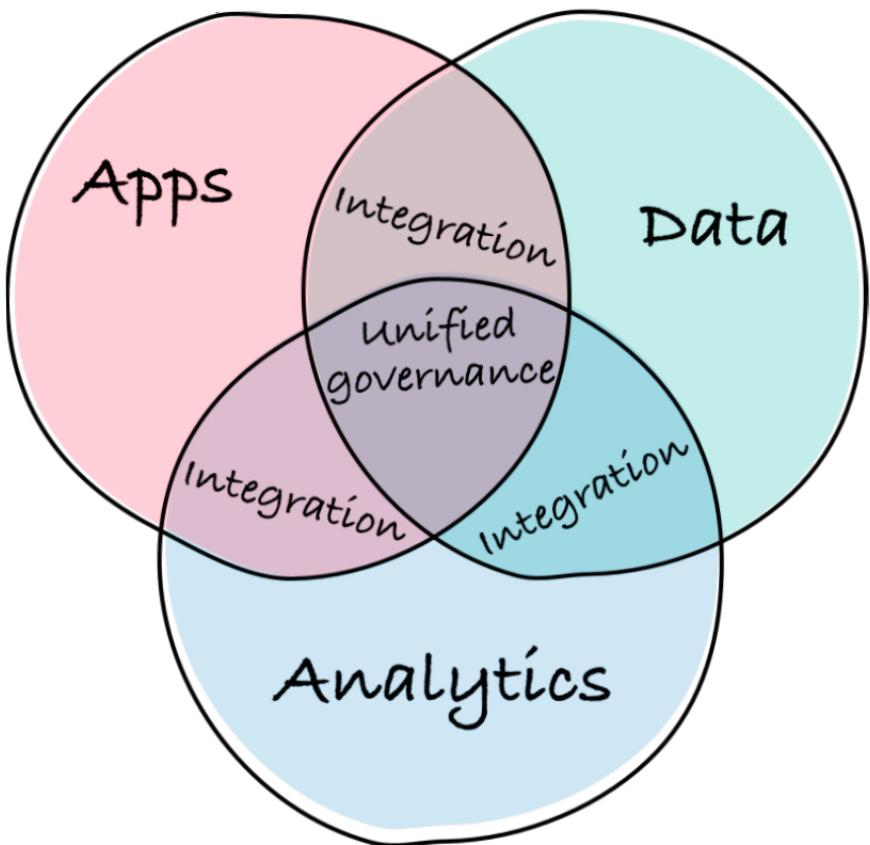
**Leverage proven architectures.** Modern architectures are built on proven platforms, patterns from real-world implementations, and value-driven business outcomes.



Learn more

Check out the IBM Data and analytics architecture.

<https://www.ibm.com/cloud/garage/architectures/dataAnalyticsArchitecture>



Jump start your data and analytics strategy with a proven  
reference architecture.

# Multicloud is key to modernization

Multicloud deployments deliver the broadest degree of flexibility and choice, protecting past and future investments. With cloud automation and management, it is easy to provision environments for data science, data engineering, and building applications.

## PROTECT YOUR CURRENT AND FUTURE INVESTMENT

**Open, standards-based approach.** Build capabilities that are open by design, enabling client flexibility and reducing vendor “lock-in”. Enable developers and operations staff to rapidly build, run, and manage all types of applications.

**Private cloud.** Get the benefits of public cloud in your data center, including rapid deployment, scalability, ease of use and elasticity.

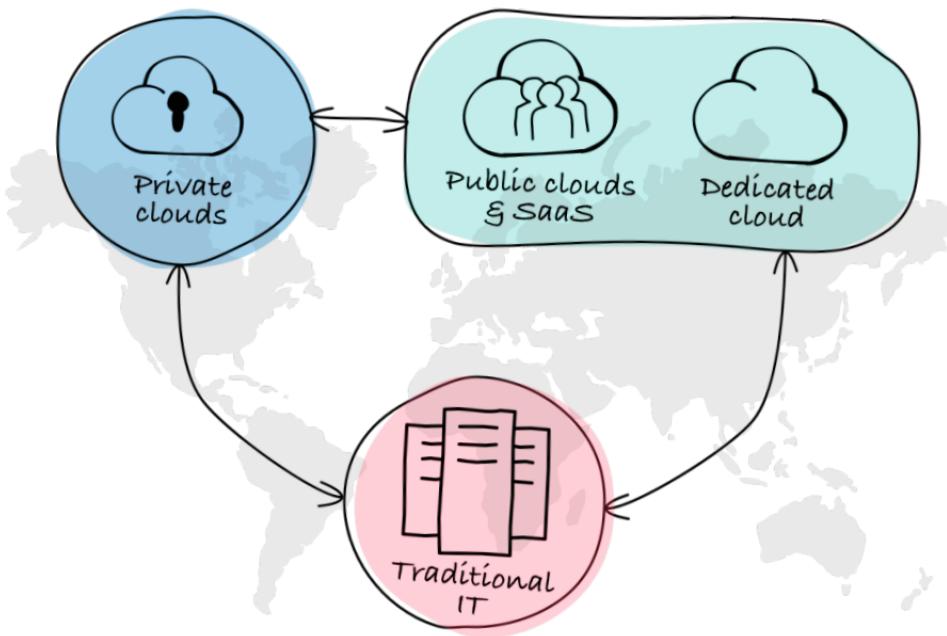
**Multicloud.** Take advantage of all the cloud has to offer and protect your existing investment by seamlessly integrating new capabilities with your existing enterprise applications and data.



Learn more

Check out the IBM Cloud®.  
<https://www.ibm.com/cloud/>

It's about choice!



**Flexible deployment based on your users, data, and workloads.**

# IBM Cloud Pak for Data

Natively built on Red Hat® OpenShift® Container Platform, IBM Cloud Pak® for Data provides optimized hardware to increase container performance and accelerates time to value of your workloads.

## PREPARE DATA FOR AN AI AND MULTICLOUD WORLD

**Eliminate data silos: Connect all data.** Handle big data and make queries across multiple data sources fast and simple without moving data. Manage all of your data wherever it lives.

**Operationalize AI with trust and transparency.** Improve outcomes with trusted data. Simplify data governance so data scientists, developers, engineers, and business experts can improve business outcomes.

**Automate and govern the data and AI.** Gain visibility into your data landscape. Benefit from tools that manage, mask, and prevent data misuse. Simplify data policy, protection, and management with automated tools and machine learning.

**Avoid lock-in and run anywhere.** Deploy IBM Watson® anywhere, on IBM Cloud, third-party clouds, or hybrid clouds.

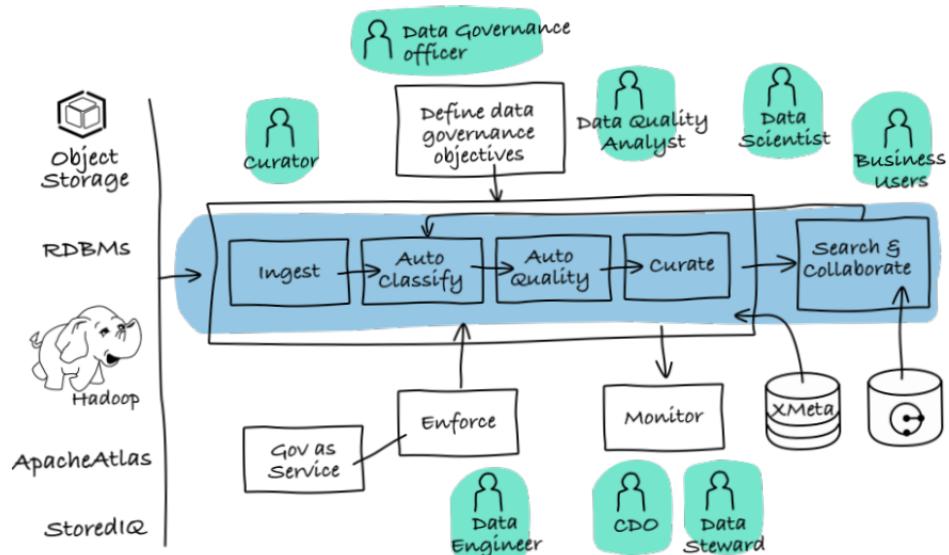


Learn more

Check out the IBM Cloud Pak for Data.  
<https://www.ibm.com/products/cloud-pak-for-data>



Eliminate data silos.  
Connect all data.



Collect and organize your data to generate insights and set the foundation of AI-driven business.

# Your launchpad: Discovery Workshop

The IBM Discovery Workshop is the first step to deliver accelerated results for your business outcomes with data and analytics.

In this collaborative working session, together we take a selected business need and develop it into an actionable use case with a high level architecture, roadmap, and action plan.

## YOUR PLAN FOR EARLY RESULTS

**Focus.** Focus on your highest priority business initiative to drive early business results and set the context for future success.

**Collaborate.** Together, in support of the focused business initiative, we capture the current state of producers and consumers of data, along with flows, systems, and pain-points. Then we build a high level architecture to support the desired outcome.

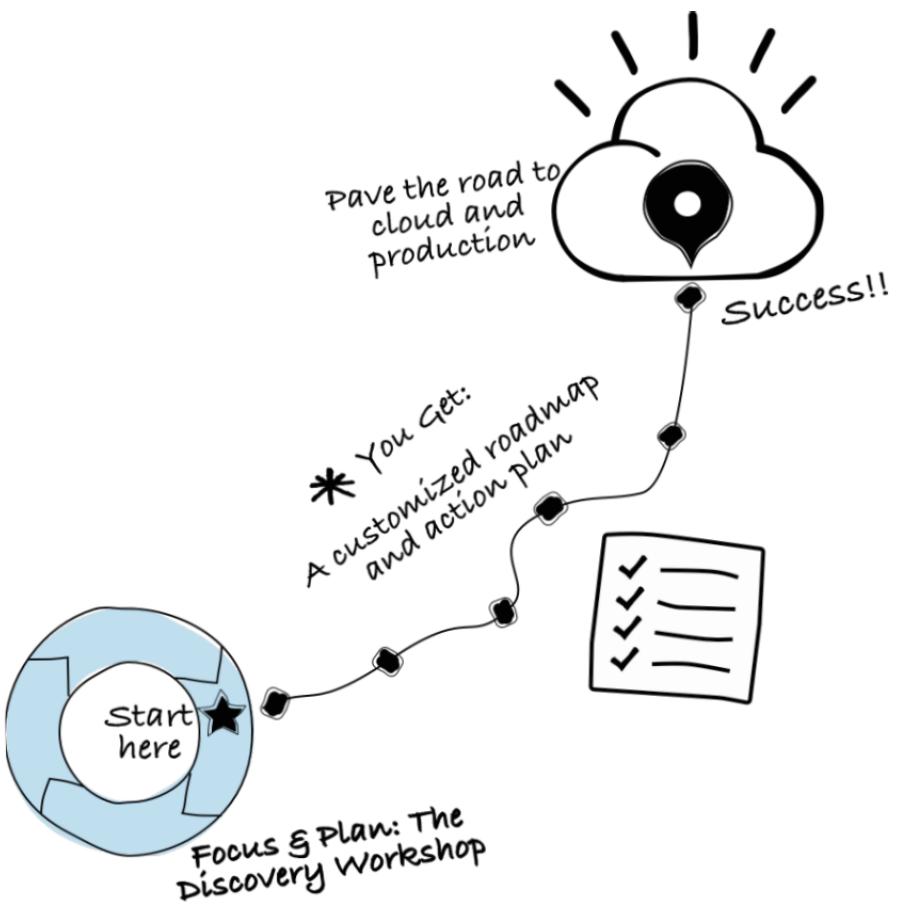
**Plan.** We develop the roadmap with key objectives and create an action plan with a prioritized list of next steps, owners, and near-term timelines.



Learn more

Check out IBM Analytics Services.

<https://www.ibm.com/analytics/services>

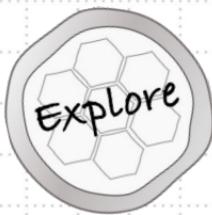


Establish a data and analytics strategy to adopt,  
transform, and deliver value—then turn it into a roadmap  
you can follow.

# Notes:

visit an IBM Garage

[ibm.com/garage](https://ibm.com/garage)



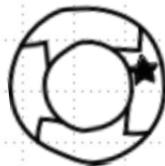
visit the IBM Data and  
Analytics architecture!!!

[ibm.com/cloud/garage/architectures/  
dataAnalyticsArchitecture](https://ibm.com/cloud/garage/architectures/dataAnalyticsArchitecture)

Learn about IBM Cloud Pak for Data!!!

[www.ibm.com/products/cloud-pak-for-data](https://www.ibm.com/products/cloud-pak-for-data)

Ask IBM to help with  
your roadmap!



[https://www.ibm.com/analytics/  
services](https://www.ibm.com/analytics/services)

Explore Red Hat OpenShift.  
Experience the platform!!

[https://www.ibm.com/analytics/  
cloud-private-for-data](https://www.ibm.com/analytics/cloud-private-for-data)



# Notices

© Copyright International Business Machines Corporation 2018, 2021.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

**IBM Director of Licensing  
IBM Corporation  
North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
US**

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions; therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

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

## Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

# IBM DATA AND ANALYTICS STRATEGY

