L4: RFX Best Practices and Samples

—

Ecosystem Engineering

Understanding Solution and Non-Functional Requirements

Solution Requirements

- Metadata Repository, Business Glossary, Data Catalog
- Discoverability
- Data Lineage, Impact Analysis, Query, Search and Reporting
- Extensibility/User Customization, Rule Management
- Metadata/Information Model
- Data Governance & Data Management
- Analytics
- Collaboration
- Data Profiling

Non-functional Requirements

General

- Availability
- Backup and Recovery
- Data Ingestion
- Data Retention and Archive
- Data and Privacy
- Accessibility

Security Requirements

- Certifications
- Roles and Responsibilities
- External Audit
- Background Checks
- Authentication
- Platform Hosting
- Anti-virus
- Encryption
- Outsourced Cloud Services
- Operating system hardening
- Virtualization hardening
- Data backup and restoration

Prepare for Prospecting Checklist for RFXs

i. Understand the use cases and pain points
ii. Quickly determine the best way to prove IBM solution
iii. Identify what outcome the client expects to achieve to qualify how we can validate our capabilities by using Proof of Technology (POT) or Proof of Concept (POC) or Minimum Viable Product (MVP) or a combination of both

i. Based on the question description tailor the presentation to line up with the RFP's requirement
ii. Include success stories (Refer the links in next slide)
iii. provide a summary of the proposed solution and how it meets the requirements in the RFP including: Functional and non-functional requirements; Security requirements; and Service Levels.
iv. Provide an overview of the logical design showing the relevant demarcations in scope and responsibility.
v. Explain the redundancy and resiliency inherent in the proposed solution, including the extent to which the infrastructure and components.
vi. Outline the primary elements within the proposed solution, the extent to which they would be "out of the box" and the extent to which configuration or customization would be required.
vii. describe any key risks with the proposed solution and how IBM would propose to mitigate them.
viii. Responses should be in detail and mention public links to navigate through ,don't leave any blank or don't reply with any open ended like Yes or No
ix. Please describe product and service strategy for IBM Data Governance services for the next three years and how the strategy is aligned to the objectives of customers.
x. Please outline key differentiators and why customer should select the proposed solution.

Data Governance & Privacy RFXs

Keywords for search:

data breach, self service analytics, AI projects, data governance, data stewardship, data literacy, data protection, data acceptability, data lake, regulatory and compliance, data science, data fabric, data mesh, data lineage, data ops, data hub

No.	Scenario	Paint Points	Possible Past actions to solve pains	Expected Result	Use cases to refer
1	Define the enterprise data strategy and enable the easy access to data that facilitates an analytics-driven decision making	• Embed governance across the entire enterprise • Simplify the access to data for all data citizens • Enable the creation of an easy-to-use data marketplace	i. Stitching together disparate point solutions for governance ii. Leveraging a central repository for data	i. Intelligent self-serve data marketplace	Life Sciences
2	Most interested in data governance for insights initiatives, and they have the financial resources to progress ahead any opportunity	• Excessive wait times for data and heavy reliance on IT • The time their team is spending searching for, cleaning and preparing data instead of spending that time on doing analytics and AI projects	i. Merge several data sets (mainly XLS files). Not sure whether these were trusted data sets	i. Self-service access to data and tools ii. Collaborate in data science process	Provider Payer
3	Ensures that the organization conducts its business processes in compliance with laws and regulations, professional standards, international standards, and accepted business practices	• Providing the business access to data while safeguarding confidential data and ensuring regulatory compliance	i. Protects personal and confidential information to ensure privacy and compliance	i. A distributed active governance layer for all data initiatives that reduces compliance and regulatory risks by providing trust and transparency.	Financial Services Insurance

User Roles and specific tooling requirements:

Watson Knowledge Catalog within Cloud Pak for Data is designed precisely to address the needs of these personas and data-driven challenges

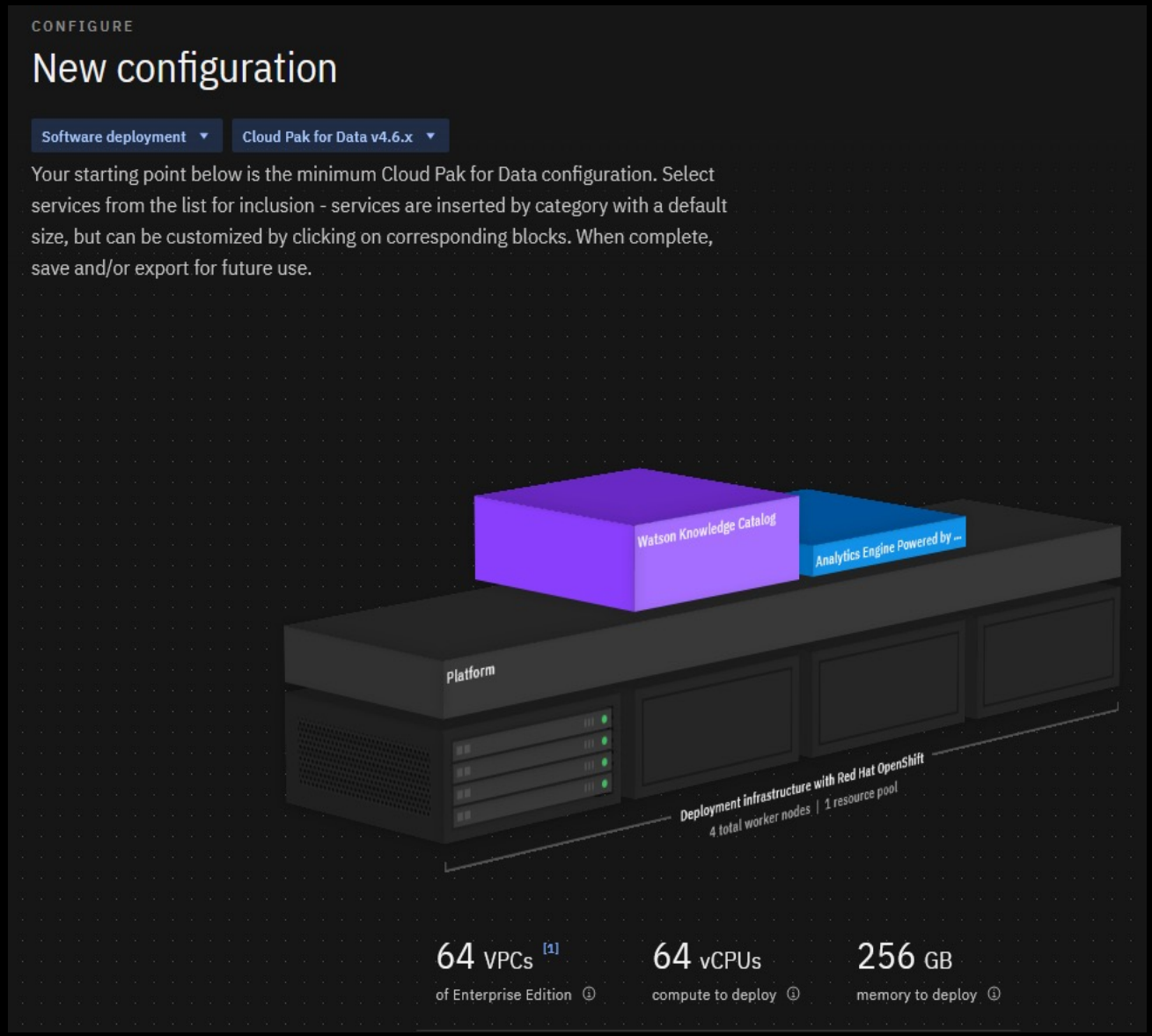
No.	Role	Sample Use case	Acceptance Criteria
1	Data Steward	As a Data Steward, I want to be able to ensure the correct trade date is queried when I receive a request from another business unit.	A data catalog solution that has straightforward query functionality. Data elements are easy to identify, and the associated source/API information is readily available for cases where the data needs to be exposed to another application.
2	AML Officer	As an AML Officer, I want to run a lineage report to receive a back-dated statement.	A tool that provides clear data lineage visualizations to assist with understanding of statements and instill confidence in the accuracy of reports.
3	Security and Privacy Analyst	As a Security and Privacy Analyst, I want to be able to respond to regulatory requests and be confident in the quality of the data.	A data governance solution that allows for routine data quality checks. There is consistency on how data elements are defined, and tables are structured. Policies on data redaction / obfuscation are clearly defined and consistent.
4	Business Analyst/Data Scientist	As a Business Analyst, I want to be able to understand the meaning behind data without tracking down an SME or looking at the code.	A solution with a business-friendly User Interface. Resources at the organization with varying levels of technical expertise can use the tool and find the information they need with ease. Business requirements documentation is readily available, and visualizations are provided to enhance understanding of data without the need for technical analysis.

Sizing

Configure services and deployment infrastructure via an interactive 3D visualization to produce high-level license and capacity estimates.

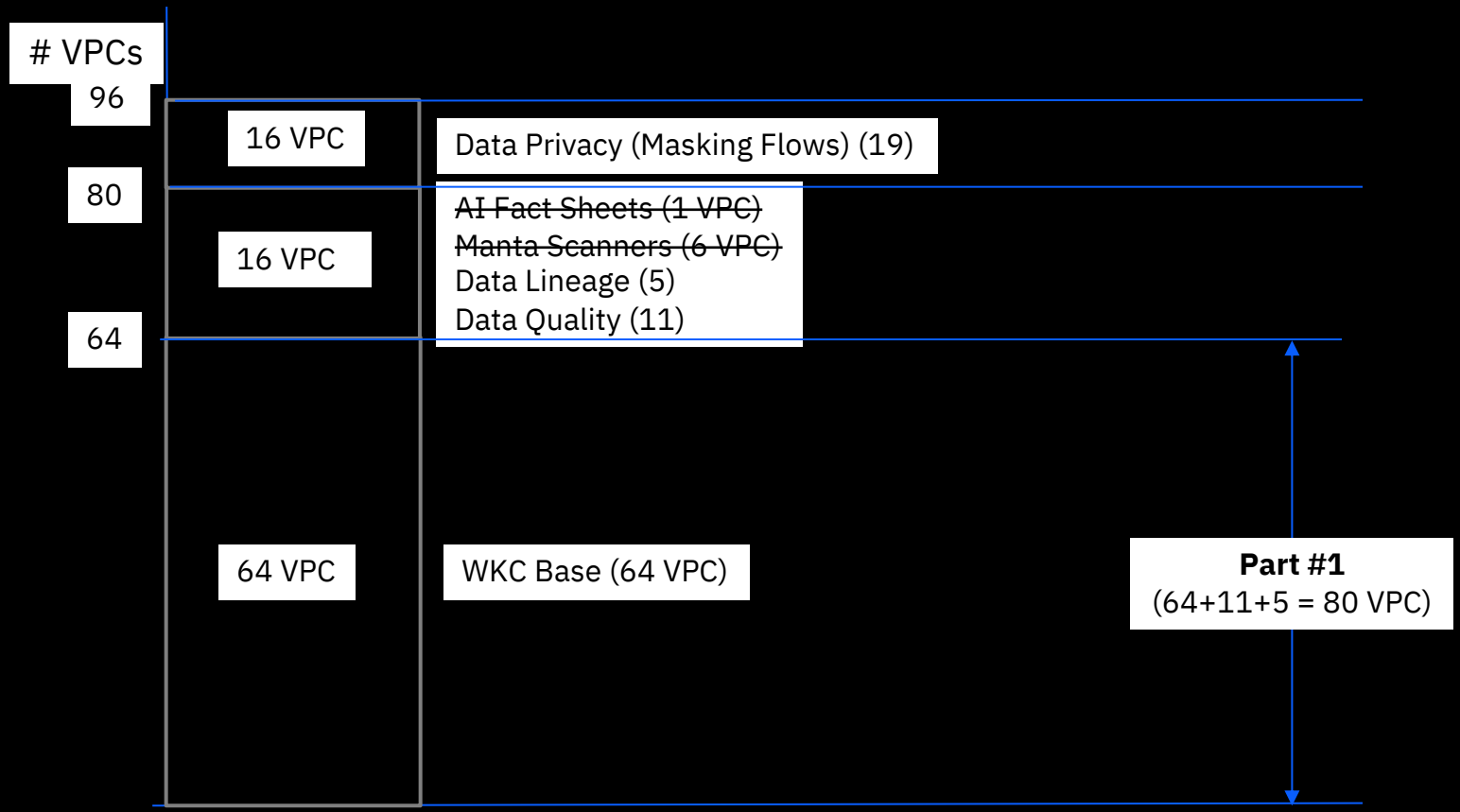
[IBM Sales Configurator]

NOTE: If you can't access the Sales Configurator, request access at the link above and an admin will grant it within 24 hours.



Watson Knowledge Catalog

Smaller clients are driven by one or more features



WKC optional Installations

Reduce VPC footprint by installing only what customer need

- Any combination on deployment rounded up to the nearest 64 VPCs (based on node deployment)
- All optional install components require WKC base as a prerequisite for deployment
- Deploy a Data Quality package by selecting WKC base + the data quality option
- Deploy the data privacy by selecting WKC Base + Masking flow

	Small	Medium	Large
WKC base	64	86	121
Optional Installations :			
Data Quality	11	17	27
Business data lineage	5	7	8
MANTA Scanners	6	8	10
AI Factsheets	1	2	4
Masking Flow	19	38	76

Sizing assumptions & BOQ

- Single pod per service
- Requires Analytics Engine Powered by Apache Spark sized as at least "Installation only (Small)".
- Data Assets: 10,000 tables, 30 columns per table, 300,000 columns across all tables, and 1,000,000 rows per table.
- Governance artifacts: 30 categories, 500 terms, 0 Reference Datasets, 20 data protection rules, 200 data classes (including out of the box), and 10 classifications.
- Auto-Discovery, QuickScan and Data Quality: 50 data classes per DQ project, 1,000 columns analyzed per day, 5,000 rows per table or sampling size, 100 DQ rules executed per day, default OOTB Data Quality dimensions only, and default Data Frequency distribution only.
- Metadata Import: 2,000 columns imported per day.
- Lineage: 1,000 assets used in lineage, and 10 interactions with lineage per day.
- 100 concurrent read-only users (browse, search, view).

- Business Data Lineage: +9 VPCs, +5 vCPUs, +20 GB memory
  - Data assets: 10,000 tables, 30 columns per table, 300,000 columns across all tables, and 1,000,000 rows per table.
  - Governance artifacts: 30 categories, 500 terms, 0 reference datasets, 20 data protection rules, 200 data classes (including out of the box), and 10 classifications.
- MANTA Scanners: +9 VPCs, +6 vCPUs, +24 GB memory
  Run lineage analysis and import that involves up to 4 data sources and 24 MB of scripts (10,000 assets)
- NEW Data Quality: +18 VPCs, +11 vCPUs, +34 GB memory
  150 Data Quality rules executed per day

BOQ

Description	Qty	Metric
IBM Cloud Pak for Data Enterprise Edition per Virtual Processor Core License + SW Subscription and Support 12 Months	64	VPC
IBM Cloud Pak for Data Enterprise Edition Non-Production per Virtual Processor Core License + SW Subscription and Support 12 Months	48	VPC
Manta Production (1 RU = 30000 scripts)	1	RU
Manta Non-Production (1 RU = 30000 scripts)	1	RU

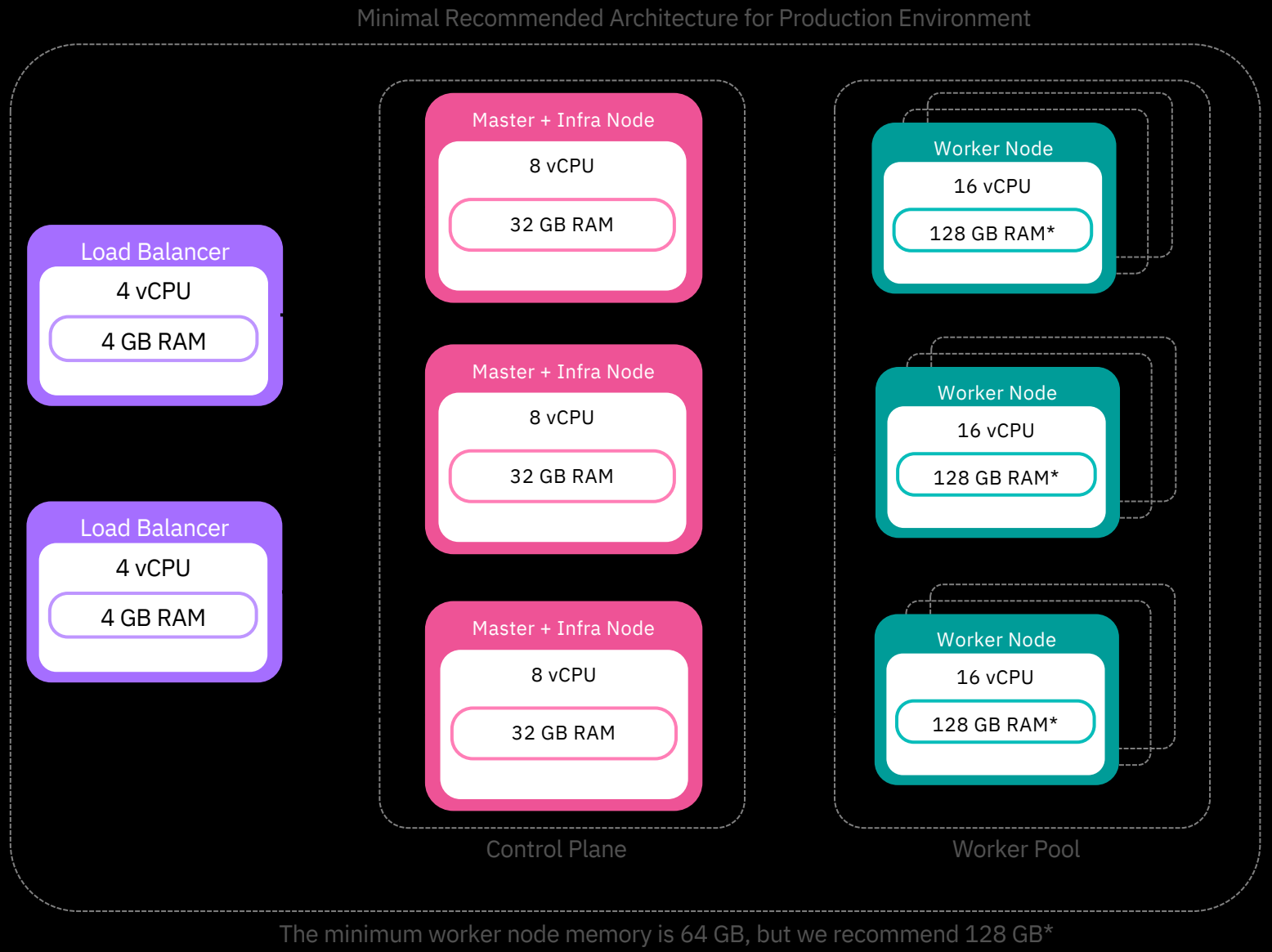
- Engage with technical sales representative to size and price for Cloud Pak for Data. For estimates, use the T-shirt sizing available in the IBM Sales Configurator.

Recommended Deployment Architecture

Production Environment

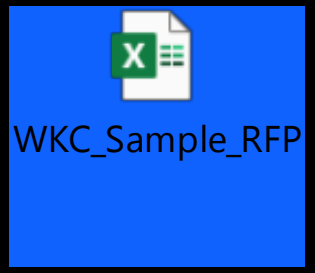
Production Environments focus on fault-tolerance, resiliency, and high availability - this is the recommended architecture for a production environment

Documentation



Sample RFP for Data Governance

Please carefully read the instructions before utilising this asset to complete a client RFP. It is strongly recommended that you alter the wording of your answers to provide a uniquely tailored answer for your client



Gartner competitive summary

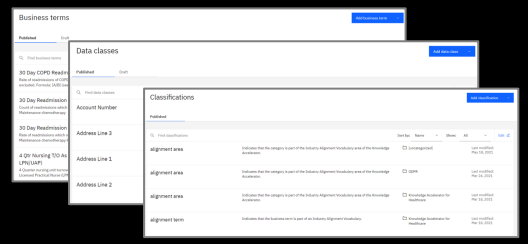
		Gartner Peer Insights				Gartner Metadata Management				Gartner Data Quality						
Vendors	**Platform?**	**Overall Customer**	**Product Capabilities**	**Service & Support**	**Integration / Deployment**	**Data Governance**	**Risk And Compliance**	**Data Analysis**	**Data Value**	**AI and ML**	**Analytics and BI**	**D & A Governance**	**Data Integration**	**Data Migration**	**Ops/Trans Data Quality**	**Total (excludes DQ)**
IBM	Yes	4.4	4.5	4.3	4.3	3.39	3.51	3.63	3.44	4.33	4.29	4.18	4.28	4.33	4.31	**3.93**
Informatica	Yes	4.2	4.3	4.3	3.9	3.64	3.85	3.75	3.64	4.44	4.29	4.15	4.4	4.39	4.37	**3.94**
SAP	Yes	4.2	4.1	4.2	4.1	3.34	3.4	3.46	3.41	4.27	4.13	4.03	4.24	4.25	4.27	**3.77**
Collibra	No	4.2	4.4	4.2	4.1	3.74	3.68	3.79	3.84	Unranked in 2020 Data Quality analysis, has since acquired OwlDQ and have gained significant capabilities in DQ space						**3.99**
Alex Solutions	No	4.8	4.7	4.6	4.5	3.74	3.81	3.8	3.81	Has pre-built data quality capabilities, unranked by Gartner						**4.2**
Alation	No	4.4	4.4	4.4	4.4	3.73	3.72	3.72	3.64	Has data quality integrated in profiling, unranked by Gartner						**4.05**
Ataccama	Yes	4.5	4.4	4.5	4.3	Has catalog, major RFP Competitor unranked in 2020 by Gartner				4.15	4.01	3.81	4.13	4.11	4.10	**NA**
Erwin	No	4.4	4.6	4.6	4.5	3.70	3.65	3.60	3.65	Has data quality as part of Erwin Data Modeler, unranked by Gartner						**4.08**
Data.world	No	4.7	4.6	4.9	4.6	3.79	3.82	3.66	3.76	Has data quality capabilities, unranked by Gartner						**4.22**

AUTO GOVERNANCE IN 3 SIMPLE STEPS

Step 1

Glossary

Gates

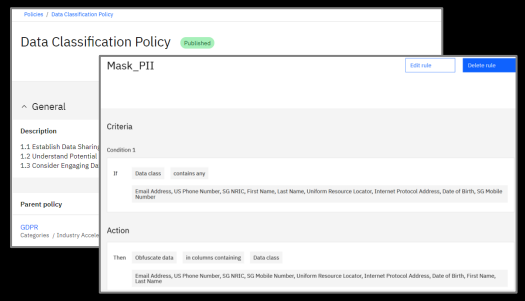


- ✓ BUSINESS TERMS
- ✓ CLASSIFICATIONS
- ✓ DATA CLASSES
- ✓ REFERENCE DATA

Step 2

Policies & Rules

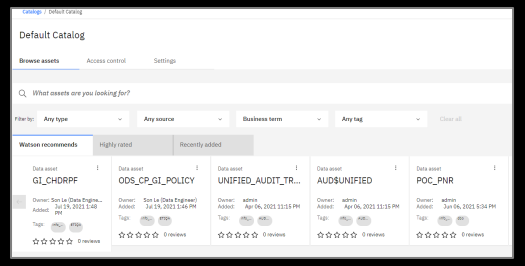
Gates



- ✓ POLICIES
- ✓ RULES

Step 3

Data Sources



- ✓ CATALOG

WATSON KNOWLEDGE CATALOG

ACCELERATORS & AUTOMATION

DATA FABRIC

With invisible omnipresent governance

DATA CONSUMERS

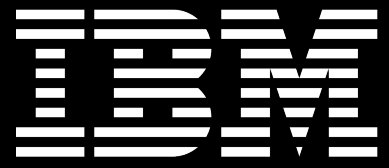
Reference Architecture on hyperscalers

Deploying IBM Cloud Pak for Data on Red Hat OpenShift Service on AWS

https://aws.amazon.com/blogs/architecture/deploying-ibm-cloud-pak-for-data-on-red-hat-openshift-service-on-aws/

Deploying IBM Cloud Pak for Data on Azure

https://learn.microsoft.com/en-us/samples/azure/azure-quickstart-templates/ibm-cloud-pak-for-data/


IBM