Feature Mapping, Component Mapping and Creating Effective Architectures

—

Ecosystem Engineering

Outline

- Understanding functional and non-functional requirements
- Scoping the POC
- Building architectures for scoped features
- Building architectures for overall requirements

Understanding Solution and Non-Functional Requirements

Solution Requirements

- Data Ingestion
- Discoverability
- Business Glossary
- Metadata Management
- Data Lineage
- Data Governance
- Analytics
- Collaboration
- Data Profiling
- ML and AI Capabilities
- Quality Management
- Master Data Management
- Installation and Configuration
- Support and Training

Note: For compliance related questions, refer https://ibm.seismic.com/Link/Content/DCJWdccm2mb9c8mGj24XBp72jd4j

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

Templates for Building Architectures

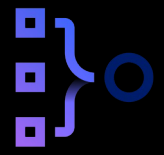
Icons

Manta

Data Virtualization

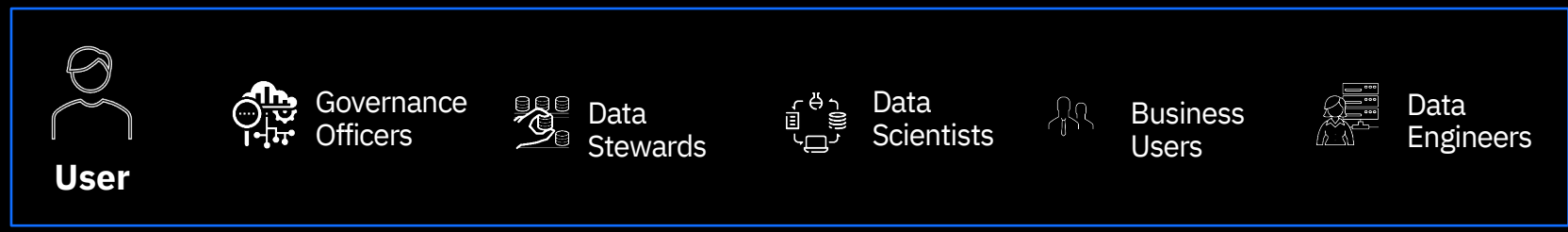
Watson Knowledge Catalog

Match 360



Watson Studio

DataStage



Feature blocks

Systems of Record

IOT

Systems of Insights

Cloud

Hadoop

Social Media

Unstructured

Other External

Industry Data & Applications

Business Glossary | Data Privacy & Protection | Data Quality Management | Metadata Enrichment | Data Catalog | Workflow Management | Data Lineage

Automated Curation Services

- Auto Discover Data
- Auto Classify Data
- Auto Detect Sensitive Data
- Auto Analyze Data Quality
- Auto Assign Business Terms

Business Glossary

- Categories
- Classifications
- Data classes
- Business terms
- Rules & Policies

Self-Service Interaction

- Search & Find Relevant Data
- Tagging, Annotations, Comments
- Self-Service Data Access
- Self-Services Data Preparation
- Projects & Collaboration

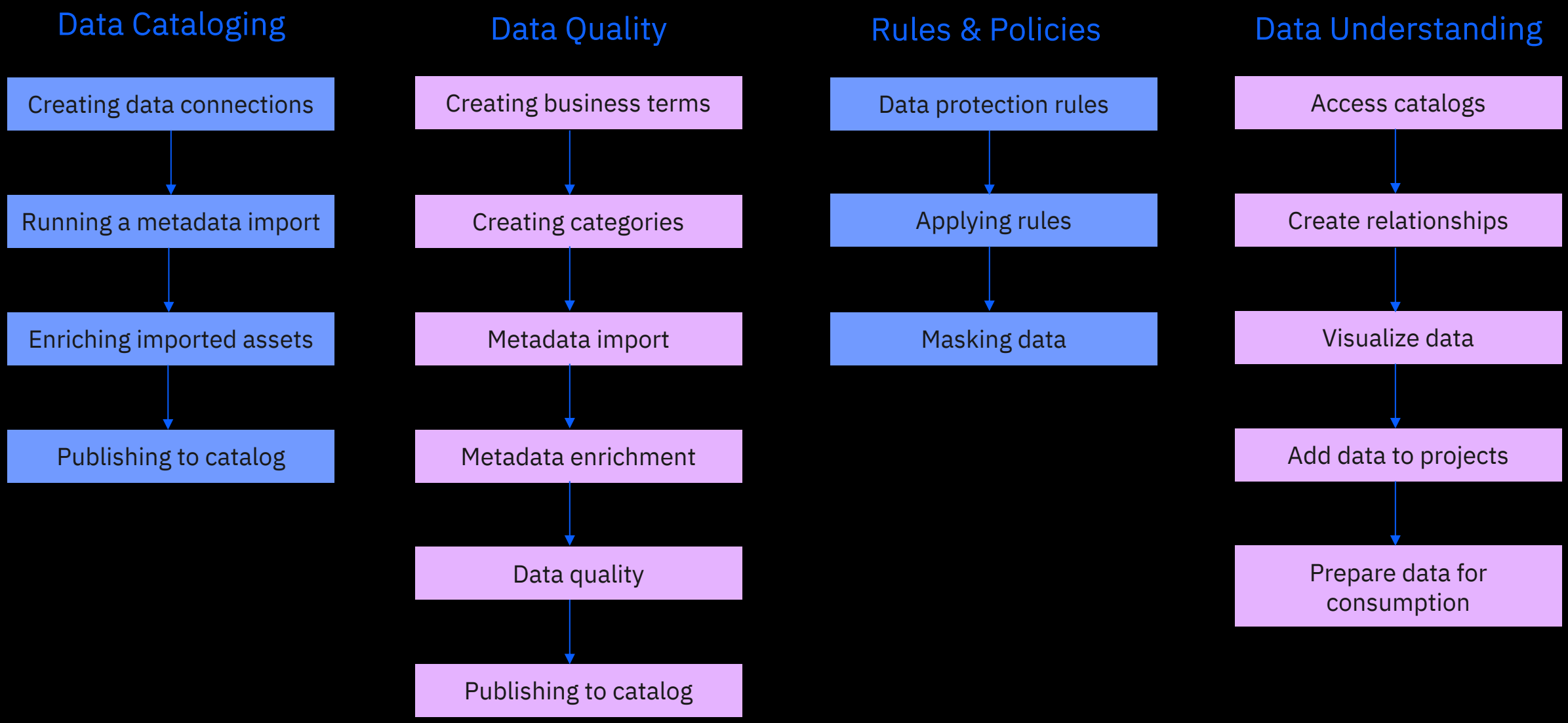
Data Lineage

- Advance Data Privacy
- Classifications
- Data classes
- Business terms
- Rules & Policies

Integrated Data Management Capabilities

Data Integration & Replication | Entity Management | Reference Data Management | Industry Knowledge Accelerators | Data Virtualization | Advanced Data Preparation

Data Governance Flow – IBM WKC



Sample Reference Building

Business Glossary

Understanding Basics of RFx

Note: If you want to understand sample requirements from RFx, how and why requirements are rated, refer to **BEST PRACTICES WHILE RESPONDING TO RFX** asset by Ecosystem Engineering - SI

- The next couple of slides are going to take sample requirements from RFx requests and show how an architecture can be built using them
- The requirements used in this are in MoSoCoW format and rated 0-4

  MoSoCoW – Must, Should, Could, Would

  This is to create an effective reference architecture.

  The requirements are classified as M, S, C, W as per the solution being proposed. As an architect/owner on the response, do you find the requirement a Must have, Should have, Could have or Would have to the solution.

  0 - 4 Rating

  This is to create an effective component architecture

  This is to define if the requirement is supported by IBM offerings. 0 means it is not supported, 4 means it's completely supported out of the box.

- If the RFx request is presented in a different format, it's highly recommended that we use the MoSoCoW and 0-4 ratings on the requirements anyways for the poc shortlisted requirements.

Sample Solution Requirements – Business Glossary

Consider the following requirements which are already filled and rated

Note: If you want to understand sample requirements from RFx, how and why requirements are rated, refer to **BEST PRACTICES WHILE RESPONDING TO RFX** asset by Ecosystem Engineering - SI

No.	Description	MoSoCoW	Rating
1	Provide an interactive user interface to define business and technical metadata, define criticality, assign roles and categorization, and other related supporting information	M	4
2	Data assets can be segmented into custom categories defined by users e.g., owned/not owned, data classification (Internal, Confidential, Personal Information, Highly Protected)	M	4
3	Able to create and maintain categories and hierarchies for classification, navigation, search and use	M	4
4	Able to bulk import and export glossaries using structured and semi-structured file formats (XLS, CSV, JSON)	M	4
5	Support for acronyms	S	3

Best practices to follow while scoping a poc:

- Ask your partner for a dump of minutes of meetings if you are not directly engaged in the pre-bid calls.
- Don’t just select the features which are rated high, select the features which are important and **MUST** to the solution as well.
- When a feature is not supported by IBM product, always check if it’s supported indirectly through a BP solution or OEM partnerships
- Unless it’s an important feature too the client, do not prioritize implementation of a requirement if it’s indirectly supported with IBM products yet ensure it’s reflected in the architecture.

Best Practices while scoping a POC

- Scope the selected functional requirements with the following template:

No.	Scenario	Data	Activities	Expected Result

- Careful with the features which are indirectly supported, make sure you have an approach in mind to address before selecting those requirements.
    - For example, in the above requirements sample for business glossary, the support for acronyms is not directly available with IBM Governance features but can indirectly be executed through tags/related terms etc.
- Define data, activities and expected results even for the requirements that are important but not supported with IBM products if it’s a requirement important to the client so that it can be used in the reference architecture

Scoping a Governance PoC

Here's a template that you could use to define your Data Governance PoCs ☺

No.	Scenario	Data	Activities	Expected Result
1	Provide an interactive user interface to define business and technical metadata, define criticality, assign roles and categorization, and other related supporting information	Prepare two table structures – i. A customer data that consists of customer id, customer name, dob, address, email, social security number ii. A transaction data consists of transaction ID, transaction account number, transaction amount, transaction date time	i. Create business and technical metadata using the interface provided for both datasets above. ii. Assign data consumers and data producers for each dataset and assign criticality for each data element. iii. Assign categorization of data elements based on pre-build data categorization iv. Define new data categorization v. Assign categorization of data elements based on custom data categorization	i. Easiness to use the user interface to perform all activities using the tool ii. Have any guidance or pre-build functionalities to support the execution of the activities above. iii. Show how business and technical data is created OTB
2	Data assets can be segmented into custom categories defined by users e.g., owned/not owned, data classification (Internal, Confidential, Personal Information, Highly Protected)	Use the above created tables itself #customer_data #transaction_data	i. Create categories and sub-categories as needed ii. Create Internal, Confidential, PI, Highly Protected classification on WKC and add them to the category	i. Easiness to use OTB features to define categories and classifications ii. Support for hierarchical model while defining categories
3	Able to create and maintain categories and hierarchies for classification, navigation, search and use	Use the above created tables #customer_data #transaction_data	i. Create a category person ii. Create sub-categories within person employee, privacy	i. Support for hierarchical definitions of categories and sub-categories ii. Cover the search and navigation while showing semantic search after all business data is attached to data assets
4	Able to bulk import and export glossaries using structured and semi-structured file formats (XLS, CSV, JSON)	Prepare three csv files – i. A file that contains categories ii. A file that contains business terms iii. A file that contains data classes	i. Create sample categories in the csv file ii. Create sample business terms in the csv file iii. Create sample data classes in the csv file	i. From GUI import the csv files i, ii, iii and show the imported categories, business terms and data classes
5	Support for acronyms	Prepare two columns – SSN, CID	i. Define a business term for SSN and add it as related term to Social Security term ii. Add CID as tag to the data asset in the catalog	i. From platform search, search for SSN and all the columns with social security number will be displayed. ii. Explain the semantic search functionality

Proof Of Concept – Feature Mapping

Understand the **activities** that are to be performed and assign a suitable feature for the activity:

i. Create business and technical metadata using the interface provided for both datasets above – **Categories, Classifications, Business Terms**
ii. Assign data consumers and data producers for each dataset and assign criticality for each data element - **RBAC**
iii. Assign categorization of data elements based on pre-build data categorization - **Categories**
iv. Define new data categorization
v. Assign categorization of data elements based on custom data categorization

i. Create categories and sub-categories as needed - **Categories**
ii. Create Internal, Confidential, PI, Highly Protected classification on WKC and add them to the category - **Classifications**

i. Create a category person - **Categories**
ii. Create sub-categories within person employee, privacy - **Categories**

i. Create sample business terms in the csv file – **Business Terms**
ii. Create sample data classes in the csv file – **Data Classes**

i. Define a business term for SSN and add it as related term to Social Security term – **Business Terms**
ii. Add CID as tag to the data asset in the catalog - **Tags**

Scoping a Governance PoC

Here's a template that you could use to define your Data Governance PoCs ☺

No.	Scenario	Data	Activities	Expected Result
1	Able to allow user to mask/encrypt data within the workflow to hide protected/sensitive information.	Prepare two table structures. 1. A customer data consists of customer ID, customer name, customer DOB, customer address, contact information 2. A transaction data consists of transaction ID, transaction account number, transaction amount, transaction date time	i. Create data protection rules - All emails must be protected; All credit card information must be protected. ii. Define masking options, one with Advance Data Protection and one without ADP.	i. Show the types of rules available ii. Show the types of data masking options available iii. Show the advance data protection options available
2	Can support data governance rules just for documentation purposes	Use the above created tables itself #customer_data #transaction_data	i. Create data governance rules - All Emails must be protected, All credit card information must be protected, All phone numbers must be protected ii. Add them to a category	i. Support for documentation rules and easiness to create them
3	Can support location-based rules for data soverginity purposes	Use the above created tables #customer_data #transaction_data	i. Create data soverginity rule - All data from US region must be masked ii. Define the rule if sourrce soverginity contains any USA, then mask columns	i. Support for data location and region related rules
4	Can create policies by grouping several rules	Use the above created tables #customer_data #transaction_data	i. Create a policy - All sensitive data must be protected ii. Add the rules created above	i. Support for grouping rules and enforcing a policy on data

Proof Of Concept – Feature Mapping

Understand the **activities** that are to be performed and assign a suitable feature for the activity:

i. Create data protection rules - All emails must be protected; All credit card information must be protected. - **Rules**
ii. Define masking options, one with Advance Data Protection and one without ADP. - **Rules**

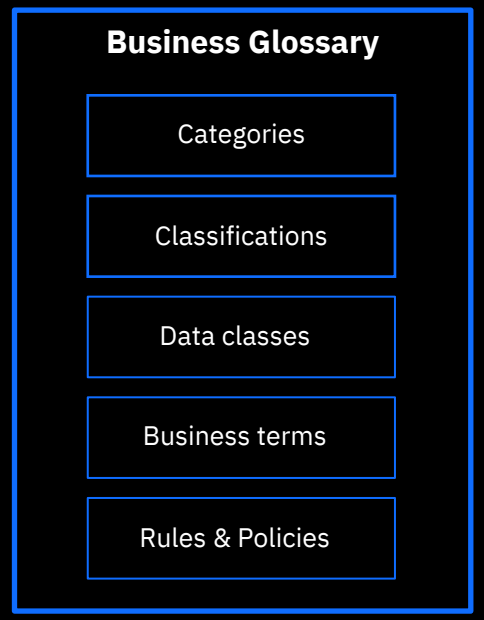
i. Create data governance rules - All Emails must be protected, All credit card information must be protected, All phone numbers must be protected - **Rules**
ii. Add them to a category - **Rules**

i. Create data soverginity rule - All data from US region must be masked - **Rules**
ii. Define the rule if sourrce soverginity contains any USA, then mask columns - **Rules**

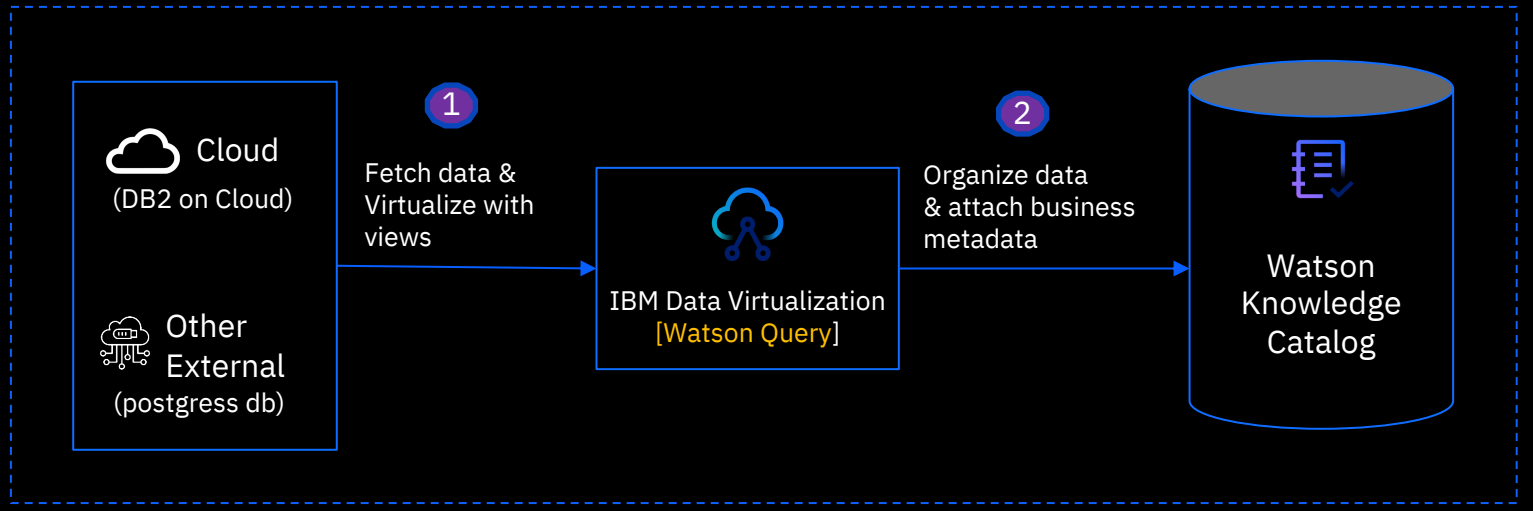
i. Create a policy - All sensitive data must be protected - **Policies**
ii. Add the rules created above

Reference & Component Architecture – Business Glossary

Reference Architecture



Component Architecture



Sample Reference Building

Metadata Management

Sample Solution Requirements – Metadata Management

Consider the following requirements which are already filled and rated

Note: If you want to understand sample requirements from RFx, how and why requirements are rated, refer to **BEST PRACTICES WHILE RESPONDING TO RFX** asset by Ecosystem Engineering - SI

No.	Description	MoSoCoW	Rating
1	Provide an interactive user interface to define business and technical metadata, define criticality, assign roles and categorization, and other related supporting information	M	4
2	Read technical objects from a data storage system to automatically catalogue the data elements for metadata creation	M	4
3	Support a hierarchical structure to catalog and categorise the new data element according to its position in the hierarchy. The system should enable users to make changes to the hierarchy, such as changing subject areas, adding subject areas, changing the assignment of the data element, etc.	M	2
4	Perform sanity check and validation on metadata, such as cleanse unused metadata, remove duplicated metadata, detect, and resolve conflicting metadata information, etc.	M	4
5	Provide user access matrix to metadata to ensure that different levels of privilege are assigned accordingly to different groups of users.	M	3

Scoping a Governance PoC – Metadata Management

Here's a template that you could use to define your Data Governance PoCs ☺

No.	Scenario	Data	Activities	Expected Result
1	Provide an interactive user interface to define business and technical metadata, define criticality, assign roles and categorization, and other related supporting information	Prepare two table structures. 1. A customer data consists of customer ID, customer name, customer DOB, customer address 2. A transaction data consists of transaction ID, transaction account number, transaction amount, transaction date time	i. Create business and technical metadata using the interface provided for both datasets above. ii. Assign data consumers and data producers for each dataset and assign criticality for each data element. iii. Assign categorization of data elements based on pre-build data categorization iv. Define new data categorization v. Assign categorization of data elements based on custom data categorization	i. Easiness to use the user interface to perform all activities using the tool ii. Have any guidance or pre-build functionalities to support the execution of the activities above.
2	Read technical objects from a data storage system to automatically catalogue the data elements for metadata creation	Prepare two CSV files with dummy data. You can use the same table structure above in #1 but add dummy values.	i. Use the tools to import the data from CSV files and automatically catalogue the data element for metadata creation.	i. Easiness to auto-catalog data for metadata creation.
3	Perform sanity check and validation on metadata, such as cleanse unused metadata, remove duplicated metadata, detect, and resolve conflicting metadata information, etc.	Use created metadata from #1, #2, #3, #4	i. Create duplicated metadata and use the tool to detect the duplication and resolve duplication.	i. Whether the tool can effectively detect duplication of metadata. ii. Whether the tool can provide useful information to help resolving the duplication in metadata.
4	Provide user access matrix to metadata to ensure that different levels of privilege are assigned accordingly to different groups of users.	Use created metadata from #1, #2, #3, #4	i. Define user roles and assign with different privileges.	i. Flexibility in assigning user matrix

Proof Of Concept – Feature Mapping

Understand the **activities** that are to be performed and assign a suitable feature for the activity:

i. Create business and technical metadata using the interface provided for both datasets above. **– Business Glossary & Metadata Import**
ii. Assign data consumers and data producers for each dataset and assign criticality for each data element. **- RBAC**
iii. Assign categorization of data elements based on pre-build data categorization **- Categories**
iv. Define new data categorization **- Categories**
v. Assign categorization of data elements based on custom data categorization **- Categories**

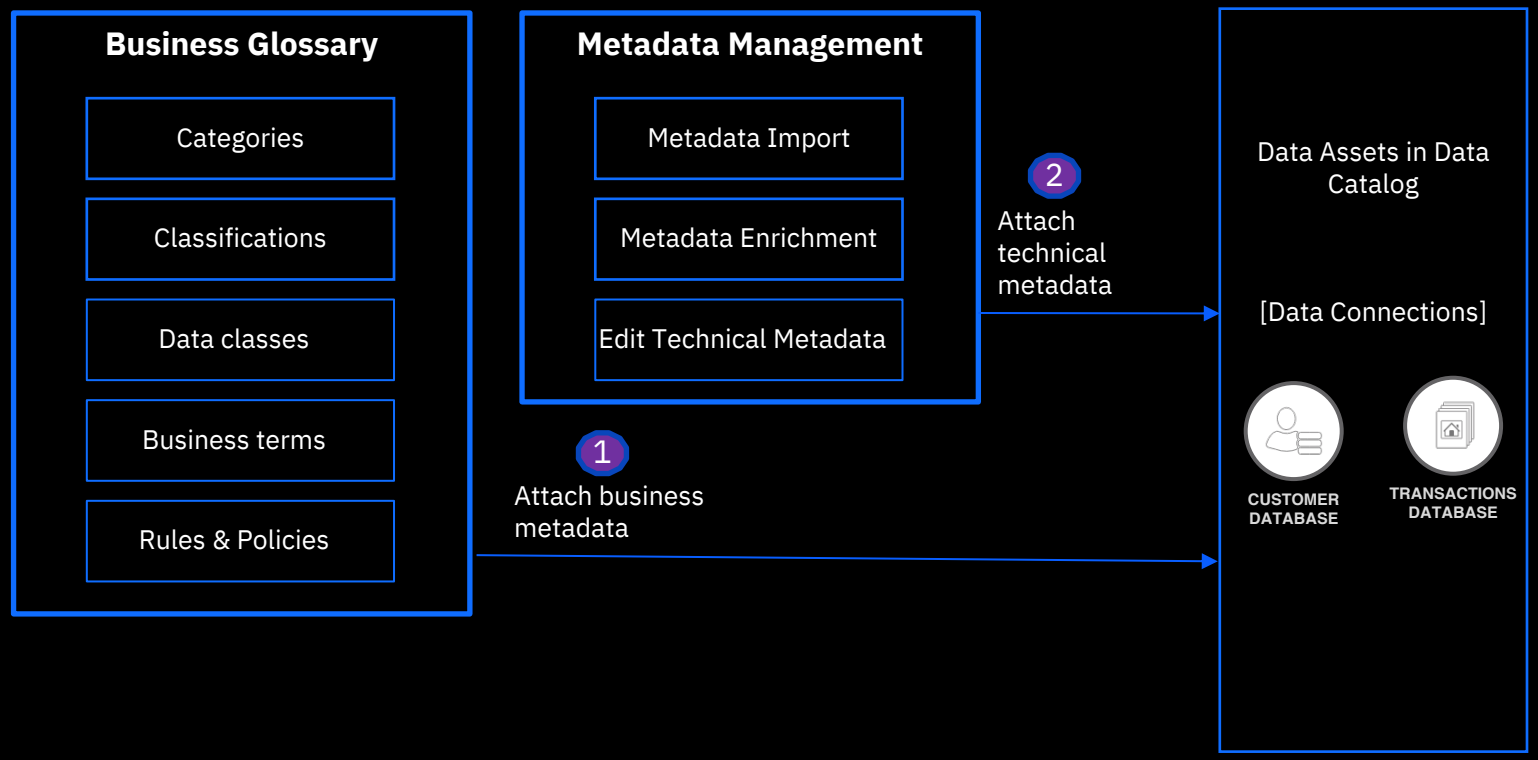
Use the tools to import the data from CSV files and automatically catalogue the data element for metadata creation. **- Metadata Import**

Create duplicated metadata and use the tool to detect the duplication and resolve duplication. **– Metadata Enrichment**

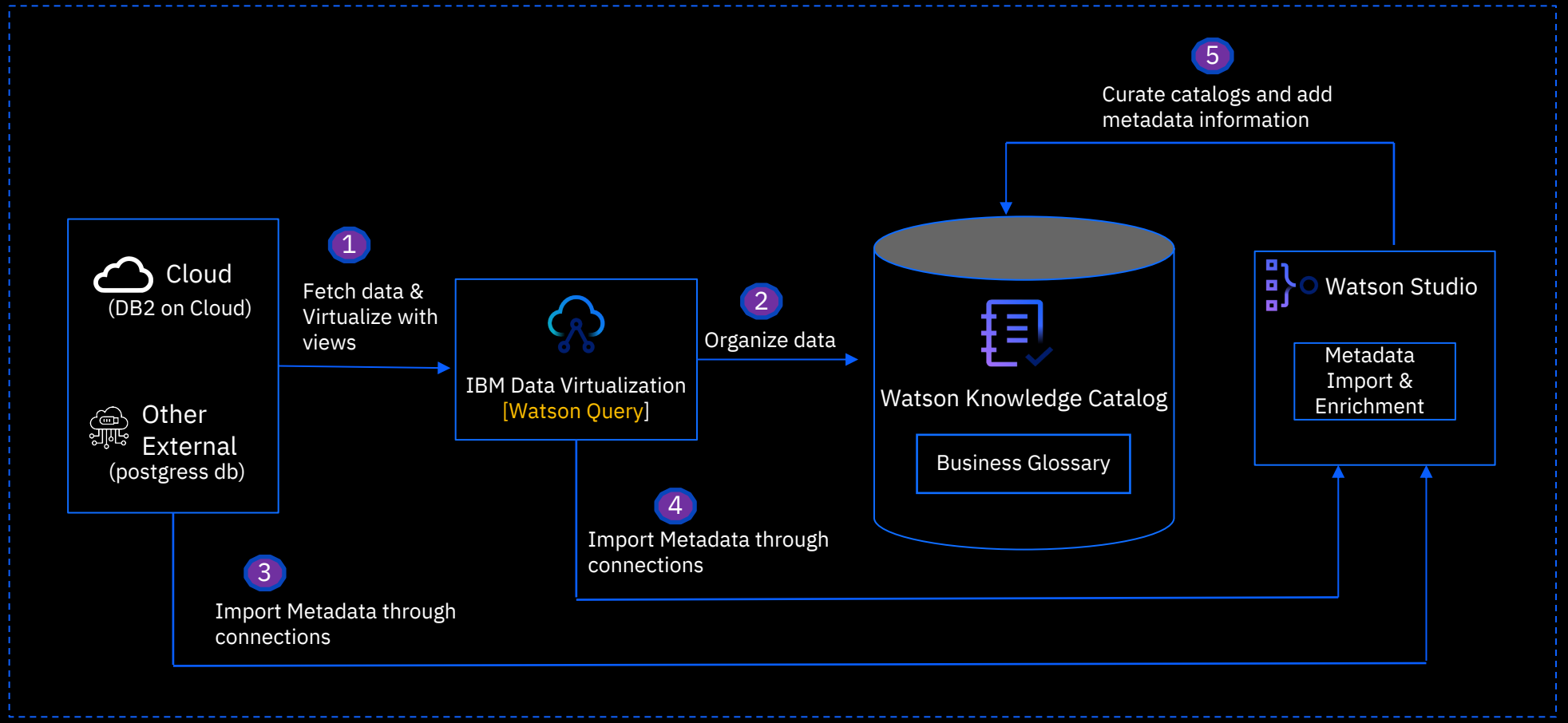
Provide user access matrix to metadata to ensure that different levels of privilege are assigned accordingly to different groups of users. **- RBAC**

Reference Architecture

Business Glossary and Metadata Management



Component Architecture
Business Glossary and Metadata Management



Sample Reference Building

Data Quality

Sample Solution Requirements – Data Quality

Consider the following requirements which are already filled and rated

Note: If you want to understand sample requirements from RFx, how and why requirements are rated, refer to **BEST PRACTICES WHILE RESPONDING TO RFX** asset by Ecosystem Engineering - SI

No.	Description	MoSoCoW	Rating
1	Provide an interactive user interface to define and implement customized data quality rules	M	4
2	Integrate with metadata systems to apply the same data quality standards across the entire data lineage, as well as ensure alignment within the same categorization. For example, the data quality rule applied on one data category is also applied to all its sub-categories.	M	4
3	Automated data profiling on new data element as well as existing data elements from the data storage systems to detect new and historical data quality issues.	M	2
4	Properly categorise data quality issues and provide a library of common data quality resolutions as rules to mitigate data quality risks.	M	4
5	Perform sanity check and validation on data quality rules, includes remove unused data quality rules, remove duplications, resolve conflicting and misaligned data quality rules, etc.	M	3

Scoping a Governance PoC – Data Quality

Here's a template that you could use to define your Data Governance PoCs ☺

No.	Scenario	Data	Activities	Expected Result
1	Provide an interactive user interface to define and implement customized data quality rules	A customer data consists of customer ID, customer name, customer DOB, customer address A transaction data consists of transaction ID, transaction account number, transaction amount, transaction date time	i. Create data quality rules using the interface provided for both datasets above.	i. Easiness to use the user interface to perform all activities using the tool ii. Have any guidance or pre-build functionalities to support the execution of the activities above.
2	Integrate with metadata systems to apply the same data quality standards across the entire data lineage, as well as ensure alignment within the same categorization. For example, the data quality rule applied on one data category is also applied to all its sub-categories.	Prepare dummy data with several data quality issues, i.e., blank data, contains incorrect format/ values	i. Choose data elements and apply the relevant data quality standards	i. Validate whether tools apply to same data quality standards within the data lineage
3	Automated data profiling on new data element as well as existing data elements from the data storage systems to detect new and historical data quality issues.	Prepare two CSV files with dummy data. You can use the same table structure above in #1 but add dummy values. Manually manipulate the dummy data so that they can have the data quality issues as we see in real business cases.	i. Use the DQ tools import the data files and run data profiling to see if all the data quality issues have been identified by the profiler.	i. Completeness and accuracy in data profiling and identifying all the data quality issues from the sample dataset. ii. Easiness of running data profiling.
4	Perform sanity check and validation on data quality rules, includes remove unused data quality rules, remove duplications, resolve conflicting and misaligned data quality rules, etc.	Use created metadata from #1, #2, #3, #4	i. Create duplicated data quality rules and use the tool to detect the duplication and resolve duplication.	i. Whether the tool can effectively detect duplication of data quality rules, especially the conflicting rules. ii. Whether the tool can provide useful information to help resolving the duplication in data quality rules.

Proof Of Concept – Feature Mapping

Understand the **activities** that are to be performed and assign a suitable feature for the activity:

i. Create data quality rules using the interface provided for both datasets above. **– Data Rules and Rulesets**

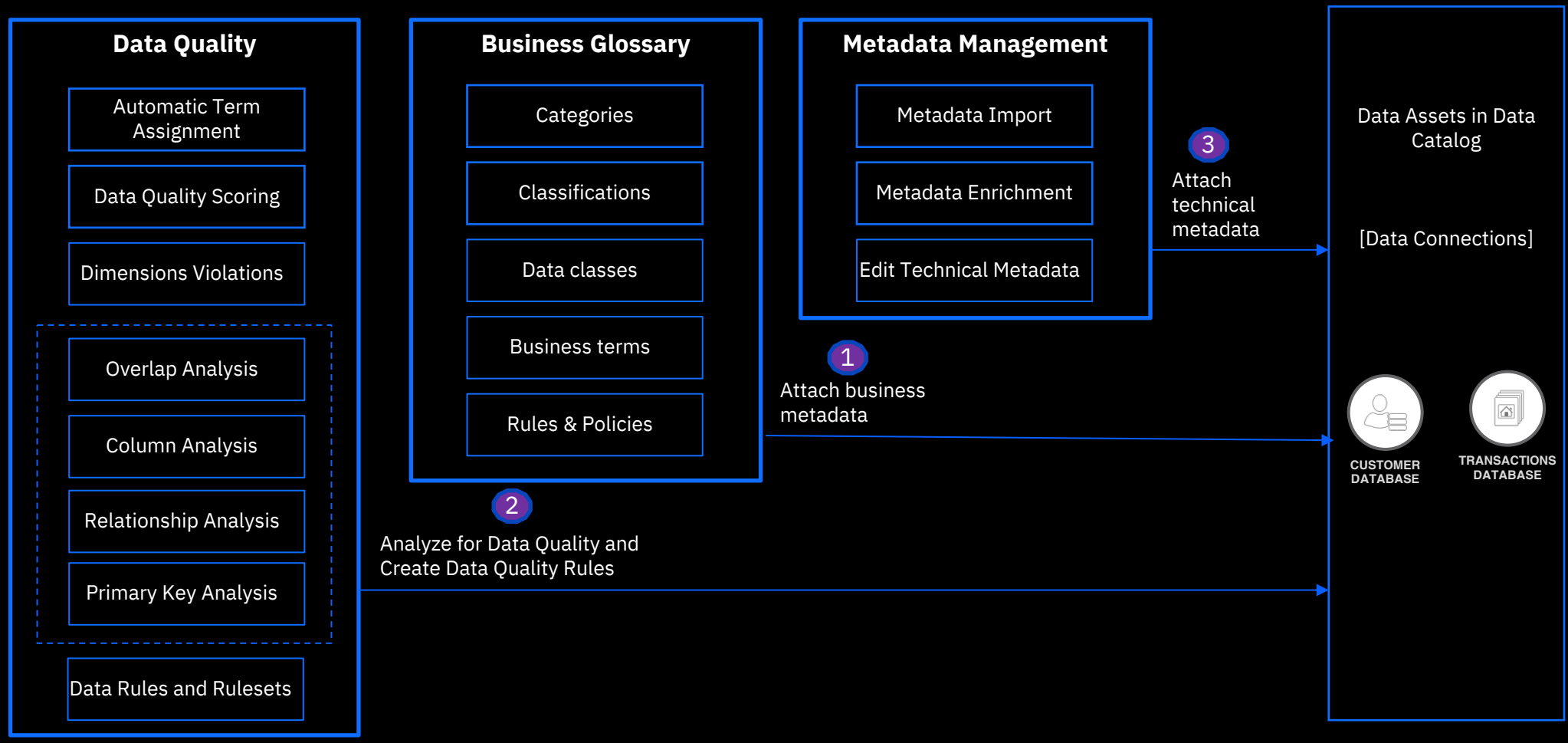
i. Choose data elements and apply the relevant data quality standards. **– Auto term assignment, data quality analysis**

i. Use the DQ tools import the data files and run data profiling to see if all the data quality issues have been identified by the profiler. **– Review data quality analysis and edit analysis results**

i. Create duplicated data quality rules and use the tool to detect the duplication and resolve duplication. **– Editing Analysis Results, Data quality rules and rulesets**

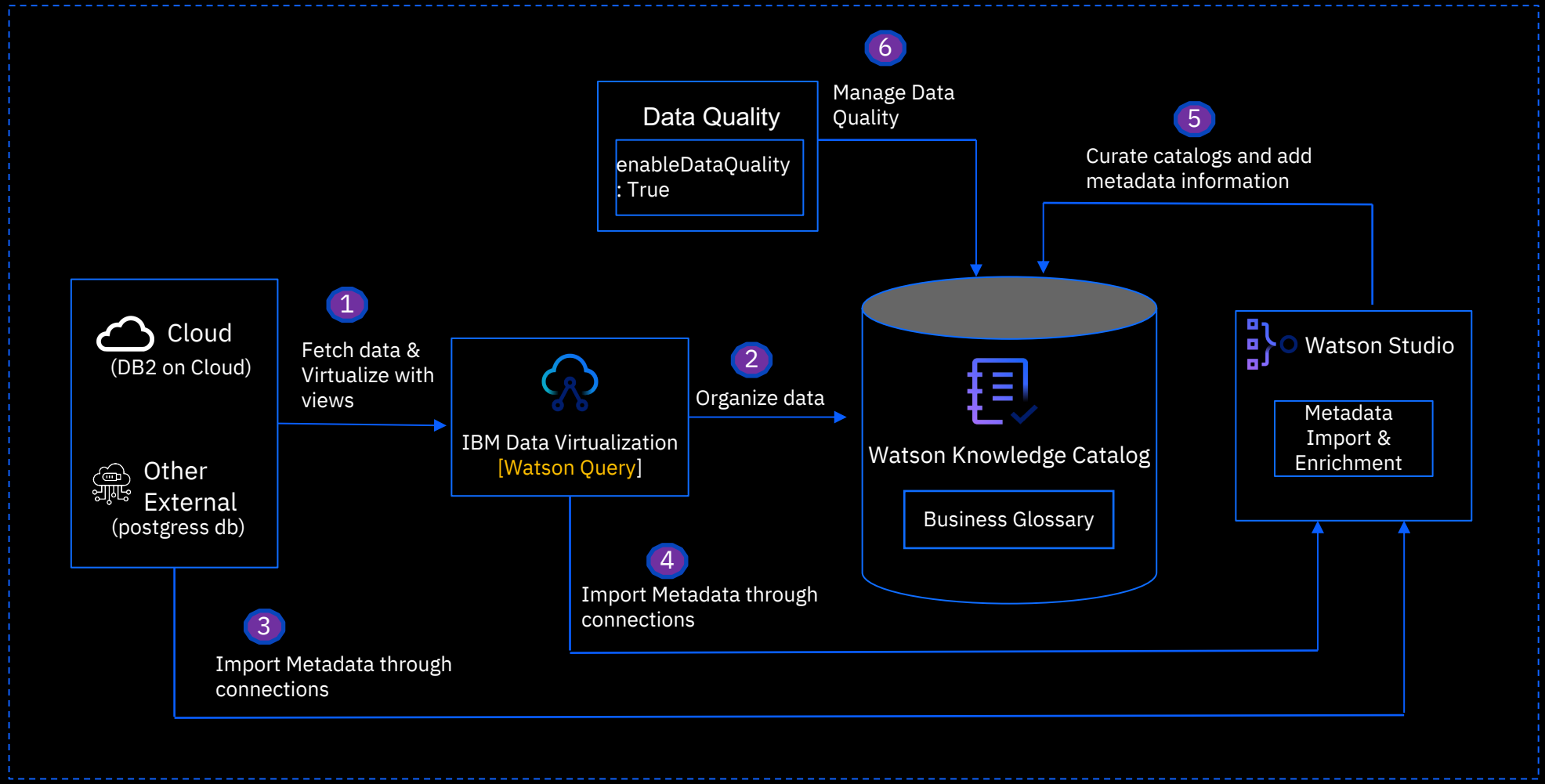
Reference Architecture

Business Glossary, Metadata Management and Data Quality



Component Architecture

Business Glossary, Metadata Management and Data Quality



Sample Reference Building

Discoverability & Governance

Sample Solution Requirements – Discoverability and Governance

Consider the following requirements which are already filled and rated

Discoverability & Governance

No.	Description	MoSoCoW	Rating
1	Able to search the catalogue using object filters as well as natural language	M	4
2	Able to search for linked or related assets	M	4
3	Able to search data dictionary, business glossary, wiki/information pages/documents using natural language, name, description, content, tags and other attributes	M	4
4	Advanced search capabilities including predictive search with iterative refinement to search results	S	3
5	Assign data ownership and stewardship to data assets	M	4
6	Able to understand created, last updated, status, and % completeness	S	3
7	Able to identify duplicate or similar data assets across the catalogue	S	3

Scoping a Governance PoC – Discoverability and Governance

Here's a template that you could use to define your Data Governance PoCs ☺

No.	Scenario	Data	Activities	Expected Result
1	Able to search the catalogue using object filters as well as natural language	1. A database with columns Customer ID, Contact Information, Phone Number, Email Address, Home Phone Number, Cell Phone Number	i. 1. Create Business Terms Contact Information, Phone Number, Email Address, Home Phone Number, Cell Phone Number ii. 2. Define relationships for business terms. For example - contact information -> type of -> phone number -> home phone number iii. 3. Search for contact information and contact informaton for mortgages	i. 1. From CP4D console, show the search results like phone number, home phone number, email addresses also getting displayed ii. 2. Search results for contact information for mortgages, show the types of filters available. iii. 3. Explain type relationships for business terms and natural language search using meaning of search
2	Able to search for linked or related assets	1. Create a data base called Property with columns property id, property value, property address, house selling price, location	i. 1. Create Business Terms Property Value, Home Price, House Selling Price, location ii. 2. Define relationships for business terms. For Property Value business term, synonym us home price and other related business terms is house selling price iii. 3. Search for property values by geo	i. 1. Show the search results for contact information which also contains related assets like phone number ii. 2. Emphasize on ranking of search results
3	Able to search data dictionary, business glossary, wiki/information pages/documents using natural language, name, description, content, tags and other attributes	Use the data assets created earlier	i. 1. For a few data assets, add metadata like tags, related attributes, synonyms ii. 2. Search for a business term on console	i. 1. Search results for All and available filters for business terms, data classes, classification, reference data ii. 2.When filtered by data, all the filters available
4	Advanced search capabilities including predictive search with iterative refinement to search results	Use the data assets created earlier	i. Search for terms created above and filter further	i. Search and filter features changing iteratively depending on the search feature ii. Search ranking and predictive terms from natural language understanding

No.	Scenario	Data	Activities	Expected Result
5	Able to understand created, last updated, status, and % completeness	Use the governed data catalog created earlier	i. Search on platform for property value	i. Show filter options available for last updated, last updated in 3 months etc. ii. Show catalog update options - last created, last updated, imported, removed etc.
6	Able to identify duplicate or similar data assets across the catalogue	Create a duplicate table and attach the relevant metadata	i. Search for duplicate table	i. Delete the duplicate table/add a relvant tag

Proof Of Concept – Feature Mapping

Understand the **activities** that are to be performed and assign a suitable feature for the activity:

i. Create Business Terms Contact Information, Phone Number, Email Address, Home Phone Number, Cell Phone Number

ii. Define relationships for business terms. For example - contact information -> type of -> phone number -> home phone number

iii. Search for contact information and contact information for mortgages **- Semantic Search**

i. Create Business Terms Property Value, Home Price, House Selling Price, location

ii. Define relationships for business terms. For Property Value business term, synonym us home price and other related business terms is house selling price

iii. Search for property values by geo **- Semantic Search**

i. For a few data assets, add metadata like tags, related attributes, synonyms

ii. Search for a business term on console **- Semantic Search**

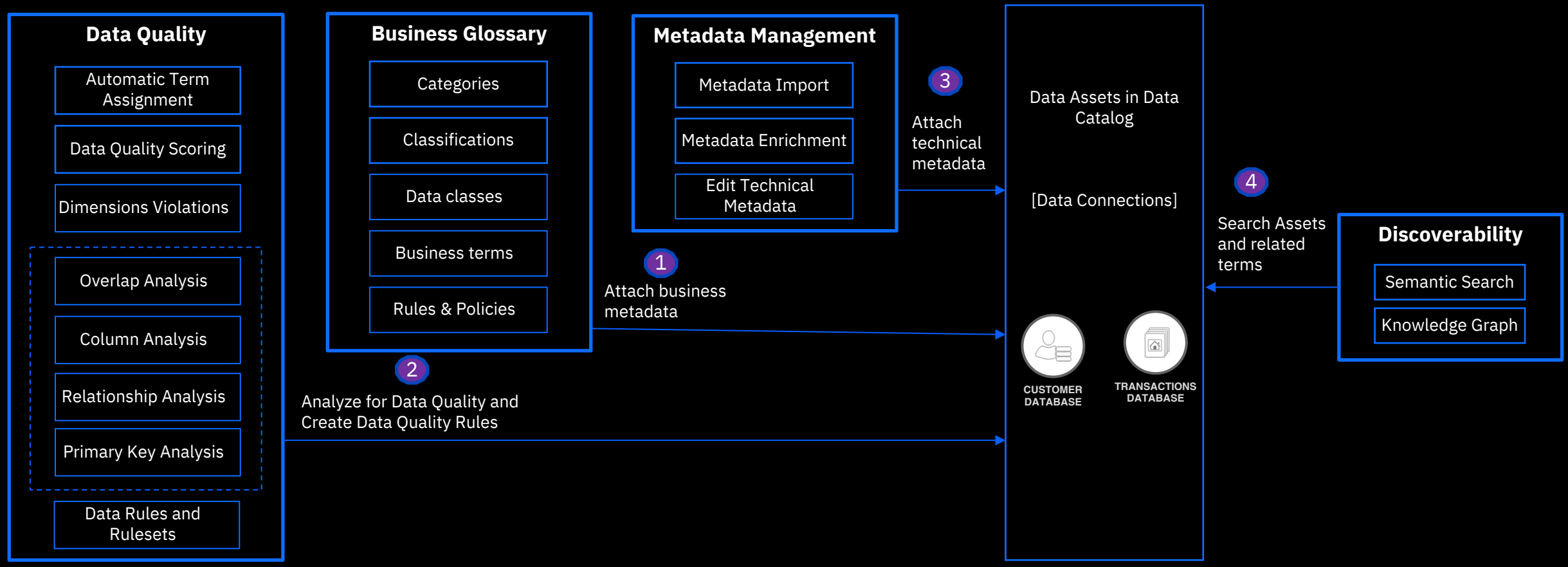
i. Search for terms created above and filter further **- Semantic Search**

i. Search on platform for property value **- Semantic Search**

i. Search for duplicate table **- Semantic Search**

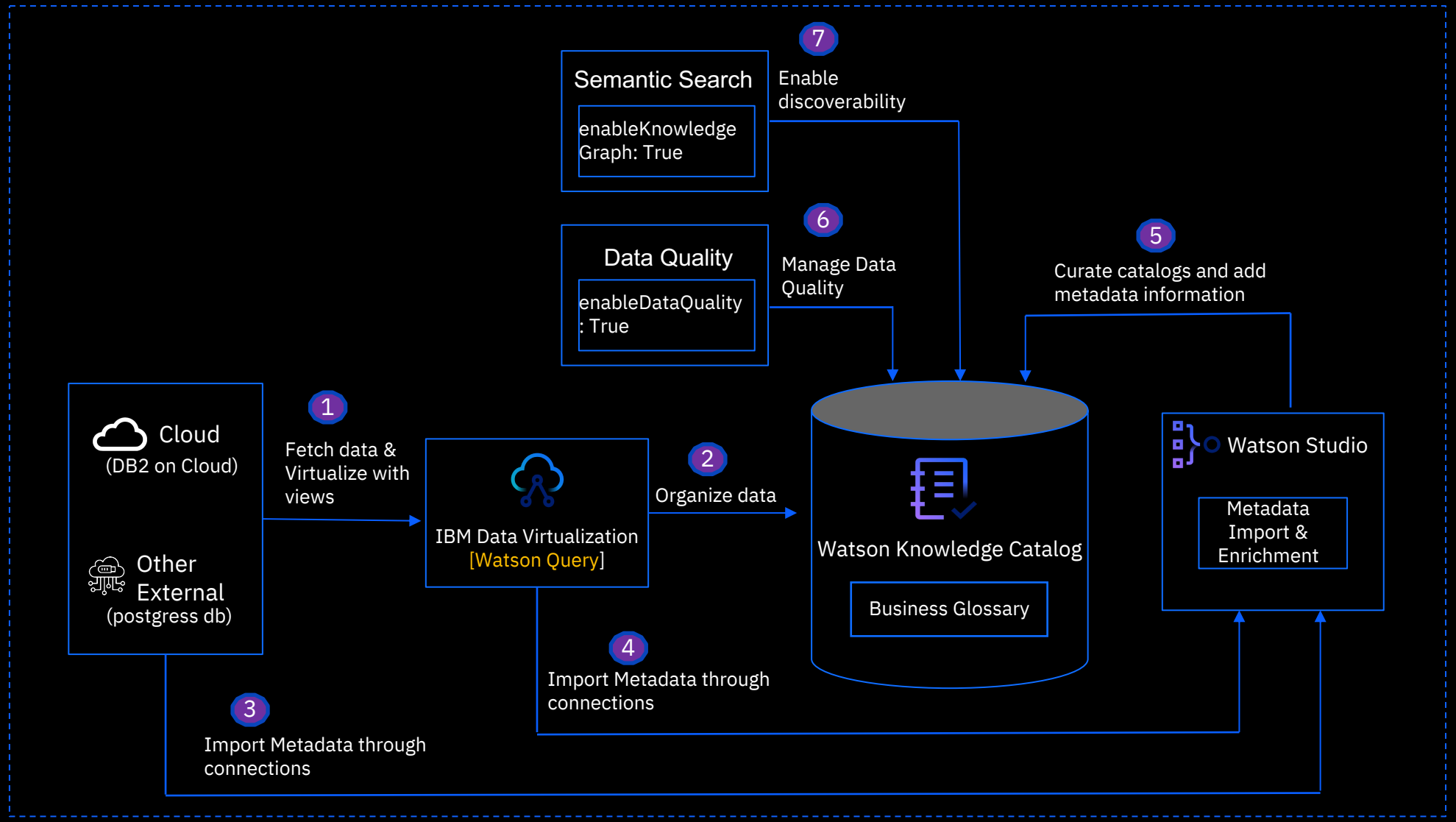
Reference Architecture

Business Glossary, Data Quality, Metadata Management and Discoverability



Component Architecture

Business Glossary, Data Quality, Metadata Management and Discoverability



Sample Reference Building

Data Lineage

Sample Solution Requirements – Data Lineage

Consider the following requirements which are already filled and rated

Data Lineage

No.	Description	MoSoCoW	Rating
1	Ability to present upstream and downstream impacts of attribute changes	M	4
2	End to end visual representations of data flows to illustrate data provenance, storage, usage and transformation	M	4
3	Able to see metadata for assets on lineage path	S	4
4	Able to automatically mark assets upstream of identified quality issues e.g., quality issues raised against a tableau dashboard	S	2
5	Able to perform analytics on lineage	M	4

Scoping a Governance PoC – Data Lineage

Here's a template that you could use to define your Data Governance PoCs ☺

No.	Scenario	Data	Activities	Expected Result
1	Ability to present upstream and downstream impacts of attribute changes			
2	End to end visual representations of data flows to illustrate data provenance, storage, usage and transformation			
3	Able to see metadata for assets on lineage path			
4	Able to automatically mark assets upstream of identified quality issues e.g., quality issues raised against a tableau dashboard			
5	Able to perform analytics on lineage			

Proof Of Concept – Feature Mapping

Understand the **activities** that are to be performed and assign a suitable feature for the activity:

i. Create. **– Data**

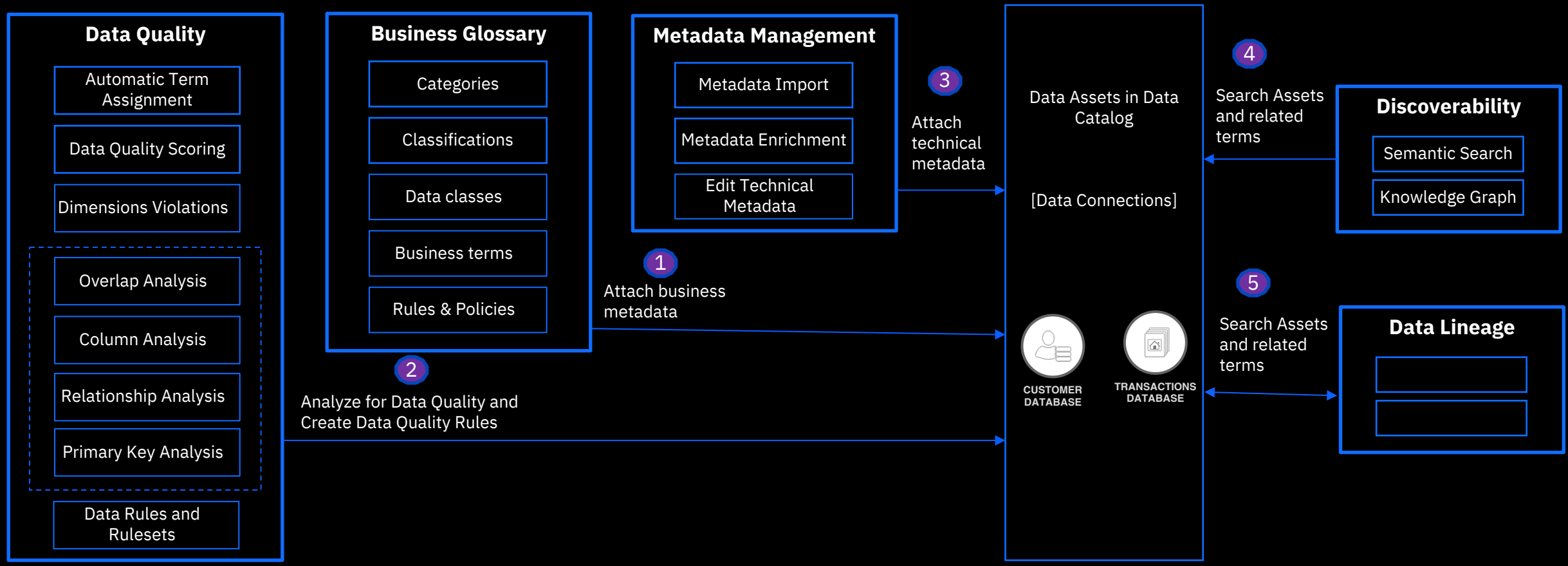
i. Create. **– Data**

i. Create. **– Data**

i. Create. **– Data**

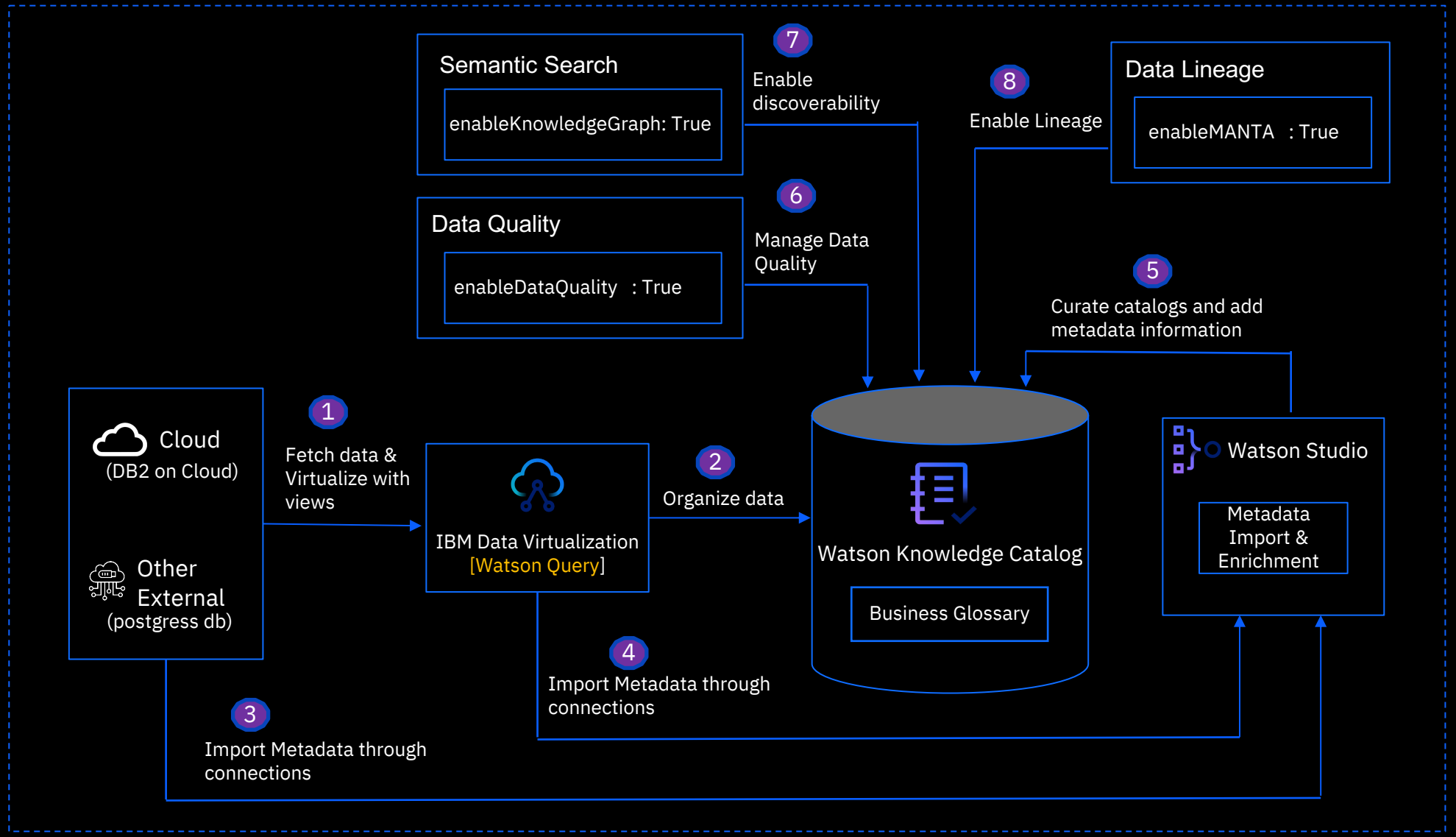
Reference Architecture

Business Glossary, Data Quality, Metadata Management, Discoverability and Data Lineage



Component Architecture

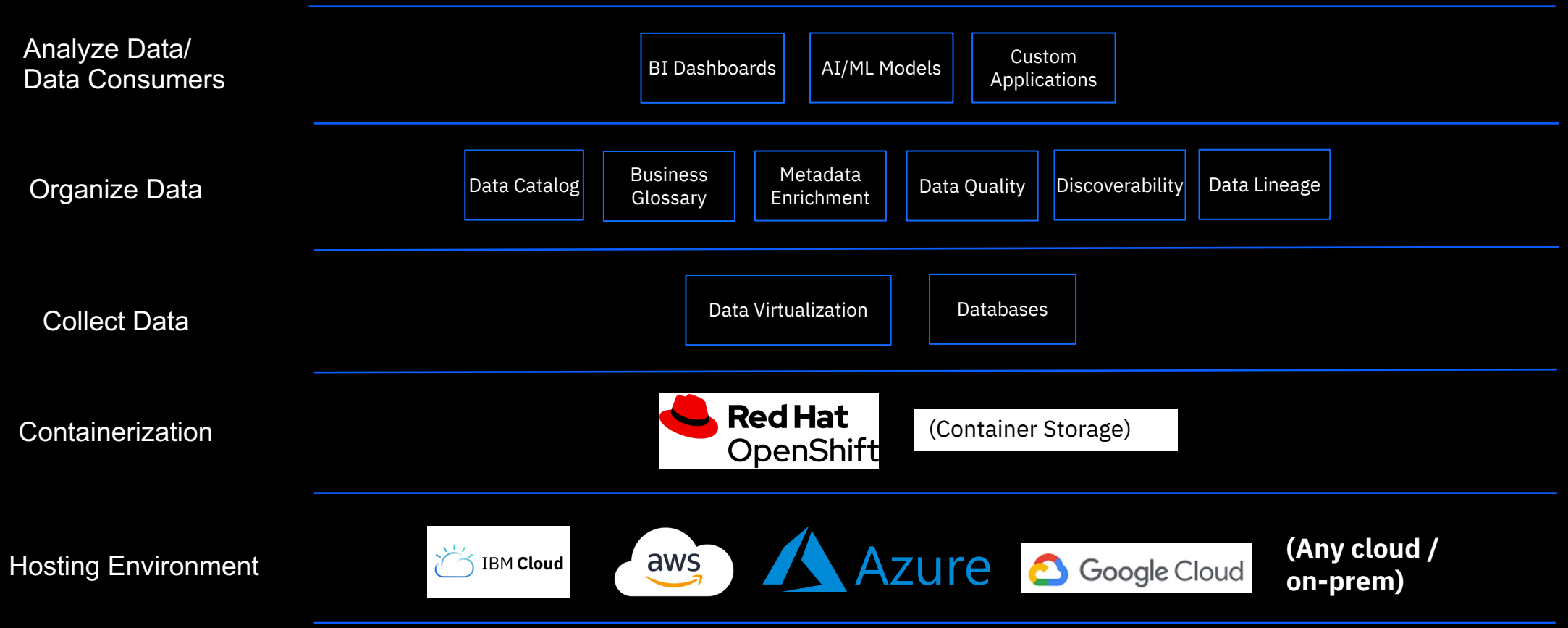
Business Glossary, Data Quality, Metadata Management, Discoverability and Data Lineage



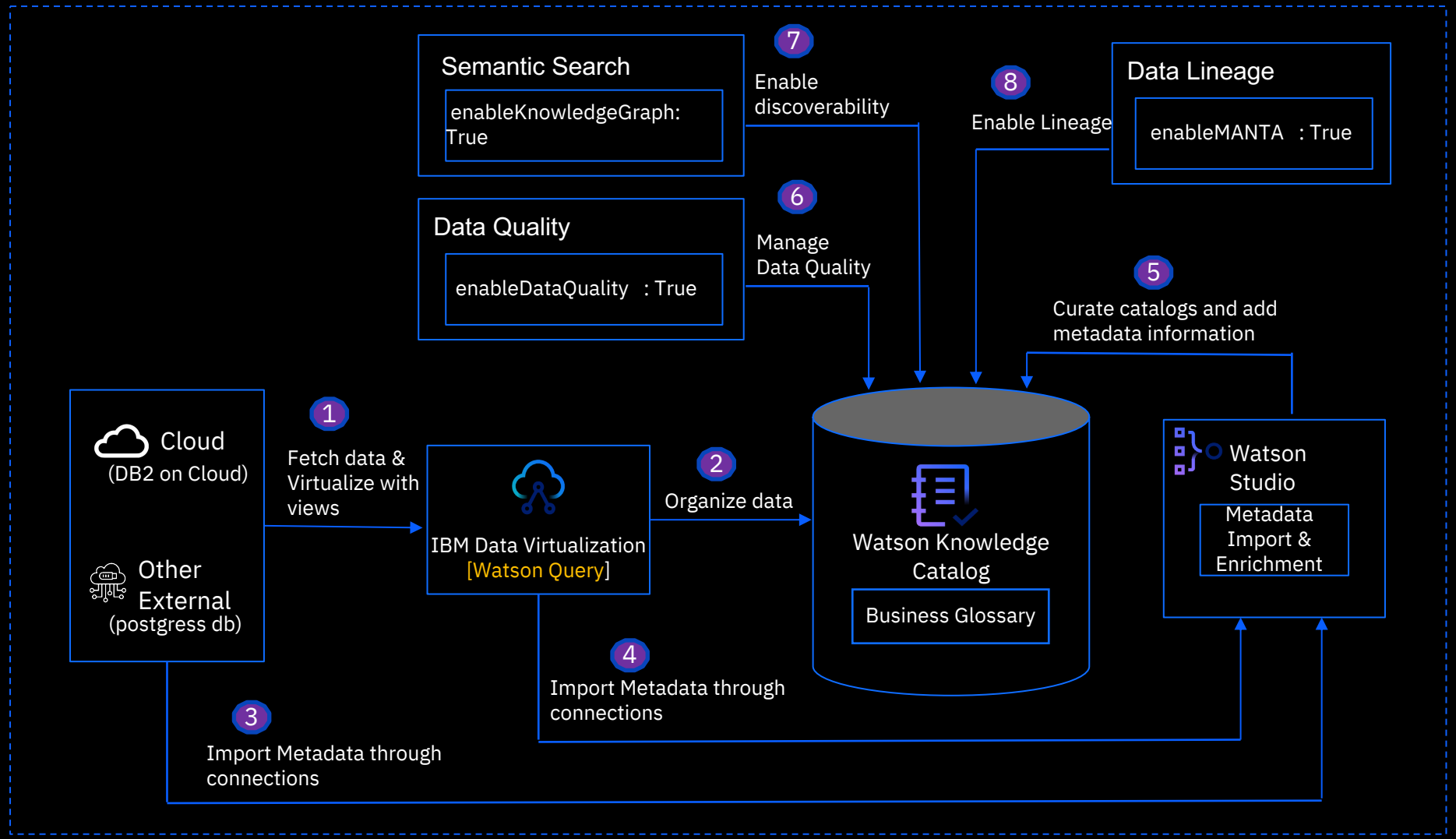
Overall Requirements

Reference Architecture Building

Arrange the identified features as a reference –



Component Architecture Building (include NFR)



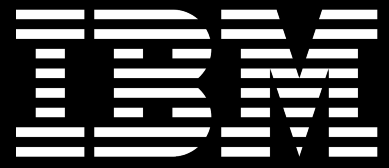
Executing and Presenting POC

PoCs by Ecosystem Engineering

Banking 360

Prime Brokerage

Sample Architectures by EESI Data & AI Squad


IBM