

IP&I Data Quality Mgt AI Use Case Briefing Pack

CIO Hackathon – 10-11 December 2025

Data Quality Mgt AI Use Case

Title:

- Creation of a Central Data Quality Management (DQM) Agentic AI solution

Content:

- Create a multi agent system that automates DQ identification, treatment strategy generation and remediation fix.
- Synthetic data has been provided with various DQ issues and anomalies planted within it.

The Ask:

- We would like you to:
 - build the 3 key agents (identifier, treatment and remediator agents) that meet the key requirements in this briefing.
 - design a set of relevant metrics that give the consumers a quick view of the materiality of DQ issues identified and build a 4th Metric Agent that can build Power BI views.
 - identify as many of the DQ issues and anomalies in the data provided as possible via the AI agents you have built, with proposed treatment strategies to fix and if possible, attempts made to clean the data.

Data Limitations:

- The synthetic data is in the format of IP&I BaNCS Life & Pensions data (modelled on the BaNCS Valuations Foundation Data Product design) and we have planted various DQ issues and anomalies within it.



The Key Agents

Intelligent, automated multi-agent system with multiple agents working together

Automated JIRA tickets to Operational teams if system fix is not possible

Multi-Agent System

The 'Identifier' Agent

To scan a data source, create and refine relevant DQ rules, identify DQ issues and anomalies, and automate manual process steps and trigger the Treatment Agent if issue is found.

The 'Treatment' Agent

To enrich an issue, determine its root cause, priority and treatment strategy options and assign it to an owner.

The 'Remediator' Agent

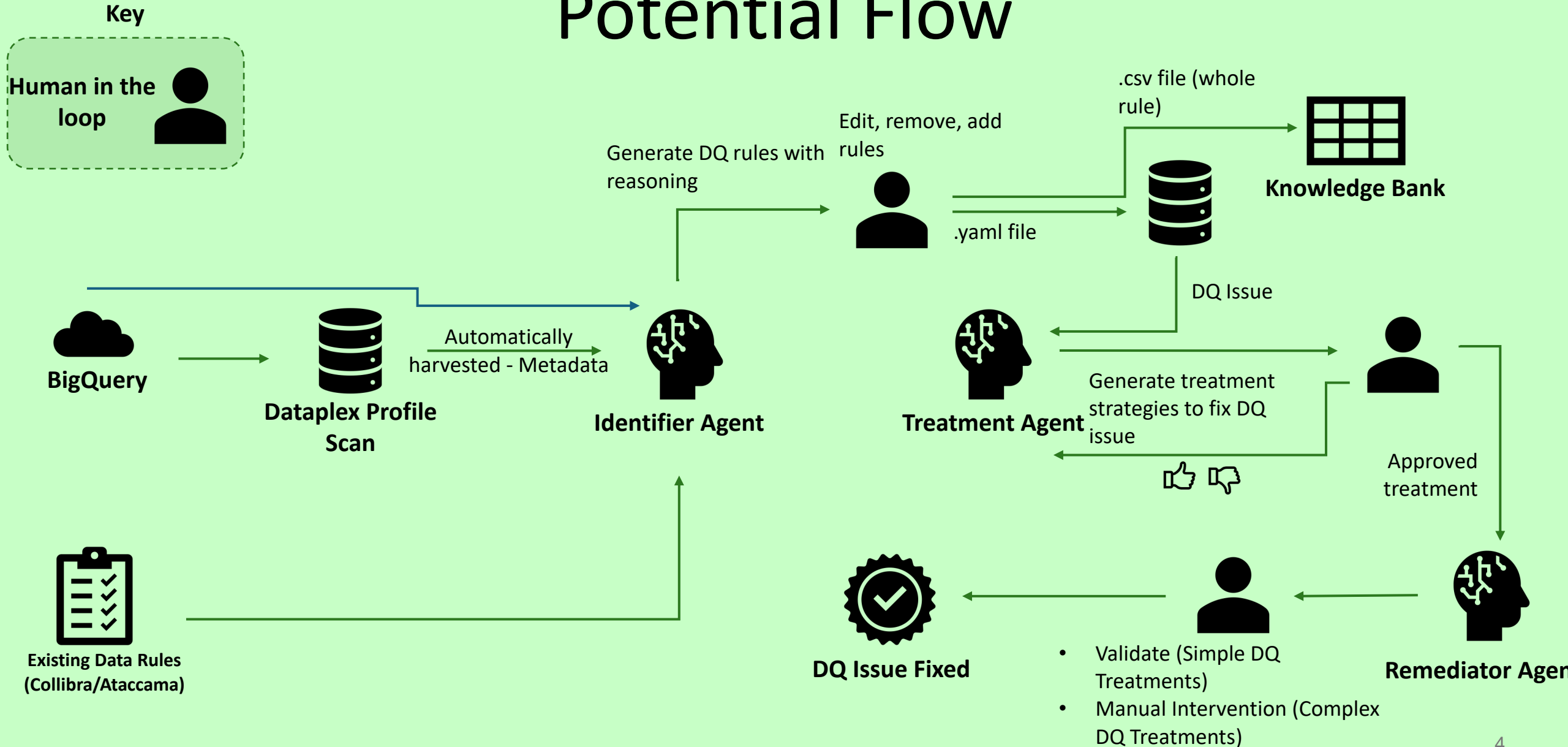
To fix directly in the source system or manage the human-in-the-loop process of fixing the issue.

The 'Metric' Agent

To create a clear suite of metrics that help inform the business of customer impact, materiality and cost of inaction.

Mini agents/workflows under each main bucket (e.g. Dataplex profile -> AI generated data quality rules -> HITL -> DQ rules)

Multi-Agent System Potential Flow



DQM Hack Challenge & Focus Areas

Your Challenge:

- build the 3 key agents (identifier, treatment and remediator agents) that meet the key requirements in this briefing.
- design a set of relevant metrics that give the consumers a quick view of the materiality of DQ issues identified and build a 4th Metric Agent that can build Power BI views.
- identify as many of the DQ issues and anomalies in the data provided as possible via the AI agents you have built, with proposed treatment strategies to fix and if possible, attempts made to clean the data.

Judges will be looking at whether you have fulfilled the following requirements in the build of the Agents:

DQM AI Agent Design

Purpose: Create a multi agent framework that automates as much of the e2e DQ process as possible with a route to live plan.

1. How well automated is the design? How well does this the design integrate with LBG infrastructure?
2. Is the design enduring, coherent and compliant?
3. Do the teams help judges understand the problem and the value of the solution?
4. Do you have a feasible route to live plan? Does it include resource, requirements and timescales?

DQ Identifier Agent

Purpose: Identify the issues planted in the anonymised data, show them in a user friendly report and fulfil the below goals.

1. Agent can recommend an initial set of valid rules
2. Agent can reference, use and refine previous rules as it learns
3. Agent is allowing generation of rules from natural language
4. Anomaly detection - Agent can call an API to create anomaly detection using natural language – based on thresholds, frequency of comparison, etc.
5. Agent can load outputs into user friendly reports.

DQ Treatment Agent

Purpose: Create value add treatment strategy options for the user to consider for each DQ issue.

1. Analyse and articulate the root cause of the issue.
2. Refer to previously defined DQ treatment strategies from Collibra to understand if there is treatment precedent for the particular issue identified.
3. If no previous treatment is available, the Agent should create a relevant set of treatment strategy options for the DQ issue for the user to consider.
4. A Knowledge Bank should be created which stores all of the previously approved treatment strategies.

DQ Remediator Agent

Purpose: Create an agent that can automate the fix in the source system or if not possible, generate a workflow ticket for relevant team.

1. Understand what would need to be built to be able to fix DQ issues in source systems.
2. Build the agent and attempt to clean the DQ issues, where possible.
3. If not possible, introduce capability to create a ticket in JIRA with the proposed treatment strategy that is sent through to the relevant team to fix.

DQ Metrics Agent

Purpose: Create an agent and design a set of metrics that provide a clear view of the materiality of the DQ issues raised.

1. Are the metrics clear in how they are calculated and their value?
2. Do they inform a neutral reader quickly whether there is an issue or not? And do they help inform the business of impact, materiality and cost of inaction.
3. Do they help provoke action in the right area?
4. Can the agent collate the information into a self serve dashboard of data quality.

Hack – Your Final Proposal



Your Submission Should Be:

- **Insightful** – Reveal something new or useful
- **Actionable** – Inform decisions or next steps
- **Clear** – Easy to understand
- **Creative** – Make smart use of the data
- **Collaborative** – Show teamwork and diverse thinking



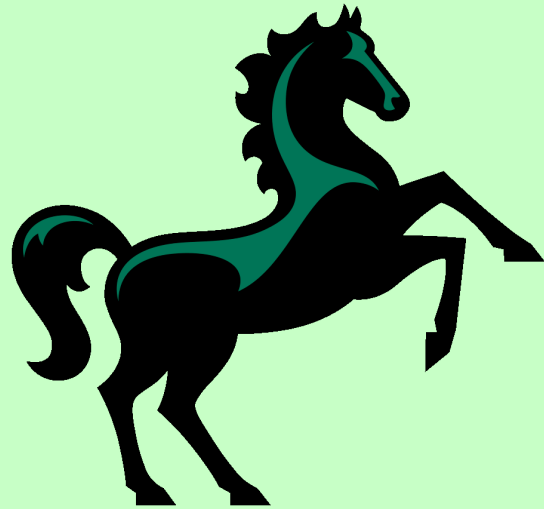
What to Create:

- A short, compelling story (5-6 slides or any format you and your team deem fit) and a prototype build of the agents whose value you can demonstrate as part of the Hackathon
- Focus on ideas, not polish
- Highlight key findings and recommendations



What to Aim For:

- Align with the principle of the brief
- Spot gaps and suggest improvements
- Show impact on IP&I and the wider Group



Good Luck!

The 'Identifier' Agent -> Code

Purpose

To scan a data source, create and refine relevant DQ rules, identify DQ issues and anomalies, and automate manual process steps and trigger the Treatment Agent if issue is found.

Trigger

When a data source is updated and/or set automatic checks

Prompt

- Data source
- List of Data Quality rules (an example of 3 DQ rules you could apply to the synthetic data are on the left but you should expand on these)
- Instructions File (Few-shot prompting, role, etc.)

Expected output

Information send to "Treatment" agent

Metrics

- Accuracy and completeness
- Complexity of data quality issue

Field Name	Column	DQ Rule	DQ Dimension
POLI_INCOME_PMT	X	value must be >=0 and <=100,000 Format allowed decimal to 2dp	Validity
CUS_DOB	G	Format DD/MM/YYYY age>18 must be in the last 130 years	Validity
CUS_NI_NO	F	must be unique	Uniqueness

The 'Treatment' Agent

Purpose

To enrich an issue, determine its root cause, priority and treatment strategy options. and assign it to an owner.

Issue Category	Specific Issue	Type	Proposed Treatment Strategy
Unfeasible Dates of Birth	Ages over 120 years or under 18 years	Alert user	Flag for review; correct or exclude from analysis/reporting; verify if policy type allows under-18 customers
	Dates of birth set in the future	Alert user	Flag and correct; set to null or request data owner confirmation
	Invalid date formats	Auto-correct	Standardise date formats; request correction if unresolved
Cross-Day Inconsistencies	Same policy number with different customer details across days	Investigate	Investigate source; reconcile records or escalate for manual review
	Policy valuation changes >20% for the same policy across days	Investigate	Audit change history; confirm with data owner or business rules

Trigger

The 'Identifier' Agent highlighting data issue

Prompt

- Metadata of Data source and Data Owner
- Metadata of data quality issue
- Expected Issue Categories (Some predefined)

Expected output

Log of issues and issues fixed by agent (at first can incorporate Human in the Loop)

Metrics

Accuracy, completeness and complexity of data quality issue (predefined scale)

The 'Remediator' Agent

Purpose

To fix directly in the source system or manage the human-in-the-loop process of fixing the issue.

Trigger

The 'Treatment' Agent highlighting data issue

Prompt

Log of issue classification to generate potential treatment strategy

Expected output

Treatment strategy for human, time from issue identification to remediation minimised

Metrics

Accuracy, completeness and remediation velocity

The 'Metric' Agent

Purpose

To create a clear suite of metrics that help inform the business of customer impact, materiality and cost of inaction..

Trigger

The 'Treatment' and 'Remediator' Agent

Prompt

- Log of issues from the 'Treatment' agent
- Log of fixes made by human by the 'Remediator' agent

Expected output

- Log of identification, treatment strategies, remediation on a dashboard
- Trending data quality metrics

Metrics

Accuracy, completeness, consistency, timeliness, customer complaint linkage, root cause clustering, issue recurrence rate, materiality index, remediation velocity – ensure they are easy to understand by business users