Kumaran’s

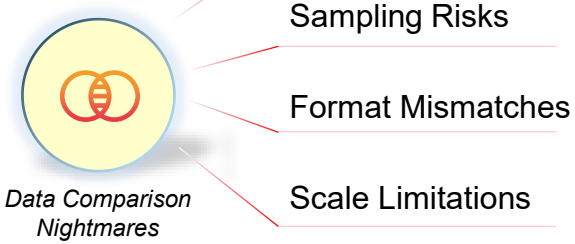
Data Quality Validator (DQV)

Engineered for Every Byte That Matters

.

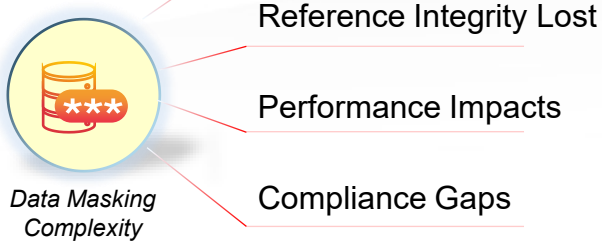
KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Pain Points in Data Processing



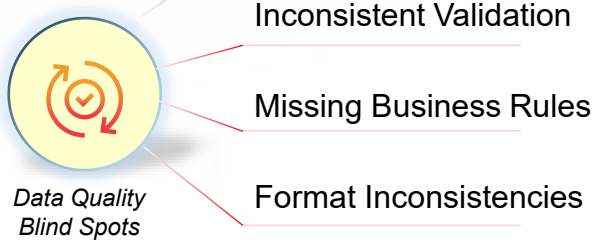
Data Comparison Nightmares

- Manual Excel Hell
- Sampling Risks
- Format Mismatches
- Scale Limitations



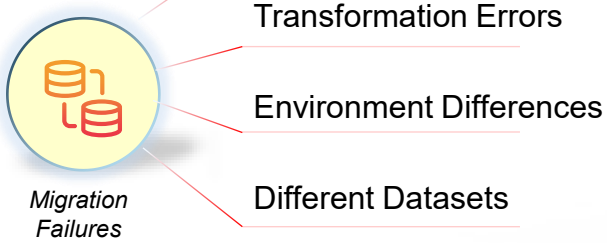
Data Masking Complexity

- Inconsistent Masking
- Reference Integrity Lost
- Performance Impacts
- Compliance Gaps



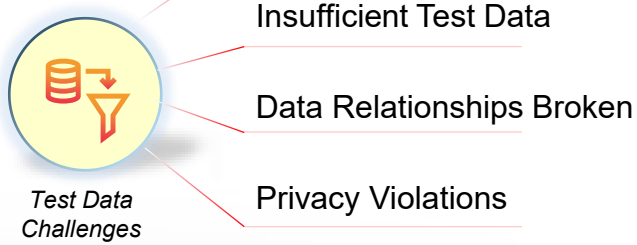
Data Quality Blind Spots

- Discovery Too Late
- Inconsistent Validation
- Missing Business Rules
- Format Inconsistencies



Migration Failures

- Data Loss
- Transformation Errors
- Environment Differences
- Different Datasets



Test Data Challenges

- Production Data Risks
- Insufficient Test Data
- Data Relationships Broken
- Privacy Violations

KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Solution Overview

DQV

- Unified Platform
- Any-to-Any Connectivity
- Enterprise Scale
- Business-Friendly

KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

DQV – Built on lessons learnt

Operations

- Comparison
- Masking
- Validation
- Transfer / Migration
- Synthetic Data Generation

Who needs DQV?

User						Use Cases
ETL Tester	✓	✓		✓	✓	✓ Validate & report quality of incoming data. ✓ Fixing the missing or invalid data. ✓ Mimic the data flow & process to get target data set. ✓ Compare source and target data sets. ✓ Generate synthetic data for testing the data process.
DBA		✓	✓			✓ Mask (in-flight, in-place, deterministic masking) data ✓ Migrate data between environments.
Developer	✓	✓		✓	✓	✓ Validate & report the quality of incoming data. ✓ Fixing the missing or invalid data. ✓ Export and import data to different formats. ✓ Compare two different data sets – Unit testing. ✓ Generating synthetic data in the absence of production grade data
Functional QA		✓		✓	✓	✓ Generate test data out of existing database records. ✓ Compare two different data sets / exported files / DB. ✓ Generate synthetic test data.
API Tester		✓		✓	✓	✓ Generate synthetic test data. ✓ Mimic the data flow & process to get target data set. ✓ Compare actual vs expected results.

KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Features & Data Sources

General Features

- Supports various Homogeneous or Heterogeneous data sources *(Extensible to new data sets)*
- Parallel Processing
- Lesser memory / CPU even for multi-million records
- Detailed field level comparison results
- Detailed mismatch reports
- Integration with CI/CD pipeline; or embeddable in application

Supported Data Sources

✓ **Databases**
- Any Relational Database (Oracle , MS SQL Server / Azure SQL, MySQL / MariaDB, PostgreSQL, etc)
- Any NoSQL database having JDBC drivers (MongoDB, DataBricks, etc)
- ODBC data sources via ODBC-JDBC driver.

✓ **Files**
- Delimited Files (CSV, TSV, custom delimiter)
- Fixed Length text files
- Excel files (old and new formats)

✓ **File Formats**
- XML
- JSON

✓ **Special Data Sources**
- Parquet files
- REST APIs
- Any AMQP supported middleware (ActiveMQ, IBM-MQ, Solace, etc)

✓ **Any other custom data sources**

KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Operations

Comparison

- Supports different structures
- Memory efficient even on large data comparison
- Detailed comparison results, highlight field level data differences
- Detailed mismatch report

Validation

- Validate the quality of data (part of profiling)
- Field level: Null values, Value in Range, List of Values, Format / pattern
- Record level: Conditions against value of other fields (example: max_val >= min_val)
- Fix the data on-the-fly for missing or incorrect values

Masking

- Deterministic masking
- In-Place / In-Flight masking

Generation

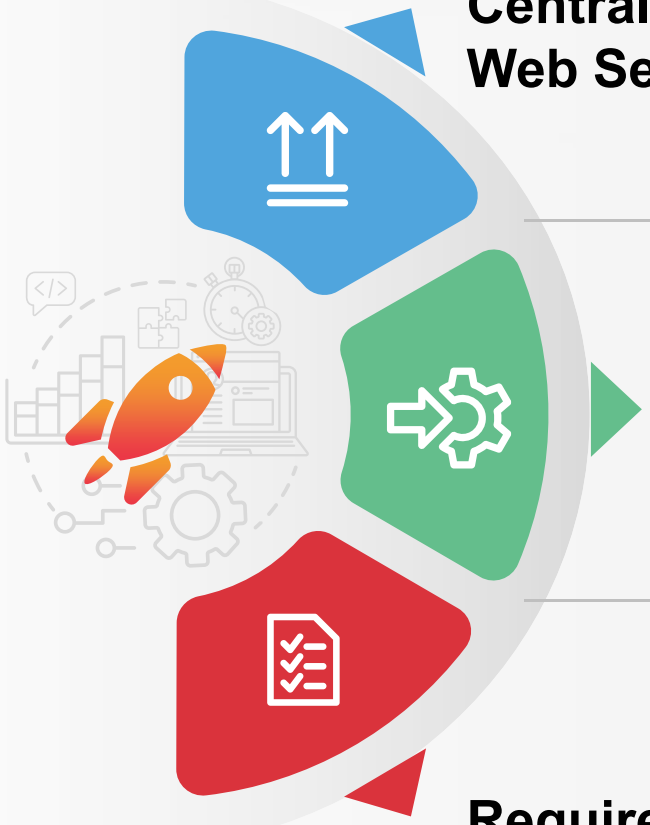
- Meaningful data generator (names, addresses, etc)
- Numerical, String generators with given range, list of values
- SIN / SSN fake value generation
- Random, Sequential selections

Transfer / Migration

- Filter & Transform data
- Aggregate the data
- Create calculated fields

KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Deployment Options



Standalone or **Centralized Web Server**

- Centrally managed web application
- Or, executable as a single JAR file; with configuration file names as parameters
- Supports parallel processing within each execution

Integration

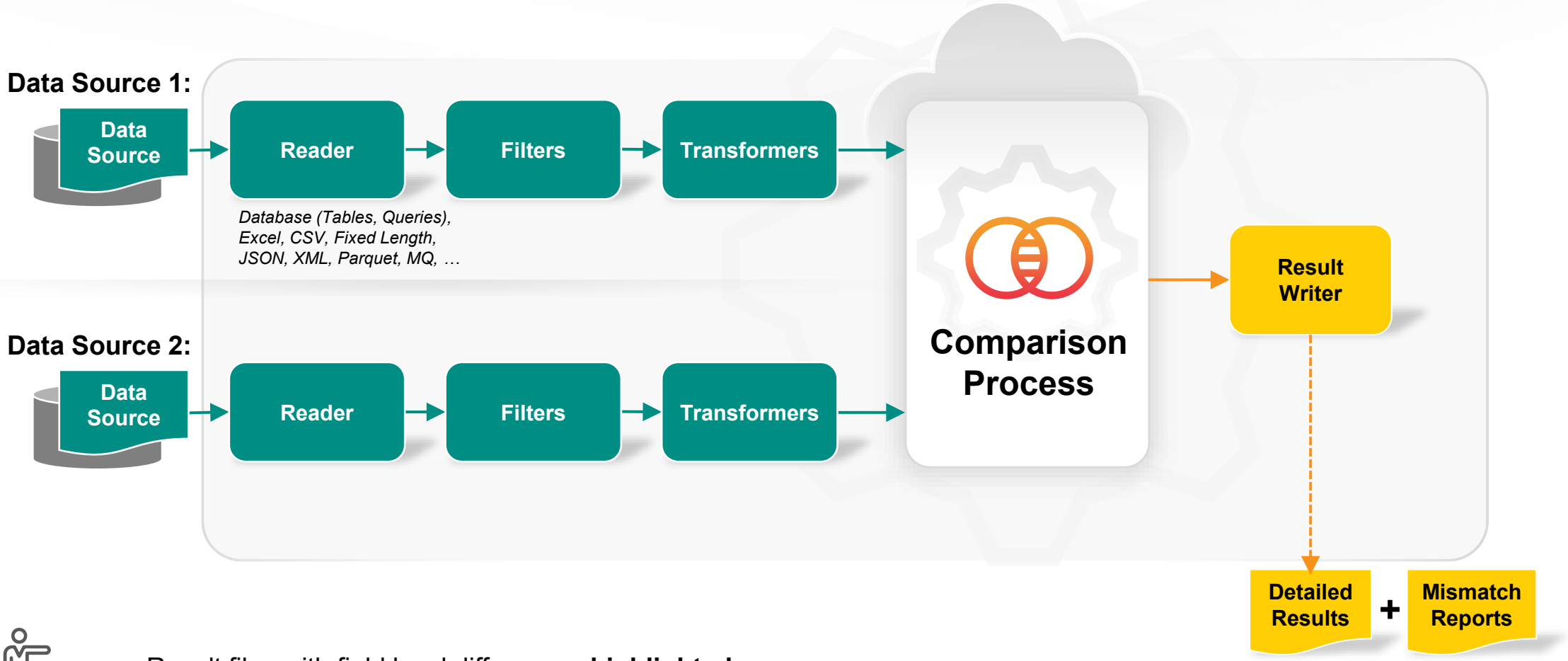
- Embeddable in the application
- Embeddable in ETL tools (Informatica, DataBricks)
- Integrated within CI/CD pipeline.
  - *Failure can be determined using exit code.*

Requirements

- Just Java is enough
  - *Minimum version 1.8.x*
- Recommended to have 4 GB RAM
  - *More the memory, quicker the process*
- Should have connectivity and access to different data sources

KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Comparison Process



- Result files with field level differences **highlighted**
- Supports comparing **entire schema** in one shot

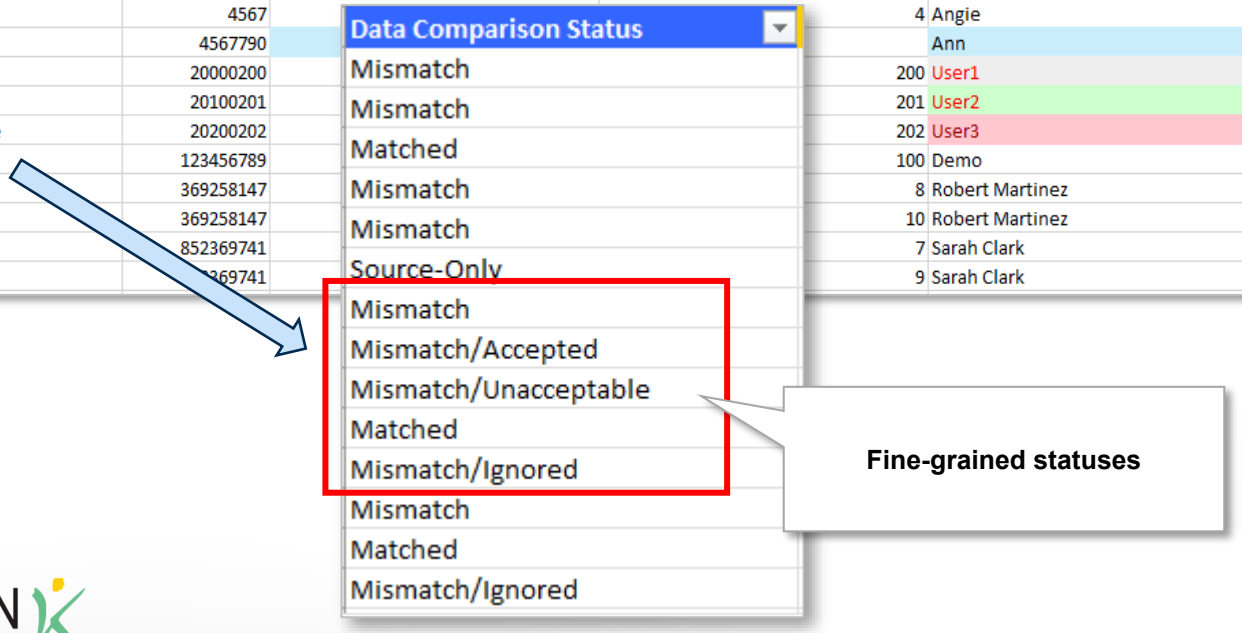
KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Comparison – Actions

DQV has advanced control of “Difference Actions” which is used to achieve finer controls over the **ignorable**, **legitimate** or **unacceptable** differences.

For the required fields, specify the condition and one of the actions → “`accept`”, “`ignore`”, or “`unacceptable`”

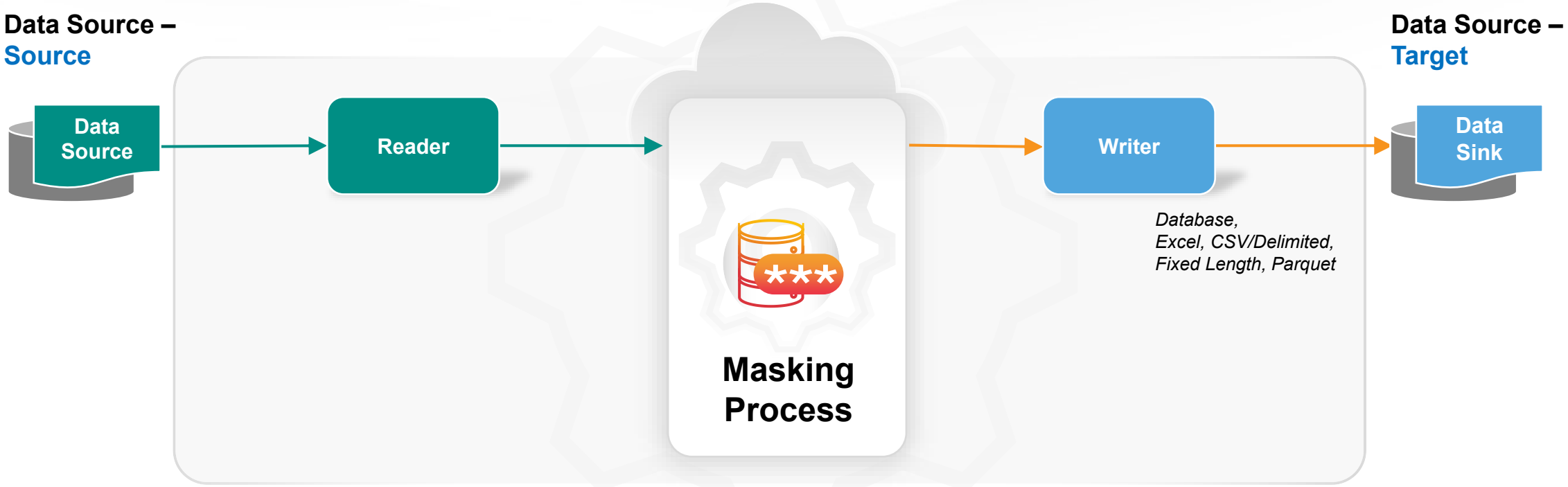
	A	B	C	D	E	F	G	H
1	K-DQV (1.0.0.505)		loan_data1.csv - Input1 (Source)	loan_data2.csv - Input2 (Target)	loan_data1.csv - Input1 (Source)	loan_data2.csv - Input2 (Target)	loan_data1.csv - Input1 (Source)	loan_data2.csv - Input2 (Target)
2	loan_data1.csv - Input1 (Source)	accountNumber						
3	loan_data2.csv - Input2 (Target)	accountNumber						
4	Data Comparison Status	accountNumb	id	id	name	name	type	type
5	Mismatch	1234	1	1	Dave	Dave	Education	Home
6	Mismatch	1234	3	3	Dave	Dave	Home	Education
7	Matched	1234	6	6	Dave	Dave	Home	Home
8	Mismatch	4567	2	2	Angie	Angie	Home	Home
9	Mismatch	4567		4	Angie	Angie	Car	Car
10	Source-Only	4567790			Ann		Personal	
11	Mismatch	20000200		200	User1	UserA	Other	Home
12	Mismatch/Accepted	20100201		201	User2	UserB	Home	Home
13	Mismatch/Unacceptable	20200202		202	User3	UserC	Education	Home
14	Mismatch	123456789		100	Demo	Demo	Other	Other
15	Mismatch/Ignored	369258147		8	Robert Martinez	Robert Martinez	Car	Car
16	Mismatch	369258147		10	Robert Martinez	Robert Martinez	Education	Education
17	Matched	852369741		7	Sarah Clark	Sarah Clark	Personal	Personal
18	Mismatch	369741		9	Sarah Clark	Sarah Clark	Home	Home



Fine-grained statuses

KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Masking Process



- **Deterministic masking** support) for PDiT requirements

Masking Features

In-Flight / In-Place data masking

- In-Flight: Mask the data ask the data is copied from source to target. Ex: PROD to DIT database
- Use any data source (DB to DB, File to DB or DB to File)
- In-Place: Read the data from table, mask and write back to the table.

Deterministic masking

- Apply the same masked value across other tables in the db.
- Maintain referential integrity across parent-child relationships.

Repeatable/deterministic masking

- Masked data can be stored in file system (Mask Info store) for re-use in another migration run or different database / schema.
- Original values are protected – i.e., only hash value of original value and the new value are stored.

Various masking **Techniques**

- Redaction
- Replacement
- Randomize
- Nullify / Clear
- One-way Hashing / Encryption

KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Masking Techniques

Redact

- Retain part of the source data and replace with redaction character ('x')
- Supports retaining or replacing source characters with alphabet, number characters.

Example:
Source Data: "john.david@abc.com", Mask: "=x*(.)=x*(@)==x*(.)="
Result: "jxxx.dxxxx@abx.com"

Replace / Variance

- Replace the values with new values which are meaningful incl. sequential numbers.

Example: "John David", Replace: "generate:name().fullName()" → "Samuel Daniel"

Example: "100", Replace: "applyVariancePct(current_value, 20)" → "85"

Randomize

- Generate random values, or from range, or from list/lookup or regex patterns

Example: "random(5)" → "39563", "getNext('country')" → "France"

Nullify / Clear

- Clear the values from the field i.e., set to NULL

Hashing / Encrypt

- One way hashing using SHA-256 algorithm
- Reversible encryption with the given secret

Generally, not recommended for sharing data.

KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Masking – Results

	Row #1
CUSTOMER_ID	1
FIRST_NAME	John
LAST_NAME	Doe
SSN	123456789
EMAIL	john.doe@example.com
PHONE_NUMBER	555-1234
DATE_OF_BIRTH	1985-03-15 00:00:00.000
ADDRESS	123 Elm St, Cityville, USA
NOTES	Username is Doe, John. SSN is 123456789. Email is john.doe@example.com, Phone number is 555-1234.

← **Source** data

Masked data →

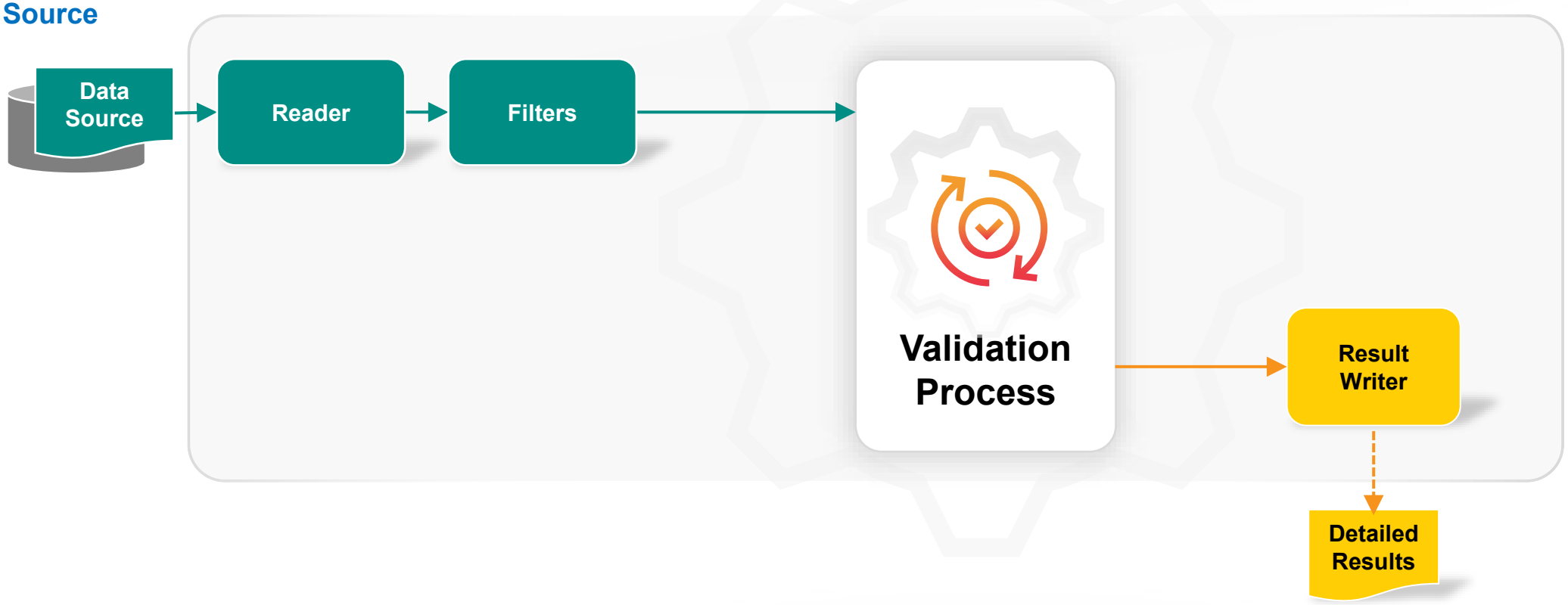
	Row #1	
CUSTOMER_ID	1	
FIRST_NAME	Gavin	← Replace
LAST_NAME	Bogan	
SSN	142965048	← Randomize
EMAIL	jxxx.xxx@exxxxxx.xxx	← Redact
PHONE_NUMBER	555-0149	← Redact & Randomize
DATE_OF_BIRTH	1985-03-15 00:00:00.000	
ADDRESS	Suite 614 829 Boehm Spring, North Annemariestad, IA 96897	← Replace
GENERATED_FIELD	Username is Bogan, Gavin. SSN is xxxxxxxx. Email is jxxx.xxx@exxxxxx.xxx, Phone number is xxxxxxxx.	

↑ Redact & Replace

KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Validation Process

Data Source –
Source



KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Validation Checks

Validation at Field level

- Data Type
- Format of data – Number, String and Date formats, patterns
- Range (Min, Max) of values for Numbers and Dates
- Allowed list of values / Lookup reference

Business Validation at Record level

- Validate relationships between fields –
  Example: “End Date >= Start Date”, Or “Request Date < Today”
- Displays custom message on failures

Validation at entire table level

- Uniqueness

Reporting Issues

- Log all errors at field, record and table level in an Excel

Fixing Values

- Should treat empty values as missing?
- Fixing Missing values
- Fixing Invalid values

KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Validation – Error Report

PK Field Name(s)

Failed field name

Actual Value

Custom error message

	A	B	C		
1	emp_id	Field	Validation Name	Actual	Error Message
2	1	emp_id	Range	1	Employee-id is not in range '2-8'.
3	1	date	Range	2025-09-08-00:00:00	date is missing.
4	1	date	Data type and format	2025-09-08-00:00:00	Format has error.
5	1	date	Data type and format	2025-09-08-00:00:00	Format has error.
6	1		Record condition	age = 25, risk = HIGH	Age must be greater than 25.
7	1		Record condition	age = 25, risk = HIGH	
8			Range	1(	
9	10	date	Range	2024-02-09-00:00:00	
10	10	date	Data type and format	2024-02-09-00:00:00	
11	10	date	Data type and format	2024-02-09-00:00:00	
12	2	date	Range	2025-05-08-00:00:00	
13	2	date	Data type and format	2025-05-08-00:00:00	
14	2	date	Data type and format	2025-05-08-00:00:00	
15	2		Record condition	age = 28, risk = HIGH	
16	3	date	Range	2026-05-08-00:00:00	
17	3	date	Data type and format	2026-05-08-00:00:00	

Record level validation

Field level validation

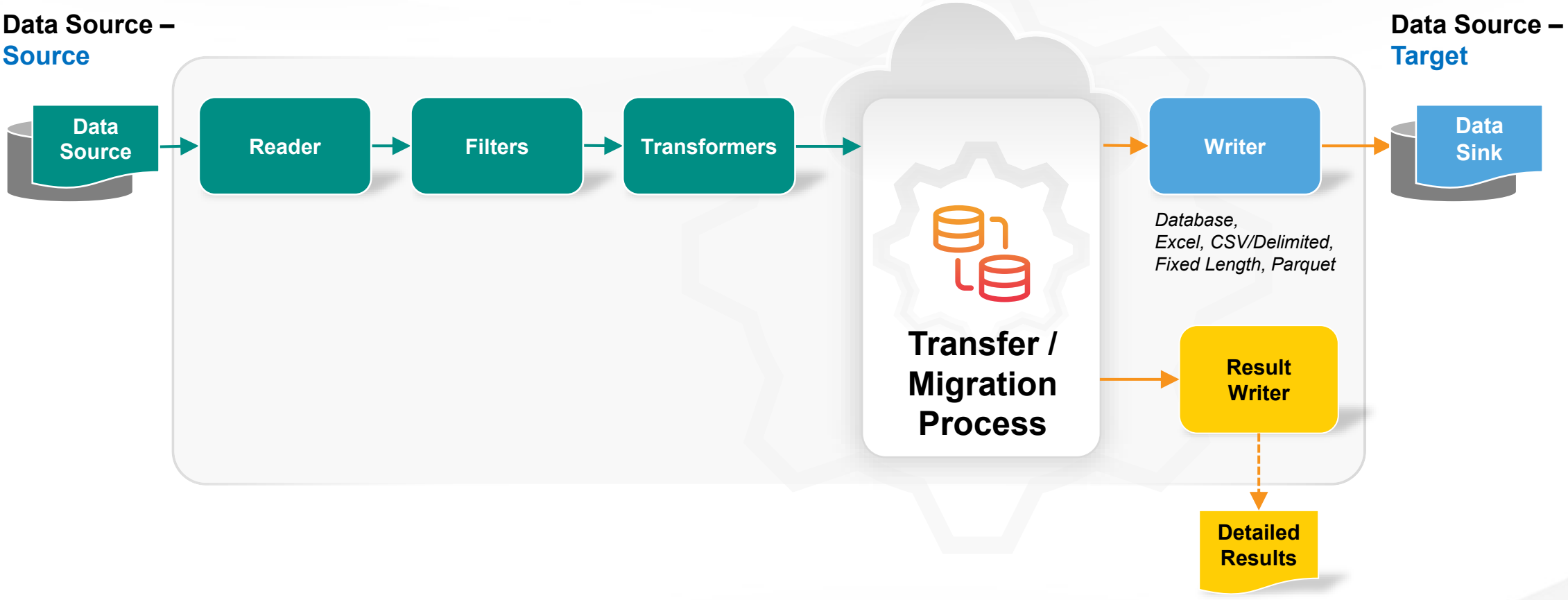
	A	B	C	D	E
1	id	Field	Validation Name	Actual	Error Message
2	11		Data set condition	emp_id = 5	employee-id must be unique
3	12	country	Lookup	US	country is not found for lookup-key : country-code .
4	12	country	Allowed values	US	country is not in given allowed values
5	13			USA	country is not found for lookup-key : country-code .
6	13		alues	USA	country is not in given allowed values
7	13		Data set condition	emp_id = 6	employee-id must be unique
8	14	country	Lookup	null	country is not found for lookup-key : country-code .
9	14	country	Allowed values	null	country is not in given allowed values
10	16		Data set condition	emp_id = 8	employee-id must be unique
11	19	country	Lookup	US	country is not found for lookup-key : country-code .
12	19	country	Allowed values	US	country is not in given allowed values
13	19		Data set condition	emp_id = 10	employee-id must be unique
14	20	country	Lookup	USA	country is not found for lookup-key : country-code .
15	20	country	Allowed values	USA	country is not in given allowed values
16	20		Data set condition	emp_id = 10	employee-id must be unique
17			Data set condition	emp_id = 2	employee-id must be unique
18			Data set condition	emp_id = 2	employee-id must be unique

Dataset level validation

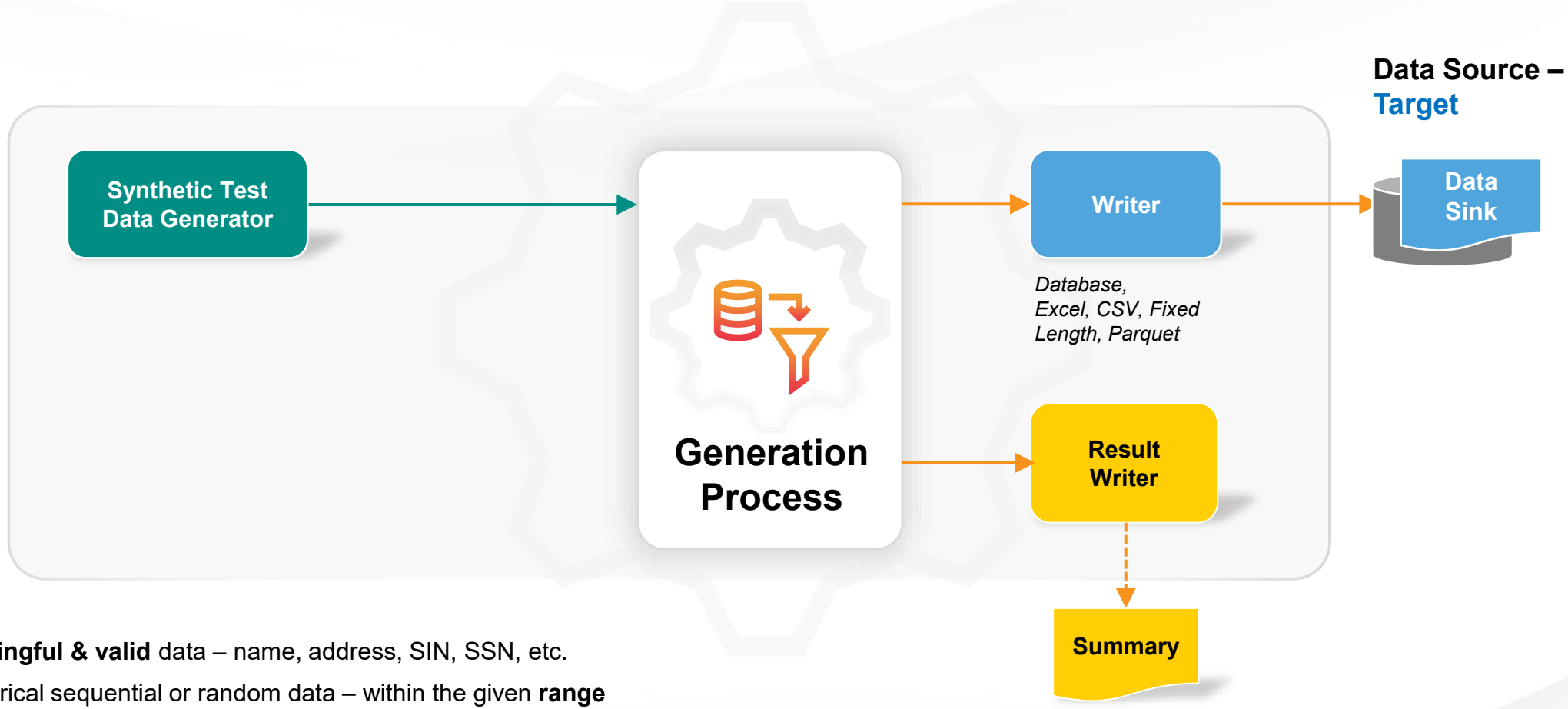
Field level validation

KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Transfer / Migration Process



Generation Process



- **Meaningful & valid** data – name, address, SIN, SSN, etc.
- Numerical sequential or random data – within the given **range**
- String data – Sequential or Random data selected from a **pre-defined list**

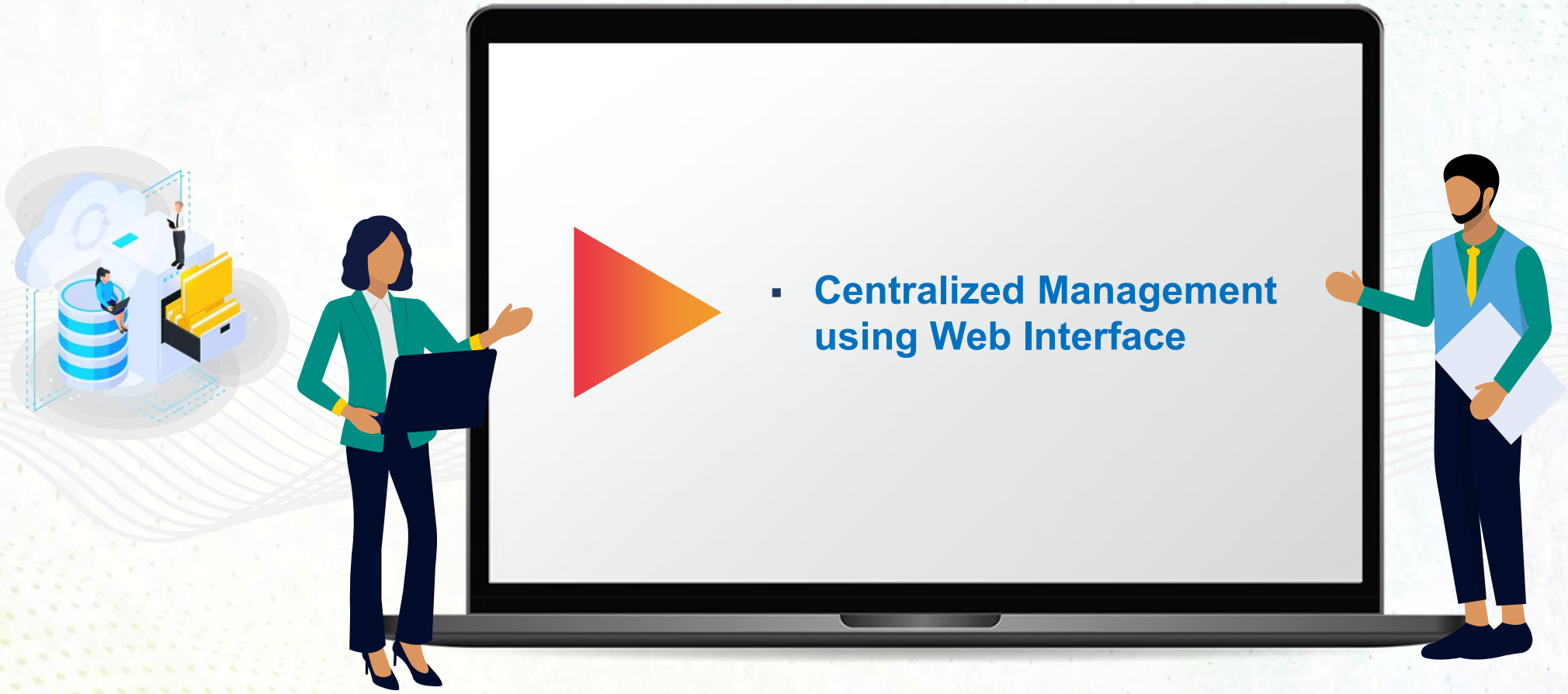
KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Synthetic Data Generation

Meaningful Fake Data	▪ Names, Addresses, etc ▪ Random number, strings ▪ Conditional data based on other field values
Pattern	▪ Regular Expression pattern values ▪ Alpha Numeric pattern
ID numbers	▪ SSN and SIN numbers
Lookups – List of values	▪ Sequentially using values ▪ Randomly using values ▪ Lookup from reference data / tables
Parent-Child relationships	▪ Generate child data for each parent record ▪ Lookup data from another reference table ▪ Generate child records as JSON/XML in parent
Masking Data	▪ Pattern Masking ▪ Random Masking ▪ Deterministic Masking

KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

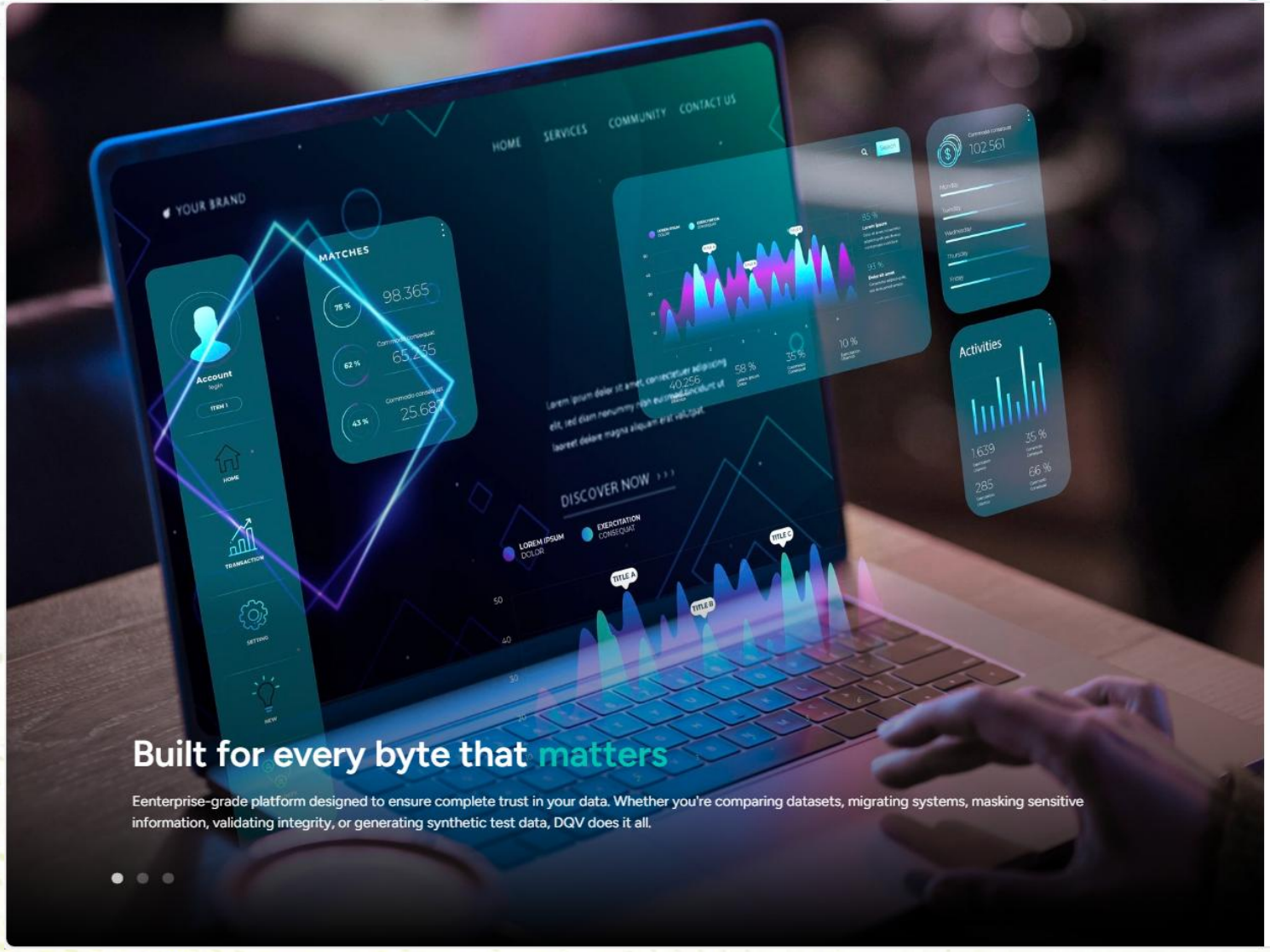
User Interface



- Centralized Management using Web Interface

KUMARAN SYSTEMS

Login



Data Quality Validator

Login to your account

Username

Password

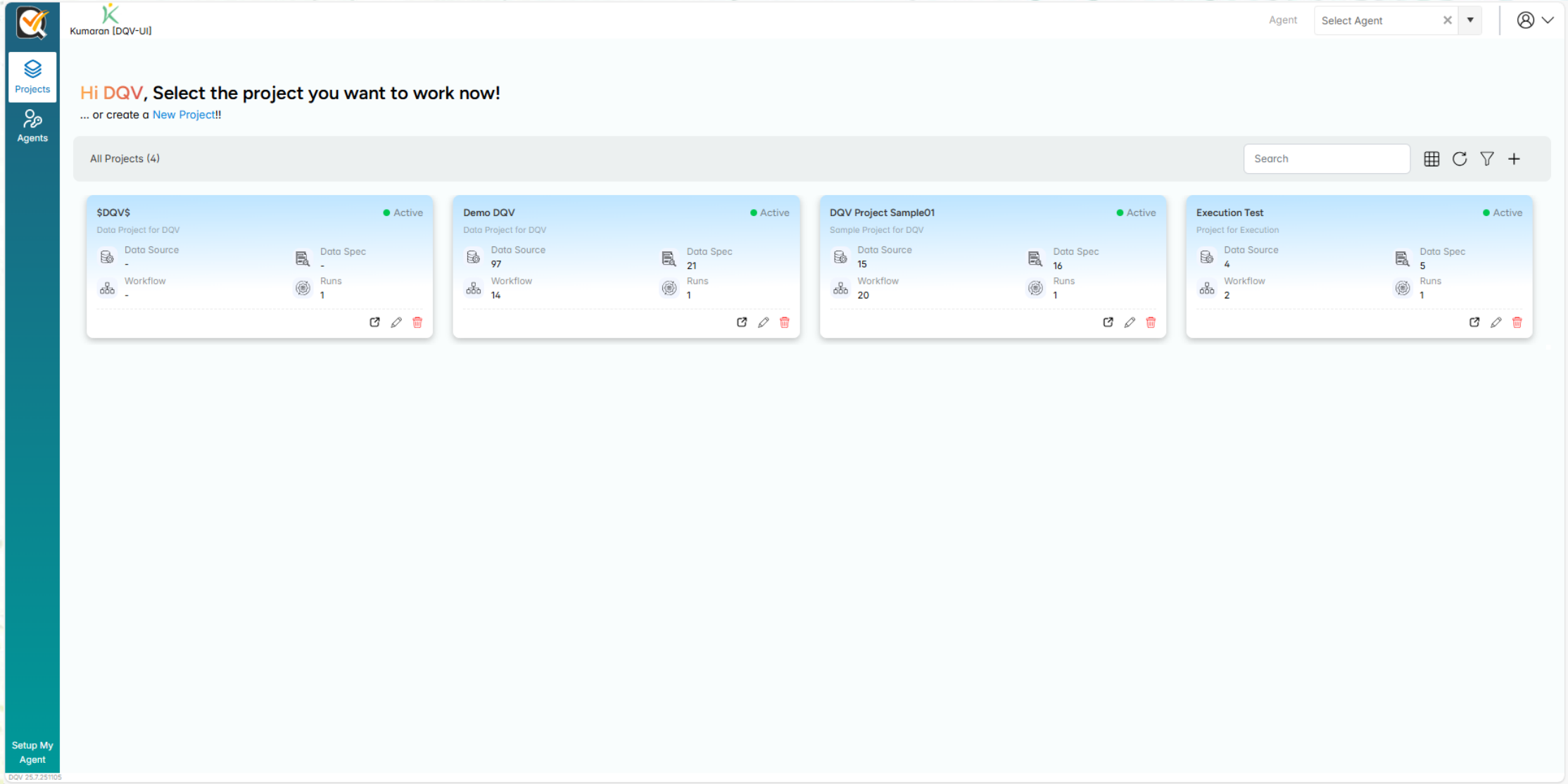
Sign in

Forgot Password?

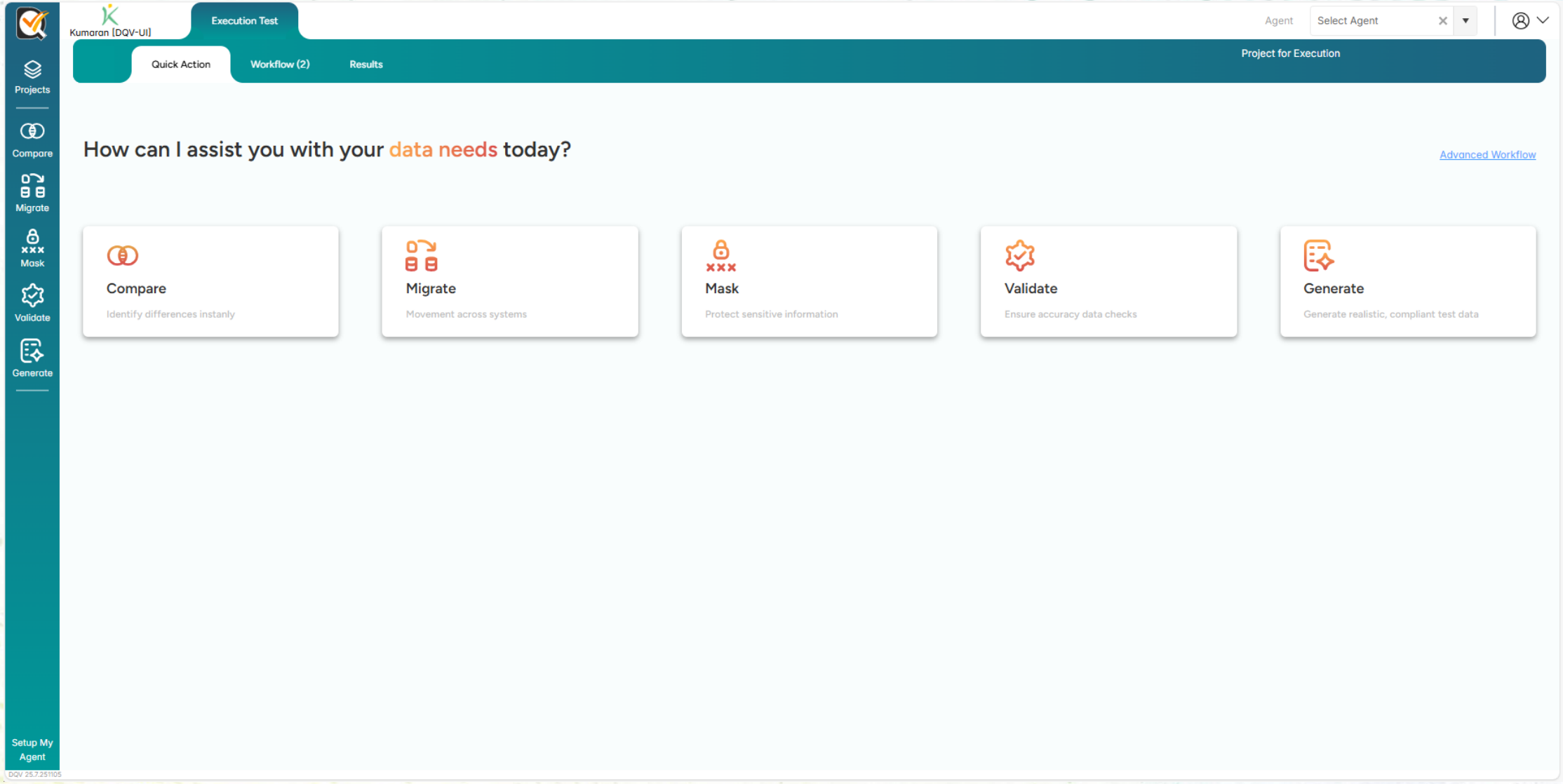
Powered by KUMARAN SYSTEMS

KUMARAN SYSTEMS

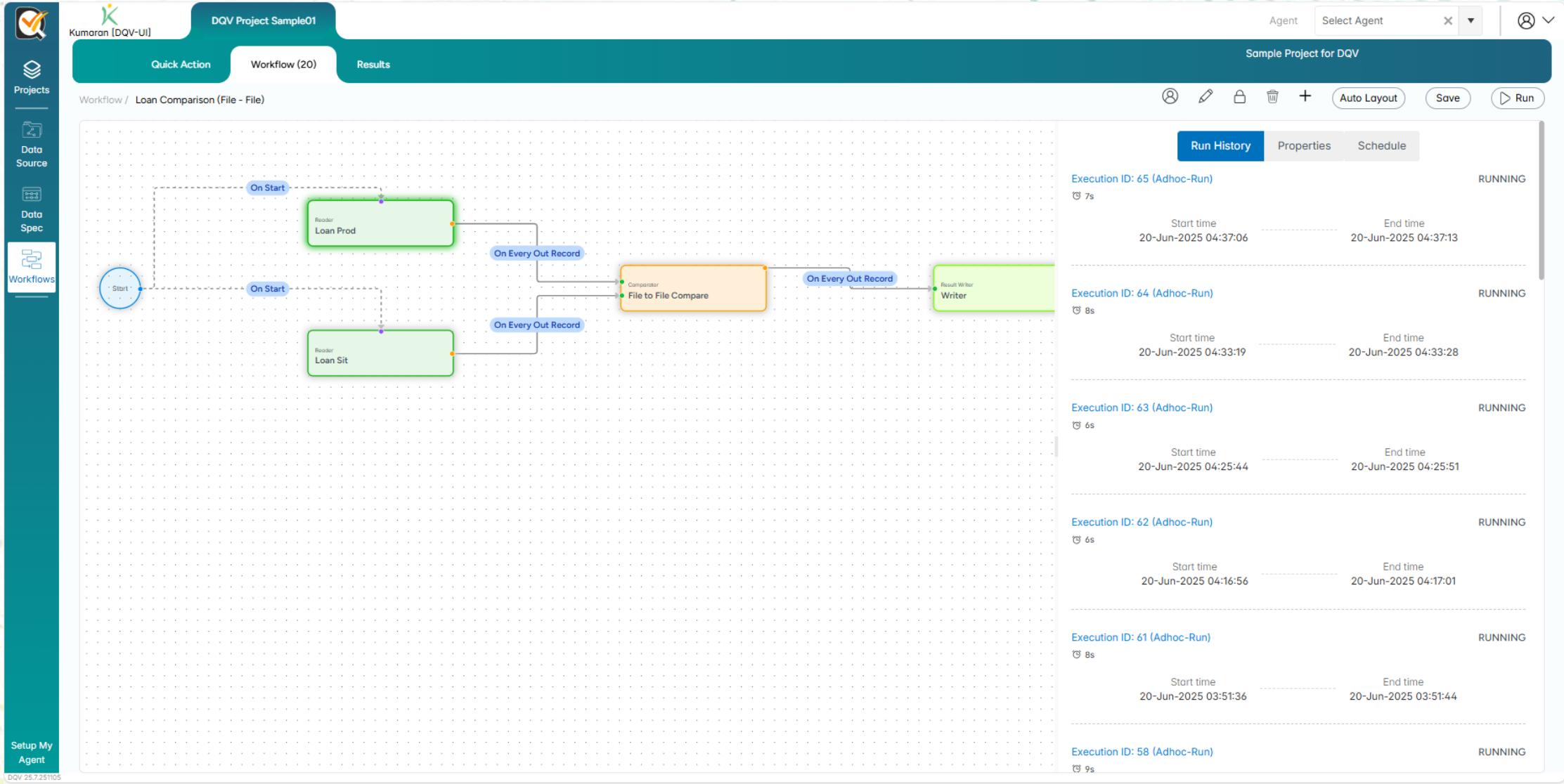
Project Selection



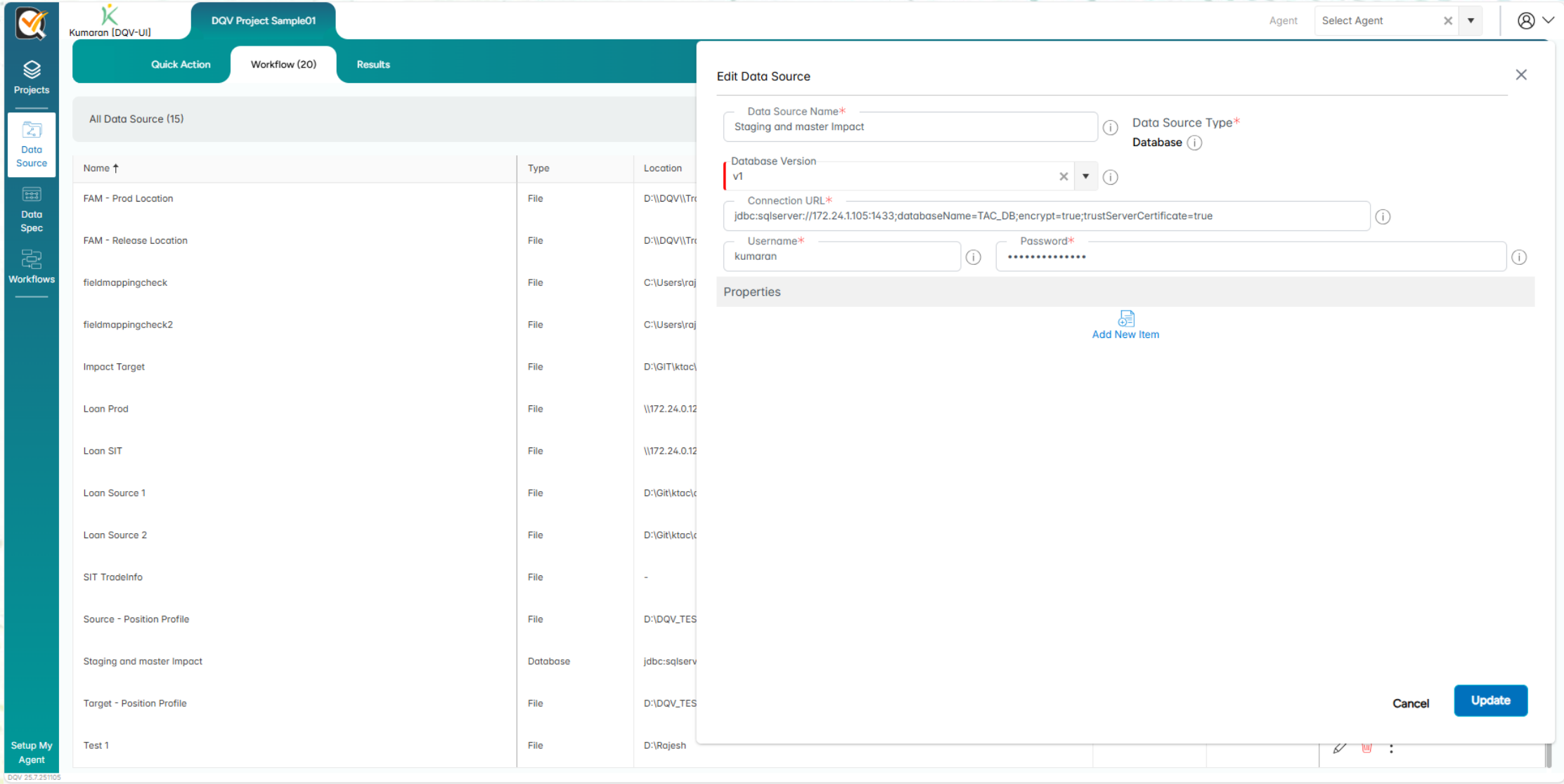
Quick Actions



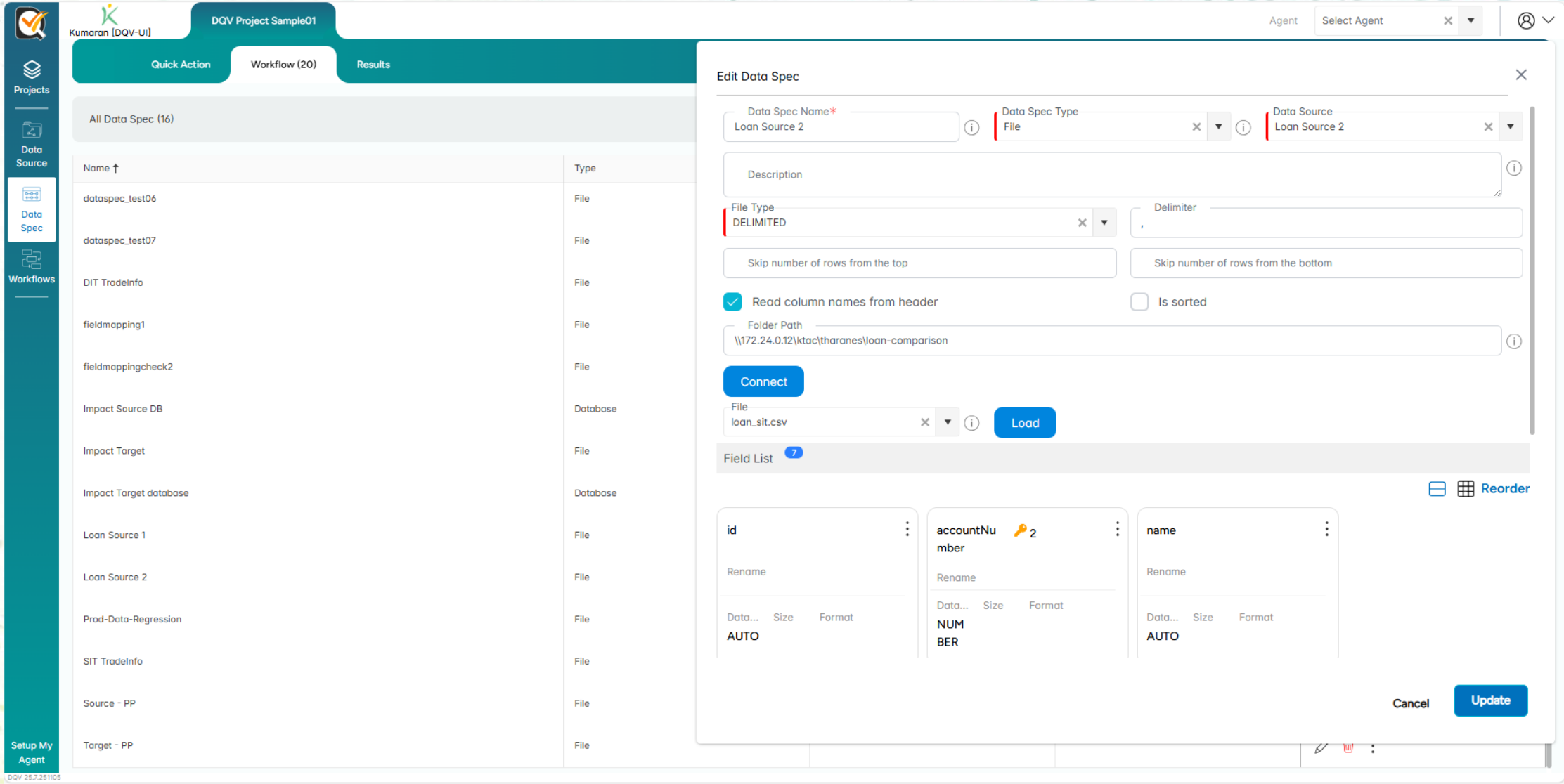
Workflow Editor



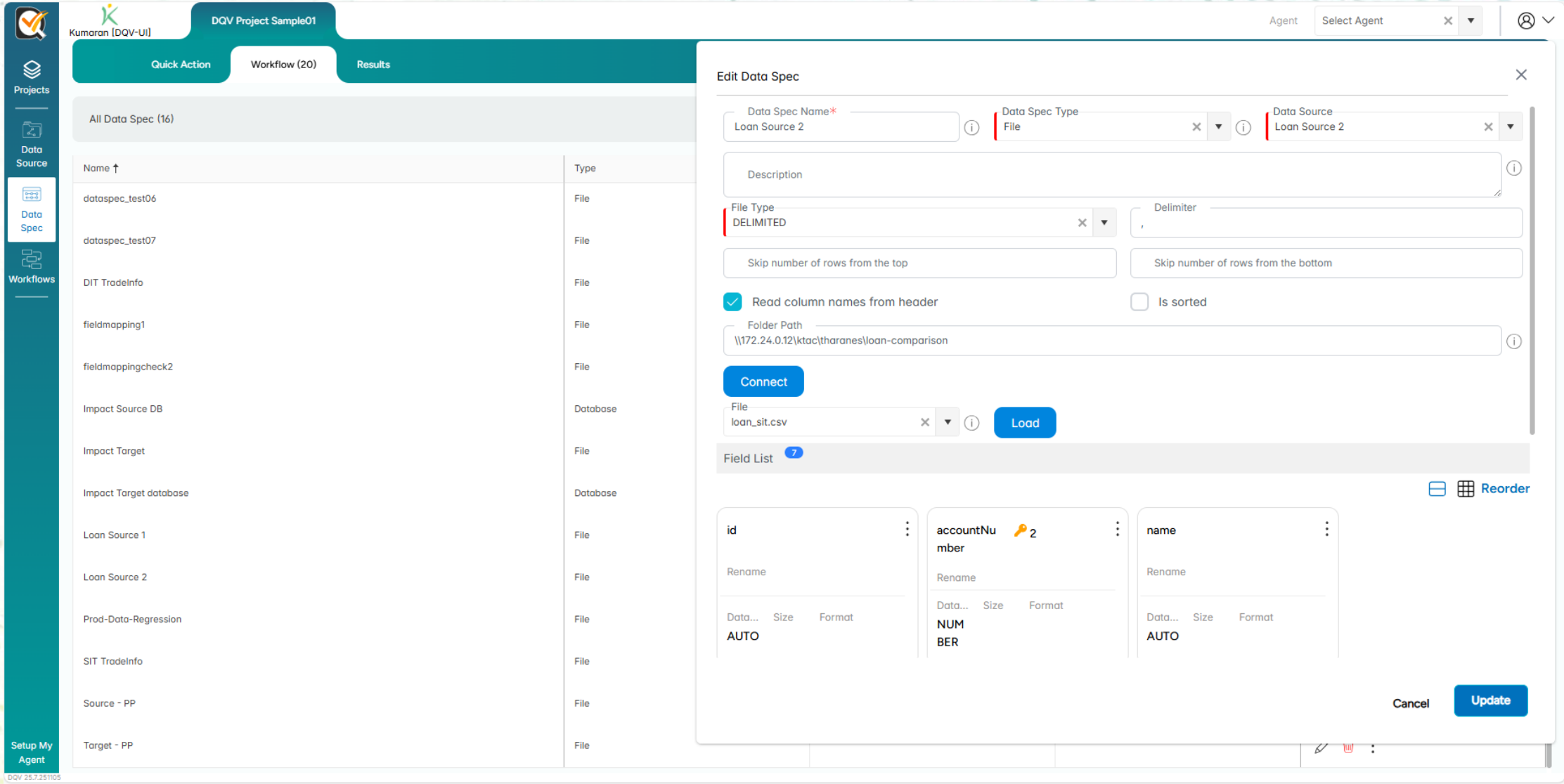
Data Source definition



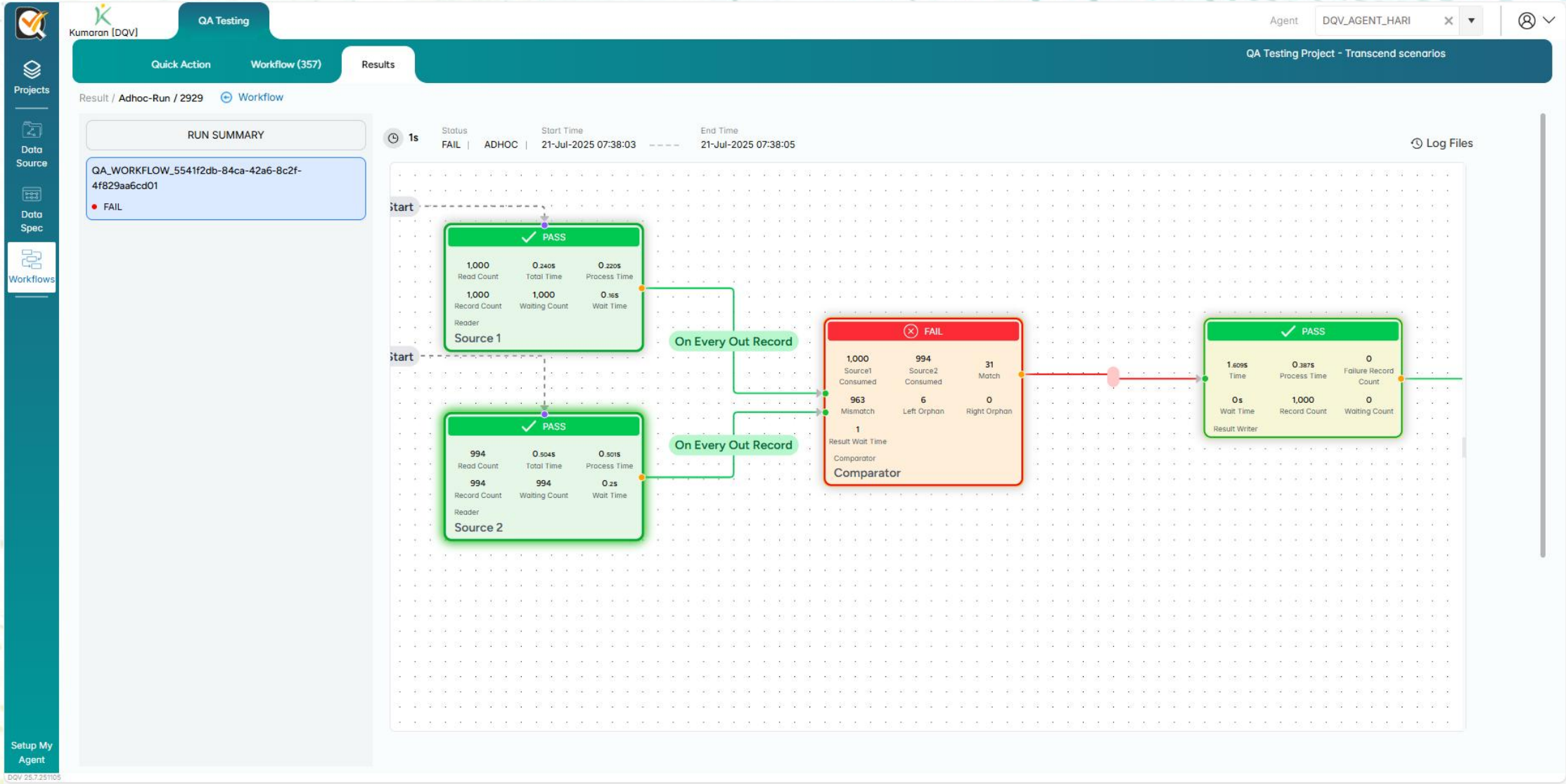
Data Specification



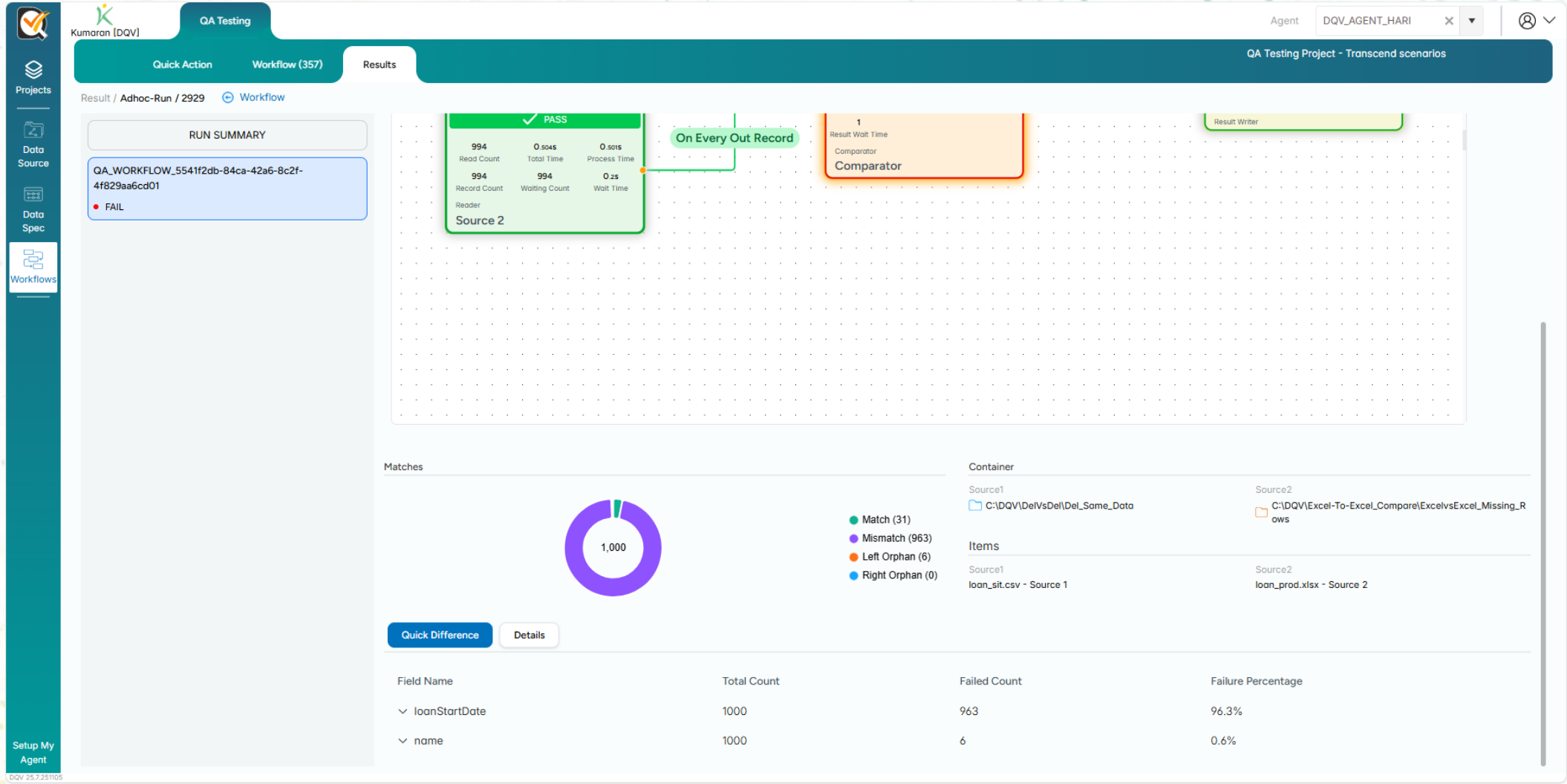
Results



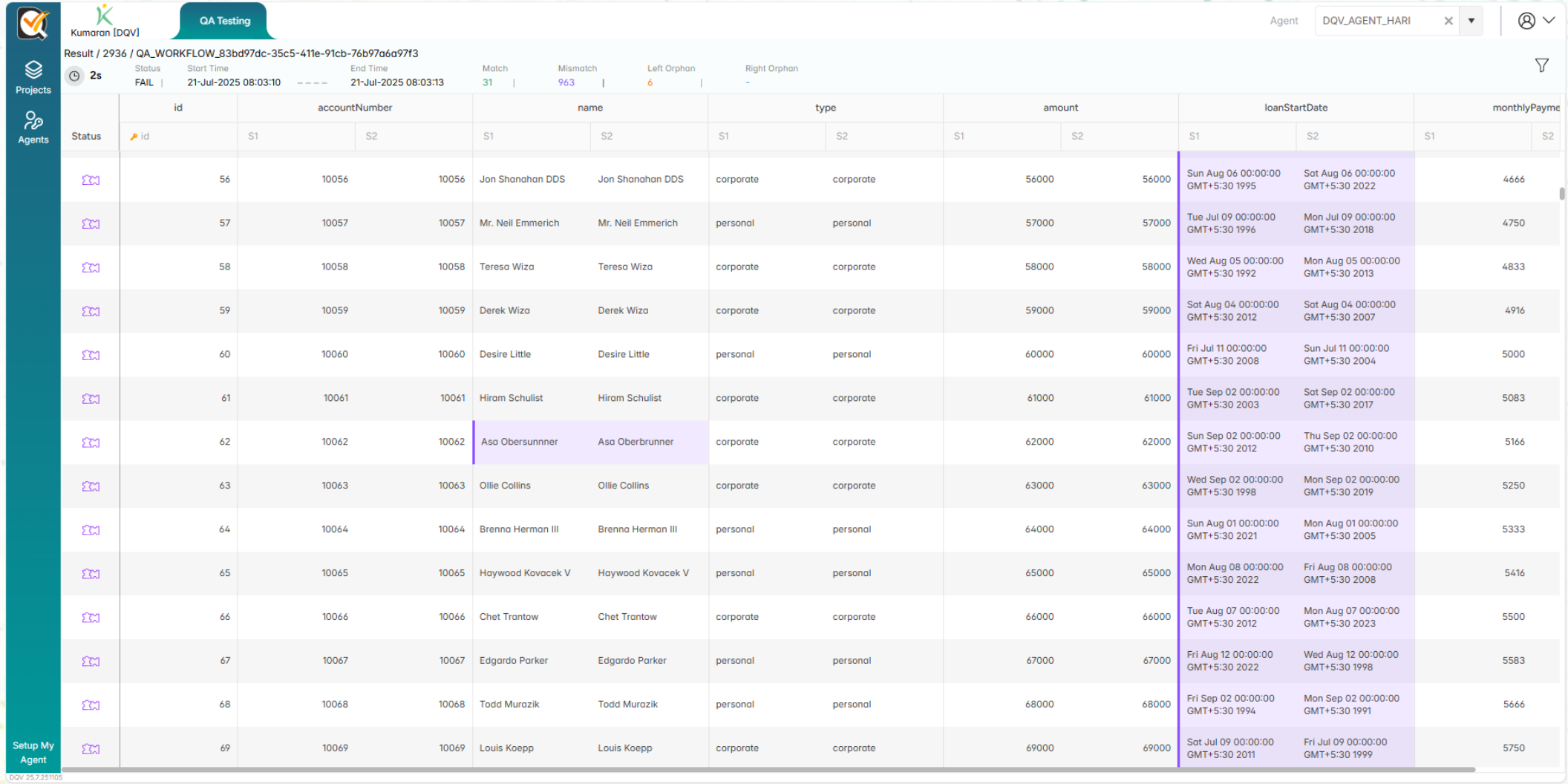
Run Results


Kumaran [DQV]
QA Testing
Agent
DQV_AGENT_HARI
Quick Action
Workflow (357)
Results
QA Testing Project - Transcend scenarios
Projects
Data Source
Data Spec
Workflows
Setup My Agent
Result / Adhoc-Run / 2929
Workflow
RUN SUMMARY
QA_WORKFLOW_5541f2db-84ca-42a6-8c2f-4f829aa6cd01
FAIL
1s
Status
FAIL | ADHOC |
Start Time
21-Jul-2025 07:38:03
End Time
21-Jul-2025 07:38:05
Log Files
Start
PASS
1,000 Read Count
0.240s Total Time
0.220s Process Time
1,000 Record Count
1,000 Waiting Count
0.16s Wait Time
Reader
Source 1
On Every Out Record
Start
PASS
994 Read Count
0.504s Total Time
0.501s Process Time
994 Record Count
994 Waiting Count
0.2s Wait Time
Reader
Source 2
On Every Out Record
FAIL
1,000 Source1 Consumed
994 Source2 Consumed
31 Match
963 Mismatch
6 Left Orphan
0 Right Orphan
1 Result Wait Time
Comparator
Comparator
PASS
1.609s Time
0.387s Process Time
0 Failure Record Count
0s Wait Time
1,000 Record Count
0 Waiting Count
Result Writer

Comparison Details



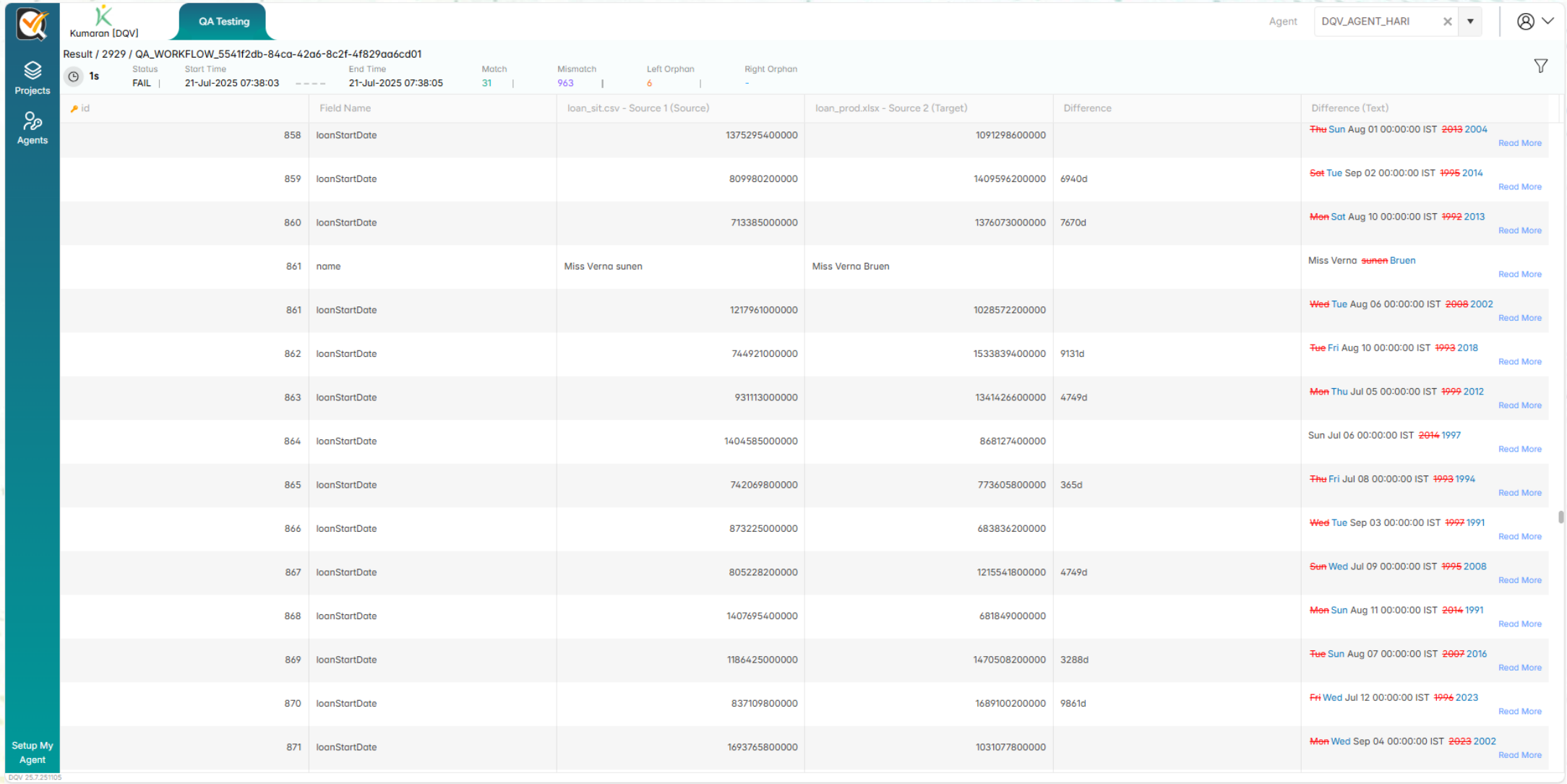
Comparison Results



Status	id	accountNumber		name		type		amount		loanStartDate		monthlyPayme	
	id	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
	56	10056	10056	Jon Shanahan DDS	Jon Shanahan DDS	corporate	corporate	56000	56000	Sun Aug 06 00:00:00 GMT+5:30 1995	Sat Aug 06 00:00:00 GMT+5:30 2022	4666	
	57	10057	10057	Mr. Neil Emmerich	Mr. Neil Emmerich	personal	personal	57000	57000	Tue Jul 09 00:00:00 GMT+5:30 1996	Mon Jul 09 00:00:00 GMT+5:30 2018	4750	
	58	10058	10058	Teresa Wiza	Teresa Wiza	corporate	corporate	58000	58000	Wed Aug 05 00:00:00 GMT+5:30 1992	Mon Aug 05 00:00:00 GMT+5:30 2013	4833	
	59	10059	10059	Derek Wiza	Derek Wiza	corporate	corporate	59000	59000	Sat Aug 04 00:00:00 GMT+5:30 2012	Sat Aug 04 00:00:00 GMT+5:30 2007	4916	
	60	10060	10060	Desire Little	Desire Little	personal	personal	60000	60000	Fri Jul 11 00:00:00 GMT+5:30 2008	Sun Jul 11 00:00:00 GMT+5:30 2004	5000	
	61	10061	10061	Hiram Schulist	Hiram Schulist	corporate	corporate	61000	61000	Tue Sep 02 00:00:00 GMT+5:30 2003	Sat Sep 02 00:00:00 GMT+5:30 2017	5083	
	62	10062	10062	Asa Obersunnner	Asa Oberbrunner	corporate	corporate	62000	62000	Sun Sep 02 00:00:00 GMT+5:30 2012	Thu Sep 02 00:00:00 GMT+5:30 2010	5166	
	63	10063	10063	Ollie Collins	Ollie Collins	corporate	corporate	63000	63000	Wed Sep 02 00:00:00 GMT+5:30 1998	Mon Sep 02 00:00:00 GMT+5:30 2019	5250	
	64	10064	10064	Brenna Herman III	Brenna Herman III	personal	personal	64000	64000	Sun Aug 01 00:00:00 GMT+5:30 2021	Mon Aug 01 00:00:00 GMT+5:30 2005	5333	
	65	10065	10065	Haywood Kovacek V	Haywood Kovacek V	personal	personal	65000	65000	Mon Aug 08 00:00:00 GMT+5:30 2022	Fri Aug 08 00:00:00 GMT+5:30 2008	5416	
	66	10066	10066	Chet Trantow	Chet Trantow	corporate	corporate	66000	66000	Tue Aug 07 00:00:00 GMT+5:30 2012	Mon Aug 07 00:00:00 GMT+5:30 2023	5500	
	67	10067	10067	Edgardo Parker	Edgardo Parker	personal	personal	67000	67000	Fri Aug 12 00:00:00 GMT+5:30 2022	Wed Aug 12 00:00:00 GMT+5:30 1998	5583	
	68	10068	10068	Todd Murazik	Todd Murazik	personal	personal	68000	68000	Fri Sep 02 00:00:00 GMT+5:30 1994	Mon Sep 02 00:00:00 GMT+5:30 1991	5666	
	69	10069	10069	Louis Koepp	Louis Koepp	corporate	corporate	69000	69000	Sat Jul 09 00:00:00 GMT+5:30 2011	Fri Jul 09 00:00:00 GMT+5:30 1999	5750	

KUMARAN SYSTEMS

Mismatch Results



KUMARAN SYSTEMS

Command Line Interface

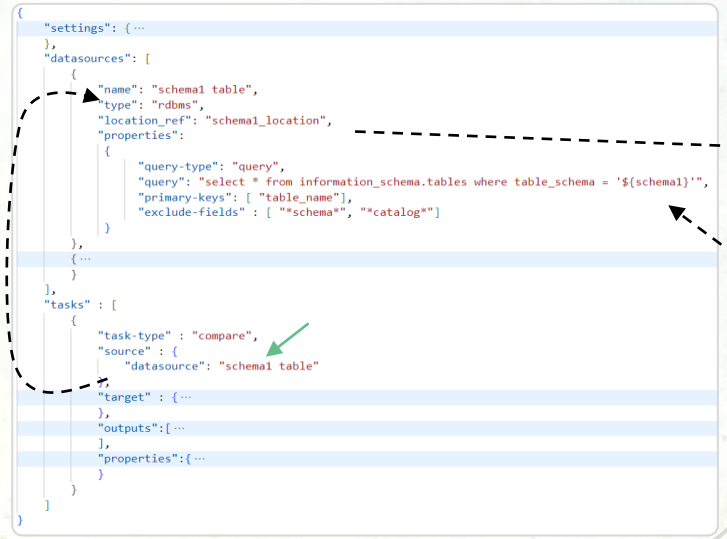
KUMARAN SYSTEMS

CLI Execution – Configuration Files

Number of configuration files can be used to extend or reuse existing configuration, or to selectively execute part of the requirements.

Task Configuration

- Contains various Tasks
  - Copy, Compare, Validate, Generate
  - Configure sub-tasks – such as filter, transform, etc.
- Each Task is associated with one or more Data Sources
  - Input, Output and Results
- A data source may be a table, schema or file, used for specific task
- Does not define the "location" such as Database Server or folder names.
- Supports parameterization
- Execute sequentially or parallelly with dependency



Location Configuration

- Defines various data source locations
  - Database server details, schema, credentials
  - File names or folder locations
- Supports encrypted password
- Supports getting parameters from environment variables
- Supports parameterization
- Parameters can be overridden from CLI arguments

Comparison – Summary Results

Source Container	Source Item	Source Field Count	Source Total Records	Source Applicable Records	Target Container	Target Item	Target Field Count	Target Total Records	Target Applicable Records	Matched Records	Mismatched Records	Left Only Records	Right Only Records	Time Taken (ms)	Status
dbo	LANGUAGESUBTOTALSSG	7	0	0	tenant	LANGUAGESUBTOTALSSG	7	0	0	0	0	0	0	1017	NONE
dbo	LANGUAGEREQUIREDCUSTOMERINFO	4	44	44	tenant	LANGUAGEREQUIREDCUSTOMERINFO	4	44	44	44	0	0	0	2032	MATCH
dbo	LANGUAGERMACONDITIONS	4	16	16	tenant	LANGUAGERMACONDITIONS	4	16	16	16	0	0	0	2037	MATCH
dbo	LANGUAGESUBCONTEXTS	3	154	154	tenant	LANGUAGESUBCONTEXTS	3	196	196	152	2	0	42	1437	MISMATCH
dbo	LANGUAGEGLOBALCONSTANTS	5	242	242	tenant	LANGUAGEGLOBALCONSTANTS	5	308	308	222	20	0	66	2722	MISMATCH
dbo	LANGUAGESTMTCONSTS	6	10	10	tenant	LANGUAGESTMTCONSTS	6	18	18	7	3	0	8	1561	MISMATCH
dbo	LANGUAGESTATES	3	10	10	tenant	LANGUAGESTATES	3	10	10	10	0	0	0	2074	MATCH
dbo	LANGUAGEVALUES	6	0	0	tenant	LANGUAGEVALUES	6	0	0	0	0	0	0	1068	NON
dbo	LANGUAGESUBTOTALS	5	92	92	tenant	LANGUAGESUBTOTALS	5	92	92	92	0	0	0	2020	MAT
dbo	ORGMACROFORMULAS	5	0	0	tenant	ORGMACROFORMULAS	5	0	0	0	0	0	0	1038	NON
dbo	NAMES	2	7	7	tenant	NAMES	2	7	7	7	0	0	0	1050	MAT
dbo	ORGMACROS	3	0	0	tenant	ORGMACROS	3	0	0	0	0	0	0	1018	NON
dbo	MODELPROPERTIES	3	21	21	tenant	MODELPROPERTIES	3	27	27	21	0	0	6	1404	MIS
dbo	MO	8	1	1	tenant	MODELS	8	1	1	1	0	0	0	1378	
dbo	LAN	5	680	680	tenant	LANGUAGETYPES	5	680	680			0	0	2034	MA
dbo	MO	3	1	1	tenant	MODELVERSION	3	1	1	0	1	0	0	1393	MIS
dbo	LAN D	4	100	100	tenant	LANGUAGEUPDATEPORTFOLIOCOND	4	100	100	100	0	0	0	2058	MATCH
dbo	LANGUAGEUPMAPPINGS	4	7756	7756	tenant	LANGUAGEUPMAPPINGS	4	7756	7756	7756	0	0	0	2088	MATCH
dbo	LITERALS	8	0	0	tenant	LITERALS	8	0	0	0	0	0	0	2032	NONE
dbo	MACROFORMULAS	3	447	447	tenant	MACROFORMULAS	3	447	447	447	0	0	0	2039	MATCH
dbo	PARAMS	6	49	49	tenant	PARAMS	6	49	49	0	49	0	0	1492	MISMATCH
dbo	MAPPEDMACROGROUPINGS	3	21	21	tenant	MAPPEDMACROGROUPINGS	3	21	21	21	0	0	0	2046	MATCH
dbo	MODELCONFIGURATIONS	3	1	1	tenant	MODELCONFIGURATIONS	3	1	1	1	0	0	0	2037	MATCH
dbo	MODELLANGMAPPING	2	4	4	tenant	MODELLANGMAPPING	2	4	4	4	0	0	0	2030	MATCH
dbo	QUICKPROJTYPEMAP	3	355	355	tenant	QUICKPROJTYPEMAP	3	355	355	355	0	0	0	1050	MATCH
dbo	PFPMAPPINGS	5	47	47	tenant	PFPMAPPINGS	5	47	47	47	0	0	0	2018	MATCH
dbo	PROJTYPEHISTORY	6	20	20	tenant	PROJTYPEHISTORY	6	20	20	20	0	0	0	2022	MATCH
dbo	PROJTYPEMAP	3	1745	1745	tenant	PROJTYPEMAP	3	1745	1745	1745	0	0	0	2024	MATCH
dbo	PROXYMAPPINGEXPR	5	75	75	tenant	PROXYMAPPINGEXPR	5	75	75	75	0	0	0	2019	MATCH
dbo	STATEMENTCONSTANTS	8	5	5	tenant	STATEMENTCONSTANTS	8	9	9	0	5	0	4	1386	MISMATCH
dbo	REPORTS	6	77	77	tenant	REPORTS	6	77	77	77	0	0	0	2022	MATCH
dbo	REPORTTYPES	2	11	11	tenant	REPORTTYPES	2	11	11	11	0	0	0	2023	MATCH

Source table information

Comparison Summary

Time taken in milli-second

Status of comparison

Target table information

KUMARAN SYSTEMS

Field-by-Field Comparison Results

	A	B	C	D	E	F	G	H	I	ST	SU	SV	SW
1						irce	Target	e	Target	Source	Target		
2	Source	SourceCategory	SourceSubCategory	UseCategory	UseSubCategory								
3	Target	SourceCategory	SourceSubCategory	UseCategory	UseSubCategory								
4	Data Comparison Status	SourceCategory	SourceSubCategory	UseCategor	UseSubCategory	Id	Id	BusinessDate	BusinessDate	AggMarketValue	AggMarketValue	Difference%	Alert
29	Mismatch					16948943	16925236	2023-04-13	2023-04-13	208127.39	225975.010	-8.575334607870335	FALSE
30	Mismatch					16959334	17090004	2023-04-13	2023-04-13	814284741	16247702282	-99.533424401754	FALSE
31	Mismatch					16994757	17013464	2023-04-13	2023-04-13	38853995	237535283.5	38.86464360411964	TRUE
32	Mismatch					16995641	16901797	2023-04-13	2023-04-13		8822257	3.4056972563219783	FALSE
33	Mismatch					16908542	17000096	2023-04-13	2023-04-13		014410.1	-58.90737684528848	FALSE
34	Mismatch					16887895	16879894	2023-04-13	2023-04-13		325449.1	7.525418551314818	FALSE
35	Target-Only						17036646		2023-04-13		1.39392E-06	0.0	FALSE
36	Source-Only					16936904		2023-04-13		37539336.57		100.0	TRUE
37	Mismatch					16996256	16960776	2023-04-13	2023-04-13	32243007.8	30393141.78	5.737262567651468	FALSE
38	Mismatch					16936617	16990149	2023-04-13	2023-04-13	25050080.19	23052924.08	7.972653559487726	FALSE
39	Mismatch					16996083	16974800	2023-04-13	2023-04-13	30619836.4	37864715.84	-23.66073855430528	FALSE
40	Mismatch					16959295	16897805	2023-04-13	2023-04-13	2038822207	2069864924	-1.5225808767610964	FALSE
41	Mismatch					16973996	169	2023-04-13	2023-04-13	74295246.24	76127449.42	-2.4661109129072116	FALSE
42	Target-Only								2023-04-13		0.000438304	0.0	FALSE
43	Mismatch							04-13	2023-04-13	5370580561	3922047819	26.971622997029325	TRUE
44	Mismatch							04-13	2023-04-13	257178312.4	240168469.2	6.614027077582023	FALSE
45	Mismatch							04-13	2023-04-13	1709579579	3018574428	-76.56823143448753	FALSE
46	Mismatch					16969551	17061454	2023-04-13	2023-04-13	335764359.6	345240911.7	-2.8223817698001965	FALSE
47	Mismatch					16905881	16888777	2023-04-13	2023-04-13	99123321.37	102880212.2	-3.7901179908846423	FALSE
48	Mismatch					16940417	17091510	2023-04-13	2023-04-13	131855221	83745984.1	36.534542937493995	TRUE
49	Mismatch					16952838	16904925	2023-04-13	2023-04-13	95.27	25372916.85	-46.05643424131524	FALSE
50	Mismatch					16892250	16882497	2023-04-13	2023-04-13	36.53	22693170.12	10.225740455721667	FALSE
51	Mismatch					16933710	16924988	2023-04-13	2023-04-13	433.8	698079869.3	-17.37417088607095	FALSE
52	Mismatch					16955788	16956897	2023-04-13	2023-04-	449.4	71413.95003	26.716890994363716	TRUE
53	Mismatch					16959963	17050299	2023-04-13	2023-04-13	265.6	340935343.4	-0.44370889353080206	FALSE
54	Mismatch					16887393	16982099	2023-04-13	2023-04-13	2276263887	2288902872	-0.5552513003699233	FALSE
55	Mismatch					16893847	17024131	2023-04-13	2023-04-13	4315300.714	4323780.309	-0.1965006887468311	FALSE

KUMARAN SYSTEMS

Comparison – Mismatch Report

Field Name having diff.

Source Value (Green)

Target Value (Brown)

Difference Value (Number fields only)

Key Fields (Gold)

	A	B	C	D	E	F	G	H	I
1	SRC_SYST	SRC_SUBSYST	UNIQUE_ID	CUSIP	LEGAL_ENTITY_I	Field Name	Rpt20230605.xlsx - Source	Rpt20230605.xlsx - Target	Difference
2							11579	DUMMYKMART	-
3							15583-	14508-	-
4							14536-	14508-	-
5							15506-	14508-	-
6							569.317532	552.2362617	17.08127031
7							569	552	17
8							2578.893	1719.262	859.631
9							2578	1719	859
10							1210237.13	1160438.161	49798.96956
11							1210237	1160438	49799
12							26237.80996	26212.43603	25.37393204
13							26237	26212	25
14							3590056.481	3646553.383	-56496.90248
15							3590056	3646553	-56497
16							1252356.13	1259104.767	-6748.637
17							1252356	1259104	-6748
18							1034208.455	1037918.861	-3710.406
19							1034208	1037918	-3710
20							971360.0925	1298816.866	-327456.773
21							971360	1298816	-327456
22							766168.3895	775949.9473	-9781.557818
23							766168	775949	-9781
24							3900066.898	3853982.58	46084.318
25							3900066	3853982	46084
26							5114982.88	6721750	-1606767.12

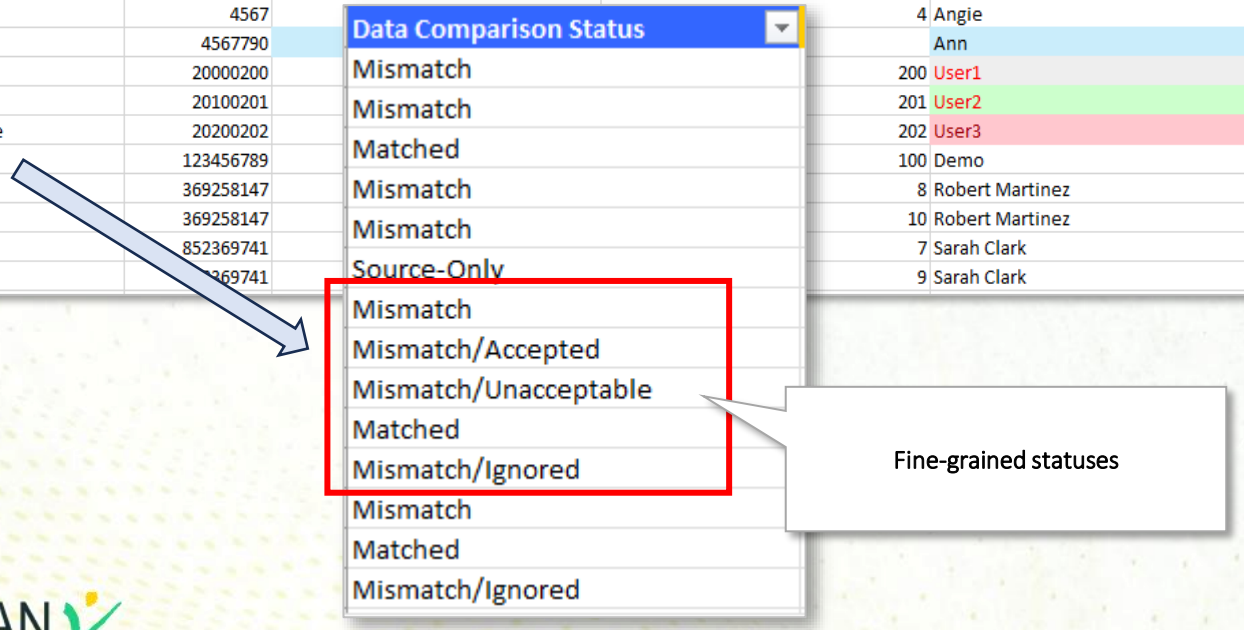
KUMARAN SYSTEMS

Comparison – Actions

DQV has advanced control of "Difference Actions" which is used to achieve finer controls over the **ignorable**, **legitimate** or **unacceptable** differences.

For the required fields, specify the condition and one of the actions → "`accept`", "`ignore`", or "`unacceptable`"

	A	B	C	D	E	F	G	H
1	K-DQV (1.0.0.505)		loan_data1.csv - Input1 (Source)	loan_data2.csv - Input2 (Target)	loan_data1.csv - Input1 (Source)	loan_data2.csv - Input2 (Target)	loan_data1.csv - Input1 (Source)	loan_data2.csv - Input2 (Target)
2	loan_data1.csv - Input1 (Source)	accountNumber						
3	loan_data2.csv - Input2 (Target)	accountNumber						
4	Data Comparison Status	accountNumb	id	id	name	name	type	type
5	Mismatch	1234	1	1	Dave	Dave	Education	Home
6	Mismatch	1234	3	3	Dave	Dave	Home	Education
7	Matched	1234	6	6	Dave	Dave	Home	Home
8	Mismatch	4567	2	2	Angie	Angie	Home	Home
9	Mismatch	4567		4	Angie	Angie	Car	Car
10	Source-Only	4567790			Ann		Personal	
11	Mismatch	20000200		200	User1	UserA	Other	Home
12	Mismatch/Accepted	20100201		201	User2	UserB	Home	Home
13	Mismatch/Unacceptable	20200202		202	User3	UserC	Education	Home
14	Mismatch	123456789		100	Demo	Demo	Other	Other
15	Mismatch/Ignored	369258147		8	Robert Martinez	Robert Martinez	Car	Car
16	Mismatch	369258147		10	Robert Martinez	Robert Martinez	Education	Education
17	Matched	852369741		7	Sarah Clark	Sarah Clark	Personal	Personal
18	Mismatch	369741		9	Sarah Clark	Sarah Clark	Home	Home



KUMARAN SYSTEMS

Comparison – Mismatch Summary

	A	B	C	D	E	F	G	H	
1	Comparison Nar	Source Container	Source Item	Target Container	Target Item	Field Name	Total count (Sourc	Failed coun	
2	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	SourceSourceSystem	138	16	
3	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat		138	13	9.42%
4	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat		138	84	60.87%
5	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat		138	137	99.28%
6	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat		138	118	85.51%
7	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat		138	111	80.43%
8	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	UseRollup20Name	138	106	76.81%
9	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	UnitOfPrice	138	10	7.25%
10	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	UseTxnType	138	1	0.72%
11	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	SourceIsin	138	125	90.58%
12	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	SourceHaircut	138	40	28.99%
13	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	SourceProductType	138	43	31.16%
14	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	SourceCptyName	138	52	37.68%
15	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	UseDirtyMarketValue	138	121	87.68%
16	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	SourceAssetExchangedDPACategory	138	16	11.59%
17	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	SourceSecurityDescription	138	125	90.58%
18	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	SecurityKeyType	138	127	92.03%
19	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	SourceUnitOfPrice	138	122	88.41%
20	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	SourceAccountUnit	138	119	86.23%
21	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	SourceAuRollup9Name	138	117	84.78%
22	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	UseUnitOfPrice	138	122	88.41%
23	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	RemainingSourceQty	138	79	57.25%
24	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	UseHQLA	138	49	35.51%
25	Comparison-1	SIT1	_20230413-01-01.dat	SIT2	20230413-01-01.dat	AllocationId	138	137	99.28%

Number of records having mismatch for this field

Total number of records compared

Source & Target table details

Mismatched field name

% of failures for this field

KUMARAN SYSTEMS

Transfer / Migration – Results

A	B	C	D	E	F	G	H	I	J
Source Contain	Source Item	Source Total Record	Source Applicable Record	Target Contain	Target Item	Target Processed Record	Target Failed Record	Time Taken (m	Status
[redacted]	[redacted]	0	0	[redacted]	[redacted]	0	0	0	SKIPPED
[redacted]	[redacted]	482174	482174	[redacted]	[redacted]	482174	0	148516	SUCCESS
[redacted]	[redacted]	57658	57658	[redacted]	[redacted]	57658	0	29615	SUCCESS
[redacted]	[redacted]	1400525	1400525	[redacted]	[redacted]	1400525	0	570127	SUCCESS
[redacted]	[redacted]	147207	147207	[redacted]	[redacted]	147207	0	870549	SUCCESS
[redacted]	[redacted]	104600	104600	[redacted]	[redacted]	104600	0	73076	SUCCESS
[redacted]	[redacted]	583532	583532	[redacted]	[redacted]	583532	0	1090238	SUCCESS
[redacted]	[redacted]	452338	452338	[redacted]	[redacted]	452338	0	740113	SUCCESS
[redacted]	[redacted]	67853	67853	[redacted]	[redacted]	67853	0	238999	SUCCESS
[redacted]	[redacted]	447	447	[redacted]	[redacted]	447	0	2051	SUC
[redacted]	[redacted]	139140	139140	[redacted]	[redacted]	139140	[redacted]	538567	SUC
[redacted]	[redacted]	739087	739087	[redacted]	[redacted]	739087	[redacted]	253046	SUC
[redacted]	[redacted]	1147	1147	[redacted]	[redacted]	1147	[redacted]	3090	SUC
[redacted]	[redacted]	876	876	[redacted]	[redacted]	876	[redacted]	2110	[redacted]
[redacted]	[redacted]	5139843	5139843	[redacted]	[redacted]	5139843	0	2429185	SUC
[redacted]	[redacted]	86370	86370	[redacted]	[redacted]	86370	0	27652	SUC
[redacted]	[redacted]	684895	684895	[redacted]	[redacted]	684895	0	1153723	SUCCESS
[redacted]	[redacted]	2000054	2000054	[redacted]	[redacted]	2000054	0	1244377	SUCCESS
[redacted]	[redacted]	3145	3145	[redacted]	[redacted]	3145	0	308	SUCCESS
[redacted]	[redacted]	4372820	4372820	[redacted]	[redacted]	4372820	0	1761	SUCCESS
[redacted]	[redacted]	0	0	[redacted]	[redacted]	0	[redacted]	0	SKIPPED
[redacted]	[redacted]	4211935	4211935	[redacted]	[redacted]	4211935	[redacted]	764	SUCCESS
[redacted]	[redacted]	0	0	[redacted]	[redacted]	0	[redacted]	0	SKIPPED
[redacted]	[redacted]	30046	30046	[redacted]	[redacted]	30046	0	19731	SUCCESS
[redacted]	[redacted]	0	0	[redacted]	[redacted]	0	0	0	SKIPPED
[redacted]	[redacted]	0	0	[redacted]	[redacted]	0	0	2018	NONE
[redacted]	[redacted]	151756	151756	[redacted]	[redacted]	151756	0	488977	SUCCESS
[redacted]	[redacted]	0	0	[redacted]	[redacted]	0	0	3003	NONE
[redacted]	[redacted]	0	0	[redacted]	[redacted]	0	0	0	SKIPPED
[redacted]	[redacted]	153916	153916	[redacted]	[redacted]	153916	0	181620	SUCCESS

Source table information

Failed Records

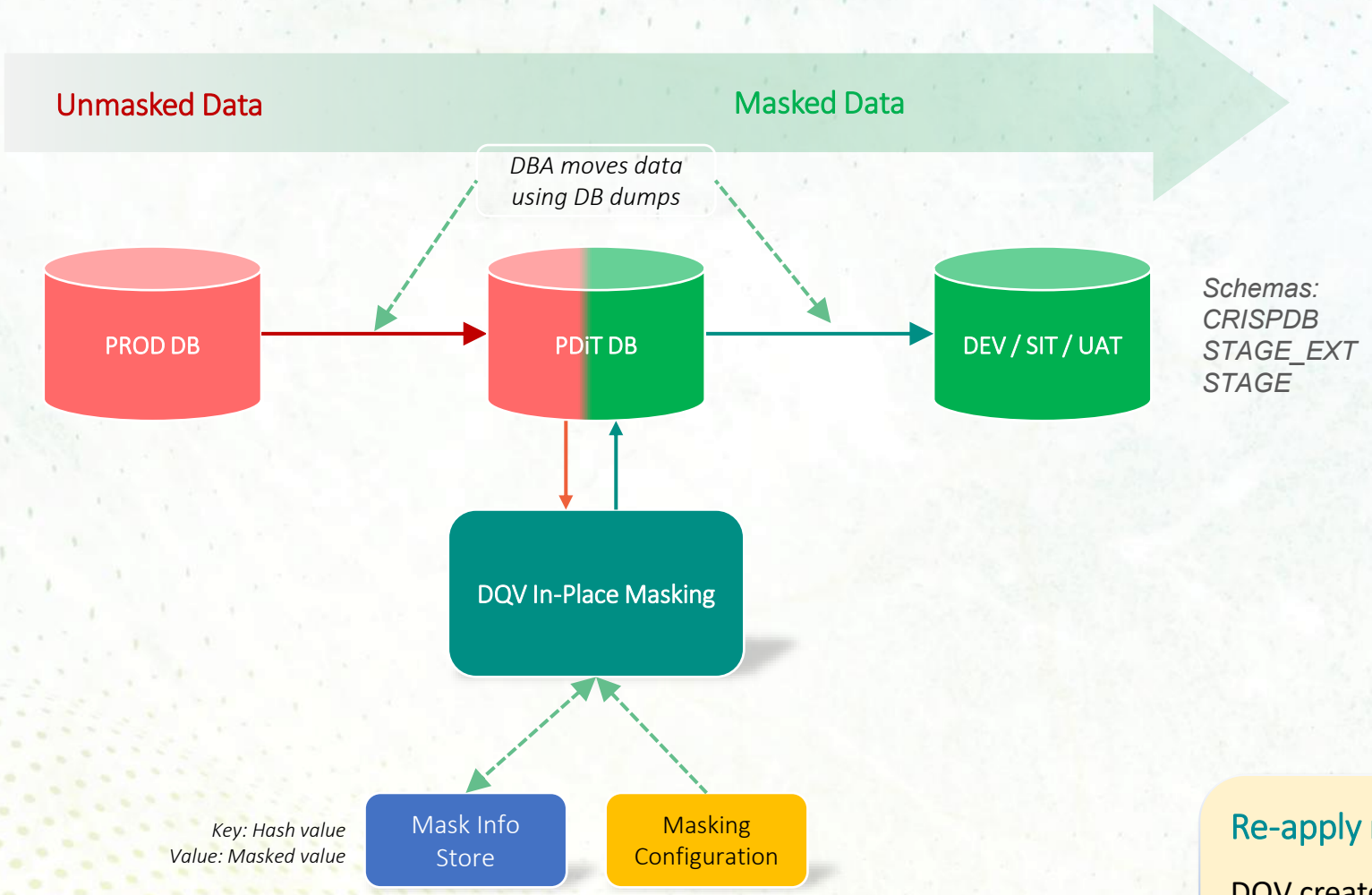
Time taken in milli-second

Status of migration

Target table information

KUMARAN SYSTEMS

Incoming Feed Masking - Initial



Activities

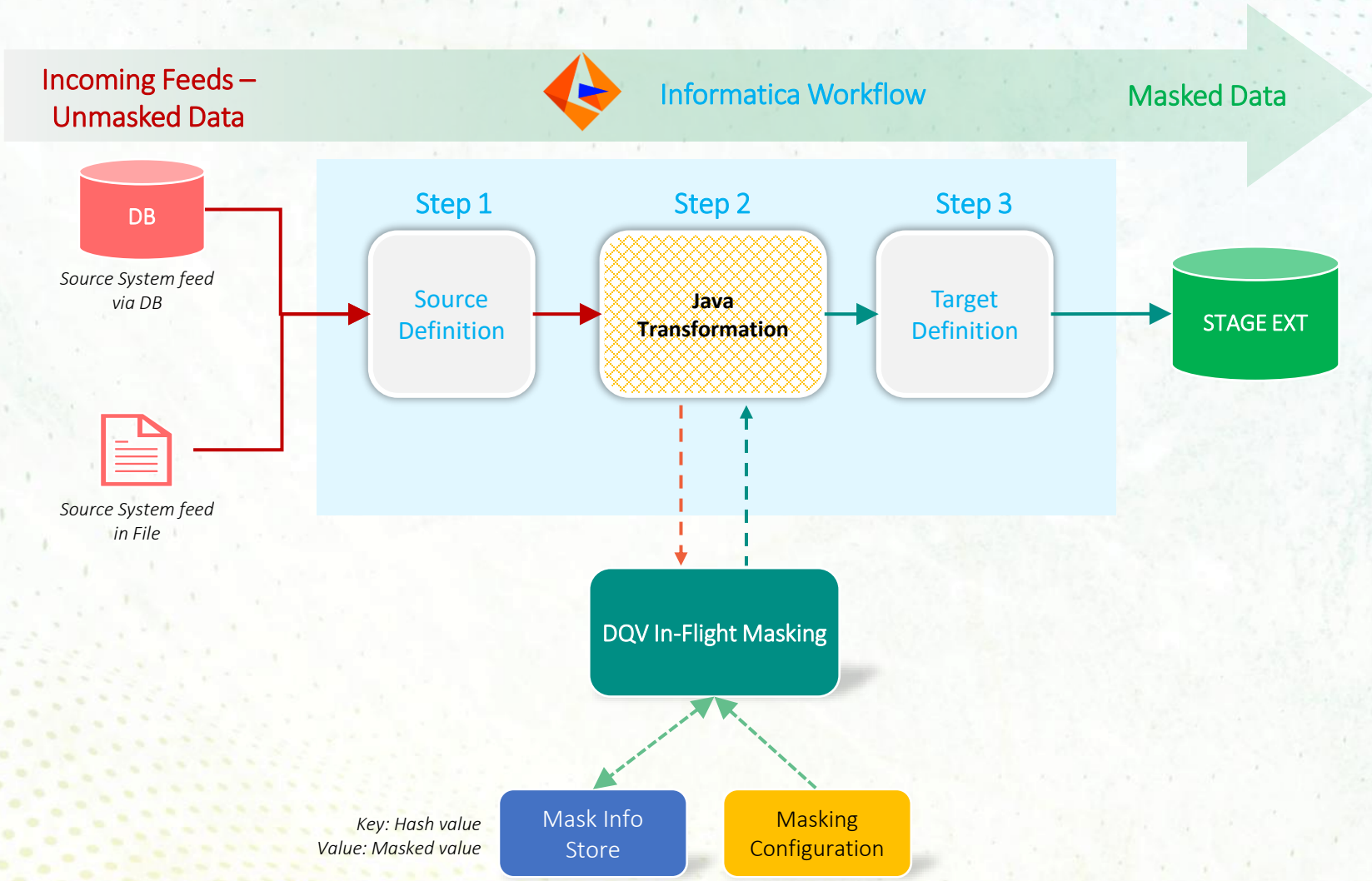
- ✓ **Step 1:** DBA move the production data to PDiT environment using traditional approaches such as dump file export / import.
- ✓ **Step 2:** DQV is executed to do in-place masking.
- ✓ **Step 3:** DBA uses the masked data for refreshing other lower environments.

Re-apply masking

DQV creates a Mask Info file which can be used for incremental / sub-sequent masking purpose for the same environment.

KUMARAN SYSTEMS

Incoming Feed Masking - Incremental

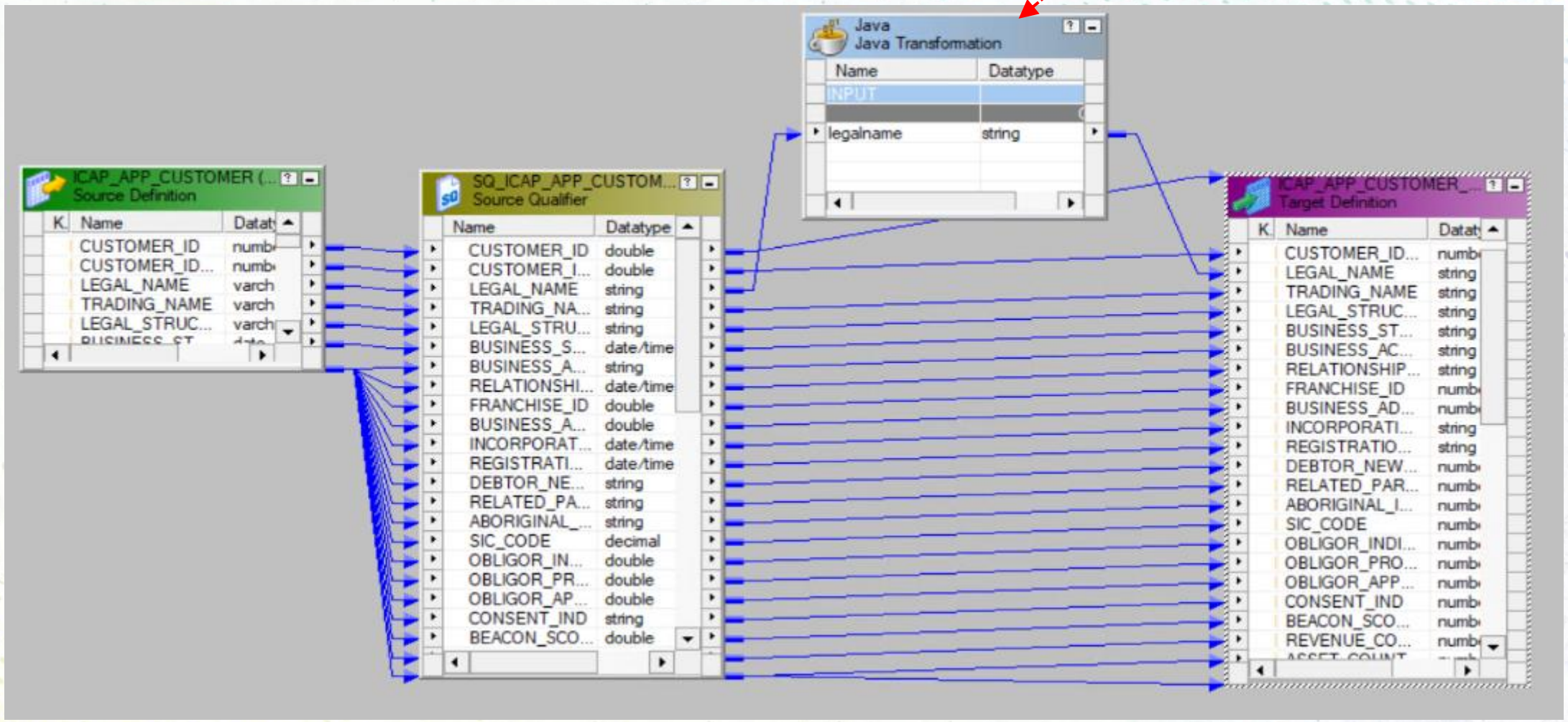


Activities

- ✓ **Step 1:** Informatica workflow reads the source system file or database
- ✓ **Step 2:** Informatica Transformation used DQV library which applies masked values **in-flight**.
- ✓ **Step 3:** Informatica writes the data to the target tables.

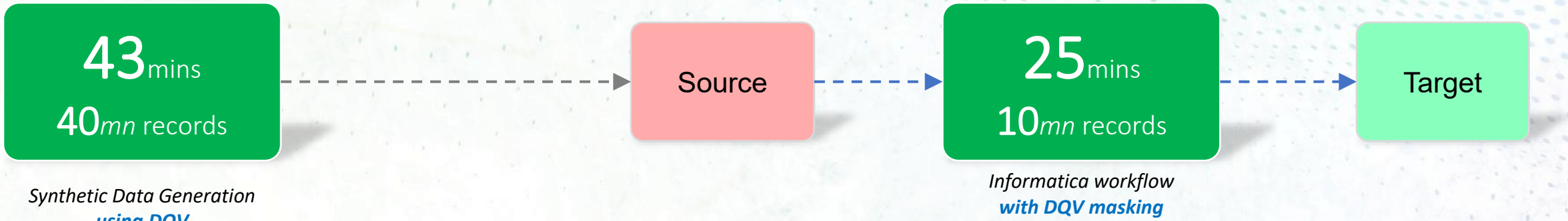
KUMARAN SYSTEMS

Informatica Workflow Change



KUMARAN SYSTEMS

Statistics



Synthetic Data Generation
using DQV

Informatica workflow
with DQV masking

Address	*7 **(5)** columns*	5*mn*
Customer	*47 **(1)** columns*	5*mn*
Facility	*23 columns*	10*mn*
Instrument	*3 columns*	20*mn*
Total		**40*mn***

Informatica workflow
without DQV masking

KUMARAN SYSTEMS

Leading enterprise Data Migration – Case Study

Problem statement

- There are about **23 millions** of records across 15 tables that are required to migrated from IBM DB2 Database to MS SQL Database Server for the Tax Estimator and Business Cash Flow Planning applications.
- SSMA was tried but it requires special privileges on the MS SQL server and requires additional software installation. Moreover, SSMA does not offer data comparison.

Portfolio

- Portfolio: **Front Line Technology**
- Owner: **Warren Chan / Montaz Chowdhury**
- Implementation Period: **May 2025** (PROD)

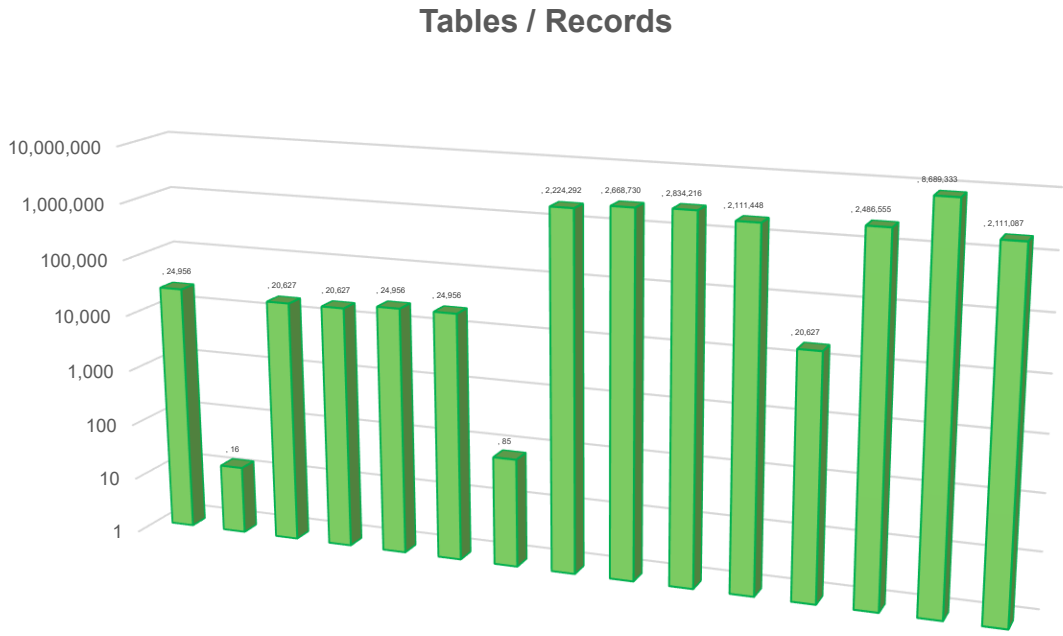
Approach

- DQV is used to migrate data in all DIT / SIT / PROD environments. Once migrated, the data is compared with DQV.
- Utilize DQV to compare the database structures of DB2 and SQL Server tables, enabling quick identification of any differences in table structures, or the definitions of views and procedures.

Advantages & Benefits

- Quickly migrated the data from source to target database without any errors.
- Field level highlight of differences are provided by DQV within a short period of time.

Data Migration



PROD Table Statistics

15 Tables	**15** Fully matched
23 million Records Compared	

Comparison Statistics

0 Mismatched Records	**0 %** Mismatch

KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL

Leading Enterprise – Case Study

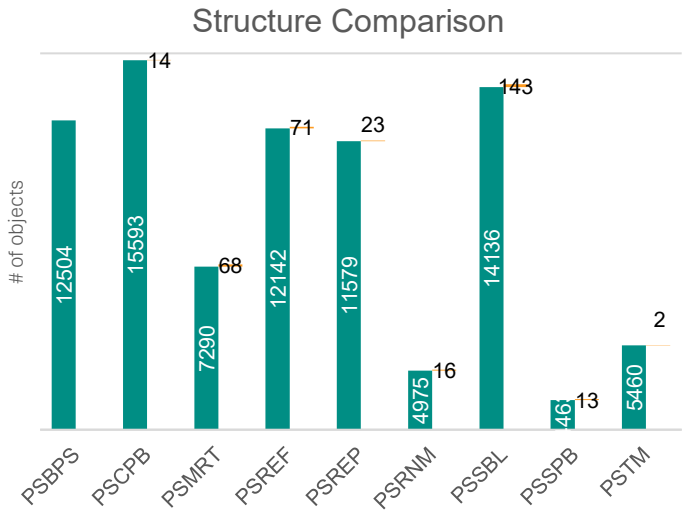
DQV will be utilized to ensure that the changes made to the applications do not disrupt the existing production functionality..

Problem statement

- There are about 800 batches being executed in Prime Services Group.
- **Data/Report Validation**: Whenever there is a change to the batch logic, regression is required to ensure all ETL jobs are working as intended, but this could not be done due to time crunch / lack of tools.
- The ETL batch configurations are stored in database tables but are not consistently promoted across different environments, making data comparison a tedious process.
- **Schema Validation:** The data model and procedure changes done by the development team is not properly promoted to other environments and caused errors.

Approach

- DQV would be used to compare outputs of different ETL jobs after executing them in both SIT-A and SIT-B environments. SIT-B is the production mirror environment.
  - ✓ Data Load jobs – Compare the data in the tables that are loaded from files
  - ✓ Data Extract jobs – Compare the data in the files that are exported from tables
  - ✓ File Delivery jobs – Compare the file's meta-data properties (size, check-sum, last modified, etc) to ensure that the files are exactly same
- Utilize DQV to periodically compare the database structures of SIT-A and SIT-B, enabling quick identification of any differences in table structures, or the definitions of views and procedures.



Data Comparison - PSTM

79 Tables	32 Fully matched	39 Mismatch

200 million Records Compared

187 k Mismatched Records

0.09 % Mismatch

01h 45m Execution Time

Advantages & Benefits

- Field level highlight of differences are provided by DQV within a short period of time.
- Integration with DevOps pipelines to reduce manual intervention.
- Kumaran agreed to automate in low budget using Fixed bid mode.
- Assisted by the DQV product team in implementing enhancements to address unique testing scenarios.

Future Scope

- Comparison of JSON structures.
- Integration with API and MQs and perform validations.
- Implementation / Integration with attestation process.

KUMARAN SYSTEMS
ENGAGE | EMERGE | EXCEL