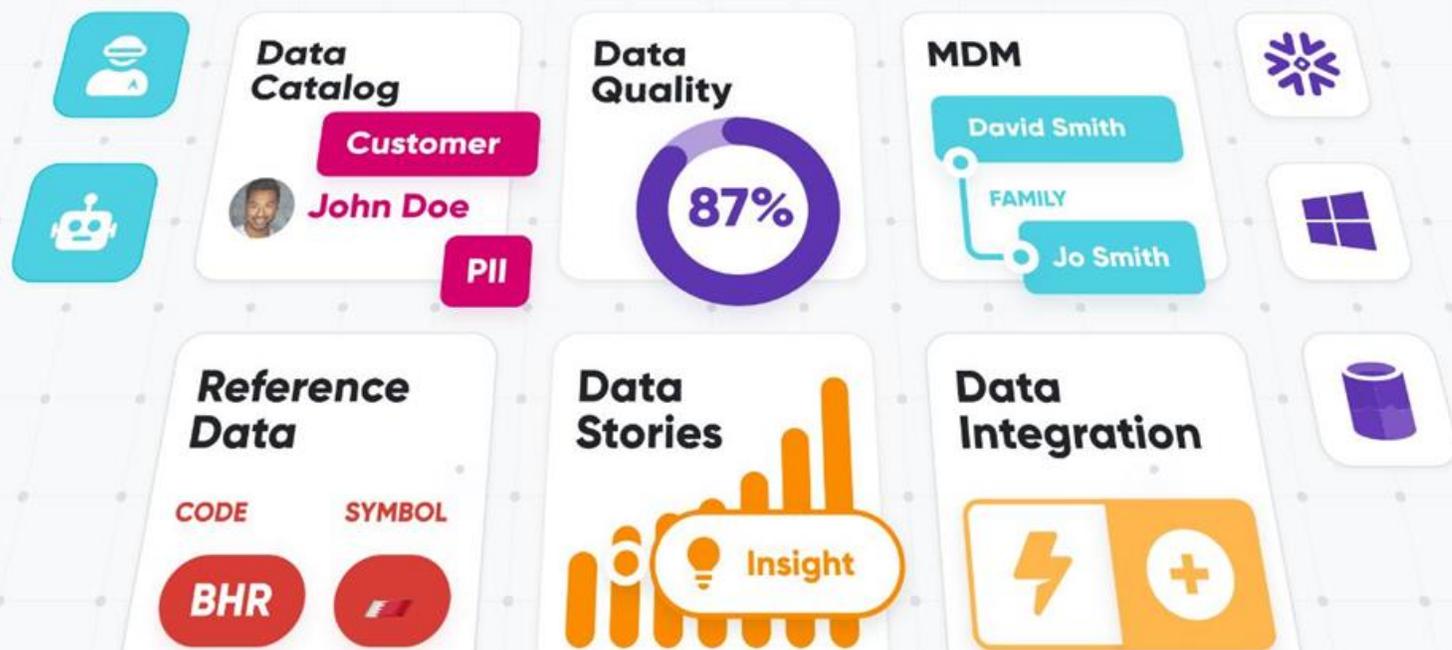


Data Quality Fabric

Make your enterprise data instantly available to humans and machines as high quality data products, with governance and compliance ensured automatically.



TODAY'S

Agenda

- 1. Company Introduction**
- 2. From Integrated to Self-driving**
- 3. Solution Proposal**
- 4. Demo**

ATACCAMA

Company introduction

Ataccama Fast Facts



Offices

Toronto, London, Sydney,
Prague
New York, Valencia, Paris,
Sofia



350+ Global Customers

6 US distinct countries
+ 15 states

60,000+ DQA Downloads



Gartner Recognition

Both DQ and MDM Gartner
Magic Quadrant

Top positions in Peer Insights

Company Highlights



New Offices

Sydney, Paris



Over 200% growth

In 3 Years



New Partners

Technology, Platform, SI



Professional Services

Capacity doubled



Cloud Offering

Managed Services, PaaS
Marketplaces



SELECTED CUSTOMERS

Financial Services

- Allianz
- Argo Group
- Aviva
- Bank of Montreal
- Beazley
- Fifth Third Bank
- First Data Corporation
- GE Capital Corporation
- HSBC
- Industrial Alliance
- KBC Group
- Raiffeisenbank Group
- RSA
- Scotiabank
- Societe Generale
- TD Bank



SELECTED CUSTOMERS

Pharma Healthcare Public

- BCBS Association
- BCBS Michigan
- BCBS Minnesota
- Celgene Corporation
- City of Winnipeg
- Daiichi Sankyo
- Edwards Lifesciences
- Glaxo Smith Kline
- Hospital Corporation of America
- Otsuka Pharmaceutical
- MedAmerica Inc.
- Memorial Health System
- Novozymes



SELECTED CUSTOMERS

Telco
Retail
Other



- American Airlines
- Auchan
- AVAST
- Canadian Tire
- CD Cargo
- Darden Restaurants
- GAF
- Helmerich and Payne

- Husky Energy
- Intersnack
- Telefonica O2
- RNLI
- SOCAN
- T-Mobile
- Toronto Public Library
- Vodafone

Ataccama partner ecosystem



**Technology, OEM and
cloud partners**



**Consulting Partners
Value-Added Resellers**

OEM distribution
Software compatibility & joined value proposition
Customized Cloud-based solutions
Data enrichment services

Integration & consulting services
Reselling Ataccama SW, custom-tailored solutions
Alliance Partner Program members or
Opportunity-Based Global and region-specific

TIBCO™

collibra™

ADASTRA

UNISYS

TATA
TATA CONSULTANCY SERVICES

LOGQATE

cloudera®
Ask Bigger Questions

TERADATA.
Raising Intelligence

wipro

Capgemini

KPMG

Deloitte.

aws

Azure

Google Cloud

accenture

VIQTOR DAVIS.
DATA CRAFTSMANSHIP

Cognizant

Ataccama Service Offering

Ataccama is more than software. Our expanding professional services team ensures your success and maximum value out of your Ataccama platform investment. Our services cover the entire solution lifecycle.



Enablement

- Training
- Installation
- Knowledge base
- Starter pack



Strategic

- Architecture
- Governance
- Advanced configuration
- Trouble shooting



Rollout

- Configuration
- Integration
- Consulting support



Improvement

- Tuning
- Optimization
- Health-check
- Upgrade



Support

- Helpdesk
- Operations
- Online documentation
- Community

Enterprise Data Quality Fabric

Data Catalog

Customer

 **John Doe**

PII

Data Quality

87%

MDM

David Smith

FAMILY

Jo Smith

Reference Data

CODE SYMBOL

BHR 

Data Stories

Insight

Data Integration

Make high quality data instantly available to

Data engineers

Data stewards

Value creators

Machines & algorithms

With governance and compliance ensured automatically.

Selected Use Cases

Data Governance



Data Profiling



Data Catalog



Metadata Repositories

Data Quality



DQ Monitoring & Reporting



Data Visualization



DQ Firewall



Data Cleansing & Enrichment



Manual Issue Resolution



Data Integration

Multi-domain MDM



Reference Data Management



Multi-Domain Mastering



Master Domain Monitoring



Master Data Authoring



Householding & Aggregation



Consent Management

EPIM



Product Information Management



Media Asset Management

Platform Differentiators



Unified Platform with Complementary Modules

Active metadata usage – Extendibility – Embedded Governance



Self-Service Solution

AI & Machine Learning – Self-Service – Collaboration – UX/UI



Enterprise-Proven Usage

Scalability – Performance – High Availability – Auditing – Identity Management



Robust Data Processing Engine

Any Data / Any Domain – Integration – Performance – Scalability

ATACCAMA ONE

Security

ENCRYPTION

Transit -> TLS 1.2

Rest -> native to Azure/AWS RDBMS

AUTHENTICATION

AD via SAML (incl. Azure/AWS)

AUTHORIZATION

Keycloak (part of Ataccama ONE platform) manages mapping from AD groups/roles/claims

LOGGING

Full audit log

OWASP

Addressed during the development process

QA process and penetration tests prior to every major release

INFRASTRUCTURE

Existing Azure resources used to

protect the platform interfaces (including WAF and DDoS protection capabilities)

Deployment options



Platform as a Service

Focus on your data initiatives while we take care of the infrastructure, monitoring, maintenance, and security.



Marketplaces

Deploy the Platform from a cloud marketplace of your choice with a just few clicks.



On Prem & Hybrid

Deploy on your own infrastructure, private & public clouds and their combinations.

Built for cloud data management, deploy natively in the cloud, big data platforms, or hybrid environments



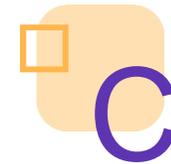
Platform as a Service

Focus on your data initiatives and bring value to your organization while we take care of the infrastructure, maintenance, and security



Marketplaces

Deploy the platform from a cloud marketplace of your choice.



Private Cloud

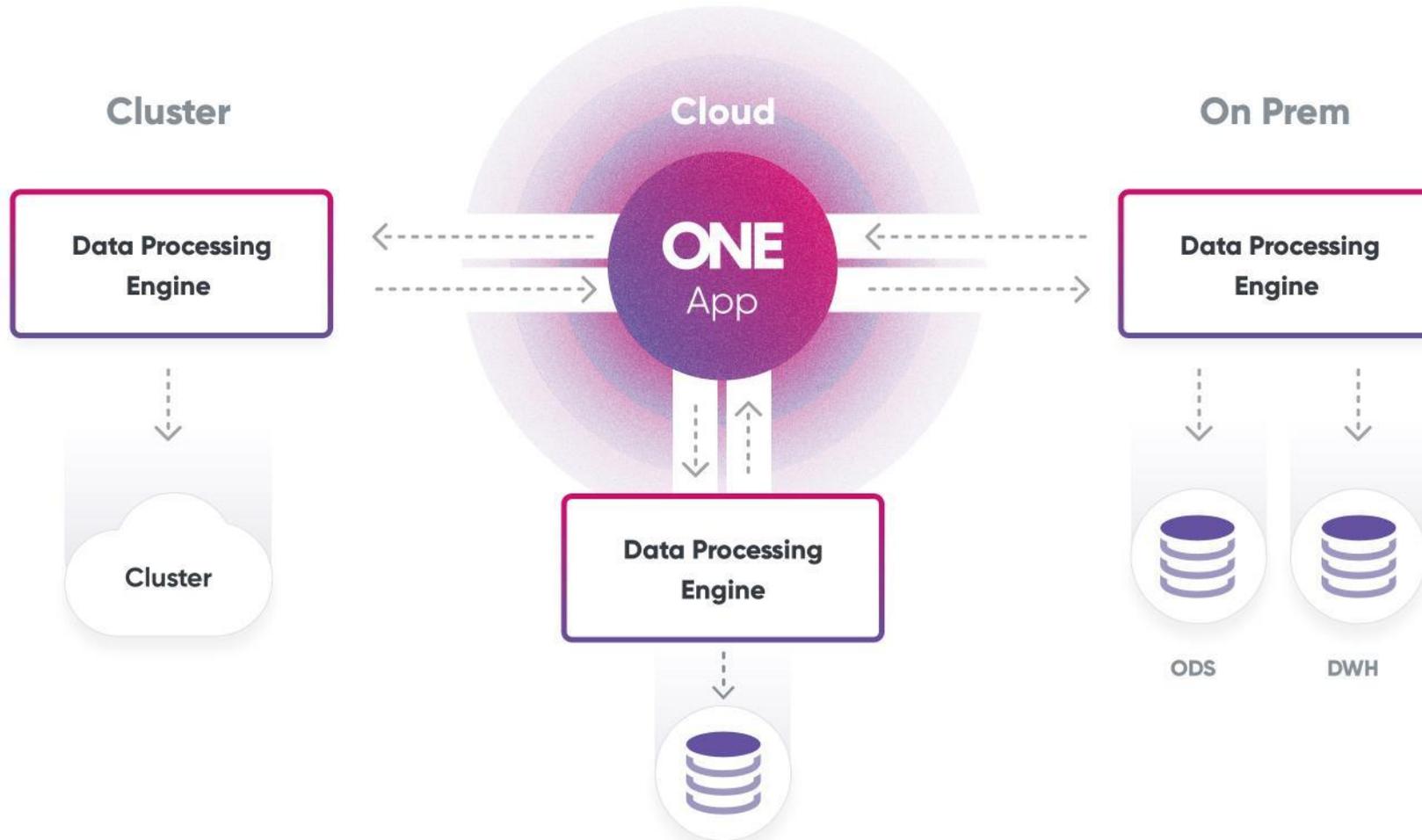
Deploy our containerized solution into your private cloud with Docker & Kubernetes.

Supporting all major cloud platforms



Hybrid deployment

ATACCAMA ONE



Data Quality Management

**Data
Quality**



87%



Automated DQ

Smart metadata-based configuration

AI-powered anomaly detection

Automated rule suggestions



Extensive capabilities

300+ built-in data transformation functions

DQ monitoring and reporting

Data enrichment & validation

Data quality firewall

Issue tracking and resolution



Powerful backend

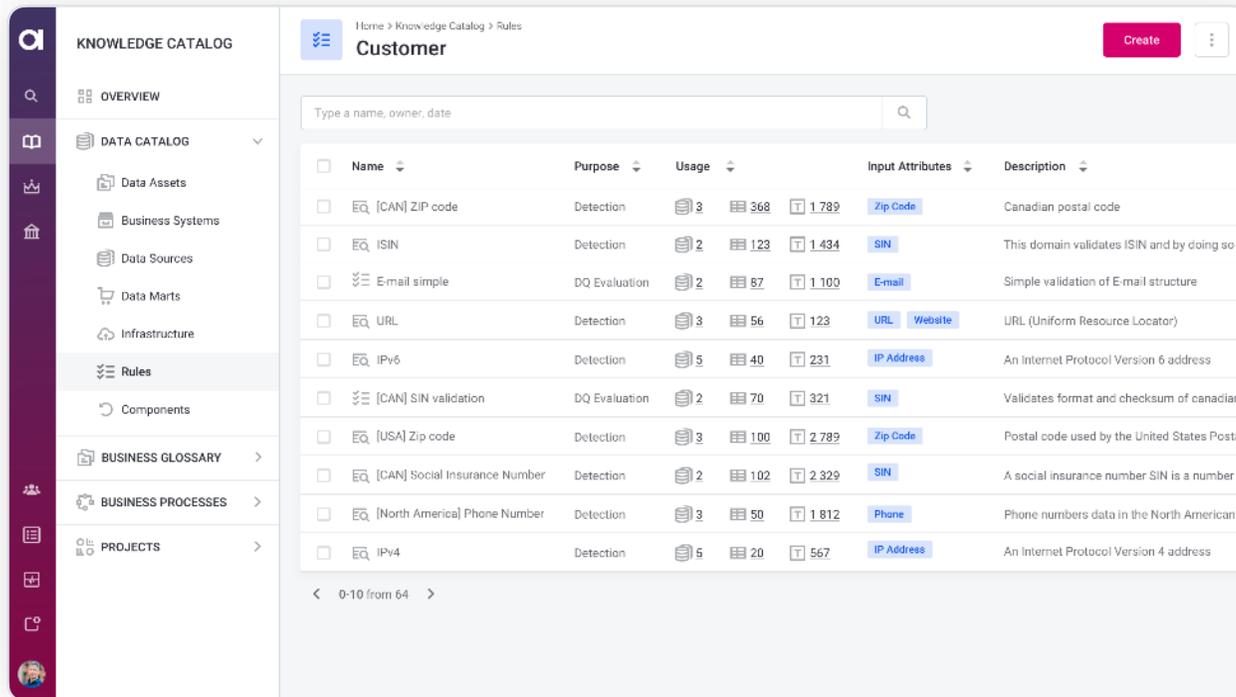
Super-fast data processing on any platform

Processing data without moving it

Extensive data ingestion and providing options

Multi-environment deployment

Rule management



The screenshot displays a web interface for rule management. On the left is a navigation sidebar with sections: KNOWLEDGE CATALOG, DATA CATALOG, BUSINESS GLOSSARY, BUSINESS PROCESSES, and PROJECTS. The 'Rules' section is selected. The main content area shows a table of rules for the 'Customer' project. The table has columns for Name, Purpose, Usage, Input Attributes, and Description. A search bar is at the top of the table. A 'Create' button is in the top right corner.

Name	Purpose	Usage	Input Attributes	Description
EQ [CAN] ZIP code	Detection	3 368 1,789	Zip Code	Canadian postal code
EQ ISIN	Detection	2 123 1,434	SIN	This domain validates ISIN and by doing so te
EQ E-mail simple	DQ Evaluation	2 87 1,100	E-mail	Simple validation of E-mail structure
EQ URL	Detection	3 56 123	URL Website	URL (Uniform Resource Locator)
EQ IPv6	Detection	5 40 231	IP Address	An Internet Protocol Version 6 address
EQ [CAN] SIN validation	DQ Evaluation	2 70 321	SIN	Validates format and checksum of canadian S
EQ [USA] Zip code	Detection	3 100 2,789	Zip Code	Postal code used by the United States Postal
EQ [CAN] Social Insurance Number	Detection	2 102 2,329	SIN	A social insurance number SIN is a number is
EQ [North America] Phone Number	Detection	3 50 1,812	Phone	Phone numbers data in the North American fo
EQ IPv4	Detection	5 20 567	IP Address	An Internet Protocol Version 4 address

- > Easy-to-use **collaborative rule management**
- > Full rule **history** and rule **version comparison** available
- > **Multiple approaches** to the validation rule implementation
- > Central rule library, shared in various projects

Advanced rules configuration

The screenshot displays the Ataccama ONE IDE interface for configuring a rule component named 'reader_test.comp'. The main workspace shows a data pipeline with the following components and connections:

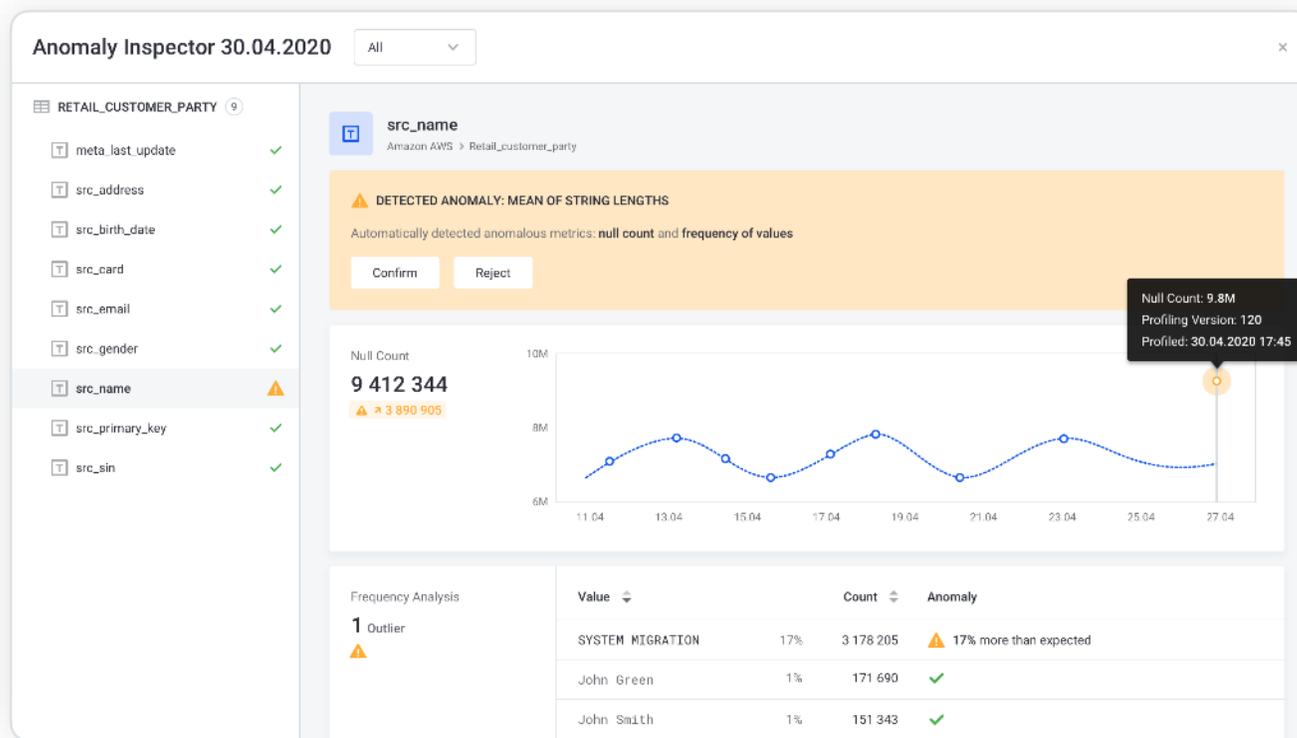
- Inputs: 'person' (Catalog Item Reader), 'retail_customer_party' (Catalog Item Reader), and 'address' (Catalog Item Reader).
- Join: A 'Join' component receives input from 'person' and 'retail_customer_party'.
- Filter: A 'Filter' component receives input from the 'Join' component.
- Group Aggregator: A 'Group Aggregator' component receives input from the 'Filter' component.
- Join 2: A 'Join 2' component receives input from the 'Group Aggregator' and another source.
- Outputs: 'new_person' (Integration Output) and 'out_results' (Text File Writer).

The left sidebar shows a project tree with various components, including 'rule', 'component', 'profilingConfiguration', 'loadMonitoringProject', 'lookupItem', 'person', 'role', 'configuration', 'navigationNode', 'rdmConfiguration', 'executorJob', 'baseJob', 'catalogItemRelationshipSuggestion', 'mmd', 'mmdStatistics', 'connection', 'location', 'termDetectionRules', 'termMetadataRules', 'termValidationRules', 'termStandardizationRules', 'dqEvalTermAggr', and 'catalogItem' (with sub-items like 'tableCatalogItem', 'retail_customer_party(draft)', 'fileCatalogItem', 'rdmCatalogItem', 'dqDimensionResult', 'ruleInputGroup', 'ioAttribute', 'ruleStep', and 'ruleTag').

The bottom console window shows the following log output:

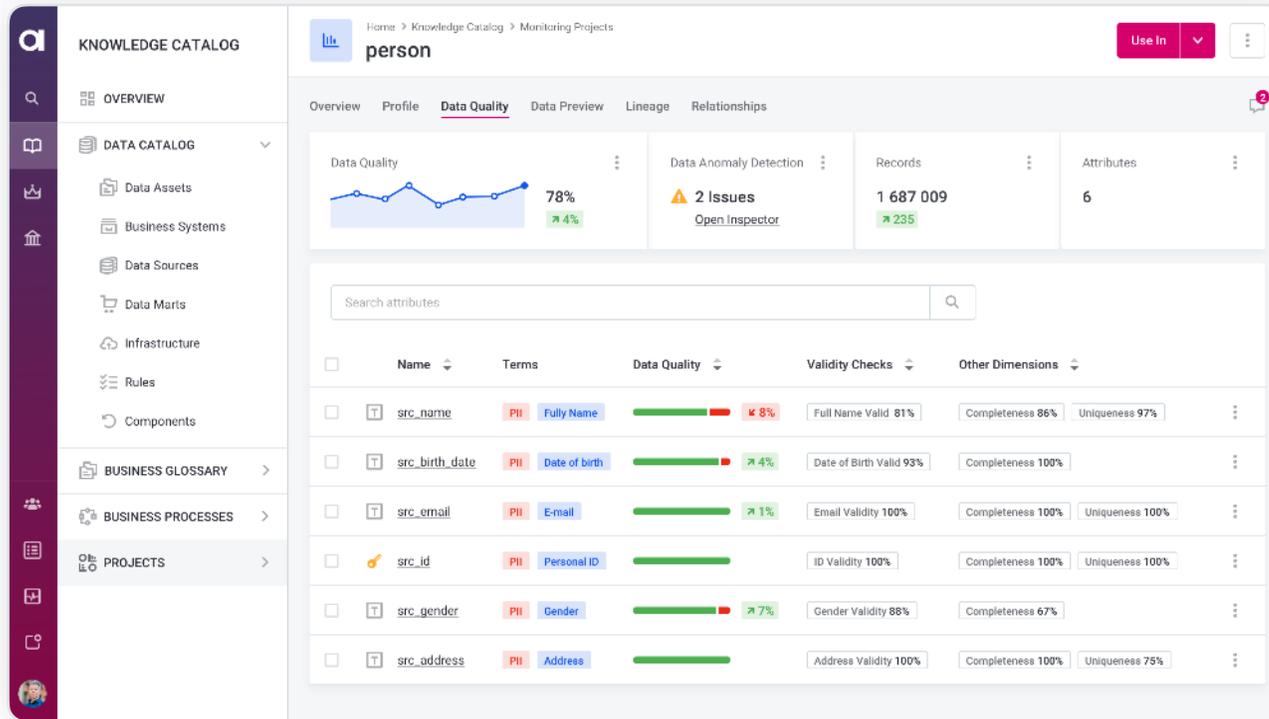
```
<terminated> load_all.ewf [Local Workflow] C:\WZ\ONEm4_rc2\jre\bin\javaw.exe (May 11, 2020, 2:44:22 PM - 2:46:53 PM)
11.05.2020 14:45:56 [INFO] Task 'add term relations' : RUNNING
2020-05-11 14:45:57 [INFO] com.ataccama.dqc.tasks.io.mmm.writer.MMMWriterInstance messageId=Output from One Metadata Writer does not have properly mapped generated GID column. expected GID column=
11.05.2020 14:46:02 [INFO] Task 'add term relations' : FINISHED_OK
11.05.2020 14:46:02 [INFO] Link(add term relations->Add detection rules) evaluated to ACCEPTED
11.05.2020 14:46:02 [INFO] Task 'Add detection rules' : IN_QUEUE
11.05.2020 14:46:02 [INFO] Task 'Add detection rules' : RUNNING
11.05.2020 14:46:48 [INFO] Task 'Add detection rules' : FINISHED_OK
11.05.2020 14:46:48 [INFO] Link(Add detection rules->assign detection rules to terms) evaluated to ACCEPTED
11.05.2020 14:46:48 [INFO] Task 'assign detection rules to terms' : IN_QUEUE
11.05.2020 14:46:48 [INFO] Task 'assign detection rules to terms' : RUNNING
2020-05-11 14:46:49 [INFO] com.ataccama.dqc.tasks.io.mmm.writer.MMMWriterInstance messageId=Output from One Metadata Writer does not have properly mapped generated GID column. expected GID column=ruleInstanceGid
11.05.2020 14:46:53 [INFO] Task 'assign detection rules to terms' : FINISHED_OK
11.05.2020 14:46:53 [INFO] Processor has finished with state: FINISHED_OK
```

Anomaly detection



- > Fully automatic AI-powered process
- > Detects irregularities in data without any rule configuration
- > Learns from user feedback and improves over time

Dashboards and reporting



- > Smart rule suggestions and project auto-configuration
- > Easy self-service project deployment and operation
- > Utilizing data Catalog, Profile and Rules repositories
- > Multiple export options for visualizing results

Automated Data Quality



Automated data cleansing and standardization

Standardize format, validate data, flag invalid values, calculate DQ scores



Data enrichment

- Translate values to a „master“ set.
- Enrich data using external APIs and resources
- Manage Reference sets and codebooks



DQ firewall

Prevent poor quality data from entering your system using Ataccama as a validation authority

Issue Resolution

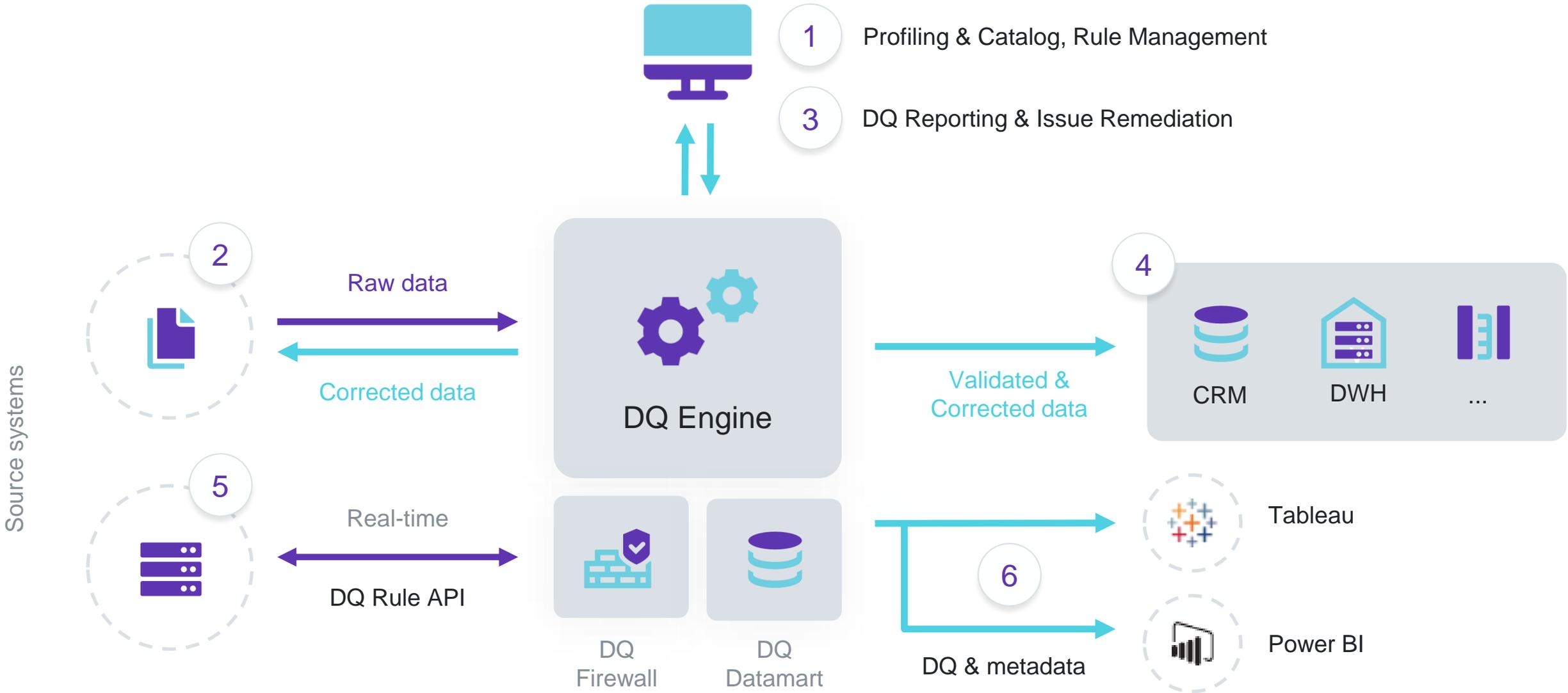
The screenshot shows an 'Issues Tracker' application interface. At the top, there are navigation tabs: 'For Resolution' (selected), 'For Supervisor', 'For Approval', 'Approved', and 'Closed'. Below the tabs is a search bar with the placeholder text 'Filter by name, owner, creation date...'. The main content is a table with the following columns: ID, Severity, System, Entity, Score, #Rec., Data, Description, Assignee, and Changed. The table contains 10 rows of data, each representing a data quality issue. The severity icons range from green (info) to red (warning or error). The bottom of the table shows a pagination control: '< 0-10 from 100 >'. A vertical sidebar on the left contains various navigation icons.

ID	Severity	System	Entity	Score	#Rec.	Data	Description	Assignee	Changed
4616	Warning	Marketing	Customer	100	2	Mrs. Jane Jones	Manual merge	Jane Smith	2019-01-06 17:21:31
4618	Warning	Marketing	Customer	100	2	Ms. Angela Thompson	Manual telephone selection	John Taylor	2019-01-06 17:21:31
4619	Warning	Marketing	Customer	100	4	Jane Holmberg	Manual telephone selection	John Taylor	2019-01-06 17:21:31
4621	Info	Marketing	Customer	500	1	203-918-116	Cleansing: Phone	Rachel Adams	2019-01-19 15:38:36
4622	Info	CIF	Contact	0	1	9842120332	Cleansing: Phone	Rachel Adams	2019-01-19 15:38:36
4624	Info	CIF	Contact	0	1	Ms. Regina Winters	Manual email verification	John Taylor	2019-01-19 15:40:09
4625	Error	Marketing	Customer	2000	1	Jane Black	Manual email verification	John Taylor	2019-01-19 15:43:00
4626	Error	Marketing	Customer	505	2	Jane Holmberg	Manual email verification	John Taylor	2019-01-24 17:21:09
4627	Error	Marketing	Customer	500	2	Wesley Nives	Manual email verification	Rachel Adams	2019-01-24 17:21:09
4629	Warning	CIF	Customer	100	2	Mrs. Regina Winters	Manual email verification	Rachel Adams	2019-01-24 17:21:09

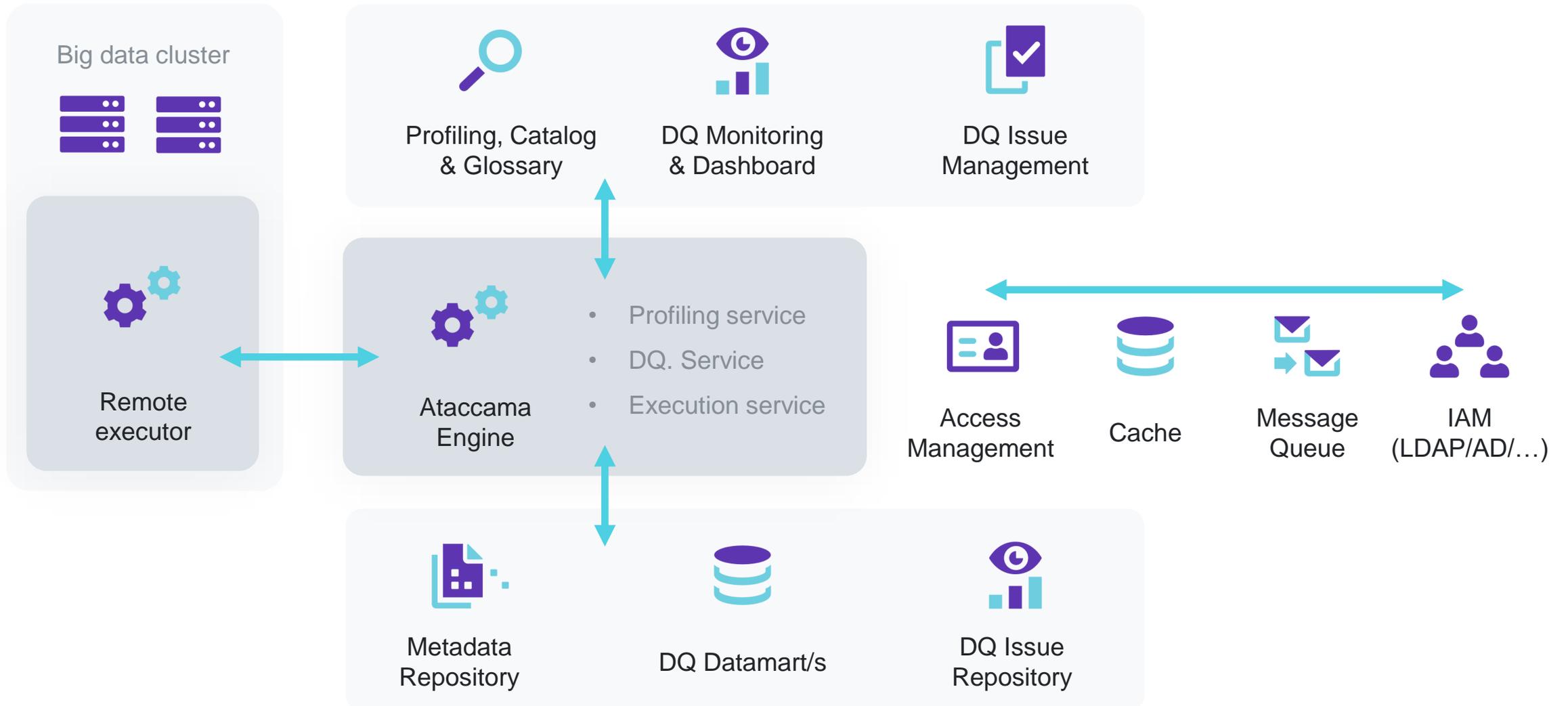
- > Manage and resolve DQ exceptions
- > Cleansing
- > Match & Merge
- > Drill-down to source level
- > Complete customizable workflow for data stewards
- > DQ Firewall to prevent new errors
- > Full audit log of actions
- > Progress reporting for managers

Architecture

Data Quality Process



High Level Architecture





KNOWLEDGE CATALOG



OVERVIEW



DATA CATALOG



Data Assets



Business Systems



Data Sources



Data Marts



Infrastructure



Rules



Components

BUSINESS GLOSSARY

BUSINESS PROCESSES

PROJECTS



Home > Knowledge Catalog > Monitoring Projects

person

Use In



Overview Profile Data Quality Data Preview Lineage Relationships



Data Quality



Data Anomaly Detection

2 Issues
[Open Inspector](#)

Records

1 687 009
↑ 235

Attributes

6

Search attributes



<input type="checkbox"/>	Name	Terms	Data Quality	Validity Checks	Other Dimensions
<input type="checkbox"/>	<input type="text" value="src_name"/>	PII Fully Name	72% (8% decrease)	Full Name Valid 81%	Completeness 86% Uniqueness 97%
<input type="checkbox"/>	<input type="text" value="src_birth_date"/>	PII Date of birth	76% (4% increase)	Date of Birth Valid 93%	Completeness 100%
<input type="checkbox"/>	<input type="text" value="src_email"/>	PII E-mail	99% (1% increase)	Email Validity 100%	Completeness 100% Uniqueness 100%
<input type="checkbox"/>	<input type="text" value="src_id"/>	PII Personal ID	100%	ID Validity 100%	Completeness 100% Uniqueness 100%
<input type="checkbox"/>	<input type="text" value="src_gender"/>	PII Gender	85% (7% increase)	Gender Validity 88%	Completeness 67%
<input type="checkbox"/>	<input type="text" value="src_address"/>	PII Address	100%	Address Validity 100%	Completeness 100% Uniqueness 75%



person

Use In ▼



- Overview
- Profile
- Data Quality
- Data Preview
- Lineage
- Relationships



Data Quality



78%

↗ 4%

Data Anomaly Detection



2 Issues

[Open Inspector](#)

Records

1 687 009

↗ 235

Attributes

6

Search attributes



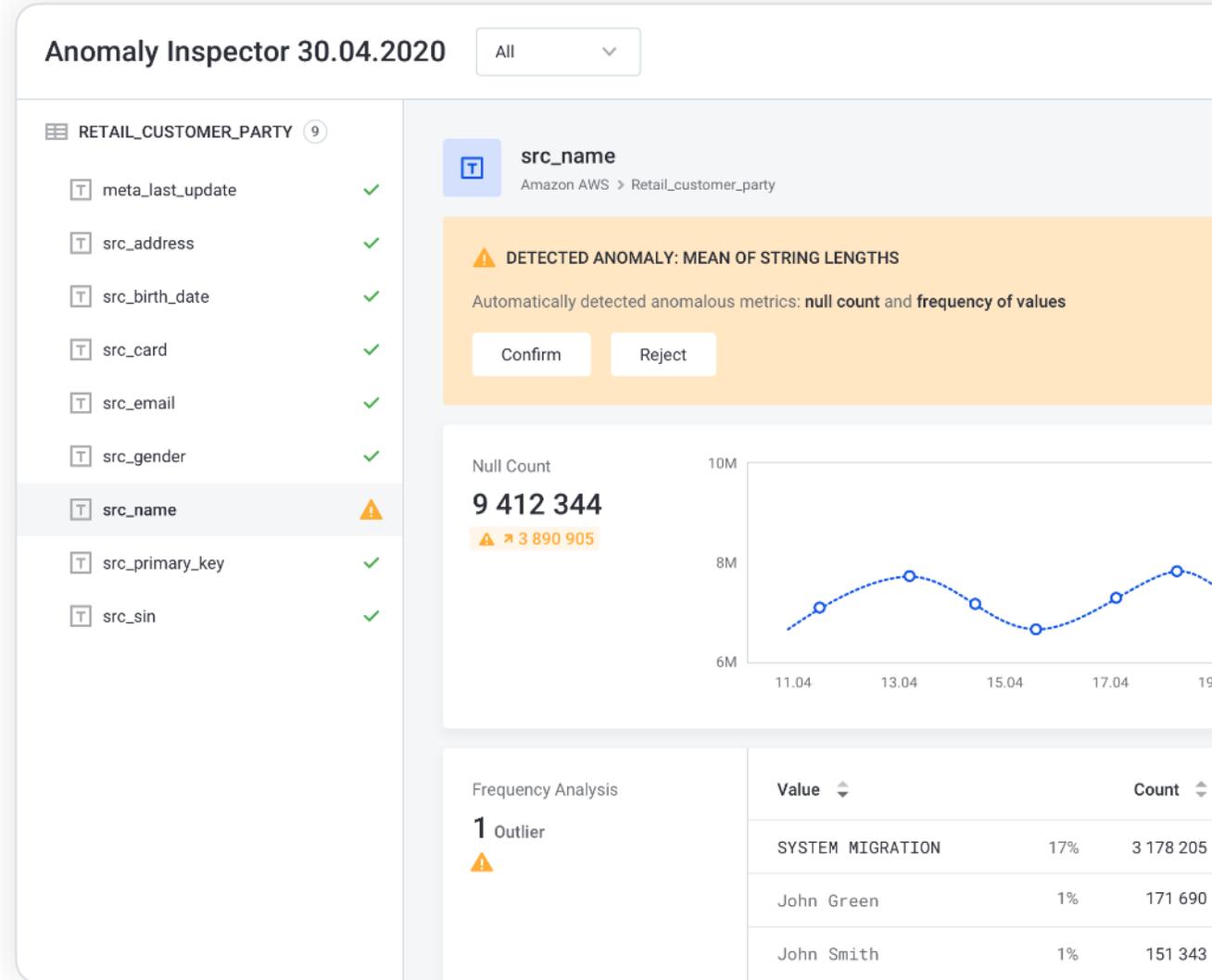
<input type="checkbox"/>	Name ⌵	Terms	Data Quality ⌵	Validity Checks ⌵	Other Dimensions ⌵
<input type="checkbox"/>	<input type="text" value="T"/> <u>src_name</u>	PII Fully Name	<div style="width: 92%;"><div style="width: 92%;"></div></div> ↙ 8%	Full Name Valid 81%	Completeness 86% Uniqueness 97%
<input type="checkbox"/>	<input type="text" value="T"/> <u>src_birth_date</u>	PII Date of birth	<div style="width: 96%;"><div style="width: 96%;"></div></div> ↗ 4%	Date of Birth Valid 93%	Completeness 100%
<input type="checkbox"/>	<input type="text" value="T"/> <u>src_email</u>	PII E-mail	<div style="width: 99%;"><div style="width: 99%;"></div></div> ↗ 1%	Email Validity 100%	Completeness 100% Uniqueness 100%
<input type="checkbox"/>	<input type="text" value="K"/> <u>src_id</u>	PII Personal ID	<div style="width: 100%;"><div style="width: 100%;"></div></div>	ID Validity 100%	Completeness 100% Uniqueness 100%

Null, empty, or...?

Anomaly detection

AI-powered

Works without rules





KNOWLEDGE CATALOG



OVERVIEW



DATA CATALOG



Data Assets



Business Systems



Data Sources



Data Marts



Infrastructure



Rules



Components

BUSINESS GLOSSARY

BUSINESS PROCESSES

PROJECTS

Home > Knowledge Catalog > Data Sources > MDM

Retail_customer_party

Overview Profile Data Quality Data Preview Lineage Relationships

Filter attributes, values, masks

Name

Terms

Insights



meta_last_update

Outlier Value

Detected in given attribute



src_first_name

Last Name



src_last_name

Last Name



src_gender

Gender



src_birth_date



src_email

Email



src_primary_key



meta_last_update



Insights

DETECTED ANOMALY: 2 OUTLIER VALUES

Null, N/A and 1900-01-01 are outliers



Make anomaly detection more precise by confirming or dismissing this suggestion.

Confirm

Dismiss

Count

Not Null 1 001 Null 876

Distinct 757 Duplicate 244

Unique 523 Non-Unique 234

Frequency Analysis

- DATA CATALOG
 - Data Assets
 - Business Systems
 - Data Sources**
 - Data Marts
 - Infrastructure
 - Rules
 - Components
- BUSINESS GLOSSARY >
- BUSINESS PROCESSES >
- OLAP PROJECTS >

Filter attributes, values, masks

Name	Terms	Insights
<u>meta_last_update</u>		Outlier Value Detected in given attribute
<u>src_first_name</u>	Last Name	
<u>src_last_name</u>	Last Name	
<u>src_gender</u>	Gender	
<u>src_birth_date</u>		
<u>src_primary_key</u>		

⚠ DETECTED ANOMALY: 2 OUTLIER VALUES

Null, N/A and 1900-01-01 are outliers

Make anomaly detection more precise by confirming or dismissing this suggestion.

2 minutes ago

⚠ Detected Anomaly on Retail_customer_party table: 2 outlier values

Null, N/A and 1900-01-01 are outliers on meta_last_update attribute

Count ⓘ

Not Null	1 001		Null	876
Distinct	757		Duplicate	244
Unique	523		Non-Unique	234

Frequency Analysis ⓘ

Automated DQ rules

Self-learning engine

The screenshot displays a Knowledge Catalog interface. The left sidebar contains navigation options: OVERVIEW, DATA CATALOG (expanded), Data Assets (selected), Business Systems, Data Sources, Data Marts, Infrastructure, Rules, and Components. Below these are BUSINESS GLOSSARY, BUSINESS PROCESSES, and PROJECTS. The main content area shows a search for 'Data Assets' with the filter 'with Customer'. A table lists the results with columns for Name, Terms, and Data Quality.

<input type="checkbox"/>	Name	Terms	Data Quality
<input type="checkbox"/>	Master customer	PII Customer	<div style="width: 100%;"><div style="width: 100%;"></div></div>
<input type="checkbox"/>	Master person	PII Customer Employee	<div style="width: 100%;"><div style="width: 100%;"></div></div>
<input type="checkbox"/>	customer_full	PII Customer	<div style="width: 100%;"><div style="width: 100%;"></div></div>
<input type="checkbox"/>	Customers 2019	PII Customer	<div style="width: 100%;"><div style="width: 100%;"></div></div>
<input type="checkbox"/>	Customer campaigns	Customer Campaign	<div style="width: 100%;"><div style="width: 100%;"></div></div>
<input type="checkbox"/>	customers_2019_ext	PII Customer	<div style="width: 100%;"><div style="width: 100%;"></div></div>
<input type="checkbox"/>	Person	Customer Employee	<div style="width: 100%;"><div style="width: 100%;"></div></div>
<input type="checkbox"/>	src_person	PII Customer Employee	<div style="width: 100%;"><div style="width: 100%;"></div></div>



Gender validation



- Overview
- Implementation**
- Usage
- Relationships
- Access
- History
- Settings



Description

Business rule validating the allowed values for gender data. Checks include reference data and legacy core system values.

Summary

INPUT ATTRIBUTES

A Gender

LOGIC

- 1 Not from reference data ^ v 👁 🗑
- 👁 Testing a custom validation
- 3 Checking against the legacy core system
- 🔧 4 Additional scoring

Input Attributes 1 i

Gender

String v



Related Terms Gender Customer

+ Add Attribute

Rule Logic i Validation Rule v

1 Not from reference data </> 📄 🔗 ⋮

WHEN

Gender v

Not from reference data v

Gender_RD + ⋮

THEN

Result Invalid v

Score 0



Catalog > Data Sources

DWH



Overview Data Browser Data Lineage Access History



DWH

Core

Address

Address city name

Address street name

Address zip code

Party

VAT Number

Customer Number

Company ID

First Name

Last Name

Segment

Segment ID

Segment

DWH

Datamart

Client

Address city name

Address street name

Address zip code

VAT Number

Customer Number

Company ID

Customer Name

Full Name

Segment

Country



Full Name



Original Name: attr124

Glossary Terms

Full Name

Product

Count

Not Null	100%	<div style="width: 100%;"></div>	Null	0%
Distinct	82%	<div style="width: 82%;"></div>	Duplicate	18%
Unique	72%	<div style="width: 72%;"></div>	Non-Unique	10%

Frequency Analysis

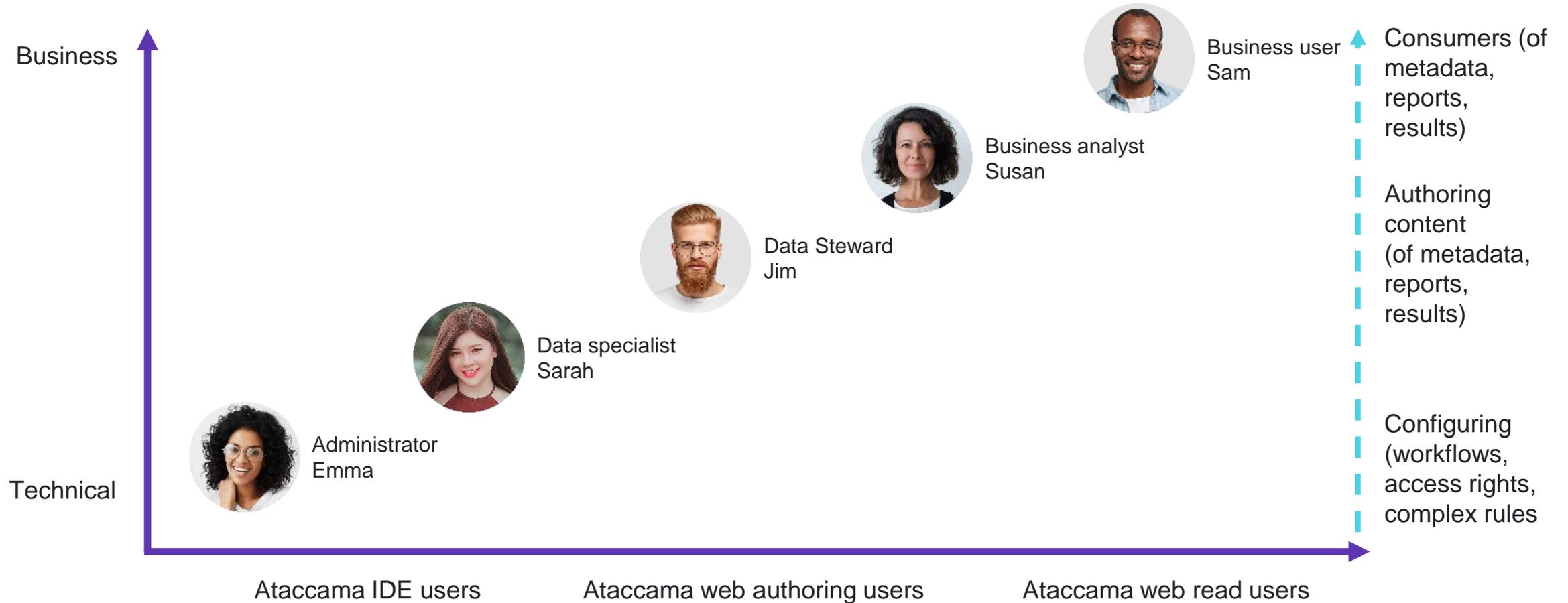
Outlier Value

Detected in given attribute

Value	Count
Maud Lynch	1 578 ⚠

Alexander McBride	1
-------------------	---

Ataccama ONE Personas



Data Quality: Example scenarios

Scenario 1

Catalog / DQ1:

Data
categorization

Scenario 2

DQ2:

Data quality
monitoring

Scenario 3

DQ3:

Data
corrections

Data Quality: Example scenarios

Scenario 1

Catalog / DQ1:

Data
categorization

Scenario 2

DQ2:

Data quality
monitoring

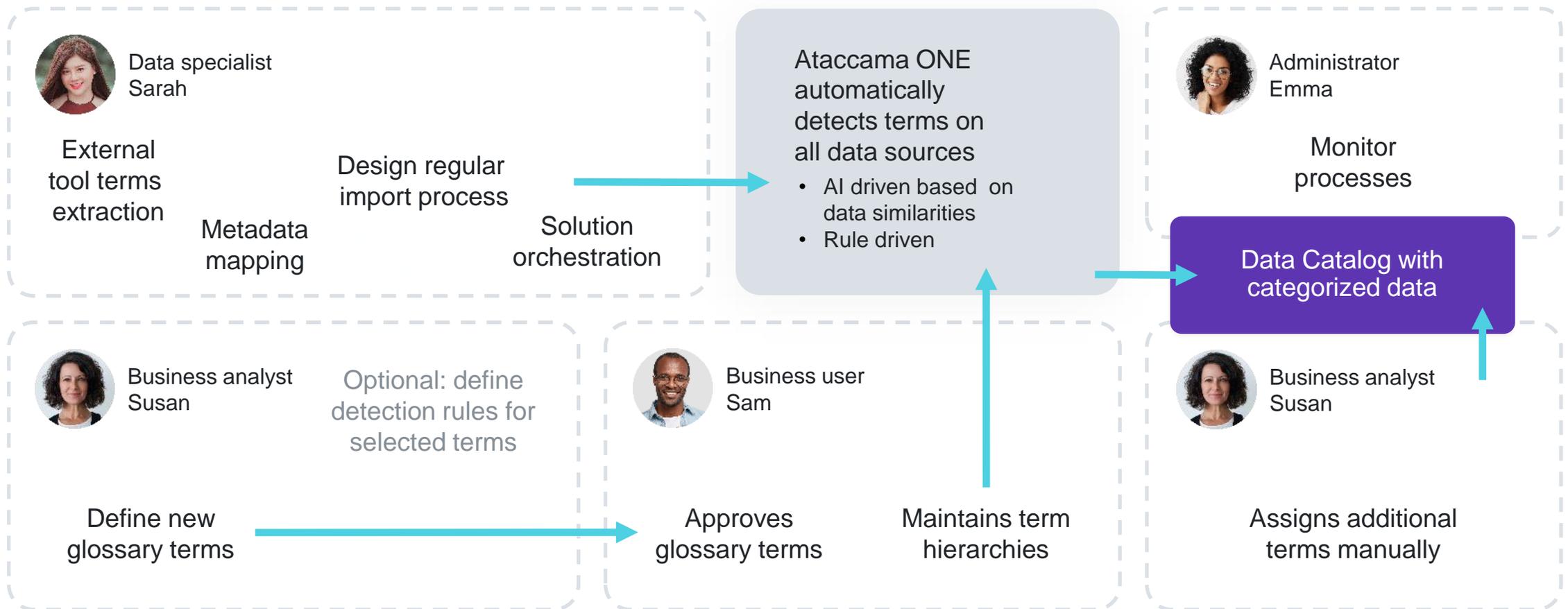
Scenario 3

DQ3:

Data
corrections

Scenario Catalog/DQ1: Data Categorization

Categorize data based on both new terms and terms imported from external sources



Scenario Catalog/DQ1: Data Categorization

Categorize data based on both new terms



Data specialist
Sarah

External
tool terms
extraction

Metadata
mapping

Design regular
import process

Solution
orchestration



Business analyst
Susan

Optional: define
detection rules for
selected terms

Define new
glossary terms



gl

- Graphical interface for composing import/mapping (meta)data processing logic
- Wide variety of connectors allowing to consume metadata from large variety of sources and applications
 - 300+ available algorithms + rich expression (function) language
 - Dedicated algorithms for metadata import (Gen2)
- Easy to use yet powerful and flexible
- Ability to implement even a complex logic
- Parametrization of designed import solutions
- Workflow designer allowing integration into desired landscape



Scenario Catalog/DQ1: Data Categorization

C

- Overview of term occurrences and their characteristics in all systems
- Capturing one or more hierarchies/relationship types
- Optional association with Detection and Data Quality rules
- Flexible underlying model allowing to define custom entities and attributes (Gen2)
 - Incl. Definition of custom relations and hierarchies

The screenshot displays the Business Glossary interface. The top part shows the 'Contacts' term details, including general information and relationships. Below this is a table of catalog items for '[USA] City'.

Name	Terms	Anomalies	Overall Validity	# Attributes	# Records	Origin	Location
address	Personal Data [USA] City [USA] Zip code		62%	4	29	Europe	CRM > test-db > NYC
accounts	Contacts Personal Data First name [USA] City		90%	26	284	Europe	CRM > test-db > Customers
Active Stores	Contacts Personal Data First name [USA] City		100%	25	1 973	Europe	CRM > test-db > Customers
transactions	Country		85%	8	121 625	Europe	CRM > test-db > All_Test
customers_main	Contacts Personal Data First name [USA] City		4%	17	1 000	Europe	CRM > test-db > All_Customers



Business analyst
Susan

Optional: define
detection rules for
selected terms

Define new
glossary terms



Business user
Sam

Approves
glossary terms

Maintains term
hierarchies



Business analyst
Susan

Assigns additional
terms manually

Scenario Catalog/DQ1: Data Categorization

Categorize data based on

external sources



Data specialist
Sarah

External
tool terms
extraction

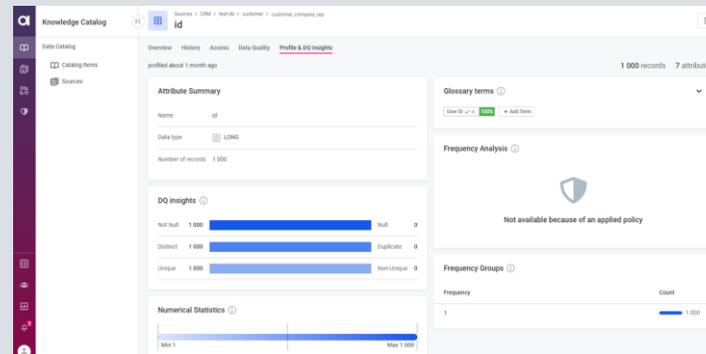
Metadata
mapping



Business analyst
Susan

Define new
glossary terms

- Powerful search allowing to filter and find relevant data sets
- Context helping with the categorization
 - Automated profiling capturing prominent data characteristics
 - Automatically detected Data anomalies (Gen2)
 - Automatically determined DQ
- Collaborative environment
 - Configurable approval workflow
 - Granular permissions and policies



Administrator
Emma

Monitor
processes

Data Catalog with
categorized data



Business analyst
Susan

Assigns additional
terms manually

Scenario Catalog/DQ1: Data Categorization

Categorize data based on

external sources

- Full overview and control over running jobs
- Easy processing issue troubleshooting
 - Option to integrate with external reporting systems
- Easy installation management



Data specialist
Sarah



Administrator
Emma

External
tool terms
extraction

Metadata
mapping

Monitor
processes

Data Catalog with
categorized data



Business analyst
Susan



Business analyst
Susan

Define new
glossary terms

Approves
glossary terms

Maintains term
hierarchies

Assigns additional
terms manually

ID	Name	Status	TransactionID	Engaged	Started	Finished
10	processMastermasters	FINISHED_OK	5738	2021-02-24 21:42:01	2021-02-24 21:42:01	2021-02-24 21:42:01
9	processDelta	FINISHED_OK	5474	2021-02-24 21:41:56	2021-02-24 21:41:56	2021-02-24 21:41:57
8	setMasterOverrides	FINISHED_OK	5431	2021-02-24 21:41:10	2021-02-24 21:41:10	2021-02-24 21:41:10
7	setMasterOverrides	FINISHED_OK	5405	2021-02-24 21:41:10	2021-02-24 21:41:10	2021-02-24 21:41:10
6	setMasterOverrides	FINISHED_OK	5369	2021-02-24 21:41:10	2021-02-24 21:41:10	2021-02-24 21:41:10
5	setMasterOverrides	FINISHED_OK	5340	2021-02-24 21:41:09	2021-02-24 21:41:09	2021-02-24 21:41:09
4	setMasterOverrides	FINISHED_OK	5332	2021-02-24 21:41:09	2021-02-24 21:41:09	2021-02-24 21:41:09
3	setMasterOverrides	FINISHED_OK	5323	2021-02-24 21:41:09	2021-02-24 21:41:09	2021-02-24 21:41:09

Data Quality: Example scenarios

Scenario 1

Catalog / DQ1:

Data
categorization

Scenario 2

DQ2:

Data quality
monitoring

Scenario 3

DQ3:

Data
corrections

Scenario DQ2: Data Quality Monitoring

Measure the current quality



Business analyst
Susan



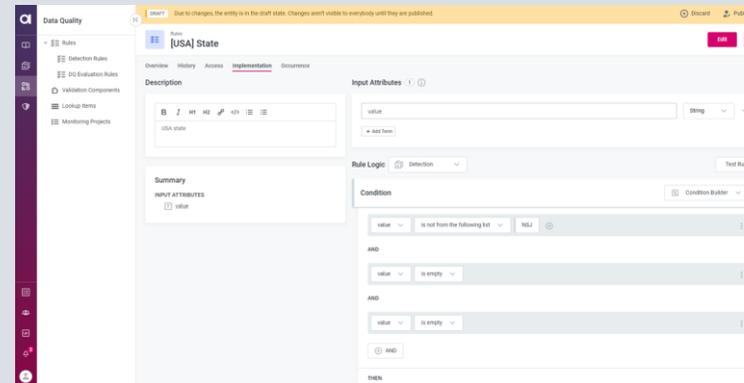
Business user
Sam

Defines & manages
data quality rules



Implements rule logics

- Central rule library with all captured (configurable) rule details
 - Rules for both automated DQ evaluation enriching Data Catalog and targeted DQ producing individual reports
 - Reusable in other data processes
- Customizable approval workflow
- Flexible rule implementation for all user types
 - Sentence-like definition
 - Expression language
 - Complex logic in dedicated components
- In-built library of frequently used rules



Business user
Sam

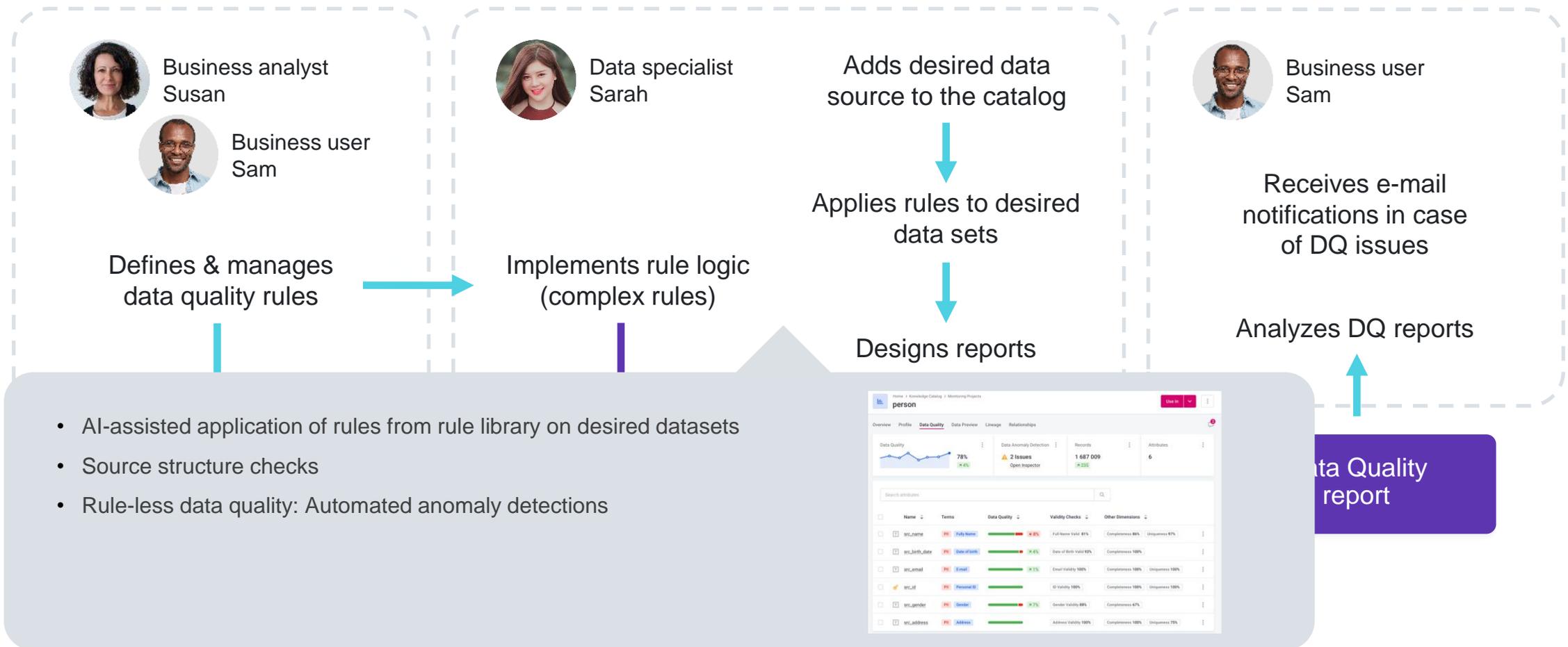
Receives e-mail
notifications in case
of DQ issues

Generates DQ reports

Generates Data Quality
report

Scenario DQ2: Data Quality Monitoring

Measure the current quality of your data and monitor its trends



Scenario DQ2: Data Quality Monitoring

Measure the current quality of your data and monitor its trends

Monitoring projects
Customer DQ report

Run monitoring

Overview Configuration & Results Report Export Notifications History Access

Overview Latest

Issues found Tables 2 Attributes 27 Checks 10

Structure Checks: No issues (2 checks on 2 tables)
Anomaly Detection: No checks applied
Data Quality: Issues In 1 table Detected (8 checks on 2 tables, 48% invalid samples)

Items to Monitor: 8 rule suggestions

Anomaly Inspector 30.04.2020

src_name
Amazon AWS - Retail_Customer_party

DETECTED ANOMALY: MEAN OF STRING LENGTHS
Automatically detected anomalous metrics: null count and frequency of values

Null Count: 9 412 344 (3 899 905)

Frequency Analysis: 1 Outlier

Value	Count	Anomaly
SYSTEM MIGRATION	17% 3 178 205	17% more than expected
John Green	1% 171 690	✓
John Smith	1% 151 343	✓



Business user
Sam

Receives e-mail
notifications in case
of DQ issues

Analyzes DQ reports

Data Quality
report

Data Quality: Example scenarios

Scenario 1

Catalog / DQ1:

Data
categorization

Scenario 2

DQ2:

Data quality
monitoring

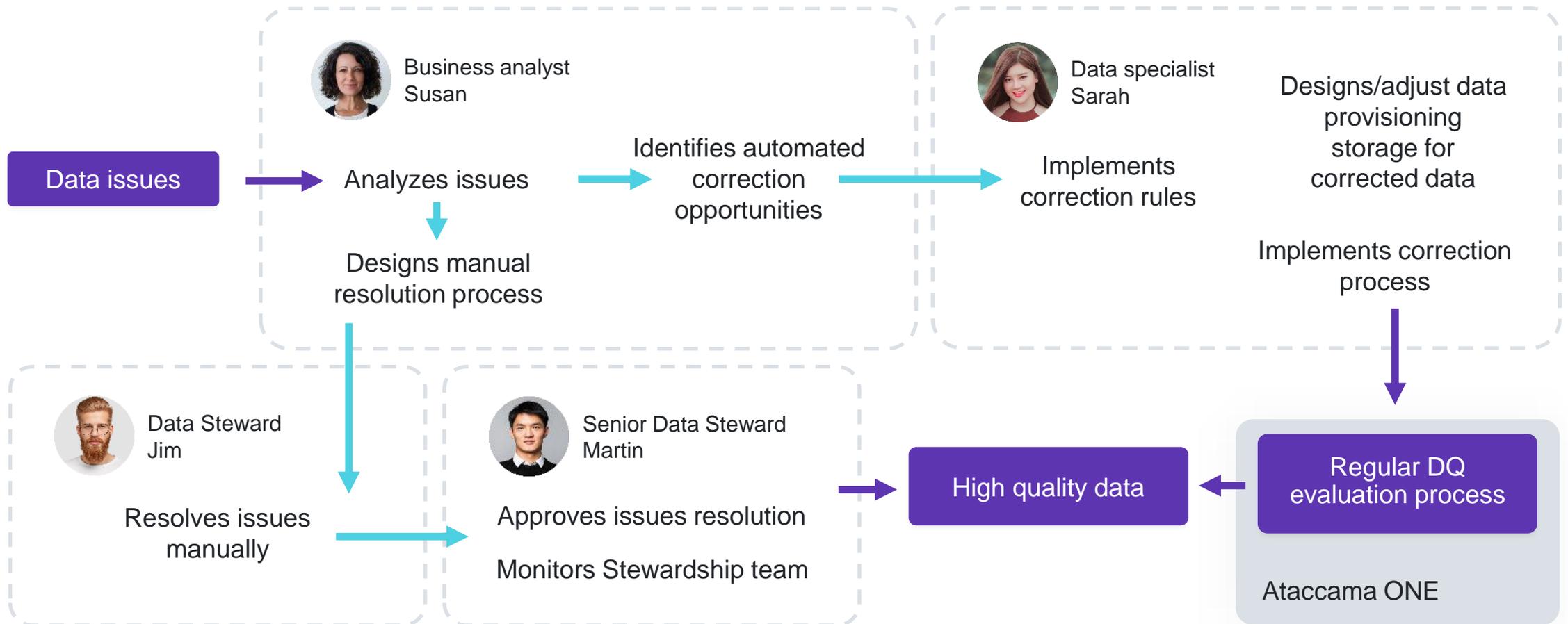
Scenario 3

DQ3:

Data
corrections

Scenario DQ3: Data Corrections

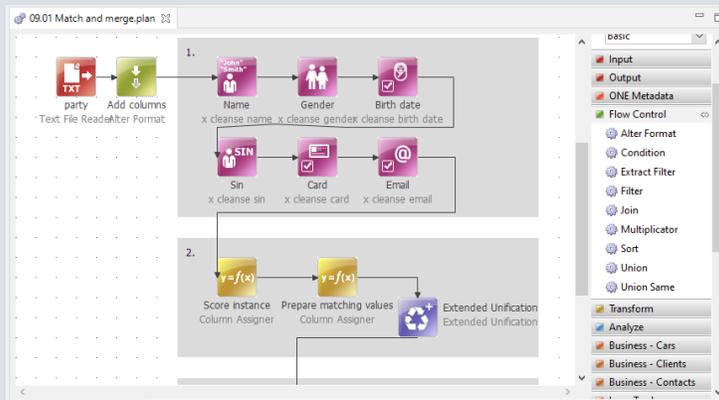
Resolve DQ problems and provide high quality data to all consumers



Scenario DQ3: Data Corrections

Res

- Graphical interface for composing import/mapping (meta)data processing logic
- Wide variety of connectors allowing to consume metadata from large variety of sources and applications
 - 300+ available algorithms + rich expression (function) language
 - Dedicated algorithms for metadata import (Gen2)
- Easy to use yet powerful and flexible
- Ability to implement even a complex logic
- Parametrization of designed import solutions
- Workflow designer allowing integration into desired landscape



data to all consumers



Data specialist
Sarah

Implements
correction rules

Designs/adjust data
provisioning
storage for
corrected data

Implements correction
process

High quality data

Regular DQ
evaluation process

Ataccama ONE

Scenario DQ3: Data Corrections

Resolve DQ problems and

- Manage and resolve DQ exceptions
 - Cleansing
 - Match & Merge
- Drill-down to source level
- Complete customizable workflow for data stewards
- DQ Firewall to prevent new errors
- Full audit log of actions

Designs/adjust data provisioning storage for corrected data

Implements correction process

Regular DQ evaluation process

ataccama ONE

Data issues

Analyze

Designs resolution

Data Steward Jim

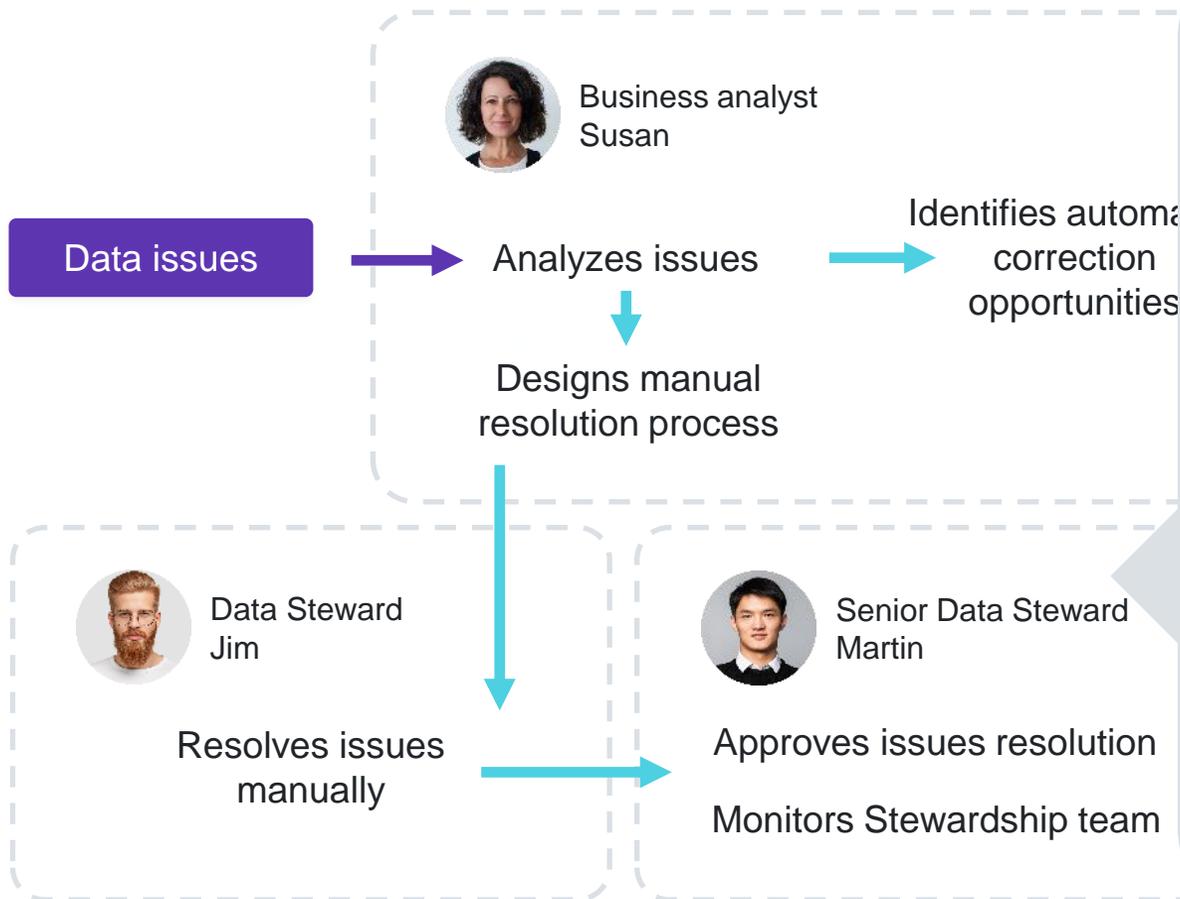
Resolves issues manually

The screenshot shows the ataccama ONE interface with a table of data quality issues. The table has columns for Id, Severity, System, Entity, Score, #Rec., Data, Description, Assignee, and Changed. A modal window is open, comparing 'Original data' and 'Corrected data' for a record with ID 18. The original data shows a mismatch between the birth date (1978-12-16) and the age (44). The corrected data shows the birth date updated to 2005-08-09 to match the age.

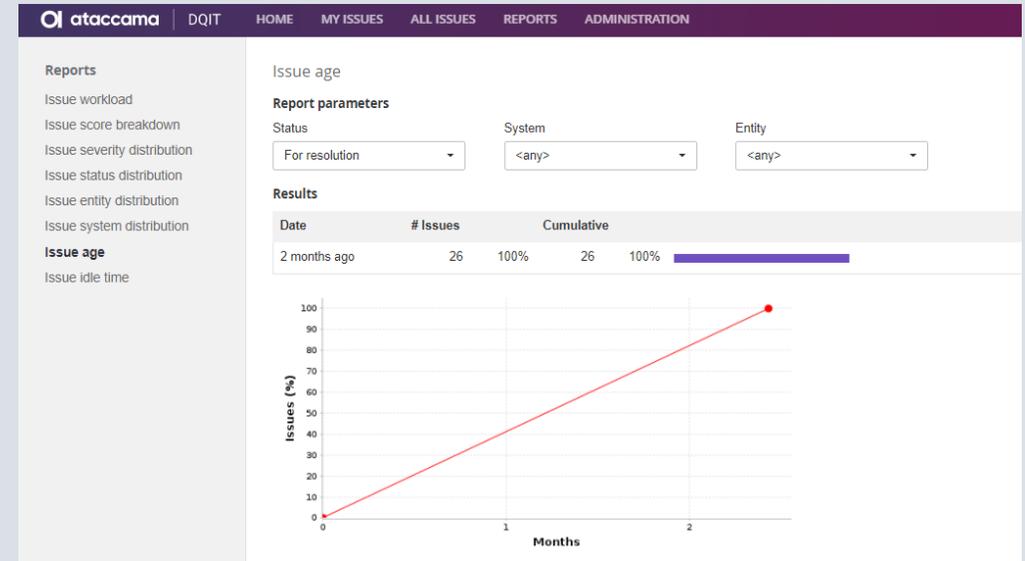
Id	Severity	System	Entity	Score	#Rec.	Data	Description	Assignee	Changed
2	Info	Marketing	Contact	0	1	Ms. Regina Winters	Manual email verification	John Taylor	2021-03-23 21:51:55
3	Info	Marketing	Contact	0	1	Jane			
4	Info	Marketing	Contact	0	1	Jane			
5	Info	Marketing	Contact	0	1	Wes			
7	Error	Marketing	Contact	500	1	203			
8	Error	Marketing	Contact	2000	1	984			
10	Warning	Marketing	Customer	100	2	Jane			
17	Warning	Marketing	Customer	100	2	Ms.			
18	Warning	Marketing	Customer	100	4	Jane			

Scenario DQ3: Data Corrections

Resolve DQ problems and provide high quality data to all consumers



- Approval workflow
 - Confirming resolution, adding comments, rejecting corrections etc.
- Progress reporting for managers



Ataccama ONE

These are not redesigned, if someone needs them let marketing know

Managing & steering
(consuming the results of Ataccama processing)



Amy

LOB Head
(business manager)

Business

Using & collaborating
(searching and using the best data)



Ginger

Business Analyst

Ensuring data is fit for purpose
(authoring business content)



Fred

Business Data Steward

Ad-hoc profiling & DQ data prep
(leveraging for best data science results)



Data Scientist

Configuring
(catalog, rules, workflows)



Jane

Technical Data Steward / Data Engineer

Monitoring
(ensuring smooth & secure operations)



Jack

IT Ops / DevOps

Ataccama Web GUI Users

Ataccama Desktop GUI Users

Ataccama Monitoring Console

LOB Head

MEET AMY



AMY

**LOB Head
(business manager)**

AMY'S WORK

Quality data for decisions

Amy needs to make sure that tools and processes are in place to ensure that business decisions are made based on validated data.

Ensure team efficiency

Amy's team needs to spend time on using good data for business, not waste time on searching for the right data or re-doing work due to poor data quality.

Know that data is secure

While Amy is not responsible for data security, she wants to be sure that only authorised team members access data based on their roles.

Equip team for self-service

Amy's team can always reach out to IT, however, it needs tools that will allow business self-service for selected tasks (like search for the best data source).

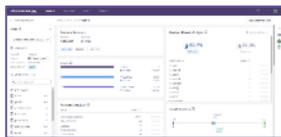
Have a peace of mind

Amy is busy with her everyday tasks. She does not have time to dig into detail, but she needs to make sure her team is working with quality data in a secure way.

AMY'S TOOL (optional)



**DASHBO
ARD**



HOW AMY CAN BENEFIT FROM ATACCAMA

- ✓ End-to-end coverage of business glossary, catalog and data quality in one application
- ✓ End-to-end process and use case coverage

- ✓ Teams equipped with collaborative and AI powered tool
- ✓ One-click search for data and its DQ confidence
- ✓ Preventative (vs. reactive) approach

- ✓ Application strictly follows access based on privileges and integrates with the Company's authentication and authorisation system

- ✓ Artificial Intelligence built in the tool learns from user interaction
- ✓ User interface is easy to navigate and easy to learn

- ✓ Get notified if something goes wrong – for example, if DQ falls under thresholds, or if DQ issues are not resolved in time

ALWAYS USE THE BEST DATA

Business Analyst

MEET GINGER



Business Analyst

GINGER'S WORK

Identify the best data for the task

Ginger needs to find the right data for analysis. Apart from being relevant, the data must have good quality. She needs a central catalog of data assets.

Collaborate with other users

The whole team needs to benefit from Ginger's knowledge, and Ginger wants the same in return. Ginger needs to collaborate easily with the others.

Get authorised access

Ginger is not supposed to know everything, but she can't be bothered to think about what she is allowed or forbidden to use.

Define what is considered good DQ for business & compliance

Ginger needs to define in plain English her requirements for DQ in relation to her tasks including statutory reporting.

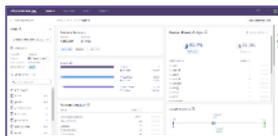
Monitor quality of critical data assets

Some of the Ginger's reports / data sources are updated regularly. Ginger needs to keep track of the source DQ that feed into those reports.

GINGER'S TOOL



CATALOG
&
GLOSSARY,
DASHBOARD



HOW GINGER CAN BENEFIT FROM ATACCAMA

- ✓ Full text search
- ✓ Filter based on glossary terms and business domains
- ✓ Filter based on data sources
- ✓ Get Data Quality confidence scores

- ✓ Comment on metadata assets
- ✓ Trigger tasks or notifications by the comments
- ✓ Help the data owner to improve the assets descriptions

- ✓ Follow the rules without knowing them
- ✓ Granular privileges
- ✓ Do everything allowed without any risks

- ✓ Save resources through streamlining DQ assessment process
- ✓ The requirements can be picked up by authoring users in the application

- ✓ Get notified if DQ drops before data feed into reports
- ✓ See results per detailed DQ conditions
- ✓ Drill down to invalid records

MAKE SURE DATA IS FIT FOR PURPOSE

Business Data Steward

MEET FRED



**Business
Data Steward**

FRED'S WORK

Manage the glossary

Importing terms, creating business descriptions or classifying data, Fred's job is to make sure the glossary reflects the business.

Analyse data profiling results

Fred often needs to analyze a new data source before ad-hoc validation of its quality of data to make sure rules will cover all problems hidden in the data.

Create rules for DQ validation

Fred needs to run simple validation of data quality without writing programming code or SQL. He needs flexible rules that can be reused across sources.

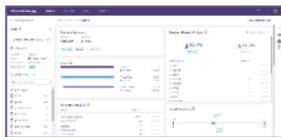
Apply DQ rules on data

Fred needs to apply DQ rules on data assets without the need to write code. Then he schedules regular DQ monitoring and reporting.

Resolve DQ issues manually

Sometimes a human eye is needed to resolve a DQ issue. Fred needs to assess each erroneous entity on mission critical data before fixing it.

FRED'S TOOL



**PROFILIN
G
RULES &
PROJECTS
ISSUE
TRACKER**

HOW FRED CAN BENEFIT FROM ATACCAMA

- ✓ Create and share glossary terms
- ✓ Automated mapping of glossary terms to other assets in the solution
- ✓ Classify data easily including sensitive data

- ✓ Easily connect to a new data source
- ✓ One-click-data profiling get insights into the data, metadata and quality of a source
- ✓ Share profiling results with others

- ✓ Create simple DQ checks in a central library
- ✓ Validation rules do not require coding skills
- ✓ Rules are reusable across datasets without additional configuration

- ✓ Apply rules on data in a simple point-and-click UI
- ✓ Establish regular DQ monitoring without diving into technical configuration

- ✓ Resolve DQ issues by simply typing in or selecting correct values
- ✓ Each resolution goes through approval process
- ✓ Choose where to provide cleansed data to

QUICKLY PREPARE GOOD QUALITY DATA SETS

Data Scientist

MEET ALEX



Alex

Data Scientist

ALEX'S WORK

Identify the best data for the task

Alex needs to find the right data for analysis. Apart from being relevant, the data must have good quality. He needs a central catalog of data assets.

Collaborate with other users

The whole team needs to benefit from Alex's knowledge, and Alex wants the same in return. He needs to collaborate easily with the others.

Get authorised access

Alex is not supposed to know about everything, but she can't be bothered to think about what she is allowed or forbidden to use.

Run ad-hoc profiling and rules

Alex needs to work with ad-hoc and new data sources. He needs to quickly understand what's in the data and if it is usable for analysis.

Run data quality preparation

Alex needs to prepare a quality data set for developing his statistical models, and he needs it quickly without extensive programming.

ALEX'S TOOL



ONE Web Apps



ONE Desktop IDE

HOW ALEX CAN BENEFIT FROM ATACCAMA

- ✓ Full text search
- ✓ Filter based on glossary terms and business domains
- ✓ Filter based on data sources
- ✓ Get Data Quality confidence scores

- ✓ Comment on metadata assets
- ✓ Trigger tasks or notifications by the comments
- ✓ Help the data owner to improve the assets descriptions

- ✓ Follow the rules without knowing them
- ✓ Granular privileges
- ✓ Do everything allowed without any risks

- ✓ One-click-data profiling to quickly assess the content and quality of a new data source
- ✓ Ad-hoc data quality rules evaluation

- ✓ Data quality preparation functionality with a wide palette of options
- ✓ Easy to use GUI to accomplish the task quickly and without writing code

CONFIGURE THE SOLUTION

Data Engineer

MEET JANE



Jane

**Technical Data Steward
/ Data Engineer**

JANE'S WORK

Create Catalog of data sources

Jane makes sure every data source across the organization is Catalogued, and the data Catalog is regularly updated for changes.

Configure DQ rules logic

Jane prepares pre-aggregated data sets for business to apply DQ rules on. Also, she configures complex DQ rules, not only monitoring but also for data cleansing purposes.

Configure orchestration

Jane ensures all jobs are scheduled. She selects the right execution scenarios (local or Spark job), create orchestration workflows and online services.

Configure issue resolution workflows

Part of the Jane's tasks is to ensure that data stewards clean the data. She needs a tool to send selected problematic records for manual resolution.

Analyze DQ issues and setup cleansing

Jane is very proactive in monitoring DQ, analyses issues and, where applicable, configures automated cleansing and remediation plans.

JANE'S TOOL



**ONE Web
Apps**



**ONE Desktop
IDE**

HOW JANE CAN BENEFIT FROM ATACCAMA

- ✓ Simple connectivity to a wide variety of data sources
- ✓ "One-click" cataloguing via ad hoc profiling
- ✓ Scheduled profiling
- ✓ Sample data selection

- ✓ Create powerful validation rules in a simple UI
- ✓ Re-use rules across multiple datasets without modifications
- ✓ Share complex rules with business stewards

- ✓ Leverage Big Data cluster power for execution
- ✓ Automate tasks and seamless solutions with orchestration workflows
- ✓ Turn any rule into an online service

- ✓ Reuse existing DQ rules to send issues for resolution
- ✓ Configure custom resolution workflow
- ✓ Choose where to send clean data after resolution

- ✓ Create cleansing, enrichment, and validation jobs in a GUI without writing a line of code
- ✓ Enjoy out-of-the box library of rules to speed up the development

IT operations

MEET JACK



Jack

IT Ops / DevOps

JACK'S WORK

Administer servers

Part of the Jack's tasks is to ensure that all jobs are executed successfully. He needs to get notified and be able to drill into the root cause in case of a problem.

Establish development pipeline

Jack needs to ensure that configuration created by stewards can be easily deployed on servers.

Ensure security and access

Jack must ensure correct users have access to applications without compromising on security.

Business continuity

Jack needs to regularly back-up the solution and if needed, restore it from back up to ensure continuity of the service.

JACK'S TOOL



Admin Center
Keycloak



ONE Desktop
IDE

HOW JACK CAN BENEFIT FROM ATACCAMA

- ✓ Easy administration console
- ✓ System health monitoring statuses
- ✓ Notifications if something goes wrong

- ✓ Integration with version control systems
- ✓ Audit log and versioning in web application

- ✓ Single sign-on through Keycloak component
- ✓ Simple user and role management
- ✓ Integration with AD/LDAP

- ✓ Configurations managed in version control system and app storage
- ✓ Data quality processing results exportable into a database or file



Learn ONE, know it all

MDM

Data Quality

Reference Data Management

Data Integration

Data Catalog

Data Profiling



Designed for data people

Data Stewards

Governance Pros

Data Analysts

Data Scientists

Data Engineers

Ataccama Cloud Platform as a Service

Focus on your data initiatives while we take care of the infrastructure, monitoring, maintenance and security.



Software

Includes licensed Ataccama products, operating system, 3rd party licenses, and monitoring and service stack.



Installation

Fully automated deployment including network setup, security, and database repositories.



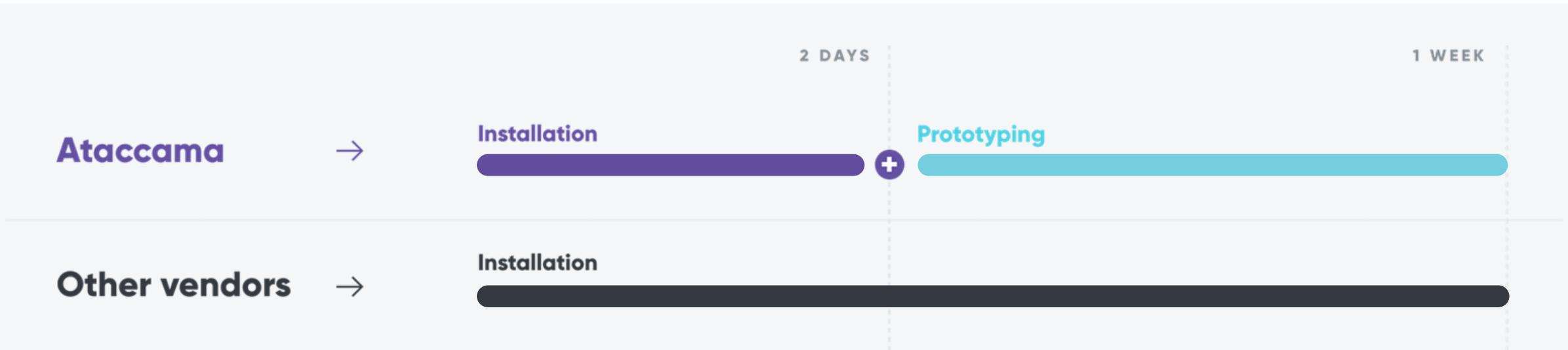
Infrastructure

Deployed with a cloud provider selected by you and in your preferred region. All costs are included in the PaaS pricing.

We work with the best to guarantee **stability,**
security and industry standards



Quick start



**24/7
monitoring**



**Security
patches**

**Disaster
recovery**

**Incident
resolutions**



**Guaranteed
SLAs & RTO/RPO
times**

**Customer
support**



**Back-up &
Restore**

Single Tenant Architecture



Your data is always safe and only yours

To assure the highest possible availability and reliability of data, Ataccama leverages native services provided by cloud vendors, such as Amazon RDS and Azure Database for Postgres. All our instances are single tenant, so your data resides in a physically separated data repository.



Ataccama is GDPR & CCPA compliant and holds ISO 27001 certification. Additionally SOC2, HIPAA and Privacy Shield are underway (estimated to complete in 2020).

Built on industry standards

We use proven, industry-standard technologies for our PaaS framework to configure, monitor, and manage the lifecycle of our applications.



And **scale** as you need

Our Kubernetes infrastructure allows you to scale as you go, and spin off new environments with ease when they are needed.

