



# EXL DQLens™

*A data quality tool suite that  
accelerates value, improves analytics,  
and eliminates technology barriers.*



According to Gartner,

“the average financial impact of poor data quality on the organization is estimated to be

**\$9.7 million per year”**

## What data quality issues lead to this cost, and why does bad data exist?

- Multiple sources of data
- Difficult to understand, track and control large volumes of data
- Lack of input controls in systems of record
- Legacy data is not updated to align with modern business processes
- Unvalidated 2nd and 3rd party data integrated with 1st party data

### Indirect Impact

- Brand reputation
- Customer satisfaction
- Missed opportunities
- Misguided business decisions

### Direct Impact

- Postage spent on sending mail to undeliverable addresses
- Legal and monetary penalties
- Unrecovered debts due to misdirected bills
- Time spent analyzing & correcting data

IBM estimates *the cost of bad data in the US alone is*  
**\$3.1 Trillion Annually**

## Building Custom Code

- Pros:
  - Low software costs
  - Custom-tailored to business requirements and technical landscape
- Cons:
  - High development costs and risks
  - Requires specialized skillsets and experience
  - Requires robust testing and quality control
  - Difficult to maintain

## EXL DQLens™

- Quicker time to value and more robust than custom code
- Less expensive than commercial tools
- Leaves the door open to adoption of commercial tools

## Buying Commercial Tools

- Pros:
  - Robust and reliable
  - Full-featured
  - Dedicated Support
  - Periodic patches and enhancements
- Cons:
  - Expensive
  - Requires specialized skillsets to operate and maintain
  - Monolithic architectures
  - Vendor lock-in



## Background and Vision

(SNI POC team to help articulate)



## POC Objectives

- (SNI POC team to help articulate)

## EXL DQLens™ Solution Capabilities



### PROFILE

Analyzes data sets and reports basic statistics



### CLEANSE

Validates and standardizes data elements



### RESOLVE

Identifies real-world entities across disparate data sets



### MONITOR

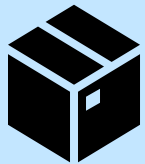
Visualizes data quality metrics and trends over time

# EXL DQLens™ Profile

*When a data set is not well-understood, EXL DQLens Profile can be used to produce summary-level statistics that characterize the data and may uncover potential data quality issues.*

*Consuming poorly understood data sets is a major risk.*

What are the most consequential questions to ask about unknown data sets



## Structure

- What attributes does the data contain?
- How is the data structured?
- What keys are used to identify or refer to the data?



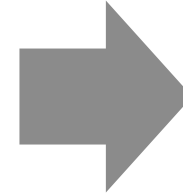
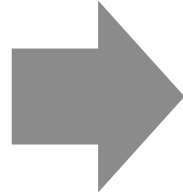
## Content

- How is the data formatted?
- Is the data complete?
- Is there duplication within the data?
- Are values reasonable for what I expect?



## Value Potential

- How does the data overlap or relate to existing assets?
- Can it be used to enrich, improve or supplement existing data?
- What new insights can be driven from this data?



The source data is in Excel format

The data is loaded into Snowflake database. The data is landed as a string datatype.

EXL DQLens™ Profile is used to analyze the data.



# EXL DQLens™ Cleanse

*Once the characteristics of a data set are understood, EXL DQLens Cleanse applies business rules to parse, standardize, and validate data to ensure consistency and fit for use.*

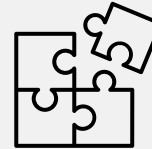


*Cleanse enables fundamental parsing, standardization and validation capabilities powered by rule and reference data libraries.*

## Key Features:

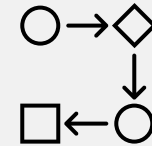
- Managed reference values and translation mappings enable scalable reusability and standardization
- Define rules once and apply them anywhere with plain SQL syntax
- Dynamically generate database views that take advantage of underlying computational power – cloud or on-prem
- Leverage libraries of rules and reference data that significantly accelerate time to value

## Capabilities



### Parsing

- Email – user, domain
- Phone – country code, area code
- Name – Prefix, First, Middle, Last, Suffix



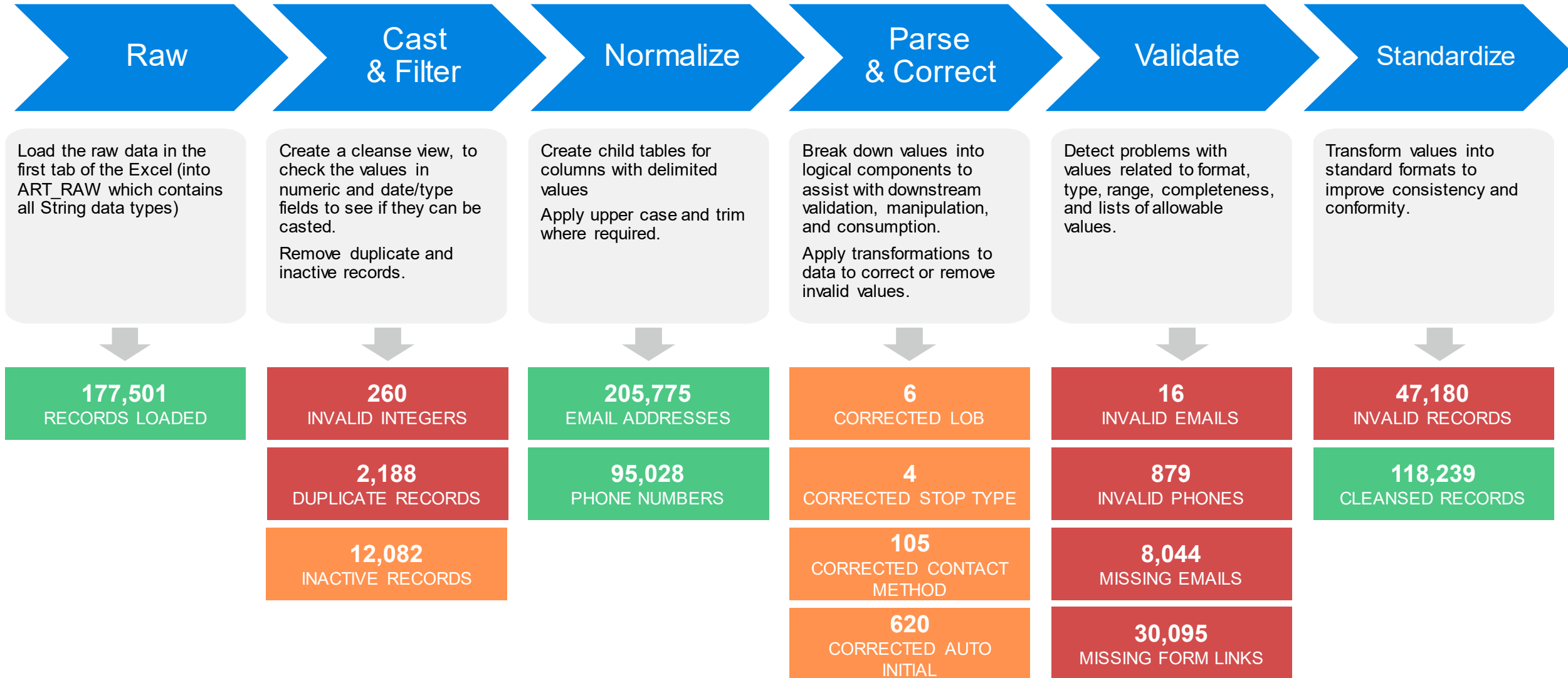
### Transformation

- Well-formatted full name
- Standardized names/addresses for matching
- Mapping to standard or corrected values



### Validation

- Ranges
- Equivalence
- Allowable values
- Patterns





## Filter Logic

Rule	Logic	Rows Filtered
In-active Records	ACTIVE_FLAG = 'N'	11,509
Duplicate Records	Duplicate on (LOCATION_ID + STOP_TYPE + TENDERING_PARTY_ID + LIABLE_PARTY_ID + LOB)	2,188

## Attributes Casted (invalid values set to null)

Attribute	Source Data Type	Target Data Type	Invalid Values
APPOINTMENT_PROFILE_ID	TEXT	NUMBER	0
LOCATION_ID	TEXT	NUMBER	90
LIABLE_PARTY_ID	TEXT	NUMBER	111
TENDERING_PARTY_ID	TEXT	NUMBER	69
UPDATED_DT	TEXT	TIMESTAMP	0



## Normalization Rules (other than standard casing and trimming)

Attribute	Logic	Normalized Values
AUTO_INITIAL	'Y' if APPOINTMENT_AUTOMATION contains "INITIAL", else "N"	165,419
AUTO_RESCHEDULE	'Y' if APPOINTMENT_AUTOMATION contains "RESCHEDULE", else "N"	165,419
APPT_REF_REQUIRED	'Y' if CUSTOMER_INPUT contains "APPOINTMENT REFERENCE REQUIRED", else "N"	165,419
INPUT_PRELOADED_DECAL	'Y' if CUSTOMER_INPUT contains "PRELOADED DECAL", else "N"	165,419
EMAIL_ADDRESS	TO, CC, and RESCHEDULE email addresses split by the ";" delimiter	205,775
PHONE_NUMBER	PHONE_NUMBER split by the ";" delimiter	95,028



## Attributes Parsed

Attribute	Parsed Components
EMAIL_ADDRESS	Account, Domain, Top-Level Domain
PHONE_NUMBER	Country Code, Area Code, Exchange Number, Line Number, Extras

## Attributes Corrected (invalid values set to null)

Attribute	Rule	Number Corrected
LOB	Manual mapping (reference data translation)	6
STOP_TYPE	Set to null if not "P" or "D" (reference data lookup)	4
CONTACT_METHOD	Manual mapping (reference data translation)	105
EMAIL_NOTIFICATION_TYPE	Set to null if not "REQUEST TO CUSTOMER"	1
AUTO_INITIAL	Set to "N" if CONTACT_METHOD is not "NONE" or "EMAIL"	620
AUTO_RESCHEDULE	Set to "N" if CONTACT_METHOD is not "NONE" or "EMAIL"	234



## Attributes Validated

Attribute	Rule(s)	Invalid Values
EMAIL_ADDRESS	Convert email into upper case and trims the spaces, then compare the top level domain name with lookup values to validate. Also ensures an email address exists when notifications need to be sent.	16
PHONE_NUMBER	Compare the area, country and exchange code with lookup values and validates the record	879
CUSTOMER_FORM_LINK	Invalid if Form Hyperlink is NULL when Include One Click Web Form is enabled	30,095
CONTACT_METHOD	Validate using lookup containing the list of expected values	105
STOP_TYPE	Validate using lookup containing the list of expected values	3
LOB	Validate using lookup containing the list of expected values	6
EMAIL_NOTIFICATION_TYPE	Validate using lookup containing the list of expected values	5



## Attributes Standardized

Attribute	Rule(s)
EMAIL_ADDRESS	Trim the email
PHONE_NUMBER	Concatenate phone area code, exchange code, line number when it is not null and put in a standardized format (area) exchange-line [extras]
CUSTOMER_FORM_LINK	Trim whitespace if not null
INTERNAL_NOTES	Trim whitespace if not null
ADDITIONAL_EMAIL_BODY_TEXT	Trim whitespace if not null
EMAIL_SUBJECT	Trim whitespace if not null
WEBSITE_DETAILS	Trim whitespace if not null



## Final Cleansed Data

1. Data was denormalized into the original format (with multiple email/phone values separated by semicolons).
2. This cleansed data was analyzed by DQLens™ Profile again to validate the results.





## From POC to Production

We look forward to the opportunity to partner with your team to turn this POC into a production solution.

The following are key areas that we might explore to do so:



### Value

We will work with your team to analyze your processes, systems and data to identify areas of opportunity where data management and analytics would enable improvement of key process metrics that result in increased revenue, reduced costs, and improved efficiency.



### Governance

We will work with your team to maximize data accessibility and value by cataloging metadata, defining standards, and establishing reports and success metrics. Data governance and stewardship processes may be operated by your team or ours.



### Rules

Several assumptions were made during this POC related to the business rules applied to the data. We will work with your team to craft rules that comply with data quality requirements and help define them if unknown.



### Automation

The data for this POC was provided in Excel format. To automate a production data flow, we will source the data directly from the system of record for processing. We will also tailor final format of the data for easy downstream consumption and/or publication back to the system of record.

# THANK YOU

EXL DQLens™



© 2021 ExIService Holdings, Inc. All rights reserved. For more information go to [www.exlservice.com/legal-disclaimer](http://www.exlservice.com/legal-disclaimer)

EXL (NASDAQ: EXLS) is a leading operations management and analytics company that designs and enables agile, customer-centric operating models to help clients improve their revenue growth and profitability. Our delivery model provides market-leading business outcomes using EXL's proprietary Business EXLerator Framework®, cutting-edge analytics, digital transformation and domain expertise. At EXL, we look deeper to help companies improve global operations, enhance data-driven insights, increase customer satisfaction, and manage risk and compliance. EXL serves the insurance, healthcare, banking and financial services, utilities, travel, transportation and logistics industries. Headquartered in New York, New York, EXL has more than 27,000 professionals in locations throughout the United States, Europe, Asia (primarily India and Philippines), South America, Australia and South Africa.

## EXLservice.com

### GLOBAL HEADQUARTERS

320 Park Avenue  
New York, New York 10017  
T +1 212.277.7100 F +1 212.771.7111

United States • United Kingdom • Czech Republic • Romania • Bulgaria • India • Philippines • Colombia • South Africa