

A man with a beard, wearing a grey sweater and dark trousers, is walking through a server room aisle. He is holding a silver laptop. The server racks on both sides are illuminated with blue light, creating a futuristic atmosphere. The perspective is from the end of the aisle, looking towards the man.

virtusa[®]
Accelerating Business Outcomes

ENHANCING DATA QUALITY

With Machine Learning &
Statistical Techniques



Data is the fuel that drives today's organizations. Ensuring top data quality continues to be a problem that plagues enterprises across all segments. Poor data quality has serious ramifications on business including lost customers owing to poor service, compliance issues due to inaccurate regulatory reporting, increased risk in fields like healthcare where timely and correct data is the difference between life and death, manual workarounds causing operational inefficiencies, etc. Owing to inadequate data governance mechanisms, shortage of data quality champions, and a variety of other system limitations, organizations still struggle to clean up and exploit their most important asset.

The cost of correcting data once it's being used across the enterprise is 100 times the cost of correcting it at the point of entry.

The cost of detecting and correcting data quality errors runs into millions. This problem is more pronounced with the growing volumes of data being generated daily from a variety of systems. Enterprises relying on correct data require a robust enterprise Data Quality Management solution, fully integrated with an enterprise data governance framework, to help rapidly identify and correct erroneous data at minimum costs.

Virtusa brings its consulting expertise that analyzes how data is being used by the business and the systems that it operates on. This helps contextualize data and detail out how data quality impacts the organization. This analysis helps us prescribe the right path forward and help orchestrate changes across business and IT systems to ensure enterprises are working with clean data.

Analyze

We work with our client SMEs to understand and analyze data problems across the company to determine the severity of the problem. This includes identifying business processes and systems involved in the creation, maintenance and deletion of data. We perform data profiling and identify data quality metrics. This also helps understand the extent of the data quality problem and be able to baseline and track it.

Recommend

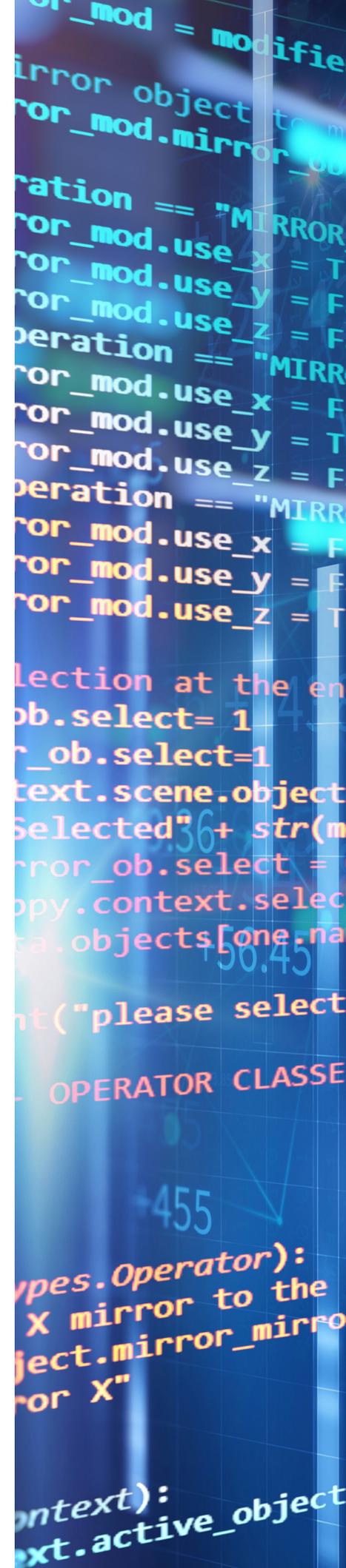
A detailed set of actionable recommendations calls out the business processes and the associated data quality challenges. This includes results of the analysis and profiling exercise carried out with a set of recommendations to address existing data quality problems, and also ways to prevent the problem in the future including process changes.

Tools

We also recommend a specific set of data quality tools and accelerators that can be deployed to help implement the recommendations. These tools can be leveraged in existing processes or may require process changes to accommodate data quality initiatives.

Pilot

We will work to implement our recommendations for a specific domain or subset of data to prove out the viability of the approach, validate the results with stakeholders before it can be rolled out enterprise wide.





We believe that data quality is everyone's responsibility and ensuring 'good data' is a way of life.

We help clients clean up their data by bringing them a unique approach that standardizes the way they collect and store data while embedding real-time data quality mechanisms into every process and application. This approach is further bolstered by our expertise in leveraging machine language and statistical methods to remediate and resolve data quality challenges at a rapid pace.

Our four-step approach brings together our experience in data governance with relevant technology to identify and correct bad data, enable data quality champions across the organization, and bolster innovation and growth. Virtusa introduces machine learning in data quality to help understand data attributes, generate recommendations, and convert the same into business rules.

Virtusa ensures continuous improvement in a way that every user is responsible for data quality.

Standardize & optimize data governance

- Leverage data discovery tools to understand current data quality management processes , data lineage and data errors
- Define processes in place to optimize data quality

1

Enable data quality champions

- De-centralize data quality control
- Enable business and non-technical users to easily profile data

2

Automate data quality with a robust DQ framework

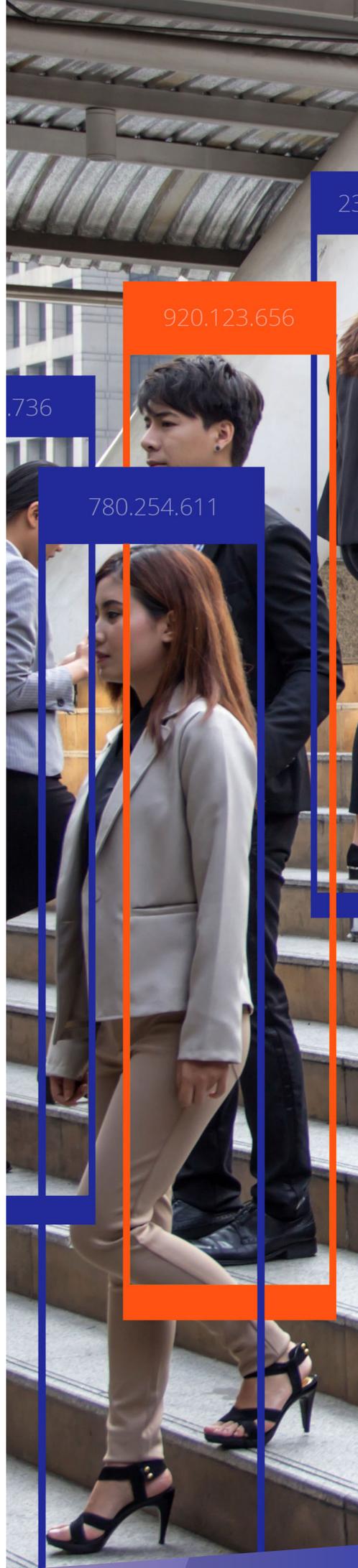
- Implement Data Quality framework with configurable rules to help remediate and cleanse data in real-time
- Support for heterogenous platforms

3

Leveraging machine learning & statistical methods

- Leverage ML to understand data context
- Proactive resolution of data errors for a variety of problems through software

4





Virtusa's unique consulting led approach includes leveraging frameworks, tools, and also leverages machine learning and statistical techniques to identify and address data issues including duplicates removal, outlier detection, data anomaly detection, fuzzy matching and cross tabulation across textual and numeric content.

Numeric Data

- **Outlier Detection** – Detect and exclude data that deviates drastically from the average of the data set
- **Anomaly Detection in Time Series Data** – Identify data points in a time series that do not confirm to the expected pattern of the series
- **Dimension Reduction** – Reduce the number of random variables within a given data set with techniques like Principal Component Analysis

Textual Data

- **Spelling Correction** – Automatically correct spellings for textual data by referring to external dictionaries/ data sources
- **Fuzzy Matching and Cross Tabulation** – Processes word-based matching queries to find similar phrases from a database and highlight discrepancy across multiple columns
- **Standardization of text entities** – Automatically identify and convert text entities into standard forms by referring to a dictionary

Mixed Data

- **Mixed Value Imputation** – Replace missing data with estimate values based on an aggregate of the data set such as mean, media, or mode
- **Duplicate Removal** – Detect and remove duplicates in multivariate data using distance measures
- **Unstructured Data Processing** – Extract structured data entities such as names, dates, numbers etc. from an unstructured data set

Virtusa enables 30M USD savings for a telecommunication services client by correcting errors during order taking

Incorrect order entry practices at the front end were causing acute inefficiencies in order processing. Nearly 35% of the the orders being taken had incorrect entries. This resulted in poor customer experience and spiraling costs owing to credit notes being raised due to incorrect pricing, third-party cost slippages, and other ancillary charges. It cost nearly 400 USD to correct every order and took 3 weeks each, resulting in a 25% increase in delivery costs.

Virtusa's helped establish data governance practices that helped the client realize an overall savings of 30M USD while ensuring continuous improvement of data quality



Reduced order fulfillment and rework costs by over 15M USD



Saved over 10M USD from credit notes owing to incorrect pricing information and incorrect data entry



Saved on other third-party charges that arose due to data inaccuracy





Virtusa helps leverages machine learning to help improve compliance for a leading American bank

As a part of their compliance, the bank is required to generate several regulatory reports that need to be periodically filed with the federal government. However, poor data quality was resulting in incorrect reports being filed. The client needed to address poor data quality problems and correct the same across the enterprise.

Virtusa established a data quality framework that analyzed previous incorrect and correct reports to identify changes that needed to be made to ensure correct data, and also help predict possible data quality errors using AI/ML. Leveraging various machine learning and statistical methods, Virtusa was able to identify and correct errors.



Deployed a data quality scoring mechanism to identify poor data quality



Significantly reduced reconciliation efforts and improved operational effectiveness



For more information write to: marketing@virtusa.com

About Virtusa

Virtusa Corporation (NASDAQ GS: VRTU) is a global provider of Digital Business Strategy, Digital Engineering, and Information Technology (IT) outsourcing services that accelerate our clients' digital transformation journeys. Virtusa serves Global 2000 companies in Banking, Financial Services, Insurance, Healthcare, Communications, Media, Entertainment, Travel, Manufacturing, and Technology industries. Virtusa engages its clients to defend and grow their business by introducing innovative products and services, creating operational efficiency using digital labor, developing operational and IT platforms for the future, and rationalizing and modernizing their IT applications infrastructure.

Founded in 1996 and headquartered in Massachusetts, Virtusa has operations in North America, Europe, and Asia.

© 2019 Virtusa Corporation. All rights reserved.

www.virtusa.com