zensar

A Ground-up Approach to **Data Quality Management**

White Paper

On January 1st, 2019, Cathay Pacific's first-class fares, listed for travels from Vietnam to anywhere else, were extremely low. While the usual business-class fare for a Vietnam to New York return travel is in the range of $16,000, passengers could book the same for about $675 on January 1st, 2019. The glitch remained active for a few hours, and many customers booked their travels at the surprisingly low fares. Moreover, contrary to expectations, the airline chose to honour the booked tickets and took the hit on the revenue. This is just one example of how poor data qualtan hurt a business.

Global businesses are suffering losses regularly because of poor data quality or incorrect data, and most of such losses are not this obvious. According to Dun & Bradstreet, poor data quality may cost businesses 15-25% of their revenues.[1] It also results in loss of productivity, opportunities, and may lead to regulatory violations.

Challenges in Managing Data Quality

In a PwC survey, ~40% of executives working in US companies (revenue of $500MN or more) cited poor data quality as the most challenging obstacle in monetizing data. Upon comprehensive research and discussions with data leaders from across organizations, we have identified the following reasons for the inherent complexity of data quality management:

01 Defining data quality is not easy:

It is probably the trickiest part of the overall process. Data quality can mean different things to different people within the same organization. Consider a scenario where you have customer data in the CRM database. A sales executive may call a customer record complete if they have an email address and phone number of the customer, but another department would call the same record incomplete for the lack of an address or preferred communication channel.

02 Late emphasis on data quality and governance:

Businesses usually think about data quality and governance years after setting up their processes and operations, once they have accumulated a considerable amount of data. In such a scenario, everything they do to manage data quality is reactive and leaves them playing catch-up with the problems.

03 The high cost of data quality management tools and making the business case:

This follows loosely from the first point. Since defining and measuring data quality is tricky, data quality management proponents find it challenging to make a business case for it. The high cost of data quality tools adds to this problem. As per Gartner, the average annual spending on on-premise data quality tools is $208,000, preventing their widespread adoption.[3]

04 Increasing variety and complexity of data generated:

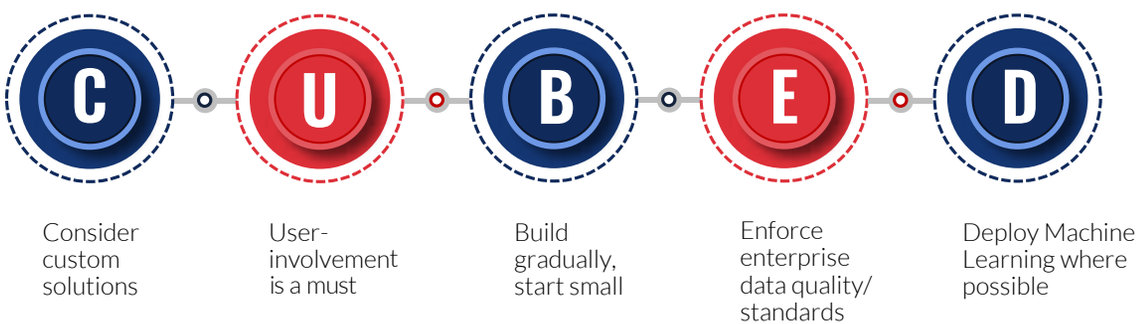
This is a known challenge with every activity in the data value chain, be it ingestion, integration, mining, or visualization. The increasing variety and complexity require a fresh and more flexible perspective and approach to data quality management.

05 Data Quality Management is much more than technology:

In discussions with clients, we have often come across this notion that managing data quality is something that concerns IT, more precisely, the data team. However, an in-depth assessment will show you that nearly all processes and operations impact data quality in any business. If the cashier in a department store makes a mistake while entering the customer's data, it will impact data quality. Similarly, if the person designing the application to capture inventory data does not ensure a basic validation of data being entered in the application, that can also cause data quality issues. **Data quality management cuts across people, processes, and technologies of modern enterprises.**

Tackling Data Quality Management Challenges the Zensar way

Based on our discussions with several clients and data leaders, we have developed the CUBED framework to help businesses manage their data quality (DQ) better.



01 Consider building a custom DQ solution when looking for options:

DQ management initiatives often start with data leaders looking at several commercial DQ solutions, weighing their pros and cons, and finding sponsors for the expected annual spend on the solution. This does not have to be the only way out, since several open-source applications and libraries can be leveraged to build a custom data quality solution suited to the organization's need.

With increasing cloud adoption, custom DQ solutions have become easier to implement. One can use the Cloud vendor's services and build a solution from scratch with nominal investment. With regular use and upgrades, the solution can be graduated into a commercial-grade application within no time.

02 User-involvement is a must for DQ initiatives' long-term success:

People who understand a business's intricacies can provide extremely valuable inputs regarding the quality of data if they are engaged in the right manner. Many data and analytics initiatives fail to yield the required results as teams developing the solutions do not engage with business users or seek their inputs. Often, a lack of involvement and inputs from business users results in developing a technically robust solution that is difficult to adopt.

03 Build gradually, start small:

Often, businesses try a big-bang implementation to tackle the data quality problem and get overwhelmed with the problem's scale. We suggest starting with one business unit, master data component or process. This approach allows businesses to use lessons and best practices from one iteration in the subsequent iterations. Moreover, small-scale implementation enables quick results and feedback.

04 Enforce enterprise DQ definitions/standards:

Businesses need to define data quality standards at an enterprise level to avoid different data quality interpretations by different people. These standards must be adhered to while designing applications that bring data into the system. Maintaining proper data quality should also be embedded in the culture of the organization. Establishing a robust enterprise data governance practice would provide the required boost to any initiative.

05 Deploy Machine Learning (ML) where possible:

Lately, involvement of ML in data quality management has increased. Nearly all commercially available solutions seem to have built-in AI capabilities. The ML methods are being used across the different DQ management activities, i.e., data enrichment, data correction, matching and deduplication, and data protection.

Leading Insurer Benefits from Zensar's Data Quality Management solution

Struggling with poor customer data quality

We learned about the client's struggle with data quality while planning a data warehouse implementation with them. They had customers across over ten geographies and multiple business lines, i.e., life, non-life, pension, etc. Several mergers and acquisitions had also brought in data from various systems and entities. As usual, their emphasis on data quality and governance was late in their data & analytics journey.

Our analysis showed that most of their data quality issues emerged from different CRM or policy management systems across geographies, each following different standards & regulations. Depending on the local needs and regulations, some business lines captured more data from each customer than other business lines. There were obvious variations in acceptable address, contact information and identification document formats in different countries.

Poor customer data quality resulting from the reasons listed above prevented the client from getting a unique record for each customer. It resulted in less-effective marketing campaigns, as the preferred mode of contact (email, phone number etc.) was unknown for many customers. Moreover, the marketing team was unable to identify the cross-sell/up-sell opportunities efficiently. To monetize customer data, the client needed a 360° view of their customers, which was difficult to achieve with such poor data quality.

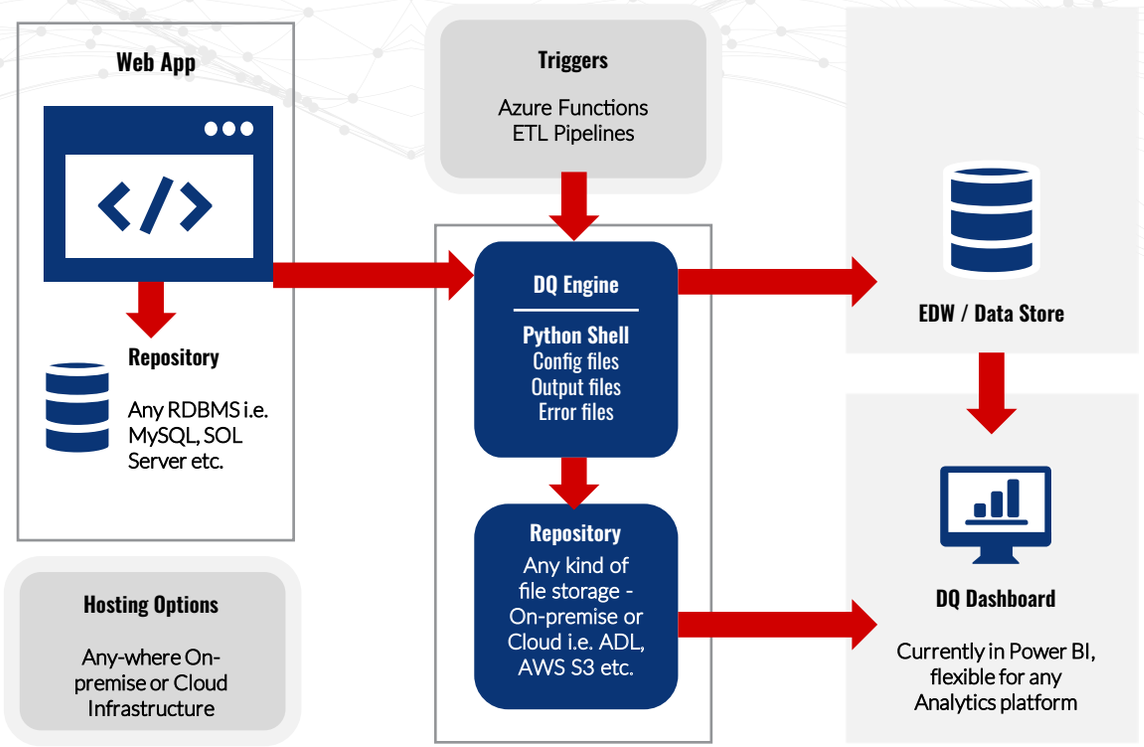
Considerations for the solution

The client considered several commercial data quality solutions. However, the cost of licenses, subscriptions, and annual support of such applications deterred them.

As an alternative, they considered screening data at every source application. However, it did not work out as there were too many source systems, most owned and supported by the vendor. The given approach would have necessitated fixing the data for existing customers as well, requiring considerable additional effort.

The decision to go with iDQ

Given the client's unique needs and budget constraints, we suggested implementing iDQ aligned with the CUBED framework. iDQ is Zensar's proprietary data quality solution. It is platform-agnostic, entirely customizable to clients' needs and technical landscape. Since it leverages open source technologies and frameworks, the maintenance and support costs are lesser than any other commercial solution. It is designed to ensure business users' involvement in implementing and maintaining data quality.



iDQ Reference Architecture

COMPONENT	DESCRIPTION
Rule Repository	It contains the metadata like rule name, error message and configurable parameters of the data validation rules.
DQ Engine	It contains the code base of functions that implement the data validation rules.
Configuration Files	Several configuration files help determine what rules are to be applied and on which attributes. They also store the parameter file that stores patterns, list of values (allowable or non-allowable).
Web UI	Web UI chooses the data validation rules to be applied and updates the configurable parameters with data validation rules.
Enterprise Data Warehouse/ Data Store	It is the single source of truth for the golden records of customers. In the case of EDW, it can also have data related to policies and claims.
Data Quality Dashboard	It visualizes data quality metrics related to completeness, conformance, error details, and overall data quality metrics.

Business Benefits of iDQ

Owing to the modular design of the solution, iDQ was easily configured to cater to the varying needs of different geographies. The Web UI allowed continuous business engagement in the DQ management process by allowing key business stakeholders to review and modify business rules when required. The data quality dashboard provided a real-time view of the overall data quality and highlighted the key areas of concern. The solution's detailed error logs enabled the client to fix more than 80% of data quality issues within a few months of implementation.

The client was able to achieve all this at a much lesser cost compared to that of any other commercial data quality solution.

Conclusion

Bad data has been costing businesses a fortune and handling it should be a top priority for them. Even a minimal increase in data quality can reduce operational risks and costs significantly. A proliferation in open source technologies and cloud-based services has made it extremely convenient to implement a custom solution tailored to the businesses' specific needs. In such a scenario, businesses that have been stalling their DQ initiative because of the challenges listed in this paper or some other hindrance should reconsider investing in Data Quality using the CUBED framework.

References

1. https://www.dnb.com/content/dam/english/dnb-data-insight/DNB_Past_Present_and_Future_of_Data_Report.pdf
2. https://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement/
3. https://www.gartner.com/smarterwithgartner/how-to-stop-data-quality-undermining-your-business/
4. https://www.wsj.com/articles/ai-efforts-at-large-companies-may-be-hindered-by-poor-quality-data-11551741634
5. https://www.kdnuggets.com/2020/03/bad-data-affecting-organizations-operational-efficiency.html
6. https://www.ibmbigdatahub.com/infographic/four-vs-big-data
7. https://onemileatatime.com/cathay-pacific-honoring-mistake-fare/
8. https://www.bbc.com/news/business-46732952

For more information on Zensar, please contact: connect@zensar.com | www.zensar.com

Authors:
Shivam Kumar
Business Development Manager,
Advanced Analytics and Data Science
shivam.kumar@zensar.com

Muneesh Bajpai
Principal Solution Architect,
Advanced Analytics and Data Science
muneesh.bajpai@zensar.com

zensar

An RPG Company

We conceptualize, build, and manage digital products through experience design, data engineering, and advanced analytics for over 145 leading companies. Our solutions leverage industry-leading platforms to help our clients be competitive, agile, and disruptive while moving with velocity through change and opportunity.

With headquarters in Pune, India, our 10,500+ associates work across 30+ locations, including Milpitas, Seattle, Princeton, Cape Town, London, Singapore, and Mexico City.

For more information please contact: **velocity@zensar.com | www.zensar.com**