

Data Quality Governance for Data Sourcing and Data Analytics Team

Anand Kumar, Arunam Technologies LLC, Columbus, OH

ABSTRACT

Having data that are consistent, reliable and well linked is one of the biggest challenges facing by financial institutions. The paper describes how SAS data management platform helps to connect people, process and technology to deliver consistent results for the data sourcing and analytics team and minimize the cost and time in the development life cycle. The paper concludes with the best practices learned from various enterprise data initiatives

INTRODUCTION

Having data that are consistent, reliable and well linked is one of the biggest challenges facing by financial institutions. over the past few years, we have had a revolution in measurement, that has allowed banks to understand in much more detail what their customers are doing, what their processes are doing, what their employees are doing. That tremendous improvement in measurement is mandating the data sourcing team to profile, cleanse, link and format data consistently to make high quality data available to the analytical model development team. The quality data not only improves the employee productivity but also promises great results to the banks to outperform competitors. The prerequisites, of course, is the SAS data management infrastructure; the ability to improve data and the data quality governance framework and the skills

CURRENT CHALLENGES

Not understanding the data and anomalies by the data sourcing team would create a big daunt on the project management and the deliverables. Typical data sourcing team priorities are how many data sources has been staged and integrated with in a defined timeline and often worry less on concentrating the data quality side of deliverable. Eventually the end user identifying the defects during the acceptance testing would create a costly re-work for the ETL team and in turn cause delays in the model development. Many data sourcing projects lack the data quality related tasks in the project plan when not measured or assessed properly would end up in date slippage, team frustrations and inaccurate results in business decisions.

Common data quality problems:

Data and its format
Temporary revision to defaults
Missing, default and overloaded values,
Missing metadata and reference data
Gaps in time series
Not knowing how much history is required

THE FINDINGS

Having a data quality framework and governance around the framework is critical to manage the defect free data as it flows through the data supply chain. There is no uniform standard for content and formats which is the cause for the same field with different format across applications and the reference data is never managed for reusability which is the cause for inconsistent codes and description across applications. Some of the key observations at various client implementations not meeting the data quality program goals are:

- Not enforcing the standards or following the standards.
- Easy attitude on the data quality issues
- No education on the importance of data quality and end user on-boarding training
- No roles and responsibilities on the data quality program charter.
- No clear goals on how the data security and access is managed for the data quality program initiatives.
- Decentralize or centralize the data quality program is unclear and how to enforce standards and enable collaboration at an enterprise level program.
- Confusion on what tool to use for data quality and cleansing efforts in the data sourcing and integration projects.

- No standard templates for delivering the data quality assessment results and no follow up actions on the improvement opportunities.
- No transparency in workflow procedures on how the quality issues are highlighted and escalated for a resolution and the remediation steps.

A PROMISE: SAS DATA MANAGEMENT PLATFORM

The data quality is pervasive and expensive. It is an important problem but when not resolved would cost hundreds of billion dollars each year. Resolving data quality problems is often the biggest effort in data sourcing and model development teams. We need a better tool which can measure the data quality and leads us to improvements in processes. SAS data management platform provides a suite of products to develop efficient processes that can assess the defects, clean, standardize, master and manage your data.

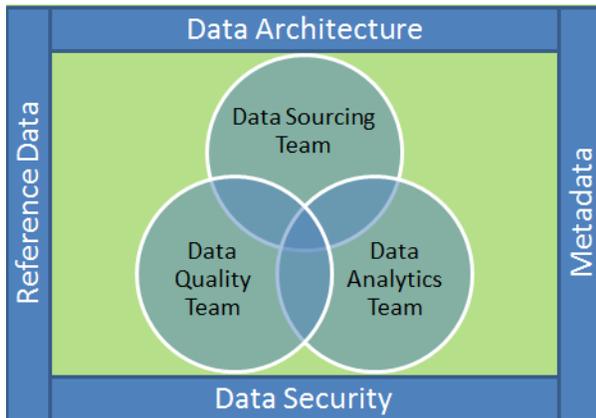


Figure 1: Team collaborations

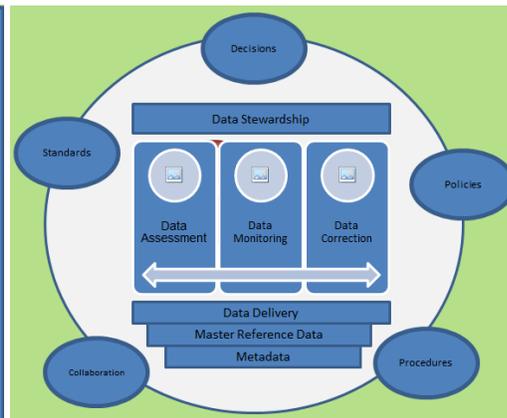


Figure 2: Data quality governance

The data quality governance defines how the program is developed and how it evolves and promotes innovation as it collaborates across the data quality, data sourcing and data analytics team in an enterprise. SAS data management platform enables to bring the people, process and technology together to deploy a robust data quality program for an efficient data delivery to the data consuming teams. As the data passes from the data producing applications, where it is collected, to the IT group, the key business data terms or the modeling variables can be defined in the SAS data management product suite Business Data Network and the business data terms can be profiled to diagnose any defects and to discover business rules, cleansing rules and matching rules so the powerful components of SAS data management suite: business rule manager, quality function nodes and entity resolution function nodes can be utilized to automate the improvement process. The Web Studio component of the SAS data management platform displays the before and after cleansing results in the form of data quality dashboards and provide the data stewards and executives at a glance summary view of the data quality scores at key data domains.



Figure 3: Sample data quality dashboards generated out of SAS data management monitor repository

BEST PRACTICES

Some of the best practices from the data quality program implementations at various client sites are:

- The data profiling reports are generated and published to the sourcing and analytics team so the consistent answer is available on any data discrepancy question.
- Dashboards on key data domains are published with the data quality scores so the executives and model development teams are aware of the data quality of the consuming data (customer, product, loan sourcing and servicing)
- Data stewards are identified at each data domain and their duties are well defined in the data quality governance charter.
- The education on the importance of data quality and end user training are part of the data quality program.
- The roles and responsibilities on the data quality program charter are well defined.
- A single collaboration portal is used to track the progress of the data quality program and to engage various groups on the activities and new initiatives.
- A data management tool is in place for the enterprise data quality program automation.

CONCLUSION

Though a minimal data quality processes in place now-a-days in the data sourcing team and analytics team, a good data management tool infrastructure and governance has to be built around the existing data quality processes to make the program reusable across the enterprise teams in a consistent repeatable manner. Having a product like SAS data management platform rightly helps to keep the data quality program up and running engaging multiple teams across the enterprise with processes automation. There is a more value in putting the data quality scores along with KPI and analytical model scores on a single dashboard so the top management gets a complete view of data defect, performance and modeling results scores in one place.

RECOMMENDED READING

Nancy Rauch, Michael Ames, Wilbram Hazejager, SAS Institute "What's New in SAS Data Management" Proceedings of the SAS Global Forum 2012 conference. Available at <http://support.sas.com/resources/papers/proceedings12/110-2012.pdf>

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Anand Kumar
Arunam Technologies LLC
247 Kastlekove Dr
Lewis Center, OH 43035
Work Phone: 614-787-0188
Fax: 1-8555ARUNAM
E-mail: anandj@arunam.com
Web: www.arunam.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.