

Data Quality Governance

For Analytics Teams

Potential
of One

Power
of
All

Data Quality Governance for Data Analytics Teams

Anand Jagarapu

Arunam Technologies, LLC.

Abstract

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 current paper attempts to show, how a strong data quality governance process with SAS® data management tools improves the cross team collaboration and promote innovation in the data analytics teams.

Key Challenges

List of key challenges in any typical data quality program initiative

- No roles and responsibilities in the data quality program charter.
- Easy attitude on the data quality issues
- No corporate education on the importance of data quality and continuous end user training.
- 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 to deliver 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 remediation steps.
- No link to the data quality and the reliability and repeatability of analytical models.

Data Quality Governance

The data quality governance framework as seen in Figure 1, 2 defines how the program evolves and promotes innovation as it collaborates across the data quality, data sourcing and data analytics team in an enterprise.

SAS data management platform supports to bring the people, process and technology together to deploy a robust data quality program for an efficient data delivery to the data analytics 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 governance product suite: business data network and the business data terms are 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 functions and entity resolution functions can be utilized to automate the data improvement process.

The monitoring component of the SAS data governance suite displays the before and after cleansing results in the form of data quality dimension dashboards and provide data stewards and executives a summary view of the scores at the key data domains to quantify the issues.

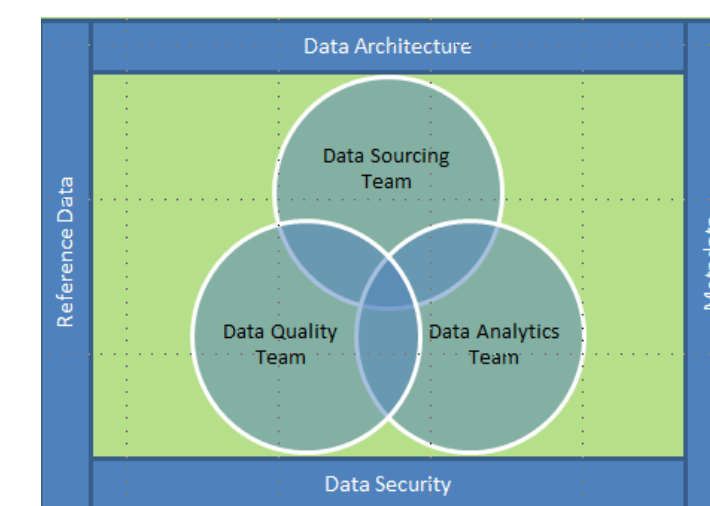


Figure 1: Team Collaborations

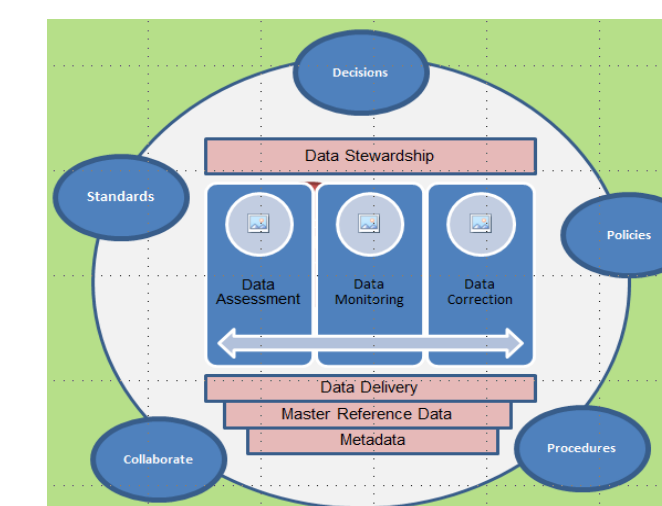


Figure 2: Data Quality Governance Framework

Results

A successful data quality governance program combined with a robust SAS data management software platform would help to capture the data quality metrics and helps to visualize how the program is running and quantify the cost justification on the overall program spending (IT and labor). It also helps the business to be more productive in minimizing the repetitive data related work.

With the reliable data, the analysts spend more time in turning the data into actionable insights applying advanced analytics and modeling.

Figure 3 show a sample dashboard view of data quality program overview with details on user onboarding, tool usage, and financial impact on the failed records and the trending patterns on the LOB data.

The data quality score linked to the enterprise key performance indicators and model prediction scores would create a collaborative work environment between the data quality, data sourcing and data analytics teams.

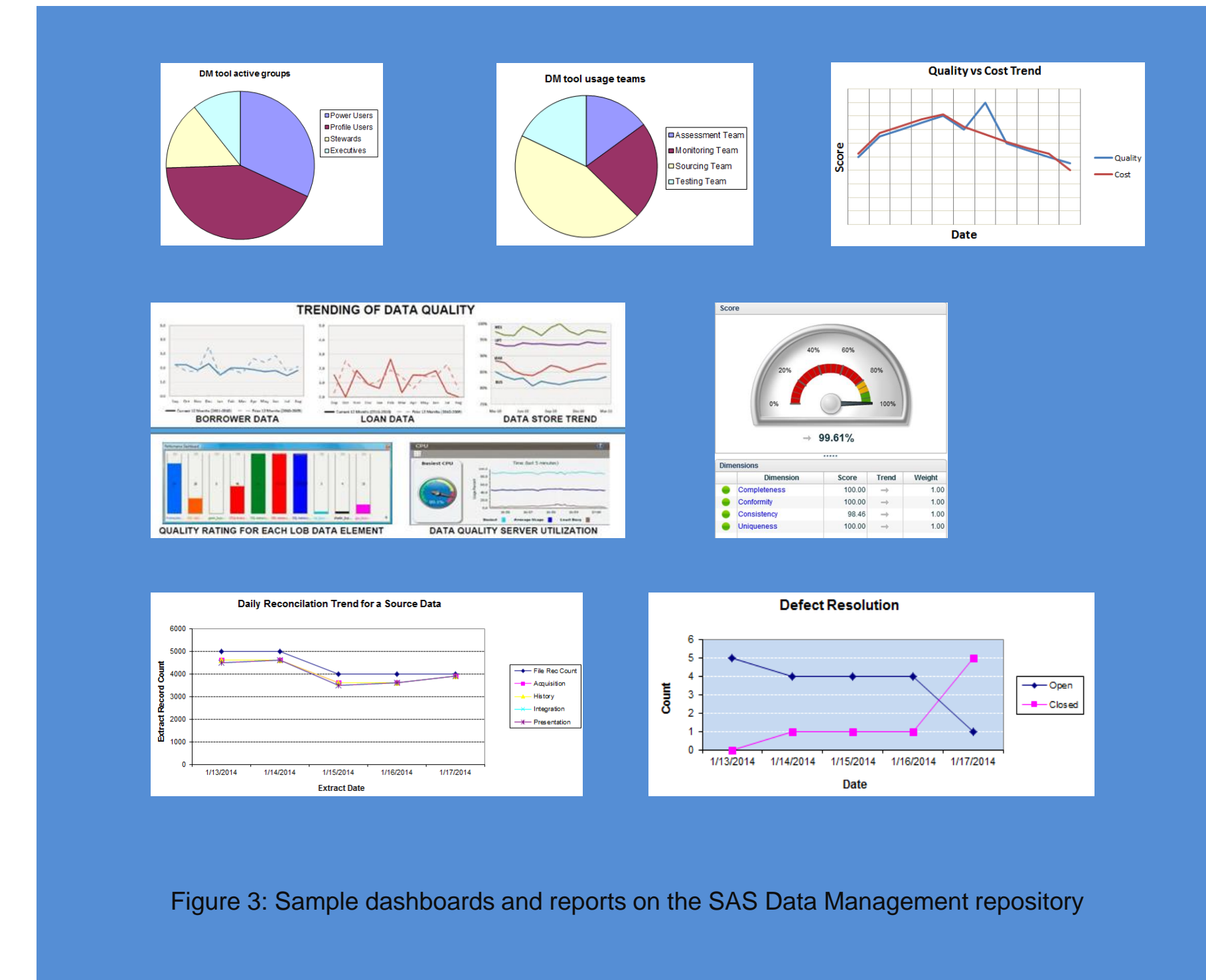


Figure 3: Sample dashboards and reports on the SAS Data Management repository

Conclusions

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 when the data quality dashboards are built and linked to the key performance indicators and modeling variables so the model development teams are aware of the issues and occupy less time in the data preparation step and stay creative in turning the data into action insights. The business can focus on frontline engagement and capabilities balancing the speed, cost and acceptance of the quality and integrated data.

References

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 Jagarapu
Arunam Technologies LLC
Work Phone: 440-227-8648



Web: www.arunam.com | Email: anandj@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.