Paper 5483-2016

Accelerate the Time to Value by Implementing a Semantic Layer Using SAS® Visual Analytics

Arun Sugumar and Vignesh Balasubramanian, Kavi Global

ABSTRACT

Although business intelligence experts agree that empowering businesses through a well-constructed semantic layer have undisputed benefits, a successful implementation has always been a formidable challenge. This presentation highlights the best practices to follow and mistakes to avoid, leading to a successful semantic layer implementation by using SAS® Visual Analytics. A correctly implemented semantic layer provides business users with quick and easy access to information for analytical and fact-based decision-making. Today, everyone talks about how the modern data platform enables businesses to store and analyze big data, but we still see most businesses trying to generate value from the data that they already store. From self-service to data visualization, business intelligence and descriptive analytics are still the key requirements for any business, and we discuss how to use SAS Visual Analytics to address them all. We also describe the key considerations in strategy, people, process, and data for a successful semantic layer rollout that uses SAS Visual Analytics.

INTRODUCTION

Digital transformation is a key priority for almost all organization, Business Intelligence, and Analytics are key components to that transformation. It is important not to cut corners while looking to accelerate analytical capabilities. Most organizations are data rich but information poor and one of the primary reasons is due to the absence of the business representation of data, the semantic layer.

Across business function and industries, poor availability and un-integrated data constitutes as the key barrier to analytics. The major oversight has been the practice of creating data silos for rapid delivery, leading to an unsustainable environment, thereby resulting in poor user experience. As organizations are racing towards establishing data lakes it becomes all the more prudent to have a strategy on how functional users are going to be enabled to leverage these data lakes for data driven decision making.

SAS® Visual Analytics enables functional users by providing the right mix of capabilities from self-service to data visualization, business intelligence, and descriptive analytics. A deployment with a tailored approach will complement the digital journey of your organization. A semantic layer will hide the underlying data complexities as well as translate all data structures to business specific terms and usage. This will eliminate the ambiguity in data usage and result in enormous productivity savings.

PLANNING YOUR SEMANTIC LAYER

Either IT, Business or both have chosen SAS® Visual Analytics as the Business Intelligence Platform for your organization. It is assumed that the size (machine configuration) and type (Symmetric multi-processing or Massively parallel processing) of the deployment has been planned according to your user base, data scope and technology stack. SAS® Visual Analytics is either being introduced or is replacing an existing BI product as the technology to provide the required analytical capabilities. The current state of digital maturity and analytics organization model determines the go forward strategy.

STRATEGY

You can pin your digital maturity in one of the four stages: Collect, Integrate, Comprehend and Innovate. Your analytics organization can fall into one of the three broad structures: Integrated, Co-ordinated (Center of Excellence) and Distributed. In the Integrated model the team serves as a hub providing analytics to the entire organization. In the Distributed model; each business service line and major function will have its own team under a team leader. In the Co-ordinated model, the members of the central team (CoE) are deployed to the business service line for a period of time. A small central group coordinates their activities.

In the hybrid model the members of the central team (CoE or Federated) are deployed to the business service line for a period of time. A small central group coordinates their activities.

It is imperative to understand your current state to determine the existing gaps in people, process, and data to arrive at a plan that would accelerate your implementation.

The Figure 1. Logical Layers, depicts the key ideal logical layers required to deliver a state of the art analytical solution. SAS® Visual Analytics is going to enable your organization to leverage the data assets for business value with the top 4 layers namely, self-service analytics, reports and dashboards, self-service discovery, and analytical models (SAS Visual Statistics add-in).

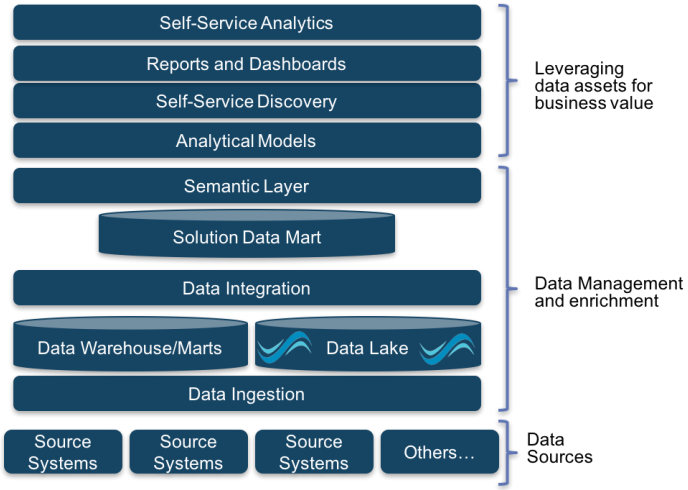


Figure 1. Logical Layers

The data availability and maturity of your organization might differ and your current state of data management and enrichment might not have all of the depicted layers. Understanding the gaps in the current state enables you to determine the level of effort and investments required in architecting your data platform.

PEOPLE

As part of the overall strategy, you have either determined or established your current analytics organization structure. The size of your organization will determine the number of individuals and the roles assigned to an individual. The following roles should be considered as an input to your analytics and data management organization staffing plan. Analytics team roles: Analytics Leader (Chief Analytics Officer), Data Scientist, User Experience Developer, Business Analyst, and Analytics Platform Administrator. Data management team roles: Data Management Leader (Chief Data Officer), Architect, Data Modeler, Data Integration Developer, and Data Management Platform Administrator. Your budget considerations should plan for adequate training for individuals assuming additional roles/responsibilities as well as complementing the internal team with a consulting partner.

The current mode of delivery, and ease of information/data availability to your information consumers and power users will determine the required changes. Having a change management plan will not only gain consensus on the desired state, but also support the transition during the delta state while SAS® Visual Analytics in being rolled out. The change management plan is a critical component that determines the adoption and success of your analytical platform. The following Figure 2. Change Management, shows the four areas, Structure (change governance team), Process, People and Culture, to address as part of your change management plan.

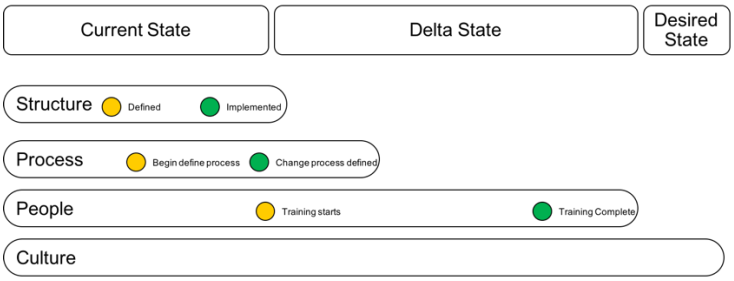


Figure 2. Change Management

PROCESS

In order to effectively introduce data-driven decision-making it is important to have an understanding of the operational process of the business and or function. This understanding would provide clarity to the various metrics, analytics, explorations and reports required to enhance and enable the process owner.

The additional component to process is the governance and working rhythm to delivery the metrics, analytics, explorations, and reports. It is imperative to have senior leadership commitment and focus, thereby mitigating potential conflicts with other initiatives, as well ensure ensuring resource availability. Having a steering and or advisory committee from middle and upper management maintains required oversight for the success of the program.

DATA

Data can either be a barrier or boon depending on the data maturity of your organization. According to a Deloitte study “One of the most frequently cited reason for not using analytics to support enterprise strategy (23%) is that the company lacks proper technology and infrastructure to capture – and then use the data”.

Data Governance

Most organizations have a very informal approach to Data Governance and have traditionally resided under development projects, leading to a very narrow focus on decision making in data-related matters. Although such a system seems to work, it ends up being the root cause for a lot of data-related issues down the road. Implementing a more formal approach to Data Governance is critical for a first-class data management practice that serves as a foundation for analytics. The below Figure 3. Data Governance Structure, portrays the roles and hierarchy for a typical data governance council.

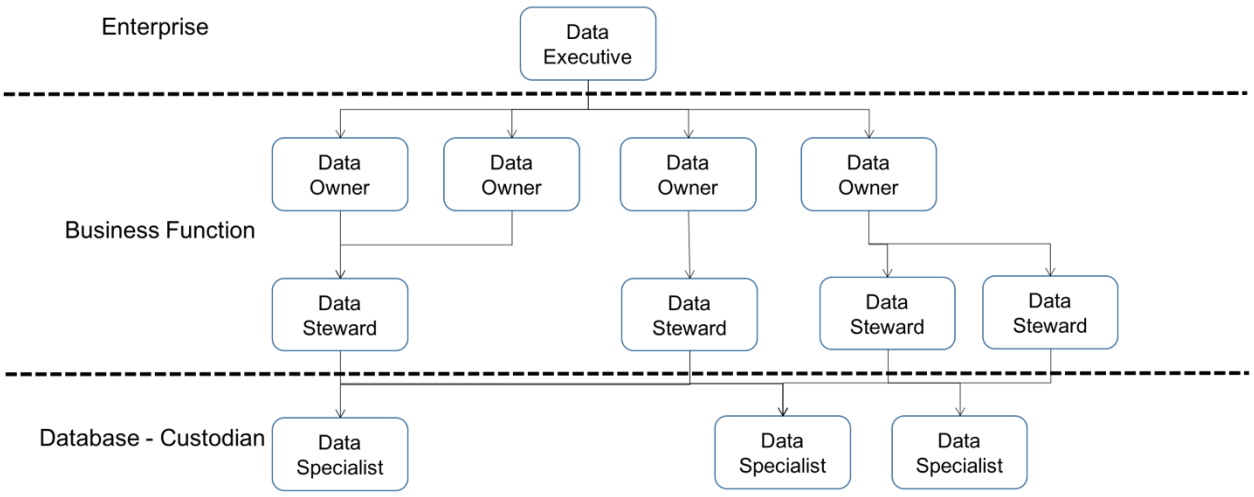


Figure 3. Data Governance Structure

Data Executive

The data executive defines the scope of governed data and drives data initiatives and investments. He or she typically leads and moderates the data council.

Data Owner

The data owners “Own” the data and typically are functional area leaders. They assist in management of data quality.

Data Steward

Data stewards are from the subject area / business Unit. They identify business requirements for data and information. They are responsible for setting priorities, maintaining a roadmap of activities to meet requirements, as well as define and classify the data (data definitions, data naming, consistent use of data, lineage and traceability of data, and etc.) including quality and security.

Data Specialist

The data specialists are from the database management area and advise on the technologies used in data management as well as implements the data delivery process.

Master Data Management

Master Data Management and Reference Data are more of a dream than a reality in most organizations. In recent years, several Fortune 500 companies have embarked on unsuccessful proof of concepts and failed attempts to establish master data management. The primary reason that most of these undertakings flop is due to a failure to address the gaps in data governance, data architecture and data quality before attempting to establish master data management. This is a case where doing things in the right order determines the difference between success and failure.

Data Quality

Most organizations today do not believe that they are ready for analytics, primarily due to a lack of trust in the quality of their data. Your data quality might not be as bad as you think, and addressing data quality is not as complicated as it sounds. The first step is to evaluate your data assets on the key dimensions of data quality namely: completeness, validity, accuracy, consistency, integrity and timeliness.

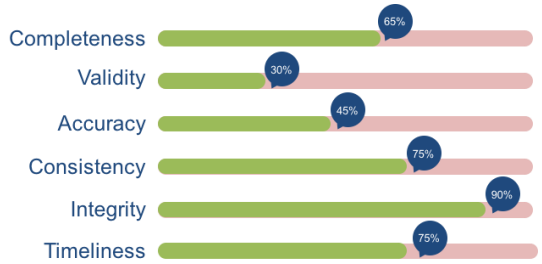


Figure 4. Data Quality Metrics

Once you have a baseline, as shown above in Figure 4. Data Quality Metrics, it becomes possible to prioritize and address the identified gaps, yielding enormous returns in enhancing the value of your data.

IMPLEMENTING YOUR SEMANTIC LAYER

INTEGRATING DATA

31% of the respondents in the Deloitte study blame un-integrated data as a challenge that impede analytics adoption. SAS® Visual Analytics star schemas can be leveraged for integrating the data. Either your organization already has an integrated dimensional data warehouse that you can leverage or you would need to build the required dimensional model. In the case of building a dimensional model we recommend you follow the Kimball’s dimensional design process:

- Select the business processes or function
- Determine the lowest atomic grain possible
- Identify the required dimensions
- Identify the facts
- Establish the required raw and calculated measures to include in your facts

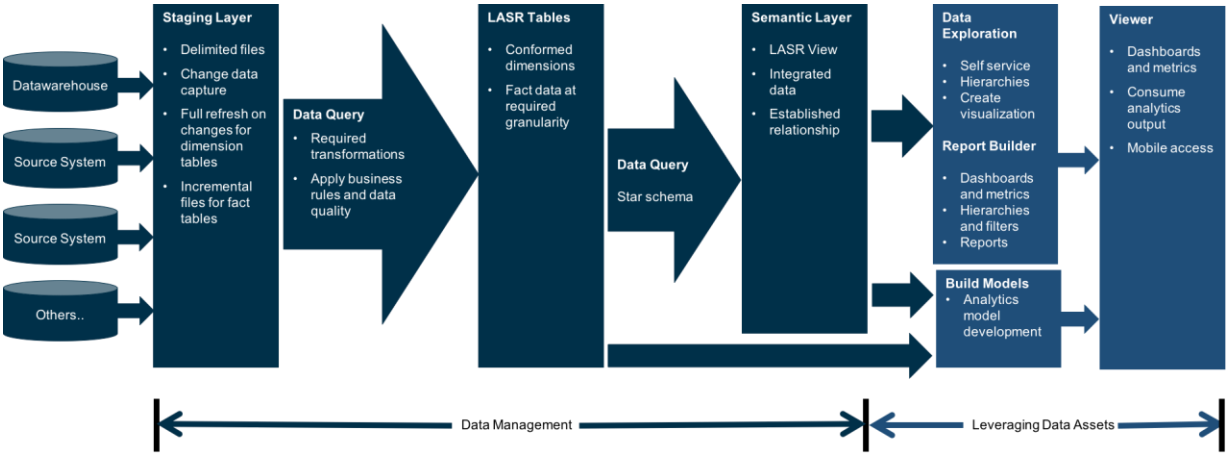


Figure 5. Data Flow Diagram

The Figure 5. Data Flow Diagram, shows a recommended approach to the data flow strategy. Once you have established the fact and dimension tables required for your star schema you would need to bring them into SAS® LASR Server as LASR tables. It would be cognizant at this stage to determine the approach for incremental data updates for your dimension and fact tables. These are typically the responsibility of the data architect and the data modeler.

Below, we discuss a few approaches to the incremental data updates:

Auto Load Dimensional Tables

The auto load process ensures synchronization between the in-memory dimension tables and the source dimension tables by comparing file and load timestamps. The default auto load schedule is set for every 15 minutes, depending on the required frequency of your change data capture it is recommended to alter your schedule to conserve system resource. The TIME_INTERVAL_MINUTES parameter shown in Figure 6. Auto Load Schedule can be changed to increase the frequency of auto loads.

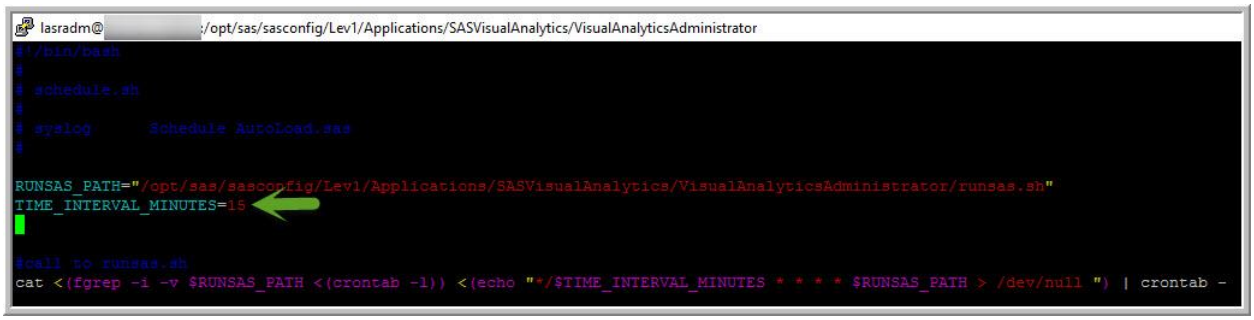


Figure 6. Auto Load Schedule

The below Figure 7: Load Status of LASR Tables, shows dimension tables that have been set for the auto load process.

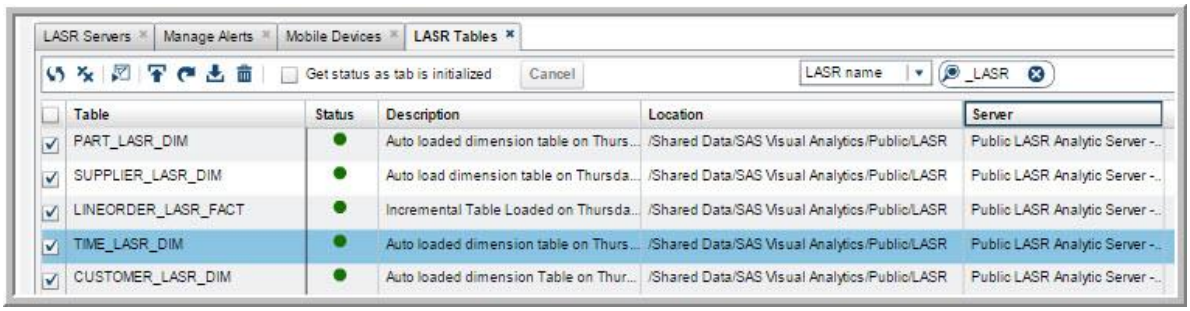
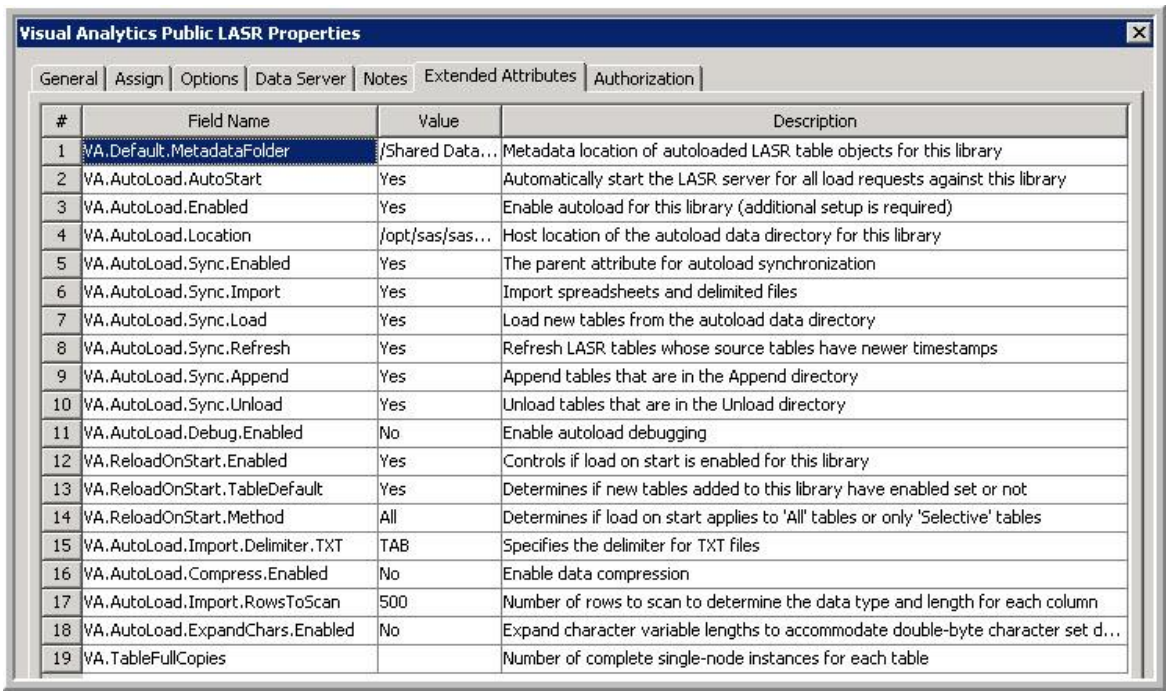


Figure 7: Load Status of LASR Tables

In SAS Management Console under Data Library Manager you can modify the the auto load options for the Visual Analytics LASR library in the extended attribute tab of the properties section. The below Figure 8. Auto Load Options, shows the various available autoload options.



#	Field Name	Value	Description
1	VA.Default.MetadataFolder	/Shared Data...	Metadata location of autoloaded LASR table objects for this library
2	VA.AutoLoad.AutoStart	Yes	Automatically start the LASR server for all load requests against this library
3	VA.AutoLoad.Enabled	Yes	Enable autoload for this library (additional setup is required)
4	VA.AutoLoad.Location	/opt/sas/sas...	Host location of the autoload data directory for this library
5	VA.AutoLoad.Sync.Enabled	Yes	The parent attribute for autoload synchronization
6	VA.AutoLoad.Sync.Import	Yes	Import spreadsheets and delimited files
7	VA.AutoLoad.Sync.Load	Yes	Load new tables from the autoload data directory
8	VA.AutoLoad.Sync.Refresh	Yes	Refresh LASR tables whose source tables have newer timestamps
9	VA.AutoLoad.Sync.Append	Yes	Append tables that are in the Append directory
10	VA.AutoLoad.Sync.Unload	Yes	Unload tables that are in the Unload directory
11	VA.AutoLoad.Debug.Enabled	No	Enable autoload debugging
12	VA.ReloadOnStart.Enabled	Yes	Controls if load on start is enabled for this library
13	VA.ReloadOnStart.TableDefault	Yes	Determines if new tables added to this library have enabled set or not
14	VA.ReloadOnStart.Method	All	Determines if load on start applies to 'All' tables or only 'Selective' tables
15	VA.AutoLoad.Import.Delimiter.TXT	TAB	Specifies the delimiter for TXT files
16	VA.AutoLoad.Compress.Enabled	No	Enable data compression
17	VA.AutoLoad.Import.RowsToScan	500	Number of rows to scan to determine the data type and length for each column
18	VA.AutoLoad.ExpandChars.Enabled	No	Expand character variable lengths to accommodate double-byte character set d...
19	VA.TableFullCopies		Number of complete single-node instances for each table

Figure 8. Auto Load Options

Incremental Load for Fact Table using Auto Load Append

There are several alternatives to appending incremental data to your fact table, one option is to use the append folder of the auto load feature. When the incremental file placed in the append folder matches the LASR table and the date timestamp is greater than the last update date timestamp of the LASR table then the data in the incremental file is appended to the LASR table. The Figure 9. Append in Auto Load Log shows the incremental records being appended and loaded to the LASR table.

```
NOTE: APPEND IN AUTOLOAD ROOT DIRECTORY

NOTE: Copying DBAPPEND.DAILY_LINEORDER to DBOUT.DAILY_LINEORDER (memtype=DATA).
NOTE: There were 988 observations read from the data set DBAPPEND.DAILY_LINEORDER.
NOTE: The data set DBOUT.DAILY_LINEORDER has 988 observations and 17 variables.
NOTE: PROCEDURE COPY used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


NOTE: LOAD TO LASR

NOTE: There were 988 observations read from the data set DBAPPEND.DAILY_LINEORDER.
NOTE: The data set LASRLIB.DAILY_LINEORDER has 1976 observations and 17 variables.
NOTE: The APPEND operation for member LASRLIB.DAILY_LINEORDER added 988 rows to the LASR Analytic Server table
      VAPUBLIC.DAILY_LINEORDER.
```

Figure 9. Append in Auto Load Log

Incremental Load for Fact Table using Staging Tables and Data Query

Another alternative is to use the staging table feature of the visual data query. This can be a best practice to leverage an existing data warehouse or source systems fact in a DBMS. This option also provides you with the flexibility to create additional calculated columns.

The input table is your fact table in the DBMS. Set the output table as the SAS LASR fact table. Check the 'Use a staging table' option and also choose your staging library as shown in Figure 10. Staging Tables.

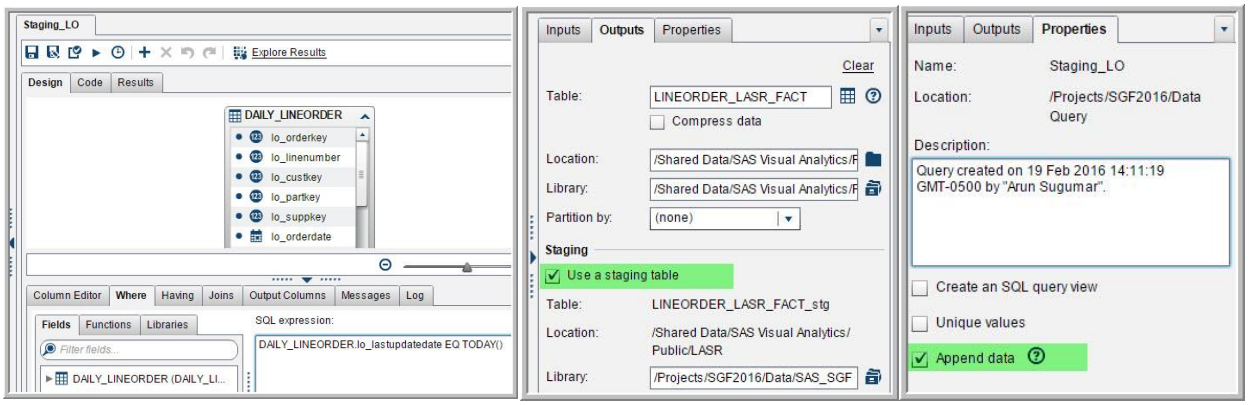


Figure 10. Staging Tables

In order to support the append feature, the staging table should either be SAS, DBMS, Teradata or Greenplum. Ensure that the Append option is checked in the properties tab. In order to query only the incremental records from the input table, you can use the where condition as shown in Figure 10. Staging Tables.

You can then schedule the data query to automate the incremental updates to your fact data using the schedule query option as shown in Figure 11. Data Query Scheduler.

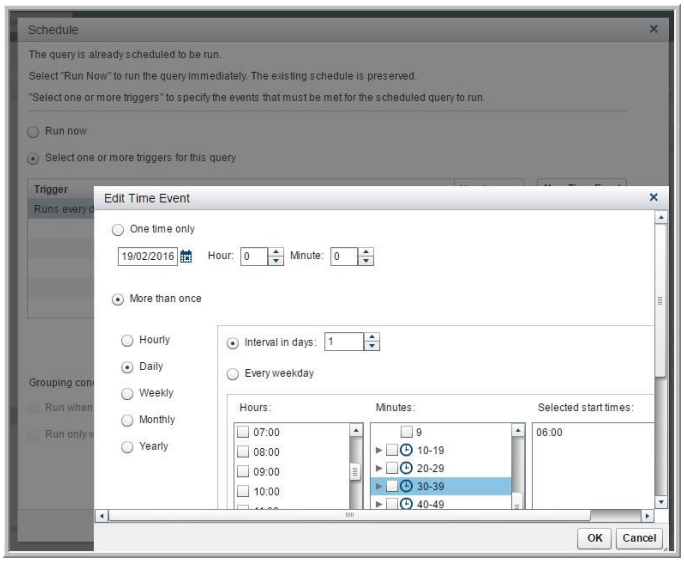


Figure 11. Data Query Scheduler

Making it easy for Users

Use the opportunity in the data query definitions to establish meaningful names and labels to the fields, define calculations, and default aggregations ensuring standardization across the users. You can also exploit the pivot by feature that makes it easy for powerful summarization as shown in Figure 12. Column Editor.

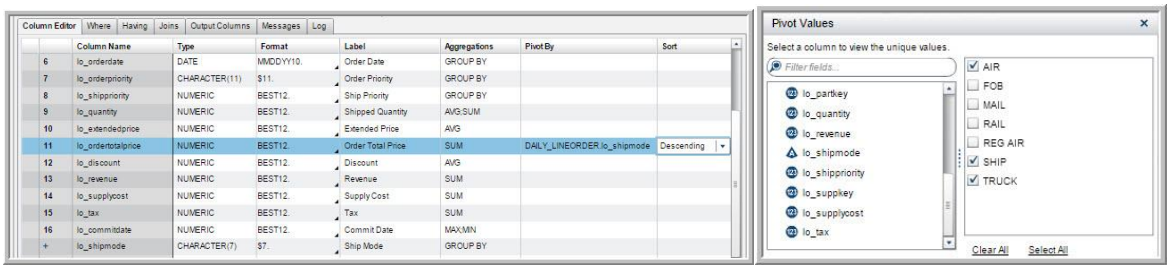


Figure 12. Column Editor

Star Schema

Establishing a star schema provides the users with the required data that is now integrated, making it easy for self service analysis. Now that you have created the dimension and fact LASR tables, you can establish the star schema as a data query view as show in Figure 13. Star Schema.

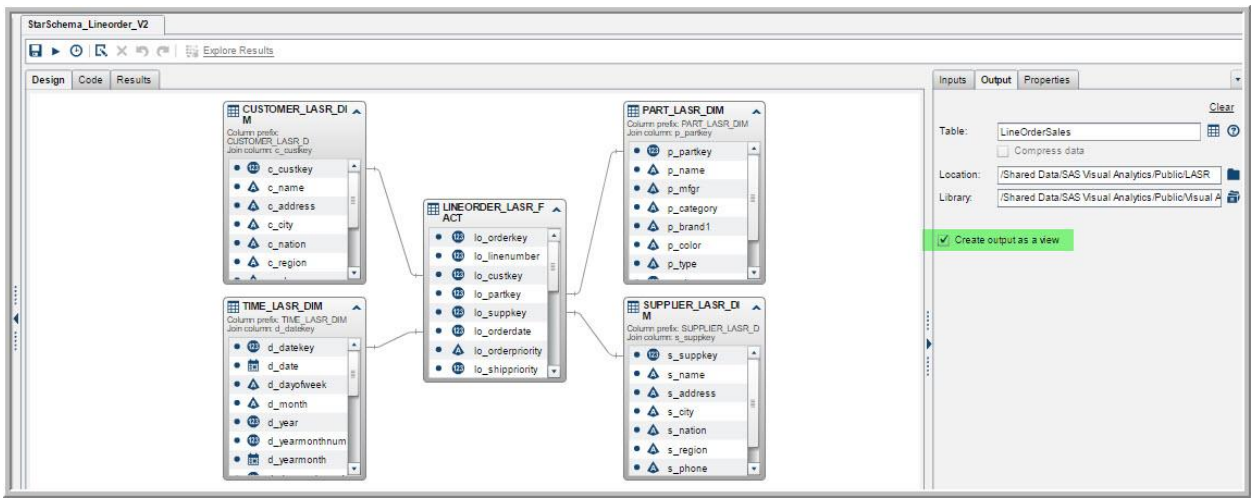


Figure 13. Star Schema

REALIZING VALUE FROM YOUR SEMANTIC LAYER

You now have your semantic layer in the form of the star schema LASR view. The attributes are named as defined in your business glossary, the data relationships have been established, the data governance approved calculated measures and metrics are made available for use, the default recommended aggregations are set and the predetermined required levels for your hierarchies are ready to go. The following value is achieved:

- User friendly access that is enabled for self-service
- Easy and timely access to required information
- Readily available business metrics
- Ability to quickly share information
- Applicable tribal data is institutionalized

All of the above results in high productivity to the user community.

EXPLORATIONS

We are going to illustrate the ease of explorations that is enabled by the semantic layer with a healthcare industry example. Interval metrics are one of the key performance indicators for hospitals and healthcare providers as part of the episode of care. There are several timestamps recorded by various hospital department visits across the patient stay. The data governance team ensures that the right timestamps are used as per the interval definitions. The calculated intervals or time spent is already made available to the user in the semantic layer as per those definitions, making data explorations quick and easy as shown in Figure 14. Exploration using Sankey.

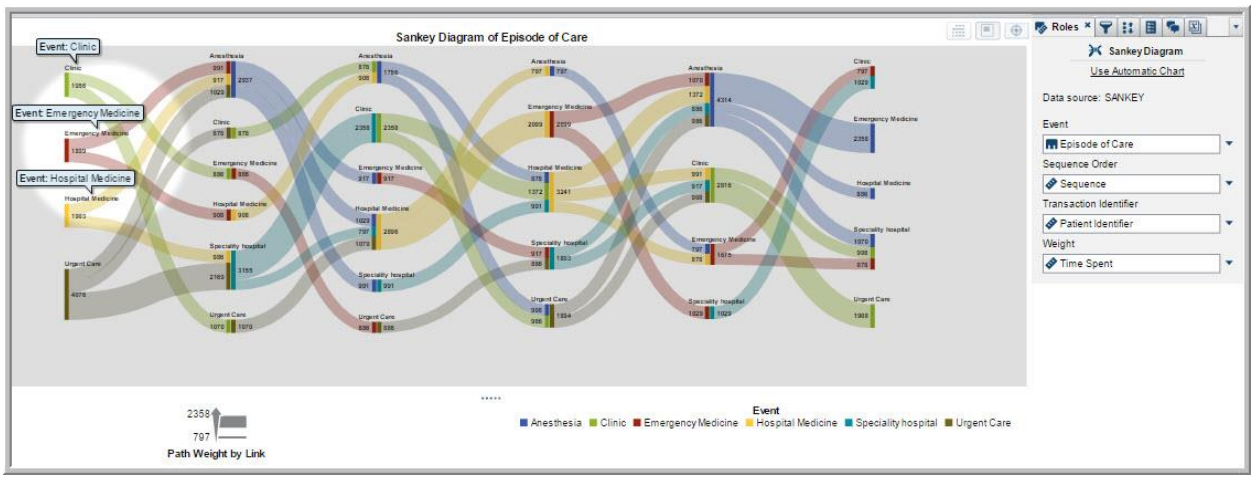


Figure 14. Exploration using Sankey

REPORTS AND DASHBOARDS

It is cognizant to ensure that the book for metrics are not just measuring things but drive decisions. This is possible by supporting metrics with the ability to create interactions that highlight root cause and provide actionable insights using the semantic layer. Attention to detail on the chart types, gauge types, layout and color spectrum crafts your stories that reveal insights hidden in your data. A primary focus in the healthcare industry is on quality of care and outcomes. Readmission of patients is an important metric that measures quality of care and the illustration Figure 15. KPI Dashboard below, shows how powerful a slider gauge can be in quickly communicating the status of a key performance indicator in relation to a target or interval.

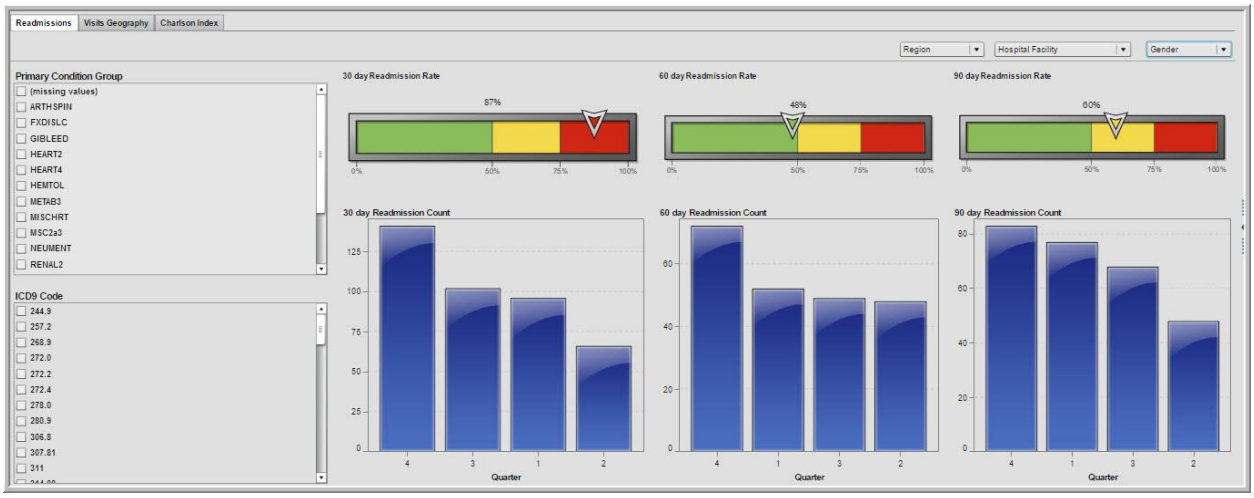


Figure 15. KPI Dashboard

ANALYTICAL MODELS

The SAS Visual Statistics add-in provides users the ability to leverage the semantic layer to quickly build analytical models. The frequently used and popular models such as Linear Regression, Logistic Regression, Decision Tree and Clustering as available in SAS Visual Analytics 7.2.

The following illustration Figure 16. Analytical Models - Clustering, shows the ease of use to quickly cluster, in this case customers, and the ability to introduce the cluster segment for further analysis and reporting.

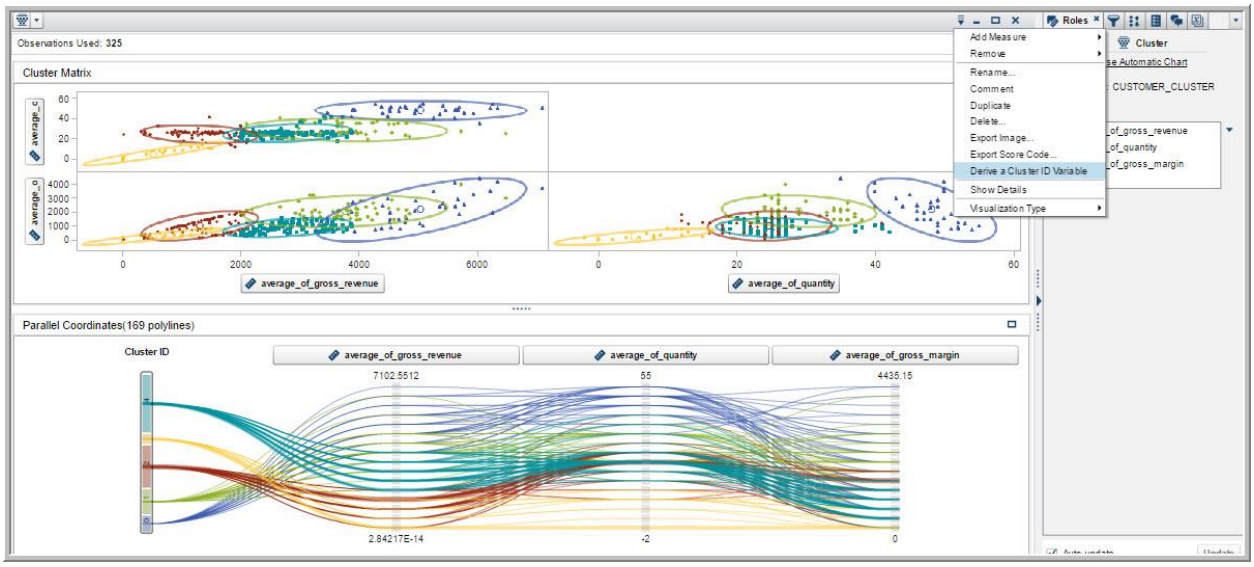


Figure 16. Analytical Models - Clustering

CONCLUSION

Given the importance of an analytics program, the key to success is a well planned and executed strategy. The value delivered can be immense when your strategy encompasses people, process, technology, data management, reporting and analytical capabilities. The Business Intelligence and Analytics tool is often considered as the magic potion that would accelerate the maturity of analytics often leading to implementation with low business impact and marginal improvements. Other crucial reason for failure is due to the bottoms up approach, geared towards small changes that do not require a strategy and generally end up with point solutions. SAS® Visual Analytics provides powerful features required in a Business Intelligence and Analytics tool, and when rolled out with a comprehensive strategy, delivers the innovation and disruption required to gain immense business value.

REFERENCES

- SAS® Visual Analytics User's Guide/ SAS® Visual Analytics Administration Guide. Available at https://support.sas.com/documentation/onlinedoc/va/
- Deloitte Analytics Advantage Study. Available at http://www2.deloitte.com/global/en/pages/deloitte-analytics/articles/the-analytics-advantage.html
- Arun Sugumar and Vimal Raj, Kavi Global. "Three S's in SAS® Visual Analytics: Stored Process, Star Schema, and Security". Paper 3433-2015. SAS Global Forum 2015.
- Kimball, Ralph and Ross, Margy 2013. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3rd ed. John Wiley & Sons

ACKNOWLEDGMENTS

We thank Kavi Global for providing us with all the necessary infrastructure and resources. We also would like to take this opportunity to thank Kishore Kondula from Kavi Global for his input on using SAS® Visual Analytics for data explorations.

RECOMMENDED READING

- *SAS® Visual Analytics User's Guide*
- Ratul Saha, Vimal Raj, and Vignesh Balasubramanian, Kavi Global. "Design for Success: An Approach to Metadata Architecture for Distributed Visual Analytics". Paper 10160-2016. SAS Global Forum 2016.

CONTACT INFORMATION

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

Arun Sugumar
Kavi Global
(847) 387-6760
arun@kaviglobal.com
http://www.kaviglobal.com/

Vignesh Balasubramanian
Kavi Global
(847) 387-6760
vignesh@kaviglobal.com
http://www.kaviglobal.com/