Paper 3433-2015

Three S's in SAS® Visual Analytics: Stored Process, Star Schema, and Security

Arun Sugumar, Vimal Raj
Kavi Associates LLC, Barrington, IL

ABSTRACT

SAS® Visual Analytics is very responsive in analyzing historical data, and it takes advantage of in-memory data. Data query, exploration, and reports form the basis of the tool, which also has other forward-looking techniques such as star schemas and stored processes. A security model is established by defining the permissions through a web-based application that is stored in a database table. That table is brought to the SAS Visual Analytics environment as a LASR table. Typically, security is established based on the departmental access, geographic region, or other business-defined groups. This permission table is joined with the underlying base table. Security is defined by a data filter expression through a conditional grant using SAS® metadata identities. The in-memory LASR star schema is very similar to a typical star schema. A single fact table that is surrounded by dimension tables is used to create the star schema. The star schema gives you the advantage of loading data quickly on the fly. A SAS application that gives the flexibility and the power of coding is created as a stored process that can be executed as requested by client applications such as SAS Visual Analytics. Input data sources for stored processes can be either LASR tables in the SAS® LASR™ Analytic Server or any other data that can be reached through the stored process code logic.

INTRODUCTION

Companies build data warehouses, data marts spending piles of money. Organizations have complex business requirements and are looking to take data driven decisions. The success of the analytics program of an organization depends on choosing the right tools which has the right capabilities built into them. SAS Visual Analytics is one such tool which has in-memory technology that helps to analyze and explore large volumes of data quickly and efficiently. The business cases used in this paper provide a platform for understanding star schema, stored process and row level permissions which would instigate the cognitive mind in the organization to achieve their business objectives efficiently.

STORED PROCESS

A stored process is a SAS program that is stored on a server. The definitions of the stored process should be provided in the SAS Metadata server. The power of the stored process exists in the manual written code which can be created using base SAS session. SAS code can contain instructions for showing report elements that include queries, prompted filters, titles, and statistical analyses. SAS Enterprise Guide has a wizard interface to create the stored process. The screen of the stored process creation wizard will collect name and location of the stored process repository, Execution SAS server details and the prompt screens for user inputs.

USING STORED PROCESS IN SAS VISUAL ANALYTICS

A typical Stored Process input parameters can be system variables or user created variables through which one can achieve many functionalities (e.g. filtering, data security, real-time report update etc.). Using stored process in a SAS Visual Analytics report is explained below using figure 1.

1. When creating the report drag the “Stored Process” from the other objects panel into the report section canvas
2. Choose the Stored Process by navigating to the metadata folder. This action embeds the SAS Stored Process into the report layout and prompts the user to enter the values
3. In the properties tab when the user opts to select “Show SAS log in output” in the right pane, then the output shows the log as well. This is very helpful for debugging the SAS code if the stored process encounters any issues.
4. When the user marks checkbox “Show Metadata View” in properties panel on the right side as shown in the figure 1 then stored process metadata information about creation, modification and server information will be available
5. Selecting the “Enable selection in the viewers” option provides a way to take input in the SAS Visual Analytics report viewer.

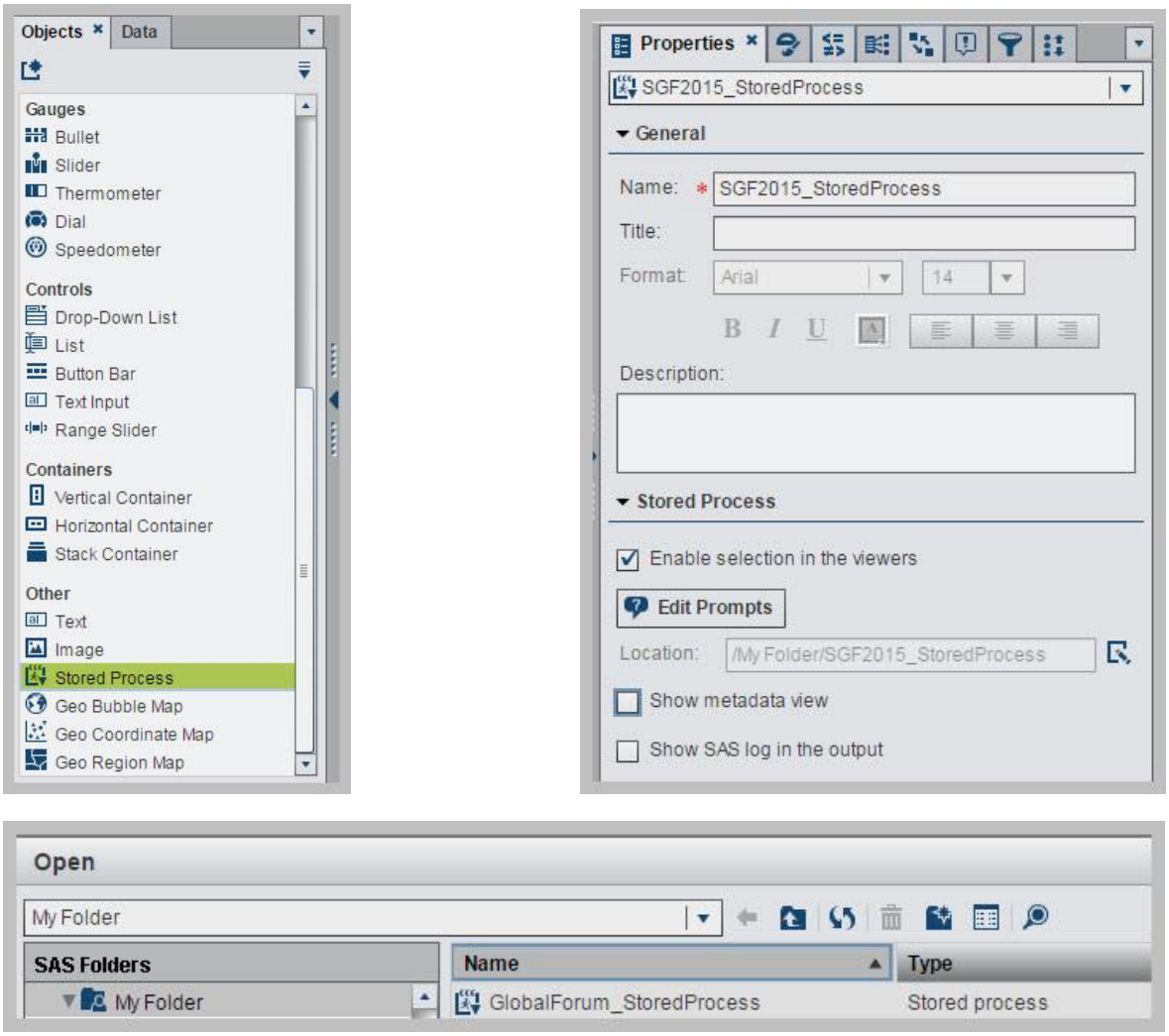


Figure 1. Adding Stored Process to SAS Visual Analytics Report

THINGS TO KEEP IN MIND WHEN WORKING WITH STORED PROCESSES IN SAS VISUAL ANALYTICS

Stored Process is very powerful as is it is a SAS code to support all the design ideas. Identifying and deciding on stored process for achieving your business objectives should be done after carefully understanding the following.

- Report which needs a precision layout cannot take in stored process as
- Interaction is a feature in SAS Visual Analytics which helps to filter the data. A Stored process cannot be the source or target of an interaction in a report. Also stored process cannot be added to a container
- The output of a stored process in a report will be an html screen
- Most of the SAS Procedures are accessible to the installs which have base SAS license. The SAS Visual Analytics can only support limited number of SAS Procedures.
- Stored processes can use any available data source (not only LASR tables). But running stored processes against large LASR tables is not a high-performance operation, because any referenced LASR tables must be read from the SAS LASR Analytic Server into a SAS session in the execution server. Using a stored process to read large tables from memory is not a high-performance operation

STORED PROCESS EXAMPLE

A dynamic interface was required to capture the Locomotive information and current rim thickness data which would go as input to the prediction model and display the results. To accomplish this task the SAS Stored process in Visual Analytics is used which provided benefits like user interaction, real time processing, what-if analysis etc.

Stored Process used in this paper takes locomotive Id, current rim thickness information and the prediction period as the inputs for the prediction model as shown in the Figure 2. "Reset to Defaults" option helps to set the default value defined in the stored process. If there is no default value saved in the stored process then existing reports use the values that were last saved with the report.

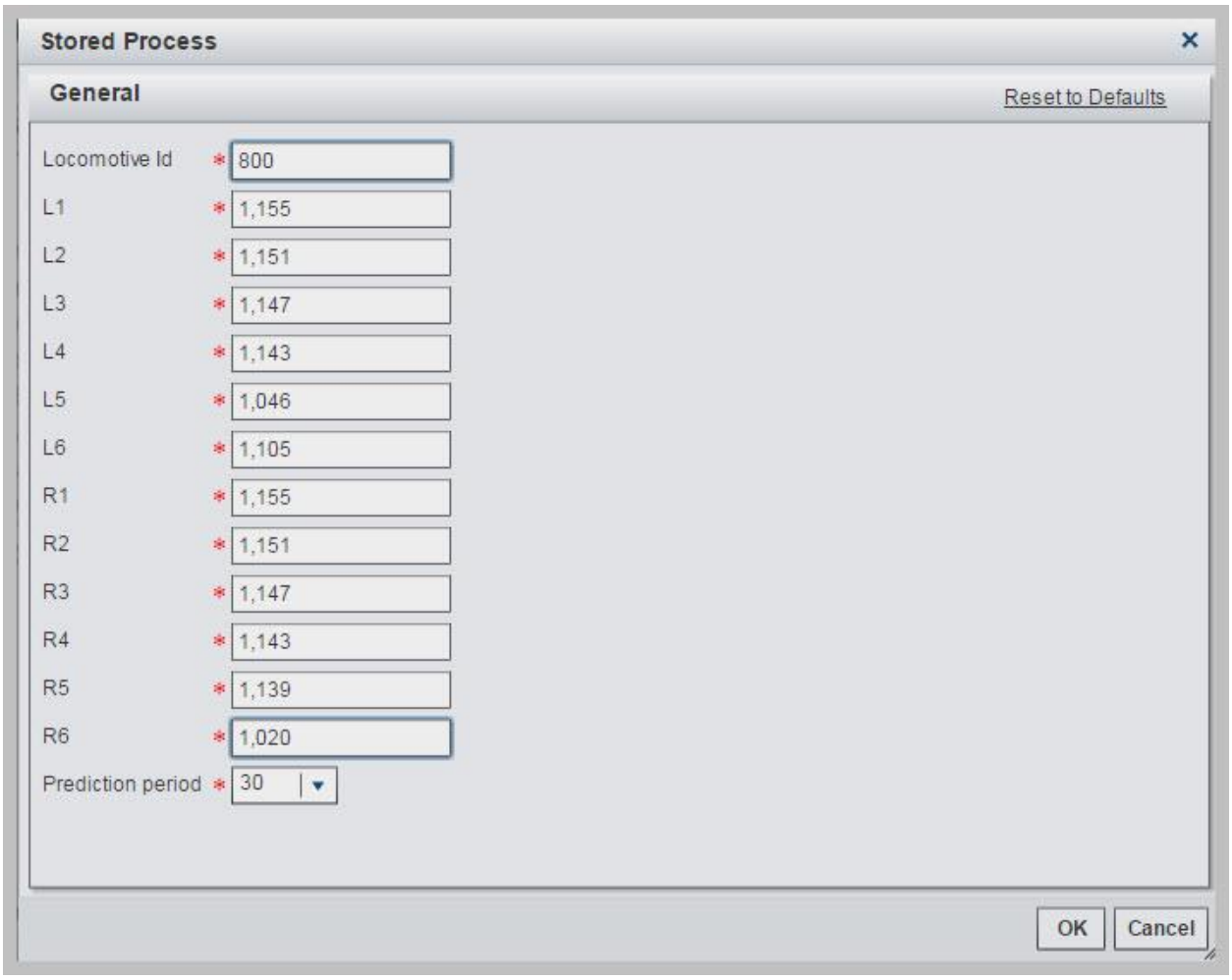


Figure 2. Providing inputs in Stored Process for predicting Rim thickness using prompts in SAS Visual Analytics 6.4

For changing the prompt values the top border of the stored process output screen is clicked to see the icon as shown in the figure 3, which enables you to select a prompt value.

Figure 3. Providing inputs in Stored Process using prompts in SAS Visual Analytics 6.4

The output screen of the stored process is presented in HTML format in the SAS Visual Analytics Report viewer. Additionally the SAS dataset from the stored process can also be loaded as in-memory LASR table. This LASR table is registered in the LASR library on the fly using %registerTable() macro in the stored process code with LIBRARY, REPOSID and TABLEID as a macro parameter. Having the output as a LASR table provides the option to use it in multiple reports.

In this example, the predicted wheel rim data is loaded as a LASR table using the stored process which is depicted below in SAS Visual Analytics Report using a Dual Axis Bar chart as exhibited in figure 4.

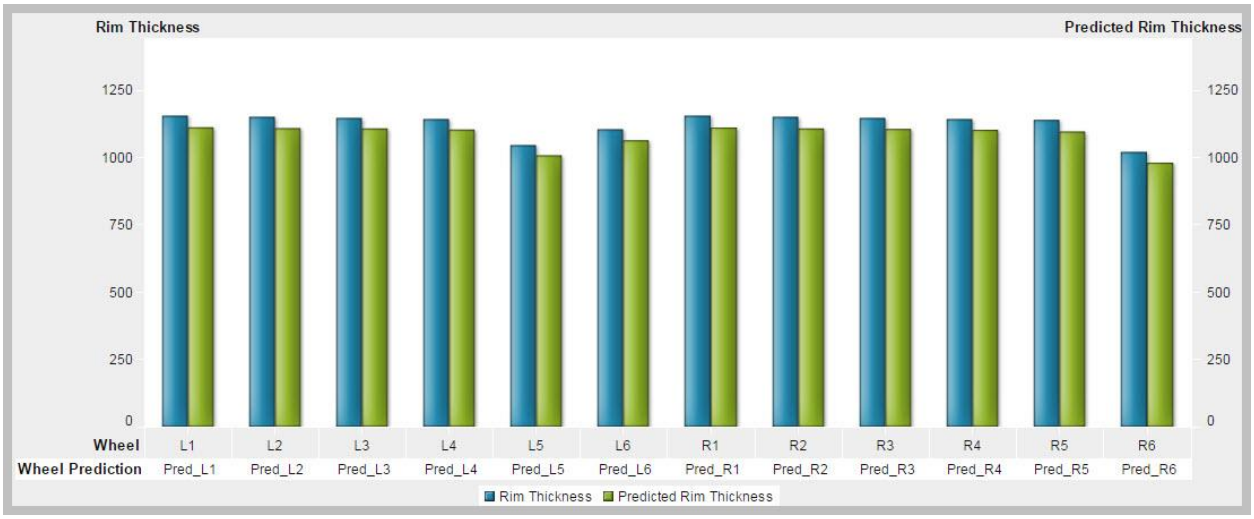


Figure 4. Output of a Stored Process chart for predicting Rim thickness using prompts in SAS Visual Analytics 6.4 after loading the data into LASR table

STAR SCHEMA

Star Schema is a data warehouse data mart schema architecture, which consists of one or more fact tables referencing to any number of dimension tables.

Fact Tables:

- A fact table typically has two types of columns: foreign keys to dimension tables and measures those that contain numeric facts.
- A fact table can contain fact's data on detail or aggregated level.
- Usually the fact tables in a star schema are in third normal form (3NF).
- Example: customers, assets, products, channels etc.

Dimension Tables:

- A dimension is a structure usually composed of one or more hierarchies that categorizes data. Flat dimension or list does not have any hierarchies.
- Dimension tables' primary keys are proper subset of the composite primary key of the fact tables.
- Dimension tables normally have numeric, descriptive / textual values and are generally smaller in size than fact table.
- Example: purchases, defects, market prices, subscriptions etc.

Star schema has below advantages used as a data warehouse schema:

- Simple and easy to understand
- Query execution is effective and faster (smaller sized normalized tables to join)
- Supports SCD types of data loading
- Reduces data redundancy

THINGS TO KEEP IN MIND WHEN WORKING WITH STAR SCHEMA IN SAS VISUAL ANALYTICS

When the warehouse architecture is based on a single key the star schema approach is feasible in SAS Visual Analytics. Also consider the following aspects before implementing star schema.

- Output of the star schema can be created as view or as full table. Views have faster creation time.
- When creating the star schema the fact table needs to be selected first before the dimension tables.
- If an input table is not present in a LASR Analytic Server library, then they are loaded into memory when the star schema is executed.
- A dimension table showing the incomplete table status icon, usually indicates that the data builder could not determine the join condition for the dimension table. The same icon provides information about how to correct the incomplete status.
- Deleting a specific dimension table column from output, can be done in the workspace.
- The dimension key and corresponding column in fact table have to be of same data type and length. If they are different in tables, either it can be modified in original data system or modify the data query to match the data type and length.
- Snow flake schemas are not supported.

STAR SCHEMA EXAMPLE

The Stored Process example used in this paper is designed on a star schema as it is most popular architecture model for a database. Below screenshot (figure 5) provides the structure for asset repairs (FACT) which is connected with the unique set of repair locations (DIM), assets (DIM), person (DIM), repairs (DIM) and date-time (DIM). This architecture helps us to effectively manage the data and query the database and generate reports across any categories defined. Each of the dimension tables are joined to the fact table with a dimension key. For example: Location DIM table is joined with Repair FACT table through Location ID key. This variable is a foreign key in FACT, but primary in DIM table.

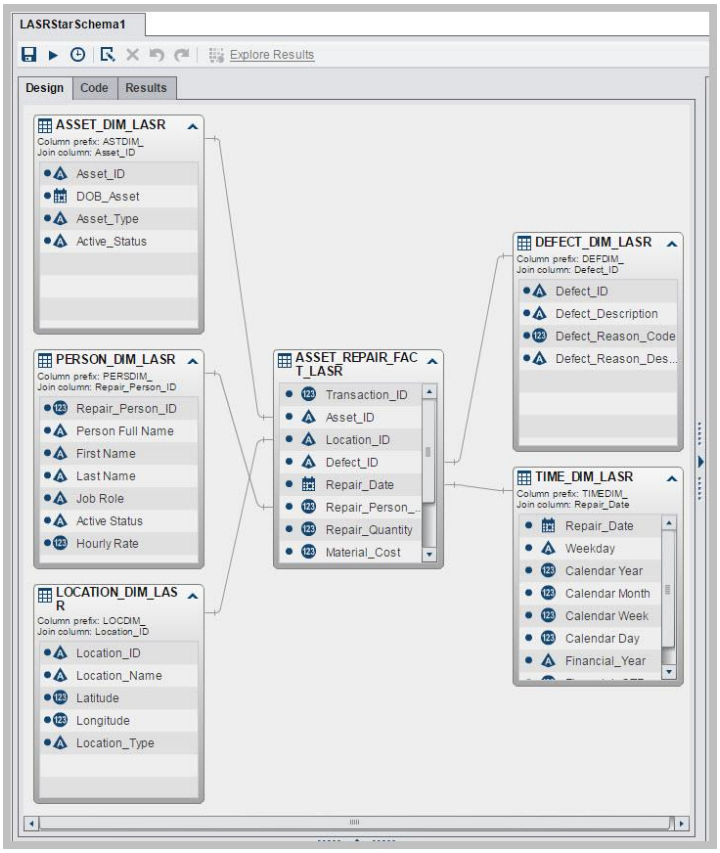


Figure 5. Visual Data Builder window to construct a star schema

Below code can be viewed through the "Code" tab in Visual Data Builder window. This is a logical representation of the schema design.

```
LIBNAME ASSTLASR SASIOLA  TAG=SGF_2015  PORT=10013
SIGNER="http://XXXXXXXXXXXX:7980/SASLASRAuthorization"  HOST="sasdemo" ;
/** LASR STAR SCHEMA CODE **/
/* Drop existing table */
%vdb_dt(ASSTLASR.OutputTable);
proc imstat;
  table ASSTLASR.ASSET_REPAIR_FACT_LASR;
  schema
    ASSET_DIM_LASR ( Asset_ID = Asset_ID
     / prefix = ASTDIM_ )
    DEFECT_DIM_LASR ( Defect_ID = Defect_ID
     / prefix = DEFDIM_ )
    PERSON_DIM_LASR ( Repair_Person_ID = Repair_Person_ID
     / prefix = PERSDIM_ )
    TIME_DIM_LASR ( Repair_Date = Repair_Date
     / prefix = TIMEDIM_ )
    LOCATION_DIM_LASR ( Location_ID = Location_ID
     / prefix = LOCDIM_ )
  ;
  run;
  table ASSTLASR.&_templast_;
  promote OutputTable;
  run;
quit;
```

The SAS Visual Data Builder helps to construct a star schema. This provides a user interface to design SQL joins which can either be performed inside the database or in the SAS environment with the help of a scheduler if data loading is time consuming. The LASR Analytics Server here provides a unique advantage in building in-memory star schema which has better performance over executing traditional SQL joins through better memory utilization.

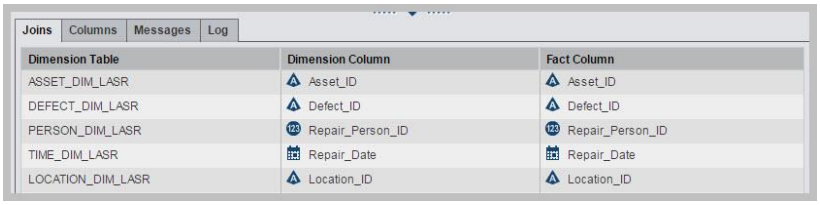


Figure 6. Visual Data Builder window showing the joins and contributing columns

The JOINS tab (figure 6) helps to review the variables on which the join happens and create the output table.

PREFIX IN STAR SCHEMA

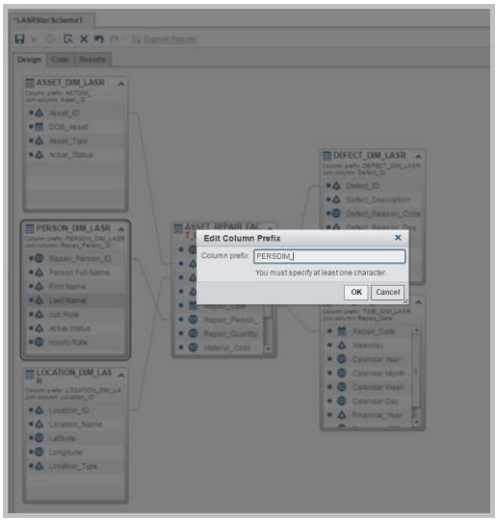


Figure 7. Visual Data Builder window helps to edit the column prefixes

The first 15 characters of a dimension table's name and the underscore character ("_") are initially set as a prefix for the column names from the dimension table. Column names for the output table are a combination of the prefix and the original column name. However, it can be changed in the design window any time per requirements. Column names from the fact table are not modified with a prefix. A column name is limited to only 32 characters. The report labels in SAS Visual Analytics Viewer will have the prefix and it is good to have the logical prefixes that will help to identify the source table from where the column is picked from.

SECURITY IN SAS VISUAL ANALYTICS

Today's environment is rich with large amount of data floating across different systems with different formats and standards. Extensive sharing and transformation makes it more complex, larger and harder to manage. Hence protection around secure (e.g. Financial Incentives, Employee Salaries) and confidential (e.g. SSN, Credit Card Number etc.) data becomes is very critical for the organization. Security around data can me implemented in many ways like real-time audits, granular audits, granular access controls, masking, encryption, end-point input validation and filtering.

In this paper, Visual Analytics is used to apply row-level security around the secure information to protect and show only permissible data for a logged user, through end-point validation and filtering methodology.

ROW LEVEL PERMISSION EXAMPLE

A new identity "Row Level Security Testers" is created SAS Management Console, which works as a user group and contains all the testers who need to test the data relevant to their own designated group. The identity permissions are set into Visual Analytics LASR Fact table. Below screenshot shows the ReadMetadata and Read permission is provided, however other activities are not permitted at this moment.

In the "Manage Environment" option This applies the first level of protection at the report / table metadata level. Anyone except this identity would not be able to access the table. Users with this identity will be able to only read the data.

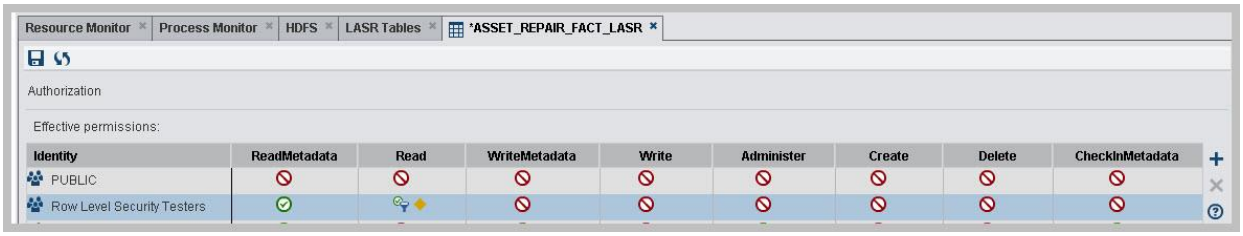


Figure 8. Entity permission setup window on the fact table

Next level of security is applied based on data filtering at row level. The Fact table has a variable named Repair_Person_ID. This contains the ID of the person designated for that record. There might be one or more users who are entitled to view a single record. In that case, their names would be comma separated.

In Visual Analytics this permission conditions can be provided in the window provided below. When the report is viewed by any user under "Row Level Security Testers" group, their login ID would be captured automatically as SAS.PersonName. This is a system variable which is managed at the system level. If Repair_Person_ID (data field) contains the logged in person's ID (a system field), then the records would be picked up in output table to view.

The edit permission condition and the corresponding code is shown in below screenshots.

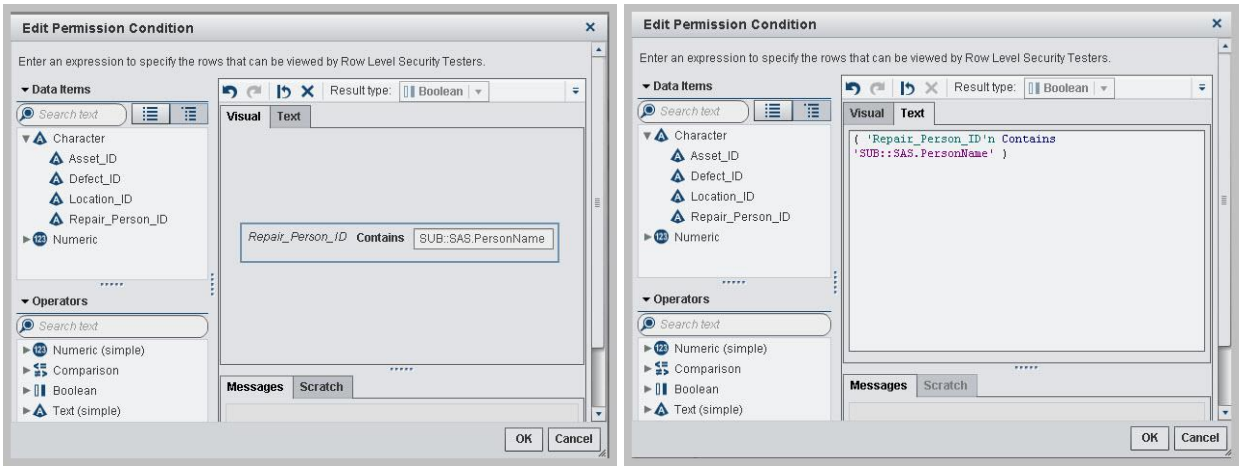
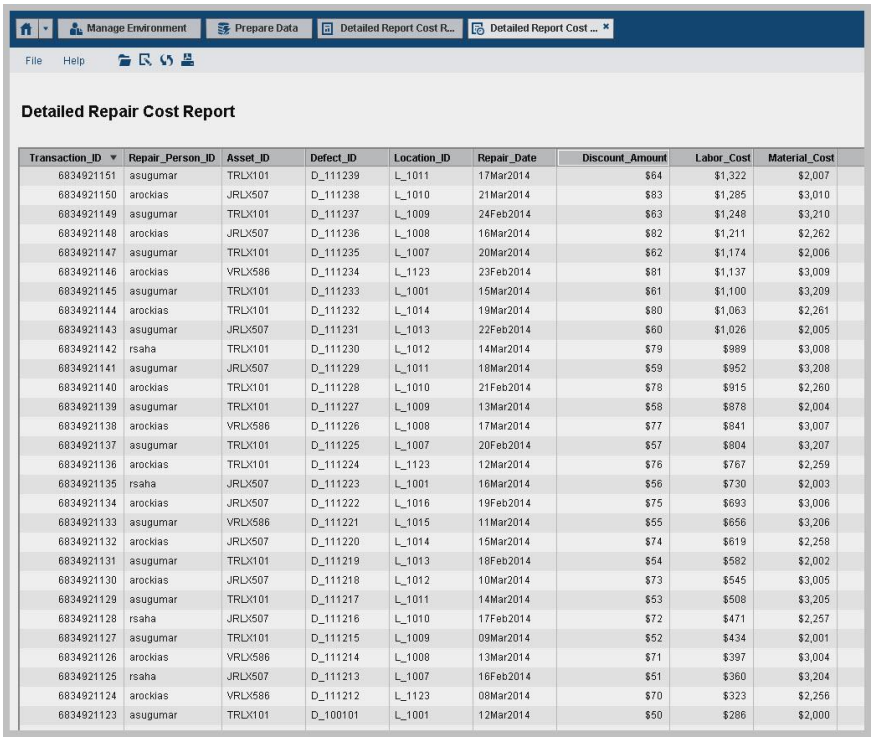


Figure 9. Setting up the row level security based on data level columns

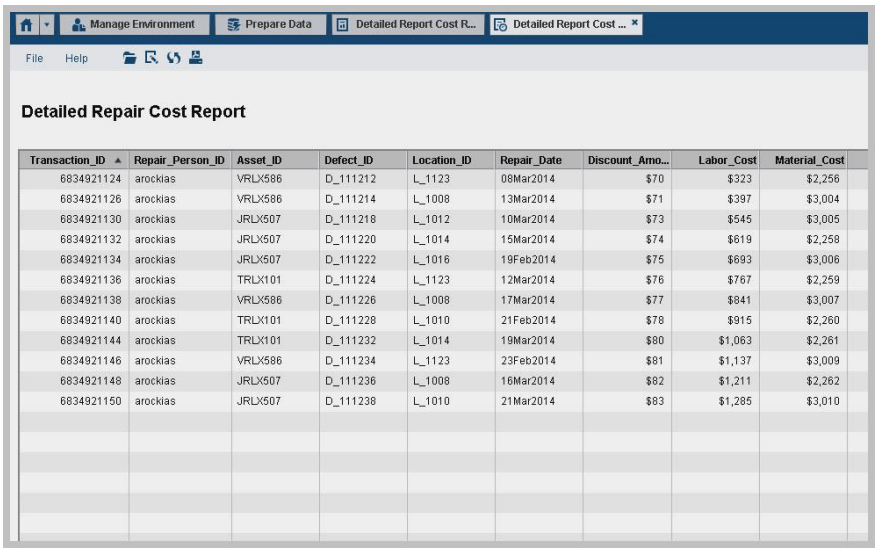
Once the metadata permissions and edit permissions are set, the report can be viewed. Below screenshot is the output from the stored process before the row-level security is applied. It shows all the records irrespective of any Repair_Person_ID.



Transaction_ID	Repair_Person_ID	Asset_ID	Defect_ID	Location_ID	Repair_Date	Discount_Amount	Labor_Cost	Material_Cost
6834921151	asugumar	TRLX101	D_111239	L_1011	17Mar2014	$64	$1,322	$2,007
6834921150	arockias	JRLX507	D_111238	L_1010	21Mar2014	$83	$1,285	$3,010
6834921149	asugumar	TRLX101	D_111237	L_1009	24Feb2014	$63	$1,248	$3,210
6834921148	arockias	JRLX507	D_111236	L_1008	16Mar2014	$82	$1,211	$2,262
6834921147	asugumar	TRLX101	D_111235	L_1007	20Mar2014	$62	$1,174	$2,006
6834921146	arockias	VRLX586	D_111234	L_1123	23Feb2014	$81	$1,137	$3,009
6834921145	asugumar	TRLX101	D_111233	L_1001	15Mar2014	$61	$1,100	$3,209
6834921144	arockias	TRLX101	D_111232	L_1014	19Mar2014	$80	$1,063	$2,261
6834921143	asugumar	JRLX507	D_111231	L_1013	22Feb2014	$60	$1,026	$2,005
6834921142	rsaha	TRLX101	D_111230	L_1012	14Mar2014	$79	$989	$3,008
6834921141	asugumar	JRLX507	D_111229	L_1011	18Mar2014	$59	$952	$3,208
6834921140	arockias	TRLX101	D_111228	L_1010	21Feb2014	$78	$915	$2,260
6834921139	asugumar	TRLX101	D_111227	L_1009	13Mar2014	$58	$878	$2,004
6834921138	arockias	VRLX586	D_111226	L_1008	17Mar2014	$77	$841	$3,007
6834921137	asugumar	TRLX101	D_111225	L_1007	20Feb2014	$57	$804	$3,207
6834921136	arockias	TRLX101	D_111224	L_1123	12Mar2014	$76	$767	$2,259
6834921135	rsaha	JRLX507	D_111223	L_1001	16Mar2014	$56	$730	$2,003
6834921134	arockias	JRLX507	D_111222	L_1016	19Feb2014	$75	$693	$3,006
6834921133	asugumar	VRLX586	D_111221	L_1015	11Mar2014	$55	$656	$3,206
6834921132	arockias	JRLX507	D_111220	L_1014	15Mar2014	$74	$619	$2,258
6834921131	asugumar	TRLX101	D_111219	L_1013	18Feb2014	$54	$582	$2,002
6834921130	arockias	JRLX507	D_111218	L_1012	10Mar2014	$73	$545	$3,005
6834921129	asugumar	TRLX101	D_111217	L_1011	14Mar2014	$53	$508	$3,205
6834921128	rsaha	JRLX507	D_111216	L_1010	17Feb2014	$72	$471	$2,257
6834921127	asugumar	TRLX101	D_111215	L_1009	09Mar2014	$52	$434	$2,001
6834921126	arockias	VRLX586	D_111214	L_1008	13Mar2014	$71	$397	$3,004
6834921125	rsaha	JRLX507	D_111213	L_1007	16Feb2014	$51	$360	$3,204
6834921124	arockias	VRLX586	D_111212	L_1123	08Mar2014	$70	$323	$2,256
6834921123	asugumar	TRLX101	D_100101	L_1001	12Mar2014	$50	$286	$2,000

Figure 10. Report view without row-level security

Below screenshot is from the same Detailed Repair Cost Report after the row-level security is applied. It shows only specific records for the logged in user.

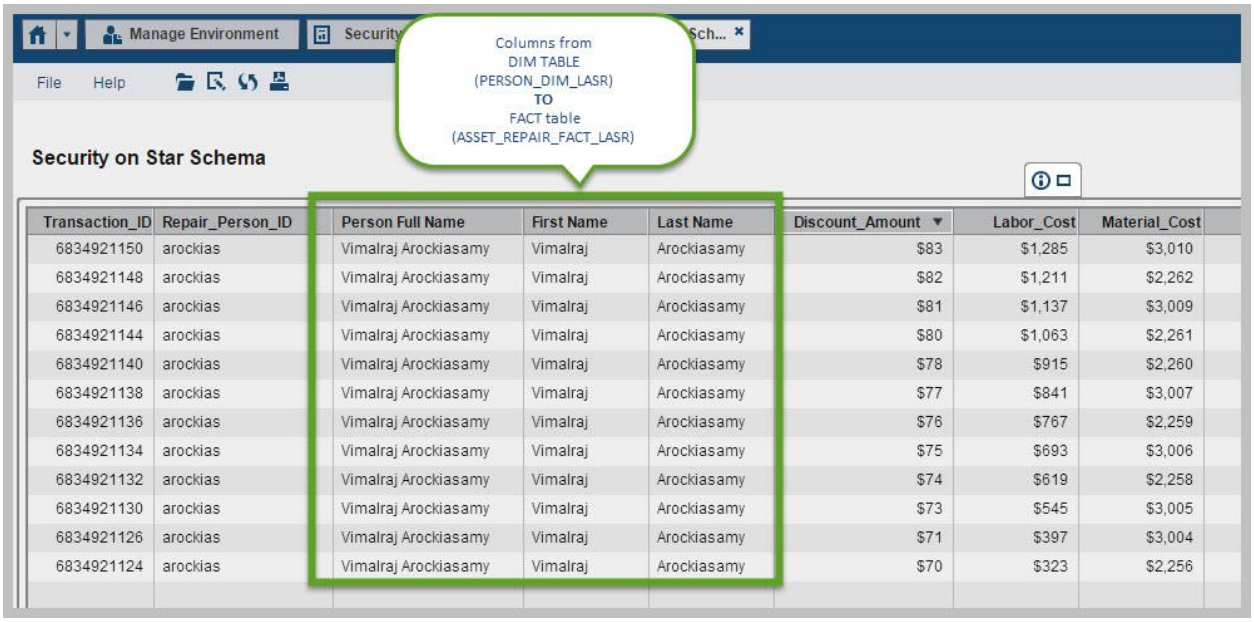


Transaction_ID	Repair_Person_ID	Asset_ID	Defect_ID	Location_ID	Repair_Date	Discount_Amo...	Labor_Cost	Material_Cost
6834921124	arockias	VRLX586	D_111212	L_1123	08Mar2014	$70	$323	$2,256
6834921126	arockias	VRLX586	D_111214	L_1008	13Mar2014	$71	$397	$3,004
6834921130	arockias	JRLX507	D_111218	L_1012	10Mar2014	$73	$545	$3,005
6834921132	arockias	JRLX507	D_111220	L_1014	15Mar2014	$74	$619	$2,258
6834921134	arockias	JRLX507	D_111222	L_1016	19Feb2014	$75	$693	$3,006
6834921136	arockias	TRLX101	D_111224	L_1123	12Mar2014	$76	$767	$2,259
6834921138	arockias	VRLX586	D_111226	L_1008	17Mar2014	$77	$841	$3,007
6834921140	arockias	TRLX101	D_111228	L_1010	21Feb2014	$78	$915	$2,260
6834921144	arockias	TRLX101	D_111232	L_1014	19Mar2014	$80	$1,063	$2,261
6834921146	arockias	VRLX586	D_111234	L_1123	23Feb2014	$81	$1,137	$3,009
6834921148	arockias	JRLX507	D_111236	L_1008	16Mar2014	$82	$1,211	$2,262
6834921150	arockias	JRLX507	D_111238	L_1010	21Mar2014	$83	$1,285	$3,010

Figure 11. Report view with row-level security showing data only for the logged in user

STAR SCHEMA ROW LEVEL SECURITY

Row level permissions applied on the output LASR table of the star schema is exhibited in figure12. The highlighted columns are from the person_dim_lasr table (DIM) which were brought by joining the repair_person_id and shown here as part of the final output of the star schema.



Transaction_ID	Repair_Person_ID	Person Full Name	First Name	Last Name	Discount_Amount	Labor_Cost	Material_Cost
6834921150	arockias	Vimalraj Arockiasamy	Vimalraj	Arockiasamy	$83	$1,285	$3,010
6834921148	arockias	Vimalraj Arockiasamy	Vimalraj	Arockiasamy	$82	$1,211	$2,262
6834921146	arockias	Vimalraj Arockiasamy	Vimalraj	Arockiasamy	$81	$1,137	$3,009
6834921144	arockias	Vimalraj Arockiasamy	Vimalraj	Arockiasamy	$80	$1,063	$2,261
6834921140	arockias	Vimalraj Arockiasamy	Vimalraj	Arockiasamy	$78	$915	$2,260
6834921138	arockias	Vimalraj Arockiasamy	Vimalraj	Arockiasamy	$77	$841	$3,007
6834921136	arockias	Vimalraj Arockiasamy	Vimalraj	Arockiasamy	$76	$767	$2,259
6834921134	arockias	Vimalraj Arockiasamy	Vimalraj	Arockiasamy	$75	$693	$3,006
6834921132	arockias	Vimalraj Arockiasamy	Vimalraj	Arockiasamy	$74	$619	$2,258
6834921130	arockias	Vimalraj Arockiasamy	Vimalraj	Arockiasamy	$73	$545	$3,005
6834921126	arockias	Vimalraj Arockiasamy	Vimalraj	Arockiasamy	$71	$397	$3,004
6834921124	arockias	Vimalraj Arockiasamy	Vimalraj	Arockiasamy	$70	$323	$2,256

Figure 12. Star schema Report view with row-level security showing data only for the logged in user

CONCLUSION

SAS Visual Analytic server and SAS Visual Analytics provides a visual and interactive environment where users can explore multiple data sources and identify patterns and trends. Exploring the data and presenting to the business visually can deliver the insights. Identifying the reporting requirements of the organization and developing it the right way using with the best tool is critical for the success. Star schema, stored process and row level permissions are forward looking functionalities which will seek more attention in future release of SAS Visual Analytics.

REFERENCES

- *SAS Visual Analytics - User Guide*
- Creating and Using SAS® Stored Processes with SAS® Enterprise Guide® - *http://www2.sas.com/proceedings/sugi30/135-30.pdf*

ACKNOWLEDGMENTS

We thank Kavi Associates LLC for providing us with all the necessary infrastructure and resources. We also wanted to thank Ratul Saha and Hari Hara Sudhan from Kavi Associates. We are also grateful to SAS for giving us this opportunity to present and publish our work and learnings.

RECOMMENDED READING

- *SAS Visual Analytics - User Guide*
- Creating and Using SAS® Stored Processes with SAS® Enterprise Guide® - *http://www2.sas.com/proceedings/sugi30/135-30.pdf*

- Big Data Everywhere! Easily Loading and Managing Your Data in the SAS® LASR™ Analytic Server http://support.sas.com/resources/papers/proceedings14/SAS347-2014.pdf

CONTACT INFORMATION

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

Arun Sugumar
Kavi Associates LLC
(847) 387 6760
arun.sugumar@kaviglobal.com
http://www.kaviglobal.com

Vimal Raj
Kavi Associates LLC
(847) 387 6760
vimal.raj@kaviglobal.com
http://www.kaviglobal.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.