Combating Fraud and Financial Crime Effectively by Empowering Data Scientists Using SAS® Viya

Stephen Overton, 6 Degree Intelligence

ABSTRACT

Building a productive financial crime monitoring system to detect suspicious activity begins and ends with the data. Access to production data ensures that data scientists and strategy developers can design scenarios with actual behavior patterns and data profiles on hand - not by using theoretical or mocked activity, which is not representative of real-world actions to be analyzed and modeled. Architecting a specialized analytical SAS® Viya sandbox environment, with Read-Only access to production data, provides a powerful toolbox to productively analyze and prepare data; design, build, and tune analytical models; and visualize results efficiently. This paper provides best practices for leveraging SAS Viya technologies within an analytical sandbox environment, including SAS® Visual Analytics, SAS® Visual Statistics, SAS® Data Preparation, and SAS® Visual Data Mining and Machine Learning. This paper also demonstrates concepts and deployment strategies to operationalize financial crime detection strategies in SAS® Visual Investigator.

INTRODUCTION

This paper is intended for architects, power users, SAS programmers, managers, or any persona which seeks to maximize productivity with SAS technology and deliver effective results. Decisions makers will gain knowledge about ways to design and plan strategically, while programmers and analysts will gain tactful insights and a vision to solve problems and leverage SAS Viya efficiently in an enterprise ecosystem.

Every organization is different and has its own set of unique challenges. The goal of this paper is to provide a blueprint for success by sharing lessons learned and potential solutions for common challenges. Suggestions provided in this paper should be discussed and vetted with key stakeholders and decision makers to help navigate internal processes and to ensure policies are followed. Examples given in this paper are based on experiences developing financial crime detection systems such as Anti-Money Laundering (AML) transaction monitoring or fraud detection but could be generalized for other potential solutions.

This paper frequently refers to the concept of a SAS Viya sandbox. For the purposes of this paper, the term sandbox is meant to imply that data scientists, analysts, and programmers have full capability to "play" with data and solve complex problems with no functionality restrictions. A sandbox environment operates separately from production environments so critical operations are not impacted but production data sources are leveraged in parallel. The term sandbox is not affiliated with any official SAS products at the time of this publication.

KEY ARCHITECTURE DESIGN CONCEPTS

A SAS Viya sandbox contains all necessary tools for data manipulation, advanced modeling, and visualization. Common tools include SAS Visual Analytics, SAS Visual Statistics, SAS Data Preparation, and SAS Visual Data Mining and Machine Learning. A sandbox environment could also contain tools such as R, Python, and Java since SAS Viya supports open source technology. After describing the architectural design concepts of a sandbox environment in this section, the remainder of this paper will provide guidance and examples to maximize tools within the sandbox.

The following diagram in Figure 1 depicts a conceptual representation of a sandbox environment separate from production but sharing the same data sources.
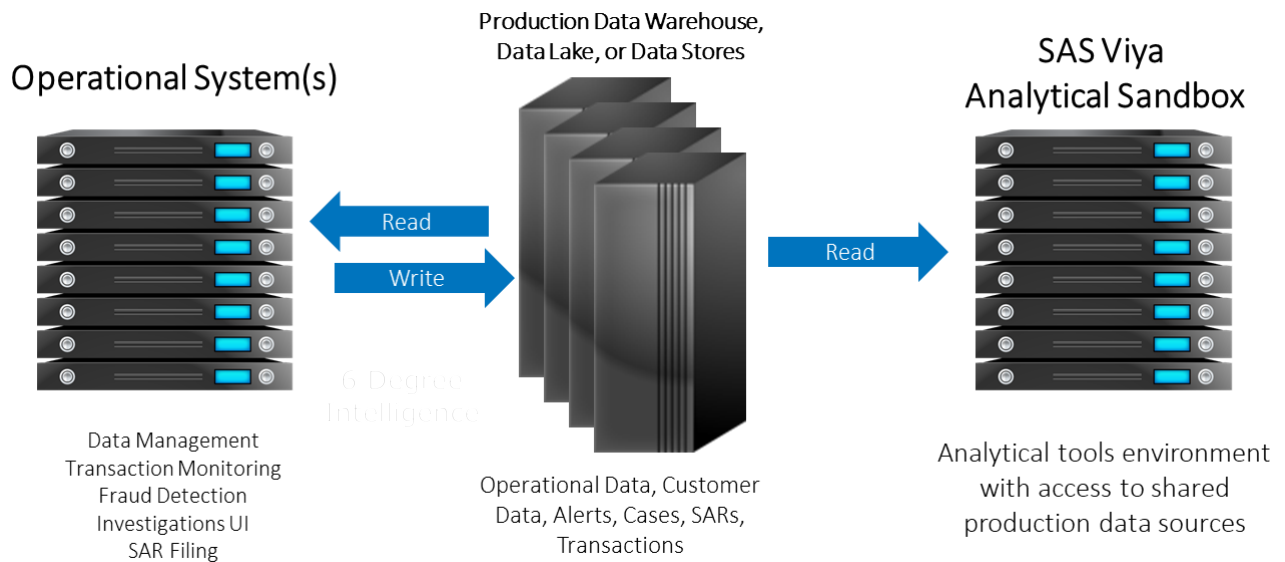

Figure 1: SAS Viya Analytical Sandbox Conceptual Diagram.

To ensure production systems are not impacted, data sources may need to be replicated to minimize impact to operational support systems. Data architects and IT administrators should assist and provide guidance during architecture planning sessions.

INTELLIGENCE LIFECYCLE

As new models are built or existing models enhanced in the sandbox environment, standard software development lifecycle (SDLC) processes should be followed to implement changes in production environments. Figure 2 describes a typical software development lifecycle with 4 total environments. Organizations may have different levels of environments based on size and complexity of the enterprise.

The overall intelligence lifecycle of insights and discoveries made in the sandbox environment flowing into production environments relies on a steady governance process typically managed by IT resources. Common outcomes from the sandbox environment may include parameter updates to existing models, new analytical models to implement, business workflow optimizations based on analysis performed, or operational reporting and executive dashboard visualizations.
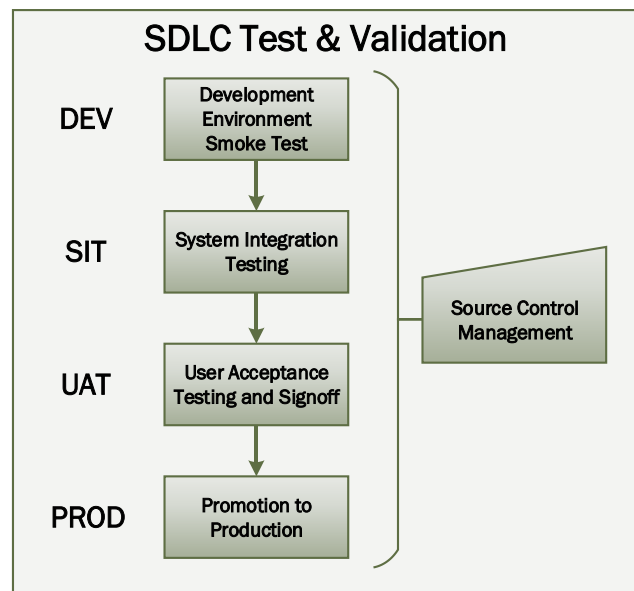

Figure 2: Software Development Lifecycle.

The following diagram in Figure 3 shows an example of how model development can occur in a sandbox environment and follow a typical promotion process through the software development lifecycle. The SIT, UAT, and PROD environments are assumed to be mirrored environments of DEV in this example.
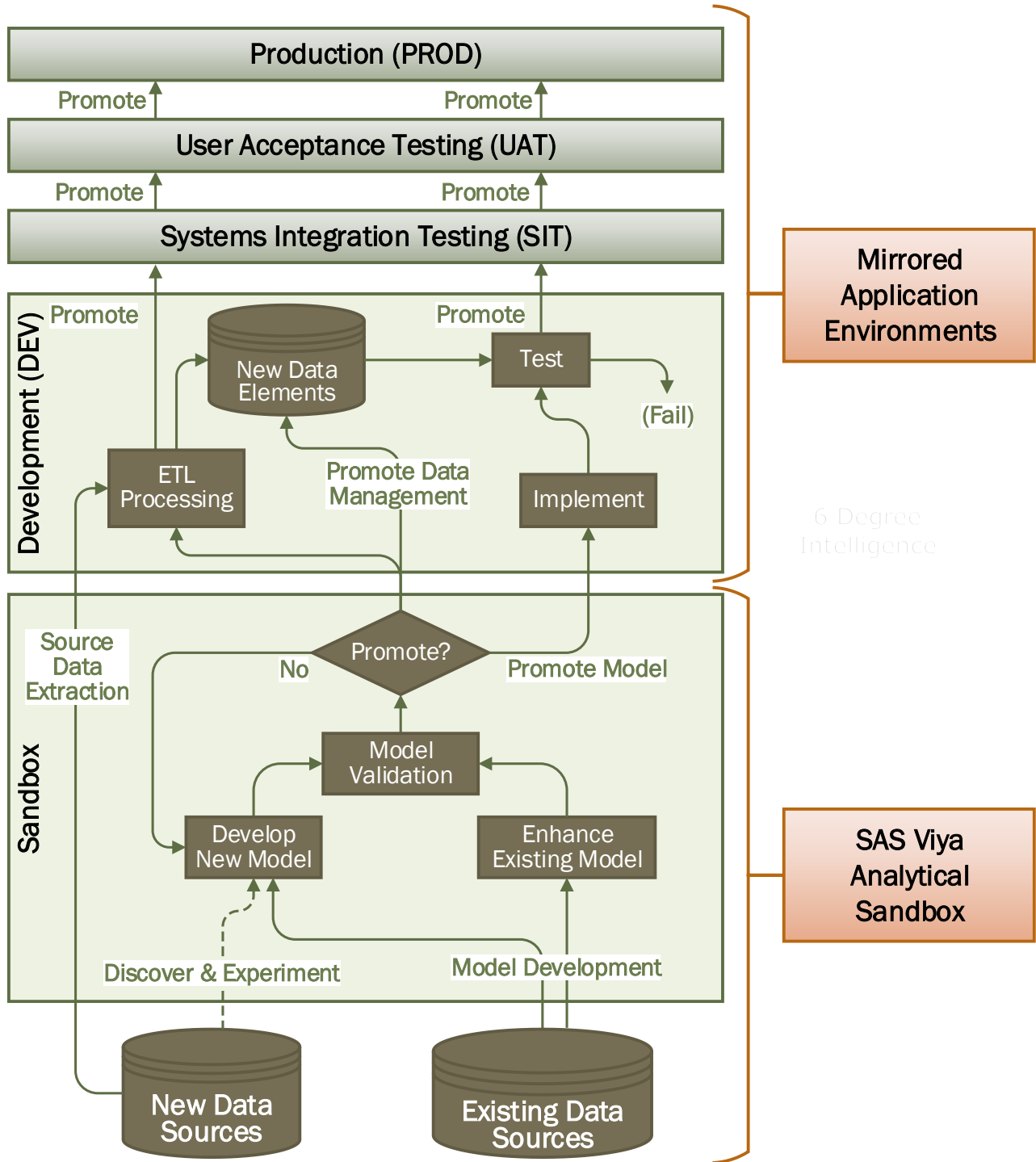
Figure 3: Example Sandbox Promotion Process.

EVERYTHING STARTS AND ENDS WITH THE DATA

The most critical piece of any project, initiative or individual analysis is the data. There can be a range of considerations depending on the maturity of the organization and resources available. The following list provides common considerations and example questions for managing data.

- Where does the data currently reside?
- Where will output data be stored?
- Who is responsible and accountable for the data?
- How is the data accessed?
- When is the data updated, what period of time does it apply to, and how is time described within the data?
- Why is the data important, what does it measure or describe in business terms?

SAS Viya has many capabilities that can be unleashed assuming the server running SAS can connect to the appropriate data sources. Ensuring users can connect to a database or necessary data source to retrieve information can be a critical point of any project.

TWO STEP APPROACH FOR MANAGING SECURE DATA ACCESS

SAS Viya has many visual interfaces which are extremely powerful and intuitive. One key requirement of tools such as SAS Visual Analytics or SAS Data Preparation is that data be registered and accessible through a CAS library. Analysis can be performed fast and easy when data is access through central CAS libraries. In situations where data sources have sensitive information, policy may require unique access for each user interfacing with the data so the database can enforce authorization permissions. Situations that may require this level of security include environments where users are spread across multiple countries or jurisdictions, if the source data has anything personally identifiable (PII data), or if there are heightened levels of security for departmental or team level access. The key architecture consideration is where the authorization occurs. The following list provides two data authorization approaches to consider.

- Data authorization managed within SAS, databases and file repositories can use a single service account or application user ID.
    - CAS libraries defined using SAS Environment Manager using standard techniques documented through SAS Support documentation.
- Data authorization managed outside of SAS, database and file repositories are accessed through unique user ID.
    - Two step CAS data management approach described in the following section.

Overall Process Flow for Two Step Access to Secure Data from SAS Viya

One method to load data from secure sources where unique user IDs are required for every connection to the data source can begin with a conceptual persona this paper defines as the Data Harvester. Data is harvested from sources using tools such as SAS Studio or SAS Data Preparation, manipulated and transformed, then loaded into a centralized CAS library for the Data Consumers to access. A Data Consumer is a conceptual persona this paper defines to include data scientists, analysts, and other developers or programmers assigned to build models or perform further analysis.

The following diagram in Figure 4 shows an example methodology of loading data in two stages to satisfy technical requirements of unique user access to source data and having a centralized CAS data management strategy for SAS Viya.
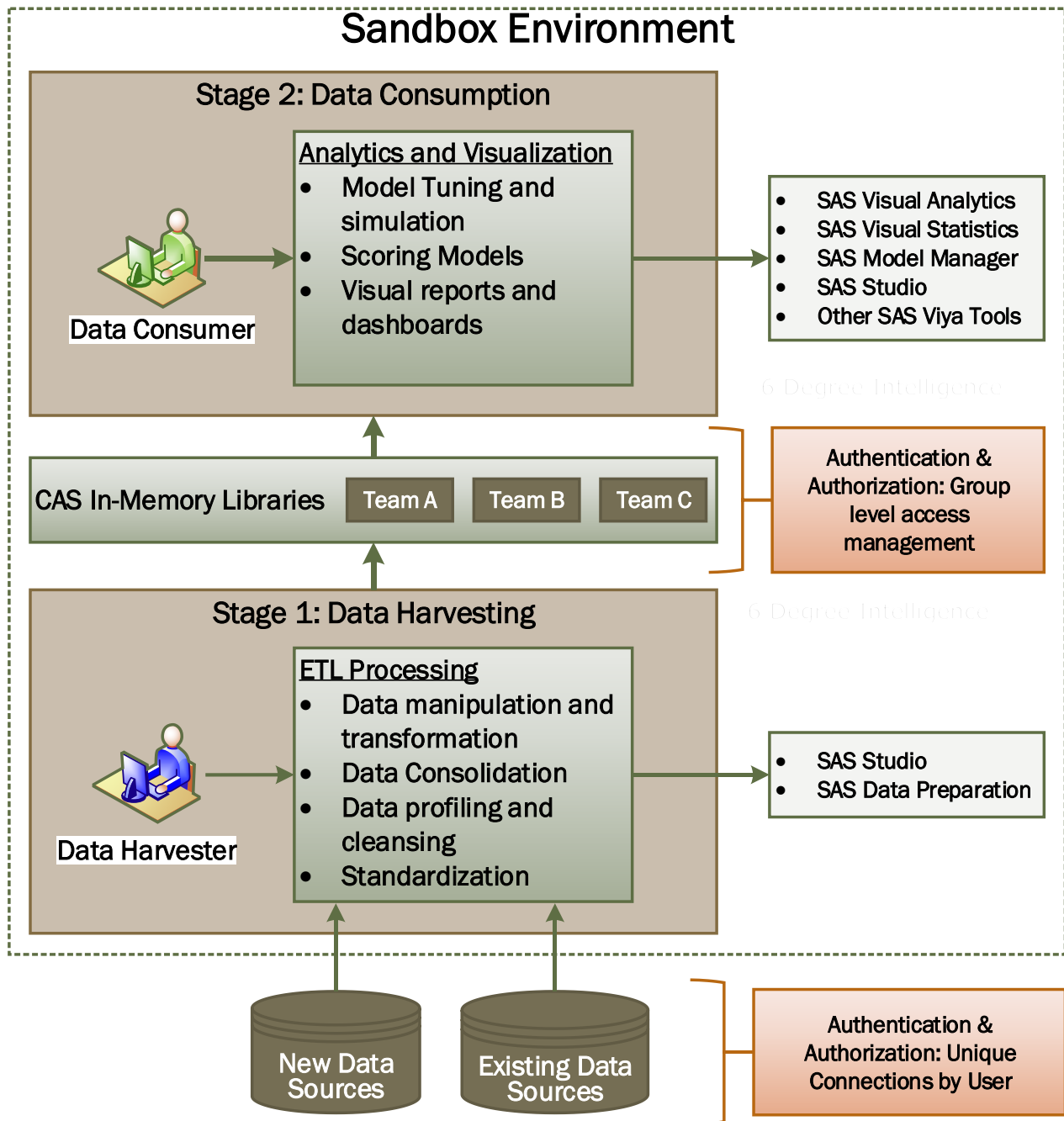
Figure 4: Example two stage approach to load secure data into CAS from SAS Viya.

Specific resources can serve across multiple personas if necessary. CAS libraries should be managed with group level access controls. It is assumed that all users within a defined group can share data within that CAS library. There can be redundancy of data across CAS libraries. Quotas can also be enforced at the CAS library level and an additional level of administrative flexibility added if libraries are managed across separate dedicated CAS servers.

SAS programming environments such as SAS Studio are easier to manage database access through individually defined SAS autoexec files or SAS macros which provide unique user access to databases.

THREE PILLARS OF THE SAS VIYA ANALYTICAL SANDBOX

Activity within a SAS Viya analytical sandbox can be grouped into three pillars of functionality: data, analytics, and visualization. These pillars help categorize the powerful tools and comprehensive abilities that can be leveraged by architecting a sandbox environment design. Figure 5 provides a visual diagram of the three pillars of a SAS Viya analytical sandbox environment.
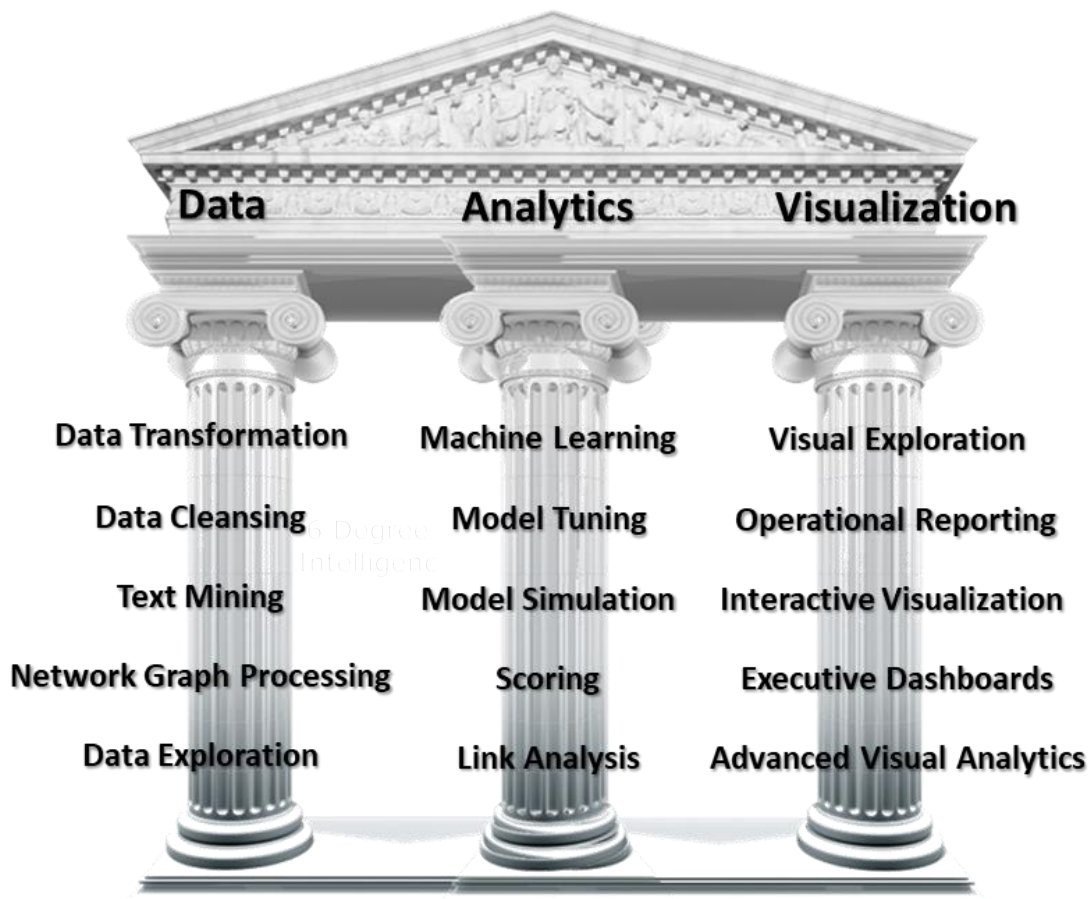

Data   **Analytics**   **Visualization**

Data	Analytics	Visualization
Data Transformation	Machine Learning	Visual Exploration
Data Cleansing	Model Tuning	Operational Reporting
Text Mining	Model Simulation	Interactive Visualization
Network Graph Processing	Scoring	Executive Dashboards
Data Exploration	Link Analysis	Advanced Visual Analytics

Figure 5: Three Pillars of the SAS Viya Analytical Sandbox environment.

DATA

As described in the previous section Everything Starts and Ends with the Data, data management is a very critical piece of any analytical ecosystem. Tools used to manage data include SAS Studio and SAS Data Preparation within SAS Viya. SAS Visual Analytics can also be used for data profiling and exploration. The following list provides example capabilities a SAS Viya sandbox can provide to manage data.

- Transform disparate sources to prepare information for modeling
- Develop ETL processes to build compliance data mart
- Profile data using tools such as SAS Data Preparation and SAS Studio
- Data quality tools using DataFlux® QKB libraries
- Ingest data in any tabular or delimited format for analysis
- Network graph theory processing

Figure 6 and Figure 7 provide a sample data profile of example bank transaction attributes.

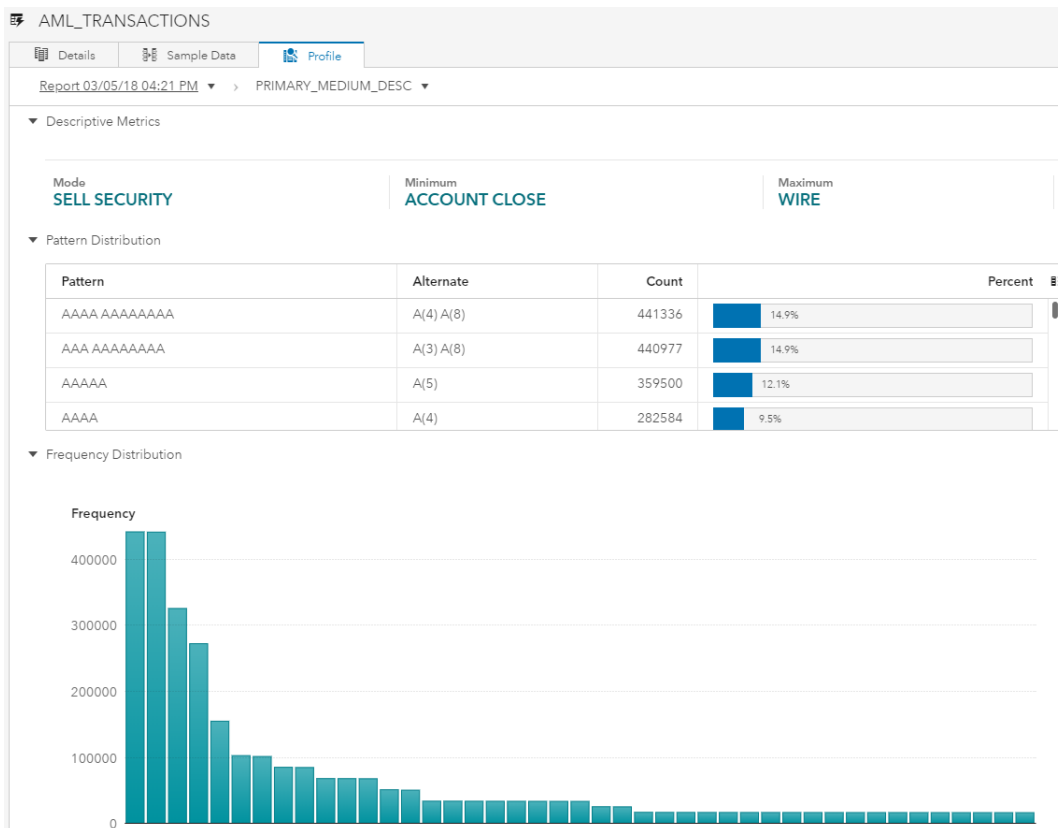**Figure 6: Data profiling using SAS Data Preparation.**

Figure 7: String value pattern profiling using SAS Data Preparation.

The following list provides example outcomes and business value that a SAS Viya sandbox could help provide for managing data within a financial crime monitoring.

- Establish feedback loop for machine learning to join response variables with predictors
- Combine disparate sources for additional dimensions within analysis
- Risk assessment validation
- Confirm known risk within data as well as identify unknown risks to the organization
- Develop and validate entity resolution approach
- Deliver higher quality data to operational monitoring systems and case management workflows

ANALYTICS

One of the primary goals of a sandbox environment is to enable data scientists and analysts. Minimizing restrictions in a controlled sandbox environment allows users to maximize knowledge and skillets as well as the technology investment. Having a sandbox environment promotes the mindset of rapid prototyping effectively because production ready data is made available to data scientists more efficiently. The faster resources can iterate through versions of models and analysis, the faster results can be obtained and optimized. The following list provides a set of common capabilities that can be leveraged within a SAS Viya analytical sandbox to improve financial crime monitoring systems.

- Develop models to predict monitoring scenario outcomes
- Perform customer segmentation analysis
- Develop risk scoring models
- Develop anomaly detection models

The following diagram in Figure 8 provides an approach for model development in a SAS Viya analytical sandbox environment. This example is tailored specifically for financial crime monitoring systems. The final phase represents deployment stages to SAS Visual Investigator. SAS Visual Investigator can also be deployed inside the sandbox environment for unit testing.
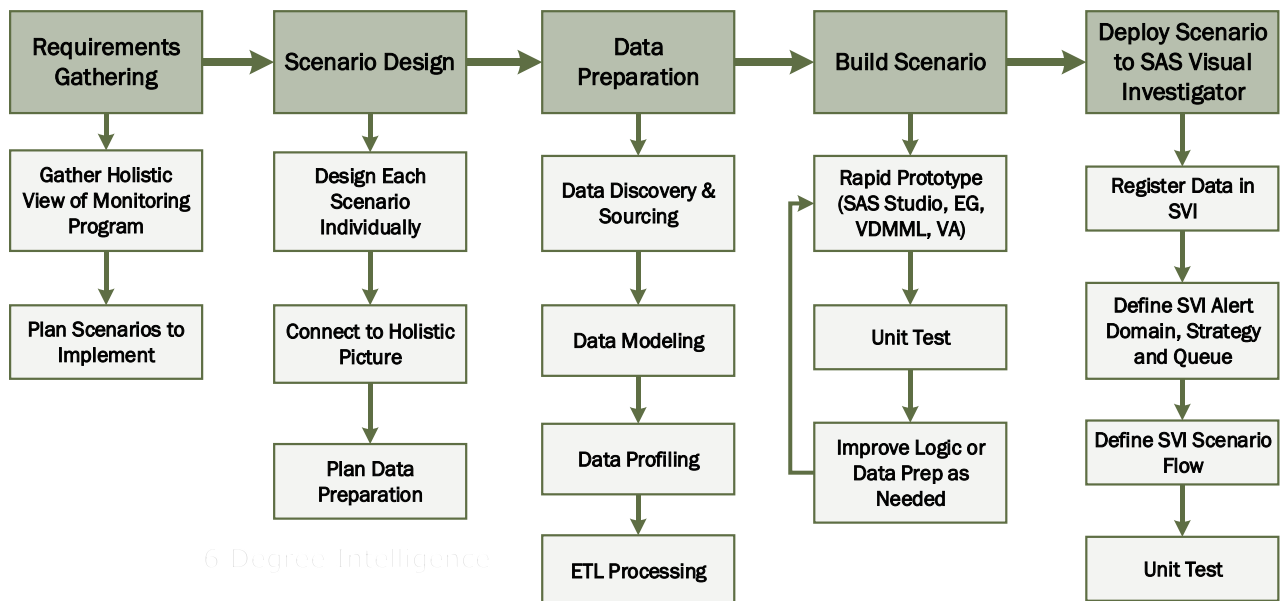

Figure 8: Example development process for financial crime monitoring systems.

The following list provides example outcomes and business value that a SAS Viya sandbox could help provide for developing analytical models within a financial crime monitoring program.

- Increase higher quality productive alerts through optimization techniques
- Improve alert triage through scoring, prioritization and ranking
- Monitor specific customer segments more effectively
- Define risk classification bands in a defendable model
- Understand drivers and contributors of risk scoring models
- Monitor for anomalous transaction behavior and develop peer groupings
- Discover new risk dimensions

VISUALIZATION

Communicating results in a sandbox environment is very effective using SAS Visual Analytics. There are many visualization objects which can be used to present simple or complex layers of information at any point in a project or analysis. SAS Visual Analytics can be used for standard reporting, operational monitoring, or executive level dashboards. The visualization pillar of the sandbox environment concept is critical because it helps decision makers leverage results to make better data-driven decisions to improve business operations.

The following chart in Figure 9 summarizes strategic goals at different points in the intelligence lifecycle. All three pillars are necessary for data to mature over time as the business grows.
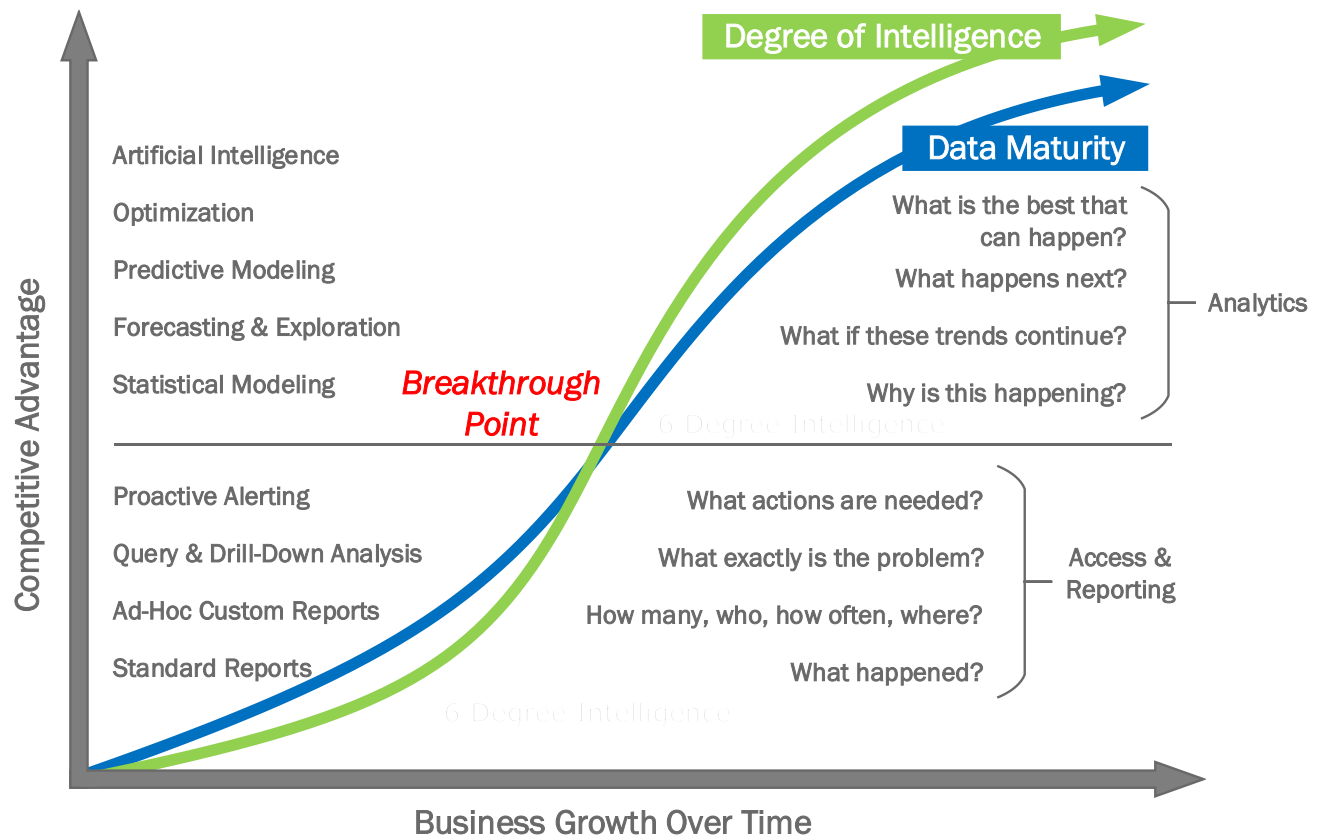

Figure 9: Intelligence growth versus data maturity curve over time.

CONCLUSION

SAS Viya has many tools which can be leveraged for endless possibilities. Using a sandbox approach helps provide focused delivery and strategic vision for managing data, developing analytical models, and communicating results to all key stakeholders.

CONTACT INFORMATION

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

Stephen Overton
6 Degree Intelligence
(919) 341-9667
soverton@6degreeintel.com
https://linkedin.com/in/overton

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.