PharmaSUG 2022 - Paper SI-099

Cyber Resiliency – Computing Platform Modernization

Susan Kramlik, Eunice Ndungu, and Hemu Shere, Merck & Co., Inc., Kenilworth, NJ, USA

ABSTRACT

This paper describes selected cyber-resiliency steps taken, challenges encountered, and lessons learned when implementing a modernization program to upgrade the statistical computing environment at our company. The components of the transformation included moving from Unix to Linux, integration with a SAS® Grid infrastructure system, and addition of web-based SAS Studio. The project's scope became larger than initially anticipated. System Development Lifecyle (SDLC) practices were also required because regulatory content is created using the computing platform. Careful planning, collaboration and alignment between the business users, IT, and Quality Assurance drove the success despite the challenges. Significant participation from the user community was also instrumental in the success of this project. Lessons learned were shared after the project deployment. The planning, scope change, adaptation, user participation, execution, and lessons learned may be of value to others undertaking similar endeavors.

INTRODUCTION

In a 2020 Harvard Business Review article[1], resilience is defined as "a company's capacity to absorb stress, recover critical functionality, and thrive in altered circumstances." It refers to an entity's ability to continuously deliver the intended outcome, despite adverse cyber events.[2] Cyber resilience is an evolving perspective that is rapidly gaining recognition. The concept brings together the areas of information security, business continuity, and organizational resilience.

The Harvard Business Review article further stated that resilience "requires a fundamentally different mental model of business — one that embraces complexity, uncertainty, interdependence, systems thinking, and a multi-timescale perspective," and that "Resilience is especially important today because the business environment is becoming more dynamic and unpredictable." It goes on to cite accelerated technological evolution as one of the reasons.

Similarly, within clinical trial analysis and reporting, there needs to be resilience due to accelerated technological evolution. That evolution impacts the way clinical trials are designed and conducted, the way data are collected and analyzed, and the need to maintain quality, efficiency, data security, and compliance with continually changing regulatory requirements. Being able to produce analyses even more quickly and efficiently while facing these challenges requires a reliable and resilient computing environment.

The Harvard Business Review article suggests ways to build structural resilience and mitigate adversity. Redundancy and migration strategy are two ways mentioned. Redundancy is a means to replicate capabilities that eliminate a single point of failure, while migration strategy is a means to maximize opportunity and minimize adversities. There are parallels to building resiliency in a statistical computing environment.

At our company, continual efforts to enhance information security, business continuity, and organizational resiliency drove the decision to make simultaneous updates to the statistical computing environment operating system and the computing software. These include redundancy as in adopting SAS Grid and SAS Studio and migration strategy as in migrating from Unix to Linux, which are all described in our paper. This paper describes selected steps taken, challenges encountered, and lessons learned when building resiliency in a clinical statistical computing platform.

RATIONALE FOR CYBER RESILIENCY

The migration from Unix to Linux was an initiative aimed at enhancing our company's data and infrastructure cyber resiliency. It was tied to the concept of staying healthy and recovering quickly. To ensure the realization of this idea, the organization embarked on a set of activities to move users from Unix to Linux, avail SAS studio as an additional editor, and leverage SAS Grid capabilities.

A. BENEFITS OF LINUX OPERATING SYSTEM

The move to the Linux operating system introduced the availability of new technology and modern infrastructure.
One of our earliest planning steps to optimize the move to Linux was to work with SAS® to determine the configuration for the new environment. A sizing questionnaire captured the relevant information needed to determine the hardware sizing requirements to support the programming community. Figure 1 below displays the components of the hardware sizing questionnaire.

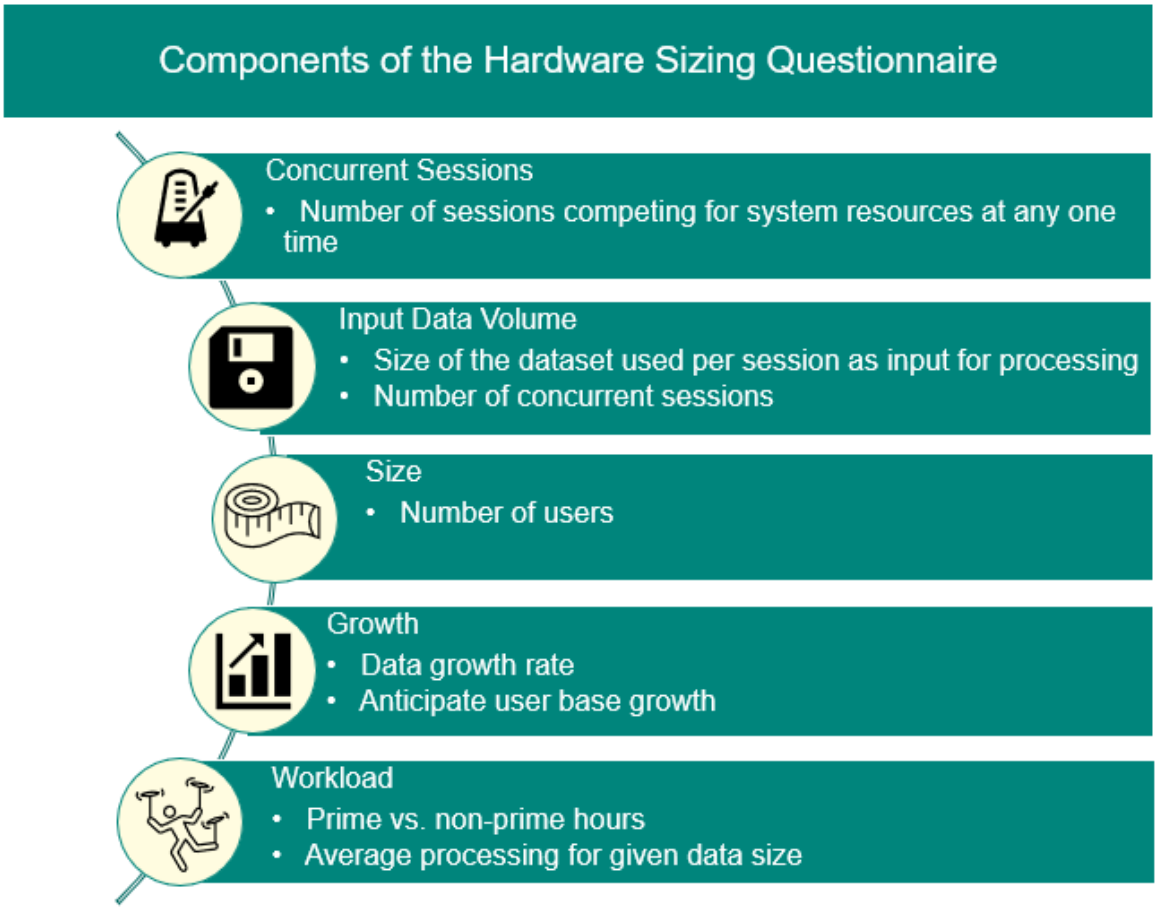

Figure 1: Components of the Hardware Sizing Questionnaire

B. BENEFITS OF THE SAS GRID

SAS Grid architecture allows for expanded, scalable, shared, centrally managed computational resources. There is expanded capability to distribute analysis and reporting activities across multiple servers. This results in reduced disruptions and downtime and minimizes risks to regulatory deliverables. The shared and highly available grid-enabled servers allow load balancing by distributing and balancing multi-user workload and reduces the risk of a single point of failure. Benefits of SAS Grid computing are shown in Figure 2. It also allows for parallel processing vignettes of code from the same program. Through SAS Grid manager central management of resources, the system can analyze growing volumes of data against available resources. This provides the capacity to seamlessly scale in response to business computational demands. Hence, the grid enhances cyber resiliency by managing multiple servers, jobs, applications and users, and overcoming disruptive outages.

The SAS Grid Infrastructure was enabled for SAS Enterprise Guide and SAS studio users. This allowed the users to leverage the server cluster benefits of the SAS® Grid such as expanded, scalable, shared, and centrally managed computational resources.

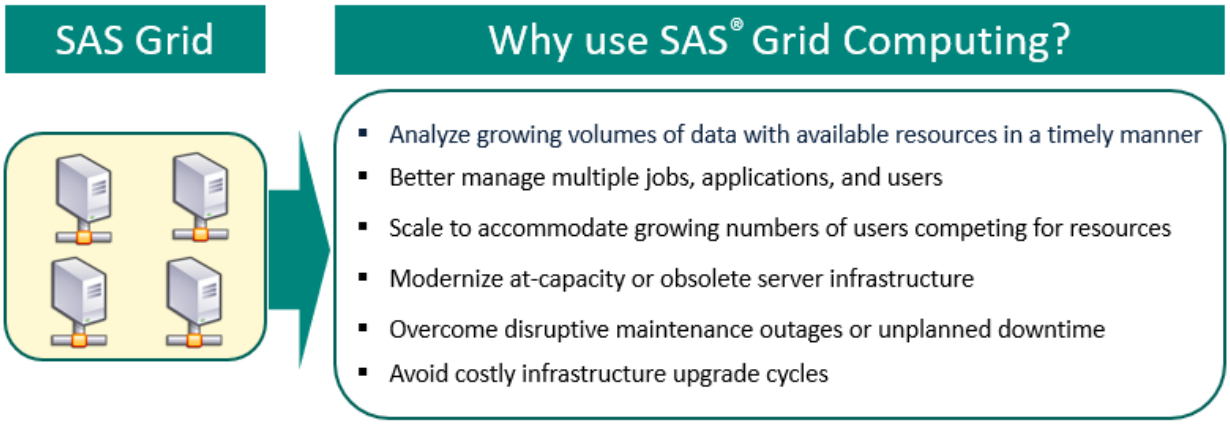

Figure 2: Benefits of Grid Computing

C. BENEFITS OF SAS STUDIO

The SAS Studio editor comes already enabled to interface and harness SAS Grid Computing's performance enhancements. It provides a web-based SAS client tool in addition to current PC SAS and SAS Enterprise Guide. Using SAS studio means removing the need to have SAS installed locally. Figure 3 highlights the benefits of using SAS Studio.[3]

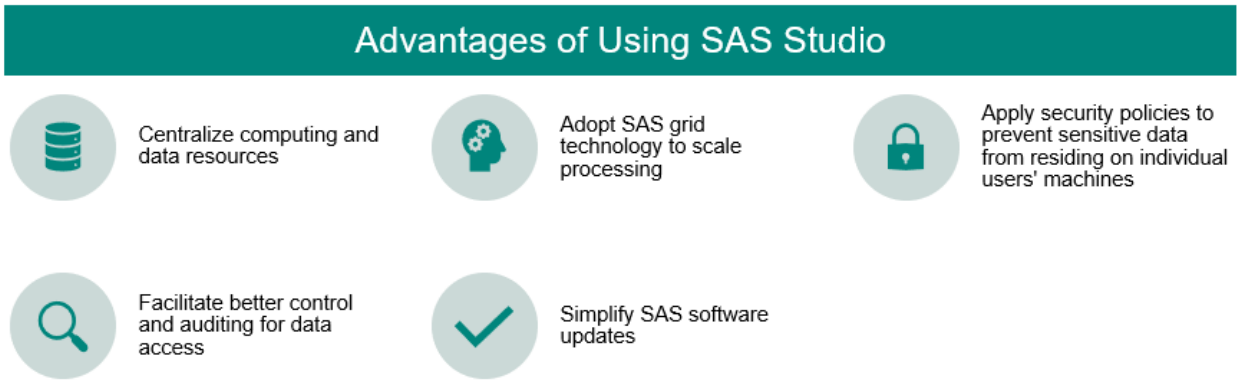

Figure 3: Advantages of SAS Studio

PLANNING & SCOPE

The project was estimated for a duration of 6-months. Various pre-planning aspects such as scoping, timeline, communication, and adoption, had been assessed. The project was evaluated as a migration of the computing server from Unix to Linux, i.e., an operating system change designated as a 'lift and shift' meaning system/user files could be copied from one environment to another. The impact to the application and users was deemed to be minimal. SAS was operationally the same across both platforms, and no impact was expected to user's code or way of working in SAS. No changes were expected from an end-user perspective, e.g., connection to Linux, the way programs are submitted, etc. would be the same and an end-user would experience no difference in working in the new computing environment. The assessment also confirmed no impact to existing functionality of the application; application software using Java, Oracle,

SQL, etc. that supported business functions was not impacted. The minor exception noted was related to user created code which needed to be evaluated for Unix system commands, which were expected to work.

CHALLENGES OF MIGRATION

As the project developed, the Linux Migration team faced challenges along the way. Table 1 summarizes the challenges encountered, reasons and suggested improvements.

Challenge	Reason	Improvement
Extended project duration	Under-scope	Thorough cross functional assessment Utilize resources with domain knowledge
Lack of comprehensive system and system utilization knowledge by individuals	Domain knowledge was not uniform across stakeholders	
Encountered SAS program issues	More thorough evaluation needed for cross environment SAS functionality	

Table 1: Summary of Challenges

Based on the initial assessment, the project was deemed an infrastructure change with a scope of 6 months. However, as the project progressed, testing teams found various issues with system functionality. The project timeline was extended to accommodate challenges, taking an additional 6 months to complete. This was evidenced by the numerous impact assessments and defects that were documented to identify unplanned events. Typically, one impact assessment is completed at project initiation that identifies all impacted components. However, additional impact assessments came about as a result of issues identified while the migration project was in progress (such as during testing or development activities).

Table 2 summarizes the most impactful issues encountered:

No	Issue	Impact
1	A legacy connection port remained in the data extraction systems when converted from Unix to Linux.	All data extraction capabilities failed.
2	EDC was not compatible with the Linux environment e.g., data representation for Unix (HP_UX_64) source dataset & macro catalog 'sasmacr.sas7bcat' used for data extraction was not compatible with Linux (LINUX_X86_64) due to a different operating system.	EDC data extraction function failed.
3	After migrating the datasets from Unix to Linux server, it was noticed that the copied datasets are under "data representation" of HP_UX_64. As the copied files are in a format native to another host, incremental updates using commands such "proc append", "proc sql update" and "insert into" failed when updating the datasets on Linux.	Update processing (row or record level update) failed.
4	Application service account name on both Unix and Linux were identical.	Confusion for ongoing work on Unix while parallel development work is ongoing on Linux.

No	Issue	Impact
5	Extraction macros did not contain conditional logic to allow use for other editors - SAS EG and SAS Studio.	Data extraction failed for SAS EG and SAS Studio.
6	System alias for SAS versions changed but change was not implemented in integrated utility.	Utility failed since aliases were not updated.
7	PC SAS cannot connect to the Grid using remote connection.	• User is not able to connect remotely and work remotely using PC SAS. • PC SAS users cannot leverage the GRID.
8	SAS Studio functionality was missing • Mapped drives were not accessible through SAS Studio • Dropdown server connection for SAS was not working	Lack of functions limits use of SAS Studio.
9	Issues with supporting utility • Displays Unix instead of Linux • Incorrect environment in utility about info • Unavailable SAS version was displayed in utility selection • .sas is displayed twice after file name in the utility transaction log	Confusion for users when using the utility.
10	SAS account name change to grid enabled account name.	User's startup program failed for remote libraries using the old account name.
11	Data needed to be anonymized for testing.	Resource constraints prevented follow through on commitment and BU scope was expanded.

Table 2: Summary of Impact

CHANGE MANAGEMENT

In order to ensure the cyber resilience solutions attained the expected outcome, we needed to implement change management. This is a structured approach that ensures change is thoroughly and smoothly implemented, and the lasting benefits of the change are achieved.

The initial phase of change management involved creating awareness of the initiative through general communication to all stakeholders. Next, the communication of the benefits was targeted at impacted roles and aimed at creating excitement around the initiative. Benefit awareness gave way to the knowledge and ability phase where Share and Learn sessions were held and training on the new environment offered, ensuring the impacted roles could ease into leveraging SAS EG and SAS studio on the SAS Grid infrastructure. Finally, we assessed adoption and continued to encourage it. Figure 4 reflects the change management details.

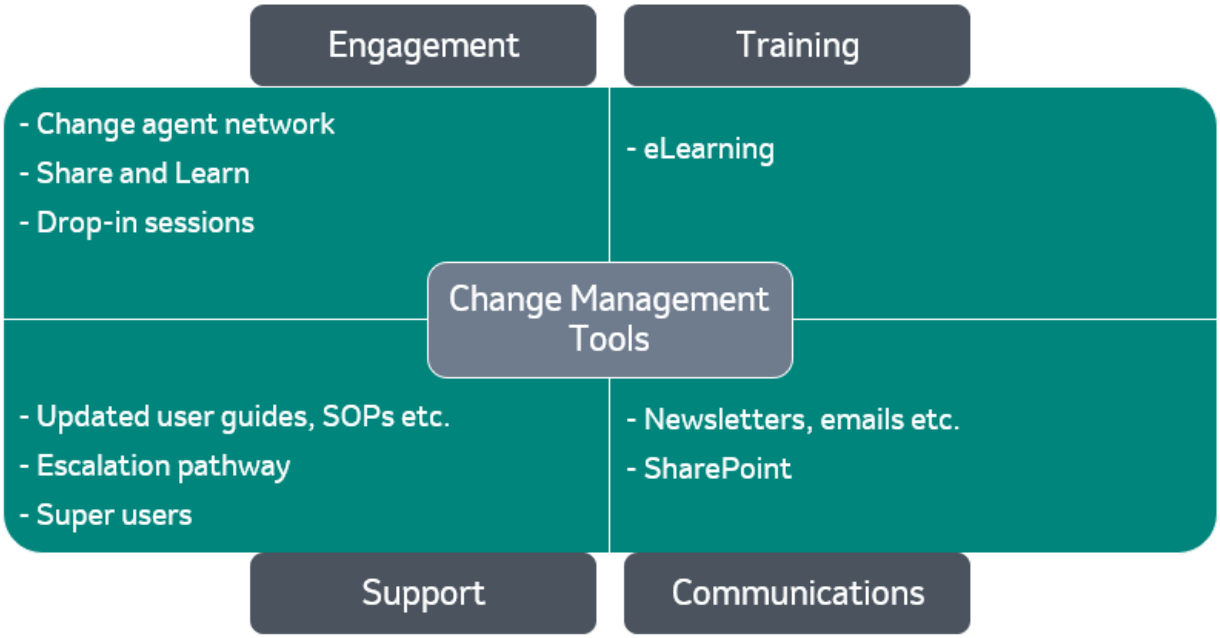

Figure 1: Change Management

REALIZED BENEFITS OF MIGRATION TO LINUX

One of the key benefits of the cyber resiliency project was availability of the SAS Grid. SAS Grid provided a multi-server environment where jobs could be spread across multiple servers, greatly enhancing delivery time. Along the same idea, a single program could be partitioned into multiple segments and executed in parallel across servers instead of code running sequentially line by line. Such parallel processing could greatly enhance end user productivity. At our company, there are two use cases which exemplify parallel execution.

A. PARALLEL PROCESSING USE CASE - SAS SIMULATION

One of our company's analysis and reporting teams needed to run SAS simulation code to determine p-values which required tremendous computing support. The team needed to run a large number of iterations, approximately 100,000 where 1,000 permutations take 1 hr. to run. In each iteration, a new set of random numbers is generated to determine a new assignment of treatment for each subject. Based on the new treatment assignment, the appropriate analysis model is processed, and results saved (including p-value).

Table 3 shows the outcome when a job is split into smaller pieces executed in parallel to reduce the computational time.

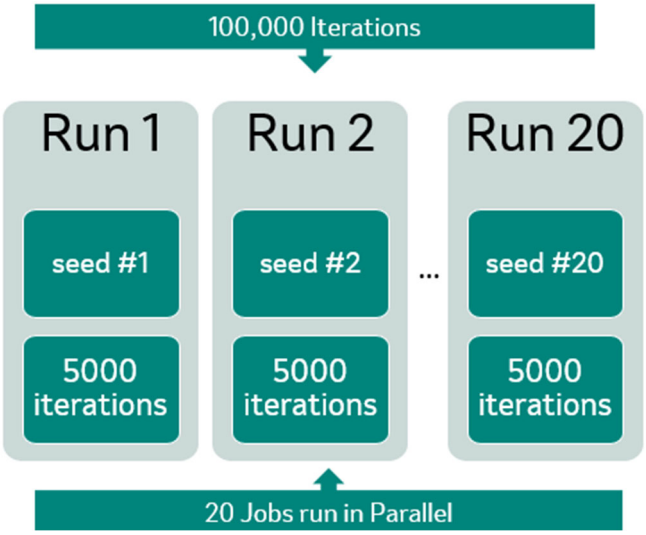

Table 3: Use Case for SAS Simulation

B. PARALLEL PROCESSING USE CASE – COMPARISON WITH SEQUENTIAL PROCESSING

We consulted with programmers to test parallel processing which was a new feature available with the Linux migration and SAS Grid implementation.

- Sequential processing runs code in sequence, while one section of code or program does not start until the previous one completes.
- Parallel Processing allows code to be partitioned and sections of the code to be run in parallel. Consequently, the time it takes to run all sections is the time needed for the longest running section. The same logic applies to running multiple programs in parallel.

The testing was done with a job that took 40+ hours to process to completion in the original Unix environment. After migrating the computing platform from Unix to Linux and special saswork setup, the same job finished in approximately 12 hours. The same program run in parallel completed in 9.8 hrs., a 18% savings. Table 4 summarizes the comparison between Parallel Processing and Sequential Processing on different environments.

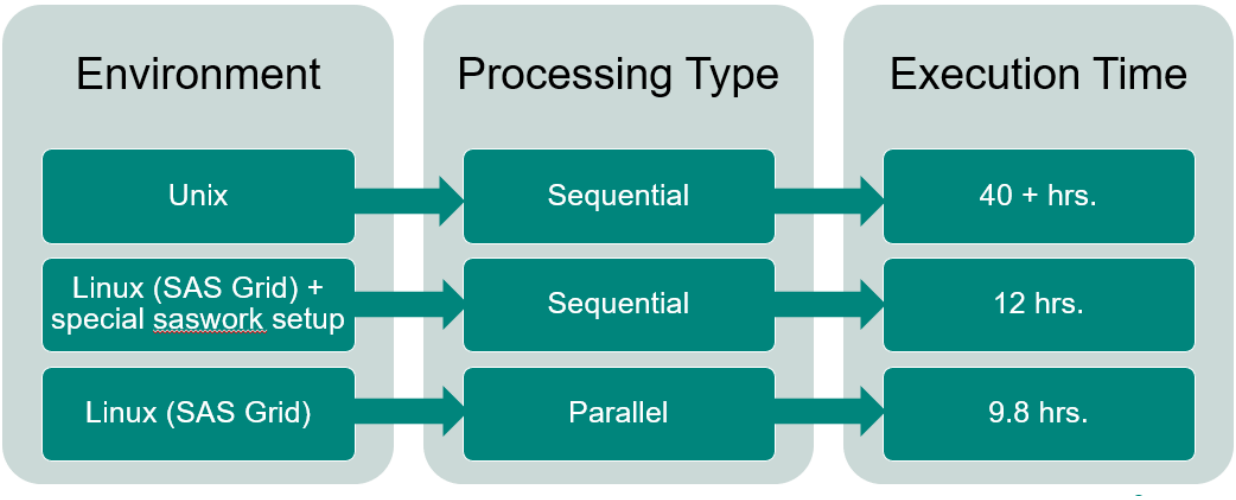

Table 4: Comparison of Parallel Processing vs. Sequential Processing

LESSONS LEARNED

Although the project team encountered obstacles as presented in this paper, the team did deliver the upgrades and the ultimate goal of enhancing our organization's data and infrastructure cyber resiliency was achieved. As consistent with the theme of improving through learning, the team reflected on the challenges and suggested preventive actions that can help with future projects.

Two main areas of improvements were highlighted:

Utilize Resources with Broad and in-Depth Domain Knowledge

- Covers all requisite aspects of the Unix to Linux migration
- Results in thorough impact assessments
- Reduces churn and provides on-time delivery
- Results in specific and more comprehensive testing
- Identifies issues early in development
- Identifies all SDLC deliverables impacted

Clear Communication across the Project Lifecycle

- Creates awareness across team members when documented decisions are changed
- Improves awareness when IT functions are used by busines unit (e.g., service account)
- Provides advance notice for impacted cross-functional areas (e.g., create training)

CONCLUSION

Building redundancy via SAS Grid reduces reliance on a limited number of servers for analysis and reporting. Similarly, use of SAS Studio, afforded redundancy by allowing another means of SAS access. The migration strategy for moving to a modern server type (Unix to Linux) introduced the availability of new technology and modern infrastructure

Simultaneous implementation of three means of improving cyber-resiliency were ultimately successful. As an organization, we felt we could apply what we learned to future efforts for large scale changes. While the execution uncovered unanticipated challenges, collaboration, partnership, and dedicated efforts of subject matter experts resulted in effective implementation and improved resiliency.

REFERENCES

[1] https://en.wikipedia.org/wiki/Cyber_resilience#cite_note-1

[2] https://hbr.org/2020/07/a-guide-to-building-a-more-resilient-business

[3] https://support.sas.com/software/products/sas-studio/faq/SASStudio_advantages.htm

ACKNOWLEDGEMENTS

Mustaq Ahmad

Merck & Co., Inc., Kenilworth, NJ, USA

mustaq_ahmad@merck.com

Matias Morrison

Merck & Co., Inc., Kenilworth, NJ, USA

matias.morrison@merck.com

CONTACT INFORMATION

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

Susan Kramlik

Merck & Co., Inc., Kenilworth, NJ, USA

susan.kramlik@merck.com

Eunice Ndungu

Merck & Co., Inc., Kenilworth, NJ, USA

eunice_ndungu@merck.com

Hemu Shere

Merck & Co., Inc., Kenilworth, NJ, USA

hemu.shere@merck.com

TRADEMARK INFORMATION