Date: 1 February 2023

Hardware Estimate for: University of Hawai'i of Manoa

Sales Tracking Number: 16403010
ServiceNow Number: RITM0515687

Proposed Software:
SAS® 9.4

- SAS® Data Management Advanced

Introduction:

==This is not an architectural design or installation document.==

==A new Sizing document is recommended for requests more than 90 days old due to potential changes in SAS software or hardware/operating system technologies.==

Critical Considerations:

==The sizing document does not guarantee any Service Level Agreement (SLA) or performance criteria.== There are too many variables outside of the control including quality and quantity of coding, contention for resources outside the cores and memory indicated by this sizing, central resources such as storage and networking utilizations, and many others. For this reason the Sizing team cannot guarantee any Service Level Agreement or performance criteria.

The estimation of resources herein is a no-fee service. **This estimate is not a performance benchmark, and does not provide any performance guarantee.** Any use or reliance on this estimate is at customer's risk. SAS disclaims any liability with regards to (i) the hardware sizing estimate provided herein; or (ii) University of Hawai'i of Manoa's reliance on the estimate. University of Hawai'i of Manoa acknowledges that it is responsible for all costs associated with procuring any hardware.

This service does not address any existing performance issues and is not intended to correct any existing performance issues. Performance issues should be investigated to ensure the root of the problem is identified. Appropriate monitoring of the environment and engagement with SAS Technical Support may provide insight into the root of any existing performance concerns.

The SAS Account team and University of Hawai'i of Manoa are responsible for ensuring the necessary licensing strategies. University of Hawai'i of Manoa is ultimately responsible to ensure appropriate

licensing of all third-party licensing, as well as any sizing details specific to any third-party products/solutions.

University of Hawai'i of Manoa acknowledges that failure to establish the SAS environment with the minimum specifications outlined within, as well as provide adequate storage and I/O, may result in undesirable end-user experiences as well as diminish SAS' ability to address performance concerns.

SAS solutions data is typically stored in a relational database. The database sizing has not been included in this estimate. University of Hawai'i of Manoa should work with their selected database provider for sizing information.

Data tier is recommended to be on a separate hardware server/partition.

The Sizing team is responsible for providing the required number of cores and memory for the required tiers of the solutions as requested. The Sizing team does not become involved in the deployment of the solution which involves the potential for combining/splitting of the required tiers of a solution.

Implementation issues should be addressed by the appropriate teams within Professional Services, Global Enablement Teams, or domain experts that can provide more specific information based on the needs of the customer and their environment.

University of Hawai'i of Manoa has an estimated workload comprised of:

System Wide

Totals for system-wide resources	
Total users registered in metadata?	7
Total concurrent sessions?	15
Average Annual Percentage Growth? (unless otherwise specified, 10% for both metrics is assumed)	Data volume 20 % Concurrent Sessions
If Web Report Studio (WRS) is used, describe the level of conditional highlighting in WRS reports (Low, Medium, High)	Medium

Prime Time

Session Type	Concurrent Sessions	Average Input Data Volume per Session(specify MB/GB)
DI Generated ETL	6	14 GB
Basic Statistics	6	14 GB
Advanced Analytics	3	14 GB

Batch

Session Type	Concurrent Sessions	Average Input Data Volume per Session(specify MB/GB)
Base SAS related to data prep / ETL	3	1GB
Business Rules Manager	3	1GB
Basic Statistics	3	1GB
Data Quality	2	1GB

Data Preparations (ETL)

SAS sessions related to ETL/data management *(During overnight BATCH Window)*	Concurrent Sessions	Average Input Data Volume per ession(specify MB/GB)	
Data volumes =< 1 GB	3	1GB	
What is your estimated percentage of annual data growth			10%

SAS sessions related to ETL/data management *(During PRIMETIME)*	Concurrent Sessions	Average Input Data Volume per ession(specify MB/GB)	
Data volumes > 10 GB	6	14 GB	
What is your estimated percentage of annual data growth			10%

SAS Analytics

Session Type	Concurrent Sessions	Average Input Data Volume per ession(specify MB/GB)
Basic Statistics (both Display Manager and Enterprise Guide)		
Data volumes > 10 GB	6	14 GB
Advanced Analytics (both Display Manager and Enterprise Guide)		
Data volumes > 10 GB	3	14 GB

Data Quality Products - Batch Processing

Data Quality Products – Batch Processing	
Please describe your entity data (Ex: 5 million customers with 22 fields totaling 372 Bytes, 100k products with 12 fields totaling 151 Bytes, etc…)	1. 50k clients with 36 fields totaling 425B 2. 70k incomes with 17 fields totaling 335B 3. 145k applications with 46 fields totaling 486B 4. 25k cases with 20 fields totaling 214B 5. 180k dependents with 20 fields totaling 183B
	4M claims, 250 fields, at 575B
Please describe the frequency and especially size (number of rows) in the batch to be cleansed. Approximately how many are new records and how many are updates	150k rows, where 50k rows are updated 100k are added
How long is the batch processing window for DM activities(Ex: 2 hours for DM)	7-hour window roughly
Can this batch be broken into sub-groups if needed, describe how it can be broken up (Ex: 5 groups of states beginning A-E, F-K, etc)	Can be broken down by time intervals (ex: 2017-2018), or by Provider ID.
How many fields will be matched (Ex: LName, Fname, City, State, Zip, Cell, ID…)	# fields 11 fields
How many fields will be used as part of your data matching criteria Of these, how many will be used for "fuzzy" (non-exact) matching	# fields First, Last, Middle, Sex, DOB Month, DOB Year, SSN, Addr1, Addr2, Addr3, Zip # fuzzy 11
Please describe the number or complexity of the clustering process: 　　Simple (under 5 match rules) 　　Moderate (5-10 match rules) 　　Complex (more than 10 match rules)	Complex

Data Quality Products – Batch Processing	

Please describe the expected number of data quality transformations[#]:
    Small (< 5 data quality transformations)
    Medium (5-10 data quality transformations)
    Large (11+ data quality transformations)
[#]A data quality transformation would be, for example, an operation to standardize, parse, identify, or change the case of data
	11+ DQ transfomrations

Will you use address validation If validation, what country for validation	Yes USA
Detail any additional needs or information you feel may be relevant to the DM processing	Initial data processing will be done at time of initial data ingest from sources to our DBMS. 1. We may need to deidentify our data for research purposes 2. Many Code sets that we have to validate: CPT, HCPCS, etc

SAS Business Rules Manager

SAS Business Rules Manager	
Please describe the expected number of business rules: -- Small (< 200 business rules) -- Medium (200-1,000 business rules) -- Large (1,000+ business rules) A business rule would be, for example, IF purchase > $100 AND customer type = 'new' THEN offer = 'Silver Card'	Medium
What are the maximum # of condition and action terms in a rule set?	>10
Do you anticipate your business rules designed to evaluate mostly text or numeric values?	Text
Do you plan to use large character strings (where large is > 100 characters) in your rules?	No
Do you plan on using lookup tables in your rules? **Note**: A Lookup table is a table managed in SAS BRM and used to validate if an input is valid, for example, if a product code is a valid product code or lookup a value.	yes
How many your lookup tables do you expect to use?	75
What is the largest lookup table you plan on referencing? -- Small: < 1,000 entries in a Lookup Table -- Medium: 1,000 – 10,000 entries in one Lookup Table -- Large: + 10,000 entries in one Lookup Table	Large

SAS Business Rules Manager	
How many lookups will be necessary for perform a standard set of Business Rules? i.e. the frequency of using lookups.	0-5 lookups frequency: 10-95%
Do you plan to use BY-GROUP rule flows?	**Yes**
For rules that summarize data using by-group processing: -- # of by terms -- # distinct by groups anticipated -- # of retained terms **Note**: retained terms are input values that are summarized during rule processing, such as total transaction amount by customer	A lot (>50), we don't know yet
How many transactions (rows) per day via batch process do you plan on applying Business Rules?	120k records/rows
How many fields will be used as part of your business rules criteria?	30-100
How many fields will be used for "fuzzy" (non-exact) matching? **Example**: using "LIKE" or pattern matching logic	5-40
Will your business rules fire on a large portion of input data or do you expect that your rules will fire only on an exceptional basis?	Large portion
What is the overall rate at which you expect rules to fire? **Note**: Rule fires occur when a business rule evaluates to true or that the condition is found in the data	Some will fire on everything, especially at ETL
How many concurrent users do you plan to use the system for development and testing business rules?	2

Hardware Estimate – Based on the above listed estimate of workloads, this solution should be deployed in an environment such as:

Production Environment:

<mark>This is not an architectural recommendation.</mark>

The customer is free to distribute and/or combine the cpu's as required by their architectural design. This is an estimate of the hardware resources required by the SAS solutions being sized. Implementation issues and architectural design for this environment should be addressed by the appropriate teams within Professional Services, Global Enablement Teams, or domain experts that can provide more specific information based on the needs of the University of Hawai'i of Manoa.

Tier	Cores / RAM
SAS Metadata Server	2 cores with 16GB RAM (8 GB RAM per core minimum)
SAS Compute Server	6 to 8 cores with 48 to 64GB RAM (8 GB RAM per core minimum)
SAS Mid-Tier Server (Web application server and components)	2 cores with 24GB RAM (24 GB RAM per server minimum)

This response is based on Intel Xeon E5-2600v4 or Gold 6200/6300 series processor with a clock speed of at least 3.30 GHz running Windows Server 2019, 64 bit operating system.

Depending on the specific chipset used, the number of cores may have been rounded up to accommodate the minimum 'cores per socket' requirement. For example, if the sizing determines that three (3) cores are required, but the chipset used is a quad-core CPU, then the sizing will be rounded up to four (4) cores. Depending on the chipset used, these minimums will typically fall on boundaries that are multiples of 4, 6, 8, 10 or 12.

The estimate is an order of magnitude estimate of the required hardware based upon the workloads as presented above. Data storage and overall I/O throughput of the environment may play an important role in the success of your project and should be carefully considered before beginning implementation.

Assumptions:

SAS is a very powerful tool and therefore can require significant resources. The following assumptions were used to arrive at the results and estimation of needed resources:

Overall Assumptions:

- This SAS hardware estimate is based on the workload descriptions provided by University of Hawai'i of Manoa. The provided workload descriptions are assumed to be consistent with the intended ordinary and typical uses of the software for the types of sessions characterized. At its core, SAS is a fourth generation programming language and allows customers the flexibility and power to use the software in novel ways that exceed the original intended ordinary and typical use of the software. Such use may result in workloads in excess of those described in the estimate and may require additional resources.

- The estimates included herein assume that the servers are to be used for the SAS workloads as described above. Addition of other third-party applications to the same server as SAS is generally discouraged; the estimated resources are for deployment of SAS applications only.

- Changes to the workload (in either number of sessions or data volumes), operating system, or preferred vendor or chipset may render this sizing void. In the event of changes, the SAS Account Team should resubmit the questionnaire with the needed updates for reprocessing. This sizing assumes the 'best practice' that all cooperating servers of the SAS environment, and any related servers such as those that house data/RDBMS used by SAS, are connected via a high-speed LAN without significant degradation due to traffic, latency, or other network device.

- Hyper-Threading creates two logical processors from one physical processor core, allowing the operating system to schedule the execution of two threads/processes per physical core. This allows more instructions to execute per machine cycle increasing processor throughput between 15 and 20%. sizing estimates are based on the number of physical cores with Hyper-Threading active.

- The proposed software components require a SAS® Metadata Server, Middle Tier and Compute Server. It is recommended that the production metadata, middle tier and compute servers run in separate operating system instances.

- A dedicated machine for the SAS Metadata Server is recommended to simplify tuning, management, and diagnostics, but is not required. The SAS Metadata Server is multi-threaded and a multi-core server is recommended.

- The estimate is based on a typically available CPU/chipset as available from the preferred hardware vendor with a given performance characteristic. Different CPU types may affect the resources required.

- SAS® Office Analytics/Data Management Advanced is a bundle which includes several imbedded SAS solutions and technologies; only those components noted within and the workload provided are expected to be used. Any additional workloads may impact the available resources.

- University of Hawai'i of Manoa will have to determine best practices for the hard drive configurations of the various tiers with their preferred hardware vendor.

- The implementation team is responsible for designing and developing the most efficient job streams possible to meet the desired processing window; however it should be noted that due to job processing pre-requisites, data updates, the completion of other system's processing may impact or even prevent the SAS environment from achieving the documented goals.

- The Sizing team strongly recommends that the environment should be closely monitored and scaled to support the required workloads to meet the business objectives. This would include not only the computing environment, but also networking, storage and I/O infrastructure.

- This sizing estimate is based on a combination of guidelines provided by SAS R&D, SAS Product Management, test data, and field experience.

- Our best practice is to provide the topology as developed by R&D and try to provide as unified a presentation of the requirements as possible. When questions on deployment arises, the Sizing team defers to the account team to make a determination, based on the customer's desires and the best practices of the solutions involved, on the most practical installation.

Hardware and Operating System Assumptions:

- Core counts are guidelines only. Complexity, user loads, and data models can require additional system resources

- Unless specified otherwise, SAS generally requires a minimum of 6 - 8 GB of RAM per core for the Production Compute, Metadata, and other tiers not used for the web-based mid-tier applications. For the Mid-Tier web server, the general requirement is a minimum of 8-12 GB of RAM per core. These requirements may vary depending on the solutions installed or the number of users/sessions supported in accordance with Operating System Guidelines and SAS recommendations, page file space should be set to 1.5 to 2 times the amount of physical memory.

- The machines should be configured for maximum memory bandwidth; this will be dependent on the actual processors/machines selected. University of Hawai'i of Manoa should work with their preferred hardware provider to achieve this objective.

- This estimate assumes minimal Java Virtual Machines (JVM) will be deployed on the SAS Mid-Tier supporting web application servers (for those solution components that provide a web-based interface such as Web Report Studio, etc.) Depending on the specifics of the customer's deployment, additional application servers and/or JVMs may be required. In these cases, memory may need to be increased to support the additional environments and/or servers. Please consult with SAS Technical Support or SAS Professional Services to determine the memory requirements for the specifics of your installation.

Storage Assumptions:

- SAS sessions can consume a wide range of I/O bandwidth across its associated file systems depending on the type of processing that is being performed. The amount of sustained I/O that is recommended for the SAS solution is a factor of the estimated number of cores on the compute tier times the average sustained I/O required per core. There may be peaks of I/O throughout the workday and possibly as jobs are executed during non-prime hours (when typically most ETL work is performed).

- SAS tends to be I/O intensive. SAS may utilize the SAS WORK disk location to temporarily house results as it calculates statistics or performs certain functions. To maintain adequate performance it is essential that I/O be as fast as possible between the SAS Compute Server and the SAS WORK disk(s). It is also critical that sufficient I/O throughput be allocated so as to maintain the cores of the Compute Server in a "busy" state.

- University of Hawai'i of Manoa may wish to consider the peak I/O throughput requirements of their system and work with their storage provider to ensure that the storage environment can provide the level of I/O required. A significant percentage of "performance problems" reported to SAS Technical Support can be directly attributed to insufficient levels of I/O throughput.

- Recommended I/O throughput rates for the SAS Data and SAS WORK file systems:

  o SAS Data – for permanent SAS data files, your application throughput requirements may dictate a minimum I/O throughput rate of **100-150 MBs/sec per core**, **minimum**, in the system. Reads and writes to the file system will occur during the ETL process.

  o SAS WORK – Chronic and heavy reads and writes are common for this file system. Depending on the intensity of your workload it may be much higher. We also

recommend that you create a second file system (UTILLOC) for the utility files created by sorts and summarizations.

- Depending on the architecture and deployment, multiple compute tiers need access to a common data area. This may require the use of a centralized storage mechanism such as a Clustered File System (CFS). It is suggested that the customer work with the SAS Account Team to engage the appropriate domain expertise to determine the optimal configuration for the specific requirements of the installation.

Please refer to these papers for more information on how to configure your storage for SAS applications:

Additional Information – Recommended Reading:

ETL Performance Tuning Tips

http://support.sas.com/resources/papers/ETLperformance07.pdf

How to Maintain Happy SAS®9 Users – IT focused

http://support.sas.com/resources/papers/proceedings16/SAS6201-2016.pdf

Windows Server 2012R2 considerations:

http://support.sas.com/kb/52/767.html

The Latest Tuning Guidelines for Your Hardware Infrastructure:

http://support.sas.com/resources/papers/proceedings14/SAS107-2014.pdf

Solving SAS® Performance Problems: Our Methodology -

https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/3490-2019.pdf

Important Performance Considerations When Moving SAS® to a Public Cloud

https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2020/4312-2020.pdf

SAS® Product Support for Virtualization Environments

http://support.sas.com/techsup/pcn/virtualization.html

Moving SAS® Applications from a Physical to a Virtual VMware Environment.

http://support.sas.com/resources/papers/MovingVirtuaVMware.pdf

Frequently Asked Questions Regarding Storage Configurations:

http://support.sas.com/resources/papers/proceedings15/SAS1500-2015.pdf

Best Practices for Configuring Your I/O Subsystem for SAS®9 Applications

http://support.sas.com/resources/papers/proceedings16/SAS6761-2016.pdf

To test if your storage can sustain the desired IO throughput before you install SAS, please use the following tools:

Testing Throughput for your SAS 9 File Systems: Microsoft Windows platforms

Note: Please work with your hardware vendor to estimate and configure your storage architecture.

Disclaimer