Paper 1098-2017

Classroom Success with SAS® Grid Manager and SAS® Visual Analytics: Coping With Big Data

John Jernigan, NC State University – Institute for Advanced Analytics, Raleigh, NC;

Ken Gahagan, SAS® Institute Inc., Cary, NC; Cheryl Doninger, SAS® Institute Inc., Cary, NC

ABSTRACT

The Institute for Advanced Analytics struggled to provide student computing environments capable of analyzing increasingly larger data sets for its Master of Science in Analytics program. For the fast-paced practicum, the centerpiece of the curriculum, waiting 24 hours for a FREQ procedure to complete was unacceptable. Practicum proposals from industry were pared down (or turned down) because the data sets were too large, depriving students of exciting and relevant learning experiences. By augmenting the practicum architecture with an 18-node computing cluster running SAS® Grid Manager, SAS® Visual Analytics, and the latest High-Performance Analytics procedures, we were able to dramatically increase performance and begin accepting terabyte-scale practicum proposals from industry. In this paper, we discuss the benefits and lessons learned through adding these SAS products to our analytics degree program including capability versus complexity tradeoffs, and the state of our current capabilities and limitations with this architecture.

INTRODUCTION

MSA HISTORY AND OVERVIEW

The Institute for Advanced Analytics was founded at North Carolina State University in 2007. The Institute's flagship program is the nation's first Master of Science in Analytics (MSA) degree. With a mission to produce the world's finest practitioners of analytics, the Institute serves as a focal point for a university-wide, interdisciplinary collaboration among a large number of faculty. The MSA degree is an intensive, full-time, 10-month learning experience with an innovative curriculum developed exclusively for students in the program. It is NC State's leading master's degree measured in terms of student outcomes, and its graduates (which number as many as 120 annually) are among the university's most sought-after and highly compensated.

MSA PRACTICUM OVERVIEW

The MSA practicum is a core educational experience in the Master of Science in Analytics program. A 5-person student team completes a real-world analytics project using large data sets and industry-standard tools over a period of 8 months. The Institute collects no fees for the data provided by sponsoring organizations, all projects fall under a confidentiality agreement, and the student team's output and intellectual property is solely that of the sponsor. In the Institute's first 10 years of operation, students have completed 134 practicum projects among more than 100 sponsors spanning virtually every industry, including some of the world's best-known brands, as well as governmental agencies and non-profit organizations.

PREVIOUS PRACTICUM COMPUTING ARCHITECTURES

The time constraints of the Master of Science in Analytics create intense challenges for a system administrator. Running the MSA practicum from September to April (i.e. 8 out of the 10 months of the degree), the student projects are a major educational foundation of the program. The MSA classroom curriculum, a separate component of the program, is already a major sink of student time and attention. Thus, it is worth emphasizing that downtime of the computing environment that hosts the MSA practicum is extremely undesirable. Students need to be able to work on their project wherever they are and whenever they decide to, as the consequences of lost productivity often mean the student team must forgo a project objective for lack of time resources to delay it. In addition to this, the only time of the year that is appropriate for R&D and improvements on this architecture is a narrow window in the summer. For all these reasons, IT issues need to be fixable immediately, and a premium is placed on systemic simplicity and flexibility.

We designed the practicum environment so that a team of 5 students can simultaneously use Remote Desktop Connection to get a Windows desktop session in their dedicated team server. Using Windows Server and a full suite of office productivity software, this is a comfortable and familiar home base for analytical activities. Most importantly, it keeps the confidential project data isolated between teams.

FIRST GENERATION

The very first architecture for the MSA Practicum (ca. 2007-2012) was a virtualized cloud environment running on blade servers. The operating systems and data stores were on a NAS system with undesirable latency, which resulted in overall performance often being no better than the students' own laptops. This was not working well, with complaints of long computation times in Base SAS and SAS Enterprise Miner, and overall sluggish system response.

SECOND GENERATION

Due to the I/O demands of Base SAS and SAS Enterprise Miner (which were in conflict with the aforementioned NAS latency), we then migrated to a bare-metal architecture hosted in-house and focused on affordable high-performance I/O. From 2012 to 2015, this system allowed the practicum to surge forward. The diagram shows the eventual configuration with a mixture of entry-level enterprise-class and desktop-class servers. Note that this is not a grid architecture, but rather a group of dedicated and isolated team servers.

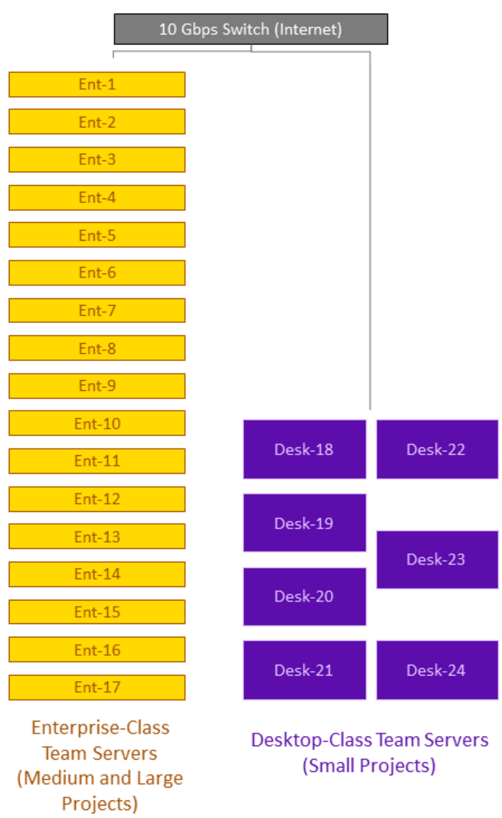

Figure 1: Second generation practicum computing architecture consisting of individual team servers and no clustering. 5-person student teams can simultaneously log in with Remote Desktop to their own team server. Upper limit on performance was not ideal; projects with initial data sets larger than ~30GB tend to frustrate students.

I/O performance was achieved in this architecture by using dedicated SSDs for both the operating system and SAS Temporary Files. Overall this performed well for smaller data sets, and was relatively simple and cost-effective.

Note about SAS Middle Tier deployment:

MSA students use SAS Enterprise Miner quite a lot, as well as SAS Forecast Studio. In a Windows Server environment, this necessitates a SAS Middle Tier installation along with a Metadata Server. The typical approach is to centralize these components for the sake of resource efficiency. However, due to the small system administration staff, and (at the time of the architecture design) a beginner-level grasp on security controls within SAS Platform Administration, we decided to install a dedicated Middle Tier and Metadata Server on each team server (i.e. a "one machine" type of planned deployment). Twenty-four student team servers meant twenty-four middle tiers, which is indeed a bit of wasted memory, but it drastically shortened R&D time. With the team server securely locked down at the operating system and network layers, each team of students could be made SAS Administrators in the context of their own team. SAS Platform Administration on this architecture takes minimal attention from us so that we can focus it on academic and user-level challenges. Additionally, we do not find that the middle tier and SAS Metadata Server cause any undue burden on I/O or CPU performance on a server with only 5 concurrent users. The tradeoff in ease of system administration for a cost of ~8GB of extra system memory usage is worthwhile from our perspective.

LIMITATIONS WITH SECOND GENERATION (PREVIOUS) PRACTICUM ARCHITECTURE

The architecture described above works very well, and affordably, for small project data sets. Less than about 30GB, we find that the student teams can wrangle their data, conduct exploratory analysis, and design and validate models without feeling like they are waiting excessively on the software. For data sets larger than 30GB, we started to hear frustration from students (keep in mind that the teams are often duplicating the data sets and analyzing them simultaneously). As one might expect, we were finding that every year the size of proposed industry projects was growing steadily, with always more data available to the students. We were unfortunately beginning to turn down (or pare down) industry projects with an excuse of "big data" challenges. The Institute for Advanced Analytics lamented these lost educational opportunities, and we set out to understand the nature of our limitations and a plan for improvement.

Early in the exploratory phase of a student project, a team might run a FREQ procedure on the entire data set and notice that it took 24 hours to complete while just a small fraction of the CPU resources were being utilized. Many traditional SAS procedures operate in single-threaded mode, meaning they can only utilize 1 processor at maximum, though a computer may have e.g. 8 processors available to the operating system. High-Performance Analytics procedures can instead operate in multithreading mode, which in this case would utilize all available processors, however we did not have access to these procedures at the time. Complicating the situation was poor I/O performance which did not allow the FREQ procedure to access as much data as it could handle. Multiple students were exploring the data at the same time. Though SAS Temporary Files and the operating system had dedicated SSDs, the data itself was stored on a single-spindle HDD. This meant performance dropped tremendously under heavy simultaneous random loads due to high disk latency. In worst-case scenarios, system memory would be depleted causing the operating system to swap to the hard drive (i.e. use it for overflow memory operations to prevent an application or system crash) at a severe performance penalty.

Regarding storage capacity, we were quickly running out of disk space as well. A team of 5 students will tend to duplicate data sets into their own Windows folders for convenience and logical separation. We find that a typical project data set will get replicated perhaps 10X among the 5 team members as they copy it into personal folders, subset it, modify it, etc. All things considered, a 2TB HDD is fine for a small project (e.g. less than 30GB), but on the order of hundreds of gigabytes, these disks fill up fast. Every year we were having to upgrade hard drives as a quick bandage on our storage capacity problem. In many cases, industry sponsors would propose a terabyte-scale project, but we would have to compromise with a sampled data set. Of course, sampling is perfectly fine conceptually, but in some project cases sampling a data set is not feasible, and the project unfortunately could not be accepted.

Another limitation from a student's perspective was due to either inefficient use of CPUs or poor disk I/O, and that was the infeasibility of quickly processing 1000's to 10,000's of model iterations. Whether it was training or validation, sometimes the student team would arrive at a final model with precious little time before the final presentations to faculty or company executives occurred. It is not uncommon for the student team to have startling analytical results to share with their sponsor companies, and persuading skeptics with statistical arguments can sometimes be easier when models have been run on *all* provided data. In the critical last stretch of a project, sometimes a rapid, massive number crunch is the key to delivering an effective argument. With single-spindle HDDs holding team data (and their tendency to degrade performance exponentially under simultaneous loads), a full-team effort to process data in a hurry would have the unintuitive consequence of actually slowing progress down.

Finally, having seen the astonishing speed of in-memory analytical techniques, we hungered for the ability to load hundreds of gigabytes, even entire terabytes, into an in-memory analytical platform.

Considering all this, we needed a leap forward in computational and storage performance as well as memory and storage capacity.

THIRD GENERATION (CURRENT) PRACTICUM ARCHITECTURE WITH SAS GRID MANAGER AND SAS VISUAL ANALYTICS

How would we grow our capabilities to continue providing our students with challenging and contemporary industry analytics projects? Generally, we could have continued to put temporary bandages on our existing system, scrapped it completely and started over, or a hybrid of the two.

The hybrid solution to our chronic resource constraints was to augment our capabilities rather than replace them. An 18-node grid was installed with drastically larger specifications than the existing team servers. The grid's new capabilities could benefit any student team by choosing to utilize it from their existing team server. Even better, all 24 student teams could access the grid capabilities simultaneously.

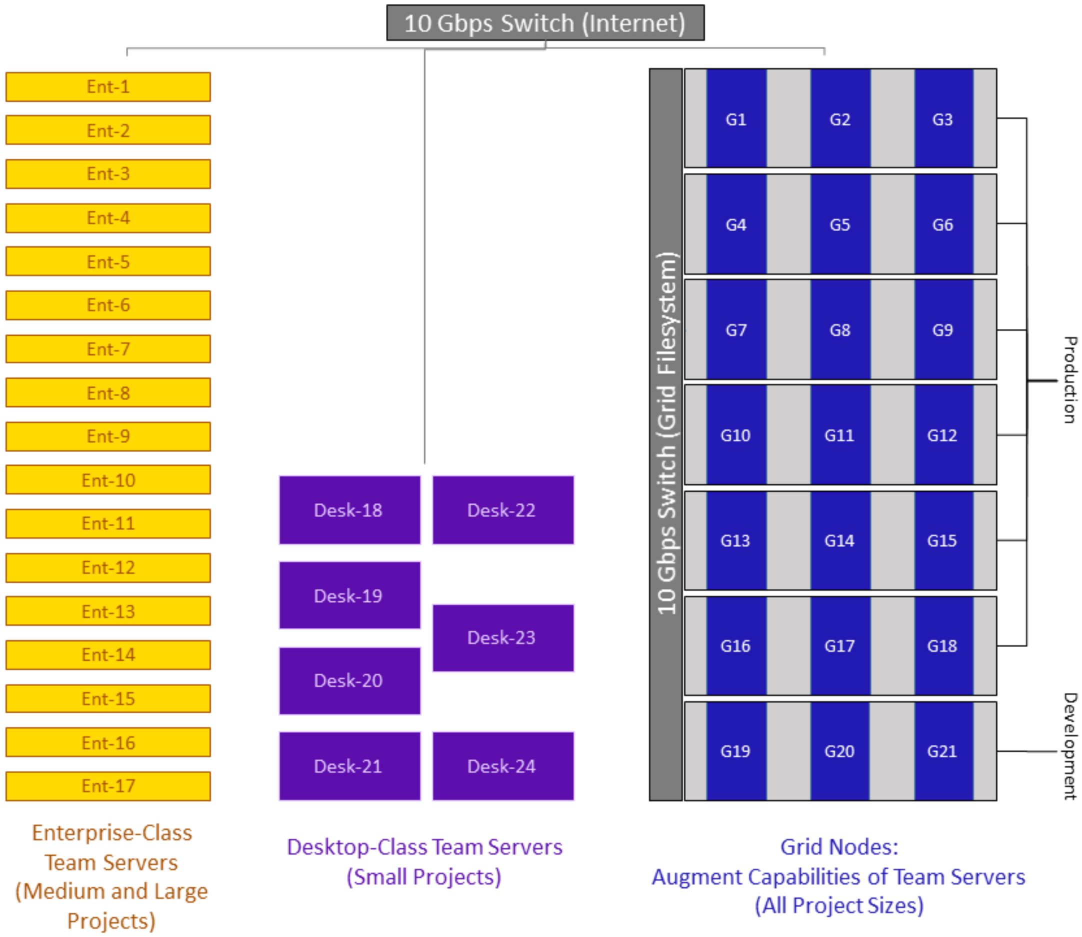

Figure 2: Third generation (current) practicum computing architecture consisting of the second generation system with the addition of an 18-node grid. Any student team can access grid capabilities simultaneously.

The difficult questions were what software to install, and what features to configure? Furthermore, how far will we push for higher performance (or more features) given the tradeoff of more complexity?

We had to be careful not to bite off more than we could chew. The Institute for Advanced Analytics had 1 full-time system administrator with an already-full slate of responsibilities. Additionally, the security policies surrounding the MSA practicum were stringent due to the presence of confidential commercial data. For example, all team data sets must be completely isolated from each other, which is to say that we had to design for 24 separate security domains and ensure that adequate preventative security controls were implemented. These were large challenges.

Considering the tradeoffs of performance and complexity (both of administration and user education), the following technologies were implemented:

- Analytics technologies:
    - SAS Grid Manager
    - High-Performance Analytics
    - SAS High-Performance Data Mining
    - SAS Visual Analytics and SAS Visual Statistics

- System infrastructure:
  - Red Hat Enterprise Linux 6
  - IBM General Parallel File System (GPFS; currently known as IBM Spectrum Scale)
  - Apache Hadoop HDFS (in Secure Mode utilizing Kerberos)

The benefits of these technologies will be elaborated on in the following section.

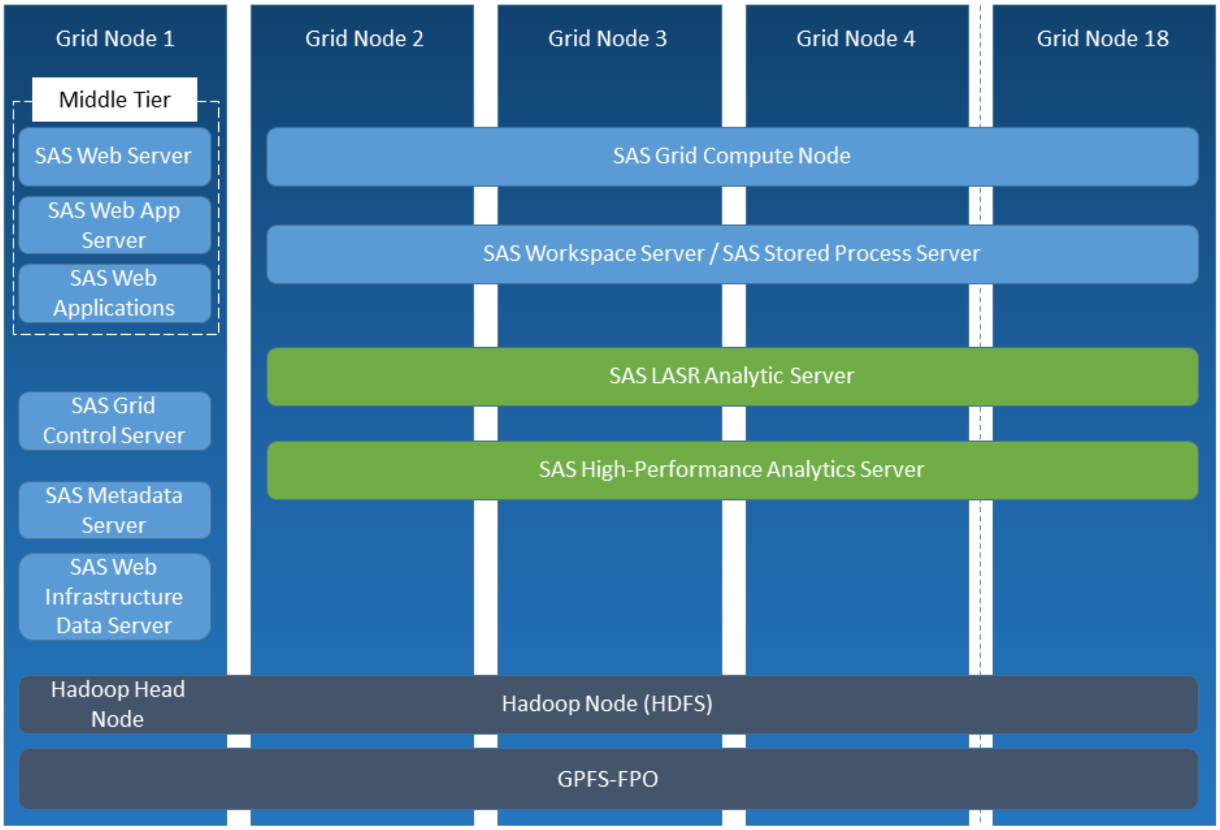

Figure 3: Software architecture of third generation (current) practicum grid. Note the shared deployment of both SAS® Grid Manager and SAS® LASR Analytic Server (e.g. for SAS® Visual Analytics).

IMPROVEMENTS AND NEW CAPABILITIES

The new Linux grid architecture is an augmentation of our existing system rather than a complete replacement. Aside from drawing further value from our previous investment, we found this approach provided two important benefits. Most importantly, it provides flexibility. Each team still has their own isolated Windows environment on bare metal which, in the name of “getting things done”, frees up all lanes for IT staff’s troubleshooting, risky experimentation, and oddball configurations (all of which have proven difficult to avoid in shepherding a student team through their project) on an *individual* team basis which limits potential collateral damage. To make our second point, we must state our biggest concern with unfamiliar centralized systems: should there be lengthy system-wide downtime on the grid architecture (e.g. an emergency filesystem reconfiguration, which has happened) the students must still have a productive, albeit smaller and with fewer features, environment on their Windows servers. One of course argues that *any* of these systems is susceptible to a system-wide failure of some sort, but we at the Institute for Advanced Analytics acknowledge our lack of troubleshooting expertise (at the time) with the centralized Linux grid and new SAS technologies when giving our initial trepidations. The point here is that the overall MSA practicum is fault-tolerant to the new additions and capabilities of the Linux grid architecture.

We chose to implement several SAS solutions on the same Linux grid to get the most value from our equipment, trying to keep the system no more complicated than necessary. The architecture is scalable if we grow and, in most aspects, redundant for high-availability.

New Software Available to Students:

- SAS Grid Manager and High Performance Analytics:
  - SAS Grid Manager is a rock-solid foundation for providing multi-user access to the grid along with workload balancing and high-availability. By itself, it does not make single-threaded procedures or data steps run across multiple processors. To take advantage of the massive number of processors in the grid, we needed High Performance Analytics to provide high-performance procedures. Capable by default of utilizing multiple CPUs on a single grid node (i.e. up to 32) in Symmetric Multiprocessing (SMP) mode, or with extra parameters set, of utilizing in Massively Parallel Processing (MPP) mode up to all CPUs across all nodes (i.e. up to 288). These two technologies together make a powerful and reliable combination.
- SAS High-Performance Data Mining:
  - SAS Enterprise Miner, when connected to the new SAS Grid Manager, can utilize SAS High-Performance Data Mining nodes even without having them licensed, but with the caveat that they will only work in SMP mode. That is, they can utilize all of the processors on a single grid node (i.e. up to 32). With HPDM licensed, however, those same SAS High-Performance Data Mining nodes can execute in MPP mode (i.e. utilize potentially all processors across the entire grid, i.e. up to 288).
- SAS Visual Analytics and SAS Visual Statistics
  - Operating on top of the SAS LASR Analytic Server, SAS Visual Analytics can process data in-memory at a great increase in speed and efficiency due to avoiding relatively slow disk I/O. The user interface itself offers capabilities and experiences not found in Base SAS, with its point-and-click visual nature.
  - A huge benefit of the SAS LASR Analytic Server is that libraries loaded in-memory can be accessed by SAS Grid Manager jobs if using High-Performance Analytics procedures. In the case that you want to use Massively Parallel Processing (MPP) mode with a high-performance procedure to utilize hundreds of CPUs in the grid, having each node involved access SAS libraries in-memory while avoiding relatively slow disk I/O is tremendously appealing.
- IBM General Purpose File System (GPFS; now known as IBM Spectrum Scale)
  - GPFS is a distributed file system that makes the entire grid's collective storage available to all nodes. It provides redundancy and high-availability, and with a 10Gbps dedicated network, good performance.
  - This is the primary filesystem for our SAS solutions.
- Apache Hadoop HDFS
  - Why a second distributed filesystem? It is the only way we know for SAS Visual Analytics to be able to save results out of memory and back to the disk. Likewise for any other in-memory libraries loaded into SAS LASR Analytic Server that were used by High Performance Analytics procedures. Our students naturally want to save the results of their modifications and analysis, and due to the shared nature of the grid resources, they cannot expect data loaded into memory will persist long-term.

- Due to the MSA's security policy requiring each of the 24 student team projects to be securely isolated from each other in the filesystem, we configured HDFS in Secure Mode using a local Kerberos instance. The major SAS services were also configured to integrate with Kerberos and HDFS. This is a heavy, complicated operation compared to the rest of our installation. However, the benefits of in-memory SAS libraries are for naught if they cannot be saved back to the disk somehow.

New Hardware Resources Available to Students:

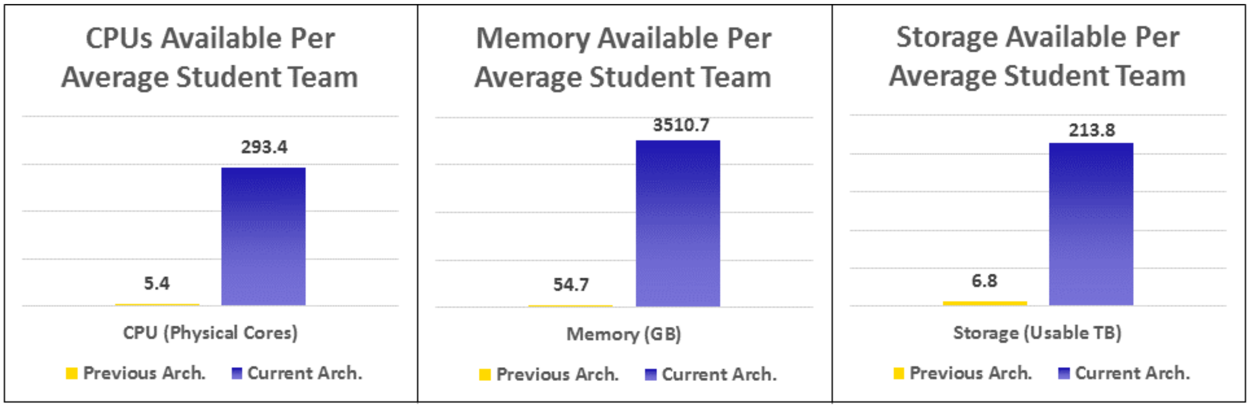

Figure 4: Comparison of resources available to an average student team between previous and current architectures. Since there are 2 classes of team server hardware with different specifications (see Appendix A), the architecture values reflect an average value of each resource available per team.

In the figure above, we see the current architecture provides a drastic increase in hardware resources available to the student teams (e.g. memory availability increased 6,318%). Of course, each team has to share the grid resources with their peers, but in the MSA practicum we tend to find that teams need large resources sporadically rather than consistently. This resource sharing is working, at least for now, without overt attention to resource management.

Positive Consequences of New Architectural Capabilities:

- What constitutes "big data" to the MSA practicum is now very different. Instead of 1 TB for an initial data set being a high threshold for acceptability, we are not even sure what our upper limit is yet. We no longer have to turn down or slim down good industry projects because of their size. In fact, the choke point is looking like the data transfer logistics between the industry sponsor and the Institute's equipment, rather than student analytical capability.
- Now that we have High-Performance Procedures, we can fully utilize SMP and MPP modes of processing when we choose the appropriate procedures (i.e. compute with all processors within a single node, or among all the processors across multiple nodes). The speedup allows for faster development all around, along with the inevitable down-to-the-wire number crunching at the last minute. Note that the optimum number of nodes with which to process is rarely the maximum available due to storage I/O limitations. For example, to compute on all 288 CPUs across 18 nodes with a large data set, it must ultimately be transferred to each node's memory. The transfer logistics (i.e. overhead) can cause higher amounts of parallelization to results in longer processing times. We are continuing to develop best practices in this regard.
- Where we had concerns about memory constraints before, they no longer remain with 3.5 TB of memory available to the SAS LASR Analytic Server. This unlocks the door for rapidly investigating very large data sets with SAS Visual Analytics and particularly with the clustering and regression modeling (e.g. GLM) capabilities of SAS Visual Statistics.
- Whereas we attributed poor disk I/O to multiple random loads on a single HDD spindle, with 270 spindles across the Linux grid (see Appendix A for specs), that is no longer an issue. Our

particular IBM GPFS performance is not on-par with high-performance solid-state drives, but it holds up well under a multi-tenant environment. We are continuing to explore the performance characteristics of this filesystem under realistic heavy loads.

- With 120 usable TB in GPFS and another 87 TB in HDFS, we do not worry about having enough storage for large projects. This is more than plenty for our foreseeable future.
- Any student team who wants to run a very large number of model iterations needs only to use the appropriate High-Performance Analytics procedures and configure them accordingly for SMP or MPP mode for a SAS Grid Manager job. Our students have tremendous new computing capability available to them.

LESSONS LEARNED AND CURRENT CHALLENGES

The addition of the grid nodes into our latest architecture introduced significant complexity from a system administration and, to a lesser extent, student user perspective. The following are some of the issues and challenges we surmounted or continue to face in integrating the new system architecture.

- More server parts (i.e. number of disks, storage controllers, etc.) mean there are more physical pieces that can fail. Troubleshooting is more complicated with high-end enterprise servers and requires more skill of the system administrators. All of this can generate undesirable downtime for our students. The lesson we learned is to have spare server parts on-hand, either on the shelf or running in non-production equipment. We can then swap them out immediately (which aids tremendously in troubleshooting), and then do a warranty replacement at normal pace.
- The MSA program has an intense academic schedule that leaves little time for R&D and major maintenance upgrades each year. By partitioning some of the nodes to be a dedicated development environment, we were able to increase the amount of R&D we could accomplish without impacting the production system's stability. This translates to new capabilities for our students, and the latest software upgrades available from SAS.
- MSA students are typically not computer science experts. They had to learn the SAS language earlier in the academic year, and though they become Advanced SAS certified, they're not writing expert code effortlessly. They sometimes write convoluted models that are not easily parallelizable (by manually executing pieces simultaneously) by the SAS Grid Manager system. Instead of training students to write better code, we try to leverage the use of High-Performance Analytics procedures to let the procedures do the parallelizing instead of the student.
- Operating two disparate environments for each student team (a dedicated team Windows server "home base" and a communal, though securely partitioned, Linux grid environment) introduces some complexity and confusion to the students. For one thing, students are unlikely to use only one of the two environments, and that means their data is duplicated in multiple filesystems. They need to keep track of what data is where and in what state of existence. This is asking a lot of a group of busy students who may not be technically savvy. Ongoing user education and process improvements are occurring in this area, for example with logical file and folder naming or better team communication and documentation.
- As a follow-up to the Windows and Linux environment disparity, SAS software notices when a Linux-based procedure is operating on data that originated from within a Windows environment. It automatically engages Cross Environment Data Access (CEDA) which implies an inherent performance loss. For all the work of building out a higher performing grid, it is unfortunate to have some of that eaten up in compatibility operations. However, the alternative, which is to properly import data into the Linux environment in the first place, is unrealistic in practice as a default solution. For very large data sets it takes quite a long time (it does not appear to have an SMP or MPP mode for parallelization), and students can get easily confused about which data is in which state of compatibility. Our practice at the moment is to accept the performance penalty of CEDA for teams who expect to work with data across both Linux and Windows. For teams who are likely to utilize the SAS Grid Manager and/or SAS Visual Analytics environments exclusively, then we aid the students in importing the data at the beginning of the practicum.

- Moving data between the Windows and Linux filesystems eats up some of the performance gains by the Linux grid computation. If it were possible to leave the data in the grid filesystem (i.e. GPFS) throughout the entire project, that would be ideal. In practice, this does not tend to happen with our students, but is an area we continue to look into to improve the overall experience and performance. Additionally, we believe CIFS (Samba) to be the preferred method for easily allowing students to securely transfer data between the two environments.
- Resource management is not yet a concern for us, but we will eventually have to address this. Cf. Paper 289-2014: SAS® Grid Manager, SAS® Visual Analytics, and SAS® High-Performance Analytics: Sharing Hardware and More in the SAS Global Forum proceedings for more information about Linux-based approaches to resource management (e.g. quotas) for an architecture like the one in this paper.
- Even when multithreaded tools (e.g. High-Performance Analytics procedures) are available, they are rarely a direct, drop-in replacement for familiar single-threaded tools (e.g. traditional SAS procedures), and the learning curve to migrate towards them can sometimes be an impediment to our students. One often has to ask if they are really saving any time at all by spending *X* amount of time learning a more efficient analytical tool in order to save *Y* amount of processing time. Many times, for our students under tight timelines and with already-saturated mental landscapes, the math points to the slower, familiar single-threaded methods being preferable. This does not seem ideal. We expect our MSA curriculum to update in this regard, and shift the tide on this topic.

CONCLUSION

NC State University's Master of Science in Analytics (MSA) practicum aims to provide real-world experiences to students by giving them industry problems to solve that are relevant and also of beneficial consequence to the sponsors. As proposed projects got larger and larger over time, students struggled to keep up due to sluggish computing performance and the fast pace of the MSA degree. Faced with performance limitations, student learning experiences suffered. By augmenting the practicum architecture with an 18-node grid running a collection of SAS software, we drastically increased computing resources for the student teams by factors of 30X – 60X across CPU, memory, and storage capacity resources. Empowered by this new hardware, we installed SAS Grid Manager, SAS Visual Analytics and High-Performance Analytics procedures on the new grid. The integration of the new grid with the existing student team servers was complex and challenging given the MSA program's IT staffing resources and design requirements. Running multiple filesystems across 2 operating systems and managing 24 separate security domains was a mighty challenge to administer and of which to educate users. However, the benefits were very valuable to the MSA program, as the new system allowed for very large contemporary practicum projects that would otherwise be dismissed due to prohibitive "big data" challenges. The practicum team learning experience is a firm practical foundation for the MSA degree curriculum, and one of the primary reasons that MSA students have some of the best job placement outcomes throughout the whole of NC State University. With the scalable architecture described in this paper, the authors expect MSA student learning experiences to keep pace with industry project opportunities for a long time to come.

APPENDIX A: HARDWARE SPECIFICATIONS

Individual Team Servers (Desktop-Class Hardware)

- Dell Optiplex 990 Small Form Factor
- 1x Intel(R) Core(TM) i7-2600 CPU @ 3.4GHz (4 cores)
- 32GB RAM (16GB is the max "supported" by Dell but the motherboard accepts more)
- 1x 2TB hard disk drive LFF (Data)
- 1x 256GB solid state disk (OS)
- 1x 512GB solid state disk (SAS® Temporary Files)
- Windows Server 2012 R2

Individual Team Servers (Enterprise-Class Hardware)

- Lenovo ThinkServer RD550 1U
- 1x Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40GHz (6 cores)
- 64GB RAM
- 4x 4TB hard disk drive LFF in RAID 10 (Data)
- 1x 256GB solid state disk (OS)
- 1x 1TB solid state disk (SAS® Temporary Files)
- Windows Server 2012 R2

Grid Nodes (18x Production; 3x Development)

- HP Proliant SL4540 chassis with 3 nodes/chassis and 7 chassis
- 2x Intel(R) Xeon(R) CPU E5-2450 v2 @ 2.50GHz (8 core)
- 198GB RAM
- 2x 512GB hard disk drive SFF (OS)
- 15x 1TB SATA hard disk drive LFF (Data)
- 2x 10Gbps NICs (1x Internet; 1x iSCSI)
- Red Hat Enterprise Linux 6

Network Hardware

- 2x Juniper EX4550 10Gbps switches (1x for Internet; 1x for GPFS / HDFS)
    - Switches configured for jumbo frames (MTU 9000) and cut-through.

CONTACT INFORMATION

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

John Jernigan
Data Security Specialist
NC State University – Institute for Advanced Analytics
john_jernigan@ncsu.edu

Ken Gahagan
Director, Software Development, Host R&D / Performance Analysis
SAS® Institute Inc.
ken.gahagan@SAS®.com

Cheryl Doninger
Senior Director, SAS® R&D
SAS® Institute Inc.
cheryl.doninger@SAS®.com