SAS Visual Analytics (SAS VA) – Network Considerations and Data Management/Governance

Nicholson Warman, Hewlett-Packard, Cary, NC, USA

ABSTRACT

SAS Visual Analytics (SAS VA) is an entirely new approach to information analysis and management of large volumes of data, using an in-memory analytics engine to provide lightning-fast responses to analytical questions. With this new class of products come opportunities for your organization, but to make the most of these opportunities, there is a need for planned choices on data networking, data provisioning strategies, plus the need for the appropriate level of computing power, all in accordance with your corporate information governance policies and practices. As you soon realize, this isn't just Foundation SAS anymore, but a rich analytical environment that is able to rapidly analyze large amounts of data. This paper will present an approach for approaching these choices, using HP configurations as examples but with references to other alternatives, while attempting to address some of the most common questions asked by HP customers during the SAS VA pre-sales and implementation stages.

INTRODUCTION

In the past year, since the April 2012 SAS® Global Forum in Orlando where SAS VA (SAS Visual Analytics) was showcased, we have seen the release of SAS HPAS (SAS High-Performance Analytics Server) in August/September 2012 and the release of several versions of SAS VA (5.1, 5.2), bringing us to the current SAS VA 6.1 in December 2012. This new innovative approach to data management, data transfer and data presentation, represented by SAS VA and now SAS HPAS as well, means that there are opportunities for organizations to integrate a best-of-class solution into their business environment, to see and profit from the speed of insights faster than their competition. Some fundamental pre-planning can help make the most of this new solution. This paper was developed from some of the most common customer questions raised during the pre-sales and implementation phases, and to build understanding on the ecosystem represented by SAS VA.

While images and some technical details referenced in this paper are specific to HP, possible alternatives will be called out so that, regardless of the vendor technology used, an understanding of the scope, balance and possible methods of addressing the issue should result. Further, while at least one other Linux version may possibly be used, experience and the 13 HP reference configurations are based on Red Hat Enterprise Linux (RHEL) 6.3. Finally, recommendations are current to SAS VA 6.1; contact SAS with all questions on the SAS VA application.

Experience on initial installations is showing that correct networking is proving to be a key success factor for a timely and effective implementation. While two 10 Gb/sec network data ports are preferred, one is required to support of a SAS VA configuration; this only addresses the user interface connections and external data sourcing connections (hereafter referenced as the client-facing connections). This is the public face for SAS VA; there are two additional networks in place, the data network within the configuration (referenced as the SAS VA-internal network), and the management network of physical and virtual ports (referenced as the Management network). The latter networks, for internal SAS VA data and for management traffic, are essential and these networks need to be planned for by your corporate IT network management, to ensure the IP address range/capacity exists in the IT infrastructure in time for when the SAS VA configuration is to be delivered and implemented. One final item to address is the corporate information governance policy on network access.

WHAT IS DATA NETWORKING? A PRIMER FOR WORKING WITH SAS VA

At its simplest, Data Networking is a set of cables, typically copper but more often lately made of glass fibers, that are used to connect servers and data switches, so that information can flow in both directions, to meet business goals and objectives. The goal is to use the least amount of cabling to connect two or more servers together and to maximize full-duplex (simultaneous bi-directional information flow) throughput between them. Typical units of measure are in Megabits per second, Gigabits per second or Terabits per second, abbreviated as Mb/sec, Gb/sec and Tb/sec; Gb/sec on Ethernet is known as GigE, and Gigabytes per second as GB/sec. While higher bandwidth technologies are becoming available, in the range of 40 to 54 Gb/sec at the switch to switch level, for now at least, expect server connectivity to remain at most with Ethernet at 10 Gb/sec and Fibre Channel (FC) at 8 Gb/sec. Because of framing bits (outside the scope of this paper), 10 Gb/sec Ethernet has identical carrying capacity as 8 Gb/sec FC, yielding 1 GB/sec (1 gigabyte per second) of data transfer capability. The Switch (or a Fabric) used to network these servers together needs to be able to support bi-directional/full-duplex service at that rate for every processor (not just every server). More and more, we are seeing multiple processors in servers, and each processor has 2, 4, 8+ cores apiece – each core can be likened to a processor unto itself. Historical data says that to keep a

processor busy that is running Foundation SAS, a minimum network capacity of 50 MB/s per core is required for data provisioning; HP has documented system throughput on Red Hat Enterprise Linux (RHEL™) with SAS delivering over 100 MB/sec/core sustained throughput. The balance in the paradigm of processor speed, network speed and data delivery speed of storage media is constantly shifting, with new advances in storage and new processor designs together putting the pressure on the network to keep up in provisioning the processors successfully. And as new and improved networking technologies hit the market, this pushes the "choke point" back to the processors and/or the storage technologies.

Some terms and numbers which are important and which need to be adjusted to your vendor of choice, follow. Using HP equipment, up to 16 blade servers per enclosure, and up to 4 enclosures to a rack are the key numbers. If rack-mount servers are used, then depending on the form-factor used, up to 16-32 rack-mount servers are permitted per rack. The key words are UP TO as there are heating/cooling, earthquake and building code regulations, etc. which all impact how full a rack may be loaded. Industry standard racks are 72 inches tall, divided into 42 units (each unit is known as a U). Rack-mount servers come in 1U, 2U and 4U high form factors, although other sizes also exist. An HP BladeSystem enclosure is a 10U form factor, so the blades are about 4U high and about 1U wide as the servers are racked on their side in an enclosure. Rack-mount servers by comparison are the width of the rack and are quoted in *nU* for height only (that is, 1U, 2U, 4U, etc.).
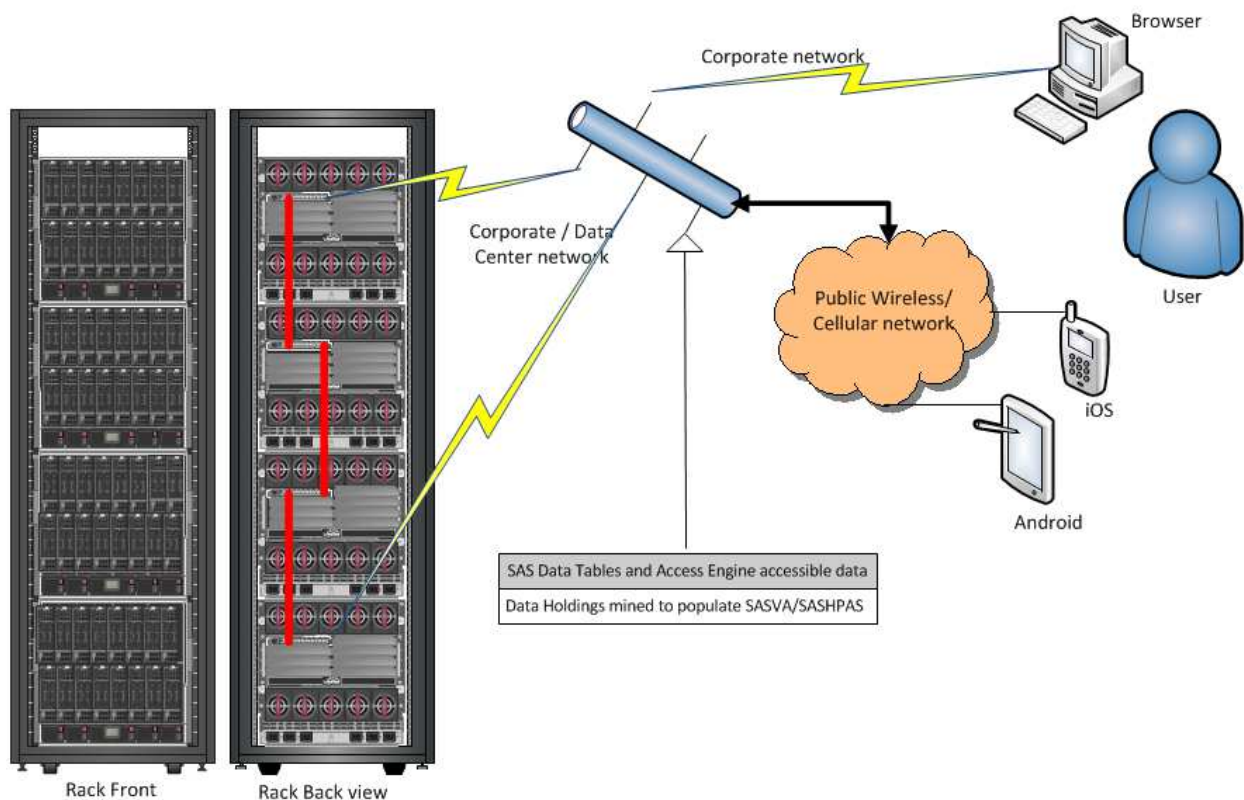
Figure 1. SAS VA network interaction view, using HP BladeSystem servers and enclosures

For SAS VA, all communications within the configuration needs to be at the fastest data rate possible. For the user interaction, however, all data communications would be handled at the customer site's standard communications speed and technical standard; most typically, this will be 10 Gb/sec Ethernet, although some European sites still only provision at 1 Gb/sec. The SAS VA application gives the user the option to interact with the system by iPad™ or Android™ tablet, as well as by traditional web browser. These data pipes are also used by the SAS VA Administrator to load data from various sources into the SAS VA configuration. This is a mix of public internet access on the same ports as those used for access to potentially sensitive internal data sources, and in the most restrictive case, this becomes an Information governance matter to address. In such cases, to alleviate concerns that Audit and Compliance officers might raise, I recommend identifying one port as externally facing, and keep the other port on a more secure channel for the data provisioning needs. So what could, in a regulated industry, become a barrier to success you instead can turn into a success by gaining their active support in identifying and rectifying a point on which they might otherwise need to raise an objection. Even if not working in a regulated industry, thinking about such possibilities can help prepare a smooth implementation of SAS VA as concerns that others may raise during implementation can be addressed positively and quickly.

To understand the networking implications better, there are some system architecture terms used in SAS VA. The design has a SAS Head Node, typically the blade in Figure 1 in the bottom left-most location in the diagram; the remaining blades are the Data Nodes. As the terminology indicates, SAS VA is a gridded solution, just not a SAS/Grid solution. From a data perspective, all analytical data are retained in the Data Nodes, and the sole use of these data stores is to persist the data holdings across restarts and reboot conditions; all analytical data work is conducted solely in-memory. For each server, the SAS, web server and other application software components reside in RHEL's EXT4 file system on a small partition of each server's disk space

There are two different but related networking strategies possible, depending on the hardware vendor's technology, and how networking is implemented in your organization. The first method is used by default with SAS VA, as implemented with HP BladeSystems, and the second approach is that used with HP's rack-mount servers and some other vendors' servers, in concept if not in fact. Since the HP BladeSystems give the easiest networking explanation, we'll start with the networking design details for these configurations, and then call out the differences for other networking solutions. When in doubt, keep in mind that the rack-mount servers typically represent in fact or in concept how virtually all servers will be networked, especially in data centers using older or basic design principles. First, we'll cover how to estimate the number of IP addresses required for any configuration, then look into how they are distributed in the various hardware variations possible, then how the network architecture will work in each of those four categories of hardware variation to deliver on these networking strategies.

IP ADDRESS REQUIREMENTS AND USAGE

IP Addresses required – the formulae

There are four possible scenarios for the IP addresses needed, and they fall into three sub-classes (i.e., client-facing, SAS VA-internal, Management network). A minimum of two client-facing ports are recommended, SAS VA-internal configurations are one per server, and the remainder are on the Management network as virtual IP addresses except for two per enclosure (OA physical ports) and one per switch; for rack-mount servers, expect two data ports for SAS VA-internal and client-facing use, plus a Management network for system management. The total IP address estimation formulae, based on known hardware variations, are:

- All data traffic remains within the HP BladeSystem server enclosure or within at most four HP BladeSystem enclosures:

  2 + 2*[number of servers] + 4*[number of enclosures]

- All data traffic crosses more than four HP BladeSystem enclosures but not more than four racks:

  2 + 2*[number of servers] + 4*[number of enclosures] + [number of racks]

- All data traffic crosses more than four racks of HP BladeSystem enclosures:

  2 + 2*[number of servers] + 4*[number of enclosures] + 2 + [number of racks]

- All data traffic is managed by a single large, enterprise-class switch, be they HP BladeSystem or rack-mounted servers (also known as pizza boxes, due to their shape):

  a) Rack-mount:

  1 + 2*[number of servers] + 2 + [number of switches]

  b) BladeSystems:

  2*[number of servers] + 4*[number of enclosures] + [number of switches]

Depending on the nature of the networking equipment being used, IP address doubling may also be required if the network environment cannot adaptively switch routing between the configuration switches, in order to ensure one network connection per server will always exist; these formulae do not address that complication, but conservatively expect to double the numbers otherwise calculated. Also, while the number of IP addresses for rack-mount servers is slightly lower, they are all physical IP addresses that are required, and each requires a cable for each port used. BladeSystems include virtual IP address, so only the IP address assignment is required, as will be explained shortly.

IP Addresses required – examples to show network segmentation and Physical/Virtual IP Addressing

Taking our four scenarios above, the following tables will show a small, medium and large example for each scenario, so the network counts are more clearly visible. As the numbers change, it more clearly shows why organizations should plan for the IP address needs when considering/ordering the hardware for their SAS VA configuration. This planning will ensure that when the ordered equipment arrives, everything is prepared to welcome the new arrival to the corporation. To be clear, SAS VA-internal addresses are found only within the configuration and its switches, and the Management network (iLO and VC) addresses may be physical or virtual, depending on whether they are addressing rack-mount or BladeSystem servers. Nevertheless, the various IP addresses are all required. As a further note, all switches used require a single IP address each, added to the OA address count. The blue-shaded boxes are virtual addresses, while the white boxes are actual physical networking ports on the various configurations. The grey-shaded boxes are the criteria on the left side of the tables and the total port count appears on the right side of each of these tables.

- All data traffic remains within the HP BladeSystem enclosure or within at most four HP BladeSystem enclosures:

Server Count (enclosures/racks)	Client-Facing addresses	SAS VA-internal addresses	Management network addresses (OA)	Management network (iLO + VC)	Port Count
8 (½/1)	2	8	1	8 + 1	20
32 (2/1)	2	32	4	32 + 4	74
64 (4/1)	2	64	8	64 + 8	146

Table 1. Formula: 2 + 2*[number of servers] + 4*[number of enclosures]

- All data traffic crosses more than four HP BladeSystem enclosures but not more than four racks:

Server Count (enclosures/racks)	Client-Facing addresses	SAS VA-internal addresses	Management network addresses (OA+switches)	Management network (iLO + VC)	Port Count
96 (6/2)	3*	96	12 + 2	96 + 12	221
144 (9/3)	3	144	18 + 3	144 + 18	331
192 (12/4)	4	192	24 + 4	192 + 24	441

Table 2. Formula: 2 + 2*[number of servers] + 4*[number of enclosures] + [number of racks - 1]

Note that in this scenario, the number of client-facing IP addresses increases with the number of racks and switches.

*   While this number is not required to scale as server counts increase, in order to preserve our **_n + 1_** redundancy, at least one port per switch should be used to ensure that at least two ports total can connect successfully to the customer network; in the case of only two racks, a minimum of two ports would be required for one or both switches to maintain this minimum connection requirement.

- All data traffic crosses more than four racks of HP BladeSystem enclosures:

Server Count (enclosures/racks)	Client-Facing addresses	SAS VA-internal addresses	Management network addresses (OA+switches)	Management network (iLO + VC)	Port Count
320 (20/5)	4	320	40+7	320 + 40	731
384 (24/6)	4	384	48+8	384 + 48	876
512 (32/8)	4	512	64+10	512 + 64	1,166

Table 3. Formula: 2 + 2*[number of servers] + 4*[number of enclosures] + 2 + [number of racks]

The number of client-facing IP addresses is fixed with this scenario, as there are two IRF 'North' switches now included, and they each would provide two client-facing ports, so if one switch failed, we would still have two ports intact to the client-facing network.

- All data traffic is managed by a single large switch, regardless of whether the servers are HP BladeSystems or rack-mounted servers:

a)   *Rack-mount:*

Server Count (racks)	Client-Facing addresses	SAS VA-internal addresses (2 ports per server)	Management network addresses (switches)	Management network (iLO @ 1/server)	Port Count
16 (1)	2	16 + 1	1	16	36
160 (10)	2	160 + 1	1	160	324
480 (30)	2	480 + 1	1	480	964

Table 4. Formula: 1 + 2*[number of servers] + 2 + [number of switches]

Since there is a single switch handling all the network traffic, only two client-facing IP addresses are required; the recommendation is not to drop below two client-facing ports. Rack-mount servers support two IP addresses and one OA address – Virtual Connect is only found in BladeSystems. The number of servers varies according to a number of factors, but typical numbers of rack-mount servers per rack run from 16 to 32 (16 shown in these examples), depending on the U height of the servers. Again, different vendors have different density packing rules, due to issues of ventilation and heat management within the rack. Net weight can also be a factor limiting the number of servers in a rack as well. Further, only one port per server is used

on the SAS VA-internal network, except for the SAS Head Node which uses two (one for client-facing and one for SAS VA-internal network use). Attempting to use both ports on the SAS VA-internal network, to preserve redundancy, will make the configuration's network plan extremely complicated. Finally, every port will need a cable to connect it to the data switch.

b) BladeSystems:

Server Count (enclosures/racks)	Client-Facing addresses	SAS VA-internal addresses	Management network addresses (OA+switches)	Management network (iLO + VC)	Port Count
16 (1/1)	2	16	2+1	16 + 2	39
160 (10/3)	2	160	20+1	160 + 20	363
480 (30/8)	2	480	60+1	480 + 60	1,083

Table 5. Formula: 2*[number of servers] + 4*[number of enclosures] + 3*[number of switches]

Note that, by comparison to the Rack-mount option, BladeSystems allow for full redundancy in networking with typically far fewer cables (sum of the white boxes on each line) but a slight increase in the total number of IP addresses to be assigned (virtual addresses.

HP BLADESYSTEMS – SIMPLE NETWORKING (FOUR OR FEWER ENCLOSURES)

The enclosures used by HP BladeSystems include a mid-plane fabric switch (up to 7 Tb/sec fabric) and may contain up to 16 servers per enclosure. Other vendors use the term Chassis and they are interchangeable in concept, but only if a switch is included with the same capabilities provided by a BladeSystem enclosure can they be considered functional equivalents. Rack-mount servers are also nicknamed "pizza box servers", due to their shape, and are the other form factor used for servers. In all cases, the servers are accessed from the front of the rack, and the back of the rack (also called the back-plane) is where power and networking connections are made.

One cable per enclosure (minimum) is required to interconnect these enclosures (see Figure 1), with one or two additional cables required to complete the VC domain. These cables together make up to 64 blades appear as a single networked entity, only requiring the physical and virtual Management network and client-facing ports; all SAS VA-internal traffic is contained within this configuration and is not exposed to the corporate network, in theory; in practice a SAS VA-internal traffic network may be needed to sequester the intra-configuration traffic, depending on the switching technology in use at the customer site. This capability, to interconnect multiple enclosures into what appears as a seamless whole, is known as laddering or Intelligent Resilient Framework (IRF). The terms frequently associated with IRF relationships are:

- East-West, denoting
    - intra-enclosure data traffic
    - Inter-enclosure connections (so long as only one IRF hierarchy level)
    - inter-switch traffic (at the same level in the IRF hierarchy)
- North-South, denoting
    - inter-enclosure connections
    - enclosure to switch data traffic (moving up the IRF hierarchy)
    - switch to consolidating switch (moving up the IRF hierarchy).

The mid-plane fabric switches (together with the Middle-of-Rack/Top-of-Rack [MoR/ToR] switches) work together to deliver a single view of the servers and adaptively ensure that a route exists from any server to every other server, as though they were all parts of a single enclosure. To the outside network, the two client-facing connections represent the only system addresses, and are for data loading and query/response for the analyses desired. The effective view of this design is that, if you could see the logic expressed in a diagram, you could think of the external IP Addresses being at the center of a hub, with all the configuration's IP addresses surrounding it, each equally accessible; in fact, the hub needs to be in three segments to represent each of the implicit/explicit networks in use. Note that there is no mention of any switches. That is because the switches present to the external viewer the appearance that there are no switches in use – they are for want of a better term, totally invisible in most senses, except for each having their respective management network IP address. How IRF and similar switching technologies work is clearly beyond the scope of this paper, and different vendors have their own implementations of similar capabilities and logic. See Figure 2.

The customer's data network must further be able to permit iPad™/Android™ tablets and/or smartphones to access these IP addresses, and these same IP addresses must be able to reach the data source(s) used to populate the SAS VA configuration. This conflict on scope of access (public versus private) is the first of the SAS VA ecosystem issues to be tackled from an information governance perspective. From a network governance perspective, do not be surprised if one IP address is granted external access only, while the other IP address is limited to the access of specific servers from which data will be extracted to populate the SAS VA configuration.
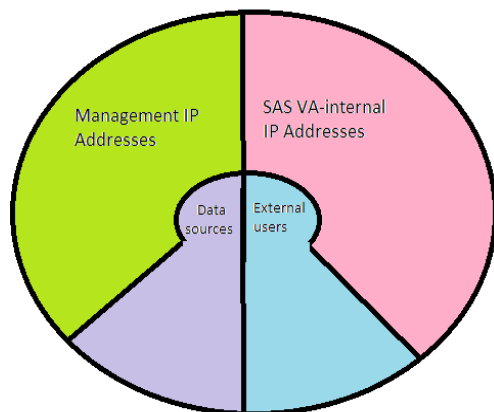
Figure 2. The Three Categories of Networks in SAS VA (Four if Data and External Users separated)

The reason for physical and virtual Management network addresses is due to the design of the BladeSystems. There are typically two Onboard Administrator (or OA) ports with each SAS VA enclosure, and via these ports, each server's own on-board server management facility (integrated Lights Out or iLO) is accessed. So while each server/blade requires its own iLO address, these can be considered virtual addresses and they must exist but do not represent a physical plug/port; they are logically accessed via the OA ports. Likewise, the Virtual Connect Domain addresses are virtual as they must exist to manage the inter-enclosure data traffic, but are not themselves physical ports.

HP BLADESYSTEMS – MOR IRF NETWORKING (FOUR[+] ENCLOSURES; UP TO FOUR RACKS)

In these configurations, there is a Middle-of-Rack (MoR) switch to which each enclosure is wired, rather than wiring each enclosure to the other enclosures in that rack. These switches are known as IRF East-West switches, and for redundancy sake, each enclosure is attached to the MoR switch in their physical rack and in the logical rack to its East. By analogy, using continental US time zones and always assuming that you move to the East, if you are in the east coast rack (think New York City and Eastern time), and wish to connect to its East switch, the switching wraps to the West coast (think San Francisco and Pacific time), and that becomes the east coast's East switch. Up to four racks can be connected in this fashion using IRF East-West. Network connections for the client-facing and Management network networks are all made to these MoR switches, and by design, from ANY East-West MoR switch, you can see every port on every server, as though the network had been collapsed to that one point. These redundant connections guarantee that any point failure has no impact on the whole configuration, even if an entire switch should fail. This is the reason for the redundancy in mapping these ports and switches, so if a component should fail, the probability that the user would know that anything has happened is vanishingly small.

HP BLADESYSTEMS – MOR/TOR IRF NETWORKING (FIVE-EIGHT RACKS)

These configurations are effectively the same as the four or fewer racks configurations, but there is a Top-of-Rack (ToR) switch added for each group of four racks with their existing MoR switches. Each ToR switch is known as an IRF North switch and is used to integrate the connections of the MoR switches "to its South" and to the client-facing network. To preserve the *n + 1* redundancy, this means that two IRF North switches are required in the ToR position, and each of the up to eight MoR switches are wired to both the IRF North switches to maintain a redundant connection paths with the client-facing and Management networks. Again, regardless of the number of switches in any particular configuration, the appearance is that the first switch seen (one of the IRF North switches) is connected to every port of every server in the configuration. This persists even if an intervening path should fail; with the redundancy levels designed into the SAS VA configurations, the switches will talk amongst themselves and negotiate the most efficient path to the sought network port.

ONE SWITCH TO RULE THEM ALL

Since a switch with 100s to 1000s of ports is required to achieve this consolidation of network traffic, list price can start at about US$500,000 and be into the millions of dollars. Note that switch failures are comparatively rare, so one enterprise class switch does not compromise overall redundancy; at worst, a bank of ports could be shut down due to a port or component failure. Cabling is also the most expensive of the four plans, as every server needs to be connected to a single switch which means that the cable lengths required are the maximum possible lengths, compared to the MoR and ToR switching used in the above scenarios. This is because each cable will need to stretch across up to eight racks and from the bottom to the top of those racks; the cabling plan presented in the first three scenarios (with assistance from a couple of data center network engineers) for SAS VA configurations is to

deliver a minimalist topology, yet yield maximum preservation of network redundancy. That's a fancy way of saying that you are protected from small problems and issues related to network equipment and cables.

Regardless of which approach is used, a hierarchy of data switch for N+1 redundancy of network connectivity or one large switch to consolidate all traffic in one place, for some customers, one approach will be preferred and for another customer, corporate standards will dictate the options and choices that are available. Consider these as a starting position when discussing the data network ecosystem with your network engineer(s). From this discussion, the ecosystem in which SAS VA is being planted can be optimally tuned to meet the competing organizational objectives.

HOW SWITCHES AND SERVERS ARE CONNECTED – CABLING AND CONNECTORS

In recent years, there has been a shift from the multiple connector standards, so that there are typically two types of network cable connectors: RJ45 and SFP+. RJ45 connections are familiar to every desktop user whose computer is wired to the corporate (or home) network. For RJ45 use, you can have inexpensive CAT 1-6a types of cabling (often simply referred to as UTP for Unshielded Twisted Pair or Ethernet cables – neither term is wholly accurate) where cables that can handle greater bandwidths had myriad different "standard" connectors. The SFP+ (Small Form factor Package enhanced) connectors have proven to be the best approach for high bandwidth and data reliability, although there is a perceived price increase in their use. The reality is that RJ45 connections need T-class connectors, and in most switches, this consumes about 15 watts of power *per connection.* SFP+ by comparison handles Ethernet on glass (single-mode or multimode), active cables (optical and electrical) and passive copper cable, typically in the sub-1 watt range, per connection.

As you can quickly appreciate, the direct power savings alone quickly lead one to the SFP+ connectors, especially when many switches limit the number of T-class connectors that may be used due to the heat load that each connector generates. Using solely heat considerations, you can handle 10-20 SFP+ connectors for the same electricity and heat cost as a SINGLE RJ45 connection. Further, the processing complexity associated with RJ45 connections to perform the needed bit error rate detection/correction, to separate the data from the "noise", is a limiting factor on how much bandwidth can be driven through such a connection; the technical term is latency. It is for these reasons that RJ45 connections are typically capped at 1 Gb/sec, where SFP+ allows in excess of 40 Gb/sec today. Some of the newer cabling options are equipped with SFP+ connectors on both ends, regardless of whether glass fiber or copper cable forms the medium of the physical "cable". This is why SPF+ has quickly become the universal standard within data center environments. The SFP+ approach allows for [Ultra] Short Range (USR/SR – limited to about 100-300 meters/1500-4000 yards) through Extended or Long Range (ER/LR – good for up to 40 km / 25 miles). Since 2 meters is the "run" from top to bottom of a rack, either USR or SR SFP+ cabling gives plenty of possible length to have a cable handle the "run" from bottom of one rack to the top of a rack that is eight racks away.

As for the perceived cost imbalance, depending on the source of these cables, Ethernet cables can run US$ 10 per cable, where SFP+ cables of a similar length could be about US$ 100-120. However, when you consider latency and bandwidth, a single SFP+ cable with traffic at 10 Gb/sec can handle the work of 10+ Ethernet cables each at 1 Gb/sec, with lower latency and high data accuracy. As the 40 Gb/sec Ethernet becomes more common, expect even better price comparisons between RJ45 and SFP+ / QSFP+ (QSFP+ is the current moniker for 40 Gb/sec SFP+ connectors and cables).

NETWORKING AND CABLING CONCLUSION

Understand that while this all may sound complex, assistance from vendors during the hardware ordering process will make this much clearer and simpler to comprehend than it first appears. As well, your corporate IT Network team should be well versed in these types of discussions and concepts, and their assistance will be necessary to make the addition of SAS VA to your information ecosystem a success. As always, you, your network engineer(s) and your vendors/hosting service will need to work through these scenarios to match the needed networking plan with your needs and requirements. Based on shared experiences, the organizations that plan for the introduction of SAS VA during or before ordering the equipment find that the experience is comparatively painless and the time from the order being placed until the users are working with SAS VA can be weeks to months shorter. It is only when the needed networking arrangements have not been made by the customer do SAS VA implementations experience notable delays.

SOME OF THE TYPICAL QUESTIONS ON IMPLEMENTING SAS VA FOR THE FIRST TIME

The following questions are samples of those asked during the pre-sales and implementation phases with HP by customers. Wording has been retained as originally written, but some notes have been added to clarify the context in which the question was posed. As you can appreciate, these questions also shaped this paper, by indicating points where customers were expressing uncertainty. Through these discussions, and now this paper, the hope is that these questions largely can be retired, to be replaced by new questions as they arise in future.

- Which entity to monitor (File System, CPU, Memory, IOPS)?
  - There are a couple of issues here. First, IOPS is generally an unsatisfactory measure to use with Foundation SAS, due to the largely serial-sequential nature of its I/O use. The bigger issue is that disk and file system usage should tend to be irrelevant as the work is performed in-memory by the CPU. So CPU usage and memory usage are about the only two indicators of system health with an assurance of relevance. While you rarely should need to know details on network capacity and loading, if you will be implementing SAS VA on organization networks that are already over-subscribed, this information may additionally prove to be invaluable. You will probably want to know the network traffic footprint for SAS VA in your organization in such a case.
- Are node names and IP addresses fixed or can they be modified?
  - So long as host node names conform to DNS naming standards (for example, no hyphens, underscores or other punctuation), the name is a networking requirement for the RHEL services, and is not a SAS requirement. IP addresses are likewise not fixed, so they can be whatever meets your operational needs and environment.
- Is only IPv4 supported [by SAS VA6.1 and earlier], so for IPv6 data centers, does SAS VA need to be accommodated solely through IPv6 – IPv4 remapping?
  - This is largely as SAS issue and should be directed to your SAS Representative. At this time, however, there is no known reason why a purely IPv6 address space could not be used with SAS VA. However, testing is needed to verify that assertion, and SAS will need to speak to that issue and their testing schedules, plans and results. This issue is most frequently raised in Europe where the available pool of IPv4 network addresses has [effectively] been exhausted, depending on the level of computerization in any particular country. In many cases, IPv6 needs are raised by Europeans and hosting services, with both groups needing more IP addresses in what is a very small (comparatively) pool of possible addresses defined under IPv4.
- How do we handle the regulated industries requirement for "permanent" retention (seven or more years) and the ability to reproduce the data used for any management query at any point in time?
  - The approach used with most data warehousing and analytical suites is to use a data staging area to prepare the data, and only when identified as complete, current and correct, is the data loaded into that system for use. There is no reason that the SAS VA Administrator cannot likewise import these data sources into your SASVA configuration. Since this is a well-established and vetted approach for use with data warehousing and analytical suites, and which further has the favor of the Audit Committee (and the Compliance VP in regulated industries) of a number of organizations (probably including your own organization), this approach has even further advantages. It means that you have a standard set of data your organization already knows and understands, it can be easily backed up and even distributed to other data centers as part of your disaster recovery plans. Using this investment in infrastructure already made in your organization should make faster and smoother the implementation of SAS VA. This approach has yet another bonus awaiting you, as you are now not connecting the public interface of SAS VA to corporate databases, but rather to a "static" and vetted collection of data ready for use by SAS VA. This should help reduce your information governance discussions, since so many potential objections can be eliminated by the use of the "tried and true" approaches already in use within your organization, but now extended to its new SAS VA environment.

CONCLUSION

As is always the case, check with your SAS Representative on software questions and issues; they are a valuable resource for you and your organization. While HP and other vendors can speak to how the software application behaves with their hardware, realize that our strength is in understanding how our technologies work and how to help you meet your needs. And never forget your corporate IT team who are there to assist you in making the implementation of SAS VA a memorable experience. Using existing organizational approaches and approved practices can speed your implementation of SAS VA, and with a little ingenuity, this should assist you in gaining support for "the use of accepted practices and standards" in support of your SAS VA implementation. And by planning the network environment before the configured equipment arrives, the time from equipment arrival to full use of SAS VA can be reduced by potentially weeks to months.

ACKNOWLEDGMENTS

CONTACT INFORMATION

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

Name:	Nicholson (Nick) Warman, MBA, PMP, I.S.P., ITCP
Enterprise:	Hewlett-Packard
Address:	Room S0056, Building "S"
	600 SAS Campus Drive
City, State ZIP:	Cary, NC 27513
Work Phone:	+1.919.531.5724
	+1.404.648.9222
E-mail:	nick.warman@hp.com, nick.warman@sas.com