Delivering High Availability Services Using a Multi-Tiered Support Model

Phillip J. Windley, Ph.D.
Chief Information Officer

Delivering high availability (HA) services in a networked environment requires more than buying the right hardware and software. HA service delivery is predicated upon having the correct platform (hardware and software), people, processes, and organization. This document describes a model based on a hierarchical, or multi-tied support (MTS) model for organizing to deliver high availability services.

The motto of the MTS model might well be "its not how many mistakes you make, but how *fast* you respond to them." The MTS model creates a very fast response vehicle for all problems that may affect service delivery by explicitly identifying each tier's function and defining the demarcation between tiers.

There is a temptation to believe that HA service delivery can be done solely by having better technology. We have found that it is often the case that technical solutions for HA service delivery complicate the system to such an extent that HA service delivery is impossible in the face of even minor technological problems.

Throughout this document, the reader should remember the basic premise that HA services are not so much about technology, but about the organization, people, and processes that surrounds that technology. Erik Brynjolfsson, a noted MIT management professor, performed a study which indicated that "on average $10 of organizational capital [are] associated with every $1 of technology capital.[1]" This paper is about how to build that organizational capital to deliver HA services.

This document serves as an overview for the MTS model. As such, the goal of this document is to describe the organization and processes necessary to deliver HA services and the relationship between them. Other documents provide detailed information about many of the organizations and processes described here.

---

1 Kindley, Mark, "Hidden Assets," CIO Insight, October 2001, Number 6, pg 26.

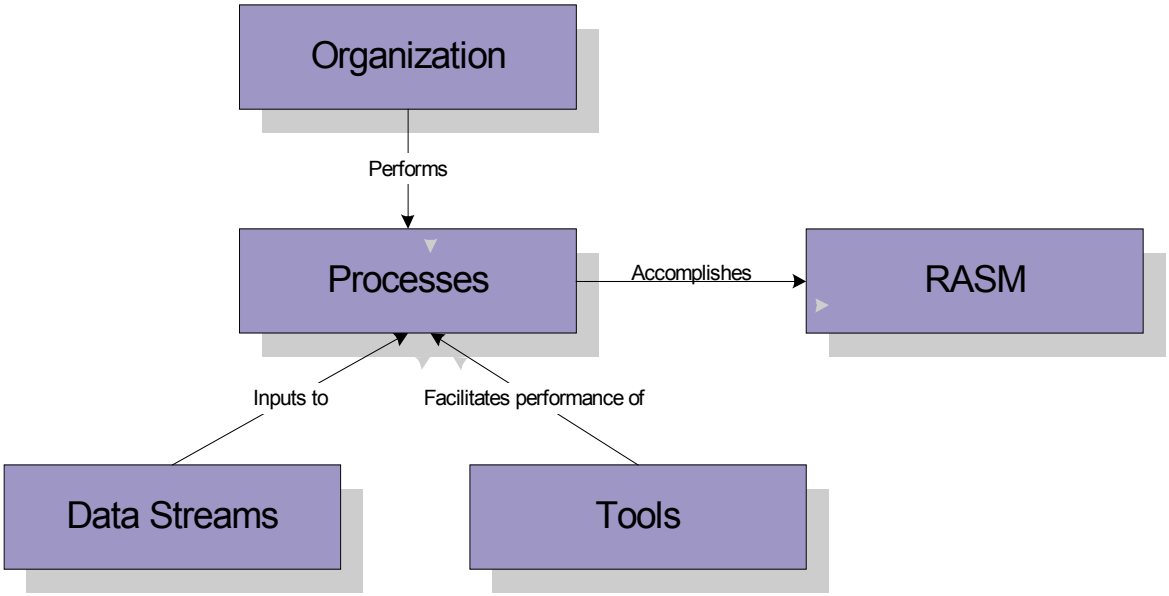

Illustration 1: Factors for RASM

High Availability

In this document, high availability service is meant to include the following attributes:

- Reliability—Reliable services perform their functions consistently. Overall system performance is one component of reliability.
- Availability—Service availability can be defined in terms of the percentage of time that the service is ready for use.
- Servicability—This attribute applies to services that are easy to maintain, troubleshoot, and upgrade.

Harris Kern[2] calls these three attributes RAS and refers to them as a group. To this list of attributes, I would add a fourth:

- Managable—managable services can be easily changed and updated to meet new business requirements.

These attributes are, to some extent, interdependent. In the manner of Kern, we will refer to them as RASM.

Ultimately, these attributes are the function of the effectiveness of processes used to achieve them. Processes, in turn, are carried out by an organization using some set of tools and various streams of data. The relationship is depicted

2 Kern, Harris, et. al., *IT Organization*, Prentice Hall PTR, 2000

in Illustration 1.

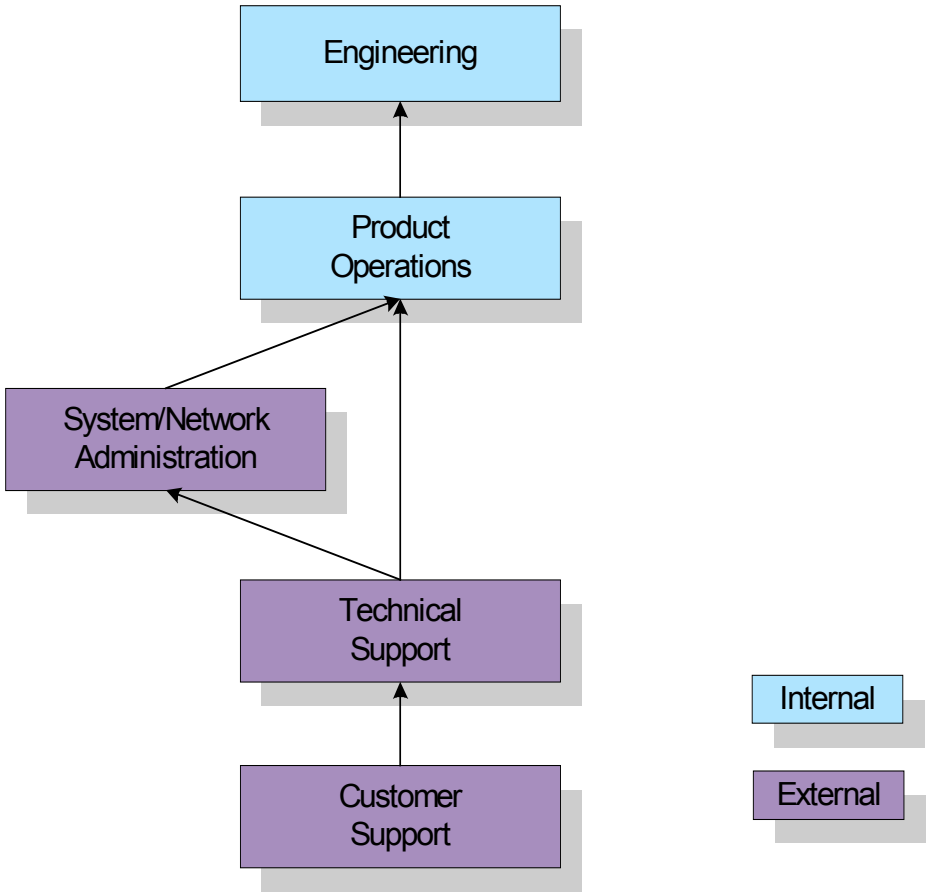

Illustration 2: Tiered Support Model

Tiered Support

Tiered support provides a model for providing high-availability service products in a networked environment. In the tiered support model, organizational functions are placed in an *escalation hierarchy* with clear lines of demarcation between functions to ensure that responsibilities and duties are clearly defined. A visual representation of the tiers and their escalation hierarchy are shown in Illustration 2.

The following descriptions show the five levels in the tiered support model, describe the escalation strategy for each level, and give a short description of the skills, titles, and job descriptions for people in each role. In general, as we move up the escalation hierarchy, the skills and knowledge required to do the job successfully become more specialized.

Tier 1: Customer support/Help desk *(Customer Support Representative)*. This

group is the first line of support and is generally receiving customers calls first. They perform triage, solving straightforward problems themselves (possibly using some kind of knowledge management tool) and pass the rest onto Tier 2. In some rare cases, they may escalate directly to Tier 3. Workers in this group are marginally skilled with little technical training. This support level is often outsourced because the skills necessary to do the job effectively are not as directly tied to the specific products and services being offered as they are for the other tiers.

Tier 2: Technical support *(Technical Support Engineer)*. This group handles escalations from Tier 1 and escalates to either Tier 3 or 4. TSEs try to find the solution to a customer's problem either on their own or using other resources at their disposal. For example, they typically have direct lines into Tiers 3, 4 and 5 and a technical understanding of the products. Technical support personnel have beginning level technical skills and often move to other technical positions in tiers 3, 4, or 5 as their skills develop.

Tier 3: System/Network Administration *(System Administrator, Network Administrator)*. System administration in Tier 3 handles escalations from Tier 2 and escalates to Tier 4. Tier 3 personnel focus on the infrastructure that HA services rely on to function properly. The NOC (see discussion below) is the primary tool that network and server engineers use to monitor system performance.

Tier 4: Product Operations Engineering *(Product Operations Engineer)*. Product operations engineers handle escalations from Tier 2 and Tier 3 and escalate to Tier 5. The product operations group is made up of seasoned, technically skilled engineers who focus on the *operation* of individual products. Usually, their functional specialty is systems engineering and they are probably exeperienced employees who have worked for a time in Tier 3. They may have other functional specialties, such as database administration or network engineering, depending on the product they support.

The focus of a ProdOps engineer is different from Tier 3 where the focus is on infrastructure; the product operations engineer is directly responsible for the availibility of their product. The ProdOps Engineers work closely with Tier 2 and 3 personnel to ensure they are well trained and can answer as many questions and solve as many problems as possible. The ProdOps engineers work closely with Tier 5 to ensure that products are designed and built so that they can be monitored effectively and that there are workable strategies for scaling and maintaining the product. Product reliability and availability is their primary concern. ProdOps engineers monitor and control rollout of new production code coordinating the efforts of the NOC, Software Quality Assurance, and Engineering.

Tier 5: Engineering *(Software or Systems Engineer)*. Engineering is in charge building and maintaining products. They handle escalations from Tiers 2 and 4. Engineering is in charge of the bug tracking process and uses it as a means for monitoring the health of the products they have built and are maintaining. They do not set priorities for bug fixes, however. That is done by a group with representatives from Product Management, Technical Support, Software Quality Assurance, Product Operations, and Engineering (see the section on Teams below).

Supporting Systems

The organization described above relies on several systems to function. These systems are described in this section.

Data Center. A properly designed, maintained, and operated data center is a crucial piece of infrastructure for delivering HA services. The data center should provide, at a minimum, reliable back-up power, a controlled environment (temperature, humidity, etc.), physical security, and redundant network connections with sufficient bandwidth for planned operations.

Network/Server Operations Center. The NOC (or SOC) is a system that monitors the health of the network and servers. The NOC is typical co-located with the data center, although this is not strictly necessary. For HA service delivery, the NOC is manned by personnel from Tier 3 on a 24x7 basis. The job of the NOC personnel is to ensure that the network and servers continue to run as well as to identify problems and issues that may affect uptime. Their focus is on systems issues. Strict standards about the NOC, its operation, and how it does monitoring are necessary for it to be effective and for people to build products that can be effectively monitored.

The NOC is described as a system here since it serves as a conduit through which many of the interactions between tiers will happen. The NOC will typically employ numerous other systems in performing its function. For example, the NOC may use SNMP monitoring tools to monitor the health of the network as well as hardware and software systems.

Trouble Ticket System. The NOC tracks issues and problems using a trouble ticketing system that is visible by all tiers. Trouble tickets differ from bugs in that they are usually operational in nature and more urgent. Trouble ticket items may be associated with a bug report when Tiers 4 or 5 are involved in the solution. The trouble ticket system must be integrated with the CRM system so that trouble tickets associated with customers can be linked back to them. This ensures that there is a history of a customer's interactions with Tiers 1 and 2 and provides a complete record of issues surrounding that customer.

Customer Relationship Management (CRM) System. CRM systems provide a place for tracking customer interactions with Tiers 1 and 2 and for tracking

responses to problems by Tiers 3, 4 and 5. A critical part of the CRM system is a trouble ticket module where individual customer issues can be tracked as discussed above.

The CRM system will also be used by sales, the CRM group (described below) and marketing. Each customer record in the CRM system will describe the complete history of that customer's interactions with the company. The CRM system serves as a central communication point between all of the parts of the organization who have interest or interaction with customers.

Bug Tracking System. In the same way that the CRM system serves as a point of communication for the customer facing parts of the organization, a bug tracking system serves as the central communication point among groups who must solve technical problems.

Bugs are tracked separately from trouble tickets since multiple trouble tickets may lead back to a single bug. Trouble tickets and bugs should be able to be linked for reporting and ticket clearing purposes. Each bug has an associated severity and priority. The severity is an objective *measure* of the impact that the bug has on HA service delivery and the proper function of the product. Severity is set, according to a pre-determined scale, by the person reporting the bug. Priority is a subjective *judgment* of the order in which resources should be applied to fixing problems.

Other Organizational Members

In addition to organizations that are part of the support hierarchy, there are a number of other organizations that have an important role in delivering high-availability service products.

Product Management: Product management manages the product development process and controls the nature of services offered, including the user interface, customer care, and even financials and marketing. Product management owns the product and drives the specification process. A product manager chairs the product team for each product or service offered and coordinates the work of all team members to achieve product goals.

Customer Relationship Management: The CRM organization (sometimes called “large accounts”) serves as the customer's advocate inside the organization. Each large customer is assigned to a CRM and that CRM looks out for the customer's interests in the organization. CR managers work closely with Tiers 4 and 5 to resolve client issues and should be tied in to the CRM system in a way that alerts them to high priority problems surrounding their customers and accounts.

Software Quality Assurance: SQA is not part of the support escalation path and thus does not have a place in the support hierarchy, but they play an

important role in product support. SQA sits between engineering and the production servers to ensure that code is properly tested, according to the pre-published validation plan, before it is released to production. SQA works with the product team to create validation plans for each product release. They manage the QA environment. In short, SQA is responsible for the production acceptance process and systems that support it.

Release Management: Release management is charged with maintaining the integrity of the production environment through careful control of the release process. Release management establishes policies that control the release process and chairs the release teams.

Teams

Each of the organizations described above participate on one or more teams that coordinate the activities of the organization for HA service delivery. Table 1 gives the membership of each of the organizations mentioned above for the various teams. The teams and their purposes are described in the paragraphs that follow.

	Product Team	***Release Team***	***Bug Team***	***Deployment Team***
Product Management	◊	●	●	
Customer Relationship Management	●	●	●	
Customer Support	●			
Technical Support Engineering	●		●	
Systems Administration		●		●
Product Operations Engineering	●	●	●	◊
Release Management	●	◊	●	
Software Quality Assurance	●	●	●	
Engineering	●	●	◊	
Finance	●			
Marketing	●			
Sales	●			

Table 1: Team Membership. Team chairs are denoted by ◊

Product Team: The product team is made up of representatives from every organization mentioned in this document except the NOC along with some organizations not discussed in this document (such as finance, marketing, and sales). Certainly, the NOC could have a representative on the team without harm, but the product operations group should be able to represent the NOC's interests. The Product Team is charged with developing the plan to build, deploy, operate, and maintain the product throughout its lifetime. There is one

product team for each product or set of related products. The product manager who is responsible for a given product chairs the product team for that product.

Release Team: The release team is charged with managing releases from the various QA environments to the production environments. The release team is chaired by a member of the release management organization and is made up of representatives from product management, the NOC, product operations, software quality assurance, and engineering. The release team signs off on the readiness of a product for release and planning and coordinates the actions required to deploy the product.

Bug Team: The bug team tracks bugs that are reported from any level or source and sets the priority of the bug so that resources can be assigned to rectify the bug. The bug team meets regularly to discuss the contents of the bug tracking system and assign priority for fixing bugs. Ultimately, the product manager for a particular product has final say in assigning the bug priority for that product. A representative from the engineering organization chairs the bug team.

Deployment Team: The deployment team must not be confused with the release team. Where the release team is concerned with product releases, the deployment team is concerned with deploying gear and other infrastructure components as part of the system and product development process. In a small operation, the team is staffed as needed with personnel from Tiers 3 and 4. In a large operation that is continuously deploying gear, this team may be staffed with dedicated engineers. The deployment team is typically working at the direction of the product operations engineer in charge of the product for which the deployment is being done. The design for the physical system was likely done by the product operations engineer working closely with engineering and drawing on technical experts in the engineering organization.

Performance Measurement

Central to the discussion of delivering high availability services is the assumption that we know what high availability is and that we measure and track the organization's progress and performance. A system for measuring the most critical aspects of HA service delivery is critical.

There are various names for the tools used to measure performance. Among the most popular are ‘‘dashboard’’ or ‘‘scorecard.’’ Many companies employ balanced scorecards that include financial and other information relative to determining overall performance of the firm. Dashboards and scorecards represent the summary information presented to management. To create such a summary, however, large numbers of measurements have to be taken, tracked, and compiled.

Not all of the measurements made are directed solely at evaluating HA service

delivery performance; many are useful for critical processes such as capacity planning. If possible, the dashboard should be aggregated from signals coming from instrumented processes so that the data is continuously fresh.

Once the right set of measurements are being made, the next step is to find ways to combine, contrast, normalize, and report these statistics so that they give a meaningful picture of the overall state of the system and trend lines for determining the progress that the organization is making toward delivering HA services.

Conclusion

Delivering HA services involves more than just buying the right hardware and getting the architecture right. In fact, in my experience, this can often make things worse, not better. The trick to deliving HA services is having the right people, in the right organization, with the right tools, doing the right things. In future papers, we'll discuss other organizations involved in product development, performance measures, and processes for delivering HA services in more detail.

Acknowledgments

The author acknowledges the help of Steve Fulling, Barry Dixon, Daniel Bray, Kristen Knight, Stefanie Rubin, Wade Billings, Leanne Isaac, and Chris Heim in the development of the model presented here.