

服务级别管理：最佳实践白皮书

目录

[简介](#)

[服务级别管理概述](#)

[关键成功因素](#)

[绩效指标](#)

[服务级别管理流程](#)

[实施服务级别管理](#)

[定义网络服务级别](#)

[创建和维护SLA](#)

[服务级别管理绩效指标](#)

[书面服务级别协议或服务级别定义](#)

[绩效指标指标](#)

[服务级别管理审核](#)

[服务级别管理摘要](#)

[相关信息](#)

简介

本文档介绍高可用性网络的服务级别管理和服务级别协议(SLA)。它包括服务级别管理和绩效指标的关键成功因素，以帮助评估成功。本文档还为遵循高可用性服务团队确定的最佳实践指南的SLA提供了重要的详细信息。

服务级别管理概述

过去，网络组织通过构建坚实的网络基础设施并积极处理个别服务问题，来满足不断扩展的网络需求。发生故障时，组织将构建新的流程、管理功能或基础设施，以防止再次发生特定故障。但是，由于更改率更高且可用性要求不断提高，我们现在需要改进的模型来主动预防意外停机并快速修复网络。许多服务提供商和企业组织都试图更好地定义实现业务目标所需的服务级别。

关键成功因素

SLA的关键成功因素用于定义成功构建可获得的服务级别和维护SLA的关键要素。要获得SLA的认可，流程或流程步骤必须提高SLA的质量，并从总体上提高网络可用性。关键的成功因素也应可衡量，以便组织能够确定它相对于定义的程序的成功程度。

有关详细信息，请参阅[实施服务级别管理](#)。

绩效指标

绩效指标为组织衡量关键成功因素提供了机制。您通常每月查看这些内容，以确保服务级别定义或

SLA工作正常。网络操作组和必要的工具组可以执行以下指标。

注意：对于没有SLA的组织，我们建议您执行服务级别定义和服务级别审核以及指标。

绩效指标包括：

- 记录的服务级别定义或SLA，包括可用性、性能、被动服务响应时间、问题解决目标和问题上报。
- 每月举行网络服务级别审查会议，以审查服务级别合规性并实施改进。
- 绩效指标指标，包括可用性、性能、按优先级的服务响应时间、按优先级解析的时间以及其他可衡量的SLA参数。

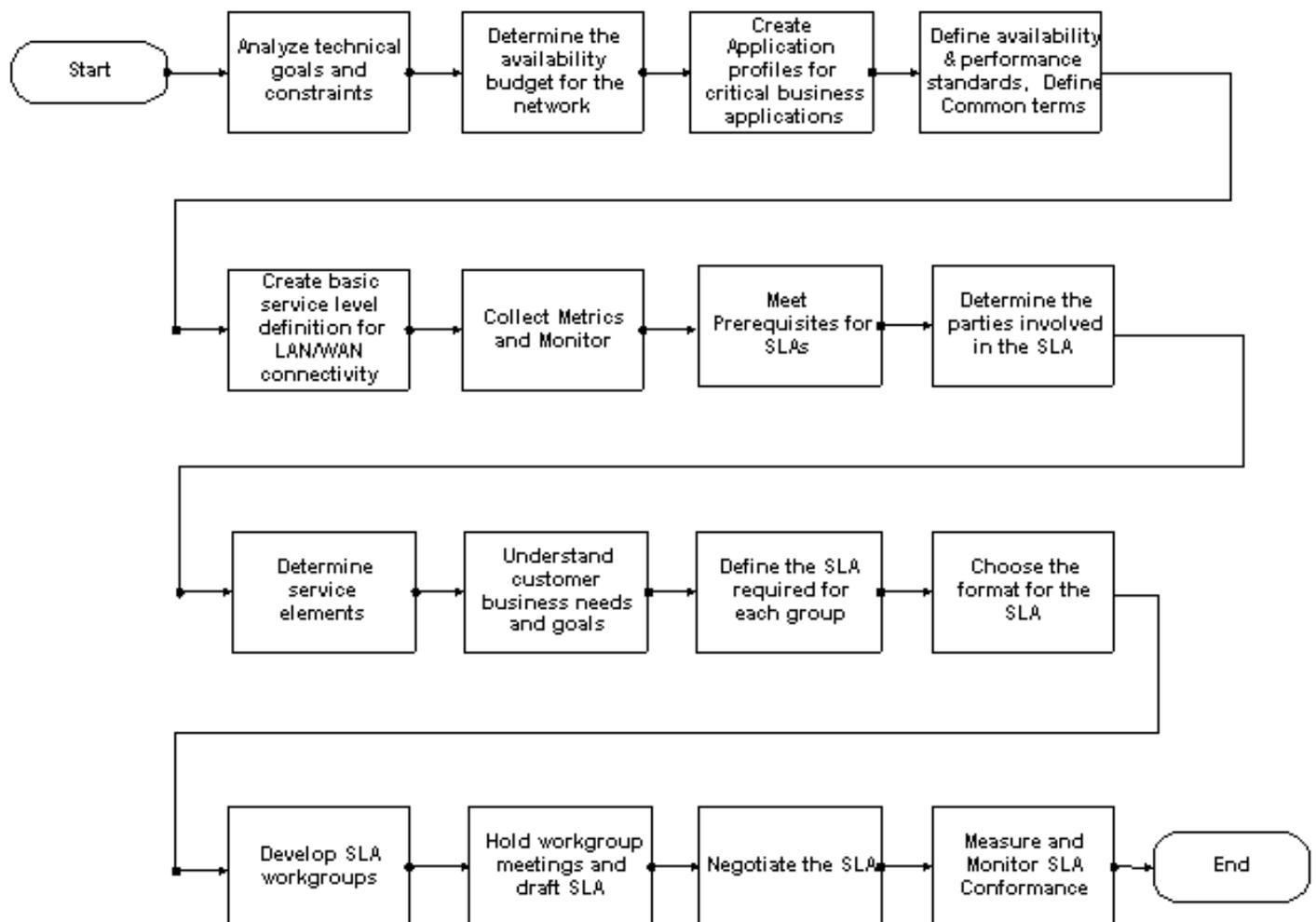
有关详细信息，请参阅[实施服务级别管理](#)。

服务级别管理流程

服务级别管理的高级流程包含两个主要组：

1. [定义网络服务级别](#)
2. [创建和维护SLA](#)

单击下图中的对象，查看该步骤的详细信息。



实施服务级别管理

实施服务级别管理包括以下两个主要类别的16个步骤：

- [定义网络服务级别](#)
- [创建和维护SLA](#)

[定义网络服务级别](#)

网络管理员需要定义支持、管理和衡量网络的主要规则。服务级别为所有网络人员提供目标，并可用作衡量整体服务质量的指标。您还可以将服务级别定义用作网络资源预算的工具，并作为需要为更高QoS提供资金的证据。它们还提供了一种评估供应商和运营商绩效的方法。

如果没有服务级别定义和衡量标准，组织就没有明确的目标。服务满意度可能由应用、服务器/客户端操作或网络支持之间差别很小的用户决定。预算可能会更加困难，因为组织不清楚最终结果，最后，网络组织在改进网络和支持模式时往往更为被动，而不是主动。

我们建议采用以下步骤来构建和支持服务级别模型：

1. [分析技术目标和限制条件。](#)
2. [确定可用性预算。](#)
3. [创建应用配置文件，详细描述关键应用的网络特征。](#)
4. [定义可用性和性能标准并定义通用术语。](#)
5. [创建服务级别定义，包括可用性、性能、服务响应时间、解决问题的平均时间、故障检测、升级阈值和升级路径。](#)
6. [收集指标并监控服务级别定义。](#)

[步骤 1：分析技术目标和限制条件](#)

开始分析技术目标和限制条件的最佳方法是集体讨论或研究技术目标和要求。有时，邀请其他IT技术人员参加此讨论会有所帮助，因为这些人员具有与其服务相关的特定目标。技术目标包括可用性级别、吞吐量、抖动、延迟、响应时间、可扩展性要求、新功能介绍、新应用引入、安全性、可管理性甚至成本。然后，组织应调查在可用资源下实现这些目标的限制条件。您可以为每个目标创建工作表，并说明限制条件。最初，似乎大部分目标都无法实现。然后开始确定目标的优先顺序，或降低仍能满足业务需求的预期。

例如，您的可用性级别可能为99.999%，即每年停机5分钟。实现此目标存在许多限制条件，例如硬件中的单点故障、远程位置中的平均修复时间(MTTR)故障硬件、运营商可靠性、主动故障检测功能、高更改率和当前网络容量限制。因此，您可以将目标调整到更易实现的水平。下一节中的[可用性模型](#)可以帮助您设定现实目标。

您可能还考虑在限制较少的某些网络区域提供更高的可用性。当网络组织发布可用性服务标准时，组织内的业务组可能会发现该级别无法接受。因此，这是开始SLA讨论或资金/预算模型的自然点，这些模型可以满足业务需求。

确定实现技术目标所涉及的所有限制或风险。根据最大风险或对预期目标的影响确定限制的优先顺序。这有助于组织确定网络改进计划的优先顺序，并确定解决限制的难易程度。约束有三种：

- 网络技术、恢复能力和配置
- 生命周期实践，包括规划、设计、实施和运营
- 当前流量负载或应用行为

网络技术、恢复能力和配置限制是与当前技术、硬件、链路、设计或配置相关的任何限制或风险。

技术限制涵盖技术本身所构成的任何限制。例如，当前的技术不允许在冗余网络环境中使用次秒级的收敛时间，这对于在整个网络中维持语音连接至关重要。另一个示例可能是数据在地面链路上传输的原始速度，大约为每毫秒100英里。

网络硬件恢复能力风险调查应集中于硬件拓扑、层次结构、模块性、冗余和MTBF（沿网络中定义的路径）。网络链路限制应侧重于企业网络链路和运营商连接。链路限制可能包括链路冗余和多样性、介质限制、布线基础设施、本地环路连接和长距离连接。设计限制与网络的物理或逻辑设计有关，包括从设备可用空间到路由协议实施的可扩展性的所有内容。所有协议和介质设计都应考虑配置、可用性、可扩展性、性能和容量。还应考虑网络服务限制，如动态主机配置协议(DHCP)、域名系统(DNS)、防火墙、协议转换器和网络地址转换器。

生命周期实践定义了用于一致部署解决方案、检测和修复问题、防止容量或性能问题以及配置网络以实现一致性和模块化的网络的流程和管理。您需要考虑这一领域，因为专业知识和流程通常是导致无法使用的最大因素。网络生命周期是指规划、设计、实施和运营的周期。在这些区域中，您必须了解网络管理功能，如性能管理、配置管理、故障管理和安全。思科NSA高可用性服务(HAS)服务提供网络生命周期评估，显示与网络生命周期实践相关的当前网络可用性限制。

当前流量负载或应用限制仅指当前流量和应用的影响。

遗憾的是，许多应用程序存在严重的限制，需要谨慎管理。当前应用的抖动、延迟、吞吐量和带宽要求通常有许多限制。应用程序的编写方式也可能造成限制。应用程序分析有助于您更好地了解这些问题；[下一部分](#)介绍此功能。调查当前的可用性、流量、容量和整体性能还有助于网络经理了解当前的服务级别期望和风险。这通常通过网络基线建立过程来实现，该过程有助于定义所定义时间段内的网络性能、可用性或容量平均值，通常为一个月。此信息通常用于容量规划和趋势分析，但也可用于了解服务级别问题。

以下工作表使用上述目标/约束方法作为防止安全攻击或拒绝服务(DoS)攻击的示例目标。您还可以使用此工作表帮助确定服务范围，以最大限度地减少安全攻击。

风险或限制	约束类型	潜在影响
可用的DoS检测工具无法检测所有类型的DoS攻击。	技术/恢复能力	高
没有对警报做出响应所需的员工和流程。	生命周期实践	高
当前网络访问策略不到位。	生命周期实践	中
当前带宽较低的Internet连接可能是带宽拥塞用于攻击的一个因素。	网络容量	中
当前用于防止攻击的安全配置可能不彻底。	技术/恢复能力	中

步骤 2：确定可用性预算

可用性预算是网络在两个定义点之间的预期理论可用性。准确的理论信息在以下几方面非常有用：

- 组织可以将此用作内部可用性的目标，并且可以快速定义和修复差价。
- 网络规划人员可以使用这些信息来确定系统的可用性，以帮助确保设计符合业务要求。

导致不可用或中断时间的因素包括硬件故障、软件故障、电源和环境问题、链路或运营商故障、网络设计、人为错误或流程缺乏。在评估网络的整体可用性预算时，应仔细评估这些参数。

如果组织当前衡量可用性，则可能不需要可用性预算。使用可用性度量作为基线来估计用于服务级别定义的当前服务级别。但是，您可能有兴趣将二者进行比较，以了解潜在的理论可用性与实际测量结果相比较。

可用性是指产品或服务在需要时运行的可能性。请参阅以下定义：

1. **可用性1** — $(\text{总连接中断时间}) / (\text{总服务中连接时间}) - [\text{Sigma}(\text{在X个停机持续时间i中断时受影响的数量连接})] / (\text{在X个运行时间中的数量连接})$
2. **不可用1** — 由于(硬件故障、软件故障、环境和电源问题、链路或运营商故障、网络设计或用户错误和流程故障)导致的可用性或总中断连接时间
3. **硬件可用性**要调查的第一个方面是潜在的硬件故障及其对不可用性的影响。要确定这一点，组织需要了解所有网络组件的MTBF和MTTR，以了解两点之间路径中所有设备的硬件问题。如果网络采用模块化和分层结构，则几乎任意两点之间的硬件可用性都相同。MTBF信息适用于所有思科组件，并可按要求提供给本地客户经理。Cisco NSA HAS计划还使用一个工具来帮助确定网络路径上的硬件可用性，即使系统中存在模块冗余、机箱冗余和路径冗余。MTTR是硬件可靠性的一个主要因素。组织应评估修复损坏硬件的速度。如果组织没有备件计划，并且依赖标准的Cisco SMARTnet™协议，则可能的平均更换时间约为24小时。在具有核心冗余和无访问冗余的典型LAN环境中，4小时MTTR的大致可用性为99.99%。
4. **软件可用性**下一个调查领域是软件故障。出于测量目的，思科将软件故障定义为由于软件错误而导致的设备冷启动。思科在了解软件可用性方面已取得重大进展；但是，较新的版本需要时间来衡量，并且被认为比一般部署软件更不可用。IOS版本11.2(18)等一般部署软件的可用性已超过99.9999%。这是根据思科路由器上实际的冷启动计算的，使用6分钟作为修复时间(路由器重新加载的时间)。由于复杂性、互操作性和故障排除时间的增加，具有各种版本的组织的可用性预期会略有降低。具有最新软件版本的组织将具有更高的不可用性。不可用性的分布也相当广泛，这意味着客户可能在接近一般部署版本时体验到显著的非可用性或可用性。
5. **环境和电源可用性**您还必须考虑可用性中的环境和电源问题。环境问题与将设备保持在指定工作温度所需的冷却系统故障有关。许多思科设备在严重不符合规格时将直接关闭，而不会冒损坏所有硬件的风险。出于可用性预算的目的，将使用电源，因为它是此领域不可用的主要原因。虽然电源故障是确定网络可用性的一个重要方面，但由于无法准确进行理论功率分析，此讨论受到限制。组织必须评估的是根据其地理区域经验、电源备份功能和实施的流程来大致衡量其设备的电源可用性，以确保为所有设备提供一致的质量电源。对于保守的评估，我们可以说，拥有备用发电机、不间断电源(UPS)系统和优质电力实施流程的组织可能经历6个9的可用性，即99.9999%，而没有这些系统的组织则可能经历99.99%的可用性，即每年约36分钟的停机时间。当然，您可以根据组织的感知或实际数据将这些值调整为更现实的值。
6. **链路或运营商故障**链路和运营商故障是影响WAN环境可用性的主要因素。请记住，WAN环境只是其他网络，其可用性问题与组织网络相同，包括硬件故障、软件故障、用户错误和电源故障。许多运营商网络已经在其系统上执行了可用性预算，但获取此信息可能非常困难。请记住，运营商通常还有可用性保证级别，这些级别对实际可用性预算几乎无依据。这些保证级别有时只是用于推广运营商的营销和销售方法。在某些情况下，这些网络还会发布看起来非常好的可用性统计信息。请记住，这些统计数据可能仅适用于完全冗余的核心网络，而不会因本地环路访问而影响不可用性，而本地环路访问是WAN网络中不可用性的主要因素。根据实际运营商信息和广域网连接的冗余级别来估算广域网环境的可用性。如果组织有多个建筑入口设施、冗余本地环路提供商、同步光纤网络(SONET)本地接入和具有地理多样性的冗余长距离运营商，WAN可用性将得到显著增强。电话服务是WAN环境中非冗余网络连接的相当准确的可用性预算。电话的端到端连接使用与本节所述类似的可用性预算方法，其可用性预算约为99.94%。此方法已成功应用于仅略有变化的数据环境，目前用作服务提供商有线网络的数据包电缆规范中的目标。如果将此值应用于完全冗余的系统，我们可以假设WAN可用性将接近99.9999%的可用性。当然，由于费用和可用性，很少有组织拥有完全冗余且地理上分散的广

域网系统，因此请对此功能进行适当判断。LAN环境中出现链路故障的可能性较低。但是，规划人员可能希望假设由于连接器损坏或松动而造成少量停机。对于LAN网络，保守估计约为99.9999%的可用性，即每年约30秒。

7. **网络设计**网络设计是提高可用性的另一个主要因素。不可扩展的设计、设计错误和网络收敛时间都对可用性产生负面影响。**注意：**就本文档而言，不可扩展的设计或设计错误包括在以下部分。然后，网络设计会根据导致流量重新路由的网络中的软件和硬件故障限制为可衡量的价值。此值通常称为“系统切换时间”，是系统内自修复协议功能的一个因素。只需使用与系统计算相同的方法即可计算可用性。但是，除非网络切换时间满足网络应用要求，否则此操作无效。如果切换时间可接受，请将其从计算中删除。如果切换时间不可接受，则必须将其添加到计算中。例如，在估计或实际切换时间为30秒的环境中，IP语音(VoIP)可能是一个示例。在本例中，用户只需挂断电话，可能会重试。用户肯定会将这段时间视为不可用，但尚未在可用性预算中进行估计。通过沿冗余路径查看理论软件和硬件可用性来计算由于系统切换时间而导致的不可用性，因为此区域将进行切换。您必须知道冗余路径中可能发生故障并导致切换的设备数量、这些设备的MTBF和切换时间。一个简单的示例是，两个冗余相同设备的MTBF为35,433小时，切换时间为30秒。将35,433除以8766（每年平均小时数以包括闰年），我们发现设备每四年会失败一次。如果我们使用30秒作为切换时间，则可以假设每台设备每年平均会因切换而出现7.5秒的不可用性。由于用户可能正在穿越任一路径，因此结果将翻倍至每年15秒。如果以每年的秒数计算，则由于切换而导致的可用性量可以计算为99.99999785%的可用性。在其他环境中，这可能会更高，因为在可能进行切换的网络中，冗余设备的数量会增加。
8. **用户错误和流程**用户错误和进程可用性问题是导致企业和运营商网络不可用的主要原因。大约80%的不可用性是由于不检测错误、更改故障和性能问题等问题造成的。组织在确定可用性预算时不会希望使用四倍于所有其他理论上的不可用性，但有证据表明，在许多环境中，情况都是如此。下一节将更彻底地介绍这一不可用性方面。由于用户错误和流程导致从理论上无法计算不可用性的数量，因此我们建议您从可用性预算中删除此项，并且组织应努力实现完美。一个警告是，组织需要了解自身流程和专业水平对可用性的当前风险。一旦您更好地了解这些风险和制约因素，网络规划人员可能希望将这些问题导致的某些不可用性因素考虑在内。Cisco NSA HAS计划可以调查这些问题，并帮助组织了解由于流程、用户错误或专业知识问题而导致的潜在不可用性。
9. **确定最终可用性预算**您可以通过将每个先前定义区域的可用性相乘来确定整体可用性预算。这通常用于任何两点之间连接相似的同构环境，例如分层模块化LAN环境或分层标准WAN环境。在本例中，可用性预算是针对分层模块化LAN环境完成的。该环境对所有网络组件使用备用发电机和UPS系统并正确管理电源。组织不使用VoIP，也不希望将软件切换时间考虑在内。估计如下：
两个端点之间的硬件路径可用性= 99.99% 可用性使用GD软件可靠性作为参考的软件可用性= 99.9999% 可用性备份系统的环境和电源可用性= 99.999%的可用性LAN环境中的链路故障= 99.9999% 可用系统切换时间不考虑因素= 100% 可用性用户错误和流程可用性假设完美= 100% 可用性组织应争取的最终可用性预算等于 $0.9999 \times 0.999999 \times 0.999999 \times 0.999999 = 0.999896$ 或99.9896%的可用性。如果我们考虑了由于用户或流程错误而导致的潜在不可用性，并假设由于技术因素导致的不可用性是4倍，我们可以假设可用性预算为99.95%。此示例分析表明，LAN可用性将平均降低99.95%到99.989%。这些数字现在可用作网络组织的服务级别目标。您可以通过测量系统中的可用性并确定由于上述六个区域中的每一个而导致的不可用百分比来获得额外价值。这使组织能够正确评估供应商、运营商、流程和员工。该数字还可用于设定企业内的期望值。如果数量不可接受，则预算额外资源以获得所需的级别。对于网络管理员来说，了解任何特定可用性级别的停机时间可能非常有用。考虑到任何可用性级别，一年内的停机时间为分钟： $\text{一年内停机的分钟数} = 525600 - (\text{可用性级别} \times 5256)$ 如果使用99.95%的可用性级别，则这相当于 $525600 - (99.95 \times 5256)$ 或262.8分钟的停机时间。对于上述可用性定义，这等于网络中所有服务连接的平均停机时间。

步骤 3：创建应用配置文件

应用配置文件可帮助网络组织了解并定义各个应用的网络服务级别要求。这有助于确保网络整体支持各个应用要求和网络服务。当应用或服务器组指向网络作为问题时，应用配置文件也可作为网络服务支持的书面基准。最终，应用配置文件通过将性能和可用性等应用要求与实际网络服务目标或当前限制进行比较，帮助将网络服务目标与应用或业务要求保持一致。这不仅对服务级别管理很重要，而且对于整体自上而下网络设计也很重要。

在您向网络引入新应用时创建应用配置文件。您可能需要IT应用组、服务器管理组和网络之间的协议，以帮助为新服务和现有服务强制创建应用配置文件。完整的业务应用和系统应用应用配置文件。业务应用可能包括电子邮件、文件传输、网络浏览、医疗成像或制造。系统应用可包括软件分发、用户身份验证、网络备份和网络管理。

网络分析师和应用或服务器支持应用应创建应用配置文件。新应用可能需要使用协议分析器和具有延迟仿真的WAN仿真器来正确描述应用要求。这有助于确定必要的带宽、应用可用性的最大延迟和抖动要求。只要您有所需的服务器，就可以在实验环境中完成。在其他情况下（如使用VoIP），网络要求（包括抖动、延迟和带宽）将得到很好的发布，无需进行实验室测试。应用配置文件应包括以下项目：

- 应用程序名称
- 应用类型
- 新应用程序？
- 业务重要性
- 可用性要求
- 使用的协议和端口
- 估计用户带宽(kbps)
- 用户数和位置
- 文件传输要求（包括时间、卷和终端）
- 网络中断影响
- 延迟、抖动和可用性要求

应用配置文件的目的是了解应用的业务需求、业务关键性和网络需求（如带宽、延迟和抖动）。此外，网络组织应了解网络中断的影响。在某些情况下，您需要重新启动应用或服务器，这会显著增加整体应用停机时间。完成应用配置文件后，您可以比较整体网络功能，并帮助将网络服务级别与业务和应用要求保持一致。

步骤 4：定义可用性和性能标准

可用性和性能标准为组织设定了服务期望。这些定义可针对网络的不同区域或特定应用进行定义。性能也可以根据往返延迟、抖动、最大吞吐量、带宽承诺和整体可扩展性来定义。除了设定服务期望值外，组织还应谨慎定义每个服务标准，以便使用网络的用户和IT团队能够充分了解服务标准及其与其应用程序或服务器管理要求的关系。用户和IT组还应了解如何衡量服务标准。

之前服务级别定义步骤的结果将有助于创建标准。此时，网络组织应清楚了解网络中的当前风险和限制条件，了解应用行为，并进行理论可用性分析或可用性基线。

1. 定义将应用服务标准的地理或应用区域。这可能包括园区LAN、国内WAN、外联网或合作伙伴连接等方面。在某些情况下，组织可能在一个区域内实现不同的服务级别目标。这在企业或服务提供商组织中并不罕见。在这些情况下，根据个别服务要求创建不同的服务级别标准并不罕见。这些服务标准可在一个地域或服务区域内分类为金牌、银牌和铜牌服务标准。
2. 定义服务标准参数。可用性和往返延迟是最常见的网络服务标准。最大吞吐量、最低带宽承诺、抖动、可接受错误率和可扩展性功能也可根据需要包括在内。查看测量方法的服务参数时要小心。无论参数是否移至SLA，组织都应考虑在出现问题或服务分歧时如何衡量或证明服务参

数的合理性。

定义服务区域和服务参数后，使用前面步骤中的信息构建服务标准矩阵。组织还需要定义可能令用户和IT组感到困惑的区域。例如，往返ping的最大响应时间与在特定应用的远程位置按Enter键时的最大响应时间非常不同。下表显示了美国内的绩效目标。

网络区域	可用性目标	测量方法	平均网络响应时间目标	接受的最大响应时间	响应时间测量方法
LAN	99.99%	受影响的用户分钟数	5毫秒以下	10 毫秒	往返ping响应
WAN	99.99%	受影响的用户分钟数	100毫秒以下（往返ping）	150 ms	往返ping响应
关键广域网和外联网	99.99%	受影响的用户分钟数	100毫秒以下（往返ping）	150 ms	往返ping响应

步骤 5：定义网络服务

这是基本服务级别管理的最后一步；它定义了您为实现服务级别目标而实施的被动式和主动式流程和网络管理功能。最终文档通常称为运营支持计划。大多数应用支持计划仅包括被动支持要求。在高可用性环境中，组织还必须考虑主动管理流程，这些流程将用于在发起用户服务呼叫之前隔离和解决网络问题。总体而言，最终文档应：

- 描述用于实现服务级别目标的被动式和主动式流程
- 如何管理服务流程
- 如何衡量服务目标和服务流程。

本节包含针对许多服务提供商和企业组织需要考虑的被动服务定义和主动服务定义示例。构建服务级别定义的目标是创建满足可用性和性能目标的服务。为此，组织必须考虑当前技术限制、可用性预算和应用配置文件来构建服务。具体而言，组织应定义并构建一种服务，以便在可用性模型分配的短时间内一致、快速地识别并解决问题。组织还必须定义一种服务，该服务可以快速识别并解决潜在的服务问题，如果忽略，这些问题会影响可用性和性能。

您不会在一夜之间达到期望的服务级别。专业知识不足、当前流程限制或人员配备不足等缺点可能会妨碍组织实现所需的标准或目标，即使在之前的服务分析步骤之后也是如此。没有精确的方法可以将所需服务级别与所需目标精确匹配。为此，组织应测量服务标准并测量用于支持服务标准的服务参数。当组织未能达到服务目标时，它应寻求服务指标来帮助理解问题。在许多情况下，可以增加预算以改进支持服务并进行必要的改进，以实现所需的服务目标。随着时间的推移，组织可能会对服务目标或服务定义进行多次调整，以调整网络服务和业务需求。

例如，当目标在99.9%的可用性时，组织可能实现99%的可用性。在查看服务和支持指标时，组织的代表发现硬件更换大约需要24小时，比最初估计的时间长得多，因为组织的预算只有4小时。此外，该组织发现主动管理功能被忽略，冗余网络设备未修复。他们还发现，他们没有人力去改进。因此，在考虑降低当前服务目标后，组织为达到所需服务级别所需的额外资源编入了预算。

服务定义应包括被动支持定义和主动定义。被动定义定义了组织在从用户投诉或网络管理功能中发现问题后将如何应对问题。主动定义描述组织如何识别和解决潜在网络问题，包括修复损坏的“备用”网络组件、错误检测以及容量阈值和升级。以下各节提供被动和主动服务级别定义的示例。

被动服务级别定义

以下服务级别区域通常使用帮助台数据库统计信息和定期审计进行测量。下表显示组织问题严重性的示例。请注意，该图表不包括如何处理新服务请求，新服务请求可能由SLA或其他应用程序分析和性能假设分析处理。通常，严重程度为5级的服务可能是请求新服务（如果通过相同的支持流程处理）。

严重性1	严重性2	严重性3	严重性4
严重的业务影响 LAN用户或服务器分段关闭 关键WAN站点	因丢失或降级而对业务造成的高影响，可能的变通方法是将园区LAN停机；5-99个用户影响国内广域网 站点关闭国际广域网 站点关闭关键性能影响	某些特定网络功能会丢失或降级，例如冗余丢失 园区LAN性能会影响LAN冗余丢失	对组织没有业务影响的功能 查询或故障

定义问题严重性后，定义或调查支持流程以创建服务响应定义。通常，服务响应定义需要分层支持结构与帮助台软件支持系统相结合，以通过故障单跟踪问题。还应提供每个优先级的响应时间和解决时间、按优先级划分的呼叫数以及响应/解决质量等指标。要定义支持流程，它有助于定义组织中每个支持层的目标及其角色和职责。这有助于组织了解每个支持级别的资源要求和专业知识级别。下表提供了分层支持组织的示例，其中包含问题解决指南。

支持层	责任	目标
第1层支持	全职帮助台支持应答支持呼叫、发出故障单、处理最多15分钟的问题、记录故障单并上报到适当的第2层支持	解决40%的来电
第2层支持	队列监控、网络管理、工作站监控为软件确定的问题发出故障单实施从第1层、供应商和第3层上报接听呼叫在解决问题之前承担呼叫所有权	100%的呼叫在2级级别得到解决
第3层支持	必须立即为第2层提供所有优先级1问题的支持同意帮助解决SLA解决期内第2层未解决的所有问题	无直接问题所有权

下一步是创建服务响应和服务解析服务定义矩阵。这设定了解决问题的速度目标，包括硬件更换。在此领域制定目标非常重要，因为服务响应时间和恢复时间直接影响网络可用性。问题解决时间也应与可用性预算保持一致。如果可用性预算中没有考虑大量严重性较高的问题，则组织可以努力了解这些问题的根源，并找到潜在的补救方法。请参阅下表：

问题严重性	帮助台响应	第2层响应	现场第2层	硬件备件更换	问题解决方法
1	立即上报到第2层	5分	2小	2小时	4小时

	, 网络运营经理	钟	时		
2	立即上报到第2层 , 网络运营经理	5分 钟	4小 时	4小时	8小时
3	15分钟	2小 时	12小 时	24小 时	36小 时
4	15分钟	4小 时	3天	3天	6天

除服务响应和服务解决外，还要构建上报矩阵。上报表有助于确保可用资源专注于严重影响服务的问题。总的来说，当分析师专注于解决问题时，他们很少专注于为问题提供更多资源。定义何时应通知其他资源有助于提高管理中的问题意识，并通常有助于制定未来的主动或预防性措施。请参阅下表：

运行时间	严重性1	严重性2	严重性3	严重性4
5分钟	网络运营经理，第3层支持，网络总监			
1小时	更新到网络运营经理，第3层支持，网络总监	更新到网络运营经理，第3层支持，网络总监		
2小时	上报给副总裁，更新给总监、运营经理			
4小时	对VP、总监、运营经理、第3层支持的根本原因分析需要CEO通知	上报给副总裁，更新给总监、运营经理		
24小时			网络运营经理	
5天				网络运营经理

到目前为止，服务级别定义侧重于运营支持组织在发现问题后如何对问题作出反应。运营组织多年来一直在制定运营支持计划，其中包含类似于上述信息。但是，这些情况下缺少的是组织如何识别问题以及他们将识别哪些问题。更复杂的网络组织尝试通过为主动发现的问题的百分比制定目标来解决此问题，而不是通过用户问题报告或投诉重新发现的问题。

下表显示组织可能希望如何衡量主动支持功能和主动支持的整体情况。

网络区域	主动问题识别率	反应问题识别率
------	---------	---------

LAN	80 %	20 %
WAN	80 %	20 %

这是定义更主动支持定义的良好开端，因为它简单且相当容易测量，特别是当主动工具自动生成故障单时。这还有助于将网络管理工具/信息重点放在主动解决问题上，而不是帮助根本原因。但是，此方法的主要问题是它不定义主动支持要求。这通常会在主动支持管理功能方面造成差距，并导致额外的可用性风险。

主动服务级别定义

一种创建服务级别定义的更全面的方法，包括更详细地介绍如何监控网络以及运营组织如何对定义的网络管理站(NMS)阈值作出7 x 24响应。鉴于管理信息库(MIB)变量的庞大数量以及与网络运行状况相关的可用网络管理信息量，这似乎是一项不可能完成的任务。它也可能极其昂贵且资源密集。遗憾的是，这些异议妨碍了很多人实施主动服务定义，从本质上讲，该定义应简单、相当容易遵循，并且仅适用于网络中最大的可用性或性能风险。如果组织随后看到基本主动服务定义中的价值，则只要您实施分阶段的方法，随着时间推移，可以添加更多变量，而不会产生重大影响。

在所有运营支持计划中包括主动服务定义的第一个方面。服务定义只是说明运营组如何主动识别和响应网络不同区域的网络或链路中断情况。如果没有此定义（或管理支持），企业或机构将期望获得可变的支持、不切实际的用户期望，并最终降低网络可用性。

下表显示组织如何为链路/设备关闭条件创建服务定义。该示例显示的企业组织可能根据一天中的时间和网络区域有不同的通知和响应要求。

网络设备或链路关闭	Detection Method	5 x 8通知	7 x 24通知	5 x 8分辨率	7 x 24分辨率
核心 LAN	SNMP 设备和链路轮询、陷阱	NOC 创建故障单，寻呼 LAN 值班传呼机	自动寻呼 LAN 值班传呼机，LAN 值班人员为核心 LAN 队列创建故障单	NOC 在 15 分钟内分配 LAN 分析师，根据服务响应定义进行维修	优先级 1 和 2 立即调查和解决优先级 3 和 4 排队早间解决
国内广域网	SNMP 设备和链路轮询、陷阱	NOC 创建故障单，寻呼 WAN 值班传呼机	自动寻呼 WAN 值班传呼机，WAN 值班人员为 WAN 队列创建故障单	NOC 在 15 分钟内分配 WAN 分析师，根据服务响应定义进行维修	优先级 1 和 2 立即调查和解决优先级 3 和 4 排队早间解决
外联网	SNMP 设备和链	NOC 创建故障单、寻呼合作	自动页合作伙伴税寻呼机，合作伙伴税人员为	合作伙伴分析师在 15 分钟内由 NOC 分配	优先事项 1 和 2 立即调查和解决；优先级

路轮 询、 陷阱	伙伴责 任寻呼 机	合作伙伴队 列创建故障 单	，根据服务 响应定义进 行维修	3和4队列 ，用于早间 解决问题
----------------	-----------------	---------------------	-----------------------	------------------------

其余的主动服务级别定义可分为两类：网络错误和容量/性能问题。只有一小部分网络组织在这些领域具有服务级别定义。因此，这些问题会被忽略或偶尔处理。在某些网络环境中，这可能不错，但高可用性环境通常需要一致的主动服务管理。

网络组织往往因为多种原因而难以应对主动服务定义。这主要是因为他们尚未根据可用性风险、可用性预算和应用问题对主动服务定义执行需求分析。这导致对主动服务定义的要求不明确，收益不明确，尤其是因为可能需要额外的资源。

第二个原因涉及平衡可与现有或新定义的资源进行的主动管理量。仅生成对可用性或有严重潜在影响的警报。您还必须考虑事件关联管理或流程，以确保不会针对同一问题生成多个主动故障单。组织可能难以应对的最后一个原因是，创建一组新的主动警报通常会生成大量以前未被发现的邮件。运营团队必须为最初的大量问题和其他短期资源做好准备，以修复或解决这些以前未发现的情况。

主动服务级别定义的第一类是网络错误。网络错误可以进一步细分为系统错误，包括软件错误或硬件错误、协议错误、介质控制错误、准确性错误和环境警告。制定服务级别定义首先要大致了解如何检测这些问题情况、谁将查看这些情况，以及发生这些情况时会发生什么。如果需要，将特定消息或问题添加到服务级别定义。您可能还需要在以下方面做更多工作以确保成功：

- 第1级、第2级和第3级支持职责
- 将网络管理信息的优先级与运营团队能够有效处理的主动工作量进行平衡
- 培训要求，以确保支持人员能够有效处理已定义的警报
- 事件关联方法，确保不为同一根本原因问题生成多个故障单
- 有关特定消息或警报的文档，有助于在第1层支持级别识别事件

下表显示了网络错误的服务级别定义示例，该示例清楚了解谁负责主动网络错误警报、如何识别问题以及发生问题时会发生什么。组织可能仍需要如上所述的额外努力来确保成功

S

错误类别	Detection Method	阈值	执行的操作
软件错误 (软件强制崩溃)	使用系统日志查看器每日查看系统日志消息按第2层支持完成	优先级0、1和2的任何事件级别3或更高的100次以上	检查问题、创建故障单，并在出现新问题或需要注意问题时派送
硬件错误 (硬件强制崩溃)	使用系统日志查看器每日查看系统日志消息按第2层支持完成	优先级0、1和2的任何事件级别3或更高的100次以上	检查问题、创建故障单，并在出现新问题或需要注意问题时派送
协议错误 (仅IP路由协议)	使用系统日志查看器每日查看系统日志消息按第2层支持完成	每天10条优先级0、1和2的消息超过100次出现3级或以上	检查问题、创建故障单，并在出现新问题或需要注意问题时派送
介质控制	使用系统日志	每天10条优先	检查问题、创

错误 (仅限FDDI、POS和快速以太网)	查看器每日查看系统日志消息按第2层支持完成	级0、1和2的消息超过100次出现3级或以上	建故障单，并在出现新问题或需要注意问题时派送
环境消息 (电源和温度)	使用系统日志查看器每日查看系统日志消息按第2层支持完成	任何消息	创建故障单并派遣新问题
准确性错误 (链路输入错误)	5分钟间隔的SNMP轮询NOC接收的阈值事件	输入或输出错误任何链路上的任意5分钟间隔内出现一个错误	为新问题创建故障单并派遣到第2层支持

主动服务级别定义的其他类别适用于性能和容量。真正的性能和容量管理包括异常管理、基线建立和趋势分析以及假设分析。服务级别定义只是定义性能和容量异常阈值以及将启动调查或升级的平均阈值。然后，这些阈值可能会以某种方式应用于所有三个性能和容量管理流程。

容量和性能服务级别定义可分为以下几类：网络链路、网络设备、端到端性能和应用性能。在这些领域开发服务级别定义需要深入的技术知识，以了解设备容量、介质容量、QoS特征和应用需求的具体方面。因此，我们建议网络架构师通过供应商的输入来制定与性能和容量相关的服务级别定义。

与网络错误一样，为容量和性能制定服务级别定义首先要大致了解如何检测这些问题情况、谁将查看这些情况以及发生这些情况时会发生什么。如果需要，可以将特定事件定义添加到服务级别定义。您可能还需要在以下方面做更多工作以确保成功：

- 清楚了解应用性能要求
- 根据业务需求和总体成本对组织有意义的阈值进行深入技术调查
- 预算周期和周期外升级要求
- 第1级、第2级和第3级支持职责
- 网络管理信息的优先级和重要性与运营团队能够有效处理的主动工作量之间保持平衡
- 培训要求，确保支持人员了解消息或警报，并能够有效处理已定义的条件
- 事件关联方法或流程，确保不为同一根本原因问题生成多个故障单
- 有关特定消息或警报的文档，有助于在第1层支持级别识别事件

下表显示了链路利用率的服务级别定义示例，该示例清楚地了解谁负责主动网络错误警报、如何识别问题以及发生问题时将发生什么。组织可能仍需要如上所述的额外努力来确保成功。

网络区域/介质	Detection Method	阈值	执行的操作
园区LAN主干和分布层链路	5分钟间隔的SNMP轮询核心和分布层链路上的RMON异常陷阱	5分钟间隔内50%利用率通过异常陷阱实现90%利用率	向性能电子邮件别名组发送电子邮件通知，以评估QoS要求或计划升级以解决周期性问题
国内WAN	SNMP轮询，间隔为5分	5分钟间隔内75%的利	向性能电子邮件别名组发送电子邮件通知

链路	钟	用率	，以评估QoS要求或计划升级以解决周期性问题
外联网WAN链路	SNMP轮询，间隔为5分钟	5分钟间隔内60%的利用率	向性能电子邮件别名组发送电子邮件通知，以评估QoS要求或计划升级以解决周期性问题

下表定义了设备容量和性能阈值的服务级别定义。确保您创建的阈值在防止网络问题或可用性问题方面有意义且有用。这是一个非常重要的领域，因为未检查的设备控制平面资源问题可能会严重影响网络。

Cisco 7500	CPU、内存、缓冲区	SNMP轮询，间隔为5分钟 CPU的RMON通知	在5分钟间隔内CPU占75%，通过RMON通知内存占99%在5分钟间隔内内存占50%缓冲区占99%的利用率	向性能和容量电子邮件别名组发送电子邮件通知，以解决问题或计划以99%的价格升级RMON CPU，放置故障单和第2层支持寻呼机
Cisco 2600	CPU、内存	SNMP轮询，间隔为5分钟	5分钟间隔内CPU占75% 5分钟间隔内内存占50%	通过电子邮件通知性能和容量电子邮件别名组以解决问题或计划升级
Catalyst 5000	背板利用率、内存	SNMP轮询，间隔为5分钟	50%利用率的背板 75%利用率的内存	通过电子邮件通知性能和容量电子邮件别名组以解决问题或计划升级
Light Stream®	CPU、内存	SNMP轮询，间隔为5分钟	CPU利用率为65%内存利用率为50%	通过电子邮件通知性能和容量电子邮件别名组以解决问题或计划升级

1010ATM交换机				
------------	--	--	--	--

下表定义了端到端性能和容量的服务级别定义。这些阈值通常基于应用要求，但也可用于指示某些类型的网络性能或容量问题。大多数具有性能服务级别定义的组织只创建少量的性能定义，因为从网络中的每个点到其他点的性能测量需要大量资源，并会产生大量网络开销。这些端到端性能问题也可能出现在链路或设备容量阈值中。我们建议按地理区域进行一般定义。如有必要，可能会添加一些关键站点或链路。

网络区域/介质	测量方法	阈值	执行的操作
园区LAN	无问题预期无问题难以测量整个LAN基础设施	10毫秒或更短的往返响应时间	通过电子邮件通知性能和容量电子邮件别名组以解决问题或计划升级
国内WAN链路	仅使用互联网性能监控器(IPM)ICMP回应，从SF到NY和SF到芝加哥的当前测量	平均5分钟内75毫秒的往返响应时间	向性能电子邮件别名组发送电子邮件通知，以评估QoS要求或计划升级以解决周期性问题
旧金山至东京	从旧金山到布鲁塞尔的当前测量值使用IPM和ICMP回应	5分钟内平均250毫秒的往返响应时间	向性能电子邮件别名组发送电子邮件通知，以评估QoS要求或计划升级以解决周期性问题
旧金山至布鲁塞尔	从旧金山到布鲁塞尔的当前测量值使用IPM和ICMP回应	平均5分钟内175毫秒的往返响应时间	向性能电子邮件别名组发送电子邮件通知，以评估QoS要求或计划升级以解决周期性问题

服务级别定义的最后一个方面是应用性能。应用程序性能服务级别定义通常由应用程序或服务器管理组创建，因为服务器本身的性能和容量可能是影响应用程序性能的最大因素。网络组织可以通过

创建网络应用性能的服务级别定义来实现巨大优势，因为：

- 服务级别定义和衡量有助于消除组之间的冲突。
- 如果为关键应用配置了QoS，而其他流量被视为可选流量，则各个应用的服务级别定义非常重要。

如果您选择创建和衡量应用性能，则最好不对服务器本身衡量性能。这样有助于区分网络问题和应用或服务器问题。使用探测或在思科路由器上运行的系统可用性代理软件以及控制数据包类型和测量频率的思科IPM。

下表显示了应用性能的简单服务级别定义。

应用	测量方法	阈值	执行的操作
企业资源规划(ERP)应用TCP端口1529布鲁塞尔至旧金山	使用IPM测量端口1529往返性能布鲁塞尔网关到SFO网关2的布鲁塞尔到旧金山	平均5分钟内175毫秒的往返响应时间	向性能电子邮件别名组发送电子邮件通知，以评估问题或计划升级以解决周期性问题的
ERP应用TCP端口1529东京至旧金山	使用IPM测量端口1529往返性能布鲁塞尔网关到SFO网关2的布鲁塞尔到旧金山	5分钟内平均200毫秒的往返响应时间	向性能电子邮件别名组发送电子邮件通知，以评估问题或计划升级以解决周期性问题的
客户支持应用TCP端口1702悉尼至SF	使用IPM测量端口1702往返性能悉尼网关到SFO网关1的悉尼到旧金山	5分钟内平均250毫秒的往返响应时间	向性能电子邮件别名组发送电子邮件通知，以评估问题或计划升级以解决周期性问题的

步骤 6：收集指标和监控

除非组织收集指标并监控成功，否则服务级别定义本身就毫无价值。在创建关键服务级别定义时，定义如何衡量和报告服务级别。衡量服务级别可确定组织是否达到目标，并确定可用性或性能问题的根本原因。在选择衡量服务级别定义的方法时，还要考虑目标。有关[详细信息](#)，请[参阅](#)创建和维护SLA。

监控服务级别需要定期召开审查会议，讨论定期服务。讨论所有指标以及它们是否符合目标。如果它们不符合要求，请确定问题的根本原因并实施改进。您还应介绍当前在改善个人情况方面的计划和进展。

创建和维护SLA

服务级别定义是一个出色的构建块，它有助于在整个组织内创建一致的QoS并帮助提高可用性。下一步是SLA，这是一项改进，因为SLA将业务目标和成本要求直接与服务质量保持一致。然后，精心构建的SLA通过维护针对网络问题或问题的明确流程和程序，作为用户群和支持小组之间效率、质量和协同作用的模型。

SLA具有以下几个优势：

- SLA对服务建立双向责任制，这意味着用户和应用组也对网络服务负责。如果他们不帮助为特定服务创建SLA并与网络组沟通业务影响，那么他们实际上可能会对问题负责。
- SLA有助于确定满足业务需求所需的标准工具和资源。在没有SLA的情况下，决定需要使用多少人和哪些工具通常是预算猜测。服务可能过度设计，导致过度支出，或者过度设计，导致未实现业务目标。调整SLA有助于实现平衡的最佳水平。
- 记录的SLA为设置服务级别期望创造了更清晰的工具。

我们建议在创建服务级别定义后，执行以下步骤来构建SLA:我们建议在创建服务级别定义后，执行以下步骤来构建SLA:

7. [满足SLA的必备条件。](#)

8. [确定SLA所涉及的各方。](#)

9. [确定服务要素。](#)

十、 [了解客户业务需求和目标](#)

11. [定义每个组所需的SLA。](#)

12. [选择SLA的格式](#)

13. [开发SLA工作组](#)

14. [举行工作组会议并起草SLA。](#)

15. [协商SLA。](#)

16. [测量和监控SLA符合性。](#)

[步骤 7：满足SLA的必备条件](#)

IT SLA开发专家确定了成功SLA的三个前提条件。遗憾的是，未能达到这些目标的组织可能会预期SLA流程出现问题，并应考虑SLA流程中涉及的潜在问题。如果网络组织能够构建满足一般业务要求的 service 级别定义，则未能实施SLA并不有害。以下是SLA流程的先决条件：

- 您的企业必须具备以服务为导向的文化。组织必须首先满足客户的需求。您需要自上而下的优先服务承诺，从而全面了解客户需求和认知。进行客户满意度调查和客户驱动的服务计划。另一个服务指标可能是组织将服务或支持满意度作为公司目标。这种情况并不罕见，因为IT组织现在与整个组织的成功息息相关。服务文化很重要，因为SLA流程从根本上来讲是根据客户需求和业务需求进行改进。如果组织过去没有这样做，他们会发现SLA流程很困难。
- 客户/业务计划必须推动所有IT活动。公司愿景或使命陈述必须与客户和业务计划保持一致，这将推动所有IT活动，包括SLA。通常，网络已部署到位，无法实现特定目标，但网络组却看不到该目标及后续业务要求。在这些情况下，会为网络分配一定的预算，这可能会对当前需求做出过度反应或严重低估需求，从而导致故障。当客户/业务计划与IT活动保持一致时，网络组织可以更轻松地适应新应用部署、新服务或其他业务需求。双方建立了关系，并将共同的总体重点放在实现企业目标上，所有团队都以团队的形式执行。
- 您必须承诺SLA流程和合同。首先，必须承诺学习SLA流程，以制定有效的协议。第二，必须履行合同的 service 要求。在没有参与的所有人员提供大量投入和承诺的情况下，不要期望创建强

大的SLA。此承诺还必须来自管理层和与SLA流程相关的所有人员。

步骤 8::确定SLA所涉及的各方

企业级网络SLA严重依赖网络元素、服务器管理元素、帮助台支持、应用元素以及业务或用户要求。通常，SLA流程将涉及每个区域的管理。当组织构建基本的被动式支持SLA时，此方案非常有效。在SLA过程中，具有更高可用性要求的企业组织可能需要技术支持，以帮助解决可用性预算、性能限制、应用分析或主动管理功能等问题。对于更主动的管理SLA方面，我们建议由网络架构师和应用架构师组成的技术团队。技术支持可以更接近网络的可用性和性能能力，以及实现特定目标所需的功能。

服务提供商SLA通常不包括用户输入，因为创建这些SLA的唯一目的是获得其他服务提供商的竞争优势。在某些情况下，高级管理层会以非常高可用性或高性能的级别创建这些SLA，以推广其服务并为内部员工提供内部目标。其他服务提供商将专注于通过创建强大的服务级别定义来提高可用性的技术方面，这些定义在内部进行衡量和管理。在其他情况下，这两种努力是同时进行的，但不一定是同时进行的，也不一定是同一目标。

然后，根据SLA的目标选择SLA所涉及的各方。一些可能的目标包括：

- 实现被动式支持业务目标
- 通过定义主动SLA提供最高级别的可用性
- 推广或销售服务

步骤 9：确定服务要素

主要服务/支持SLA通常包含许多组件，包括支持级别、如何衡量、SLA协调的上报路径以及总体预算问题。高可用性环境的服务元素应包括主动服务定义和被动目标。其他详细信息包括：

- 现场支持工作时间和非工作时间支持流程
- 优先级定义，包括问题类型、开始处理问题的最长时间、解决问题的最长时间以及上报程序
- 要支持的产品或服务，按业务重要性的顺序排列
- 支持专业知识期望、性能级期望、状态报告和用户解决问题的责任
- 地理或业务部门支持级别的问题和要求
- 问题管理方法和程序（呼叫跟踪系统）
- 帮助台目标
- 网络错误检测和服务响应
- 网络可用性衡量和报告
- 网络容量和性能测量与报告
- 冲突解决程序
- 为实施的SLA提供资金

网络化应用或服务SLA可能会根据用户组要求和业务关键性有其他需求。网络组织必须密切倾听这些业务需求，并开发适合整体支持结构的专业解决方案。融入整体支持文化至关重要，因为不要为某些个人或团体创建高级服务非常重要。在许多情况下，这些额外要求可以归入“解决方案”类别。例如，基于业务需求的白金、金牌和银牌解决方案。请参阅以下针对特定业务需求的SLA要求示例。

注意：支持结构、升级路径、帮助台程序、衡量标准和优先级定义应基本保持不变，以保持和改进一致的服务文化。

- 突发带宽要求和功能

- 性能要求
- QoS要求和定义
- 建立解决方案矩阵的可用性要求和冗余
- 监控和报告要求、方法和程序
- 应用/服务元素的升级标准
- 为预算外需求或交叉计费方法提供资金

例如，您可以为WAN站点连接创建解决方案类别。该白金级解决方案将为现场提供双T1服务。不同的载波将提供每条T1线路。该站点将配置两台路由器，以便在任何T1或路由器发生故障时，该站点不会发生故障。金牌服务将有两台路由器，但会使用备用帧中继。此解决方案在中断期间的带宽可能有限。银牌解决方案将只有一个路由器和一个运营商服务。这些解决方案中的任何一种都将考虑问题票证的不同优先级。如果停机需要优先1或2票证，某些组织可能需要白金或金牌解决方案。然后，客户组织可以为其所需的服务级别提供资金。下表显示了一个组织示例，该组织提供三个服务级别，具体取决于企业对外联网连接的需求。

解决方案	白金	金牌	银牌
设备	用于WAN连接的冗余路由器	用于在核心站点备份的冗余路由器	无设备冗余
WAN	冗余T1连接，多个运营商	T1与帧中继备份的连接	无广域网冗余
带宽要求和突发	冗余T1，实现突发负载分担	非负载共享，仅用于关键应用的帧中继备份；仅帧中继64K CIR	最多T1
性能	100毫秒或更短的往返响应时间	响应时间100毫秒或更短，预期为99.9%	响应时间100毫秒或更短，预期为99%
可用性要求	99.99%	99.95%	99.9%
服务台关闭时的优先级	优先级 1:关键业务服务中断	优先级 2:影响业务的服务中断	优先级 3:业务连接断开

步骤 10：了解客户业务需求和目标

此步骤使SLA开发人员获得了极大的可信度。通过了解各业务组的需求，初始SLA文档将更接近业务需求和预期结果。尝试了解客户服务的停机成本。估算工作效率、收入和客户商誉损失。请记住，即使与少数人建立简单的联系，也会严重影响收入。在这种情况下，请务必帮助客户了解可能发生的可用性和性能风险，以便组织更好地了解其需要的服务级别。如果您未能完成此步骤，许多客户可能只是要求100%的可用性。

SLA开发人员还应了解组织的业务目标和增长情况，以适应网络升级、工作负载和预算。了解将要使用的应用也很有帮助。希望组织对每个应用程序都有应用程序配置文件，但如果没有，则考虑对应用程序进行技术评估，以确定与网络相关的问题。

步骤 11：定义每个组所需的SLA

主要支持SLA应包括关键业务部门和功能组表示，如网络运营、服务器运营和应用支持组。应根据业务需求以及支持流程中的部分来识别这些组。拥有来自多个群体的代表性还有助于创建一个公平的整体支持解决方案，而无需个人群体的偏好或优先级。这可能导致支持组织向各个组提供优质服务，这种情况可能会破坏组织的整体服务文化。例如，客户可能会坚称，在公司内，他的应用程序是最关键的应用程序，但实际上，该应用程序的停机成本在收入损失、工作效率损失和客户商誉损失方面远低于其他应用程序。

组织内的不同业务部门将有不同的要求。网络SLA的一个目标应是就适应不同服务级别的一种整体格式达成一致。这些要求通常是可用性、QoS、性能和MTTR。在网络SLA中，这些变量通过确定业务应用的优先级以调整潜在QoS调整、定义不同网络影响问题的MTTR的帮助台优先级以及开发解决方案矩阵来帮助处理不同的可用性和性能要求来处理。企业制造公司的简单解决方案矩阵示例可能类似于下表。您可以添加有关可用性、QoS和性能的信息。

业务部门	应用	停机成本	关闭时的问题优先级	服务器/网络要求
制造	ERP	高	1	最高冗余
用户技术支持	客户服务	高	1	最高冗余
工程	文件服务器、ASIC设计	中	2	LAN核心冗余
营销	文件服务器	中	2	LAN核心冗余

步骤 12：选择SLA的格式

SLA的格式可能因组愿望或组织要求而异。以下是网络SLA的推荐示例大纲：

1. 协议目的参与协议的各方协议的目标
2. 提供的服务和支持的产品帮助台服务和呼叫跟踪基于MTTR定义的业务影响的问题严重性定义QoS定义的业务关键型服务优先级根据可用性和性能要求定义的解决方案类别培训需求容量规划要求上报要求报告提供的网络解决方案新解决方案请求不支持的产品或应用
3. 业务策略在工作时间内提供支持非工作时间支持定义假期覆盖联系电话工作负载预测申诉解决服务授权条件用户和组安全责任
4. 问题管理程序呼叫发起（用户和自动）一级响应和呼叫修复率呼叫跟踪和记录保留主叫方职责问题诊断和呼叫关闭要求网络管理问题检测和服务响应问题解决类别或定义慢性问题处理严重问题/异常呼叫处理
5. 服务质量目标质量定义度量定义质量目标按问题优先级启动问题解决的平均时间按问题优先级平均解决问题的时间按问题优先级更换硬件的平均时间网络可用性和性能管理容量管理增长质量报告
6. 人员配备和预算人员配备模式运营预算
7. 协议维护一致性审核计划绩效报告和审核报告指标的协调定期SLA更新
8. 审批
9. 附件和展品呼叫流程图上报表网络解决方案矩阵报告示例

步骤 13：开发SLA工作组

下一步是确定SLA工作组的参与者，包括组领导。工作组可以包括来自业务部门或职能部门的用户或经理或来自地理基础的代表。这些人员将SLA问题传达给各自的工作组。能够就关键SLA要素达成一致的经理和决策者应参与。这些人员可能包括管理人员和技术人员，他们可以帮助定义与SLA相关的技术问题并做出IT级别决策（例如，帮助台经理、服务器运营经理、应用经理和网络运营经理）。

网络SLA工作组还应包含广泛的应用和业务表示，以便在包含许多应用和服务的一个网络SLA上达成协议。工作组应有权对网络的关键业务流程和服务以及各服务的可用性和性能要求进行排名。此信息将用于为不同影响业务的问题类型创建优先级，确定网络上关键业务流量的优先级，并根据业务需求创建未来的标准网络解决方案。

步骤 14：召开工作组会议并拟定SLA

工作组最初应创建工作组章程。章程应说明SLA的目标、计划和时间框架。接下来，小组应制定具体的任务计划并确定制定和实施SLA的时间表和里程碑。该小组还应制定报告流程，根据支持标准衡量支持水平。最后一步是创建SLA协议草案。

网络SLA工作组最初应每周召开一次会议以制定SLA。在创建并批准SLA后，组可能每月或甚至每季度满足SLA更新要求。

步骤 15：协商SLA

创建SLA的最后一步是最终协商和签核。此步骤包括：

- 审核草稿
- 协商内容
- 编辑和修改文档
- 获得最终批准

审核草稿、协商内容和进行修订的这一周期可能需要多个周期，最终版本才能发送给管理层供审批。

从网络管理员的角度来看，协商可衡量的可实现结果非常重要。尝试备份与其他相关组织签订的性能和可用性协议。这可能包括质量定义、测量定义和质量目标。请记住，添加的服务等同于额外费用。确保用户组了解额外服务级别的成本会更高，并让他们在关键业务要求时做出决策。您可以轻松地对SLA的许多方面（如硬件更换时间）执行成本分析。

步骤 16：衡量和监控SLA一致性

衡量SLA符合性和报告结果是SLA流程的重要方面，有助于确保长期一致性和结果。我们一般建议SLA的任何主要组件都可以衡量，并在实施SLA之前制定衡量方法。然后，在用户和支持小组之间召开月度会议，以检查测量结果，确定问题根本原因，并提出满足或超过服务级别要求的解决方案。这有助于使SLA流程类似于任何现代质量改进计划。

以下部分提供有关组织内管理如何评估其SLA及其整体服务级别管理的更多详细信息。

服务级别管理绩效指标

服务级别管理绩效指标提供了一种机制，用于监控和提高服务级别，作为衡量成功标准。这使组织能够更快地对服务问题作出反应，并更轻松地了解影响服务或其环境中停机时间成本的问题。不衡量服务级别定义也会否定任何积极主动的工作，因为组织被迫采取被动的态度。没有人会打电话

说服务运行良好，但许多用户会打电话说服务不符合他们的要求。

因此，服务级别管理绩效指标是服务级别管理的主要要求，因为它们提供了充分了解现有服务级别并根据当前问题进行调整的方法。这是提供主动支持和提高质量的基础。当组织对问题进行根本原因分析并改进质量时，这可能是改进可用性、性能和服务质量的最佳方法。

例如，请考虑以下真实场景。X公司收到大量用户投诉，称网络经常长时间停机。通过衡量可用性，公司发现主要问题在于几个WAN站点。对这些景点进行更深入的调查发现，大多数问题都出在几个广域网站点。找到了根本原因，组织解决了问题。然后，组织为可用性设置服务级别目标并与用户组达成协议。未来的测量会由于不符合SLA而快速发现问题。然后，该网络组被认为具有更高的专业性、专业知识，并且是组织的整体资产。该集团有效地从被动型转变为主动型，并帮助公司实现盈利。

遗憾的是，如今大多数网络组织的服务级别定义有限，没有绩效指标。因此，他们将大部分时间用于响应用户投诉或问题，而不是主动确定根本原因并构建满足业务需求的网络服务。

使用以下SLA性能指标确定服务级别管理流程的成功：

- 记录的服务级别定义或SLA，包括可用性、性能、被动服务响应时间、问题解决目标和问题上报
- 绩效指标指标，包括可用性、性能、按优先级的服务响应时间、按优先级解决的时间，以及其他可衡量的SLA参数
- 每月举行网络服务级别管理会议，以审查服务级别合规性并实施改进

书面服务级别协议或服务级别定义

第一个性能指标只是详细描述SLA或服务级别定义的文档。服务级别定义的主要目标应该是可用性和性能，因为这些是主要用户要求。

次要目标非常重要，因为它们有助于定义如何实现可用性或性能级别。例如，如果组织具有严格的可用性和性能目标，那么防止问题发生并在问题发生时快速解决问题就非常重要。次要目标有助于定义实现所需的可用性和性能级别所需的流程。

被动的次要目标包括：

- 按呼叫优先级的被动服务响应时间
- 问题解决目标或MTTR
- 问题上报流程。

主动辅助目标包括：

- 设备关闭或链路关闭检测
- 网络错误检测
- 容量或性能问题检测。

主要目标、可用性和性能的服务级别定义应包括：

目标

- 如何衡量目标
- 负责衡量可用性和性能的各方
- 负责可用性和绩效目标的各方

- 不符合项流程

如果可能，我们建议负责衡量的各方和负责结果的各方应有所不同，以防止利益冲突。有时，您可能还需要调整可用性编号，因为存在添加/移动/更改错误、未检测到的错误或可用性测量问题。服务级别定义还可以包括修改结果的过程，以帮助提高准确性并防止不适当的调整。有关衡量可用性和性能的方法，请参阅下一节。

被动辅助目标的服务级别定义定义了组织在确定网络或IT范围的问题后如何响应，包括：

- 问题优先级定义
- 按呼叫优先级的被动服务响应时间
- 问题解决目标或MTTR
- 问题上报流程

总的来说，这些目标定义了在任何给定时间对问题负责的人，以及负责人应放弃当前任务以处理已定义问题的程度。与其他服务级别定义一样，服务级别文档应详细说明如何衡量目标、负责衡量的各方以及不符合性流程。

主动辅助目标的服务定义定义了组织如何提供主动支持，包括识别网络关闭、链路关闭或设备关闭情况、网络错误情况和网络容量阈值。设定目标，促进主动管理，因为质量主动管理有助于消除问题并帮助更快地解决问题。这通常通过设定一个目标来实现，即创建并解决多少个主动案例，而无需用户通知。许多组织在帮助台软件中设置了一个标志，以便确定主动案例与被动案例。服务级别文档还应包含有关如何衡量目标、负责衡量的各方以及不符合性流程的信息。

绩效指标指标

我们始终建议可以衡量任何已定义的服务级别目标，使组织能够衡量服务级别，确定妨碍可用性和性能这一主要目标的根本原因服务问题，并针对特定目标进行改进。总体而言，指标只是一种工具，使网络经理能够管理服务级别一致性并根据业务需求进行改进。

遗憾的是，许多组织未收集可用性、性能和其他指标。组织将其归因于无法提供完整的准确性、成本、网络开销和可用资源。这些因素可能会影响衡量服务级别的能力，但组织应将重点放在管理和改进服务级别的总体目标上。许多组织已经能够创建低成本、低开销的指标，这些指标可能不能提供完全的准确性，但确实能满足这些主要目标。

衡量可用性和性能是服务级别指标中经常忽略的一个方面。成功采用这些指标的组织使用两种非常简单的方法。一种方法是将互联网控制消息协议(ICMP)从网络核心位置向边缘发送ping数据包。您也可以使用此方法获得性能。成功采用此方法的组织还将类似设备分组到“可用性组”，如LAN设备或国内现场办公室。这也很有吸引力，因为组织通常针对网络的不同地理或业务关键领域制定不同的服务级别目标。这允许度量组对具有可用性组的所有设备进行平均，以获得合理的结果。

计算可用性的另一种成功方法是使用故障单和称为受影响用户分钟数(IUM)的度量。此方法可列出受故障影响的用户数量，并乘以故障的分钟数。如果表示为时间段内总分钟数的百分比，则可以轻松转换为可用性。无论是哪种情况，确定并衡量停机时间的根本原因都会很有帮助，以便更轻松确定改进目标。根本原因类别包括硬件问题、软件问题、链路或运营商问题、电源或环境问题、更改故障和用户错误。

可衡量的被动支持目标包括：

- 按呼叫优先级的被动服务响应时间
- 问题解决目标或MTTR
- 问题上报时间

通过从帮助台数据库生成报告（包括以下字段）来衡量被动支持目标：

- 最初报告呼叫 (或输入到数据库) 的时间
- 处理该问题的个人接受呼叫的时间
- 问题升级的时间
- 问题关闭的时间

这些指标可能需要管理影响力，才能始终如一地在数据库中输入问题并实时更新问题。在某些情况下，组织可以自动生成网络事件或电子邮件请求的故障单。这有助于准确确定问题的开始时间。根据此类指标生成的报告通常会按优先级、工作组和个人对问题进行排序，以帮助确定潜在问题。

要衡量主动支持流程更加困难，因为它要求您监控主动工作并计算其有效性的一些度量。这方面的工作很少。但是，很明显，只有一小部分人实际上会向帮助台报告网络问题，当他们报告问题时，显然需要花时间解释问题或将问题隔离为与网络相关。并非所有主动案例都会立即影响可用性和性能，因为冗余设备故障或链路故障对最终用户影响甚微。

实施主动服务级别定义或协议的组织会因为业务需求和潜在可用性风险而这样做。然后根据主动案例的数量或百分比进行测量，而不是由用户生成的被动案例。同时，最好还要衡量每个领域的主动案例数量。这些类别包括设备关闭、链路关闭、网络错误和容量违规。还可以使用可用性建模和主动案例完成一些工作，以确定通过实施主动服务定义对可用性的影响。

[服务级别管理审核](#)

服务级别管理成功的另一个衡量标准是服务级别管理审核。无论SLA是否到位，都应执行此操作。在每月一次的会议中，与负责衡量和提供已定义服务级别的人员一起执行服务级别管理审核。当涉及SLA时，用户组也可能存在。会议之目的为审阅所衡量之服务级别定义之表现，并作出改进。

每次会议应有一个明确的议程，包括：

- 对给定期间衡量的服务级别进行审核
- 审查针对各个领域定义的改进计划
- 当前服务级别指标
- 根据当前指标集讨论需要哪些改进。

随着时间的推移，组织可能也会趋向服务级别合规性，以确定组的有效性。此过程与质量循环或质量改进过程无异。会议有助于针对个别问题并根据根本原因确定解决方案。

[服务级别管理摘要](#)

总之，服务级别管理允许组织从被动式支持模式转变为主动式支持模式，其中网络可用性和性能级别取决于业务需求，而不是最新的问题集。该流程有助于营造持续提高服务级别和提高业务竞争力的环境。服务级别管理也是主动式网络管理最重要的管理组件。因此，在任何网络规划和设计阶段都强烈建议进行服务级别管理，并且应从新定义的网络架构开始。这使组织能够在首次正确实施解决方案时，最大限度地减少停机时间或返工。

[相关信息](#)

- [技术支持 - Cisco Systems](#)