

Tableau Server 的可扩展性

关于在整个企业范围内扩展 Tableau Server 的入门概述

Neelesh Kamkolkar

数据和性能部门产品经理

纵观全球，各种规模的组织纷纷使用 Tableau 大规模地提供真正意义上的自助式分析。这些组织都在进行战略转型，着力打造一种对组织的未来至关重要的分析文化。随着这种文化逐渐成形，分析平台变得对组织至关重要，停机时间则成为切实担忧的问题。要想最大限度缩短停机时间，并确保分析平台能够满足企业不断增长的需求，务须正确进行部署和可扩展性规划。

在建立可随业务增长而扩展的企业环境时，需要考虑诸多因素。本白皮书概述了在建立 Tableau 环境时需要考虑的关键因素，并介绍了一个在规划可扩展性时可以依据的框架。此外，还分享了我们的可扩展性实验室得出的概要结果，并就如何规划部署工作和为未来增长做好准备给予了高屋建瓴的指导。

对于想要更深入了解具体试验、方法、环境、观测值和结果的受众，[此处](#)提供了与本白皮书配套并深入介绍相关技术详情的白皮书。

目录

可扩展性和可靠性.....	3
全面考量 Tableau 可扩展性的影响因素	4
数据对可扩展性规划的影响.....	6
规划 Tableau 可扩展性时可以依据的框架.....	7
示例情形.....	9
结果摘要.....	13
总结	15

Tableau 经实践检验的可扩展性和可靠性

您大可放心，Tableau Server 可通过扩展来满足您组织的需求。为了测试和验证 Tableau 10 是否达到了可扩展性方面的发布目标，我们在 Tableau Public 上部署了 Tableau Server 10。Tableau Public 是我们免费提供的平台，供用户与全世界分享可视化。在 Tableau Public 上的这次部署实际上是部署到由很多台计算机组成的一个横向群集，每周支持超过 700 万次查看，因此可以有效地测试 Tableau Server 的大规模扩展能力。

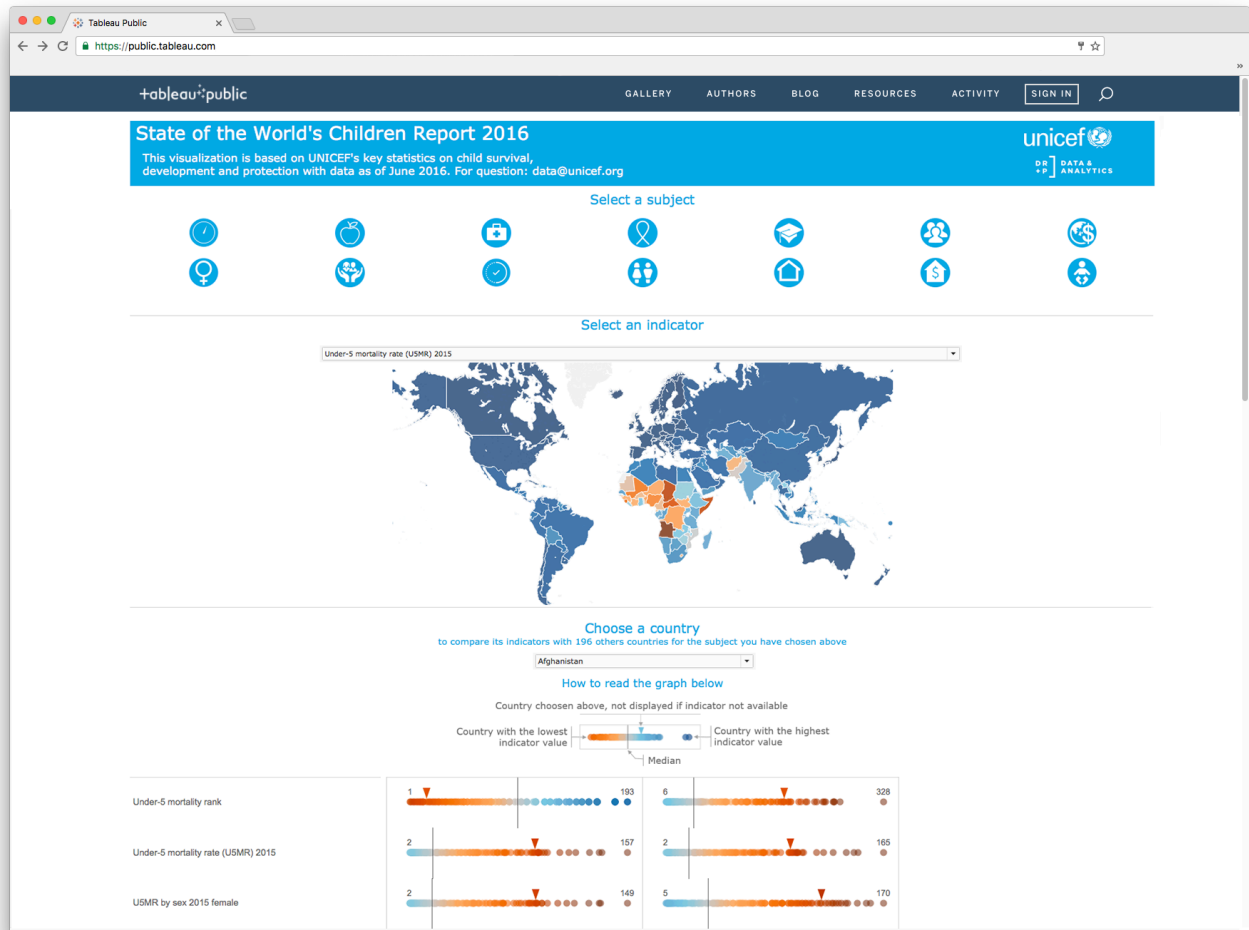


图 1：Tableau Public 库中展示了很多使用来宾权限访问的可视化。

除了少量配置以外，Tableau Public 所依托的底层体系结构与像您这样的组织在本地部署的 Tableau Server 完全相同。

而且，Tableau Public 可通过扩展支持超过 200,000 名作者、600,000 个可视化以及每周 700 万次查看。在发布过程中，在 Tableau Server 正式发布前的测试阶段，我们的整个 Tableau Public 部署实例一直都基于 Tableau 10 运行，真正做到了以极大的规模来测试平台的可靠性。

全面考量 Tableau 可扩展性的影响因素

随着您在自己的组织中逐步建立起分析文化，用户将越来越多地依靠数据洞见来作出更明智的业务决策。由于有了强大的技术和工具来探索值得信赖的数据，他们将开始洞悉到以前所不能看到的业务方面。随着用户认识到他们可以轻松解答自己的问题并进而收获到价值，他们将更多地采用分析技术，同时用户对数据新鲜度的期望也会上升。这些因素（用户对分析技术的采用率以及数据新鲜度）再加上您组织的风险承受能力，是您在设计部署策略和可扩展性计划时必须考虑的主要方面。

用户对可扩展性规划的影响

提升用户的采用率显然是实施自助式分析的一大目标。如果您能找到理想的平台并以智能且可扩展的方式加以部署，该平台就会变得对您用户的日常业务至关重要。正因为此，在制定计划时务须考虑使用模式和用户采用率。

自助服务对可扩展性的影响

定期分发报告是传统商业智能部署所采取的一种常见做法，这种做法主要侧重于报告的扩散和使用。但是，定期分发报告的做法在危急情况下不能为用户提供有力的支持；用户无法快速修改报告来提出他们自己的数据问题。实际情况反而是这样的：要想从不同的视角审视数据和提出新的问题，很可能需要回头求助于分析团队，等待他们的答复。整个过程可能耗时数天、数周甚至数月时间。

每一个建立了分析文化的组织都以自助服务为基础来整合自己的两项最为重要的资产：人员和数据。熟悉业务的人员必须能够深入剖析自己的数据，这样他们才能从各个角度对这些数据进行探索。借助 Tableau 的自助式分析平台，每个人都可以直接就他们所关注的的数据提出问题。利用 Tableau 的自助服务方法、经过实践检验的可视化分析引擎以及享有专利的 VizQL 技术，每个人都可以立即开始与数据交互和探索数据，不再受报告制作周期的束缚。

Tableau 方法



图 2：可视化的分析周期

让每个人都获得这种尽情查看数据的能力并有机会发掘潜藏的真知灼见，可以极大地调动他们的积极性。随着用户使用 Tableau 不断地发现和分享洞见，整个组织的采用率就会相应提升。因此，在为 Tableau Server 规划容量和资源时，必须考虑到用户采用率的影响。

用户采用率对可扩展性的影响

Tableau 可以进行极为严谨的分析，使用户能够就数据提出更多更有深度的问题。借助 Tableau Server，可以在一个安全可靠且受到妥善管控的环境中分享用户工作簿和真知灼见，从而使其他人可以从现有工作簿着手探索，快速开始自己的分析。创建和分享的工作簿越多，就有越多的内容可供更多人员用来快速开始自己的分析。Tableau Server 所促成的这种环境可以提高分析技术在整个组织范围内的参与度和使用率。

用户在 Tableau Server 上访问其分析结果和与之交互的频率是一项重要的规划指标。通常，大多数组织在安装 Tableau Server 后，其数据分析活动就会急剧增加。不过，您向用户提供的内容和数据越多，他们在决策过程中就会越依赖分析。估算使用频率是规划首次 Tableau 部署的一个环节。部署 Tableau Server 后，可以使用该平台随附的强大监控工具来测量频率和使用特征。通过审视使用数据，可以发现有用的信息，进而据以调整和扩展环境来满足不断变化的业务需求。

建立实践社区

很多客户为了在自己的组织内提倡和推广分析文化，纷纷创建了活跃的内部用户社区，并通过一些社区计划来宣传 Tableau 推动者、举办每周一次的午餐学习会、在入门计划中引入免费的 Tableau 培训，甚至与人才济济的 Tableau 专家社区建立联系。随着更多的人开始使用自助式分析，分享与协作成为让这种文化发扬光大的一个重要因素。这种文化的发扬光大继而又会影响用户的采用率和参与度，以及 Tableau Server 上运行的工作负载。为了确保提供最佳的用户体验和充足的容量来处理所增加的工作负载，您必须做好基础架构规划，以便为这种文化的发扬光大提供支持。

数据对可扩展性规划的影响

借助 Tableau，不论用户的数据是位于公共云、私有云中还是位于本地，抑或是在这三种位置均有分布，用户都可以连接到这些数据。利用 Tableau Server，您的用户可以实时连接到包含时效性数据的数据源。或者，对于不常变更的数据，他们也可以创建数据提取。这些数据提取可以作为已保存数据源直接存放在 Tableau Server 上。然后，您可以按所需的任意频率刷新这些数据提取，或者将它们设置为定期自动刷新。只需单击几下，数据维护者就可以在数据提取与实时连接之间切换。这样，他们就可以基于一部分离线数据进行原型设计，待重新上线后再重新连接到整个数据库。

借助 Tableau，用户可以采取最适合您业务的方式灵活优化对您数据的访问。要想使用户能够依据充分的信息作出业务决策，必须让他们在需要时能够访问他们所需的数据。通过让业务用户更接近他们的数据，不仅可以提升灵活性和敏捷性，也能在确保环境妥善扩展的过程中，让诸如数据新鲜度、数据规模和位置等因素成为举足轻重的考虑因素。

数据新鲜度

正确的业务决策都是依据新鲜数据作出的。不过在一些情况下，用户并不总是需要最新的数据。例如，每日一次的漏洞分类会议就属于这类情况。在这个示例情况下，数据只需每天刷新一次即可，因为漏洞分类会议是对过去 24 小时的数据进行回顾总结，而不是处理最新数据。

但支持队列就另当别论了，因为这种情况下您的目标可能是通过电话积极帮助和响应客户。故而，在整个工作日内数据可能每隔几分钟就需要刷新一次，而不是每 24 小时才刷新一次。

数据刷新可能会给您的基础架构带来负担。因此，在规划 Tableau 部署时，请根据数据本身以及业务需要，仔细考虑您的刷新需求，以免您的环境承受不必要的刷新负担。例如，不妨利用 Tableau Server 的内容分析功能，找出哪些工作簿根本没有用到但仍有数据提取在不断刷新。可以通过与 IT 部门和业务部门携手合作并不断地交流，根据内容分析结果和工作簿的使用情况，确定出最合适的数据刷新频率。这些简单的技巧可以帮助您提升向最终用户提供的服务整体质量。

数据规模和位置

如今，数据规模的增长速度更胜从前，与此同时，访问频率也在不断加快，数据源的类型更是越来越多样化。这就导致在内存中运行和分析所有的数据并非总是可行之举。虽然数据可能分散在本地、云端甚至是平面文件中的多个数据库中，但用户的需求是，不论数据位于何处，都能够查看它们、与它们互动以及将它们可视化。

在需要时，用户必须能够灵活使用在内存中进行分析的做法来实现高性能。他们还必须能够充分利用在数据库基础架构方面的现有投资来对实时数据库运行自己的分析。

过去，业务用户依靠具有 SQL 专业技能的人员，根据多个不同关系数据库管理系统和平面文件中的数据来生成报告。现在，借助 Tableau，不论数据规模或位置如何，业务用户都可以直接利用跨数据库联接和并集来充分发挥数据的作用，而且无需掌握专业的 SQL 技能。

对于频繁更新的数据库，用户可以建立实时连接来访问最新的数据。不过，如果担心这样会影响性能且相应的业务情形不需要实时访问数据，您可以轻松地提取这些数据以利用 Tableau 的在内存中运行技术。然后，便可以根据自己的业务需求调整该数据提取的刷新计划。

在您规划初始容量以及考虑整个组织的未来数据需求时，一个重要因素就是您选择以何种频率刷新数据以便作出准确的业务决策。

规划 Tableau 可扩展性时可以依据的框架

虽然需要依据众多可变因素来确定一项部署的可扩展性，但在您着手规划时，我们已经讨论过的可变因素才是需要估计的重要因素：

- 用户影响 – 自助服务的使用情况和用户采用率：将有多少用户使用分析技术？用户将多久使用一次分析技术来作出明智决策？用户所创建的可视化的复杂程度如何？

- 数据影响 – 新鲜度、规模和位置：您的数据规模有多大？这些数据位于何处？这些数据需要达到何种新鲜程度才能为业务决策提供准确的依据？

在规划 Tableau 部署时，应尝试量化这些矢量。本节将讨论如何结合这些因素来考虑 Tableau 的扩展事宜。

作出有效业务决策所需的分析技术使用频率	高 (每秒一次)	7.示例：全球数据探索，Tableau Public (美国总统大选)，每小时查看3万次	8.示例：销售配额仪表盘，Tableau 电视数据分析	9.示例：空中交通管制员，金融监控，交易执行
	中 (每小时一次)	4.示例：每日库存盘点，保险客户分析，营销(锁定目标)	5.示例：患者容量，经销商管理	6.示例：支持升级仪表盘，金融投资组合仪表盘，欺诈调查
	低 (每天一次)	1.示例：工程设计-产品发布讨论会，抵押贷款资产盘点，传统商业智能	2.示例：销售业绩大比拼，销售线索跟踪	3.示例：网络流量仪表盘
	低 (每天一次)	中 (每小时一次)	高 (每秒一次)	
作出有效业务决策所需的数据刷新频率				

图 3：规划可扩展性时可以依据的一个简单业务框架

了解 Tableau Server 进程以及这些进程支持各种服务器负载功能的方式，这对于调整和扩展 Tableau Server 部署至关重要。有关更多信息，请参阅 [Tableau Server 进程](#)。

在上表中，纵轴表示在 Tableau Server 中积极使用分析技术的频率。有少数与此功能关联的进程被认为是“面向用户”的进程。横轴表示为向组织提供最新数据来作出业务决策而需要达到的数据刷新频率。负责支持数据刷新的主进程是后台程序进程。要扩展 Tableau Server，第一步就是根据这些主要使用情形来增加进程实例数目。

首先，在上表中标绘出您组织的使用情况，这有助于确定从何处着手进行首次 Tableau 部署和进一步调整。Tableau Server 实际上既可以横向扩展，也可以纵向扩展。扩展部署的主要方式是根据相应的使用情况添加更多进程实例或额外的群集节点（工作节点）。

在确定性能和规模时都需要考虑的其他重要因素是：用户如何制作 Tableau 可视化以及其可视化的最终复杂程度如何。利用 Tableau Desktop 或利用 Tableau Server 上的 Web 制作功能，任何人都可以轻松制作 Tableau 可视化。但当他们创建的工作簿将得到大量使用时，他们就应该遵照一些侧重于提高性能、可扩展性和可靠性的额外最佳做法。有关更多信息，请参阅[设计高效工作簿](#)。

示例情形

以下情形通过一些示例说明了扩展 Tableau Server 的不同方式。这些情形依据的是对某个生产环境中的工作负载进行真实建模分析的结果，以及为获得可扩展性信息而在我们的可扩展性实验室运行和模拟的多项测试。

小规模单服务器部署

对于这种情形，如果您的风险承受水平允许您承受一定的停机时间，可以考虑采用一台配备 64 GB–128 GB RAM 的 8–16 核服务器。需确保有足够的磁盘 IO 带宽 (>500 MBPS) 可供 Tableau Server 使用。我们的[最低 RAM 和磁盘推荐配置](#)将能够充分满足安装和试部署的需要，但对于生产部署，我们通常建议为分配给 Tableau Server 的每个核心配备至少 8 GB RAM。

我们已经设计了默认的“开箱即用”型 Tableau Server 配置，这种配置可以针对规模较小的部署和部门级服务器进行优化。不过，在如何扩展来满足使用情形的需要方面，Tableau Server 也赋予了您灵活性。

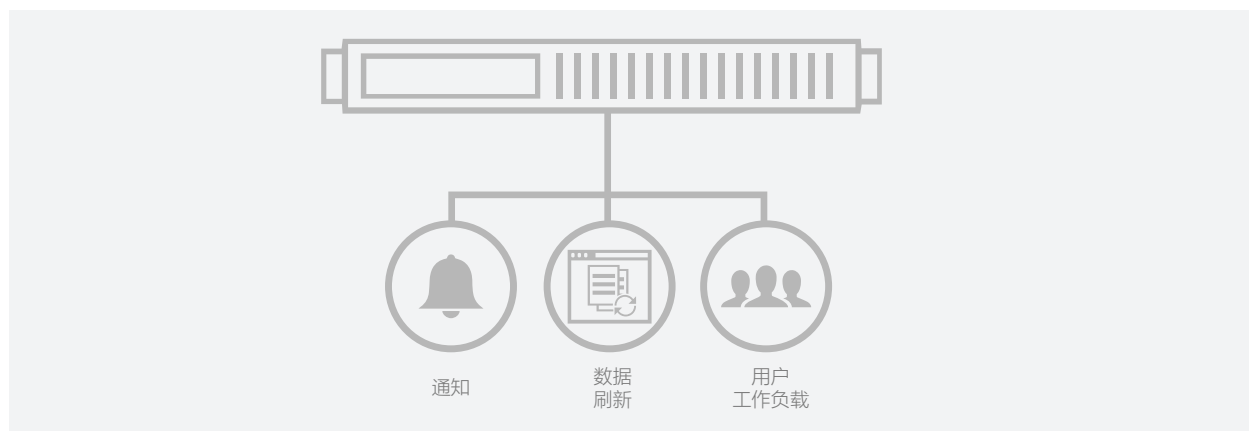


图 4：采用单台计算机时的部署图

单服务器部署易于设置、配置、维护和管理，适合只能获得最基本 IT 支持的较小团队。如需获得关于规划、安装和维护单服务器部署的端到端指导，请参阅 [Tableau Server：适用于每个人的安装指南](#)。

一般而言，对于数据刷新频率低且使用频率处于中低水平的小型组织，这种部署就足以满足其需要。您不应让这种部署以高频率刷新数据，因为负责数据刷新的进程和供分析技术使用的进程将会争用同一台计算机上的系统资源。根据我们的测量结果，当单计算机部署同时还承受数据刷新工作负载时，最终用户获得的服务质量下降了大约 10%。

务必要考虑到，当负载激增至超出您规划的使用水平时，单服务器部署所发挥的性能可能会打折扣。显然，只采用一台服务器时，万一出现故障，可用性也会受到影响。虽然 Tableau Server 无需任何配置就可以实现进程级高可用性（进程如果失败，会自动重新启动），但在这种部署情形下，万一硬件出现故障，就会导致停机。

双服务器部署

可以通过双服务器部署来减轻停机风险，同时还能提高性能。虽然双服务器部署不能提供与多服务器分布式群集相同的容错能力，但运行双服务器部署仍不失为一种经济实惠的方式：既可以保障规模，同时所具备的风险又低于单服务器部署。

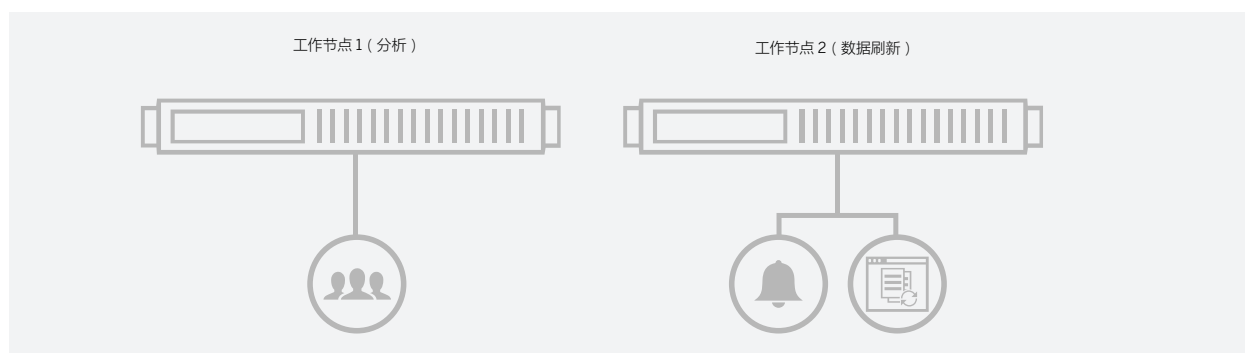


图 5：采用两台计算机时的部署图

得益于 Tableau Server 的灵活体系结构，您可以在多台服务器之间分配工作负载。可以将一台服务器用于满足最终用户分析工作负载的需要，将另一台服务器专门用于负责刷新数据和向最终用户发送通知的后台程序作业。与单服务器部署相比，投入了两台服务器的部署可以支持更频繁的数据刷新，同时又不会影响用户所获得的性能。

对于我们采用的工作负载，我们在测试过程中让 Tableau Server 持续承受可使 CPU 利用率达到 80% 的负载，结果发现，一个正确配置的双节点服务器群集可以支持 1,000 到 5,000 名用户（您测得的结果可能会与此不同）。

例如，投入了多达 16 个核心的 Tableau Server 部署可以支持一个这样的组织：拥有 1,000 名员工，大约 10% 的员工会持续活跃一个小时的时间（峰值使用时间），并且该组织的业务要求数据每 24 小时刷新一次。这种情况下，可以将其中的 12 个核心分配给分析技术使用，其余 4 个核心则用于数据工作负载和持续刷新数据。根据您的分配分析技术使用量和数据工作负载的方式，以及您在确保高性能工作簿设计时的着重点，您或许能够让此系统支持更多用户，或者可能需要再额外投入 8 个核心以用于运行非常复杂且运行缓慢的工作簿。

上述信息依据的是我们的测试结果，您所测得的结果可能会有所不同，具体取决于您的使用情形。您可能需要考虑进行一次体系结构评估，以便更好地了解您需要的规模。或者随着时间的推移，依据体系结构评估结果来确定规模需求。这种评估有助于确保您不仅仅是为某一个时间点进行规划，还会针对预计的 Tableau 使用量增长情况、业务需求以及风险缓解情况做好规划。

添加用途单一的群集节点

您还可以纵向扩展具体的 Tableau 功能组件，以便满足您的业务需求。为此，可以向群集添加用途单一的节点。

为保障新鲜数据提供支持

就 Tableau Server 而言，数据新鲜度取决于您管理数据提取刷新的方式。具体来说，一定要保障有充足的容量，从而确保数据提取及时刷新，以供业务用户使用。

对于这种情况，最重要的规模单位就是后台程序进程。添加后台程序进程可以加快数据提取刷新速度，并确保对您可视化的订阅能及时推送到用户，当然还会带来其他方面的改进。

对于要求获得实时数据或频繁执行数据提取的组织，一种常见的部署配置就是向 Tableau Server 群集添加专门用于运行后台程序进程的节点。运行此后台程序可以分担其他节点的 CPU 和处理负担，从而释放出更多周期来满足前端请求，例如可视化和制作任务。

对任务至关重要且需满足高可用性要求的部署

如果您的业务需求要求优先保证正常运行时间（可用性），则需要部署至少三个节点（从 Tableau 10 开始提出此项要求）才能确保您能够抵消硬件和软件出现故障的风险。高可用性的工作原理不在本文档的讨论范围内。您可以参阅下面这篇技术白皮书：[高可用性白皮书](#)。

投入多台服务器、需要频繁进行数据提取的部署

如果您的业务对停机风险的承受能力很低，不妨考虑部署一个包含多台服务器的分布式群集来为部署提供冗余能力。要提高可用性和规模，您可以随着时间的推移，根据需要向该群集添加更多服务器。Tableau Server 可以配置为随着您向群集添加服务器节点而相应地以线性方式进行扩展。您所能支持的用户数目将因您的工作负载和具体情况而异。不过，我们在测试过程中发现，部署在一个 32 核服务器群集上的 Tableau Server 能够支持超过 5,000 名用户，且其中可以有 414 名用户同时处于活跃状态。每个工作节点各配备 8 个核心，主节点上则仅部署存储库。

如果一个组织有 3,000 名员工，大约 5% 的员工会持续活跃一小时的时间（峰值使用时间），并且该组织的业务要求数据每 8 小时刷新一次（而不是像上述情形那样每 24 小时刷新一次），那么，如果采用低端配置，投入 28 个核心来部署 Tableau Server 就可以提供设计完善、可以快速执行的工作簿，从而满足该组织的需要。将其中的 16 个核心用于满足分析使用情形的需要便已足够，将其余的 12 个核心用于数据刷新工作负载（后台程序）则可以有助承担这些工作负载。如果采用高端配置，这可能需要投入 40 个核心来同时支持一些加载速度非常快的工作负载（不到 4 秒）和一些加载速度相对较慢的工作负载（考虑到仪表板的复杂性，加载时间为 4-10 秒）。

当您进行纵向扩展来支持更多用户时，相应的体系结构会向您现有的部署群集添加更多 Tableau Server，轻松支持这种扩展。例如，如果一个组织有 5,000 名员工，大约 5% 的员工会同时处于活跃状态，且该组织的业务目标是不到 24 个小时就要刷新一次数据，那么，投入最多 44 个核心供分析服务使用，再投入 12 个核心供数据刷新服务（后台程序）使用，总共投入 56 个核心便可满足该组织的需要。

尽管向群集添加节点和调整进程实例来响应使用情况是很简单的管理任务，但复杂的部署也需要专业 IT 人员的参与。Tableau Server 需要与数据源进行通信，而数据源可能位于网络内部或位于云端。用户可能从互联网上或者从组织内部的各种平台上通过 Tableau Server 连接到这些数据源。用户身份验证机制可能通过 SAML 与 Active Directory 或任何第三方企业身份管理套件相集成。因此，在规划复杂的 Tableau Server 部署时，应考虑 IT 部门所有资深人士的意见。

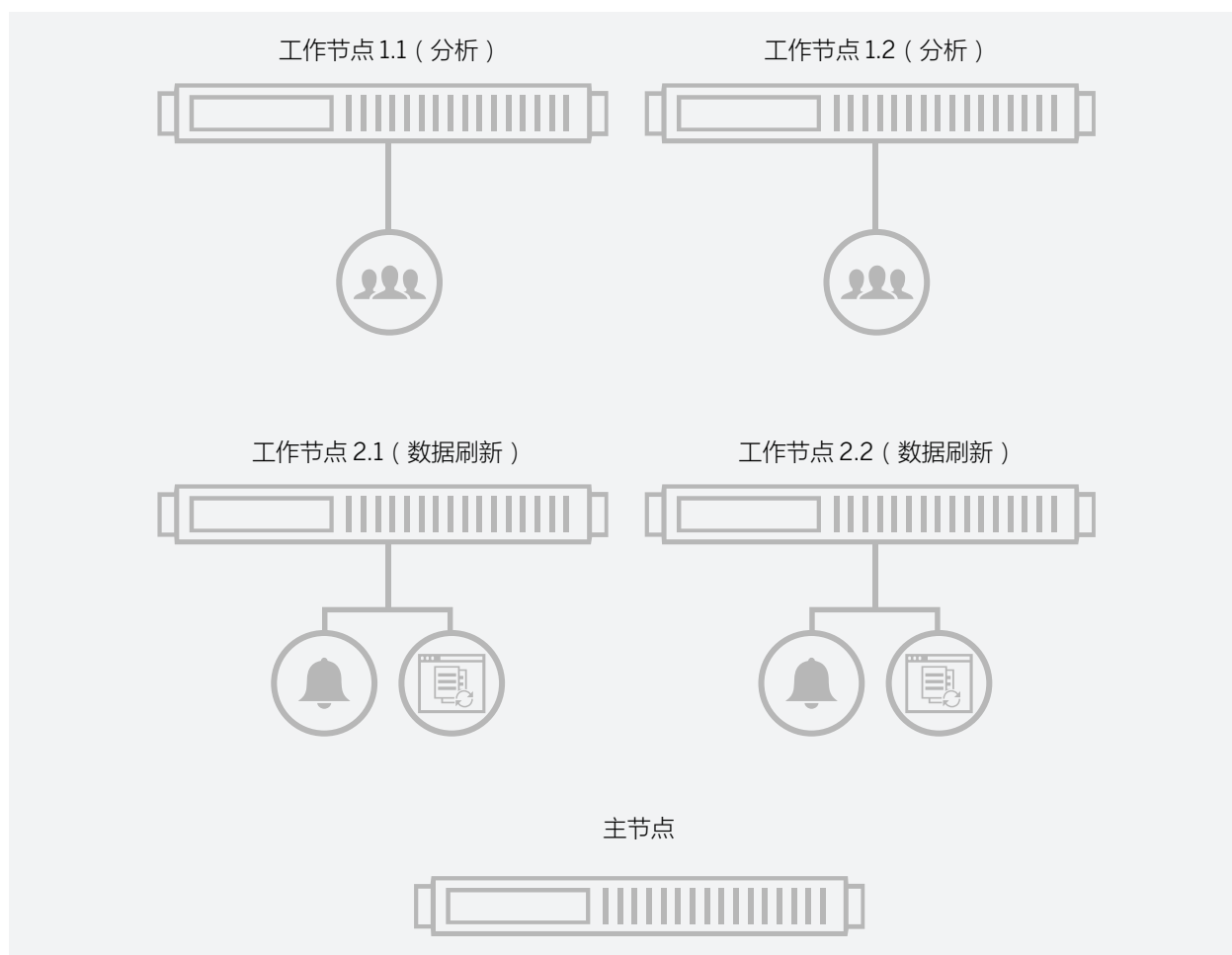


图 5：采用多台计算机时的部署图

下面是一个体系结构示意图；在这种体系结构中，多台运行 Tableau Server 的服务器可以同时满足企业内部用户和外部用户的需要。

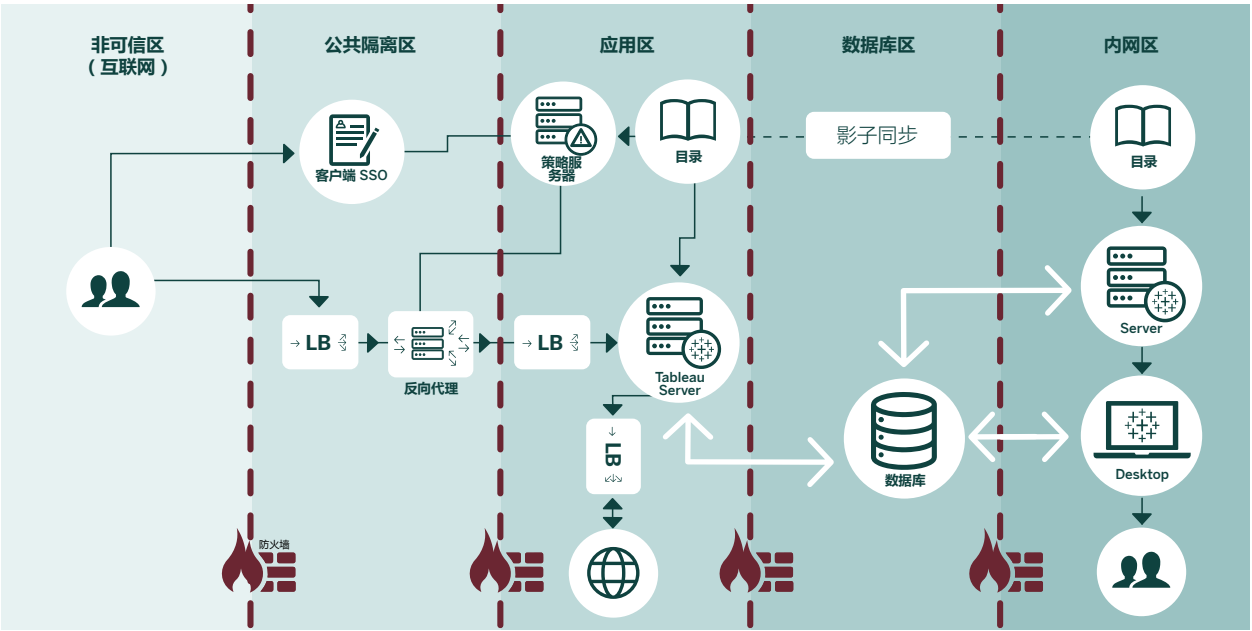


图 6：企业级部署概图

结果摘要

Tableau Server 是一款企业级分析平台，可以为整个企业提供分析服务。我们通过夜间进行多版本对比测试，并在我们的可扩展性实验室内进行测试，观察到 Tableau Server 只需横向扩展服务器群集，即可支持越来越多的用户使用相应的系统。

虽然在给予任何指导或确定规模时需要考虑到很多技术假设（如需了解更多详情，请参阅[与本白皮书配套的这篇技术白皮书](#)），但我们通过采用一个基本假设：即总用户中大约有 10% 通常在服务器上处于活跃状态（实际数字可能不尽相同），以此观察到在组织中采用 8 核单服务器部署时，Tableau Server 总共可以支持多达 1,000 名用户，在采用 32 核多节点部署时则可以支持总共约 4,480 名用户（如下图所示）。

为了支持更多负载，Tableau 体系结构提供了灵活性，允许隔离工作负载并/或采用高度可用的拓扑运行群集，以便降低业务风险。下面的可视化中的第二列就有一个工作负载隔离示例，具体是将存储库隔离起来，只允许其在自己的硬件上运行。虽然这只是可选做法，但在一些情况下，您可能会发现通过将各种 Tableau 工作负载隔离起来，使它们只能在自己的硬件上运行可以改进可扩展性。

使用 Tableau Server 10 时用户总数可以线性扩展

您获得的结果可能与此不同。除了需要考虑其他变量之外，模型还假设总用户中有 10% 处于活跃状态

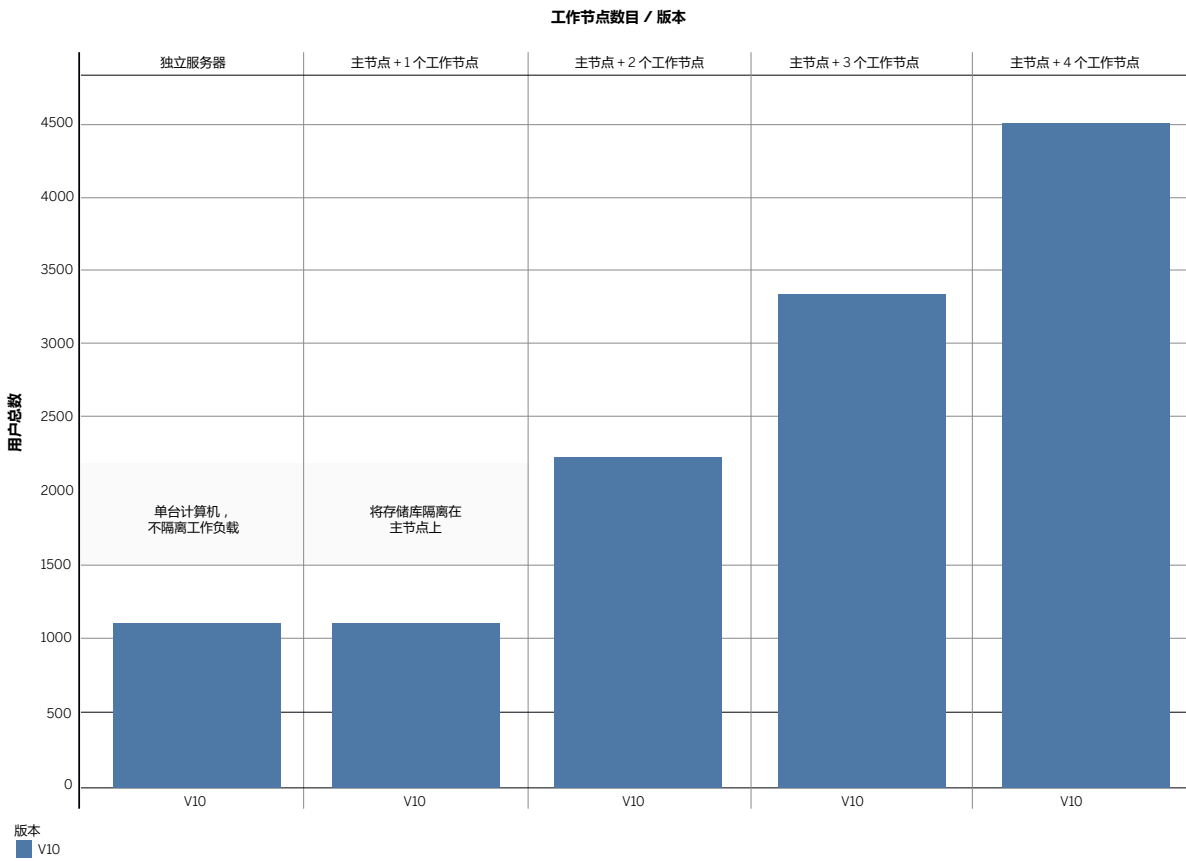


图 7：Tableau Server 的可扩展性（以用户总数衡量）：随着工作节点的增加以线性方式进行扩展。

总结

只有当自助服务和管控能力可以在整个组织中以安全可靠的方式进行扩展时，才能形成真正的分析文化。Tableau Server 具备随您的业务需求进行扩展的灵活性。不过，务必要综合考量用户采用率、数据位置、数据新鲜度以及对停机的容忍程度，从而正确评估业务需求，这样才能建立您的业务所能够依赖的环境。

在本白皮书中，我们剖析了对 Tableau Server 的基础架构可扩展性至关重要的各类因素，并分享了根据我们的内部测试得出的结果。企业部署在规模和复杂性上往往不尽相同。Tableau Server 可以满足在基础架构可扩展性、可用性、安全性、易管理性、监控以及与现有 IT 平台集成方面的企业需求，还能达到严格的产品部署要求。

关于 Tableau

Tableau 帮助人们将数据转化为可以付诸行动，发挥重大作用的见解。轻松连接到以任何形式存储在任意地点的数据。快速执行临时分析，发现隐藏的见解。通过拖放操作，创建包含高级可视化分析的交互式仪表板。然后在整个组织共享，让其他团队成员能够从自己的数据视角进行探索。从全球性企业到早期初创企业和小企业，使用 Tableau 的分析平台来查看和理解数据的人无处不在。

资源

[Tableau Server Scalability: A Technical Deployment Guide for Server Administrators](#) (Tableau Server 的可扩展性：面向服务器管理员的技术部署指南)

[Tableau for the Enterprise: An IT overview](#) (适用于企业的 Tableau：IT 概述)

[如何建立分析文化](#)

[基于 Amazon Web Services 的 Tableau](#)

