



观数台架构和扩展性 白皮书

目录

一、平台介绍	1
二、关联与内存应用程序	2
三、用户界面	4
四、观数台架构概览	6
五、多节点架构	8
六、多云端部署	10
七、观数台 API及 开发库	12
八、扩展性概览	13
九、扩展性架构	15
十、应用程序和数据的扩展	18
十一、用户的扩展	22
十二、扩展交付	24
十三、部分特性	25
十四、总结	27

一. 平台介绍

观数台通过自己的架构提供一致且可预测的性能，该架构使您的组织可以灵活地部署到您需要的位置，并可以横向和纵向扩展并与数据，应用程序和用户线性扩展。使用观数台，您的IT专业人员可以有效地维护终端用户的性能，同时对计划未来的，预期的系统使用的容量充满信心。

观数台产品家族：

观数台+： 业内领先的数据分析平台

观数台EDC：手工数据管理专家

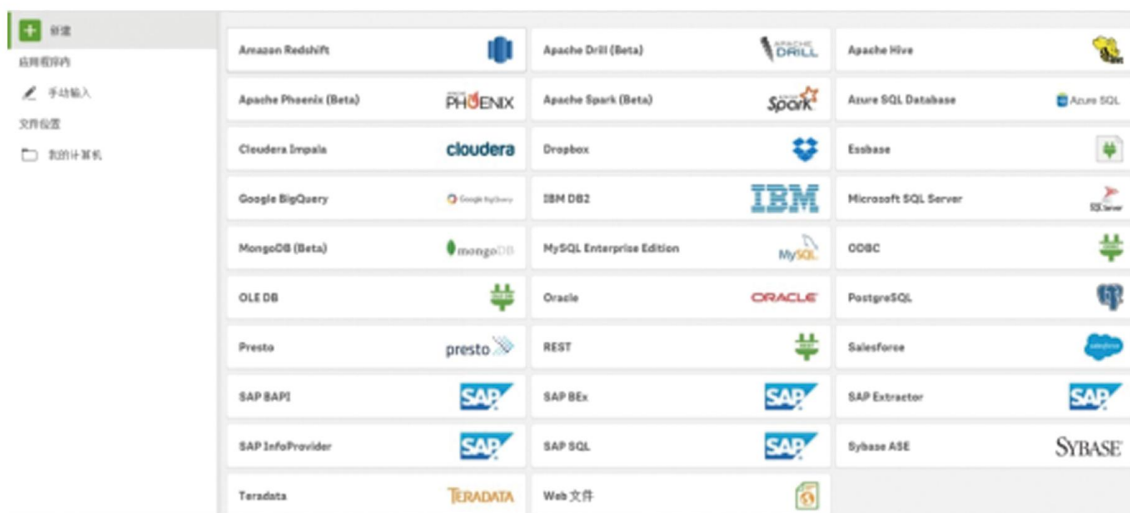
观数台Portal：跨平台多应用的管理中心

观数台平台产品架构：



二、关联，内存应用程序

观数台将内存数据存储技术与关联引擎相结合，可直观，自由地分析和探索数据。在第二代中，经过验证的观数台数据索引引擎使用户轻松地探索数据并根据来自多个数据源的数据创建可视化。这些数据源涉及从Excel 和 Access 到像Oracle、SQL Server的数据库，再到像Cloudera和Redshift的大数据源。涵盖了目前常见的、以及代表国际最新技术的数据存储系统，并紧跟新型技术的发展跟踪开发数据接口。除此之外，观数台额外提供了名为“自定义数据源”（Custom Datasource）的扩展功能，允许用户针对特殊的数据源系统开发相应的用于读取特殊格式数据的驱动程序。



观数台使用列式内存存储。唯一条目只在内存中存储一次，并且数据元素之间的关系是以指针形式表示的。这可以实现显著的数据压缩，在内存中容纳更多的数据，给予用户更快的响应时间。

在一些大数据场景中，数据应当保留在源里，这就是观数台使用按需应用程序生成的内置技术的原因。可以根据用户的选择查询数据源，但仍然可以为用户提供关联体验。观数台大数据关联索引专为进一步扩展此功能而构建，适用于您拥有的最大数据源，提供高性能，分布式并行版本的观数台关联引擎，可以直接利用分布

式数据源的性能优势。

什么是关联性？ 关联是指内存数据存储和获得专利的数据索引引擎的独特组合。关联让您的用户在有连续上下文的数据基础上查询和导航并且创建可视化内容。关联消除了传统的分层或基于查询的方法所受到的约束。

三. 用户界面

通过浏览器进入观数台。观数台网页客户端为开发、拖放内容创建和数据分析的所有方面提供可能。观数台采用响应式设计方法，通过适当的布局和信息自动显示和调整可视化，以自适应设备——无论是笔记本电脑还是台式机、平板电脑或智能手机上的浏览器。观数台采用当前的HTML5、CSS3、JavaScript®和Web套接字标准构建，使您可以在任何平台上构建并部署。

观数台具有灵活的交付方式，具体取决于情况和受众。这确保了一致且交互式的用户体验。



1. Hub

Hub是一个可让您的观数台用户访问他们有权访问的应用程序的分析门户。



2. 微信/钉钉集成

观数台可与微信/钉钉用户与软件的权限系统相结合，可以有效的进行用户的管理。软件用户只能在应用列表中看到自己有权访问的应用，同时针对拥有企业数据应用访问权限的用户，可以为其配置单点登录到该数据应用的能力，这样该用户可以实现微信到数据应用的无缝集成跳转。

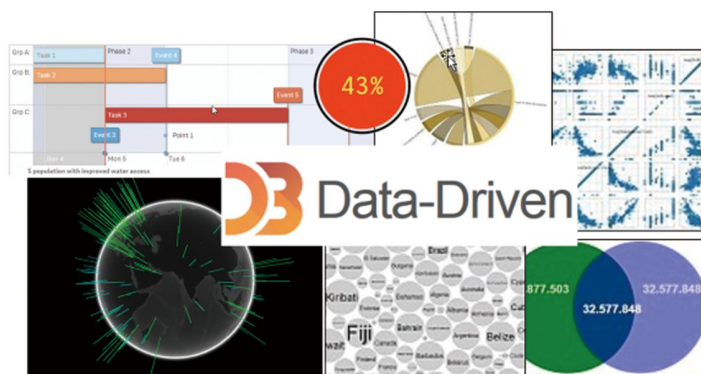
3. 管理控制中心

允许您的管理员管理观数台平台的各个方面 - 从数据连接到应用程序和任务管理，再到安全管理，监控和审计。

用户名	姓名	邮箱	电话	用户组配置	角色配置	操作
Admin	Admin	admin@gmail.com	15639253520	管理组		编辑 删除
Emma	Emma	Emma@gmail.com	15639253524	管理组		编辑 删除
Mat111	Mat111	Mat111@gmail.com	15639253525	管理组		编辑 删除
test	test	test@qq.com	13233333333		test角色	编辑 删除

4. 嵌入式分析

观数台更提供了开放的API接口，可以通过添加自己的可视化和数据源来扩展观数台。您可以将观数台可视化和数据与其他基于Web的应用程序集成。观数台具有内置的基于Web的编辑器，可用于处理这些API或使用您自己的集成开发环境（IDE）。让开发者能够构建定制引导分析应用程序、网页中的嵌入式分析和工作流应用程序，创建新型的可视化效果（D3图表）和分析，以满足定制的业务需求。



四. 观数台架构概览

观数台旨在以可扩展、安全和可管理的方式支持可视化分析。您可以在单个服务器上部署观数台，并纵向和横向扩展，以满足部署的可用性和处理要求，无论是本地部署还是云部署。部署中的每个节点都在观数台站点中执行部分或全部服务执行特定的角色：



1. 观数台代理服务

分别是通过Hub和管理控制台为您的用户和管理员提供的入口点。观数台代理服务与许多用户容器服务（例如Active Directory，SAML等）集成。观数台代理服务还管理您的会话，提供许可证，并处理与其他组件的负载平衡。

2. 观数台引擎

关联的内存数据索引引擎。观数台引擎是一个基于内存的应用层，可在运行时提供高度交互的自助式可视化，搜索和计算。

3. 观数台调度器

协调数据加载。观数台调度器支持基于时间的应用程序重新加载、依赖性，并且可以由企业调度程序（例如，CA AutoSys）。

4. 观数台存储库

是平台的配置和管理信息的中心化存储。观数台存储库管理您的用户定义，安全性以及平台的许多其他元素。数据仍然存在于PostgreSQL。

5. 观数台应用程序

包括高度压缩的数据、数据模型和展示层。您的应用程序将持久存储在文件系统中，并在用户请求时由应用程序加载到内存中。

6. 多云服务

是用于启用多云功能的可选服务

7. 辅助服务

包括用于内容传输的印刷服务和调用数据准备，应用程序迁移和按需应用程序生成的例程的观数台服务调度程序。

五、多节点架构

观数台采用模块化架构，旨在满足您组织的独特需求。虽然单个观数台服务器通常可以支持数百个并发用户，但该平台的多节点架构可以进一步扩展。

1. 分布式工作负载

您的组织可以在服务器上以不同的组合部署观数台组件以完成特定功能。该表描述了如何部署观数台的组件以允许服务器作为客户节点，开发者节点，调度器节点和中央节点运行。非中心节点一般称为边缘节点。

功能	描述	组件服务						注释
		代理服务	引擎	调度器	存储服务	存储数据库	应用程序文件系统	
客户节点	用户的消费和自助服务分析	X	X		X			代理可以单独作为它自己的服务器
开发者节点	为用户创建和重新加载新应用程序	X	X		X			对于业务用户，建议面向IT的作者使用隔离的非生产环境
调度器节点	应用程序和数据重新加载		X	X	X			引擎执行调度程序调用的重新加载
中央节点	主调度程序、应用程序升级以及混合	X	X	X	X	X	X	可以与其他节点共定位

2. 中央节点

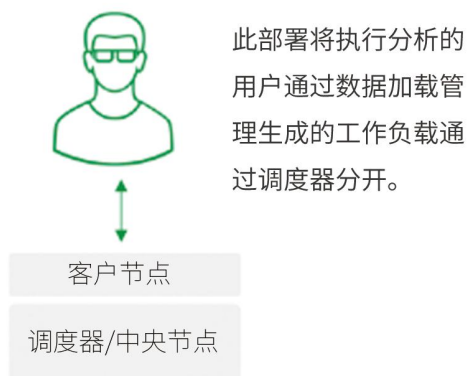
每个观数台环境都有一个中央节点，作为主调度程序。中央节点协调应用程序重新加载，在软件更新时执行应用程序升级，并将扩展对象提供给其他节点。默认情况下，中央节点还包含存储库数据库和应用程序文件系统。在多节点环境中，每个存储库服务与中央节点上的存储库数据库通信，并且每个引擎从中央节点上的应用程序文件系统加载应用程序（.QVF）。可以将存储库数据库和/或应用程序文件系统移动到专用的弹性资源。

最后，由于其有限的责任，中央节点可以共同定位在另一个节点上，最常见的是调度器。中央节点功能可以被指定在一个或多个节点上运行，并且在出现故障的情况下，其中一个备用节点将承担中央节点职责。

单节点部署



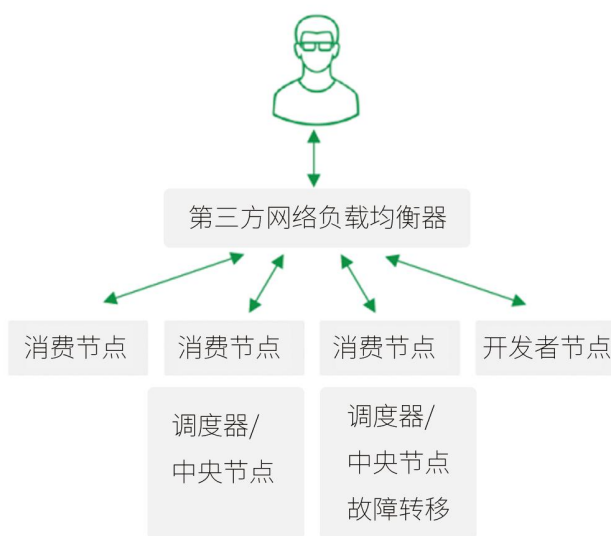
双节点部署



3. 水平缩放-多节点

使用观数台，您可以纵向和横向扩展工作负载。在单个站点（环境）中，您可以部署多个用户节点和控制节点用于自助服务分析以及用于数据加载的调度器节点，以使观数台能够横向扩展以实现弹性和不断增长的工作负载。

关于工作加载？安全规则是可配置的，并且还使您能够将工作负载专用于特定节点。例如，在用户和控制节点中，按业务单位将消费或分段应用程序中的内容创建隔离开。在调度器节点内，按复杂性或持续时间加载段数据。



六、多云端部署

观数台中的可选多云功能能让您的组织广泛扩展策略驱动的部署，将分析范围扩展到新用户，同时保留分析消费使用的灵活性。使用多云功能将观数台Windows部署的分析应用程序分发到您自己部署的弹性部署环境以及我们的全托管服务。

1. 多云端架构概览

平台

底层架构是基于微服务的，用于客户管理的观数台的弹性部署和托管的观数台云服务。可以部署在Linux®上，也可以通过Docker和Kubernetes进行部署。我们的云原生架构是弹性的，与我们提供和管理观数台云服务的方式相同，观数台用于弹性部署，可以部署到云供应商或客户管理的Kubernetes环境的托管容器服务中——使您能够从巨大的规模和无限的潜力中受益。

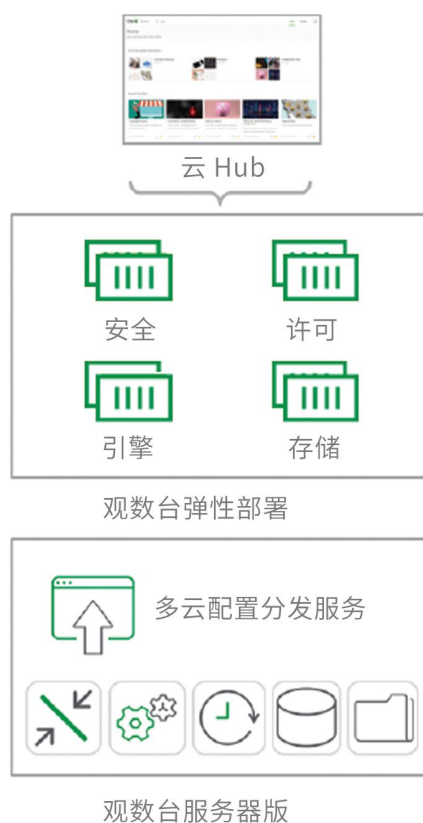
观数台遵循云原生设计的良好原则，我们的多云功能可以使用标准工具如ElasticSearch，Logstash和Kibana（ELK stack）进行记录，然后使用Prometheus，Grafana等标准工具进行呈现。



2. 组件

为了简洁，微服务根据它的功能进行了分类：

- 云 Hub 是我们基于浏览器的分析门户，用于服务消费者在观数台使用中的弹性部署。
- 安全中心管理身份验证和权限，以确保您的用户得身份认证，身份验证并且只能查看他们有权限使用的应用。
- 许可中心提供云 Hub和存储用户信息、偏好和许可证。
- 引擎中心提供观数台内存，关联数据索引引擎并且支持容器协助负载平衡和会话管理这与观数台服务器版中的引擎相同。
- 存储中心是基本的网络和数据存储容器，包括NGINXIngress，Traefik，MongoDB和Redis。
- 多云配置分发服务是额外的被激活的Windows服务。它们用于配置连接并将应用，权限，许可和其他资源分分配多云环境。



七. 观数台 APIs 和开发库

企业架构正变得越来越复杂。观数台API和开发库确保您可以将观数台完美集成到您的环境中，以满足组织所需的安全性、自动化和交付功能。

观数台 API是基于REST和JSON的;用于开发观数台的API相同。利用这种强大的能力在观数台平台上构建自己的应用程序，通过添加自己的可视化和数据源来扩展观数台，利用现有的安全供应商，甚至将观数台与基于.NET框架开发的Web应用程序或桌面应用程序集成在一起。您可以以编程方式管理许可，用户和任务。

观数台可视化 API：创建新的可视化

观数台嵌入式 API/.NET SDK：嵌入式可视化和功能

观数台QVX SDK：创建自定义数据连接器

观数台存储服务 API：管理和执行

观数台代理服务 API：身份验证/负载平衡/会话管理

观数台用户认证 API：创建自定义用户目录连接器

观数台引擎 /后端 API：与关联引擎直接通信

Picasso.js：用于构建观数台分析平台可视化的图表库

Enigma.js：用于与观数台关联引擎交互的库

Haylard.js：用于生成加载脚本的库

Leonardo-UI.js：具有观数台外观的Web UI组件库

八. 扩展性概览

扩展很重要，因为将系统部署到包含更大量数据和更多应用程序的大用户量的需求越来越多。扩展性是根据多个维度定义和测量的，并分为五个主要区域：体系架构、数据、应用程序、用户和交付。

下面来看观数台的扩展如何通过这五个类别展现的。[注意：观数台关联内存引擎的可扩展性特征无论是部署在观数台还是观数台进行云部署，都保持不变] 两个部署模型都包含在亮点中，在适当地方做了不同标记。

1. 体系架构

观数台由五个主要服务组成，可以纵向和横向扩展以满足您的需求。这种简单的架构可确保性能和扩展的可预测性。

弹性部署的观数台可以部署在Linux上，并通过Docker和Kubernetes实现。这种基于云原生的基于微服务的架构为观数台部署提供动态扩展。

2. 数据

观数台的性能与数据量呈线性关系。随着数据量的增长，分析以及多数据源的ETL性能都是可预测的。

3. 应用程序

观数台可预测资源的使用如何随应用程序扩展。每个应用程序的资源需求可以被添加，用于了解一组应用程序的要求。

4. 用户

观数台与用户数量呈线性关系。随着您的内容创建者和消费者数量的增长，响应时间和对服务器资源的影响会以可预测和线性的方式增加。

5. 交付

观数台让您的组织与最适合任务的员工保持一致。随着IT高效的管理硬件，保护和治理数据，您的用户可以创建内容，自助服务和协作。用更少的资源赋予更多人权限。

观数台 扩展性？

系统架构：性能和容量与处理器内核呈线性关系。

数据：CPU，内存和响应时间与数据成线性比例。

应用：性能是附加的，确保可预测性。

用户：CPU，内存和响应时间与用户呈线性关系。

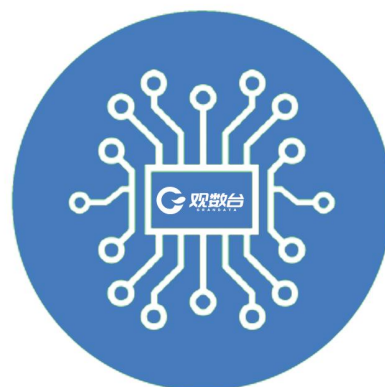
交付：更少的管理资源和自助服务内容创建，使观数台可以与您的组织进行扩展。

九. 扩展性架构

1. 观数台关联引擎

独特的关联引擎将内存数据存储技术与获得专利的关联引擎结合在一起，可让您直观地分析和自由探索数据。这个基于内存的应用程序层使我们在运行时，实现高度交互的自助式可视化，搜索和计算。

观数台会对每次点击进行计算。例如当您的用户界面需要通过聚合来显示图表对象或需要对做出的选择重新计算时，聚合会同时完成。观数台关联引擎是一个64位进程、多线程并可充分利用所有可用的处理器内核。在为不断增长的数据量和用户量执行计算时，它会线性利用可用内核。



2. 内存和缓存管理

观数台有效地使用内存，并以这三种方式存储

- 应用程序数据模型定义的数据集
- 引擎计算和用户选择的缓存结果
- 每个用户的会话状态*

观数台应用程序会保留在服务器上的磁盘。当用户请求应用程序时，如果尚未由先前用户加载，则将其加载到内存中。当多个用户在内存中使用此应用程序；它会被加载一次，而不是为访问它的多个用户重复加载。当应用程序不再使用时，内存将会复原。

当用户在应用程序中进行选择时，观数台关联引擎会实时执行计算，可以临时分配内存以执行计算。观数台会将结果存储在结果缓存中以供重用，以备任何用户进行相同的选择（前提是安全性允许相同的数据可见性）。我们的简单缓存机制可以让观数台自动管理CPU执行额外的重复工作，而无需管理。

注意：每一个用户会占用一小部分内存来保持会话状态信息。

内存管理：

应用程序

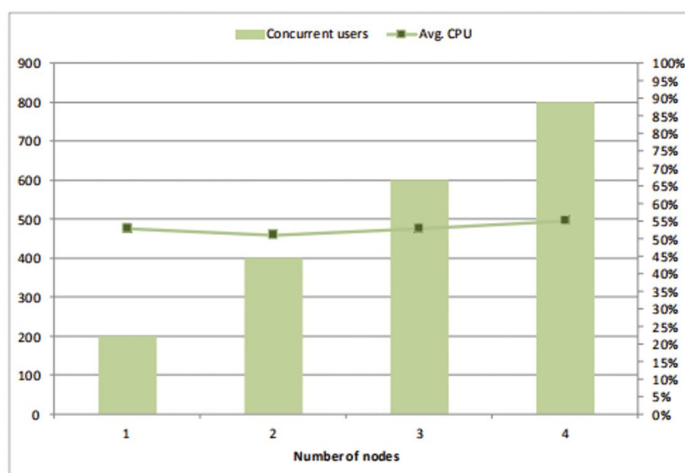
高度压缩数据；列式存储；内存中应用程序数据的单份拷贝

结果缓存

存储计算结果；自动重复使用；自我管理；与应用程序数据始终保持一致。

3. 分布式Windows架构

您可以将观数台部署到多节点上。它支持物理硬件、虚拟平台以及云环境。基于它的多节点结构，观数台随着额外引擎节点的部署线性扩展。



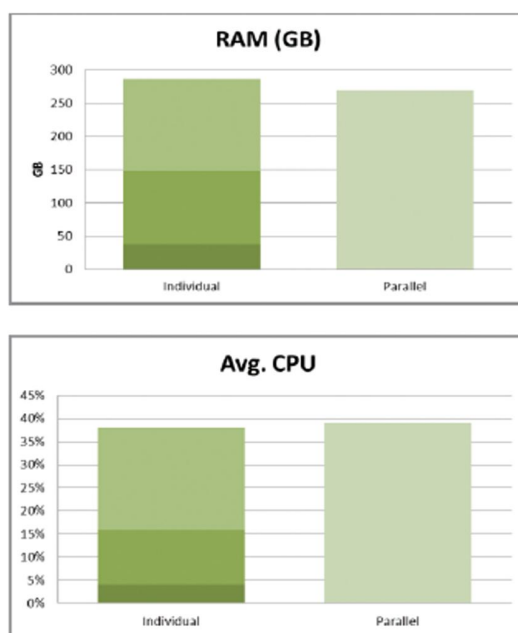
4. 多云端架构

您可以将观数台弹性部署到观数台云服务中，私有Kubernetes环境或云供应商提供的托管Kubernetes服务。所有容器都可以水平和垂直扩展，并且Engine性能随着分配给它的资源和/或容器数量线性扩展。

在任一部署模型中，观数台都实现了可扩展性，因为它是一种内存分析平台，可以克服依赖于底层数据库执行处理的基于查询的工具中普遍存在的瓶颈。如果通过添加节点或增加容器数量来扩展部署并使处理容量加倍，则可以处理两倍的用户数（假设典型用户产生的负载量相似）。此外，您还可以通过定义将哪些应用程序驻留在节点上，获得更多好处。当然，许多因素都有助于部署的扩展性和整体设计。

十. 应用程序和数据的拓展

观数台部署通常可以满足许多业务部门中用户的需求。这些增长都需要可扩展的数据和应用程序管理解决方案。观数台可预测添加应用程序或数据对资源消耗的影响。单个应用程序的资源特征（CPU和内存消耗）在部署时简单相加，能让您在发布到生产环境之前估计总需求。这些图说明了如何对资源利用总和进行预测。例如，如果您在三个不同的服务器上使用三个不同的应用程序，它们总共的内存和CPU消耗和他们在一个服务器上并行服务的消耗相同。这使您在部署时进行容量规划和预测时更有信心。

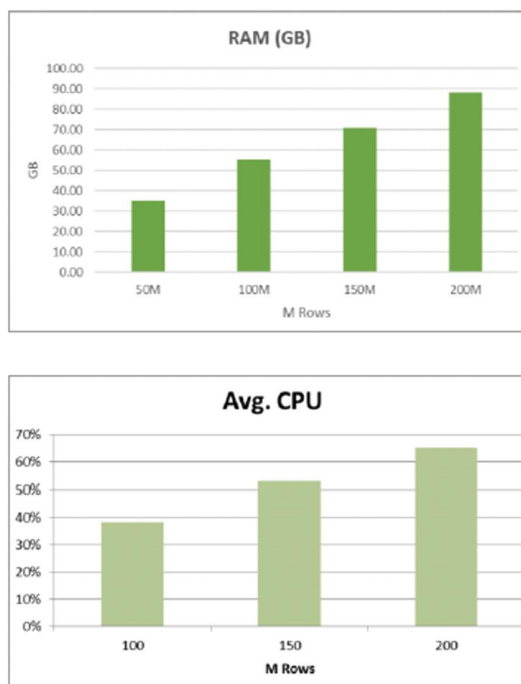


1. 应用程序性能

随着应用程序源数据量的增长，需要一定比例的内存来保存它。然而，两倍的源数据不需要两倍的内存。观数台使用高效算法存储每个唯一条目一次，然后使用指针引用数据之间的关系。当您添加更多数据时，您可以确保它尽可能高效地在内存中呈现。根据源数据的性质，内存中的压缩结果通常可达10倍。

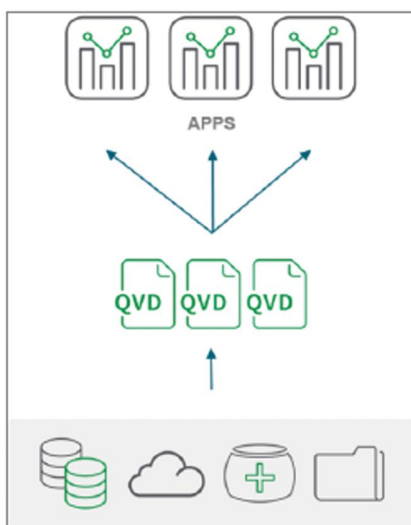
随着应用程序源数据量的增长，也需要一定比例的CPU周期来执行相同的计算。如果结果尚未存在于结果缓存中，则只要请求新的聚合就会使用CPU周期，并且响应终端用户请求所需的时间量是关于CPU执行重新计算数据的性能的函数。因此，当您向应用程序添加更多数据时，CPU将需要对更大的数据片段执行重新计算，从而导致需要更多的CPU周期。幸运的是，出于与上述相同的原因，观数台压缩确保两倍的数据并不需要两倍的CPU能力，如图所示。

观数台性能扩展与数据一致。当您向观数台应用程序添加更多数据时，相应添加的内存和CPU容量的组合，可以以可预测的方式维护终端用户的性能。CPU和内存的线性扩展可确保您获得可预测的性能和容量规划。



2. 数据层

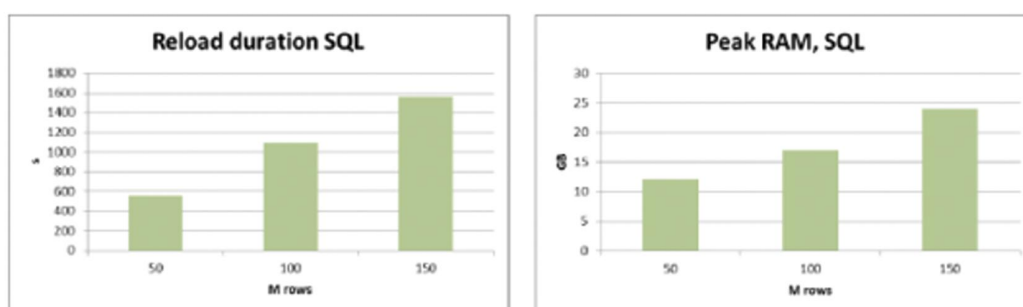
观数台允许您使用由称为.QVD文件的高度压缩文件组成的可选数据层。此中间数据层的一个好处是大型数据集的可重用性和增量负载。多个观数台应用程序可以在.QVD数据层中使用相同的数据，而无需单独查询远程数据存储，此过程以及观数台的ETL功能由观数台调度器组件自动执行。

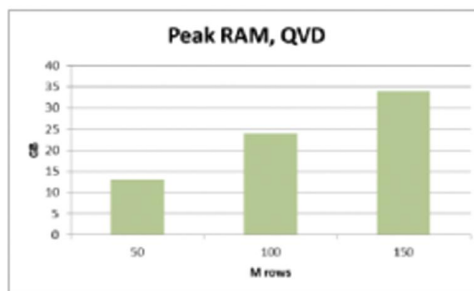
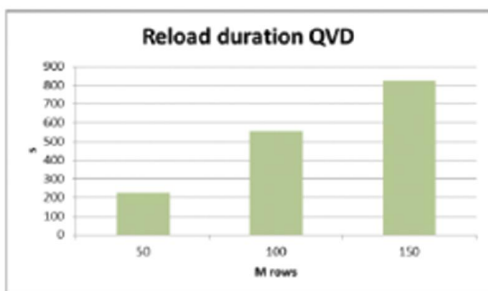


3. 源数据加载性能

观数台可以连接到几乎所有源来加载数据。重新加载持续时间取决于几个变量，包括CPU时钟频率、数据源、网络速度、磁盘速度、聚合类型以及加载脚本中的计算。然而，随着数据量的增长，重新加载数据的观数台扩展特性在几个维度上呈现线性趋势。这使重载持续时间的呈现可预测的。这些图示出了从数据库加载不同数据量的测量结果。持续时间和内存利用率均线性扩展。

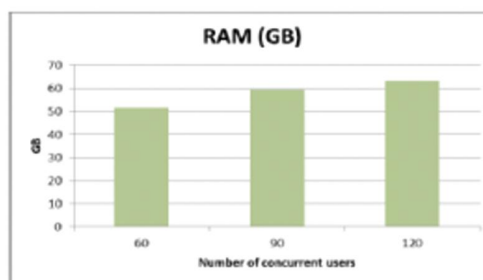
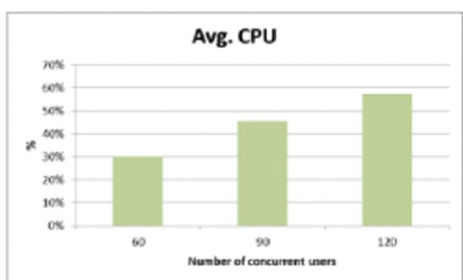
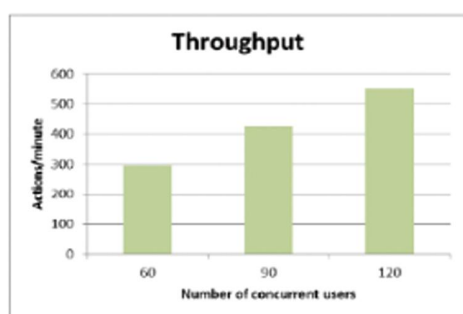
从观数台 .QVD数据层加载数据的时间明显短于直接查询源。从.QVD数据层加载数据也会随着持续时间和内存利用率线性扩展。





十一. 用户的扩展

观数台的性能与用户统一。当您向观数台添加新用户时，性能影响是可预测的，并按比例添加更多CPU和内存资源。下图显示，随着并发用户数量的增加，CPU和内存消耗量也会线性增加。



1. CPU

观数台分析引擎是一个64位多线程进程，经过优化，可在执行计算时利用多个处理器内核。用户选择的瞬间，观数台引擎会响应所有CPU内核的强大CPU功率，以尽快返回结果。在不繁忙时，CPU利用率将接近0%。CPU利用率和容量的度量是平均CPU，并且取决于引擎的设计方式，观数台可以在将用户添加到平台时线性扩展CPU利用率。

2. 内存

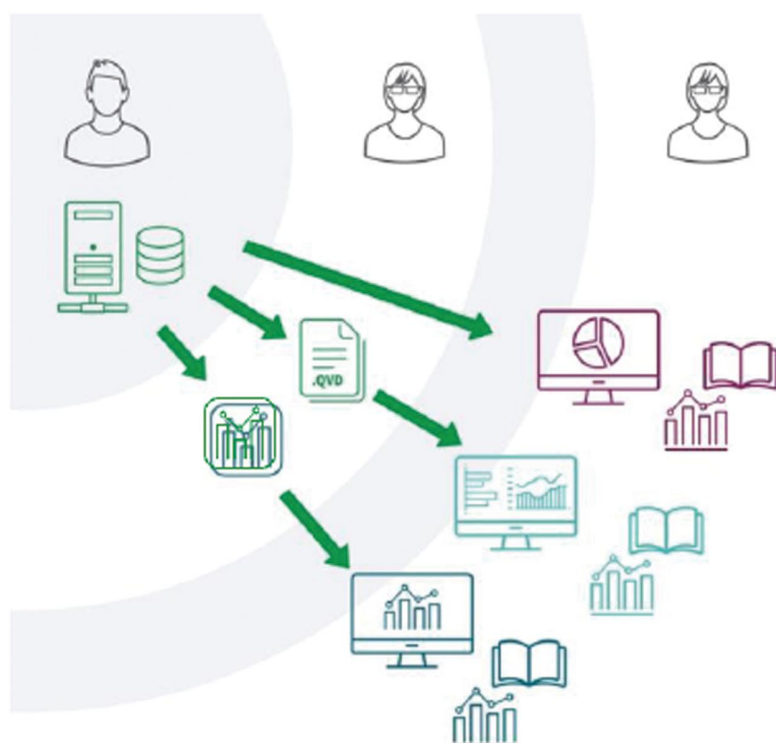
内存图表说明了标题为“内存和缓存管理”一节中突出显示的原则。随着并发用户数量翻倍，内存占用空间呈线性增长，但实际使用的内存数量却没有增加。应用程序数据一次加载，而不需为访问和分析它的多个用户重复。注意内存利用率实际上随着用户数量的增加而减慢。这是增加全局缓存的效果。

十二. 扩展交付

观数台帮助您的组织与最适合的人员保持一致。通过在虚拟、云和物理环境中部署观数台，您的IT专业人员可以高效地管理硬件。可以根据需要重新配置，横向、纵向扩展平台。您的IT专业人员可以通过集中管理数据连接来控制对数据源的访问，并将管理功能委派给不同级别的业务线管理员。

在您的用户社区中，内容创建者可以组合和准备数据，构建和使用维度和指标库，拖放可视化，创建故事和发布内容。

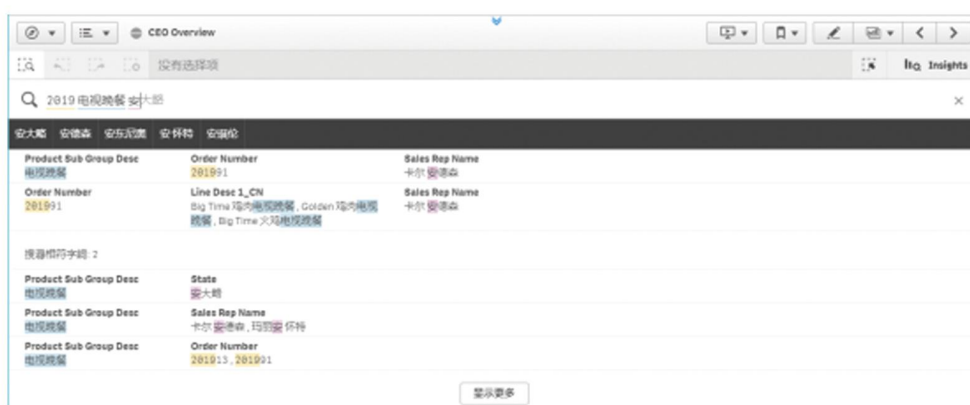
通过观数台，您的组织可以通过数据来自由地领导，通过大规模部署自助服务来发现对行业的见解。



十三. 部分特性

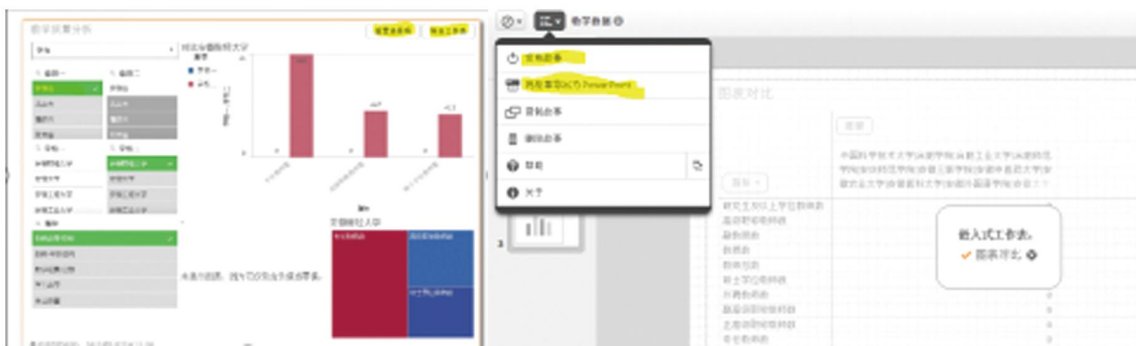
1. 强大的搜索功能

只需键入搜索字符串，智能搜索即可连接点，发现隐藏数据关系和信息并给出提示。并根据可能性自动组织条件排列组合关系



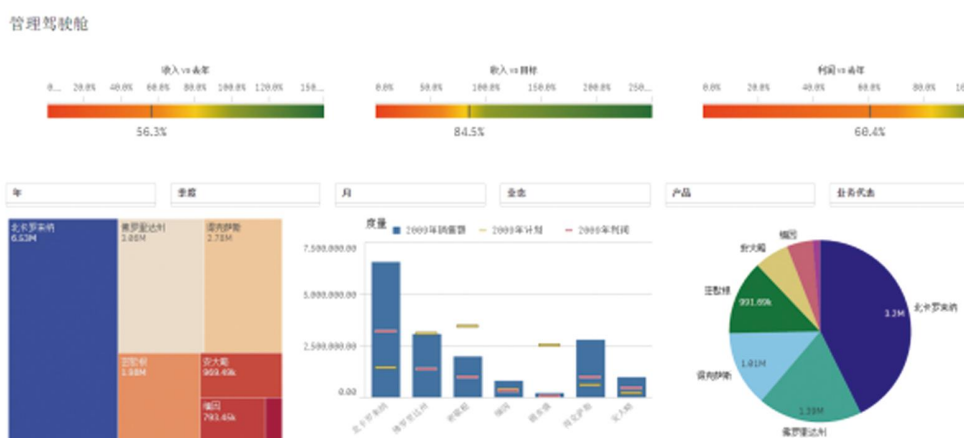
2. 快捷的数据故事和导出

分析数据同时可以根据实际需要随时制作数据故事，汇报，展示一次搞定，让分析不再枯燥，分析结果静态截图和日常界面无缝跳转，也可在数据故事里加入整页的分析报表界面，带分析步骤的业务问题溯源，半小时开紧急销售分析会议，还用担心没PPT可以用么？在线制作，PC 平板均可在线演示，可以一键导出PPT，把数据故事分享给其他同事。更可以跨过此步骤，直接在平台共享给群组。



3. 丰富的第三方组件

观数台中提供了丰富的各类常用图表展现控件，例如：柱状图、线图、柱线组合图、饼图、雷达图、气泡图、散点图、仪表盘（指针式仪表盘、数字等）、数据透视表、垂直表等图形、报表；并且所有的图形、报表对象都可通过控件属性来设置各类展现效果。



十四. 总结

观数台通过自己的架构提供一致且可预测的性能，该架构使您的组织可以灵活地部署到您需要的位置，并可以横向和纵向扩展并与数据，应用程序和用户线性扩展。使用观数台，您的IT专业人员可以有效地维护终端用户的性能，同时对计划未来的，预期的系统使用的容量充满信心。

关于观数台

亦策观数台，集合了亦策软件在商业智能（BI）领域多年的经验，精心为中国企业量身定制的本土化、轻量级、敏捷型的商业智能（BI）平台。

亦策观数台是可以允许每位用户都能深入全面地洞悉数据的下一代可视化分析平台；它将自助式BI的灵活性提升至一个新的层次，包括自助服务可视化、指导式分析应用和仪表盘、嵌入式分析和报告等。

