

Sybase IQ 15.3

实现大规模并行分析

解决方案概览





Sybase IQ 15.3解决方案 特性与好处概览

性能和扩展性

- PlexQ分布式查询平台
- 跨节点的查询并行

支持高级分析

- 对Ruby编程语言的支持
- 支持大对象数据类型的UDF
- 增强的谓词评估

数据管理和灵活性

- 多进程间通讯
- 支持Web服务
- 共享临时存储

几十年来，商业企业和政府机构力求将有关企业、客户、市场、业务流程的原始数据转换为有意义和有用的信息，以实现更加有效的战略性、操作性、战术性的决策制定。这些实践就是广为认知的商业智能，包括涵盖了数据访问与准备、报表、警告和分析等广泛行为的方法学、架构、以及技术。在这一系列行为的最高端，高级分析随着被分析的数据类型和规模的增加，已经变得更加精巧和复杂，这使得制定优化的决策更具有挑战性。

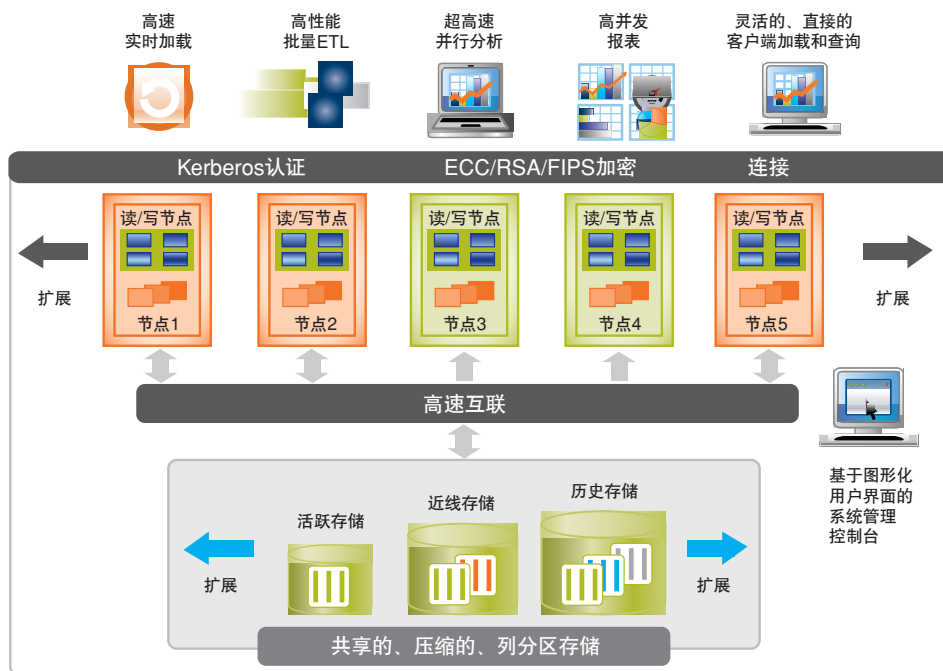
高级分析可以回答的问题代表了更高的价值和商业智能频谱中更主动的目标，并提供了新的方式去探索、理解、验证促进商业未来成功的事件与行动。高级分析可支持几乎任意的商业流程，使企业在广泛的领域开发独特的能力，包括获得更好的对客户和竞争对手的理解，预防和组织欺诈，模拟与降低风险，优化供应链执行，最大化市场投资的结果，辨别和消除利润流失，发现科学的知识。

高级分析涵盖了面向所有类型数据（业务与消费；内部与外部；结构化和非结构化）的数学模型和计算算法的应用，以将这些数据转换为可行动的信息。它帮助企业使用这些数据和知识资产获得远见以及采取适当的行动，而且分析既可以帮助人们进行决策，也可以驱动全自动的决策制定。用于高级分析的技术包括统计分析、预测和推断、数据挖掘、预测建模、以及决策优化。成功的使用这些方法需要访问大量的历史存储和当前数据、分析软件工具、以及一个存储数据和执行查询与模型以获得商业洞察力的分析服务器。而Sybase IQ是市场上第一家也是领先的列式分析服务器，它通过不断的技术创新全面提交更智慧的高级分析。

Sybase IQ—— 独一无二的设计, 分析的理想选择

Sybase IQ的设计正是为了满足企业不断增长的多样化的分析需求。它可以支持大量的用户、大规模数据, 而且可以提交对商业信息的高速访问, 速度比传统的关系型数据库快10到100倍, 甚至1000倍。相比于传统的关系型数据库以行的方式存储数据, Sybase IQ使用一个被称之为垂直分区的策略, 也就是以列的方式存储数据。使用Sybase IQ, 分析查询仅仅读取查询所需的列, 而传统的关系型数据库需要缓慢的读取整个表的每一行, 阻塞了I/O通道、内存和磁盘。

另外, Sybase IQ采用Multiplex技术, 充分利用了单个服务器(节点)执行特定功能而磁盘存储共享通用的优势。存储是共享的、压缩的、列分区存储的, 它们通过一条高速的光纤通道与Multiplex架构中的所有服务器节点互联。这些服务器节点既可以作为数据读节点也可以作为写节点(或者两者都是)运行, 支持所有连接到共享数据源的并发的数据加载和查询。这允许大量的使用场景, 比如快速的并行分析和报表, 同时支持高速的实时数据加载、批量数据集成任务、即席客户端加载和查询(见下图):



Sybase将Sybase IQ这种革命性的架构与不断为软件添加越来越多智能的开发哲学结合起来, 然后在最新的尖端硬件技术上使用这个智能软件, 以最大限度地从底层资源中挤出最后一点点性能。这从Sybase IQ 15平台的新产品发布轨迹可以得到验证, 最新的15.3版本目前已经发布。

Sybase IQ的设计正是为了满足企业不断增长的多样化的分析需求。它可以支持大量的用户、大规模数据, 而且可以提交对商业信息的高速访问, 速度比传统的关系型数据库快10到100倍, 甚至1000倍。



Sybase IQ发布了15.3版本，
它将性能、分析、灵活性提高
到更高的级别。

Sybase IQ 15平台

2年前，Sybase发布了15版本的Sybase IQ分析服务器，推进了其业已非常强势的增长。Sybase IQ 15提升了核心的引擎基础架构能力，包括数据加载和查询性能、改进的安全架构、强大的集群管理（使得在一个集群中将多个读节点和写节点组合在一起）、因数据分区而增强的管理。自那以后，Sybase陆续发布了15.1和15.2版本，每个版本都着力于增加新的核心能力以促进更深入的高级分析。

Sybase IQ 15.1引入了数据库内分析这样一个高性能的进程，允许用户定义模型和合作伙伴分析库（例如来自Fuzzy Logix和Visual Numerics的分析库）并在数据库内执行，而不是在分离的分析服务器上执行。因为数据在分析结果被物化之前从来没有离开过数据库，这缩短了从数据搜集到生成结果的周期，允许更大的数据集被分析，同时获得更准确的结果。Sybase IQ 15.1 也提升了对数据库建模、管理与监控的工具支持。

Sybase IQ 15.2进一步提升了分析支持，包括全文检索、挖掘与分析，以及通过提供流行的Web编程语言如Python、Perl、PHP的驱动丰富了对Web分析的支持。它也提供了对联邦查询的支持，从而扩展了可在Sybase IQ框架内分析的数据源的范围。有了这个功能，位于远程数据库中的数据可以被检索，并与Sybase IQ中的数据合在一起，作为一个全面的高性能分析查询和模型执行的一部分。所有这些积极的针对市场的革新使得Sybase IQ长期以来一直都是市场上第一的列式分析服务器。目前，全球有2000多家客户部署了Sybase IQ，在过去三年，每年都有大约200个新客户加入到Sybase IQ阵营。同时200多家客户已经从Sybase IQ 12.7版本移植到了15平台；而且，15版本（这次使用的是15.2版本）在2010年10月的TPC-H性能基准测试（基于IBM Power系统，运行1TB的数据）中再次排在首位。现在，Sybase IQ发布了15.3版本，它将性能、分析、灵活性提高到更高的级别。

Sybase IQ 15.3—— 核心主题与新特性

Sybase IQ 15.3基于之前的15版本,重新定义了为提升分析性能和可扩展性的大规模并行处理(MPP)。它同时继续加大对高级分析开发、建模和执行的支持,而且提供了更加强健的数据管理和灵活性。我们会接着对这三个主题分别进行阐述,包括所强调的每个主题的几个关键特性。

性能和可扩展性

在最新的版本中, Sybase IQ利用了独特的Multiplex与共享存储设计, 提供了性能和可扩展性的新高度。查询并行, 这个一直在任意服务器节点上所具有的功能, 现在通过分布式查询进程扩展到跨多个服务器节点。而且, 私有云和弹性计算也随着可动态创建以支持不同工作负载的逻辑服务器概念而被引入。

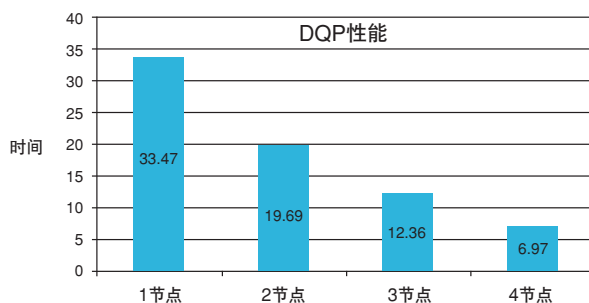
- **分布式查询进程**——Sybase IQ 15.3引入了PlexQ 分布式查询平台(DQP), 一个大规模并行处理(MPP)架构, 通过将工作分布到一个Multiplex网格配置中的众多计算机上加快了高度复杂查询的速度。不同于无共享(Shared-nothing) MPP架构, PlexQ使用一个全共享(Shared-everything)方法, 在多个计算节点中动态管理和平衡查询负载。PlexQ DQP可以将一个查询分解为很多部分并将这些部分在多个服务器节点上进行分布, 通过并行执行提高查询性能。

使用PlexQ DQP之后, Sybase IQ现在可以在计算能力、I/O吞吐、数据存储的所有维度上提供超级的性能和可扩展能力。计算能力的扩展是通过上面所述的MPP架构实现的。对共享存储层的IO吞吐通过Sybase IQ智能化的垂直分区列存储、大页面、Bitmap索引技术仍保持着高性能。存储能力的扩展可简单的通过增加更多的磁盘到多层存储环境中实现。

由于计算能力不是通过物理分布(分区)数据而获得, 全共享的方法比典型的无共享MPP方法提供了众多的好处。这包括更好的用户并行性以及同时处理多个查询, 对即席查询的更好的支持, 以及更加高效的扩展, 因为数据分区不需要随着用户扩展其环境而重新平衡或者重新配置。

- **逻辑服务器提供的弹性计算**——Sybase IQ 15.3的逻辑服务器提供了一流的资源供给能力, 通过让用户将一个个物理硬件资源组合成为一个单一的、逻辑Multiplex服务器。当用户连接到一个逻辑服务器上执行一个查询的时候, 查询的执行仅仅是分发到逻辑服务器的成员节点上, 而成员节点可以动态的增加或删除以适应单独的应用负载的需求。这种尖端的能力提交了对私有数据集云的弹性计算, 而且可以进行实时调整, 确保每个应用负载总是接收最佳数量的计算和存储资源从而以最高的性能运行。

- **证据**——来自Sybase IQ 15.3测试计划的结果显示, PlexQ DQP技术所做的承诺均得到了验证。随着给PlexQ环境递增式地增加计算能力, 一些关键的高级分析查询显示了接近线性的性能可扩展性。比如这个例子, 一个包含了大型联合与Group By 从句的复杂查询随着节点从1个到8个递增, 产生了接近线性的可扩展性。



Sybase IQ 15.3基于之前的15版本,重新定义了为提升分析性能和可扩展性的大规模并行处理(MPP)。



高级分析

Sybase IQ 15.3继续通过增加流行的Web分析编程语言的驱动, 以及扩展用户自定义函数中支持的数据类型, 扩大对分析工具支持的级别。

- **Ruby编程语言支持**——Ruby是一个面向对象的编程语言, 常用于Web应用的开发, 类似于Python或Perl。Sybase IQ 15.3现在提供了对一些Ruby API包的支持, 包括一个内建的Ruby驱动, 允许Ruby代码与Sybase IQ数据库进行交互。
- **扩展用户自定义函数的支持**——目前Sybase IQ中的用户自定义函数支持将大对象数据类型作为输入参数。与Sybase IQ大对象管理选项联合, 使用户能够将大型非结构化数据对象作为他们自定义的分析函数的输入, 从而允许大数据对象内容和上下文的分析。这些对象类型包括: 图形、地图、文本、视频和音频文件, 以及对分析进程具有重要价值的众多信息。

数据管理与灵活性

数据管理、灵活性与信息可用性将在Sybase IQ 15.3中因三个关键特性而得到进一步加强。首先是高可用性和单个服务器节点间更细粒度的通讯。其次是将Web服务作为Sybase IQ中的新数据集使用, 以及引入一个高性能的ELT (抽取、加载、转换)。

- **Multiplex进程间通讯 (MIPC)** —— 该技术是一个通讯层, 提供了Multiplex中所有节点间的事务连接。使用该技术, 每个节点可以与其他节点通讯, 而且可以在节点间打开私有通讯连接。这些可选的冗余私有交互允许数据共享, 支持分布式查询进程, 为Sybase IQ Multiplex服务器提供高可用性。
- **Web服务支持**——Sybase IQ现在提供内建的对Web服务的支持, 这使用户可以听取和管理标准的SOAP和HTTP请求。该功能现在可作为一个客户端, 允许在数据库内运行应用访问标准的Internet中可用的、或由其他Sybase IQ 数据库提供的Web服务。该功能通过使用Sybase IQ中一个新的函数命令来实现, 允许分析人员动态建立HTML内容或表格, 为数据集提供控制性供最终用户选择。
- **高性能的ELT (抽取、加载、转换)** —— 为了更快的将数据加载到Sybase IQ, 以及利用其强大的数据转换计算能力的优势, Sybase IQ 15.3正引入ELT技术。DBA现在可以通过一个抽取和加载任务将源数据和模型更快速的移动到Sybase IQ中, 数据直接流动, 无需任何改变。一旦数据被加载, 一个SQL转换任务将被定义并执行, 将数据和模型修改为支持分析查询需求的格式。无论加载还是数据库内转换, 都是高速完成, 加速了数据准备过程, 而且通过一个菜单驱动的设计非常轻松的使用, 仅仅需要几步从源到目的点击。
- **Sybase控制中心 (SCC) 增强的工具支持**——Sybase 控制中心是一个基于Web浏览器的通用Sybase服务器产品的管理和监控框架, 包括Sybase 数据库服务器Adaptive Server Enterprise (ASE)、Sybase 复制服务器、Sybase IQ。拥有丰富监控功能的针对Sybase IQ的控制插件在Sybase IQ 15.1中首次引入。它提供了强健的面向任务的监控功能, 比如全面的系统健康监控如热图表、KPI分析、节点层次与连接的向下钻取、对Multiplex和Simplex的Sybase IQ环境的历史数据分析。在Sybase IQ 15.3中, 该插件不仅升级了某些Sybase控制中心所支持的关键的管理任务, 而且增加了对所有Sybase IQ 15.3中引入的新特性的支持。尽管Sybase IQ需要非常少量的DBA支持, 新一代的Sybase控制中心仍致力于让DBA工作比从前更轻松。

总结

随着Sybase IQ15.3的发布, Sybase通过开发和扩展革新的数据库技术继续屹立于分析领域的领导者位置。大多数竞争对手的无共享MPP系统执行查询的时候严重的缺乏负载均衡。Sybase IQ PlexQ平台拥有的分布式查询处理是一个革命性的全共享的MPP架构, 通过将查询在一个PlexQ环境中的所有节点中进行分步, 加速了高度复杂查询的速度, 提升了性能级别, 同时维持部署和管理的简易性。Sybase IQ 15.3 通过增加工具支持、基于节点间的通讯管理、以及支持Web服务进一步加强了之前版本在高级分析上的优势。由于这些以及更多其他特性, Sybase IQ 15.3继续成为事务关键型、高性能的、低成本-高效率的分析的最佳平台, 让企业将信息转化为更好的决策。

好处	Sybase IQ 15,15.1,15.2特性	Sybase IQ 15.3新特性
性能和扩展性	基于每个节点的查询并行 多列索引的使用 子查询速度的提高 实时数据加载	PlexQ分布式查询平台 跨节点的查询并行 逻辑服务器的弹性计算
对高级分析的支持	数据库内分析 文本检索和分析 支持Web分析 合作伙伴分析库	支持Ruby编程语言 支持大对象数据类型的UDF 增强的谓词评估
数据管理和灵活性	范围分区 查询联邦 可配置的表空间 基于角色的管理	多进程间通讯 支持Web服务 共享临时存储

Sybase IQ 15.3继续成为事务关键型、高性能的、低成本-高效率的分析的最佳平台, 让企业将信息转化为更好的决策。



赛贝斯软件（中国）有限公司

北京总公司

北京市朝阳区建国门外大街乙
12号双子座大厦东塔9层
邮编: 100022
电话: (010)59215888
传真: (010)59215999

上海分公司

上海市黄浦区广东路500号
世界贸易大厦31楼01室
邮编: 200001
电话: (021)23210888
传真: (021)23210999

广州分公司

广州市天河区林和中路
8号海航大厦2404-2405室
邮编: 510610
电话: (020)38135388
传真: (020)85501602

成都分公司

成都市人民南路二段
18号川信大厦31-3号
邮编: 610016
电话: (028)86197488
传真: (028)86200921

深圳分公司

深圳市福田区竹子林紫竹七道
18号中国经贸大厦5楼J室
邮编: 518040
电话: (0755)82537000
传真: (0755)82537249

www.sybase.com.cn

