**openGauss不仅具有强大的计算性能、高效的数据处理能力，同时其本身也是人工智能原生（AI Native）数据库，支持参数自调优、SQL诊断、故障自诊断、全生命周期管理等功能。下面分别介绍几个openGauss的AI相关特性。**

**X-Tuner: 数据库参数调优框架**

数据库是一个非常复杂的系统，它包含大量的配置参数，控制着内存分配、IO优化、查询计划成本、并行、日志记录、数据恢复等其他行为。由于在特定负载下，不同的配置参数，数据库表现性能也不一样，因此很多数据库管理员甚至领域专家都在努力调整这些参数以获取数据库良好的表现性能。

但不幸的是，数据库参数调优是一个NP-hard问题。通常情况下人们主要基于自身经验和对当前环境的理解去调整参数，调整过程中存在大量不确定性。为了解决这个问题，数据库工程师试图去构建一个可以自动调优的智能系统去代替人类进行调优工作，比如 PostgreSQL的Postgresqltuner.pl以及MySQL的mysqltuner.pl。同样，随着人工智能技术的兴起，目前也出现了一些基于AI的调优方案，比如OtterTune. 然而，目前所有的调优方案都有以下几个限制：

数据库有几百的参数，DBAs不能一次性调整如此数量的参数；

传统的机器学习需要大量的数据训练，不考虑这种方式是否可行，仅仅数据收集就是一件不容易的事情，特别是获取拥有良好表现的数据显得尤为艰难；

一些方案采用了强化学习的方式，但是这种方案在调整时没有考虑数据库状态和数据库参数之间的关系；

由于上述方法的限制，openGauss开发了自己的数据库参数调优框架X-Tuner, 相比于传统方法，X-Tuner具有如下的功能特点：

**1. 鲁棒性强，有一定的容错性：**

X-Tuner框架在设计时包含了大量的容错机制和紧急处理机制。在系统或数据库出现故障时，算法能够正常退出，同时不会对系统带来任何影响。

**2. 灵活部署，使用方便：**

X-Tuner基于Python3.0+开发，支持Linux和Window系统，用户可以方便的进行部署。在使用方面，X-Tuner支持本地和远程两种连接模式，适用于用户的各种情况。

**3. 理解容易，便于二次开发：**

X-Tuner严格按照基准测试模块、调优算法模块、连接模块、日志模块等进行编写，层次性极强，理解容易，同时也方便用户在此基础上进行优化或编写专属于自己的功能模块。

经过实际测试证明，基于强化学习和启发式算法的X-Tuner参数调优框架，能够保障在占用内存尽可能少的情况下，极大地提升系统的性能，如图1所示，为参数调优过程的图示：

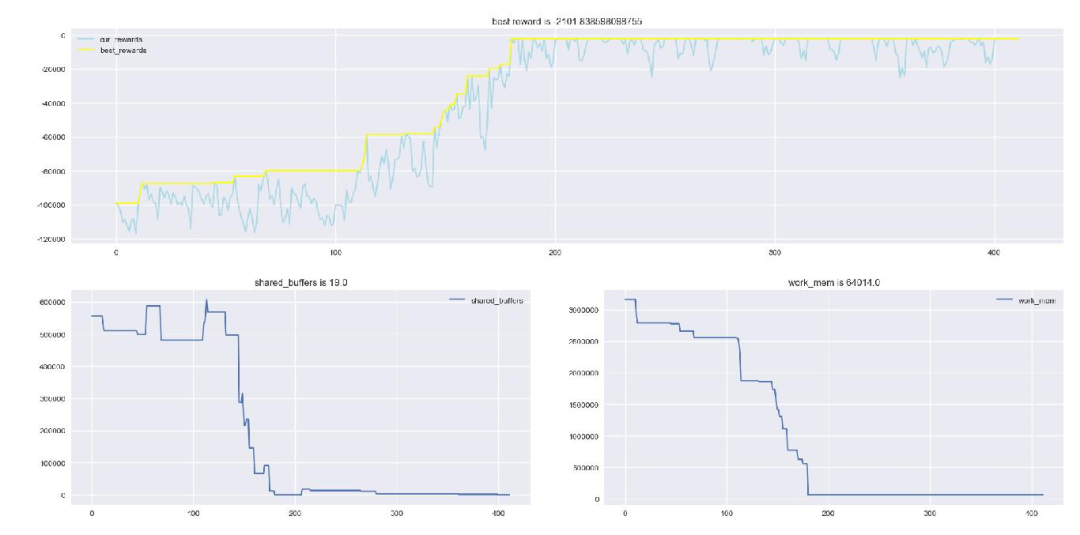


图1 X-Tuner 参数调优过程

在实际应用上，依托于 X-Tuner相关技术，华为云数据库DAS服务能够针对用户数据库的历史负载情况，智能进行参数推荐。经过实际测试，性能总体提升20%左右，能够极大地为用户节省云计算资源，降低生产成本。

**SQLDiag: SQL语句智能识别**

SQLDiag是openGauss中SQL语句执行时长预估框架。现有的预测技术主要基于执行计划的模型预测，但这些预测方案主要适用于OLAP场景，并且要求可以获取到SQL语句的完整执行计划。这对于OLTP或者HTAP这样的短查询限制极大。

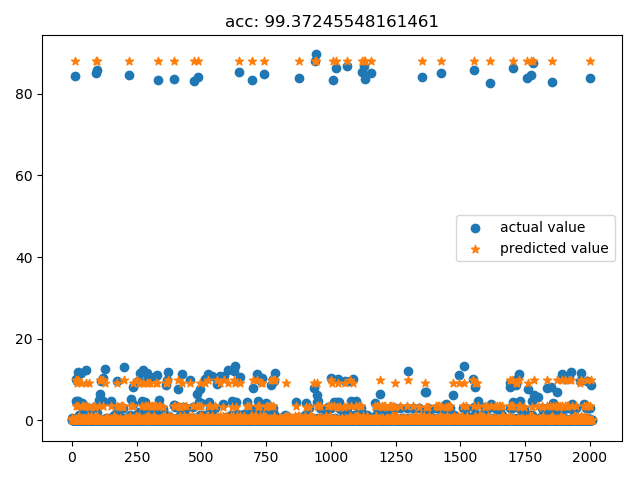
与上述方案不同，SQLDiag着眼于数据库的历史SQL语句，因为短时间内数据库SQL语句执行时长不会有太大的差距，SQLDiag可以从历史数据中检测出相似SQL语句，并基于SQL向量化技术和时序预测算法预测SQL语句执行时长，**进而识别出潜在慢SQL**。该框架有如下优点：

不需要SQL语句的执行计划，对数据库性能不会有任何的影响；

使用场景广泛，目前业内的很多算法具有很强的针对性，比如只适用于OLTP或者OLAP，SQLDiag使用场景广泛，经过改造甚至可以应用于NoSQL中；

该框架鲁棒性强，理解容易，只需要简单的修改，用户就可以设计出自己的预测模型。

经过实际测试，SQLDiag的表现结果如图2所示：

图2 SQLDiag预测结果

**智能优化器：数据库执行代价预测**

在依赖数据库的大量业务场景中，简单至一次账户登录、订单查询，复杂至上亿行的报表查询、数据挖掘，都会通过业务应用层的抽象转化以SQL语句的形式对数据库进行操作，数据库内核中的SQL引擎则会对SQL语句进行进一步优化。

来自业务逻辑的挑战：SQL语句的处理过程中，日趋复杂业务场景和商业智能工具的应用，导致生成的SQL语句自身质量良莠不齐，其中的一部分可能会耗费大量的读写和计算资源，导致服务器对其他业务语句造成阻塞，这就对数据库上层业务组件的SQL诊断能力提出了要求。

来自SQL语句的挑战：查询执行计划的优化过程中，更为复杂的查询语句也给**查询优化**带来了新的挑战。在数据库系统发展的早期查询优化主要依赖基于规则的专家系统，即一系列有着严格顺序的优化规则。这类专家系统可以被认为是数据库优化专家对于某些常见场景的经验总结。这就导致对于相同的查询语句，不论数据库中实际数据量的大小和分布情况，规则系统总会生成相同的查询计划。一旦遇到和经验规则不符的场景，数据库效率难往往难以得到保障。

不论是SQL诊断还是计划优化，其中最为核心的问题是对资源开销的评估。目前主流数据库产品中，查询资源开销主要依赖对数据库中的全量数据进行采样分析并建立统计学模型和代价模型对计划执行过程的抽象，目前面临以下挑战：

数据采样和分析的过程对服务器的磁盘读写和CPU计算资源有相当大程度的侵占，但如果为了规避这一部分的资源开销而降低采样率则会导致查询资源开销的评估准确率下降；

随着业务语句不断执行，数据的规模和分布情况会逐渐变化从而导致原先建立的数据模型失效，这两类问题在比较严重的情况下可能会引发服务器长时间无响应的情况。

那么有没有一种方法能够在尽可能少侵占数据库资源的前提下保持对资源开销预测的准确性呢？答案是有的：openGauss为数据库用户推出基于在线深度学习的查询性能预测功能。

**智能优化器功能特点：**

**1. 一键建模，机器学习小白也能驾驭**

openGauss针对SQL性能评估这一场景一站式整合了数据收集、数据清洗、数据预处理、数据编码、训练监控等流程。用户只需要根据推荐为模型配置少量参数，并在模型训练完成后调用模型预测接口直接得到性能预测结果。

**2. 粒度更细，轻松定位计划性能瓶颈**

支持更细粒度的查询计划片段级别查询性能预测，辅助定位真正的性能瓶颈点，帮助用户更好地改写语句。未来可支持基于计划片段代价性能预测的智能SQL优化。

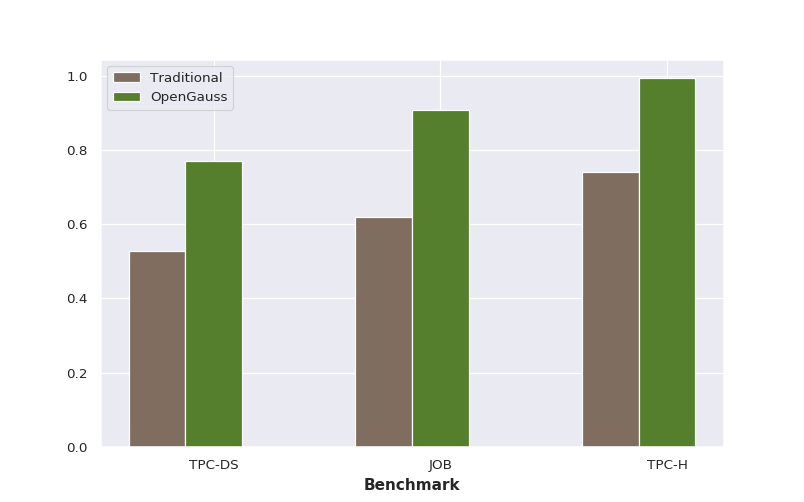
**3. 灵活部署，最小化数据库性能影响**

可根据用户需求将模型运算模块部署在云上或其他与数据库相隔离的环境。使用历史性能数据进行建模，而无需数据采样带来的额外资源开销。

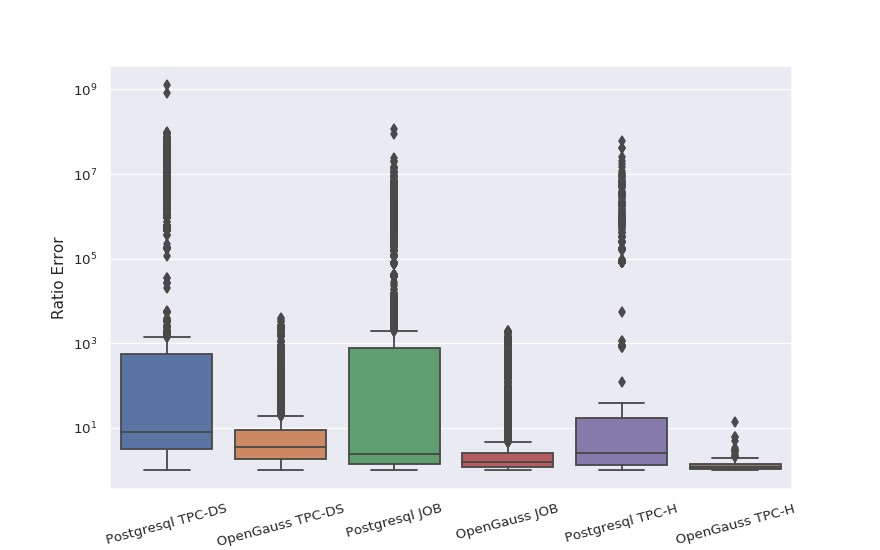
**4. 接口开放，数据科学工作者的福音**

使用HTTPS协议将数据库内核和深度学习模型相连，开放接口，支持用户将自定义机器学习作业内置于数据库函数中，一键调用。

经过实际测试，**查询性能预测准确率整体相比PG原生模型提升40%：**

图3 openGauss对比PG原生查询性能预测的表现

从预测的准确率的分布情况来看，95分位数准确率提升了3e5倍，75分位数提升了124倍，整体预测可靠性显著提升：

图4 openGauss与PG预测准确率的分布情况

**openGauss的总结与展望**

openGauss是AI结合数据库的一次粗浅尝试，或许仍有许多不足之处。上述介绍的特性是本次开源的典型AI特性，除上述特性外，仍有很多AI相关特性处于探索中。我们能够感受到，在上述应用场景中，基于AI的方法能够极大地解放人力，提高生产效率。

虽然AI与数据库相结合的道路十分艰难崎岖，甚至面临业界的质疑，但是无数研发者不曾放弃心中的信念，秉承着“万物智能”的理想砥砺前行。通过开源openGauss，我们希望以此能够起到抛砖引玉的作用，促进业内更多开发者投入到AI与数据库结合的事业中，进一步推动数据库技术的升级、激励更多有价值、有意义的AI数据库出现，从而在未来实现数据库领域智能化的伟大愿景。