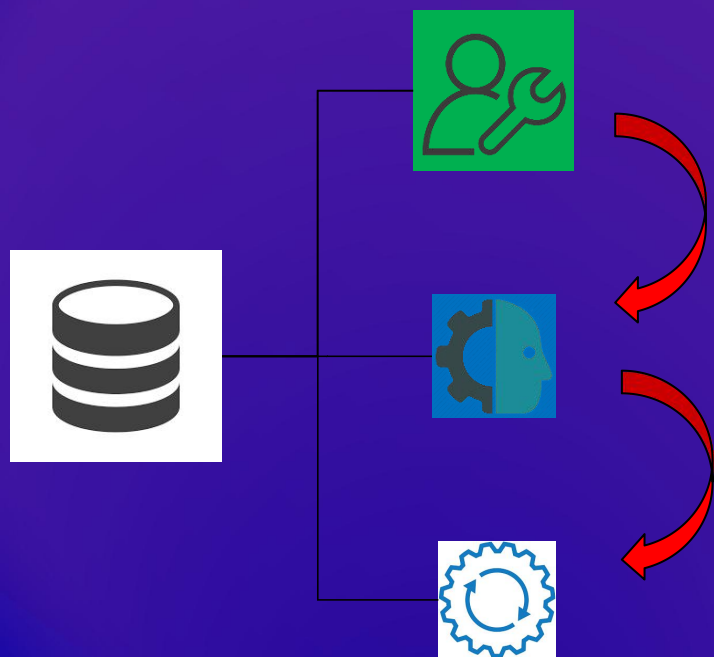




openGauss AI能力及规划

李士福

提纲



AI全景图

AI4DB介绍

DB4AI介绍

规划和路标

PART 1 AI全景图

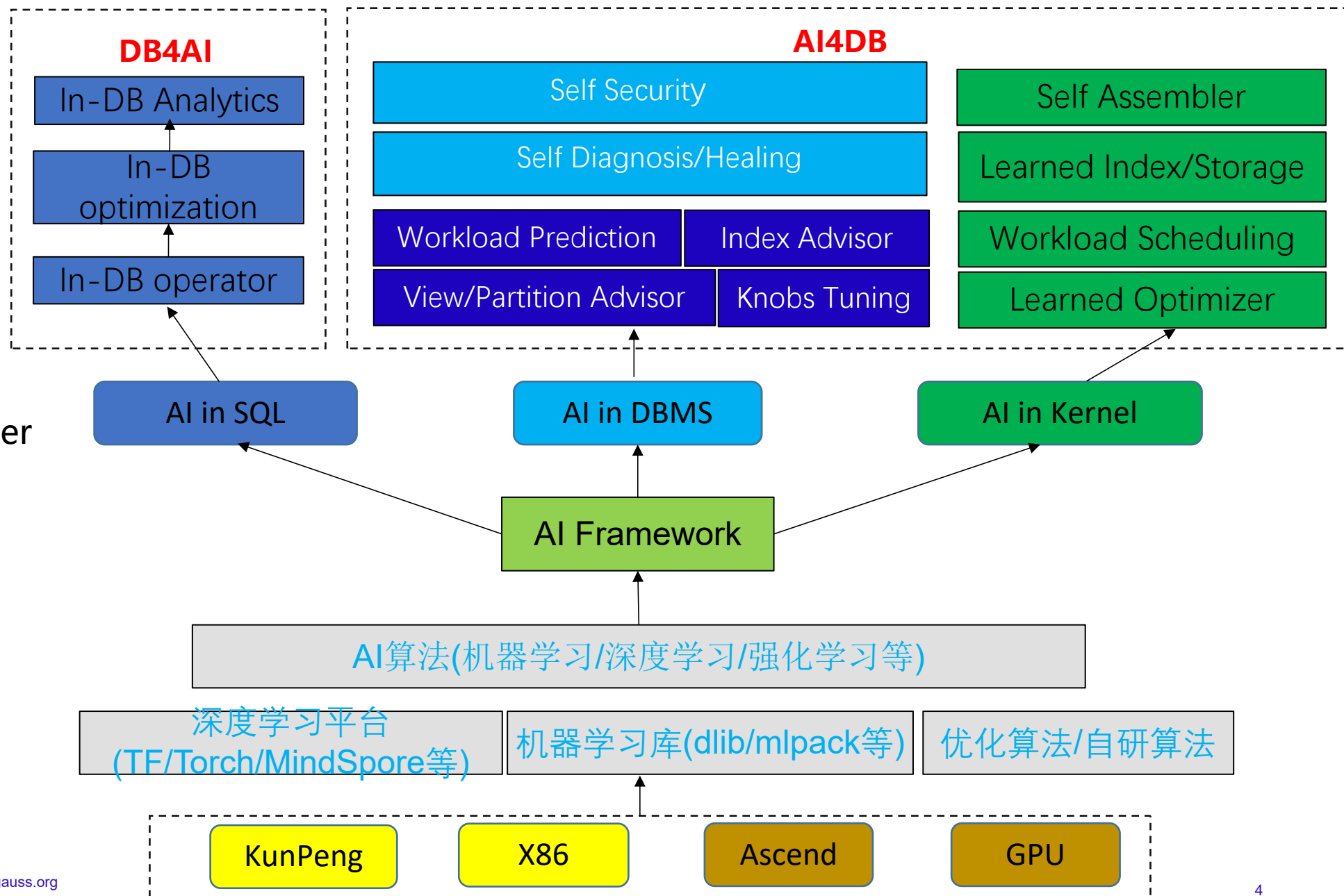
AI全景 - AI Framework for Database



全应用场景
SQL/DBMS/Kernel

全用户
User/DBA/Developer

全技术栈
AI/ML/DL



AI4DB

充分发挥AI芯片优势，提升DB智能化水平

- 1、自调优
 - 参数自调优：天→分钟级；达到、超越专业DBA
 - 负载自调优：负载预测，高效作业编排
 - 索引推荐
- 2、自监控：
 - 负载预测、空间预测、时间预测等
- 3、自诊断、自愈
 - 根因分析
 - 慢SQL诊断与建议
- 4、自安全
 - 安全规则自动识别
- 5、AI优化器（RBO-→CBO-→AI-BO）
 - 强化学习来代替启发式算法
 - 代价估计
- 6、自组装
 - 优化器自适应：强化学习来代替启发式算法
 - 利用AI探索最优路径
 - 单路径→多路径

DB4AI

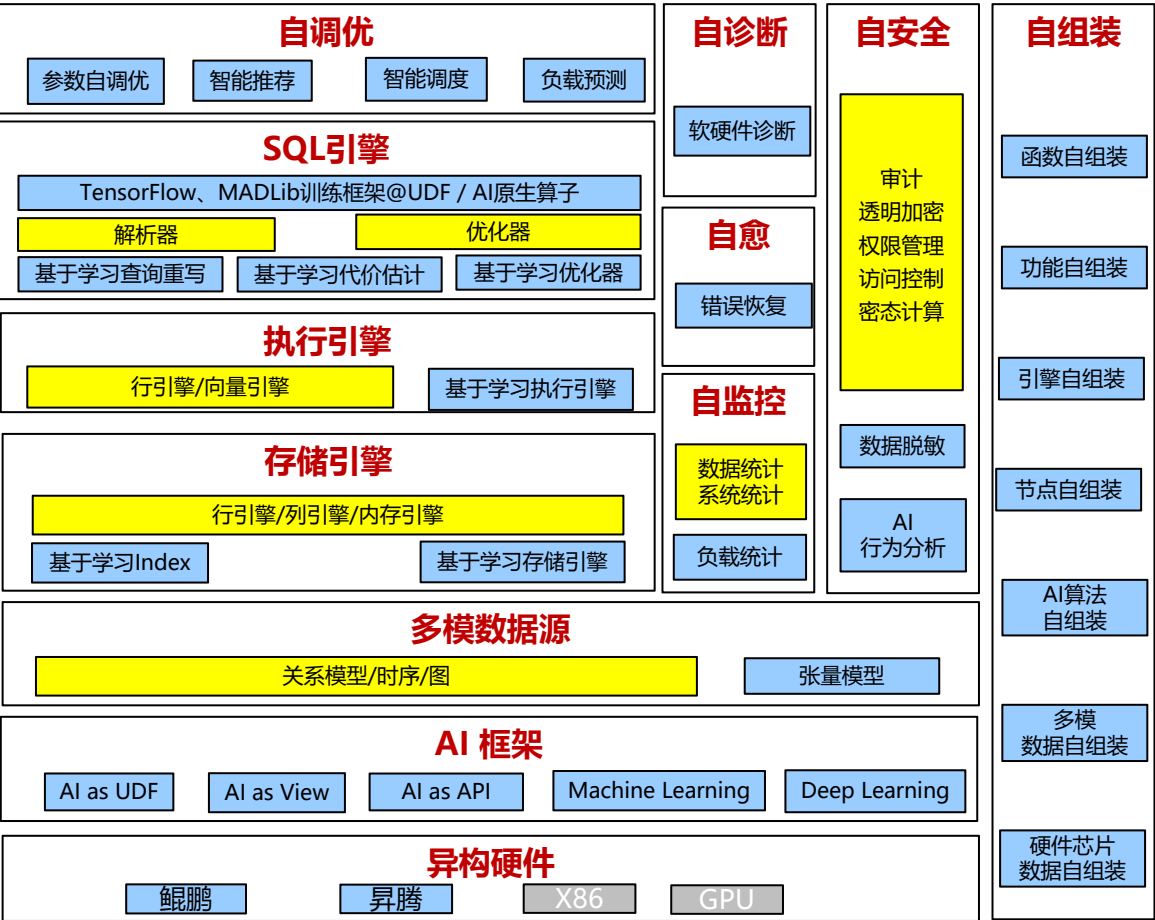
通过SQL降低AI使用门槛、推广AI使用场景

- 1、通过SQL调用复杂AI算法，无数据迁移、无缝集成多种AI深度学习框架与算法
- 2、自动选择、组装AI算法、AI芯片硬件加速
- 3、将异构芯片嵌入到数据库系统，提供端到端AI能力
 - 1980：训练模型（数据库内训练）
 - 1910：推理模型（数据库内推理）
- 4、利用DB优化能力提升训练速度
 - 索引：雨天、黑夜
 - 存储：模型参数
- 5、利用DB优化能力提升在线决策
 - 实时决策，图像识别、人脸识别、车-自动驾驶

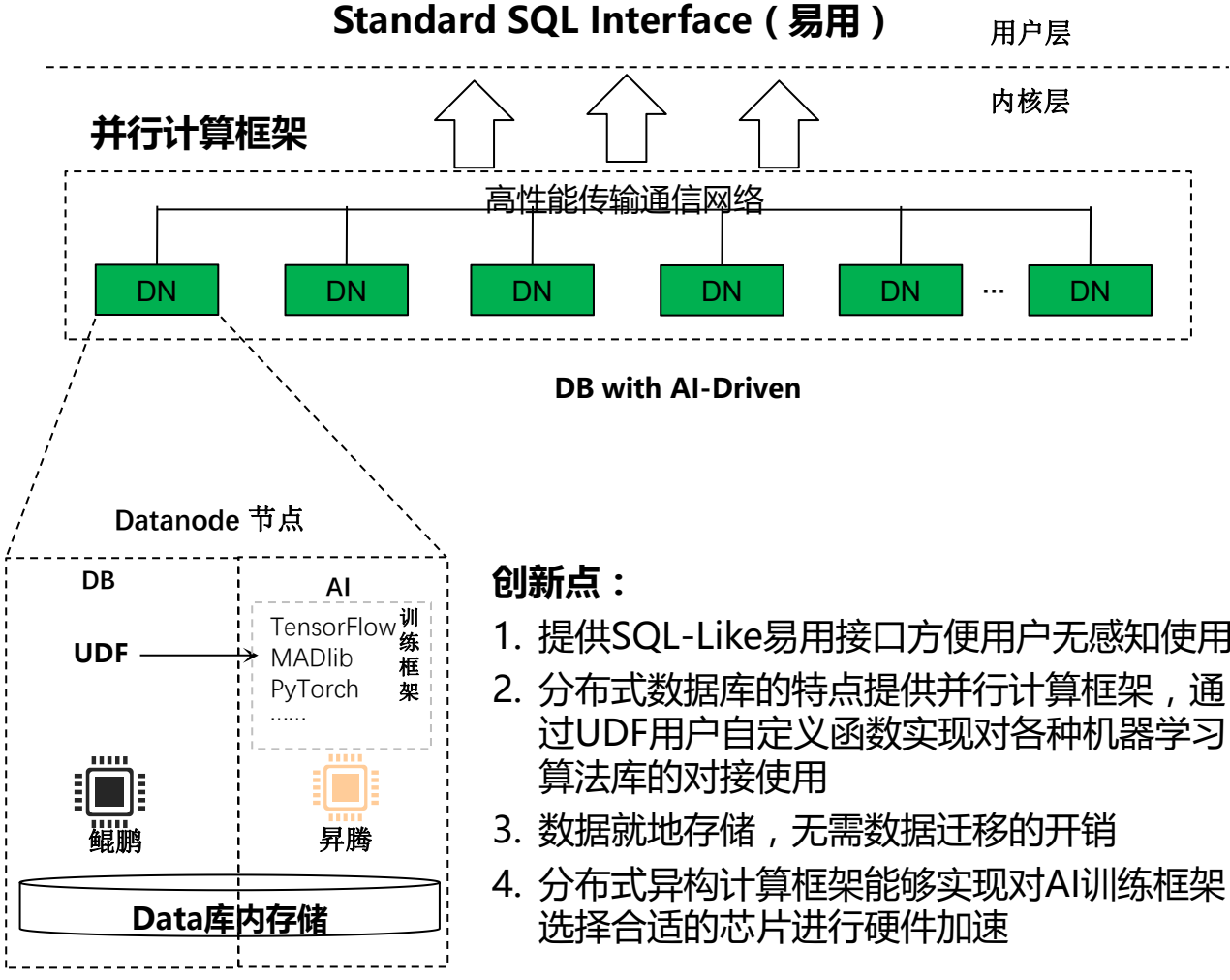
openGauss新一代数据库AI内核(AI4DB & DB4AI)

AI4DB：基于AI实现新一代的数据库引擎

AI-Advised ⇨ AI-Assist ⇨ AI-Enhanced ⇨ AI-Assembled ⇨ AI-Designed



DB4AI：简化AI开发、将AI训练推理效率最大化

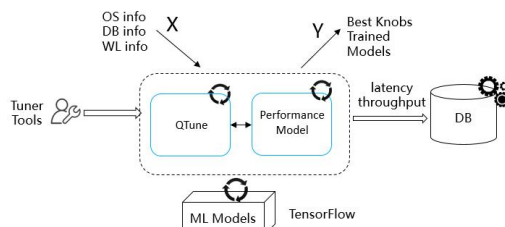


PART 2 AI4DB

AI4DB: 全面提升数据库智能化水平，应对不同行业多样化负载

1 自调优，越用越快

参数自调优&慢SQL诊断



AI优化器



① AI代价评估 ② AI调优模型 ③ AI搜索算法

云化智能调优

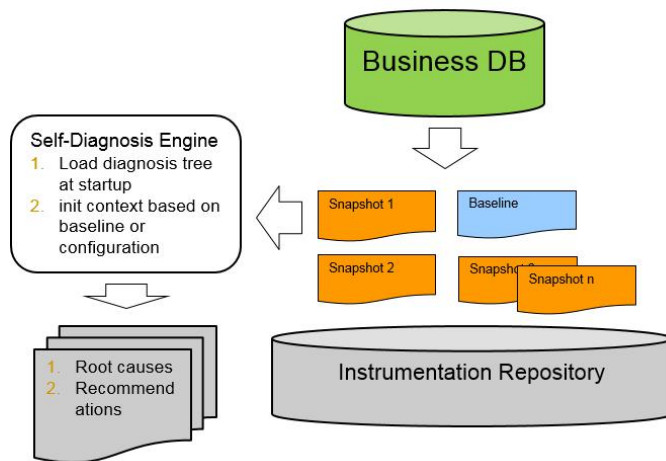
- **参数自调优**：调优时间：天→分钟级；**慢SQL诊断调优**：快速识别系统中的慢SQL，并给出调优建议
- **索引物化视图推荐**：从查询特征得出索引、物化视图
- **AI优化器**：下一代ABO优化器，实现数据库越跑越快
- **融合优化**：突破关系模型优化，实现多模数据融合优化与最优调度

云化负载预测

- **负载预测**：预测负载性能，支撑高效编排业务
- **存储空间预测**：基于历史数据增长趋势，预测可用天数，及时扩容
- **智能资源调度**：自动资源调度，智能负载均衡

2 自诊断自愈，always online

自愈自诊断架



自诊断

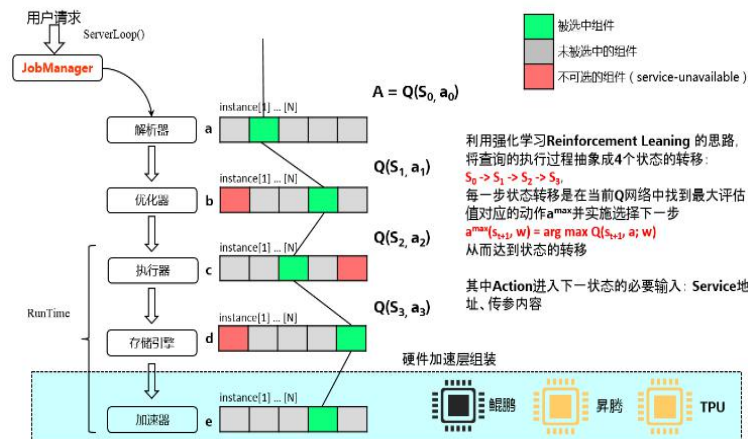
- **诊断引擎**：支持AI和Rule引擎结合的基于Trace智能诊断系统
- **诊断模式库**：支持常见故障模式库，智能可扩展的模式库

自愈&故障预警

- **故障预警**：智能预测软件&硬件故障故障检测
- **故障恢复**：基于诊断模式库，恢复流程最优化
- **统计信息恢复**：根据系统查询数据STAT老化程度自动修复统计信息
- **数据恢复**：索引、数据脏页自动修复

3 自组装，全场景

One-Stack-Fit-All 自组装框架



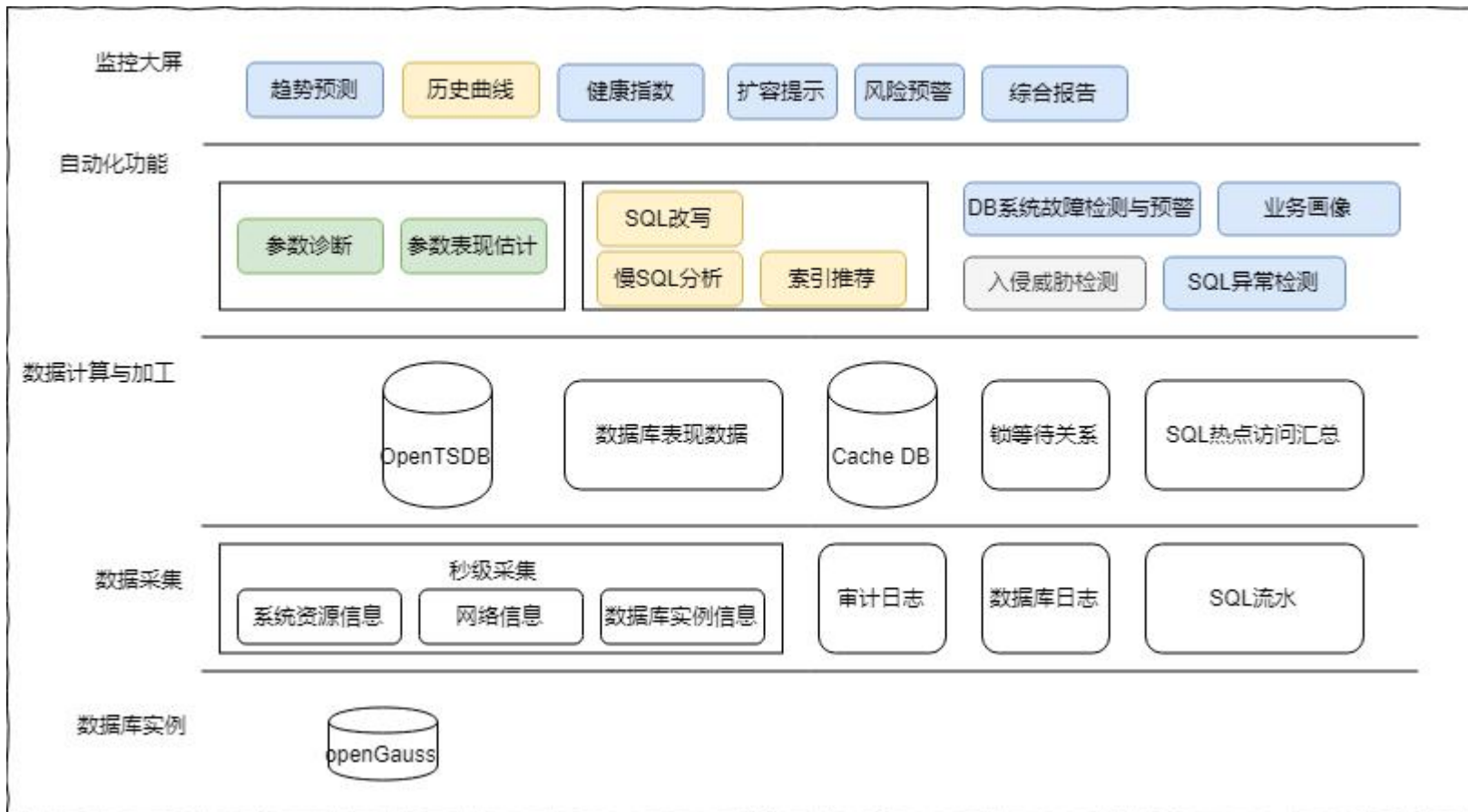
丰富的行业应用、支持多样化负载

- **业务场景和负载多样化**，单一通用数据库平台无法满足
- 现有的产品研发模式无法满足定制化交付

组件化自组装

- **组件微服务化**：超平面快速搜索理论，突破组合空间爆炸
- **自组装技术**：基于最优控制论，软件执行栈轻量化、组件组合最优化

AI4DB: DBMind智能管理平台全景规划图



功能:

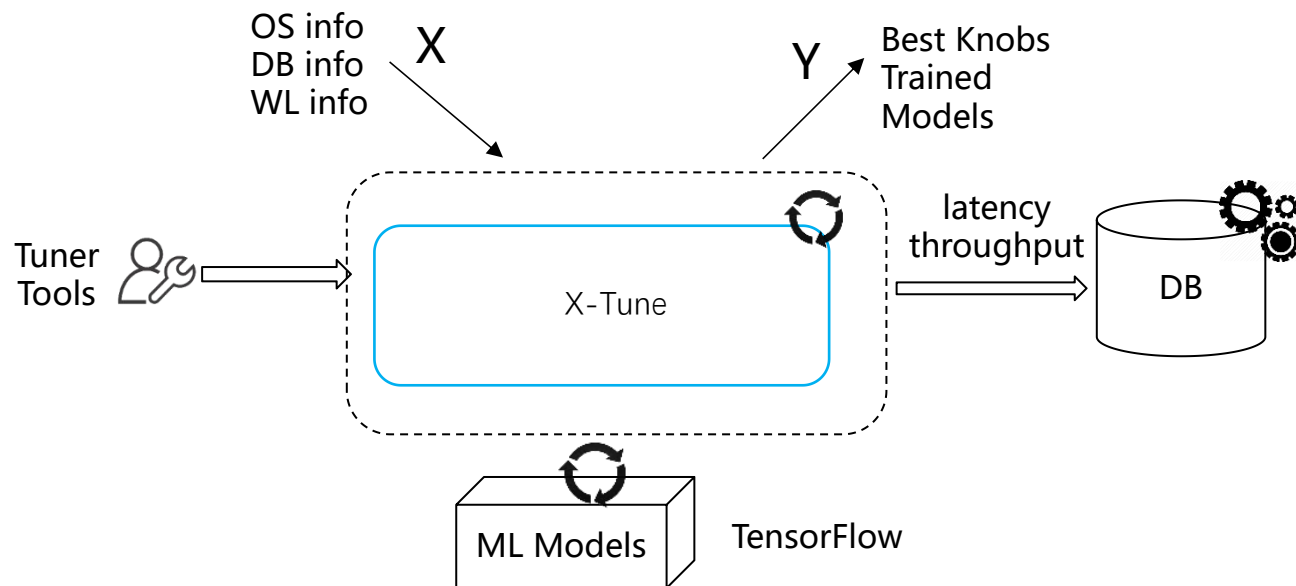
- 历史曲线
- SQL改写
- 慢SQL诊断与建议
- 索引推荐
- 参数推荐
- 参数表现估计
- 趋势预测（AI负载、性能预测）
- 扩容提示（AI存储空间预测）
- 健康指数
- 风险预警
- 综合报告
- DB系统故障检测与预警
- 业务画像
- 入侵威胁检测
-

两部分工作：

- **数据采集及计算加工**：获取数据信息进行分析和处理，为AI运维服务，提供用户多维度展现。
- **智能管理服务**：提供精确预测、智能推荐、优化SQL功能，提升用户体验，增强粘性。

AI4DB: 数据库参数智能调优（已开源openGauss）

openGauss



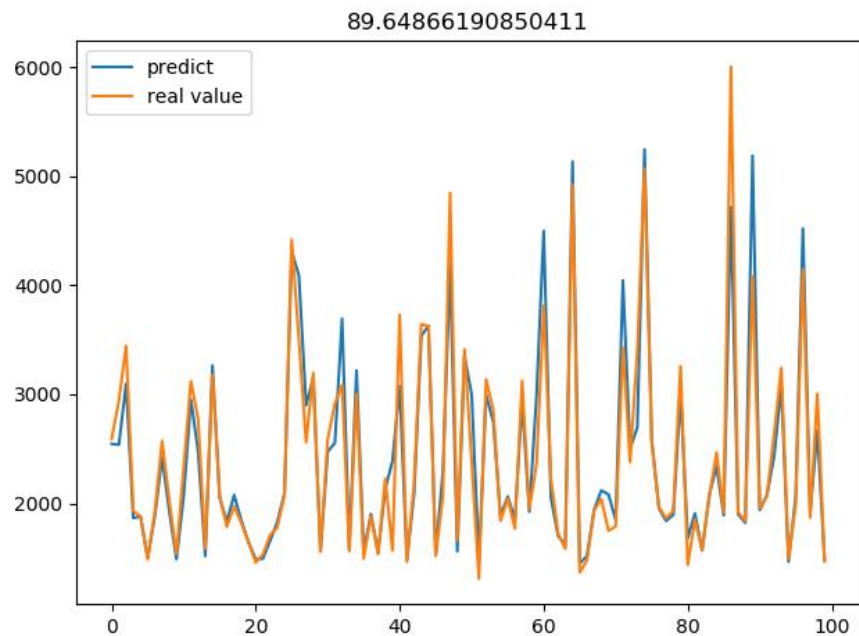
- 利用X-Tune算法，接受数据库的状态信息、作业信息，训练和预测调参结果，直接与数据库交互，迭代式反馈，直至达到SLA期望（TPS、QPS、迭代轮次等）

• 实际效果：

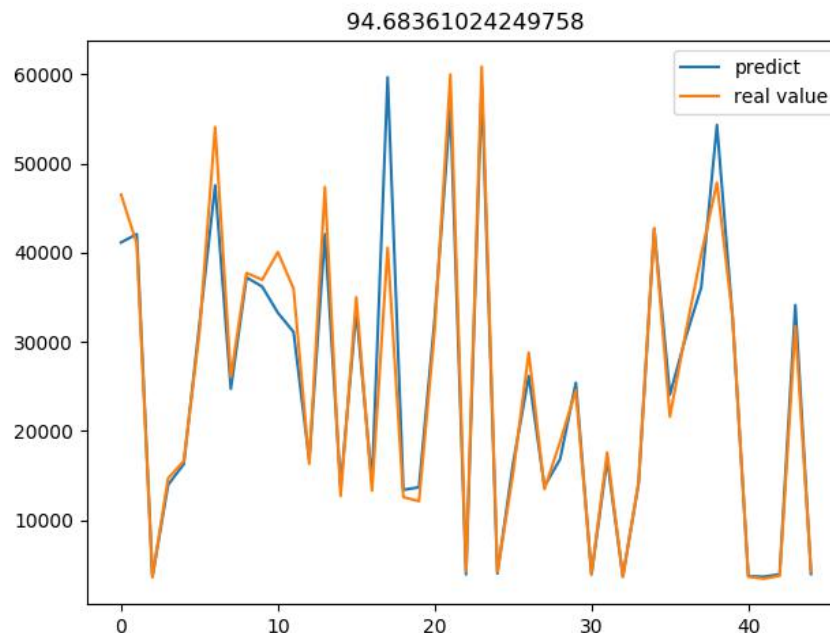
通过基于AI的数据库参数推荐调优可以显著提升系统性能，**基于华为云RDS实例测试验证性能提升比例在30%左右**，已上线华为公有云DAS。

AI4DB: 性能评估模型（已落地公有云DAS）

openGauss



2U4G-SATA Sysbench



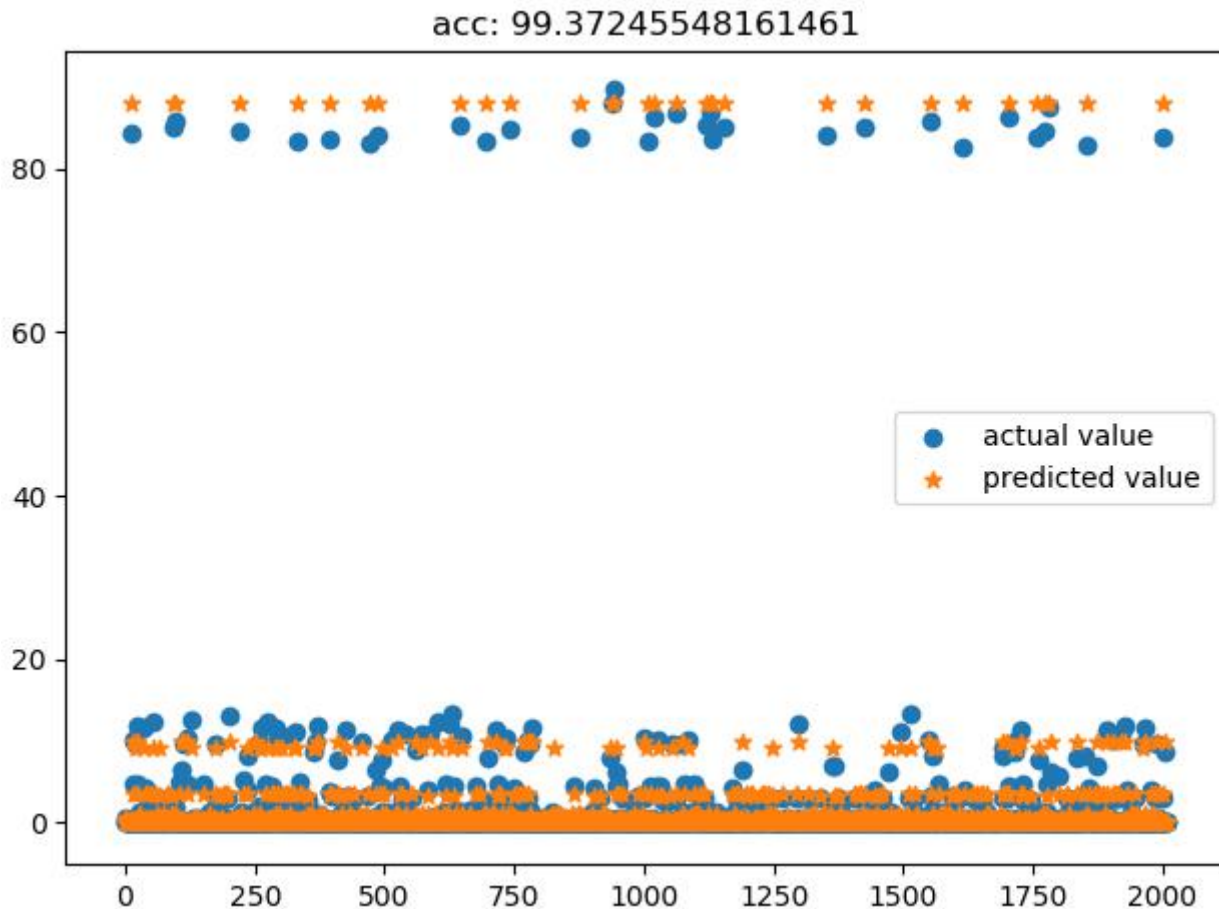
8U16G-SSD TPC-C

- 基于**深度神经网络（DNN）**算法，获取查询语句性能与执行计划、系统状态的之间关系，推测语句在DB状态变化（负载、参数）后的执行时间&资源开销。

- 实际效果：
通过基于**AI**的模型可以准确的预测系统性能，基于华为云**RDS**实例测试验证预测**R2**系数在**87%**以上。

AI4DB: 数据库SQL诊断（已开源openGauss）

openGauss



- 基于**深度神经网络（DNN）**算法，根据历史SQL流水信息，推测待上线SQL语句执行情况，提前审计，给出定性或定量时间估算。

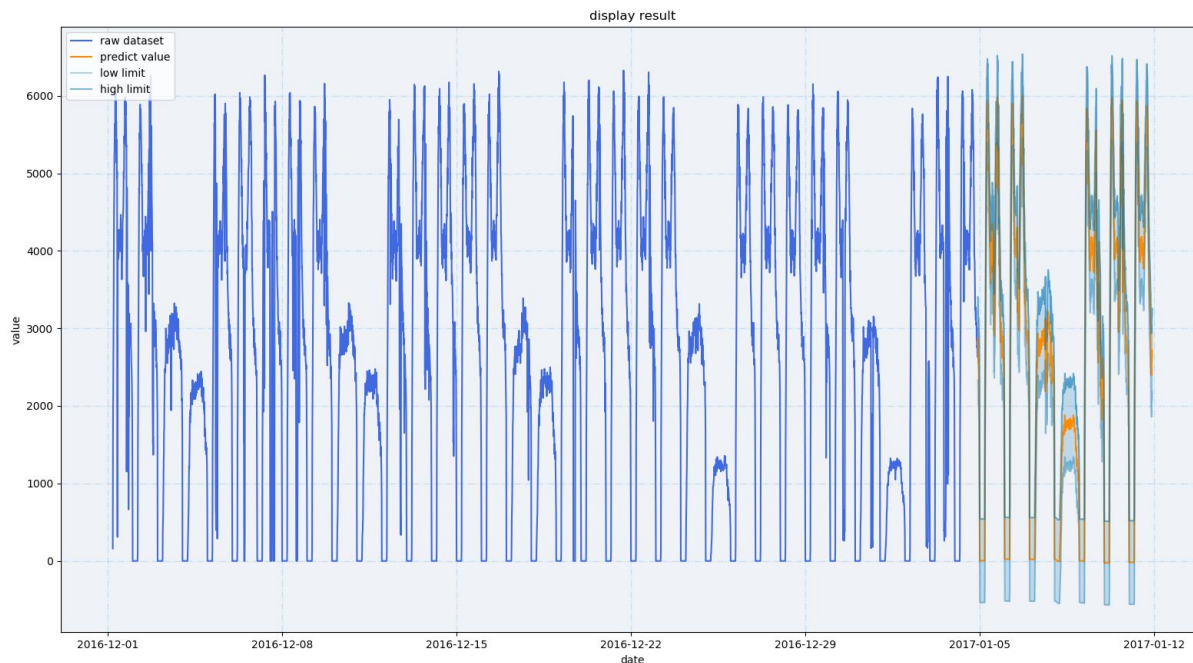
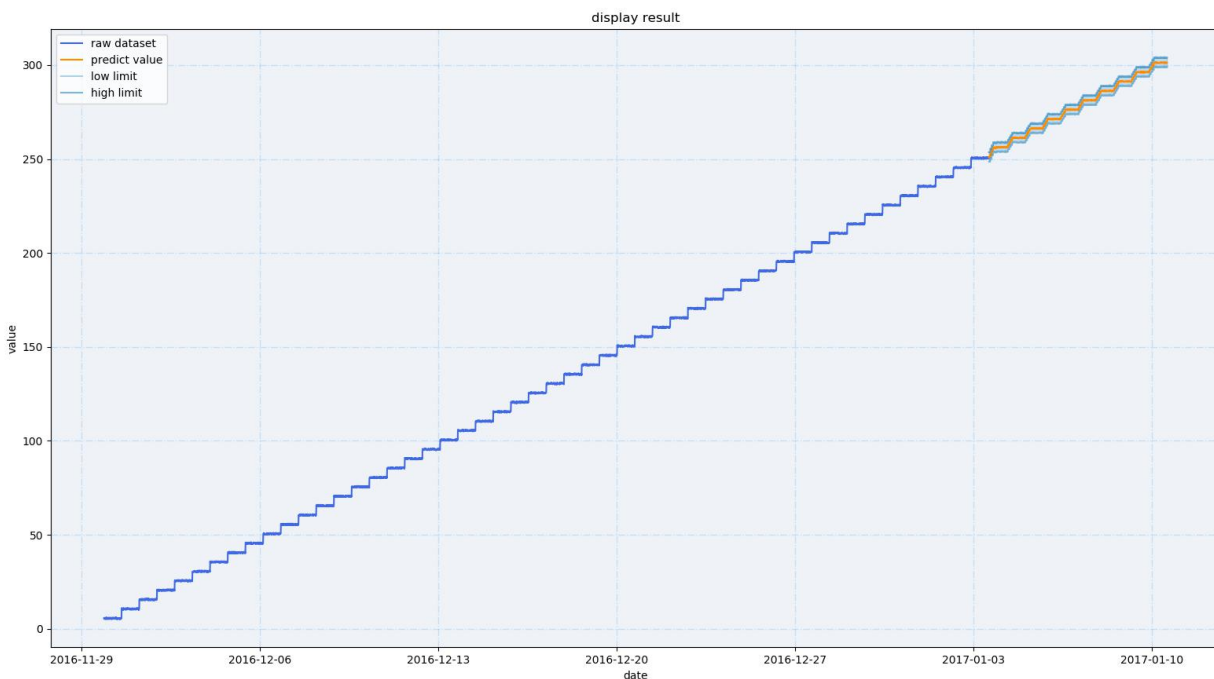
- 实际效果：

基于机器学习算法构建的AI框架定量地预测未知SQL执行时间，满足用户提前识别慢SQL的需求。
基于华为云RDS实例测试验证准确率在90%以上。

AI4DB：数据库负载预测（930开源openGauss）

openGauss

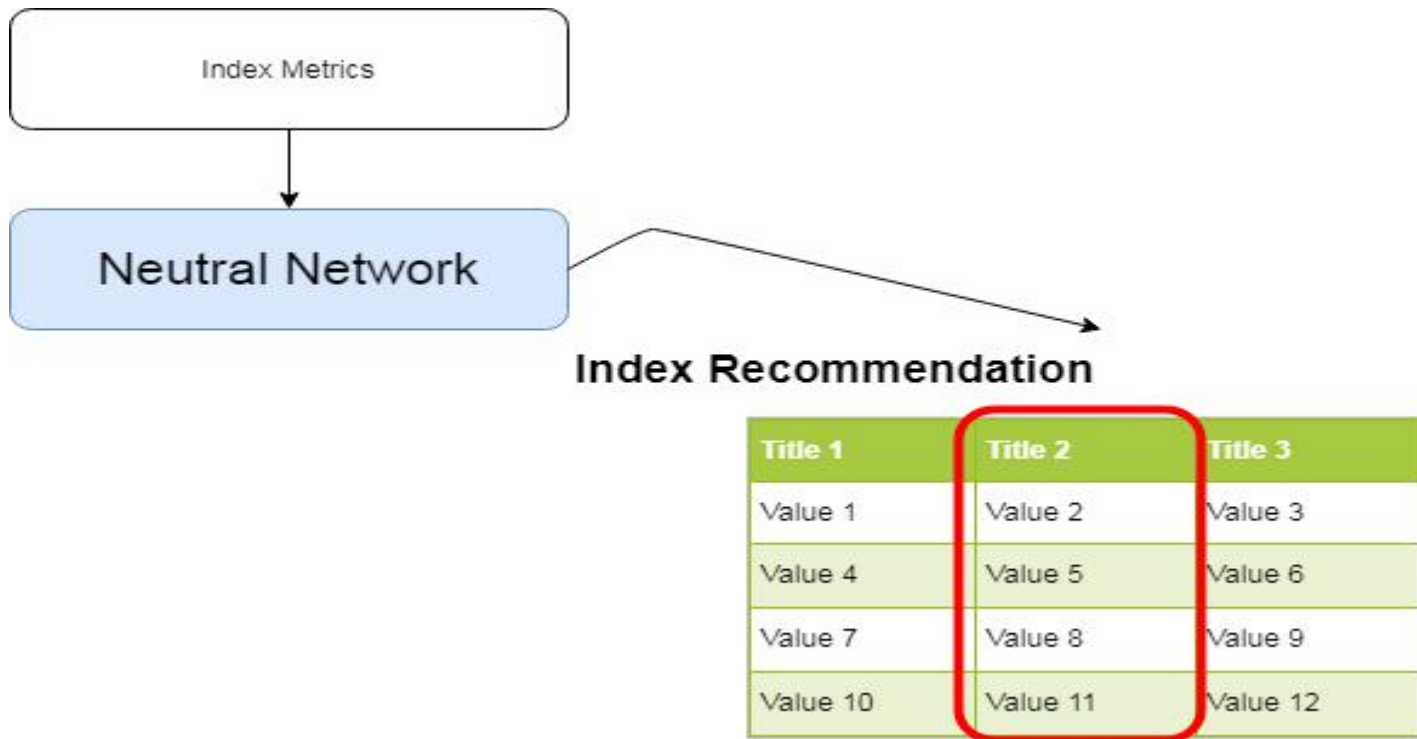
- 基于**深度神经网络（DNN）或统计学习**算法，根据历史时序数据，推测未来一段时间趋势。



- 实际效果：
构建多模式时序预测框架，帮助数据库准确预测未来负载趋势，**基于华为云RDS实例测试验证准确率在85%以上。**

AI4DB: 数据库智能索引推荐 (1230开源openGauss)

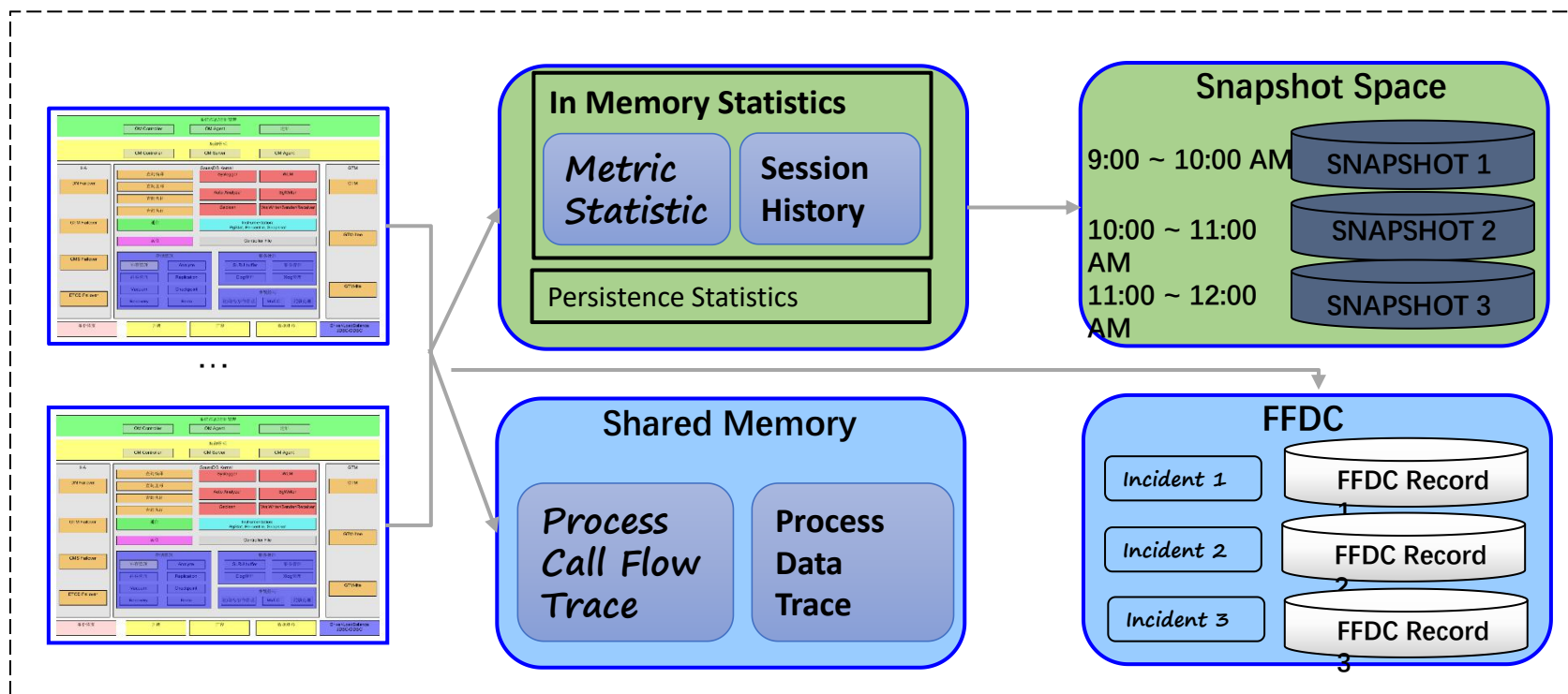
openGauss



- 用户 (DBA) 需要消耗大量时间及熟悉业务情况来设置索引，查询执行效果根据DBA能力相关。
- 通过机器学习方法，动态实现索引的动态推荐，提升作业查询效率及降低DBA运维操作。

AI4DB: 自诊断全景规划图（部分已开源openGauss）

openGauss



功能集合：

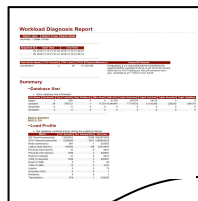
- ✓ Metric Statistic
内核监控KPI
- ✓ Session History
活跃session 历史
- ✓ Snapshot
性能快照
- ✓ Trace
内核trace
- ✓ FFDC
First Failure Data Capture
- ✓ Alarm
告警服务
- ✓ Info Collector
一键式诊断信息收集
- ✓ WDR Report
负载性能报告
- ✓ Automatic Diagnosis
自动负载诊断



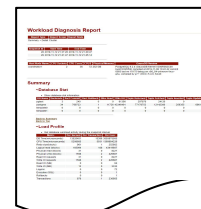
Alarm Service



One-click Info Collector



WDR Report



Automatic Diagnosis

AI4DB：自诊断之WDR报告展示（已开源openGauss）

openGauss

Workload Diagnosis Report

Report Type	Report Scope	Report Node
Summary + Detail	Cluster	

Snapshot Id	Start Time	End Time
1	2019-12-31 15:41:26	2019-12-31 15:42:38
2	2019-12-31 15:43:42	2019-12-31 15:45:18

Host Node Name	CPU%	CPU Cores	CPU Sockets	Physical Memory	GaussDB Version
coordinator1	72	36	2	252 GB	PostgreSQL 9.2.4 (GaussDB Kernel V300R002C00 build 76a388a8) compiled at 2019-12-30 20:22:22 commit 1090 last mr 1758 debug on x86_64-unknown-linux-gnu, compiled by g++ (GCC) 5.4.0, 64-bit

Summary

-Database Stat

- Show database stat information

DB Name	Backends	Xact Commit	Xact Rollback	Blks Read	Blks Hit	Tuple Returned	Tuple Fetched	Tuple Inserted	Tuple Updated	Tup Delete
postgres	12	169	2	1091	127209	322663	29278	7660	137	
template0	0	0	0	0	0	0	0	0	0	
template1	0	0	0	0	0	0	0	0	0	

[Back to Summary](#)
[Back to Top](#)

-Load Profile

- Get database workload activity during the snapshot interval

Metric	Per Second	Per Transaction	Per Exec
DB Time(microseconds)	223279	462063	924126
CPU Time(microseconds)	214555	444009	888018
Redo size(blocks)	10	23	46
Logical read (blocks)	817	1782	3564
Physical read (blocks)	0	1	1
Physical write (blocks)	16	36	71
Read IO requests	0	1	1
Write IO requests	16	36	71
Read IO (MB)	0	0	0
Write IO (MB)	0	0	1
Logons	0	0	1
Executes (SQL)	0	0	1
Rollbacks	0	0	0
Transactions	0	1	2
SQL response time P95	0	0	0
SQL response time P80	0	0	0

-IO Profile

- show io profile

Metric	Read + Write Per Sec	Read Per Sec	Write Per Sec
Database requests	16	0	16
Database (MB)	0	0	0
Database (blocks)	16	0	16
Redo requests			7
Redo (MB)			0

[Back to Summary](#)
[Back to Top](#)

Report Details

- [SQL Statistics](#)
- [Cache IO Stats](#)
- [Object stats](#)
- [SQL Detail](#)

SQL Statistics

- [SQL ordered by Elapsed Time](#)
- [SQL ordered by CPU Time](#)
- [SQL ordered by Rows Returned](#)
- [SQL ordered by Row Read](#)
- [SQL ordered by Executions](#)
- [SQL ordered by Physical Reads](#)
- [SQL ordered by Logical Reads](#)

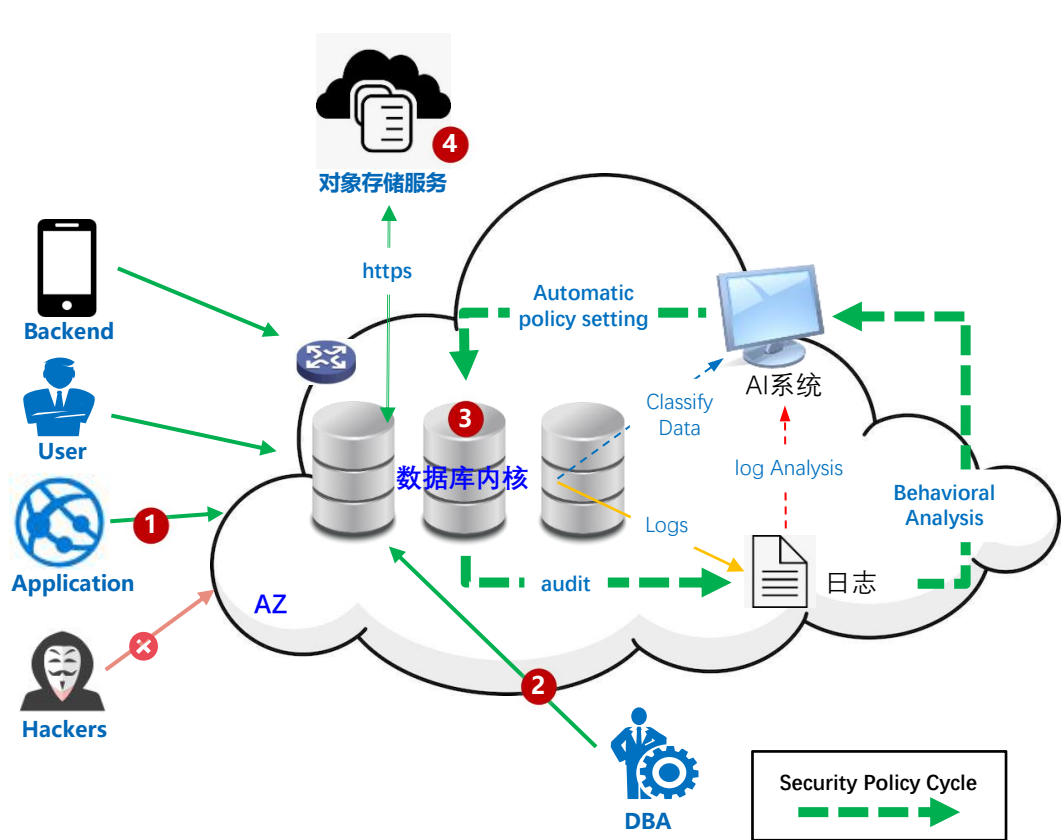
-SQL ordered by Elapsed Time

- SQL ordered by Elapsed Time
- List top 200 records
- Avg Elapse Time(us) represents the average elapsed time of sql between two snapshots

Unique SQL Id	Node Name	Total Elapse Time(us)	User Name	Calls	Avg Elapse Time(us)	Min Elapse Time(us)	Max Elapse Time(us)	Returned Rows	Tuples Read	Tuple
1676591045	coordinator1	179884	fangb	1	179884	167149	179884	421	4342	
3402541842	coordinator1	162656	fangb	1	162656	162656	163246	421	2786	
2979884330	coordinator1	156456	fangb	1	156456	136126	156456	421	4342	
3734711474	coordinator1	151838	fangb	1	151838	120278	151838	655	1541	
3739274279	coordinator1	150018	fangb	1	150018	150018	173841	241	2786	
245678917	coordinator1	145833	fangb	1	145833	132966	145833	842	3097	
1788162549	coordinator1	137360	fangb	1	137360	137360	141105	288	4016	
2323021088	coordinator1	134868	fangb	1	134868	134868	139055	294	4016	
160532468	coordinator1	124122	fangb	1	124122	124122	138384	482	1541	
4108376575	coordinator1	115154	fangb	1	115154	115154	122956	294	4016	
2372904322	coordinator1	115152	fangb	1	115152	115152	118884	294	4016	
1000539724	coordinator1	112898	fangb	1	112898	109513	112898	842	3097	
1903216855	coordinator1	112181	fangb	1	112181	86563	112181	0	3097	
82660962	coordinator1	108048	fangb	1	108048	105062	108048	588	2771	
2591130304	coordinator1	106922	fangb	1	106922	106922	110042	576	2771	
2196353822	coordinator1	105685	fangb	1	105685	102287	105685	588	2771	

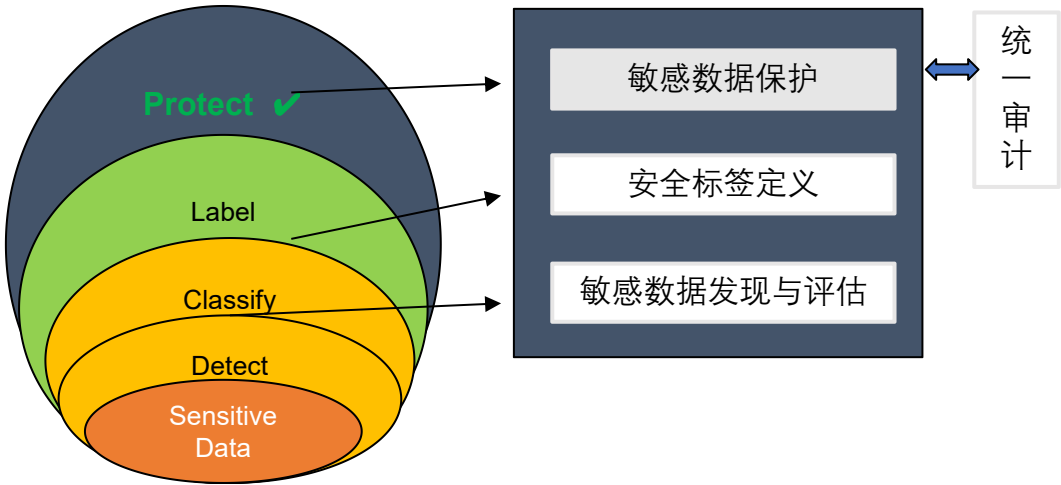
AI4DB: 自治安全架构（预研中）

- ❑ 安全功能内置：绑定鲲鹏生态，构建包含细粒度访问控制、数据脱敏、全程加密、统一审计、密钥管理等能力的内在安全防御体系；
- ❑ 智能行为分析：基于轩辕AI底座，进行数据库网络访问、日志内容、数据访问行为特征分析，提供实时AI自主安全韧性能力；



(1)访问接入；(2)DBA运维管理；(3)数据库内部运行；(4)数据存储服务；

自治安全框架



基于价值的数据库安全自治策略图

关键特征	关键技术
敏感数据发现	敏感数据发现 数据动态&静态脱敏 内置安全策略及自主化管理
用户行为审计与分析	用户行为记录框架 基于AI的行为特征提取与异常感知
数据库安全评估	安全攻击模式库 异常访问行为分析

AI4DB: AI 优化器全景规划图（预研中）

传统优化器关键问题

行数估算问题

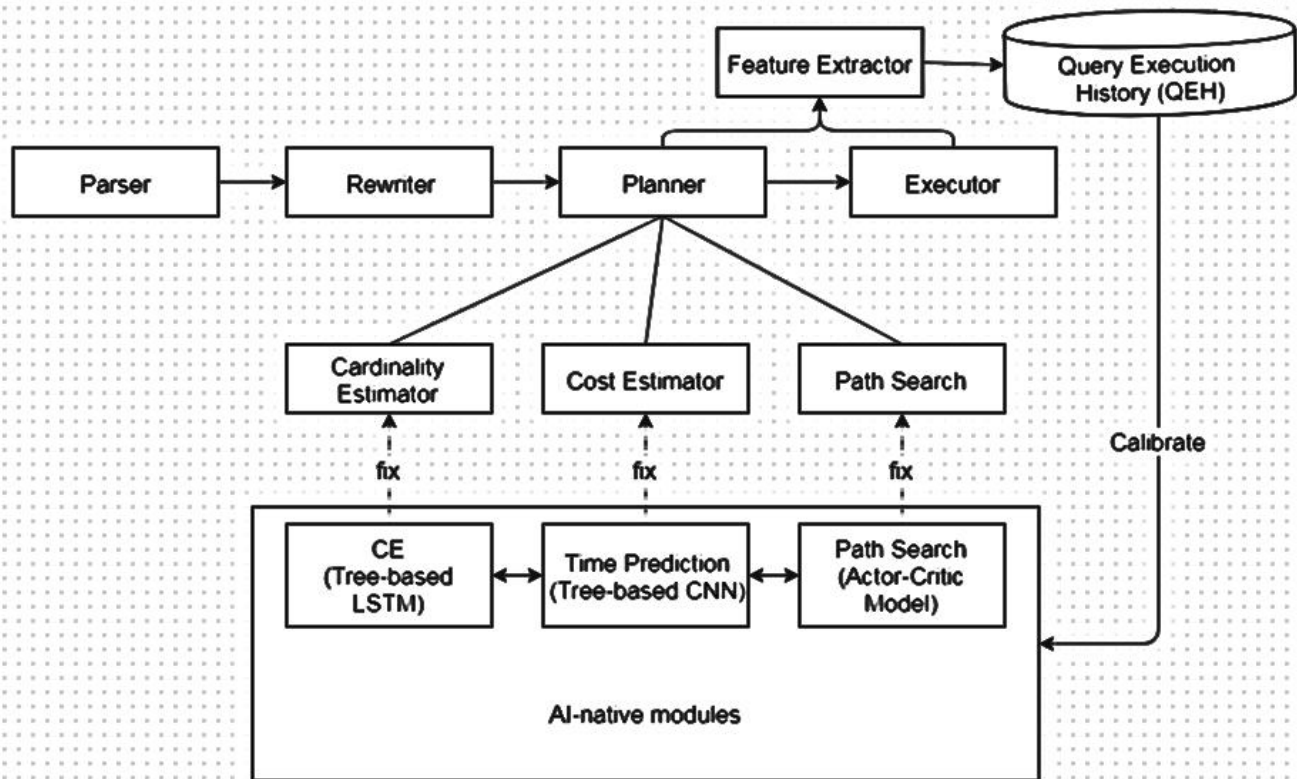
1. 多列相关性问题
2. 表达式统计信息问题
3. 中间结果集统计信息问题
4. 数据分布问题

代价模型问题

1. 算子间代价失衡问题
2. 算子并行化代价问题
3. 分布式计划并行代价问题
4. 代价-时间换算问题
5. 跨平台泛化问题

路径搜索问题

1. 搜索空间与搜索代价的平衡 (NP问题)
2. 局部最优解与全局最优解问题



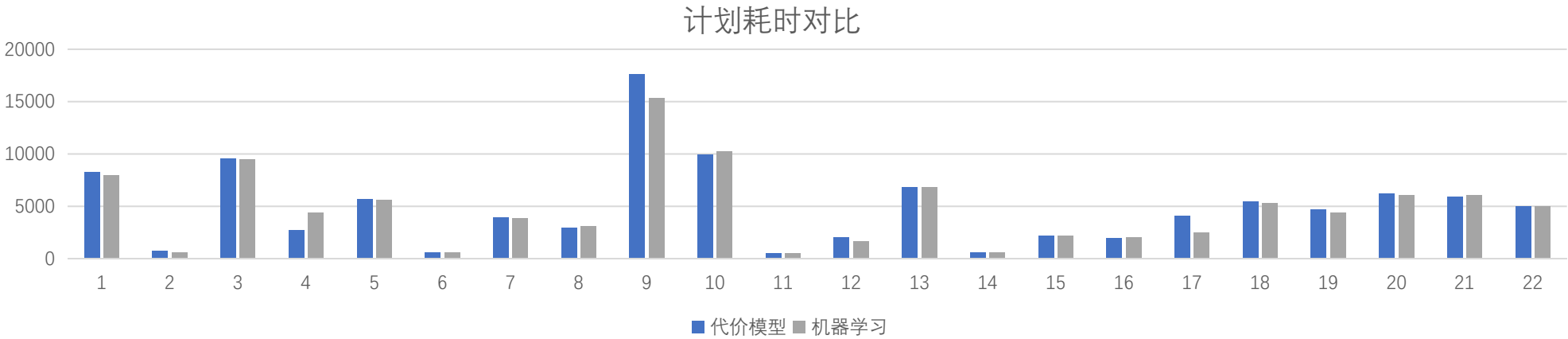
- 解决传统优化器的行数估算、代价模型、路径搜索等经典问题。
- 推动优化器朝智能化方向发展，实现自适应、自优化。

AI4DB: AI 优化器之算子代价模型修正（已预研）

- 对CStore Scan, CStore Index Scan, Vec Hashjoin三个算子的代价估算使用机器学习建模, 并且使用实际执行时间训练模型, 实际输出 startup/total time, 通过简单的线性换算转换成cost。

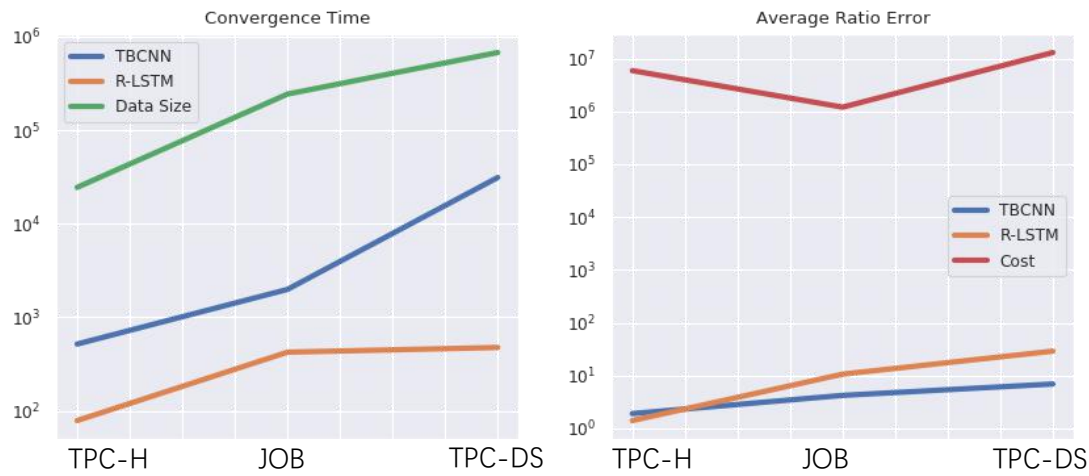
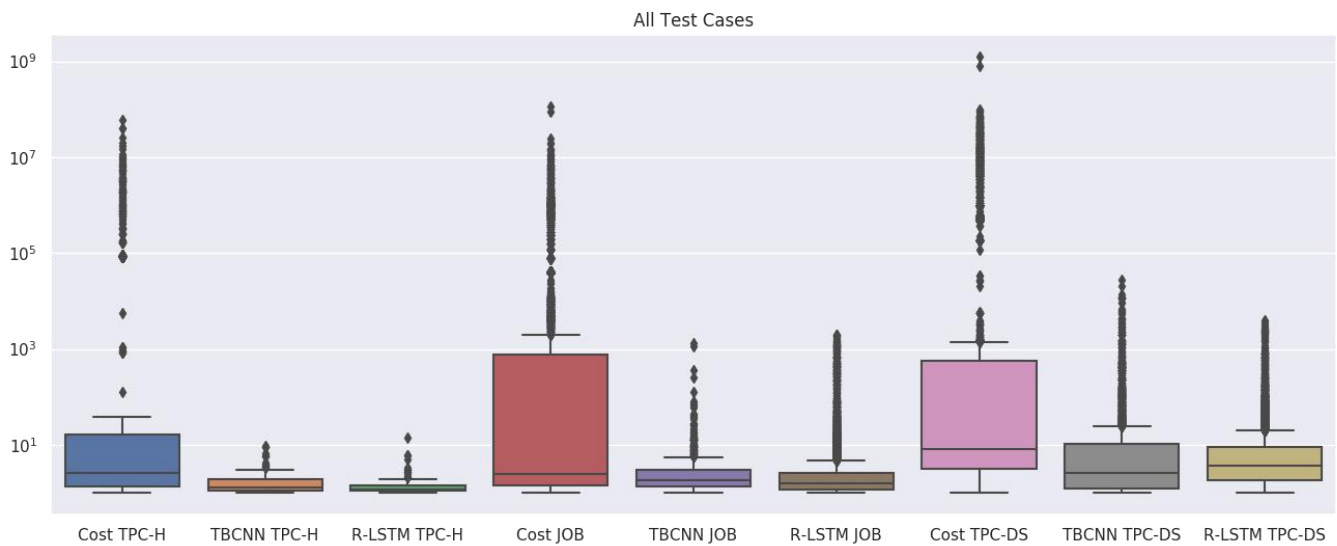
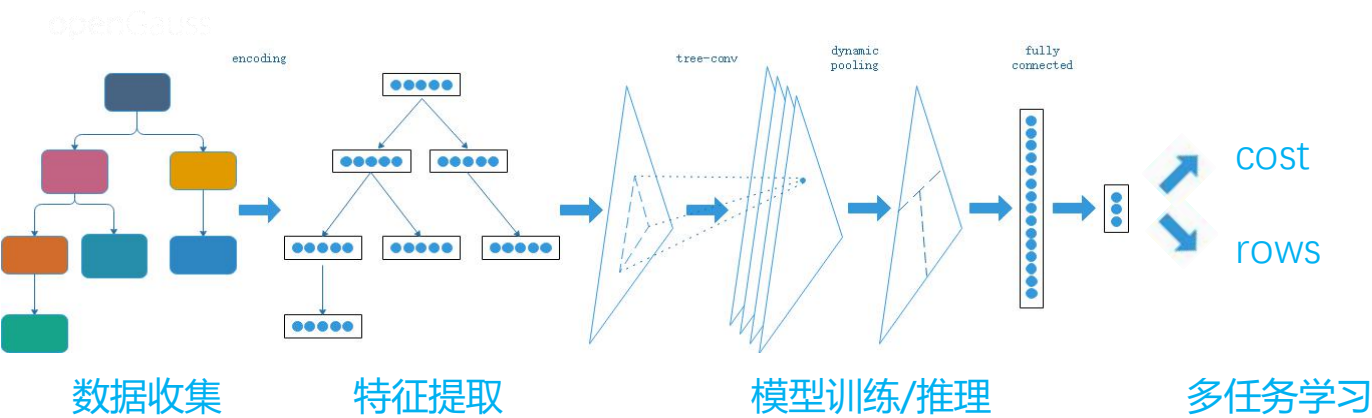
列存Scan算子特征	
input tuples	总元组数
output tuples	优化器估计的输出总元组数
ncu per col	每列的CU（压缩单元）个数
nlr ncols	过滤条件相关的所读列数
nlr avg size	过滤条件相关的所读列平均CU大
lr ncols	非过滤条件相关的所读列数
lr avg size	非过滤条件相关的所读列平均CU大小
max correlation	过滤条件相关列中数据相关性最大列的相关性（时间限制当前未加入）
max correlated col selectivity	过滤条件相关列中数据相关性最大列对应的选择率（时间限制当前未加入）

Join算子特征	
Join type	数据预处理时分为三个正交维度进行one-hot编码
DOP	算子并行度
Out tuples	Join输出的总元组数
n cond	总条件个数
Inner startup	内表启动代价
Inner total	内表总代价
Inner rows	内表输出行数
Inner width	内表宽度
Outer startup	外表启动代价
Outer total	外表总代价
Outer rows	外表输出行数
Outer width	外表宽度



- 代价模型修正后TPCH部分场景执行性能得到提升： Q2提升32%, Q9提升15%, Q12提升20%, Q17提升67%。

AI4DB: AI 优化器之基于在线学习的查询性能预测（已开源openGauss）



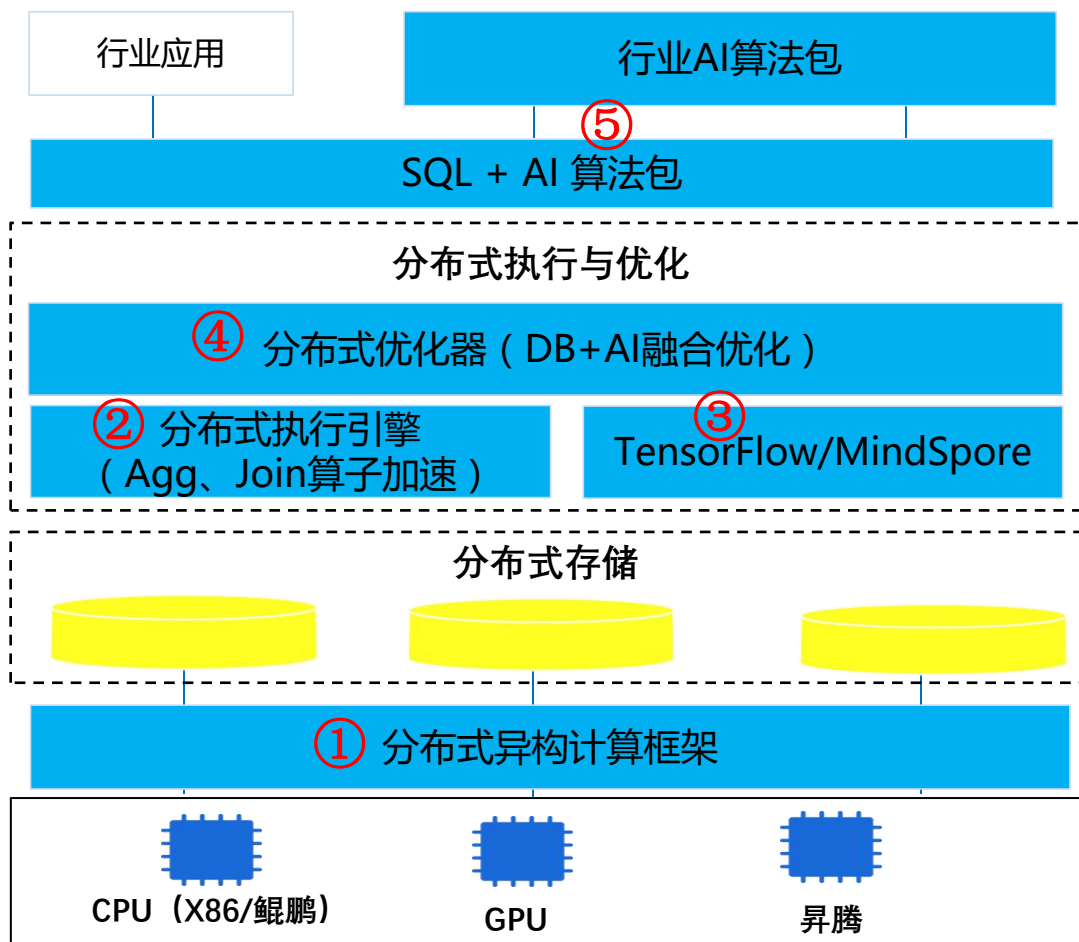
- Cost为使用代价线性拟合预测时间，拟合参数通过历史数据回归获得。
- 深度学习模型比Cost线性模型在预测准确率上有显著提升。
- R-LSTM模型收敛速度比TBCNN模型显著提升，但在较复杂场景下(如TPC-DS)容易过拟合。
- TBCNN在简单场景下和R-LSTM准确率比较接近，在复杂场景下表现出较强的泛化能力，但训练速度过慢。

	TPC-H			JOB			TPC-DS		
	Cost	TBCNN	R-LSTM	Cost	TBCNN	R-LSTM	Cost	TBCNN	R-LSTM
训练耗时(s)		590	79		1990	425		31204.5	477
平均准确率	5.90E+06	1.9(1.8)	1.4(1.3)	1.20E+06	4.2(4.0)	10.6(4.1)	1.30E+07	6.9(4.3)	28.9(7.8)
<10准确率	74%	99%	99%	62%	93%	91%	53%	74%	74%
<2准确率	42%	77%	91%	43%	53%	66%	12%	44%	29%

PART 3 DB4AI

DB4AI：基于AI芯片，释放性异构算力，降低AI开发成本，实现训练推理验证加速一体化

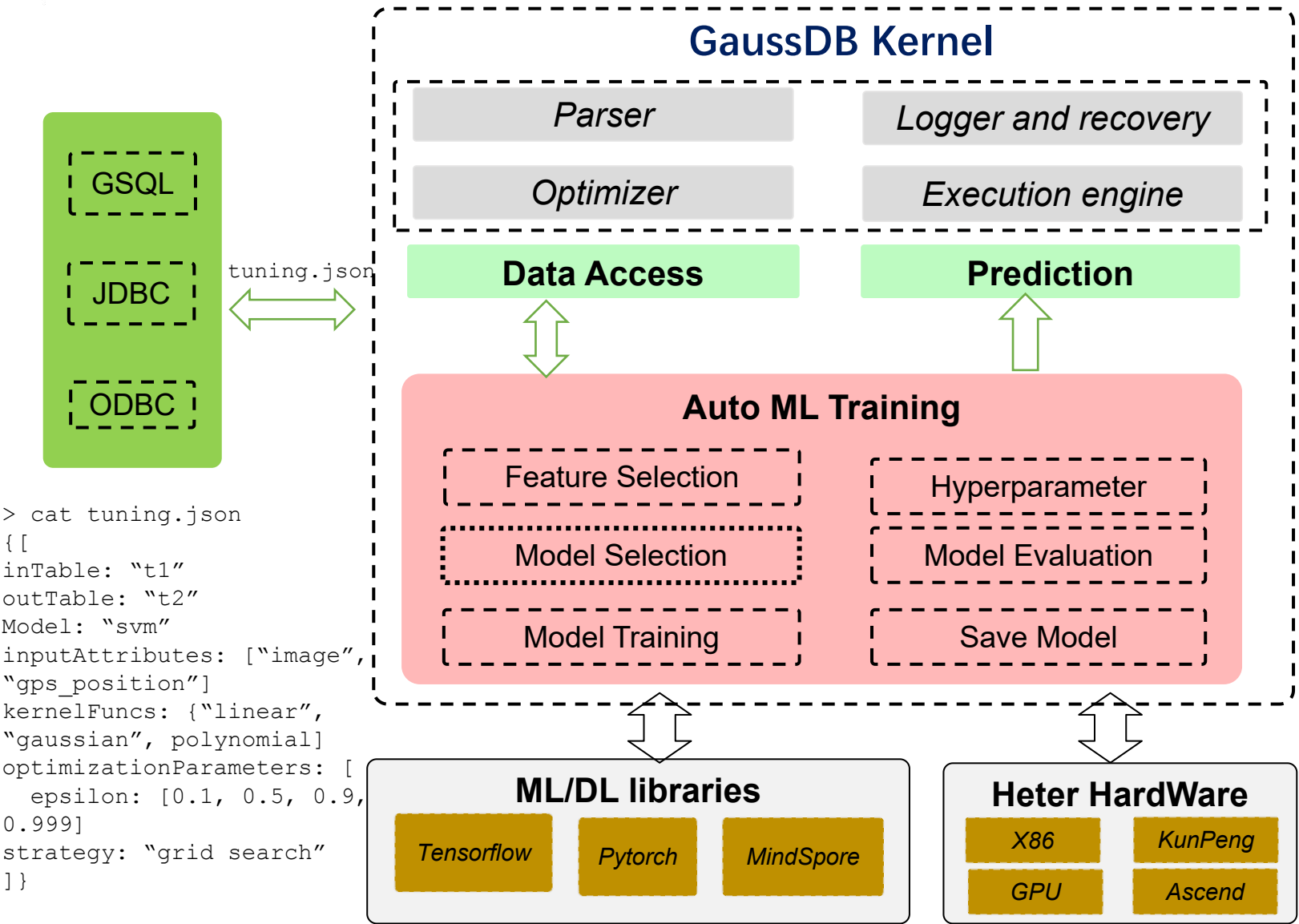
DBMind AI计算平台



关键技术方向：

1. 孵化基于CPU、GPU、昇腾芯片的异构计算框架，实现基于异构芯片的数据统一封装与最优化调度。
2. 基于GPU、昇腾芯片实现数据库重负载算子Agg、Join等的加速。
3. 基于Gauss内核DB4AI技术，结合Tensorflow/MindSpore机器学习框架，在数据库内部实现CNN、DNN等神经网络算法，并探索基于昇腾芯片的对接与加速。
4. 探索数据库与机器学习算法的融合优化技术，利用数据库优化器索引、剪枝等技术实现机器学习模型训练与推理过程的加速。
5. 探索基于数据库的内置行业AI算法包，实现数据分析由“DB+BI”向“DB+AI”的转变。

DB4AI：全流程AI，构筑数据库内端到端DB4AI系统，无需人工参与（预研中）



```
> cat tuning.json
{[
  inTable: "t1"
  outTable: "t2"
  Model: "svm"
  inputAttributes: ["image",
    "gps_position"]
  kernelFuncs: {"linear",
    "gaussian", polynomial]
  optimizationParameters: [
    epsilon: [0.1, 0.5, 0.9,
    0.999]
  strategy: "grid search"
]}
```

核心竞争力分析:

- 1. **自动特征工程:**从相关的数据表中自动提取有用且有意义的特征，减少了特征工程所需的时间。将周级的处理时间缩短至天级。
- 2. **模型最优选择:**在多个模型中，选择出最适合本数据的模型。
- 3. **超参数优化:**选出合适模型后，并且能够设定好它的最优参数。
- 4. **优化算法选择:**自动地选择一个优化算法（如SGD、L-BFGS、GD等），以便能够达到效率和精度的平衡。
- 5. **支持深度学习:**支持主流深度学习框架。并自动地选择一个优化算法，以便能够达到效率和精度的平衡。
- 6. **模型管理:**记录模型准确度等相关信息。增加模型生命周期管理，支持增量训练，模型更新，支持模型导入和导出。



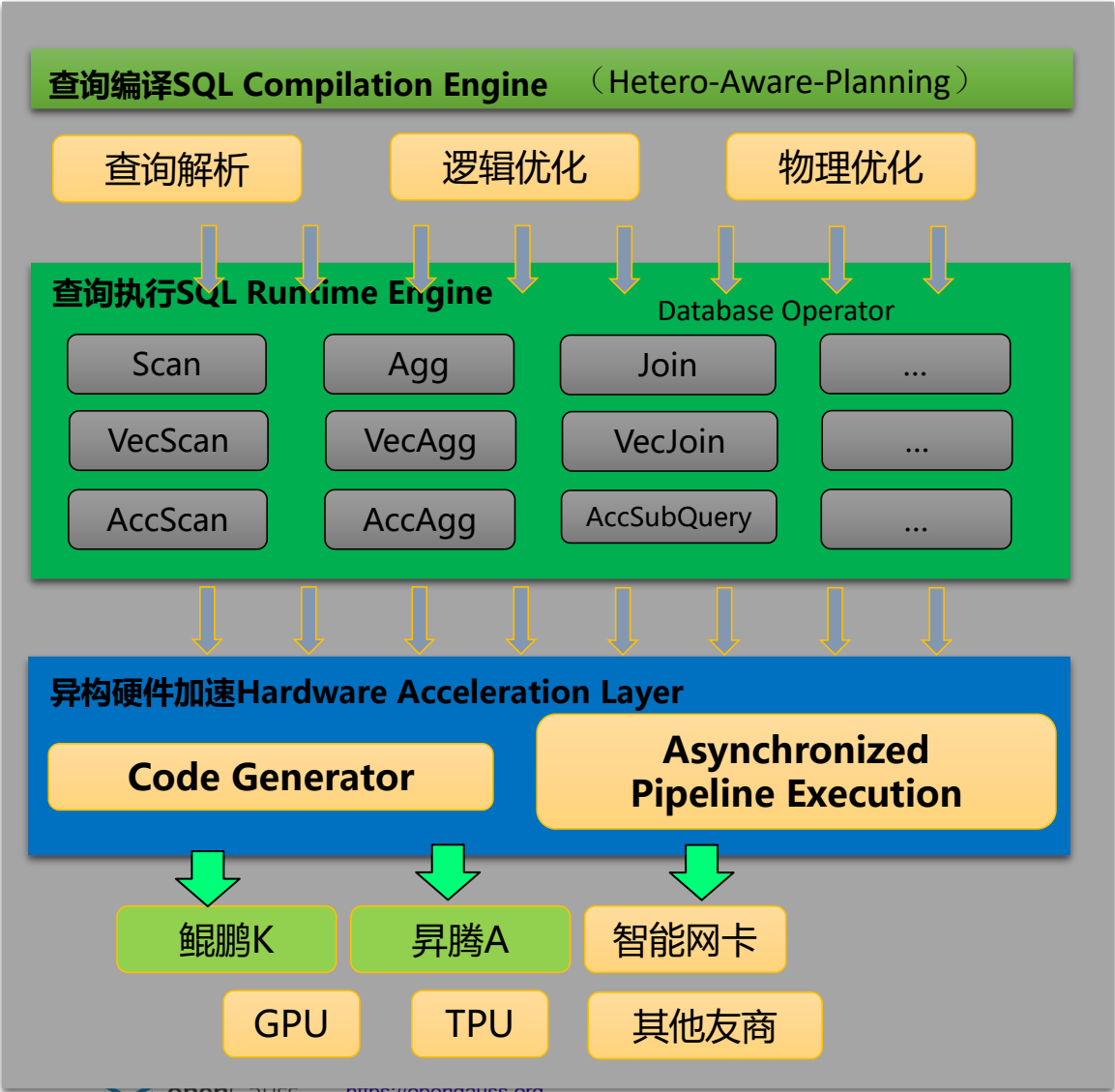
DB4AI：数据库内支持70+AI算法，满足用户AI训练和预测（计划1230开源）

当前已支持70+通用AI算法，兼容MADlib，支持新增算法N个。

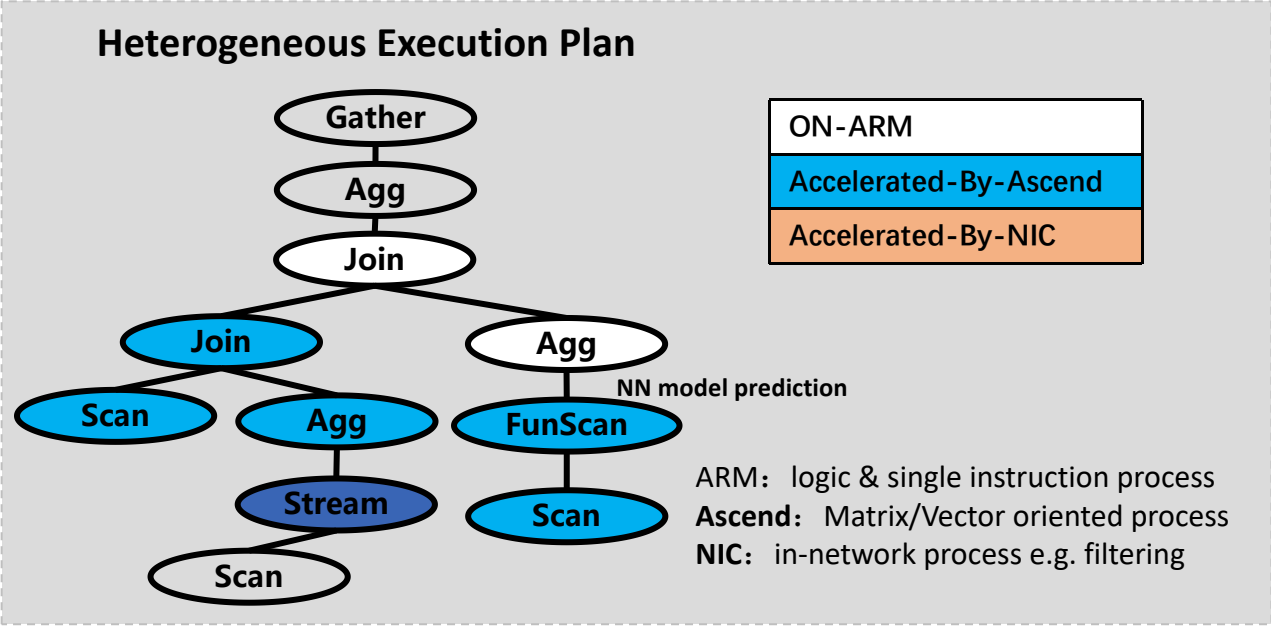
算法大类	算法名称	算法名称(中文)	特性描述	应用场景	Functions(示例)	UDF示例(已定义好，不需用户创建)
regression	Logistic Regression	逻辑回归	逻辑回归，二分类	逻辑回归主要的应用场景：寻找危险因素：例如寻找某疾病的危险因素；金融商业机构需要对企业进行评估等	logreg_train() logreg_predict() logreg_predict_prob() coxph_train() cox_zph() coxph_predict() elastic_net_train() elastic_net_binomial_prob() elastic_net_predict()	CREATE OR REPLACE FUNCTION madlib.logreg_train (source_table VARCHAR, out_table VARCHAR, ...)
	Cox Proportional Hazards Regression	Cox比例风险回归		该模型以生存结局和生存时间为应变量，可同时分析众多因素对生存期的影响，能分析带有截尾生存时间的资料，且不要求估计资料的生存分布类型。由于上述优良性质，该模型弹性回归是岭回归和套索回归的混合技术，它同时使用 L2 和 L1 正则化。当有多个相关的特征时，弹性网络是有用的。套索回归很可能随机选择其中一个，而弹性回归很可能都会		CREATE OR REPLACE FUNCTION madlib.coxph_train(source_table VARCHAR, output_table VARCHAR, ...)
	Elastic Net Regularization	弹性网络回归	弹性网络回归算法			CREATE OR REPLACE FUNCTION madlib.elastic_net_train (tbl_source TEXT, tbl_result TEXT, ...)
	Generalized Linear Models	广义线性模型	广义线性模型（二项分布，伽马分布，高斯分布，逆高斯分布，泊松分布）	在一些实际问题中，变量间的关系并不都是线性的，这种情况就应该用曲线去进行拟合	glm() glm_predict()	CREATE OR REPLACE FUNCTION madlib.glm(source_table varchar, model_table varchar, ...)
	Marginal Effects	边际效应			margins() margins_logreg() margins_mlogreg()	CREATE AGGREGATE madlib.marginal_logreg(BOOLEAN, DOUBLE PRECISION[], ...)
	Multinomial Regression	多类回归		如果目标类别数超过两个，这时就需要使用多类回归，如疗效可能是“无效”“显效”“痊愈”三类	multinom() multinom_predict()	CREATE OR REPLACE FUNCTION madlib.mlogreg_train (source_table VARCHAR , output_table VARCHAR ...)
	Ordinal Regression	序数回归	序数分类/排序学习	在统计学中，序数回归是一种用于预测序数变量的回归分析，即其值存在于任意范围内的变量，其中只有不同值之间的相对排序是显著的。它可以被认为是介于回归和分类之间	ordinal() ordinal_predict()	CREATE OR REPLACE FUNCTION madlib.ordinal(source_table varchar, model_table varchar, ...)
	Clustered Variance	聚类方差	聚类方差	Clustered Variance模块调整聚类标准误差。例如，将一个数据集复制100次，不应该增加参数估计的精度，但是在符合独立同分布假设（Independent Identically	clustered_variance_linreg() clustered_variance_logreg() clustered_variance_mlogreg()	CREATE OR REPLACE FUNCTION madlib.clustered_variance_linreg (source_table TEXT, out_table TEXT, ...)
Robust Variance	稳健方差	稳健方差	Robust Variance模块中的函数用于计算线性回归、逻辑回归、多类逻辑回归和Cox比例风险回归的稳健方差（Huber-White估计）。它们可用于计算具有潜在噪声异常值的数据集	robust_variance_linreg() robust_variance_mlogreg() robust_variance_coxph()	CREATE OR REPLACE FUNCTION madlib.robust_variance_coxph(model_table VARCHAR, output_table VARCHAR ...)	
Support Vector Machines(SVM)	支持向量机	支持向量机，可用于线性/非线性分类，也可以用于回归	用于文本和超文本的分类，用于图像分类，比起传统的查询优化方案，支持向量机能够获取明显更高的搜索准确度。这同样也适用于图像分割系统	svm_regression() svm_classification() svm_one_class()	CREATE OR REPLACE FUNCTION madlib.svm_regression(source_table text, ...)	
Linear Regression	线性回归	线性回归	应用广泛，例如经济学，金融学等	linreg_train() linreg_predict()	CREATE OR REPLACE FUNCTION madlib.linreg_train(source_table VARCHAR, out_table VARCHAR, ...)	
Tree Methods	Decision Tree	决策树	决策树模型	最为广泛的归纳推理算法之一，处理类别型或连续型变量的分类预测问题，可以用图形和if-then的规则表示模型，可读性较高	tree_train() tree_predict() tree_display()	CREATE OR REPLACE FUNCTION madlib.tree_train(training_table_name TEXT, output_table_name TEXT, ...)
	Random Forest	随机森林	随机森林是一类专门为决策树分类器设计的组合方法。它组合多棵决策树作		forest_train() forest_predict() get_tree()	CREATE OR REPLACE FUNCTION madlib.forest_predict(model TEXT, source TEXT, ...)
Other Methods	Conditional Random Field (CRF)	条件随机场	条件随机场（CRF）是一种判别的，无向概率的图形模型。线性链CRF是通过计算概率率来进行分类，可以用来处理多分类问题	在分词、词性标注和命名实体识别等序列标注任务中取得了很好的效果	crf_train_fgen() lincrf_train() crf_test_fgen()	CREATE OR REPLACE FUNCTION madlib.compute_lincrf("source" TEXT, "sparse_R" TEXT, ...)
	Naive Bayes	朴素贝叶斯		垃圾邮件过滤器	create_nb_prepared_data_tables() create_nb_prepared_data_tables() create_nb_classify_view()	CREATE OR REPLACE FUNCTION madlib.create_nb_prepared_data_tables("trainingSource" VARCHAR, "trainingClassColumn" VARCHAR, ...)
	Neural Networks	神经网络	使用的多层感知机(MLP)进行回归和分类	拥有广泛的应用场景，譬如语音识别、图像识别、机器翻译等等。在模式识别的领域中算是标准监督学习算法，并在计算神经网络及并行分布式处理领域中，持续成为被研究的课题	mlp_classification() mlp_regression() mlp_predict()	CREATE FUNCTION madlib.mlp_igd_transition(state DOUBLE PRECISION[], ind_var DOUBLE PRECISION[], ...)
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过多层节点的连接来学习复杂的非线性关系。广泛应用于图像识别、语音识别、机器翻译等。						
近邻分类方法通过计算每个训练样本与测试样本之间的距离，找到最近邻样本，从而进行分类。广泛应用于文本分类、图像分类、推荐系统等。						
支持向量机(SVM)是一种二分类模型，它通过寻找一个超平面来最大化两类数据之间的间隔。广泛应用于文本分类、图像分类、支持向量回归等。						
支持向量回归(SVR)是一种回归模型，它通过寻找一个超平面来最小化数据点到超平面的距离。广泛应用于时间序列预测、回归分析等。						
决策树是一种分类模型，它通过递归地分割数据集来构建决策树。广泛应用于分类问题、数据挖掘等。						
随机森林是一种集成学习方法，它通过构建多棵决策树并综合它们的预测结果来提高模型的泛化能力。广泛应用于分类、回归问题。						
条件随机场(CRF)是一种判别式模型，它通过联合优化所有变量的概率来建模序列标注任务。广泛应用于自然语言处理中的分词、词性标注等。						
朴素贝叶斯是一种基于概率论的简单模型，它假设特征之间相互独立。广泛应用于垃圾邮件过滤、文本分类等。						
神经网络是一种模仿人脑结构的计算模型，它通过						

DB4AI：异构硬件计算加速架构图

异构硬件下XuanYuan内核



异构硬件下的SQL执行计划

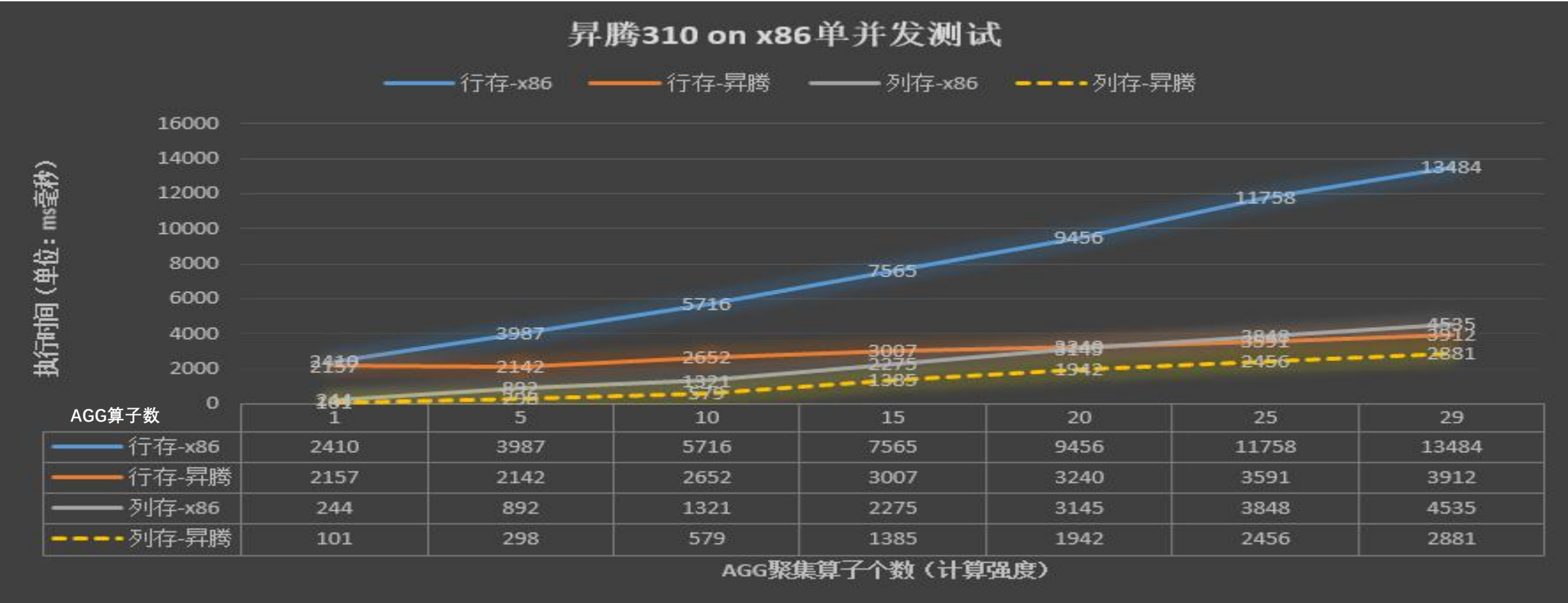


关键挑战	内核领域
Uniformed-Acceleration-Interface 数据库内核层实现统一的硬件加速接口	内核架构
Hetero Aware SQL Planning 基于代价加速设备X的选取、负载均衡、作业编排	优化器
Asynchronized Pipeline Execution 加速设备X的与Host主存之间传输连续	执行器 存储引擎

DB4AI：昇腾加速 - 单线程测试（昇腾310+x86）（已预研）

数据	行存表，800w数据，使用AGG算子数量来模拟计算量大小
测试环境	2288 v5服务器，2*昇腾310卡 48核 Intel(R) Xeon(R) Platinum 8160 CPU @ 2.10GHz 512GB内存

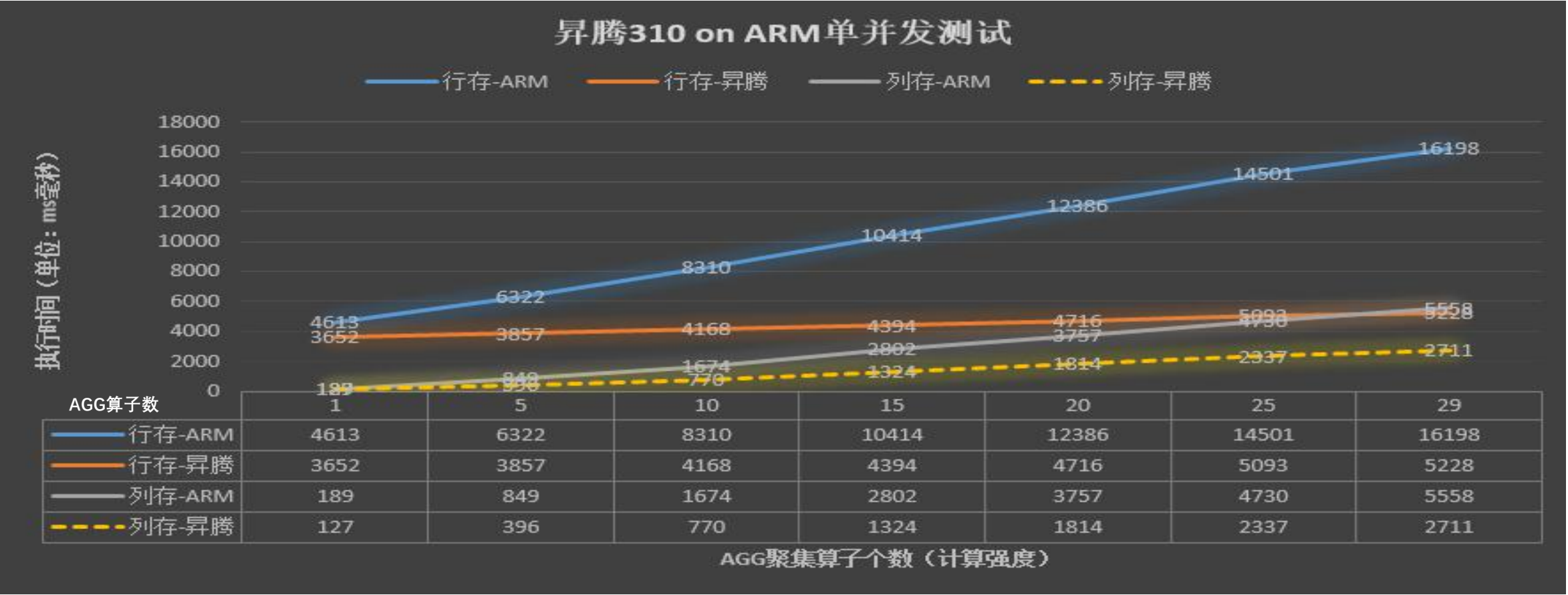
- 行存性能提升245%
- 列存性能提升57%



DB4AI：昇腾加速 - 单线程测试（昇腾310+ARM）（已预研）

数据	行存表，800w数据，使用AGG算子数量来模拟计算量大小
测试环境	2288V5泰山服务器，1*昇腾310卡 96核 ARM机器，512GB内存

- 行存性能提升210%
- 列存性能提升105%



PART 4 规划和路标

openGauss路标 (DBMind = AI4DB & DB4AI)

openGauss

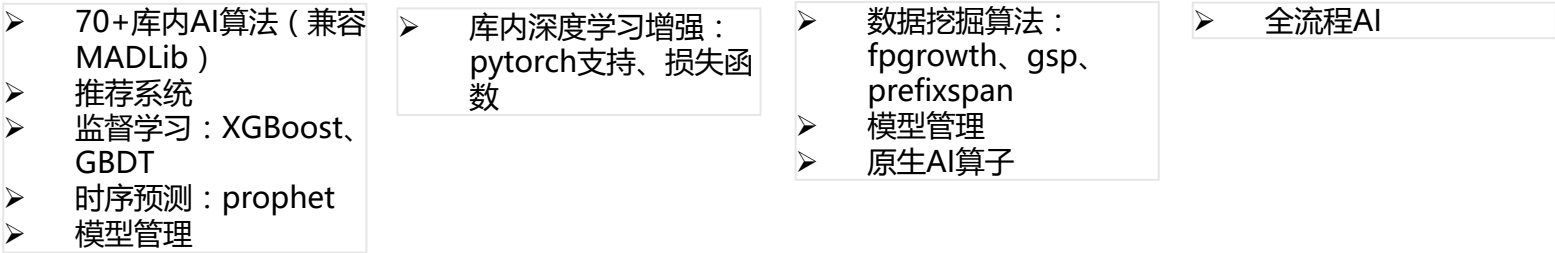
AI4DB(DB mind): 利用AI技术实现对RDBMS的自调优、自安全、自运维，做到及时响应、解放人力。

DB4AI: 利用数据库的能力，实现在数据最近端实现AI模型的训练、推理，做到SQL驱动数据、端到端一套技术栈、提高运算效能、发展生产力。

AI4DB(智能运维)



DB4AI (AI引擎)



THANKS !

谢谢观看

扫码关注微信公众号



添加微信小助手
回复“加群”进入交流群

