

ORACLE

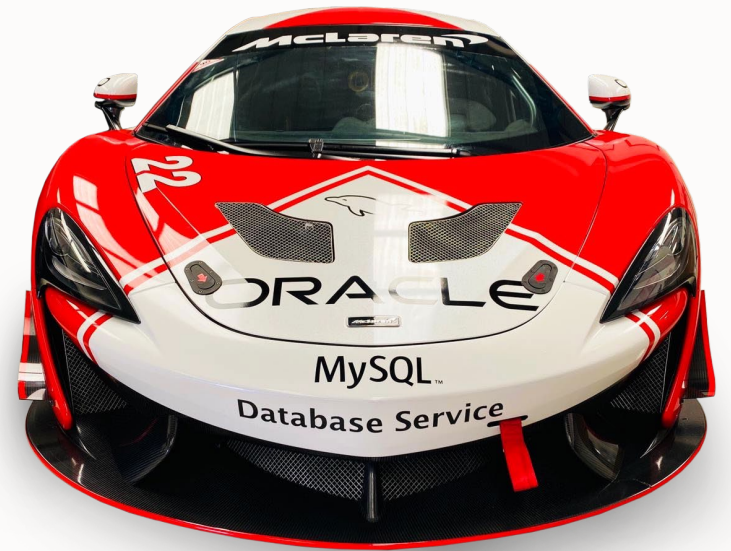
MySQL Autopilot

基于机器学习的自动化

杨鲁

yang.x.lu@oracle.com

Solution Engineer, Technology Hub, JAPAC

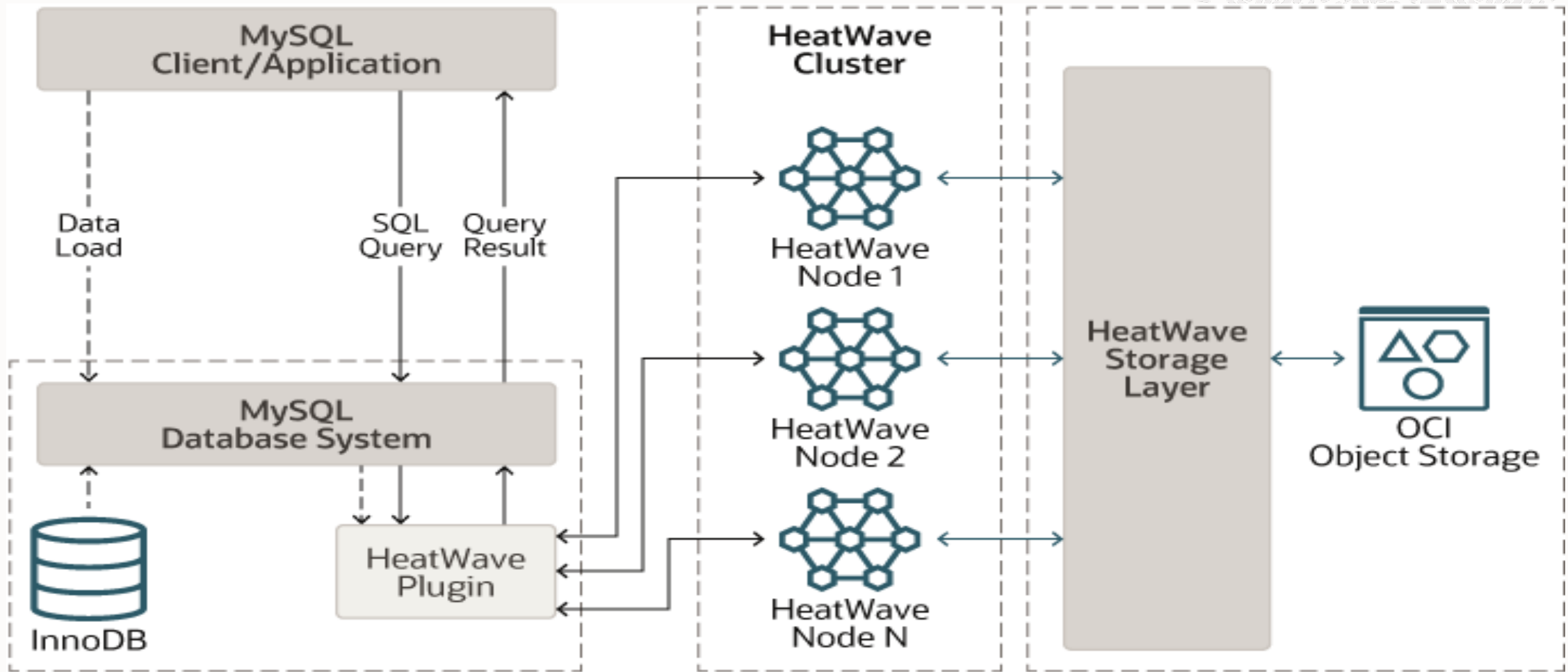


议程

- **Heatwave架构与设计**
- 自动化与机器学习
- MySQL Autopilot
- Heatwave性能与比较
- 客户评价
- 总结



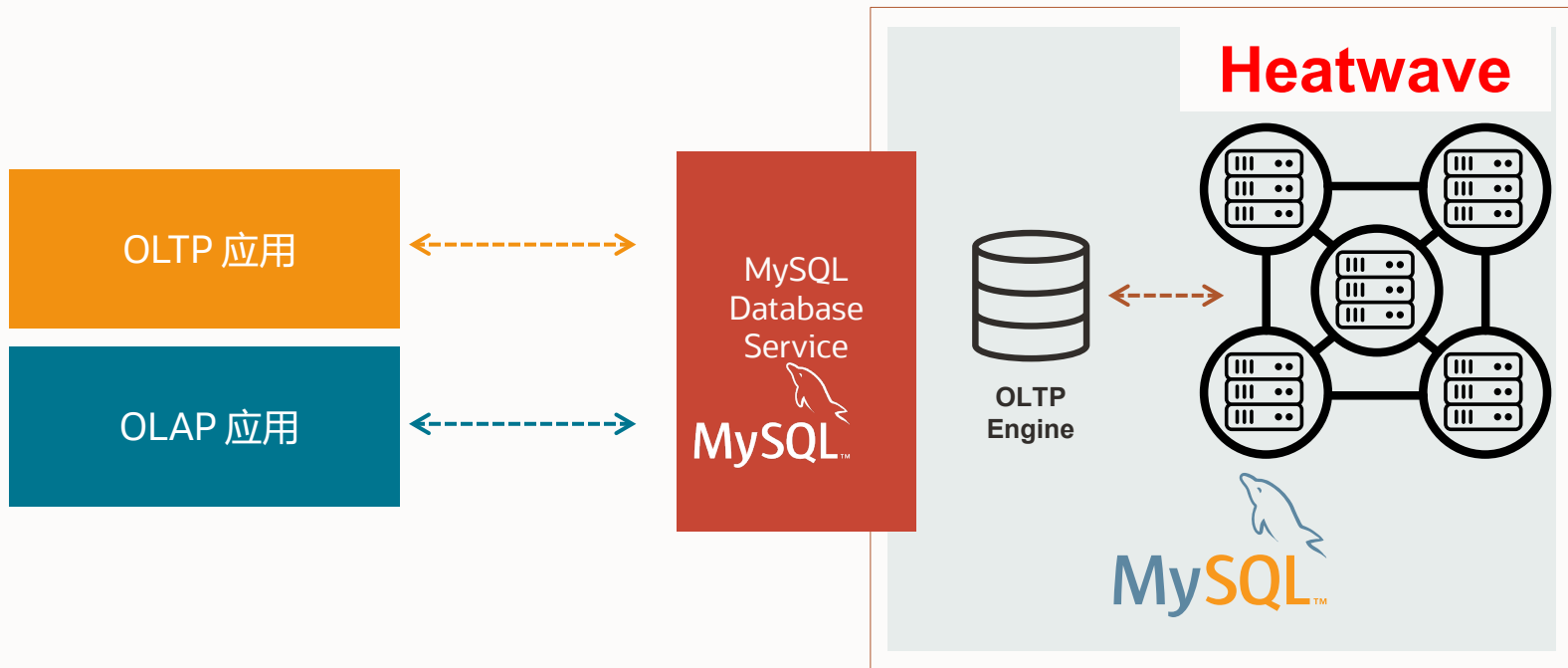
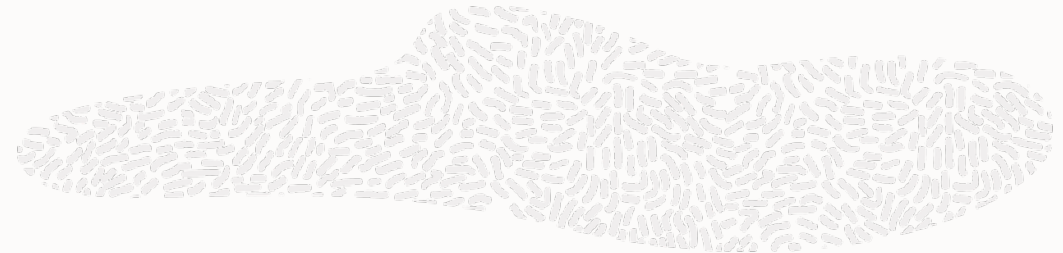
MySQL HeatWave架构图





MySQL Database Service - HeatWave

轻松对 MySQL 数据库运行高性能分析，无需 ETL



用于 OLTP 和分析应用程序的单个 MySQL 数据库

相同的 SQL——无需 ETL，以及查询加速

极致性能：比 MySQL 快 400 倍，可扩展到数千个内核



MySQL HeatWave 为云而设计



基于云从头架建:

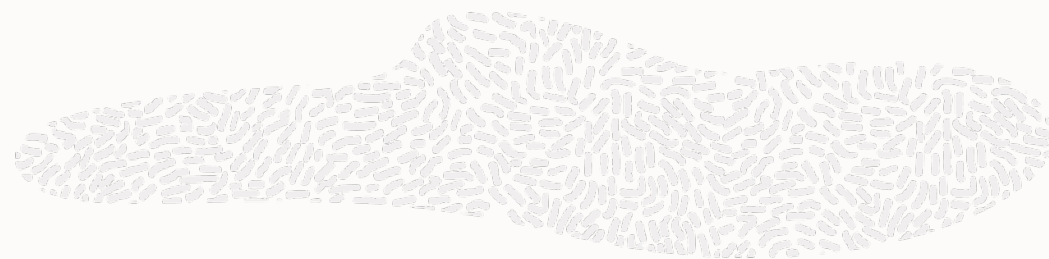
- 处理能力水平扩展
- 利用了现代云服务：低价的存储介质，网络和虚拟机
- 优化以适于OCI
- 普遍使用机器学习技术

所有这些整合到一起，使得：

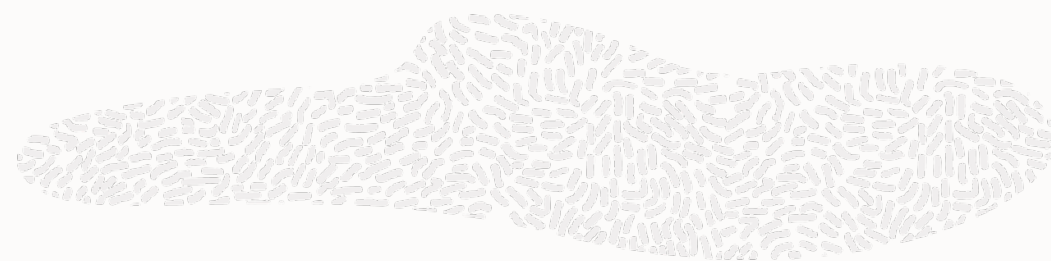
- 性能非常高
- 高扩展性
- 低成本
- 高自动化

议程

- Heatwave架构与设计
- **自动化与机器学习**
- MySQL Autopilot
- Heatwave性能与比较
- 客户评价
- 总结



云给予了自动化更多的可能



使用机器学习进行自动化有几个优点

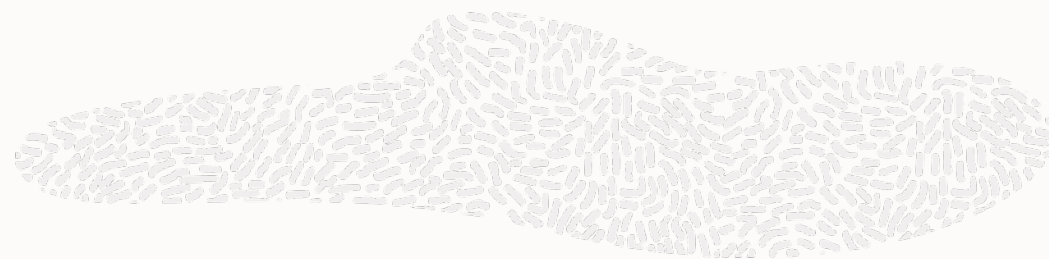
系统随着时间的推移变得更加智能



- 模型特定于实例（而不是所有用户的静态模型）
- 机器学习模型适应变化而无需重新编程规则
- 使用推荐的更改预测改进
- 能够为推荐的更改提供解释

议程

- Heatwave架构与设计
- 自动化与机器学习
- **MySQL Autopilot**
- Heatwave性能与比较
- 客户评价
- 总结



介绍

MySQL Autopilot

唯一具有基于机器学习的自动化的 MySQL 服务

无需额外付费即可使用



MySQL Autopilot 架构

新的统计信息和机器学习模型增强服务功能



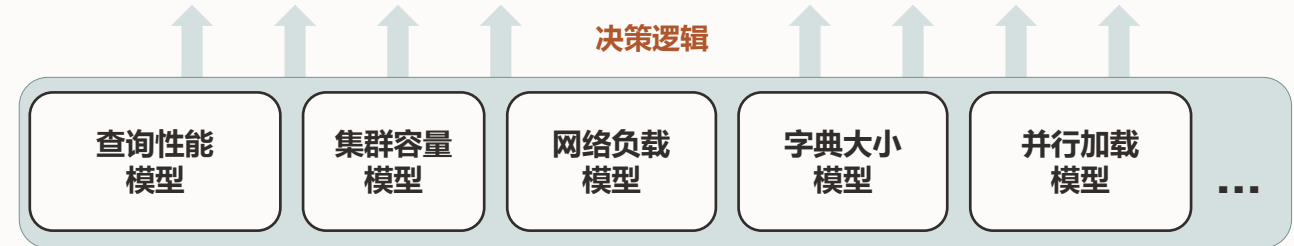
INNOVATIONS

服务生命周期



数据和工作负载驱动
机器学习决策逻辑

机器学习模型



自动机器学习
离线训练的模型

统计信息



统计信息优化
专有算法

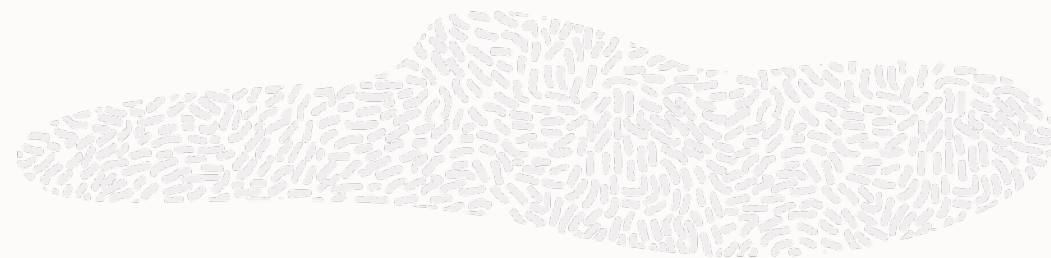
组件



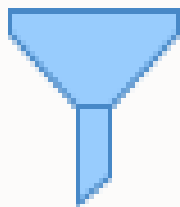
自适应采样
统计提取



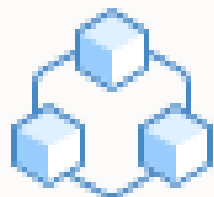
机器学习的7个步骤



Step1
数据收集



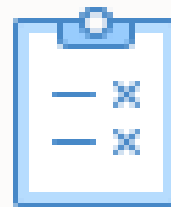
Step2
数据准备



Step3
选择一个模型



Step4
训练



Step5
评估



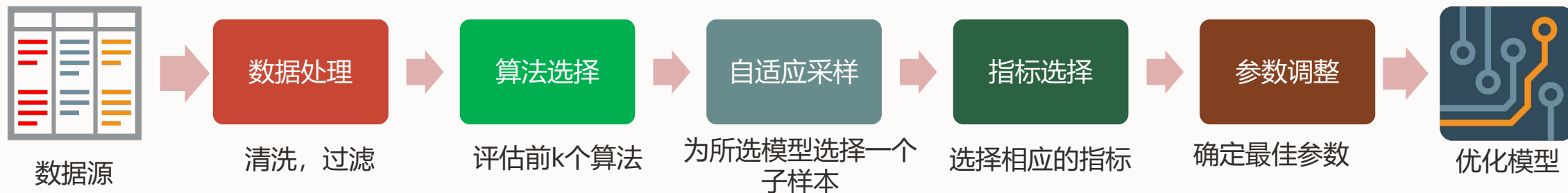
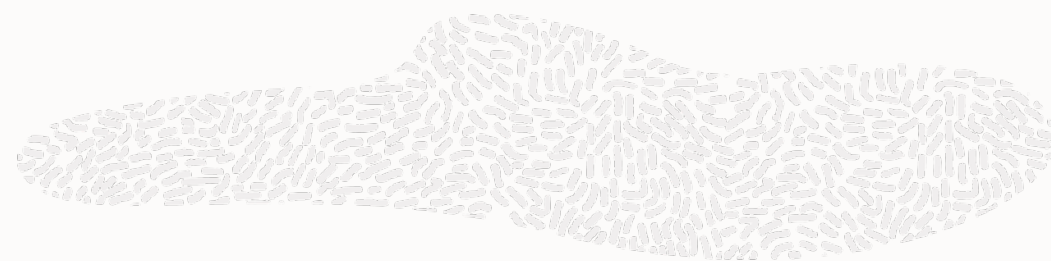
Step6
参数调整



Step7
预测



AutoML: 自动流水线以加速产生合适的模型

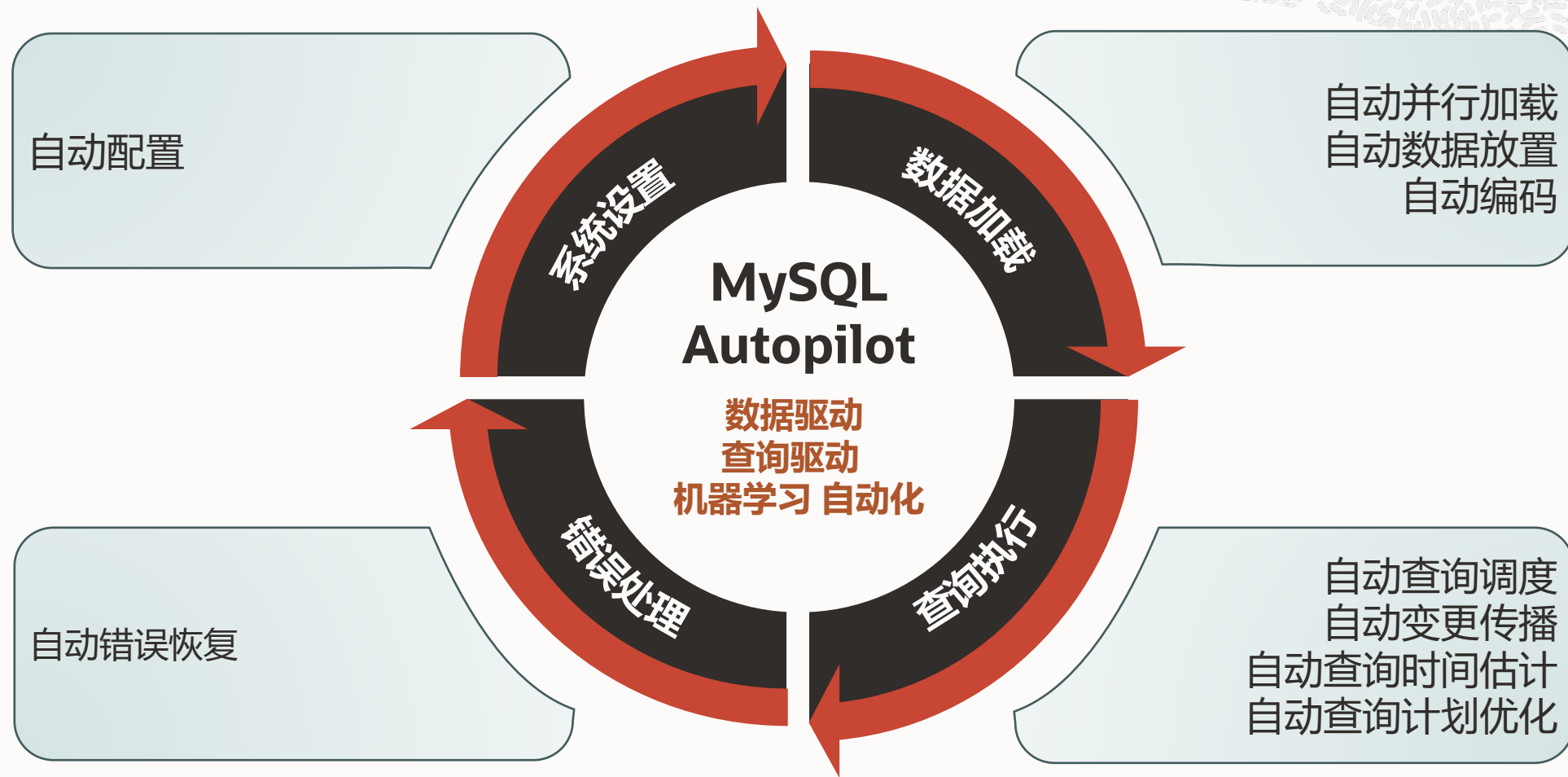


- 自动化常见的机器学习任务, 使新手最终用户和经验丰富的数据科学家都受益
- 技术集成在多个 Oracle 产品中并为每个用例提供自定义模型



MySQL Autopilot 功能

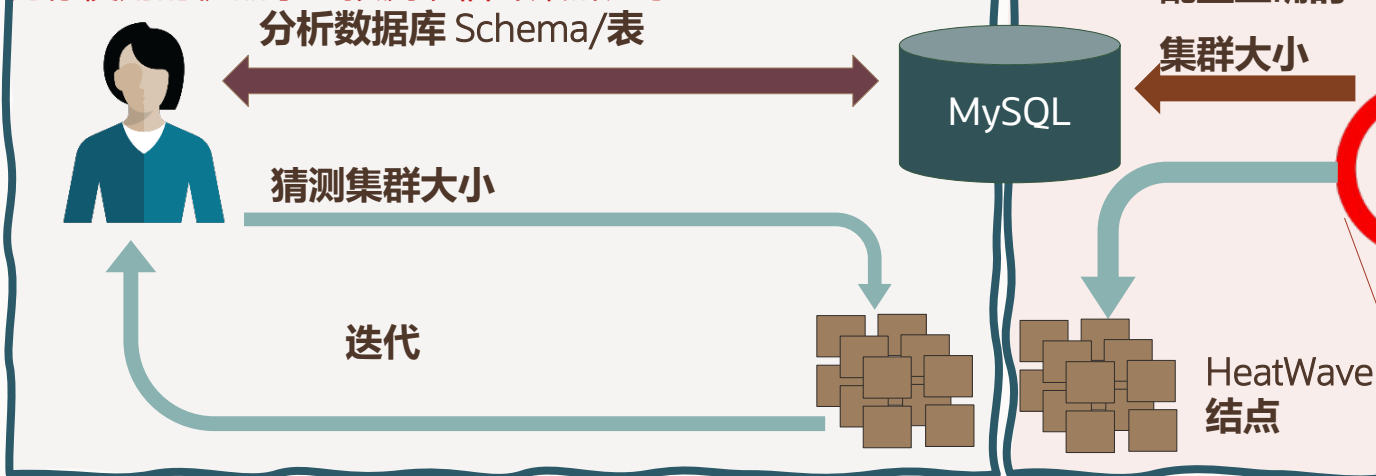
基于机器学习的自动化



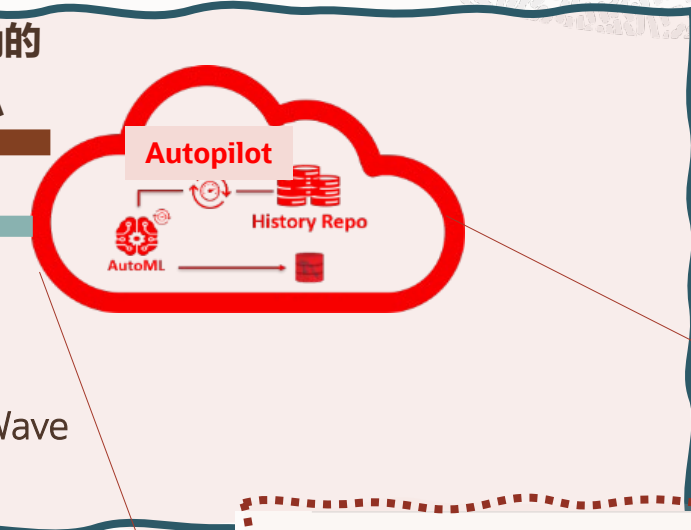
自动配置

传统的配置流程

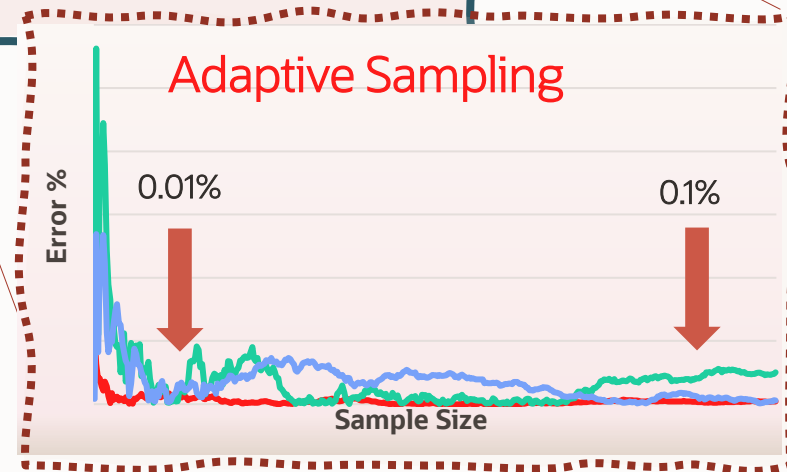
内存使用的机器学习预测以估计集群大小



自动配置



数据集	内存预测的准确性
TPCH 1024G	98.4%
TPCDS 1024G	96.9%
Cust A	98.3%
Cust B	96.9%



< 0.1% 的数据扫描用于预测





自动并行加载

减少内存，不降低加载速度



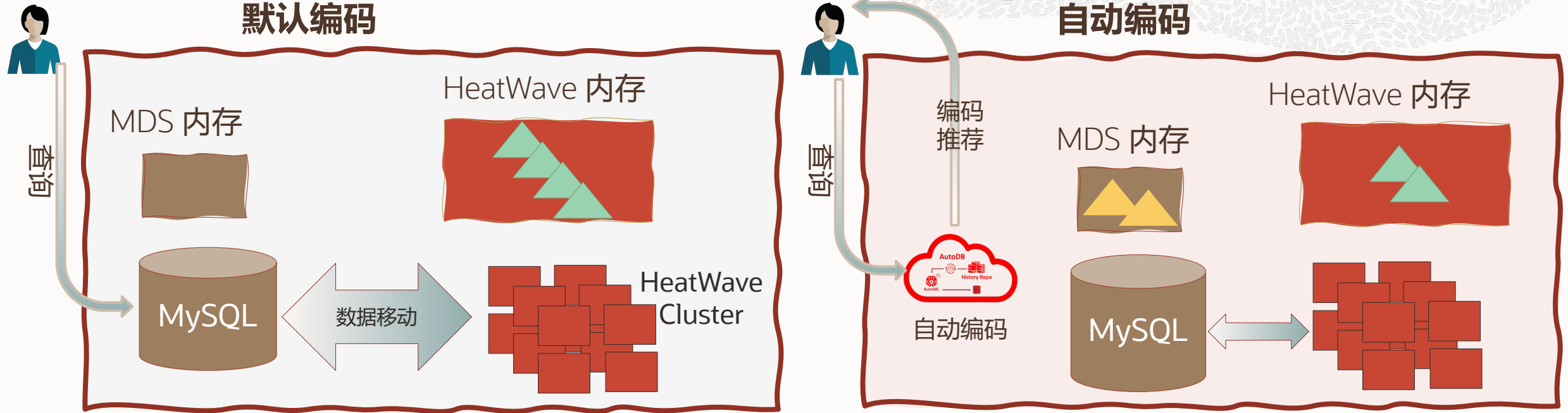
- 通过机器学习模型预测每个表的并行度
- 以最佳内存使用率实现最佳加载速度
- 在加载数据之前如加载会导致内存溢出，则提前发出警告
- 生成用于加载每个表的 SQL 语句

客户	32-线程加载时间	加载时间+Autopilot
客户 A	10.49 min	9.97 min
客户 B	8.48 min	8.38 min
客户 C	38.84 min	38.77 min



自动编码

根据工作负载优化内存使用和执行时间



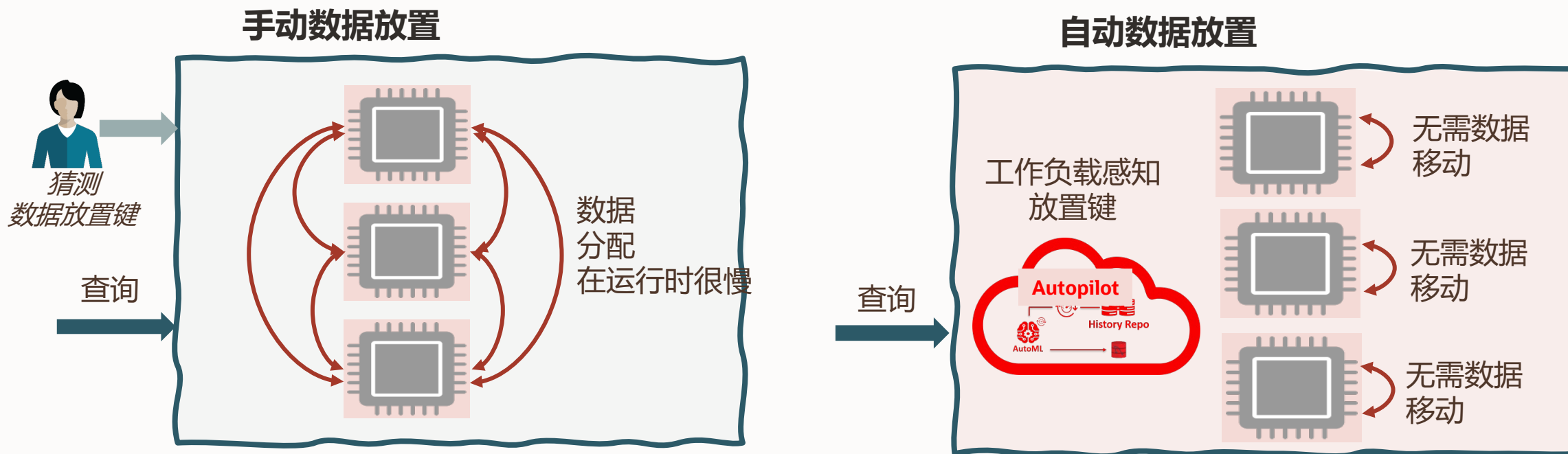
系统预测每个表中每一列的编码以优化内存使用和执行时间

测试用例	查询时间 (默认)	查询时间 (+ Autopilot)	集群大小 (默认)	集群大小 (+ Autopilot)
TPCH4096	365.3s	340.8 (-6.7%)	4287 GB	2858 GB (-33.3%)
客户 A	4.1s	2.0s (-52.1%)	19.2 GB	12.4 GB (-35.4%)
客户 B	73.4s	72.1s (-1.8%)	410.6 GB	343.0 GB (-16.5%)

自动数据放置(Placement)

最佳内存分区列的机器学习预测

- 系统根据最近的查询预测最佳列以对内存中的数据进行分区
- 预测运行时的改进

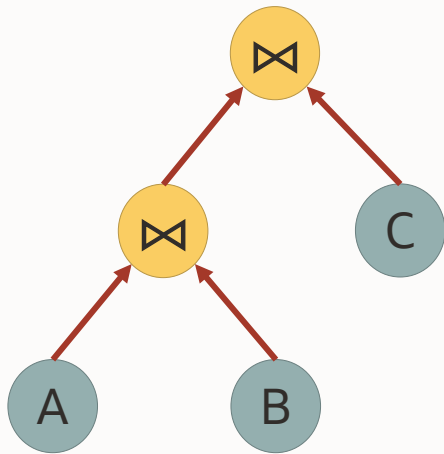


数据集	主键作为数据放置键的执行时间	预测提升 +Autopilot	实际获得提升 +Autopilot
TPCH 1024	332 sec	26%	37%
TPCH 4096	373 sec	20%	25%



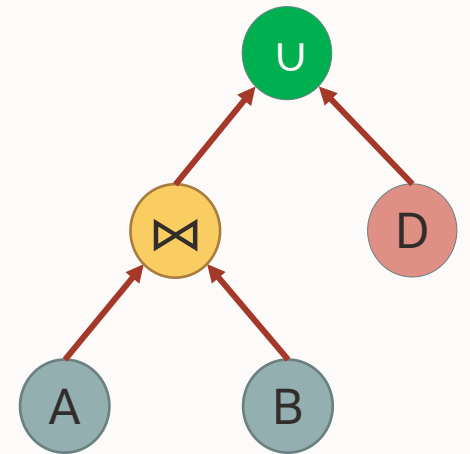
自动查询计划改进

优化器根据之前执行的查询学习和改进查询计划



结点	统计信息
A	70
B	150
A ⋈ B	1000
C	...
A ⋈ B ⋈ C	...

运行时统计信息



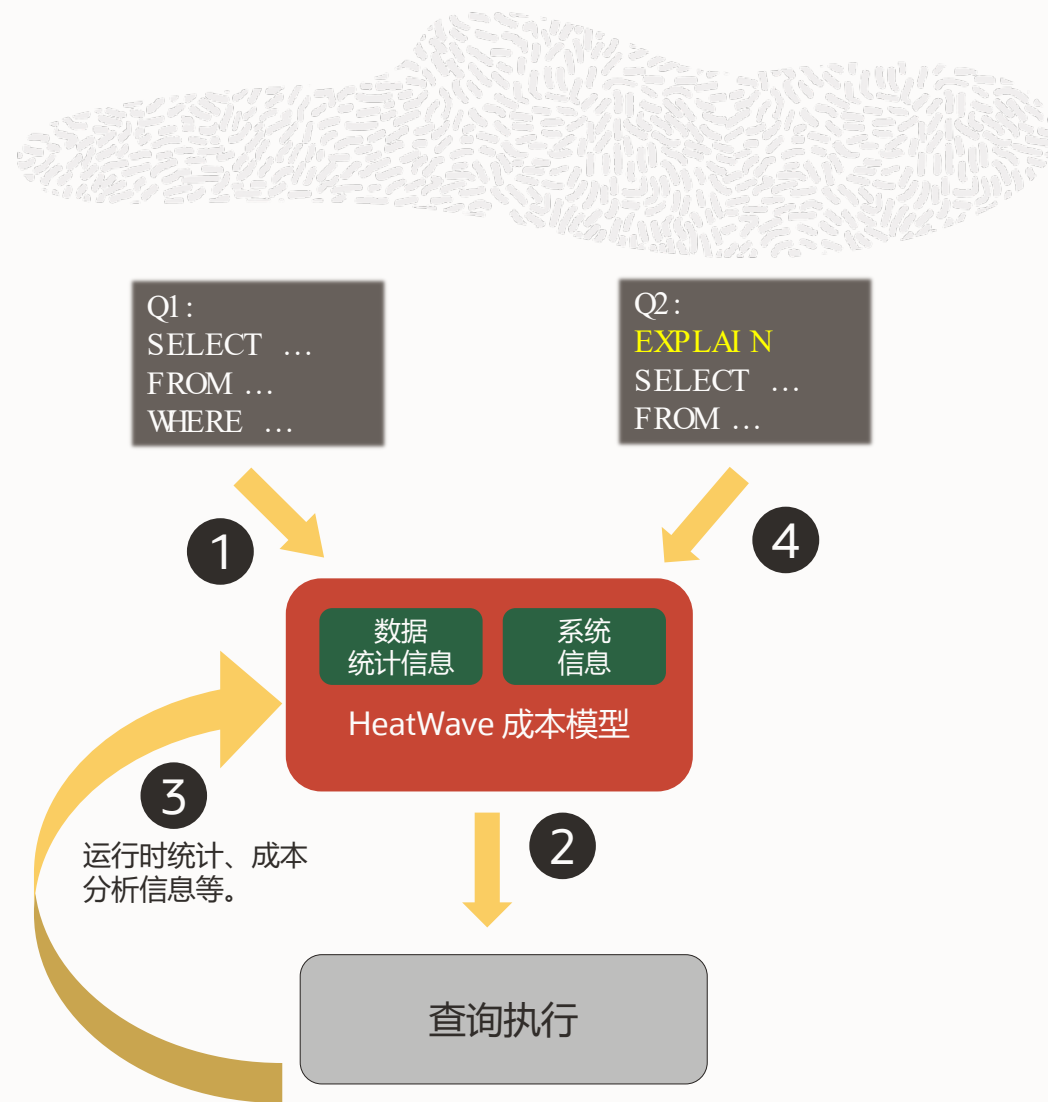
- 传统缓存技术并不智能
- 使用 Autopilot, 系统会随着运行更多查询而变得更好
- 比如Autopilot将TPCH、TPCDS 24TB性能提升40%



自动查询时间估计

数据驱动查询时间估计

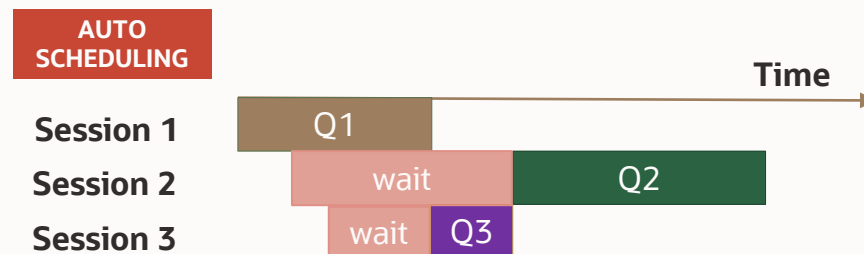
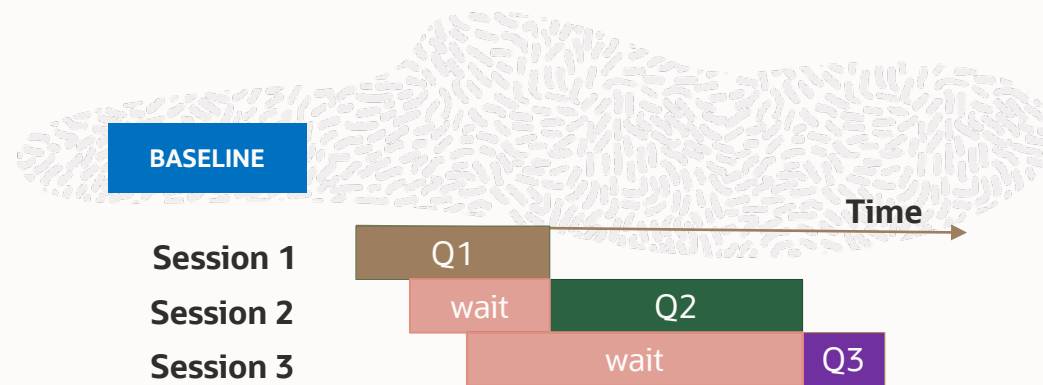
- HeatWave 查询时间估计基于其查询处理算法的分析模型
- 使用数据驱动的方法提高了查询时间估计的准确性，随着更多查询的运行，该准确性会随着时间的推移而提高
- 运行时反馈改进基数估计以帮助调整框架开销常量



自动调度

减少混合 (OLTP + OLAP) 工作负载的等待时间

- 分析查询通常比 OLTP 查询花费的时间更长
- HeatWave 预测每个查询的执行时间
- 短查询优先于长时间运行的查询
- 系统在不改变总执行时间的情况下减少较短查询的等待时间



工作负载中的短查询			
查询	基线 (sec)	自动调度 (sec)	耗时间 提升
12	26.03	7.43	
16	27.18	12.45	
10	18.93	12.62	
7	16.27	12.08	
19	18.24	13.81	
几何平均	20.89	11.43	

工作负载的长查询			
查询	基线 (sec)	自动调度 (sec)	耗时间 增加
21	62.17	65.59	
18	44.37	59.41	
9	31.42	25.59	
17	38.06	42.07	
1	22.02	23.56	
几何平均	37.34	39.72	



自动变更传播

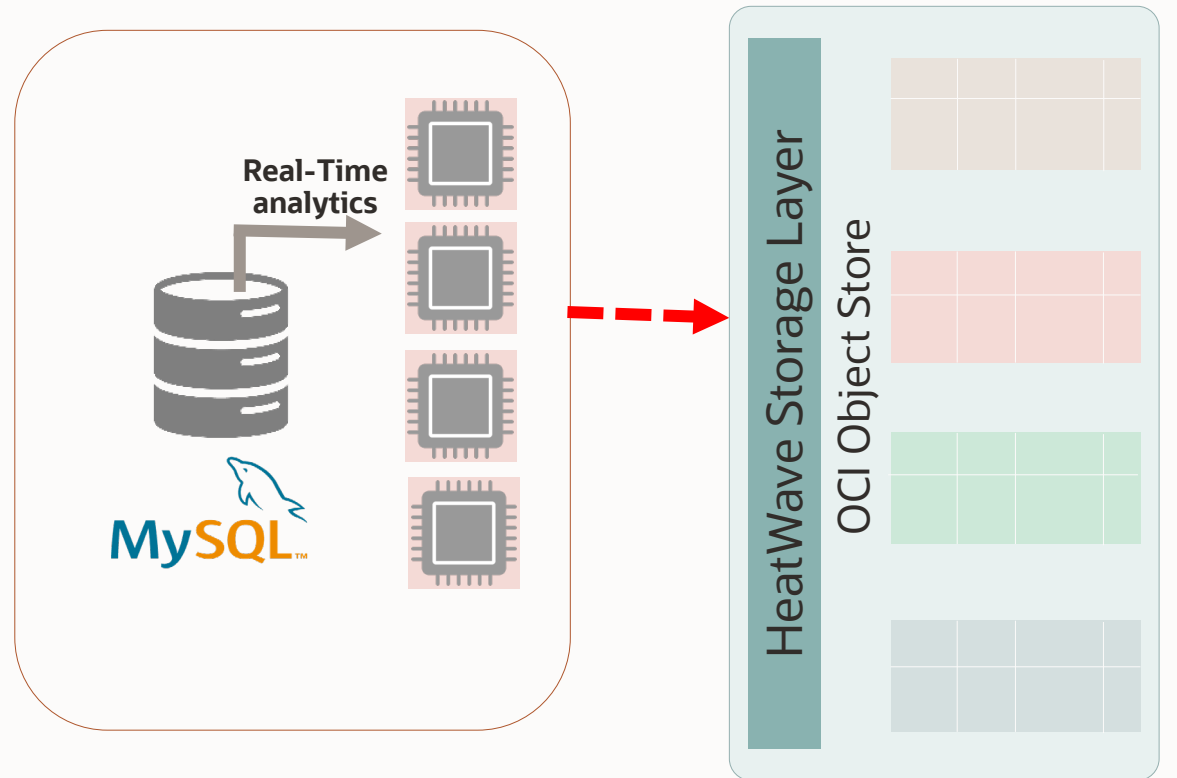
机器学习预测变更传播的最佳时间

传统做法，数据库使用固定持续时间来传播变更

对 HeatWave 集群的数据变更是实时传播的

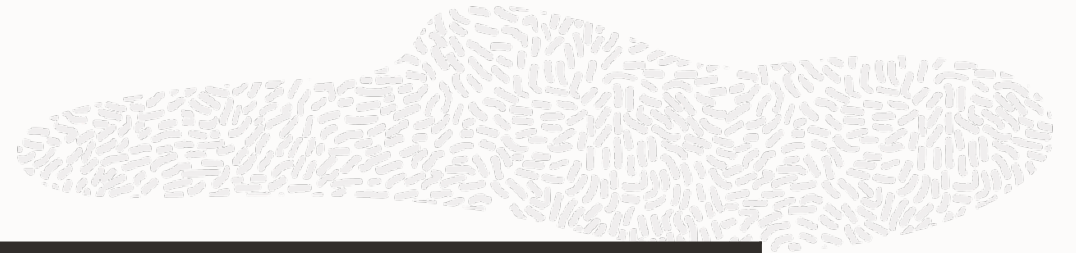
基于机器模型推送对对象存储的更改，该机器模型考虑：

- 变化率
- 变化的大小
- DML 的类型
- 对象存储带宽



自动错误恢复

系统自动从任何故障中恢复



Event	Actions
HeatWave 软件错误	<ul style="list-style-type: none">• 重启失败的 HeatWave 进程
OS/系统 错误	<ul style="list-style-type: none">• 重启失败的结点
硬件错误	<ul style="list-style-type: none">• 重新配置失败的结点
通常	<ul style="list-style-type: none">• 数据从外部存储进行重新加载• HeatWave 在线

在 HeatWave 集群恢复期间，MySQL 数据库服务是可用的



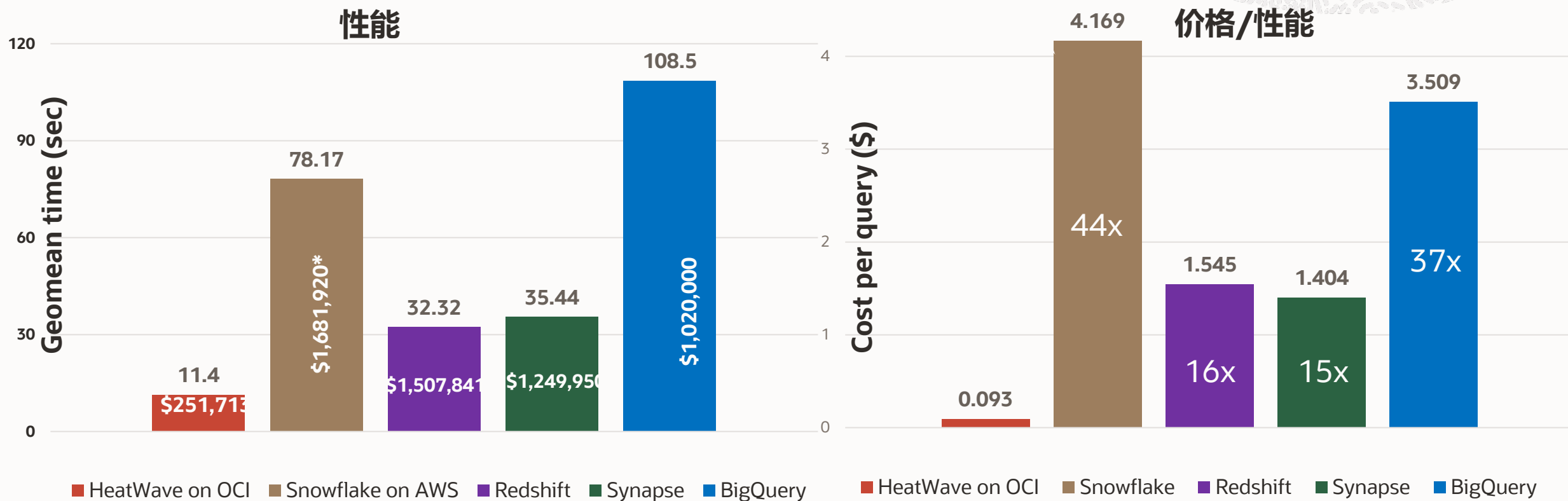
议程

- Heatwave架构与设计
- 自动化与机器学习
- MySQL Autopilot
- **Heatwave性能与比较**
- 客户评价
- 总结



性能 & 价格比较 vs Snowflake, Redshift, Synapse, Big Query

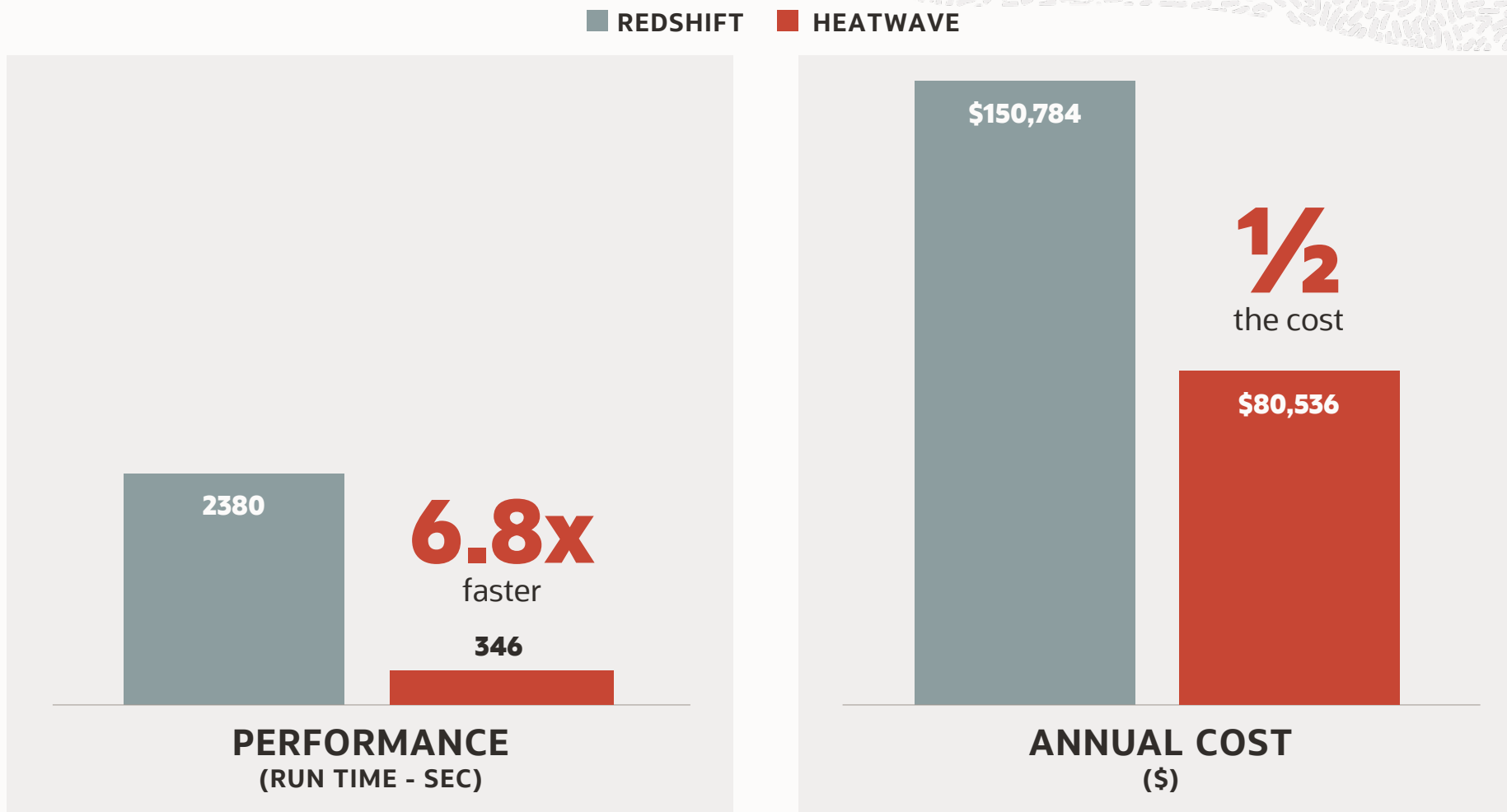
30TB TPCCH, HeatWave 比所有有竞争力的数据库服务更快、更便宜



- 第3方数字来源于 Gigaom 2020年10月的报告
- *Snowflake是使用的PAYG定价模式。其他价格基于 1 年定价



与 Redshift AQUA 相比，价格/性能提高 13 倍——高于 7 倍 10TB TPCH



25 nodes of HeatWave, 8 Nodes of RA3.4xLarge, 1 year, paid upfront, price used for AWS

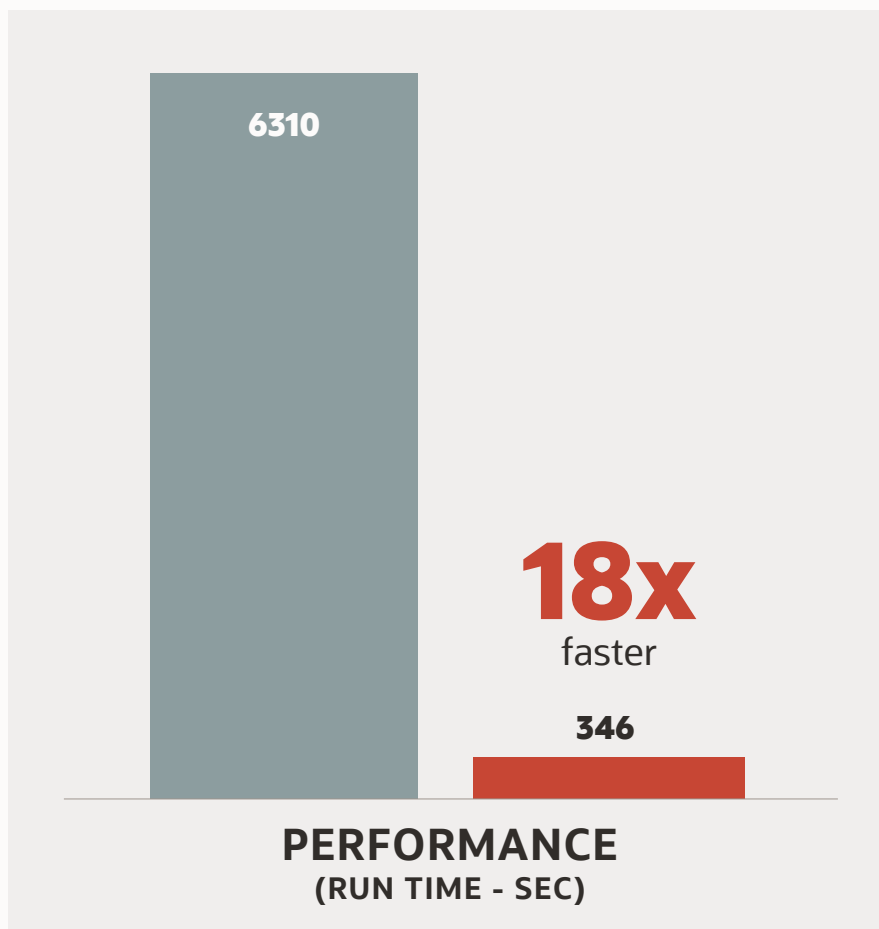
Benchmark queries are derived from TPC-H benchmark, but results are not comparable to published TPC-H benchmark results since they do not comply with TPC-H specification.



价格/性能比 Redshift AQUA 高 17 倍，成本相似

10TB TPCH

■ REDSHIFT ■ HEATWAVE



25 nodes of HeatWave, 4 Nodes of RA3.4xLarge, 1 year, paid upfront, price used for AWS

Benchmark queries are derived from TPC-H benchmark, but results are not comparable to published TPC-H benchmark results since they do not comply with TPC-H specification.



议程

- Heatwave架构与设计
- 自动化与机器学习
- MySQL Autopilot
- Heatwave性能与比较
- **客户评价**
- 总结





应用程序无变化

某些查询性能提升 1000 倍

整体工作负载性能提升 85%

成本降至 AWS Aurora 的 40%

Amit Palshikar

CTO, Red3i





OLTP 和分析查询的混合

查询时间从几分钟缩短到几毫秒

无缝迁移

HeatWave 相对 AWS Aurora 的成本降低了一半

Pablo Lemos

Co-founder

Tetris.co





MySQL HeatWave 是 Aurora 的绝佳替代品。

无需任何修改即可迁移

我们的工作负载性能比 Aurora 提高了 10 倍

迁移后成本显著下降

Kenji Suzuki
Manager, Product Development
Fan Communications

FANCOM



迁移到 MySQL HeatWave 将我们的成本降低了 3 倍

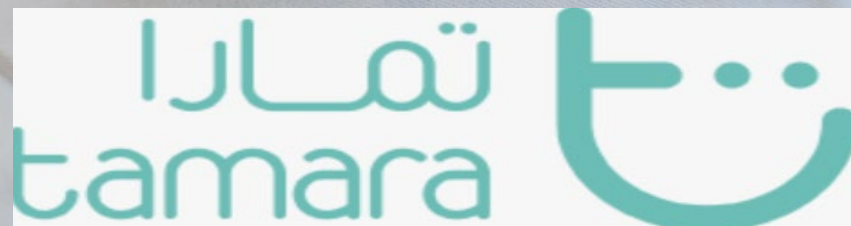
显着加速了我们之前需要很长时间的许多查询。

我们希望我们能够通过编写更复杂的查询来增强我们的应用程序，这些查询在其他云解决方案无法实现合理的执行时间。

Chien Hoang

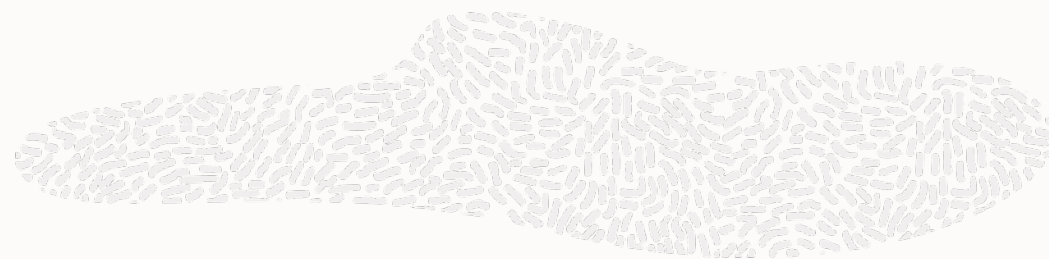
Director of Engineering

Tamara



议程

- Heatwave架构与设计
- 自动化与机器学习
- MySQL Autopilot
- Heatwave性能与比较
- 客户评价
- **总结**



MySQL Autopilot 总结

功能提供而无需额外的付费



- MySQL Autopilot 支持基于机器学习的自动化
- 自动化配置、数据加载、查询执行和故障处理
- 进一步提高 HeatWave 的性能
- 比所有其他有竞争力的数据库分析服务更快、更便宜





oracle.com/mysql/heatwave/





Test Drive MySQL Database Service For Free Today

Get **US \$300** in credits

Try **MySQL Database Service / Heatwave**
Free for 30 days.



Register Now



bit.ly/3BQt5VS



ORACLE