



MySQL HeatWave最新情報

MySQLで高速な分析処理と機械学習による予測

梶山 隆輔 / KAJIYAMA, Ryusuke

MySQL Solution Engineering Director, Asia Pacific & Japan

MySQL Global Business Unit

May 20, 2022



#oradev22



MySQL HeatWave Database Service

MySQL開発元が提供するフルマネージド・データベース・サービス

サービス概要

- Oracle Cloud Infrastructureに**最適化**されたMySQL
- 既存MySQLとの併用も可能
- 常に**セキュア**な最新MySQLを適用
- MySQL開発チームと連携した**高品質なサポート**
- **運用管理・コスト効率化**を可能にするインフラストラクチャ

適用事例

- 株式会社パソナテック
- 株式会社ファンコミュニケーションズ
- 株式会社りらく
- 中空知衛生施設組合(北海道滝川市)
- ジニアス・ソリティ株式会社 など



スタンドアロン

MySQL利用システムのクラウド化に

- ベーシックなMySQLのデータベース・サービス
- エンタープライズ向け機能をサポート

高可用性

継続運用を重視するシステムに

- 自動フェイル・オーバー、データ損失ゼロをサポート
- 3インスタンスで実現する高可用性構成 (InnoDB Cluster)
- ワンクリックで高可用性構成を構築可能

HeatWave

データ分析システムに

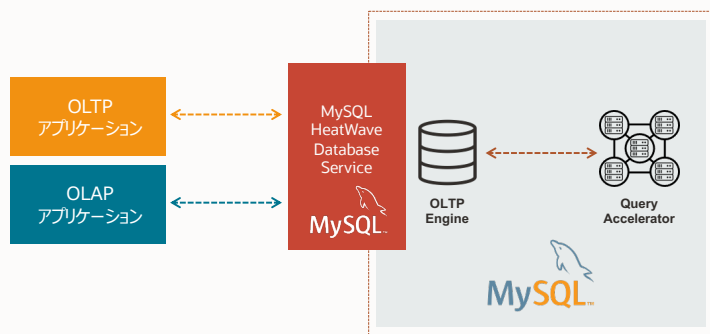
- インメモリ & カラムナ型エンジン
- 機械学習ベースの運用自動化機能をサポート (Autopilot)
- 簡単に機械学習が行えるプラットフォーム (HeatWave ML)

MySQLの高速クエリ処理アクセラレーター

HeatWaveオプション

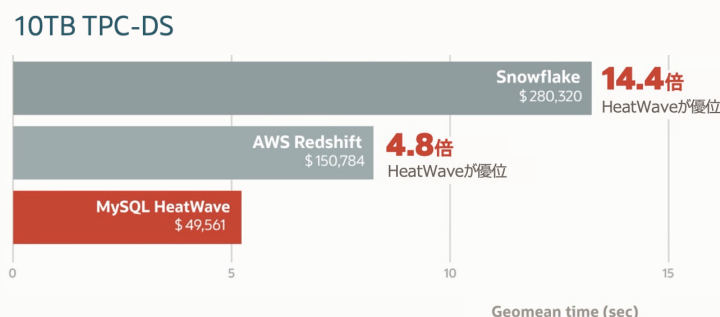
従来のトランザクション処理とETL不要な分析処理を実現

- ✓ OLTP、OLAPワークロードをMySQLインタフェースでサポート
- ✓ 従来と比較して400倍以上のクエリ処理高速化を実現
- ✓ 既存アプリケーションの変更不要



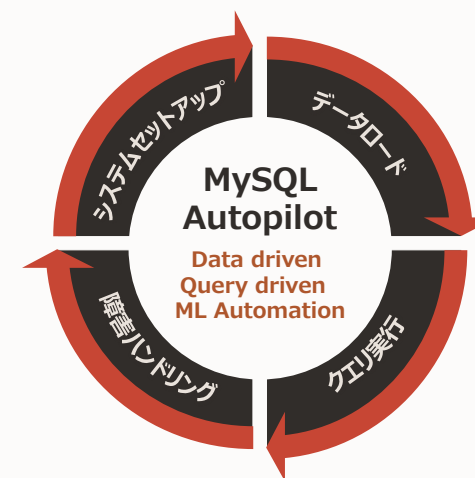
圧倒的なパフォーマンスとコスト効率化

- ✓ インメモリ・カラムナ型アーキテクチャ
- ✓ 並列・分散アルゴリズムにより、オーバーヘッドを最小化
- ✓ ノード数に比例してパフォーマンス向上するスケール・アウト機能



機械学習技術を利用した運用管理業務を効率化 **AutoPilot**

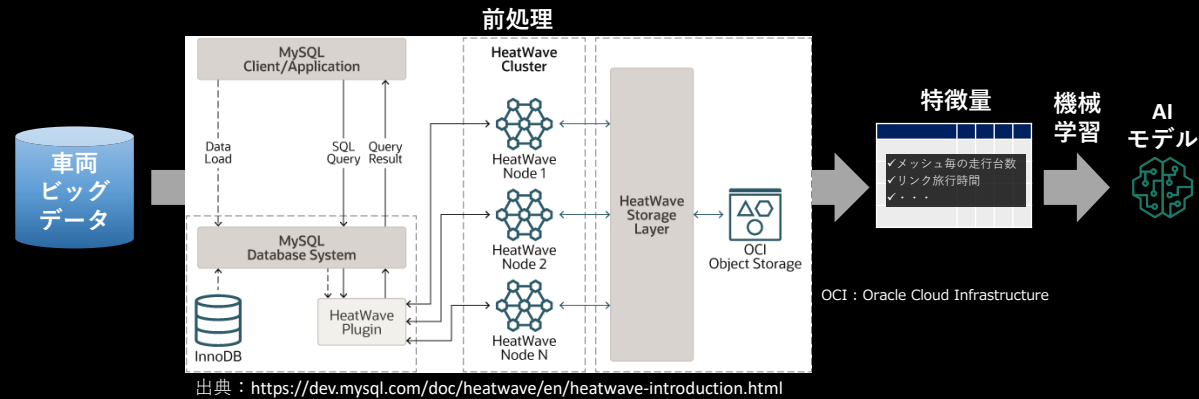
- ✓ 最適なリソース量やデータ配置を予測し実行コマンドを提案
- ✓ 煩雑なチューニング作業を自動化、ユーザが意識することなく性能向上



<https://www.oracle.com/mysql/heatwave/performance/>

- ☺ HTAPに対応したインメモリ・カラムナ型データベース
- ☺ ノード/コア/命令レベルの並列化による高速分析が可能

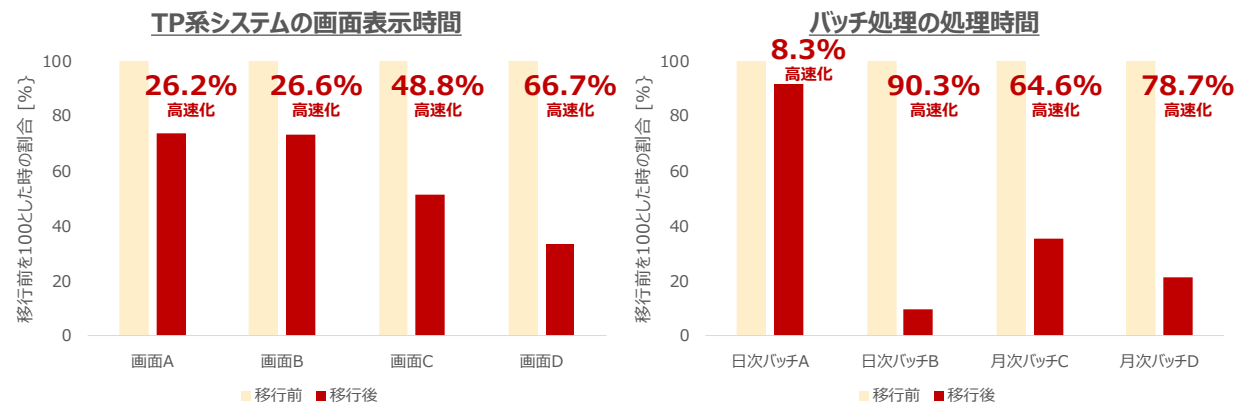
⇒AIアルゴリズム開発におけるアドホックな分析ニーズに対応できると期待



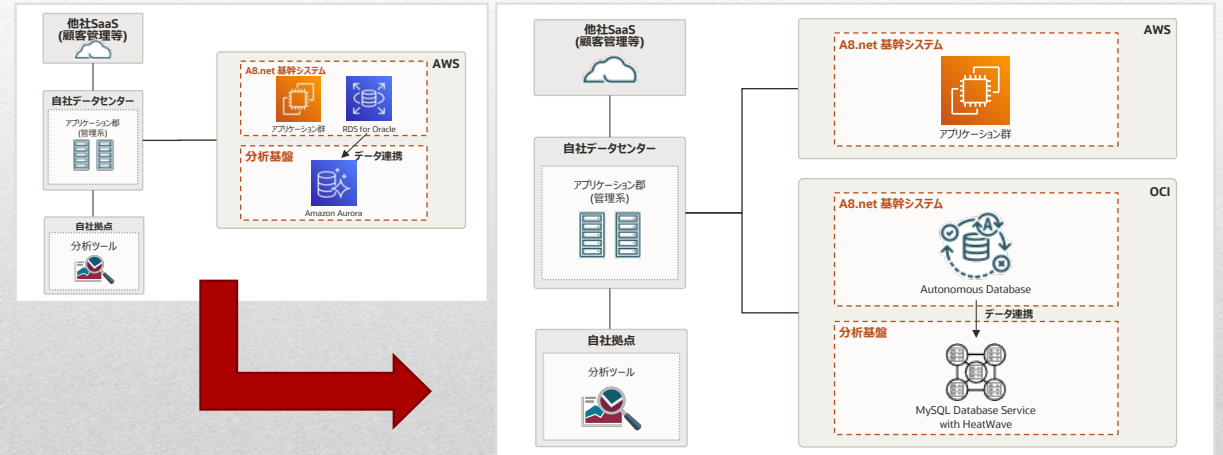
TOYOTA

3. OCI採用の理由

OCI、MDSへの移行により、日中の処理もバッチ処理もパフォーマンス改善！！



新しいシステム構成



A8.net

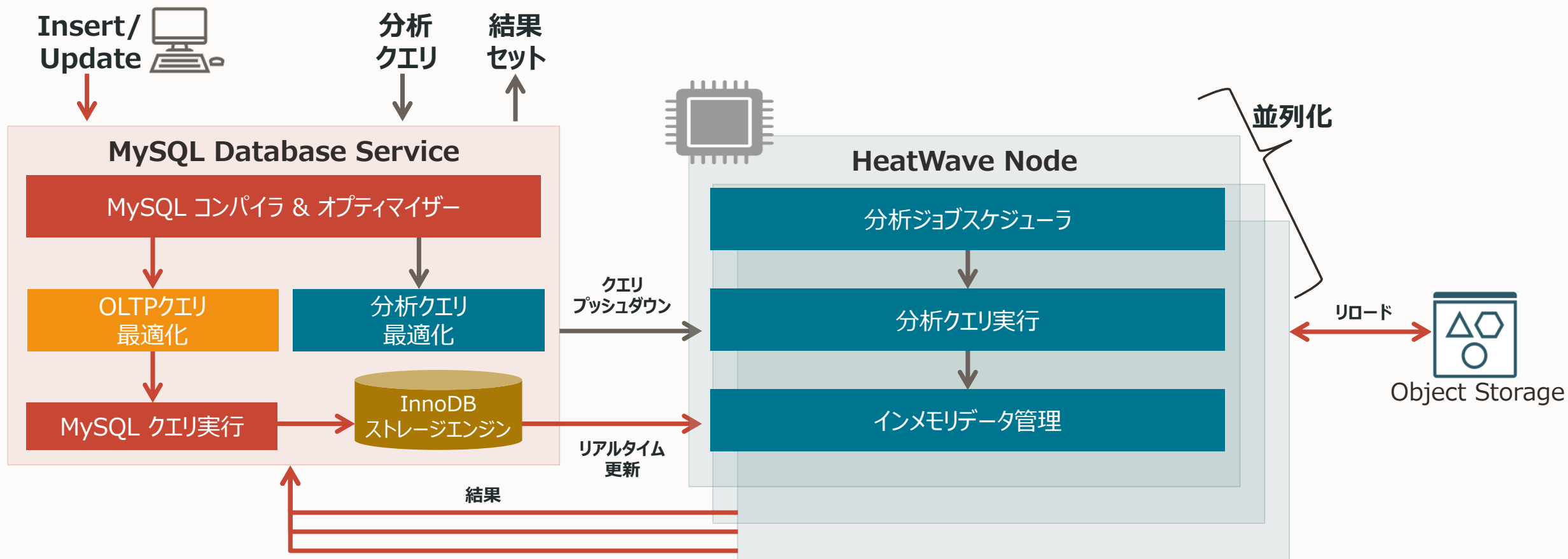
Oracle Cloud DaysでのMySQL事例講演
(2021年11月開催)

- トヨタ自動車様
機械学習の前処理のデータベースとしてHeatWaveを検証
- ファンコミュニケーションズ様
広告プラットフォーム A8.netの分析基盤をAWS AuroraからHeatWaveに移行し大幅な性能向上と費用抑制を実現
- りらく様
AWS RDSからOCI上のMySQL Database Serviceの移行し50%以上のコスト削減と処理性能の向上を実現

MySQL HeatWaveのアーキテクチャ

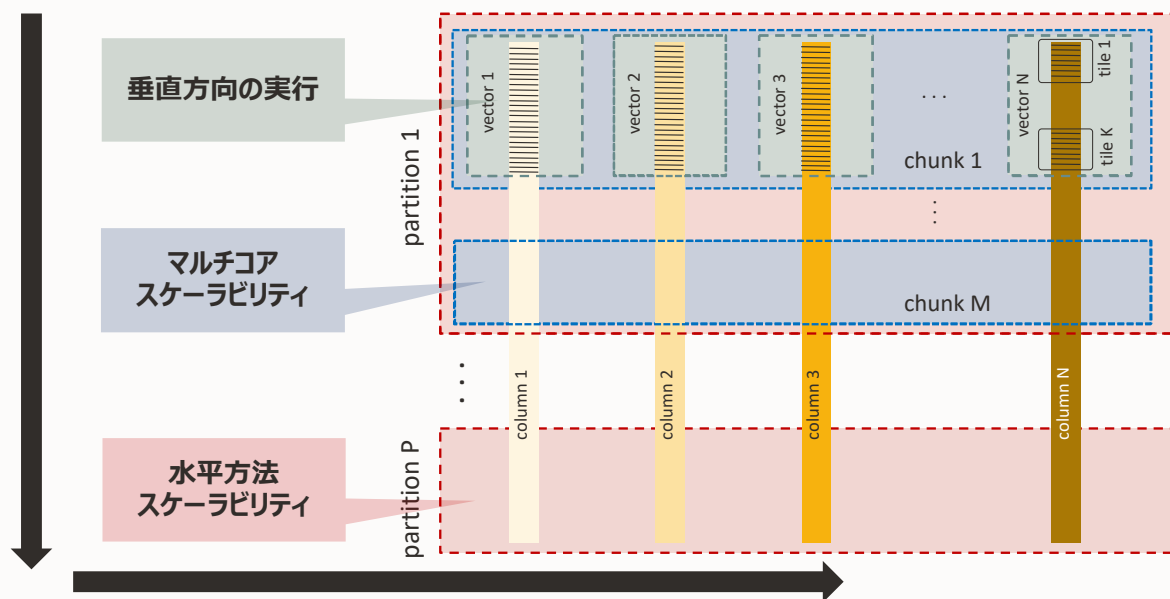
MDSに追加されたデータは分析クエリで即座に利用可能

- アプリケーション側からはMySQLに対してSQLを実行するのと同じ

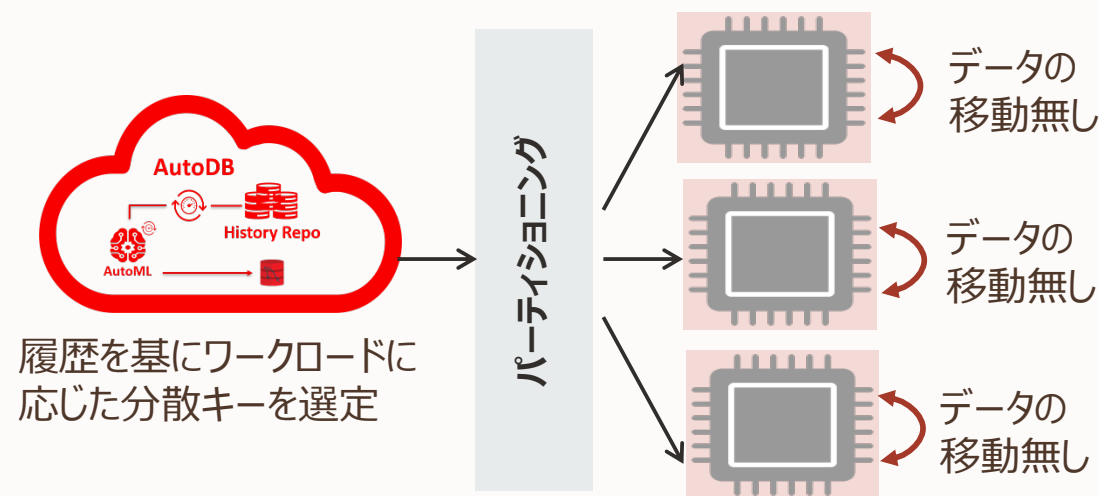


インメモリのハイブリッドカラム型データ処理

- パーティションに分割してHeatWaveノードに格納
- パーティショニングによってノード/CPUコアでの並列処理を実現



- パーティショニングのデフォルトは主キー
 - ✓ 機械学習により、パーティショニングに最適な列予測に従って変更することが可能
 - ✓ クエリ実行履歴に応じたパーティショニングが可能



HeatWaveのイノベーションは続く

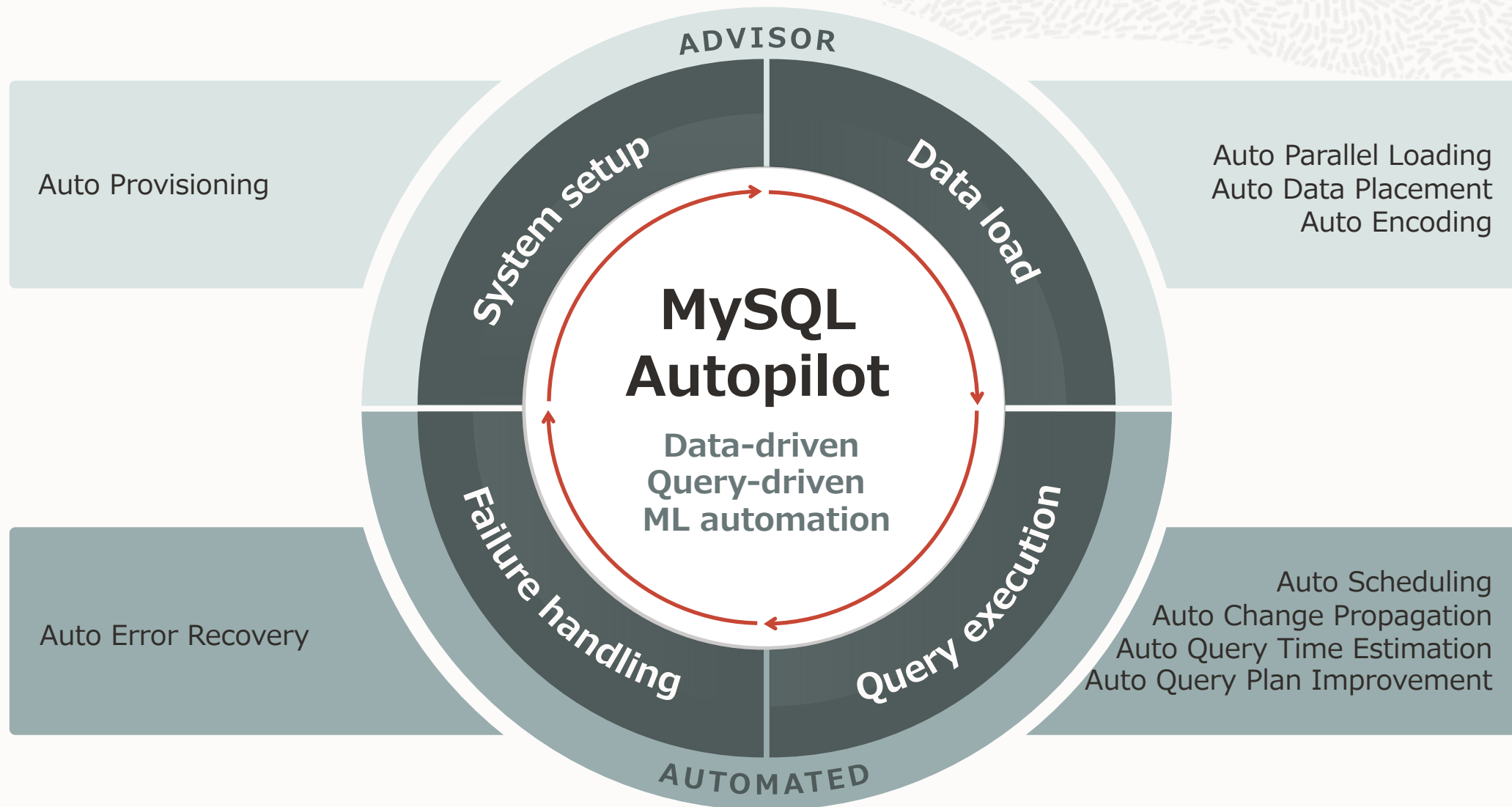
2021年8月の発表

- MySQL Autopilot
- データ管理のスケールアウトのためのストレージのサポート
- より大きなデータベースのために64ノードまで拡張可能
- TPC-Hや混合型のワークロードのベンチマークで
明らかな性能とコストの優位性

Since then we've been busy

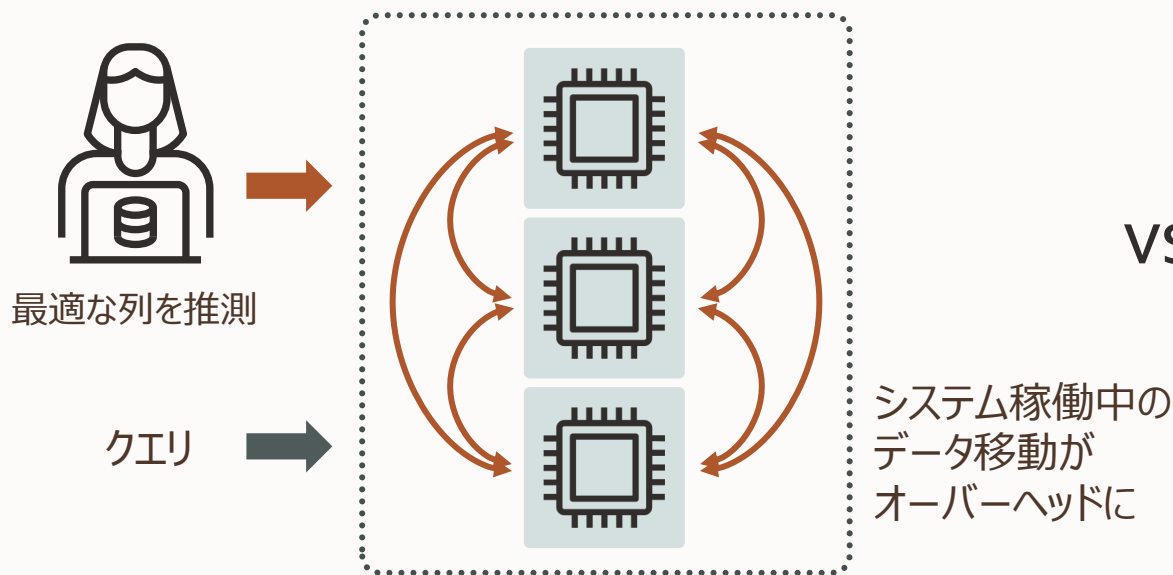
- ビューのサポート
- 日付、時間、型のキャスト
- IN句のサポート
- HeatWaveの稼働統計をパフォーマンス・スキーマに格納
- エンジン内で処理されているクエリのリアルタイムな統計
- InnoDBとHeatWaveで実行されているクエリのリアルタイムな統計
- サポートするカラム数を450から1,000に拡大
- カラムサイズの最大を8KBから64KBに拡大
- 混合型のワークロードの処理性能を2倍に向上
- MySQLサーバーでのデータマスキングをサポート

機械学習を活用したMySQL HeatWaveの自動化



最適なパーティションキーの予測 - Auto Data Placement

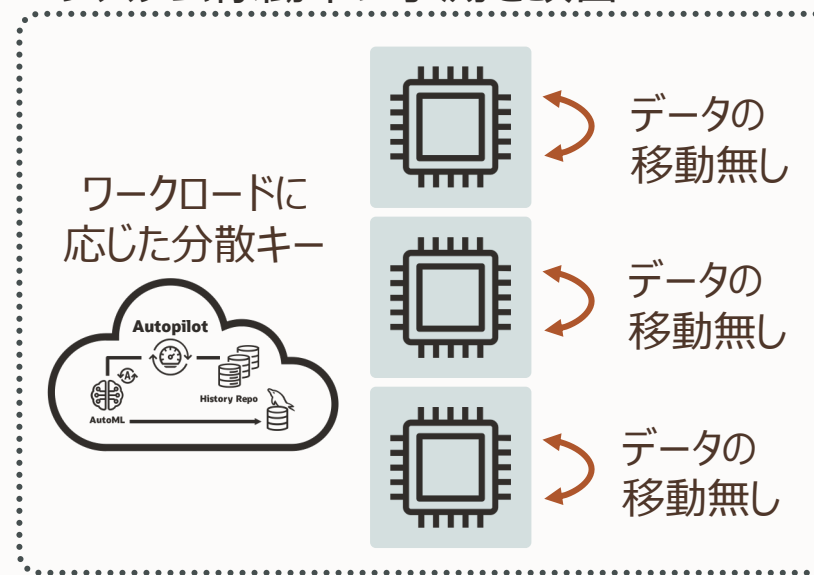
パーティションキーを推測する従来の手法では
最適なパフォーマンスとならないケースも



VS

クエリ

- 最近実行されたクエリに応じて最適なパーティションキーを予測
- システム稼働中に予測を改善

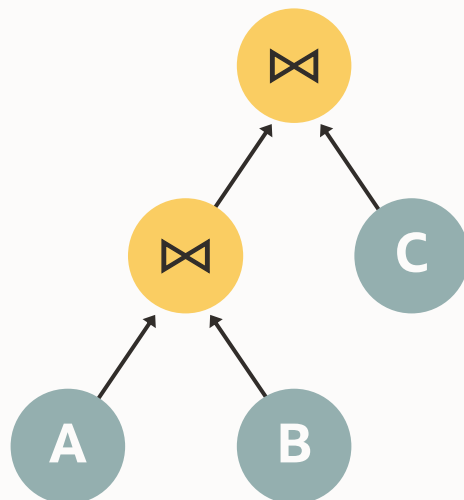


Dataset	主キーでの分割	Autopilotによる 性能向上の予測	Autopilotによる 実際の性能向上
TPCH 1024	332 sec	26%	37%
TPCH 4096	373 sec	20%	25%

SQL実行計画の自動チューニング - Auto Query Plan improvement

実行されたクエリを学習してHeatWaveオプティマイザーが実行計画を改良

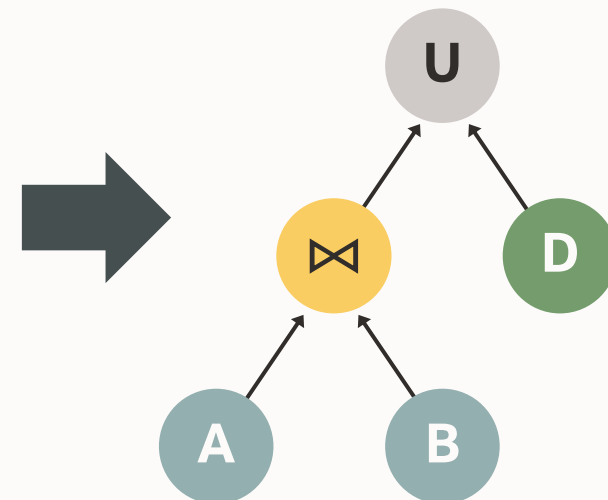
Query #1
A ⋈ B ⋈ C



Node	Statistics
A	70
B	150
A ⋈ B	1000
C	...
A ⋈ B ⋈ C	...

Runtime stats

Query #2
A ⋈ B U D

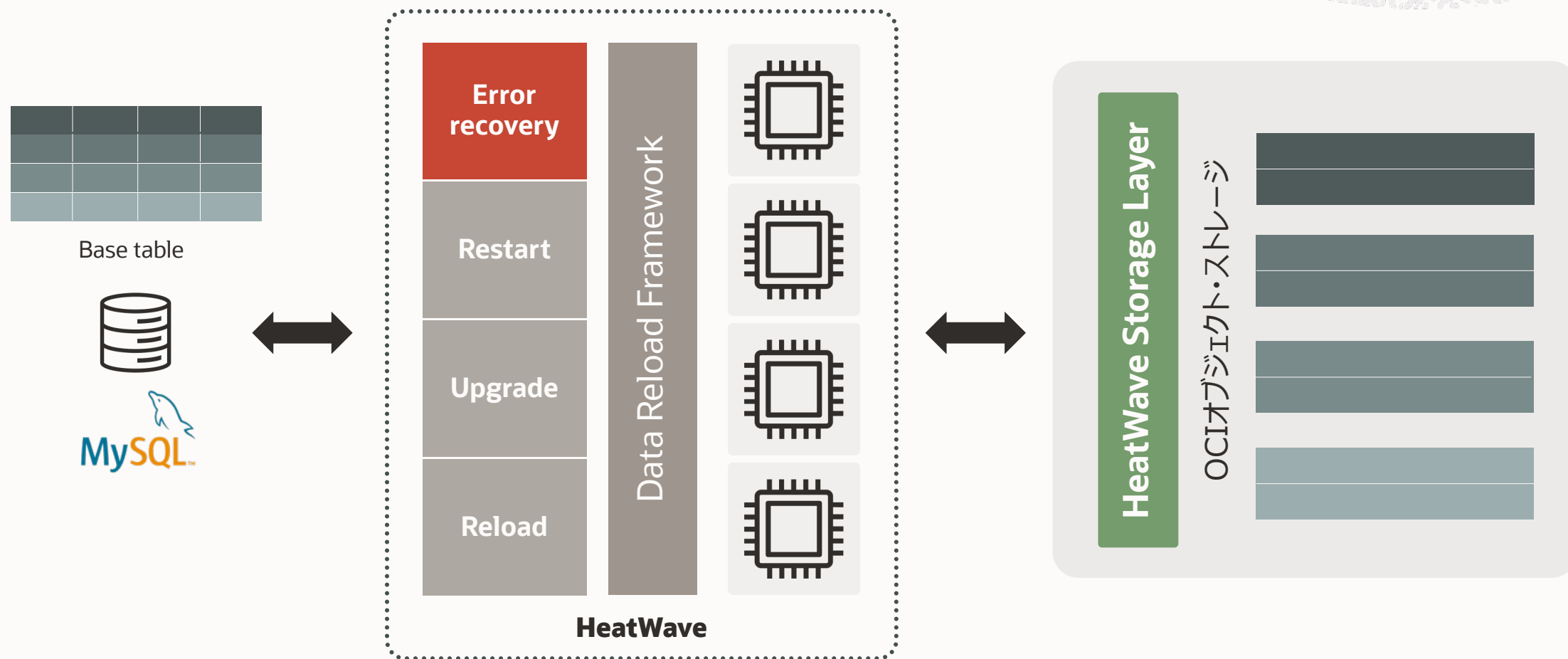


MySQL AutopilotによるTPC-H, TPC-DS 24TBの性能向上

40%

オブジェクトストレージの活用 - Scale-Out Data Management

データサイズにかかわらず一定のデータロード時間





GENIUS (SONORITY)

Copyright © 2022 Oracle and/or its affiliates

ゲーム難易度の調整やイベントの参加率をリアルタイムで分析したいと考えていました。

MySQL HeatWaveを導入したおかげで
想定最大の90倍もの高速化を実現し、
リアルタイム分析を毎時間行うことができます。

プログラム開発の負荷や事前の
想定処理時間では実現不可能と
思っていたので、本当に驚きです。

川本 昌之

ジニアス・ソリティ株式会社
取締役 / 最高技術責任者



HeatWave ML

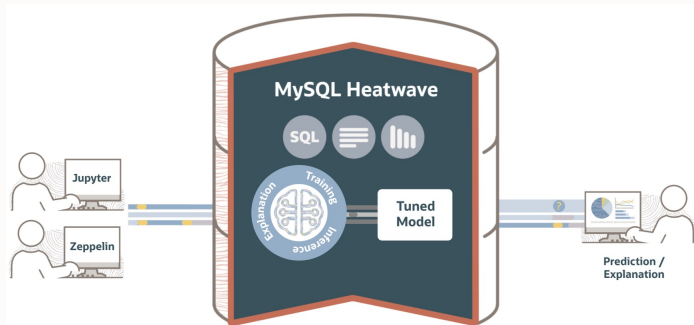


MySQLの機械学習プラットフォーム

HeatWave ML

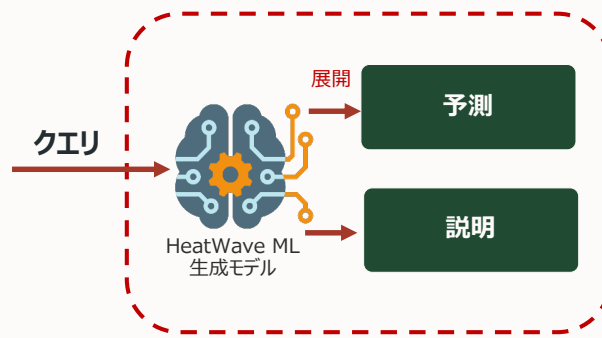
機械学習をより手軽に安全に活用できるプラットフォーム

- ✓ **追加料金不要**ですぐに利用可能
- ✓ Oracle AutoMLにより、機械学習工程を自動化
- ✓ HeatWave上で完結でき、外部にデータを出さずに実行可能



実業務に活用できる**説明可能性**をサポート

- ✓ ブラックボックス化しやすい推論の説明を明確化
- ✓ 特徴量が予測に与える影響を自動で判断
- ✓ **法令遵守、公平性**などの観点からモデルの挙動、有効性を見極め



高性能・高精度なモデル生成を誰でも実行できるインタフェース

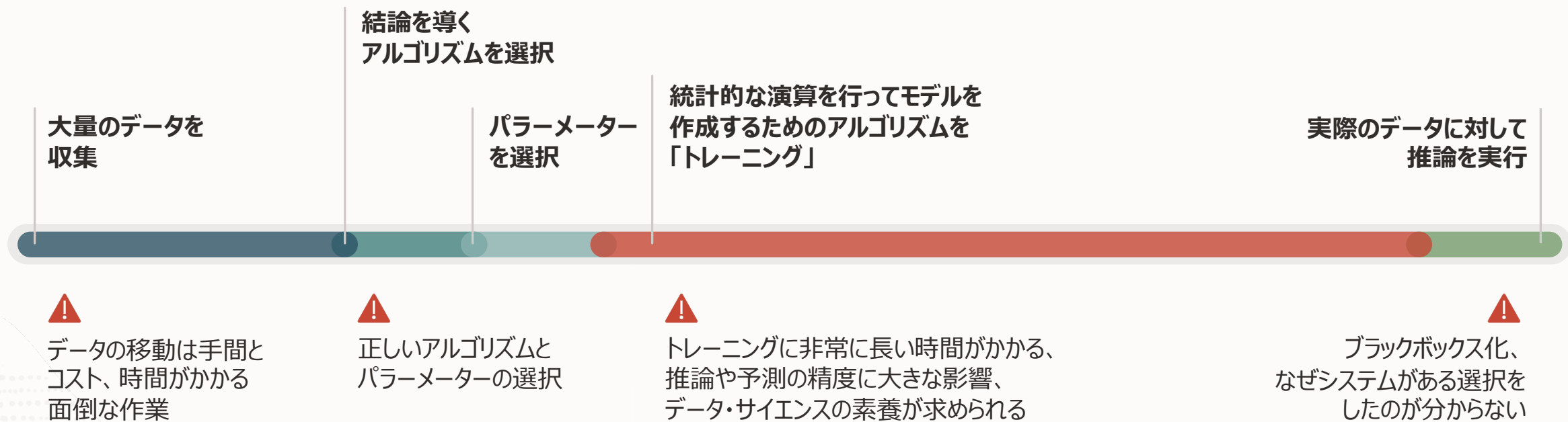
- ✓ 各工程に対応した関数を実行するだけで予測まで実行可能
- ✓ パラメータ、アルゴリズムの選択など**専門的な知識は不要**
- ✓ トレーニングの高速化により、迅速なモデル生成を支援

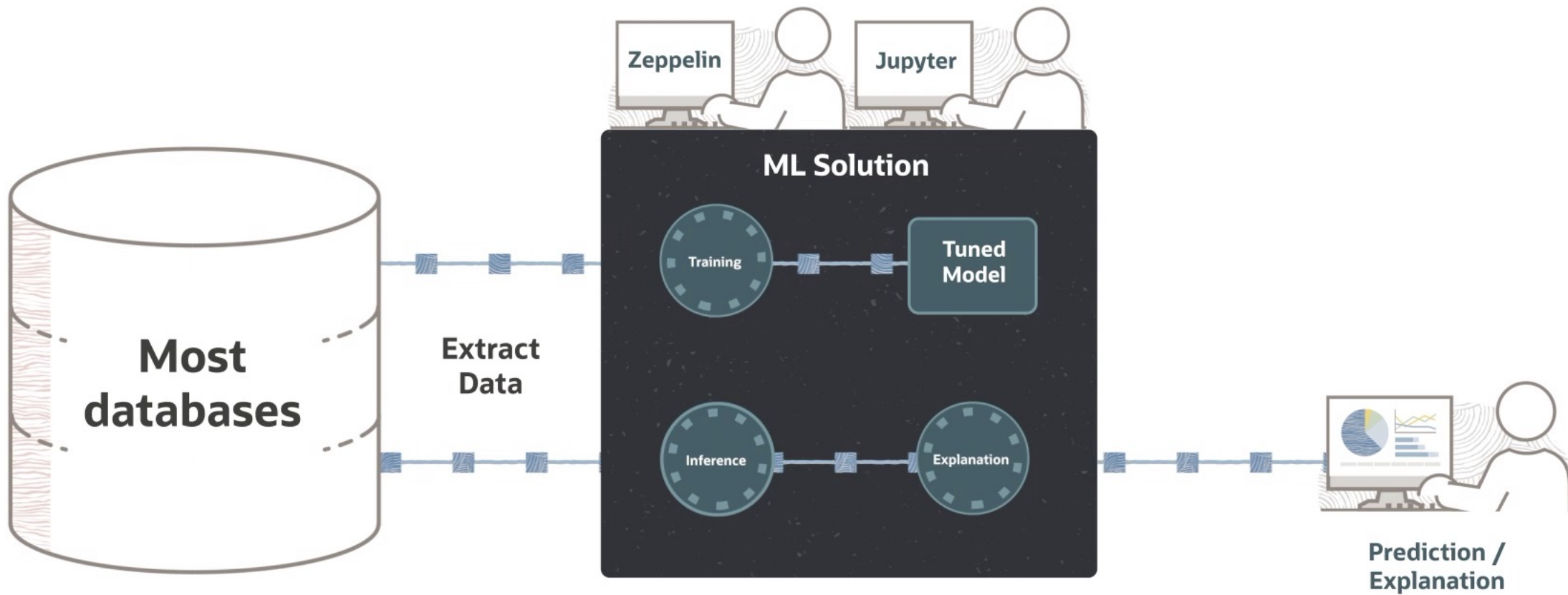
HeatWave MLは
Redshift MLより

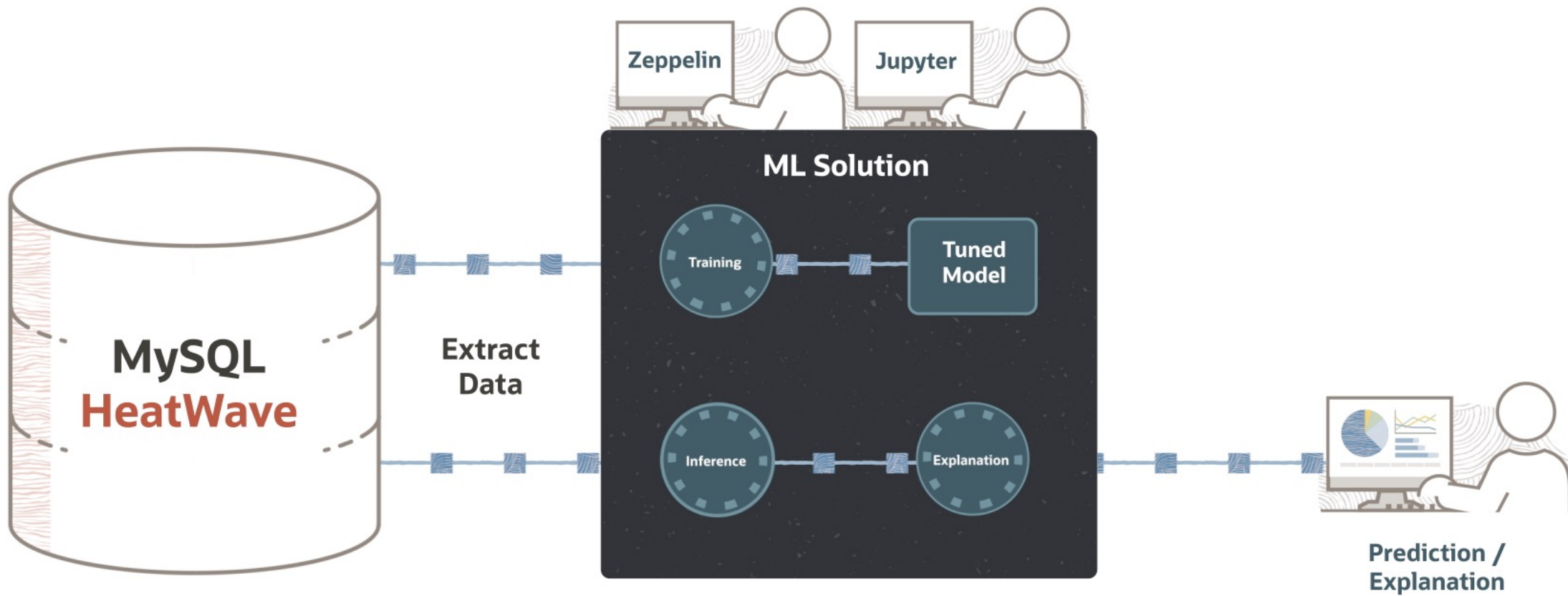
平均
25倍
高速

機械学習のきほんのき

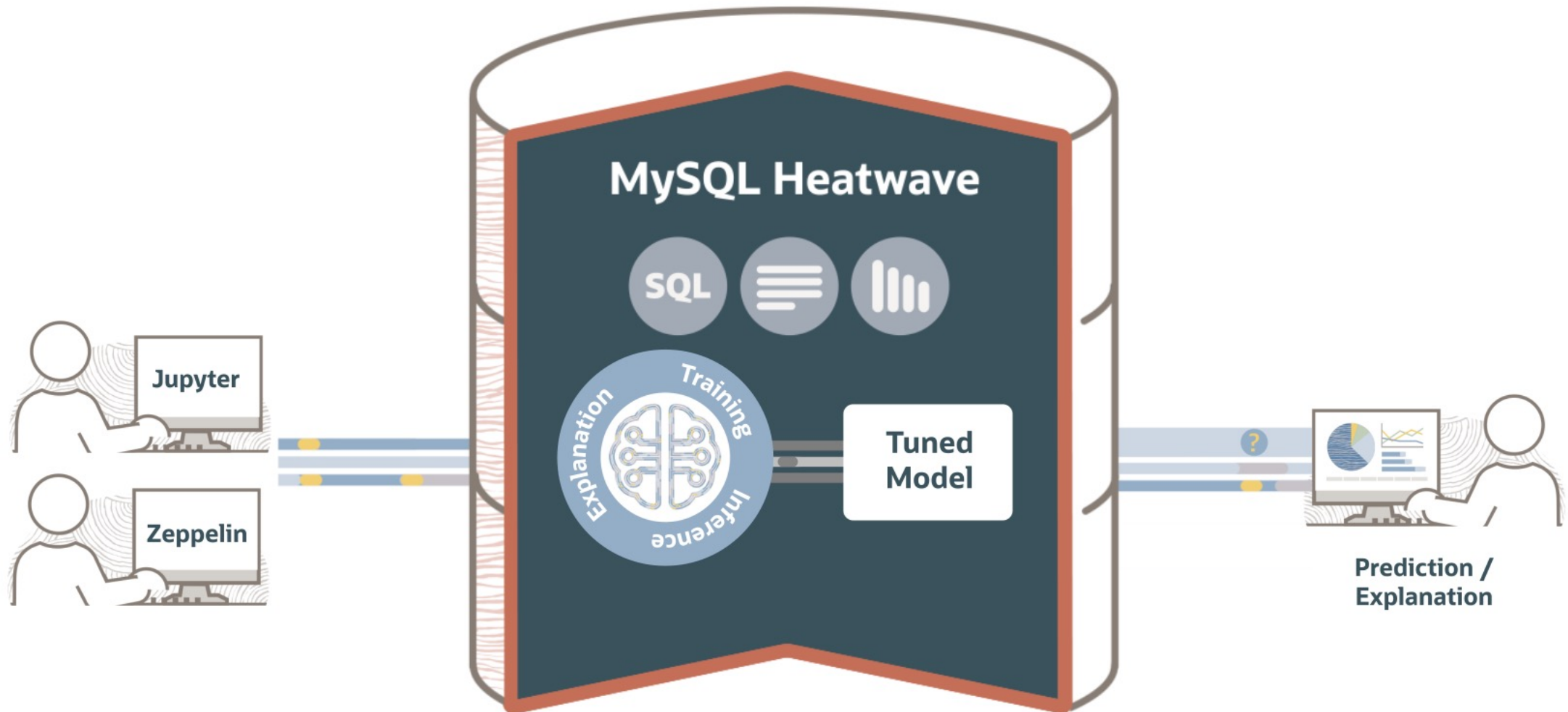
工程





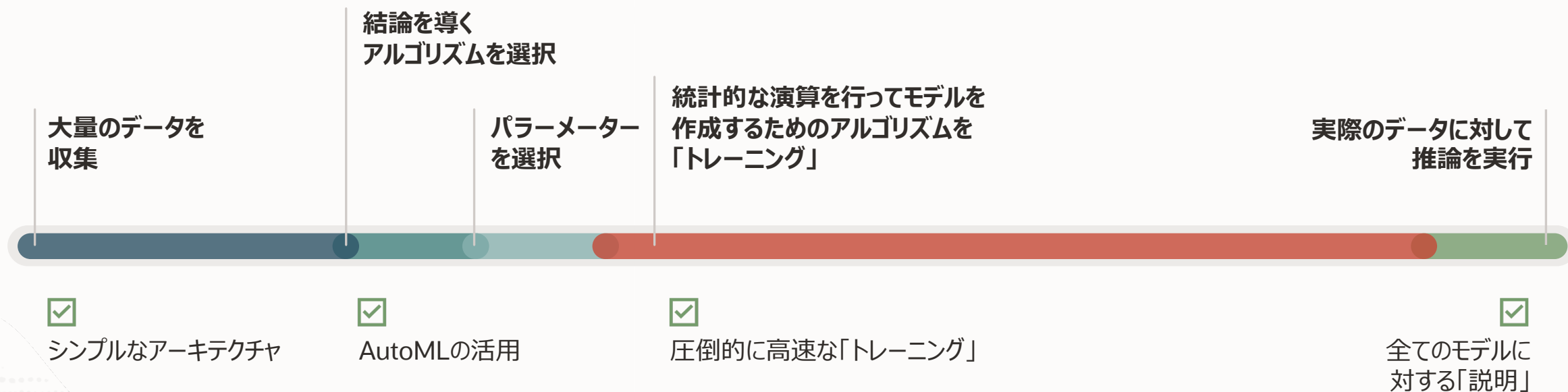


MySQL HeatWaveで機械学習は簡単に



機械学習のきほんのき

PROCESS



機械学習の悩みをHeatWaveが解決

MySQL HeatWaveは**本当**のAutoMLを提供



HeatWave MLがトレーニングを高速化できるポイント

HeatWave MLでのトレーニングは
繰り返し作業不要

高度に並列化された
ハイパーパラメーターのチューニング技法

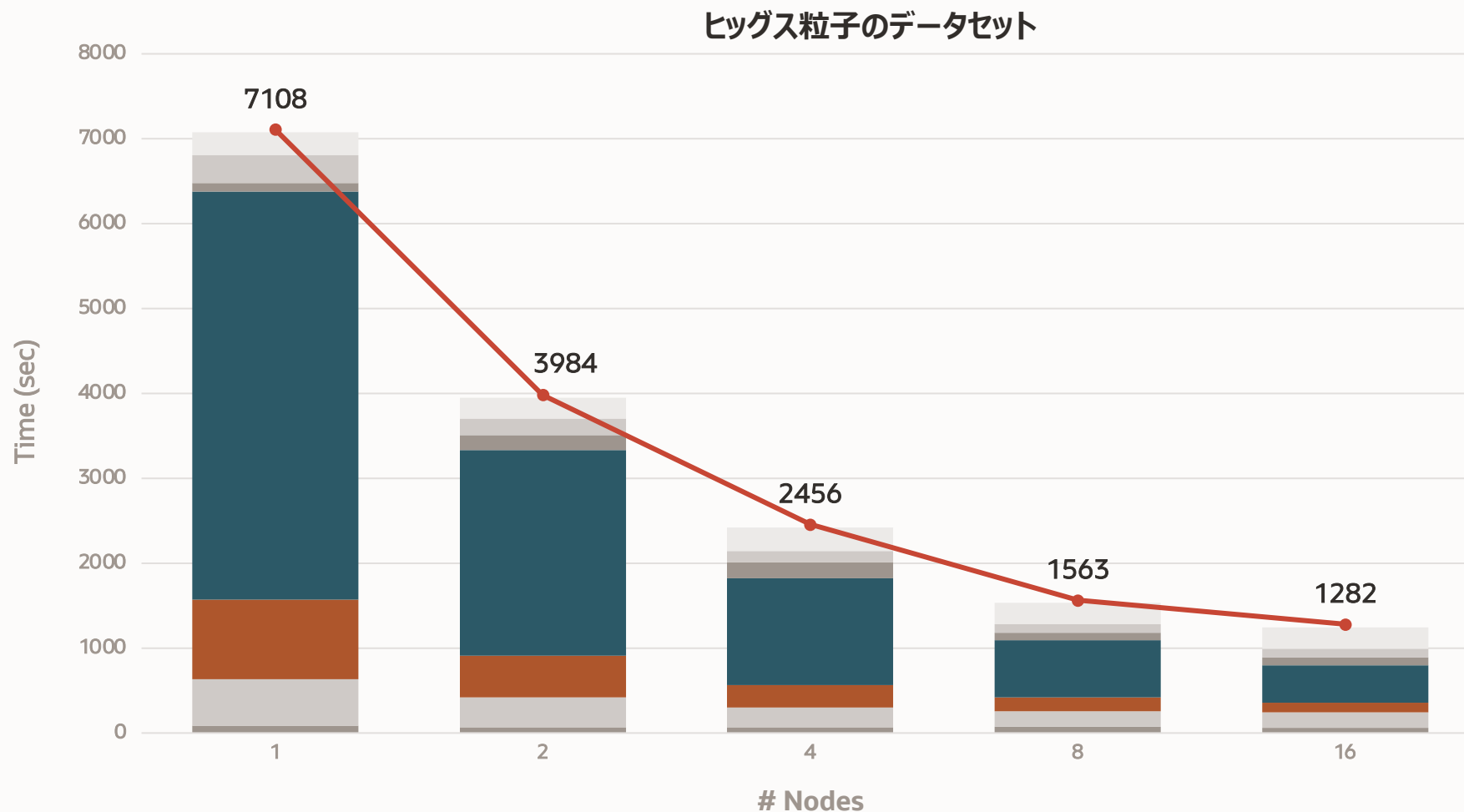
正確性を維持しながら行う
インテリジェントなサンプリング

トレーニングが**自動的に収束**

クラスターのサイズに応じて**性能がスケール**

HeatWave MLはクラスターのサイズに応じてスケール

時間のかかる工程を高速化



11倍

モデルのチューニング
を高速化

8倍

特徴量の選択
を高速化

MySQL HeatWaveは**本当**のAutoMLを提供

データセット

データの preprocessing

アルゴリズムの選択

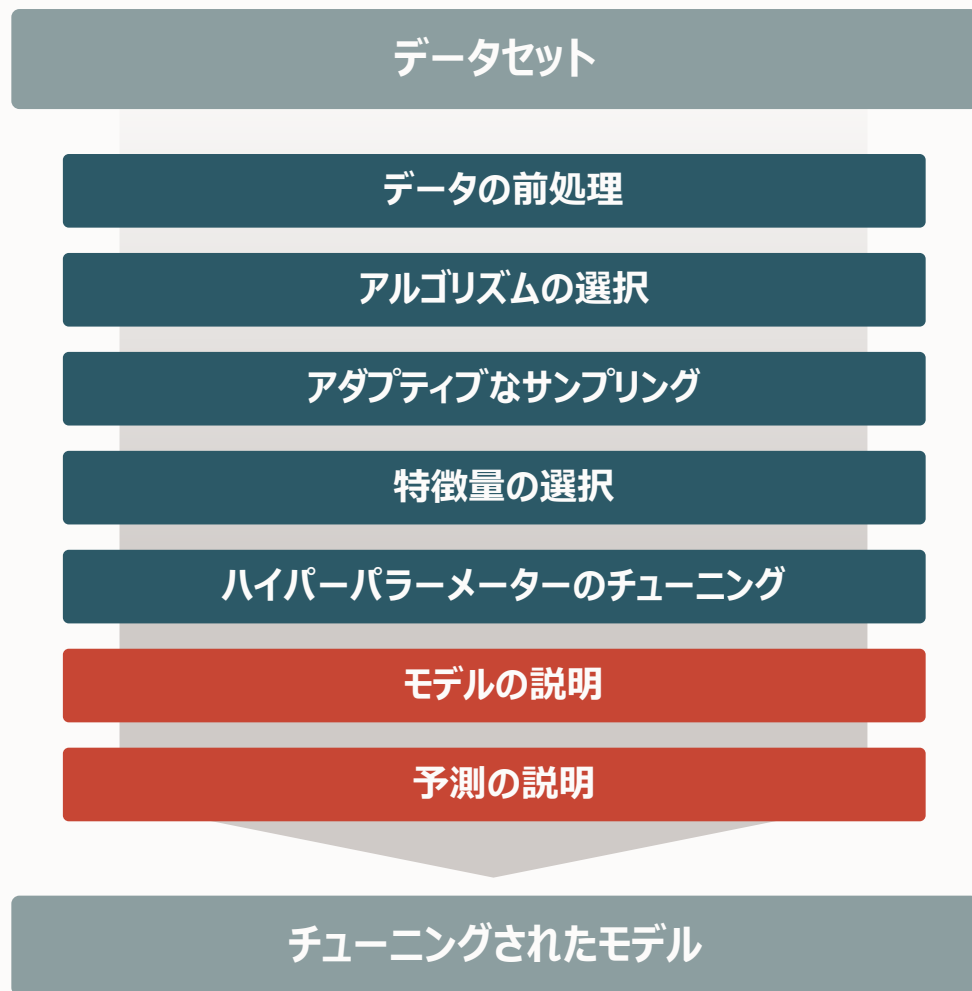
Adaptive sampling

Feature selection

Hyperparameter tuning

Tuned model

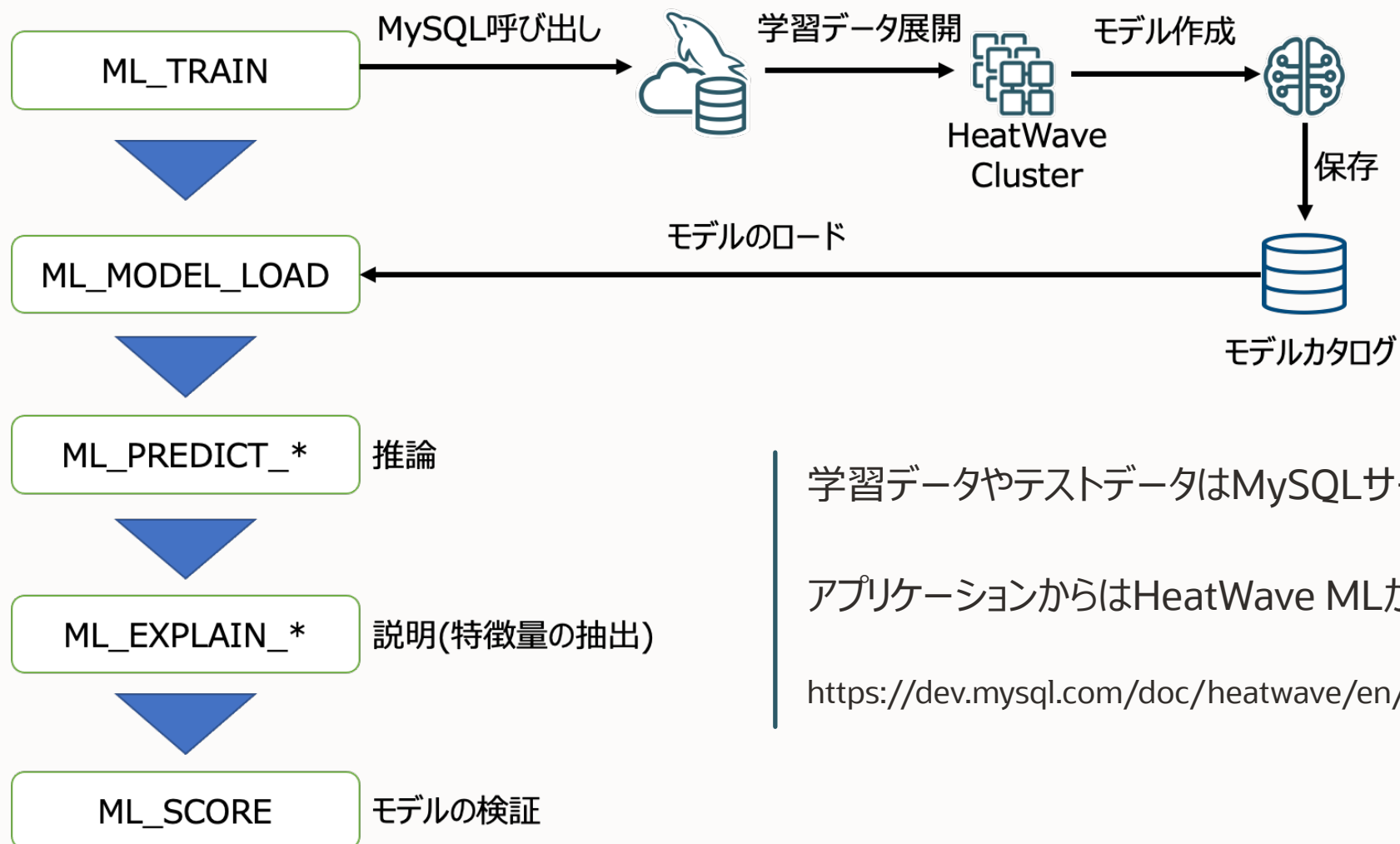
HeatWave MLで作成された全てのモデルは説明可能



HeatWave MLの説明可能性で実現

- 01 法令の遵守や規則への準拠
- 02 公正性
- 03 再現性
- 04 因果関係
- 05 信頼性

HeatWave MLでの機械学習処理の流れ



学習データやテストデータはMySQLサーバー上のテーブルとして管理

アプリケーションからはHeatWave MLが用意した関数群を実行

<https://dev.mysql.com/doc/heatwave/en/heatwave-machine-learning.html>

```
mysql> CALL sys.ML_TRAIN(mlcorpus.bank_market_train', 'y', NULL, @model);
```

```
Query OK, 0 rows affected (2 min 40.47 sec)
```

```
mysql> CALL sys.ML_MODEL_LOAD(@model, NULL);
```

```
Query OK, 0 rows affected (1.10 sec)
```

```
mysql> CALL sys.ML_PREDICT_TABLE('mlcorpus.bank_market_test_temp',  
                                @model, 'mlcorpus.predictions');
```

```
Query OK, 0 rows affected (1.74 sec)
```

```
mysql> SELECT id, duration, pmonth, balance, Prediction FROM predictions;
```

```
+-----+-----+-----+-----+-----+
| id | duration | pmonth | balance | Prediction |
+-----+-----+-----+-----+-----+
| 1 | 897 | jul | 4189 | yes |
| 2 | 768 | jul | 0 | yes |
| 3 | 223 | oct | 16873 | no |
| 4 | 1018 | jun | 844 | yes |
| 5 | 1337 | nov | 1315 | no |
| 6 | 630 | jul | 0 | no |
| 7 | 1030 | may | -311 | yes |
+-----+-----+-----+-----+-----+
```

モデルの説明

使いやすさ

品質

性能

```
mysql> SELECT model_explanation FROM ML_SCHEMA_root.MODEL_CATALOG  
WHERE model_handle=@model;
```

+-----+-----+-----+			
id	feature	impact	
+-----+-----+-----+			
1	duration	0.0658	
2	pmonth	0.0303	
3	pday	0.0152	
4	contact	0.0148	
5	poutcome	0.0138	
6	balance	0.0085	
7	housing	0.007	
8	age	0.0059	
+-----+-----+-----+			

このモデルで最も影響の大きい特徴量

直感的には影響が大きそうだが、実際には影響が小さい

モデルの説明

使いやすさ

品質

性能

```
mysql> CALL sys.ML_TRAIN(mlcorpus.bank_market_train', 'y', NULL, @model);
```

```
Query OK, 0 rows affected (2 min 40.47 sec)
```

```
mysql> CALL sys.ML_MODEL_LOAD(@model, NULL);
```

```
Query OK, 0 rows affected (1.10 sec)
```

```
mysql> CALL sys.ML_PREDICT_TABLE('mlcorpus.bank_market_test_temp',  
                                @model, 'mlcorpus.predictions');
```

```
Query OK, 0 rows affected (1.74 sec)
```

```
mysql> SELECT id, duration, pmonth, balance, Prediction FROM predictions;
```

+-----+-----+-----+-----+-----+					
id	duration	pmonth	balance	Prediction	
+-----+-----+-----+-----+-----+					
1	897	jul	4189	yes	
2	768	jul	0	yes	
3	223	oct	16873	no	
4	1018	jun	844	yes	
5	1337	nov	1315	no	
6	630	jul	0	no	
7	1030	may	-311	yes	
+-----+-----+-----+-----+-----+					

モデルの説明

定期預金開設 の要因:

- ▲ 15分の勧誘の通話
- ▲ 勧誘の電話を7月に実施
- ▼ 口座の残高は0ユーロ

定期預金拒否 の要因:

- ▼ 勧誘の電話を10月に実施
- ▼ 4分の勧誘の通話
- ▲ 口座の残高は16,873

HeatWave MLと他のクラウド・データベースのソリューションとの比較

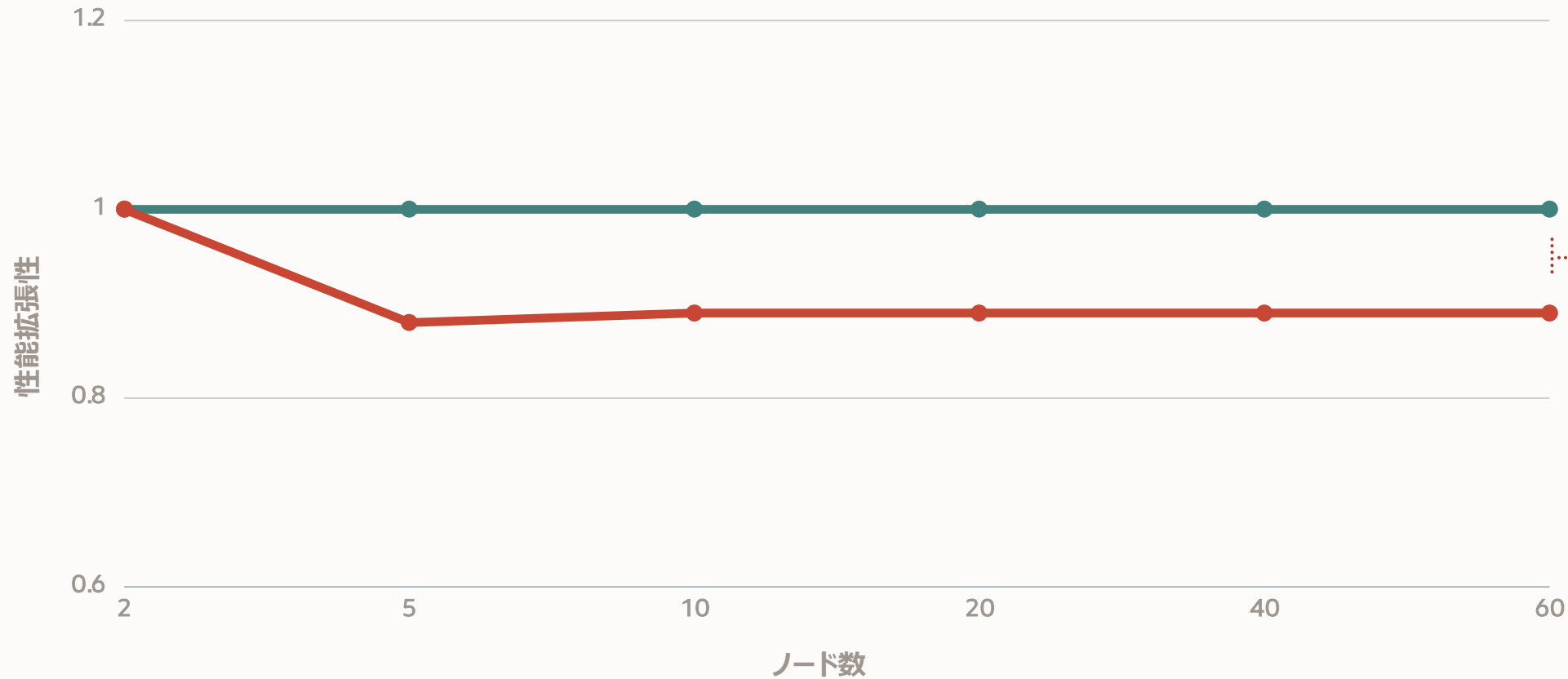
機能	HeatWave ML	Redshift ML	Snowflake ML
API	SQL	SQL	Java, Scala, Pythonでプログラミング
MLが内包されたデータベース	Yes	No	No
データとモデルをDB内に格納	Yes	No	No
全てのモデルに対する説明可能性	Yes	No	No
完全な自動化	Yes	No	No
ノードを追加して実行可能	Yes	No	No



最新のイノベーション

機械学習を実装
リアルタイムの拡張性
コスト削減
パフォーマンス

MySQL HeatWaveの高い性能拡張性



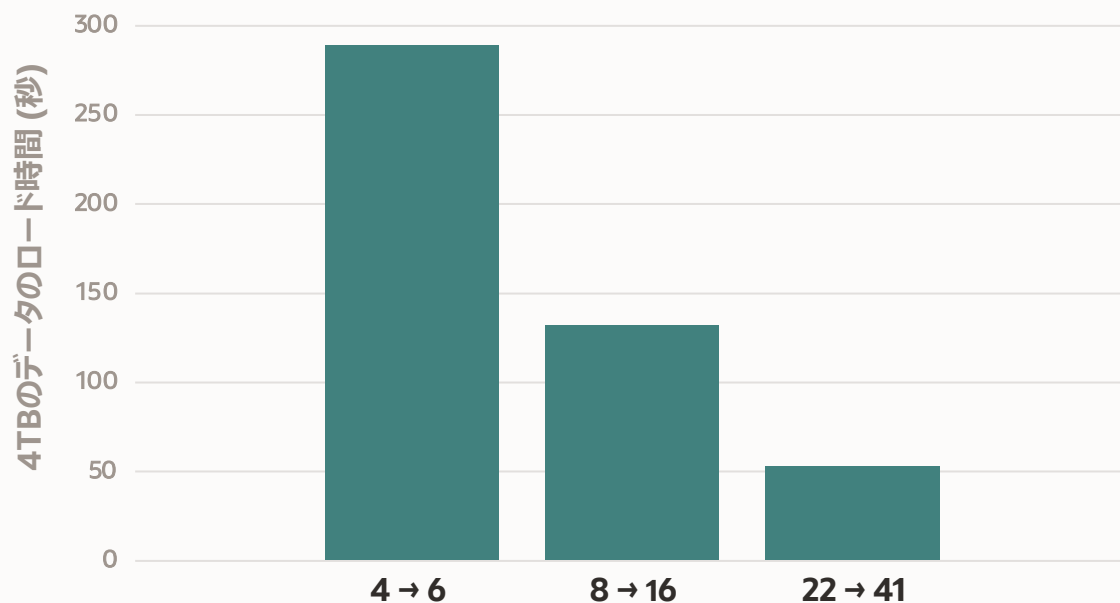
ほぼ
リニア

● 理論上の性能拡張性 ● 実際の性能拡張性

性能の伸縮性

高速、予測可能、クラスターはいつでも利用可能

大きなクラスターサイズへの変更による
データロード時間の短縮



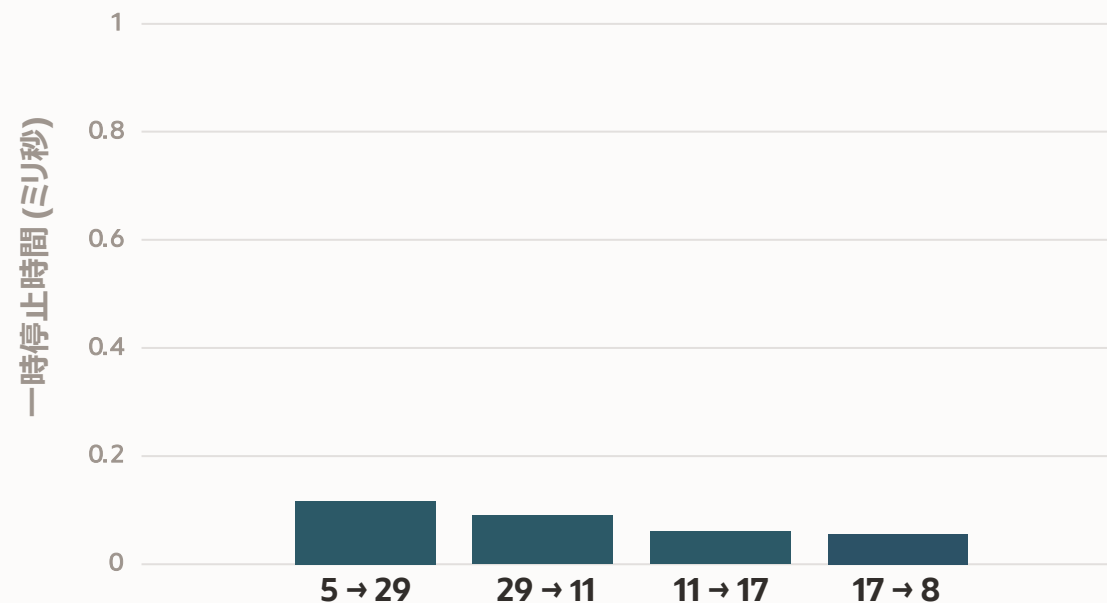
高速化

オブジェクトストレージの
ほぼ帯域性能に相当

予測可能

時間はノードあたりの
データサイズに依存

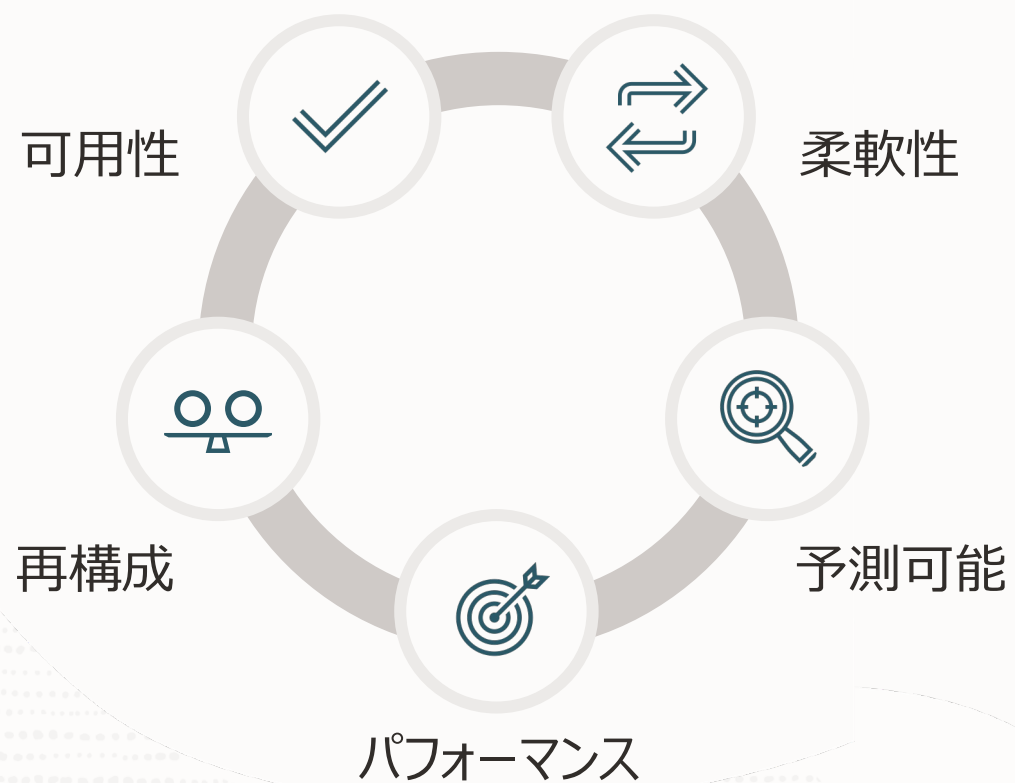
クエリの一時的停止時間



いつでも利用可能

HeatWaveノードのメモリー内に格納された
パーティション構成情報の更新の間だけ一時停止

リアルタイムの伸縮性



最新のイノベーション

機械学習を実装

リアルタイムの拡張性

コスト削減

パフォーマンス

2倍

ノードあたりのデータ量

ほぼ

半分のコスト

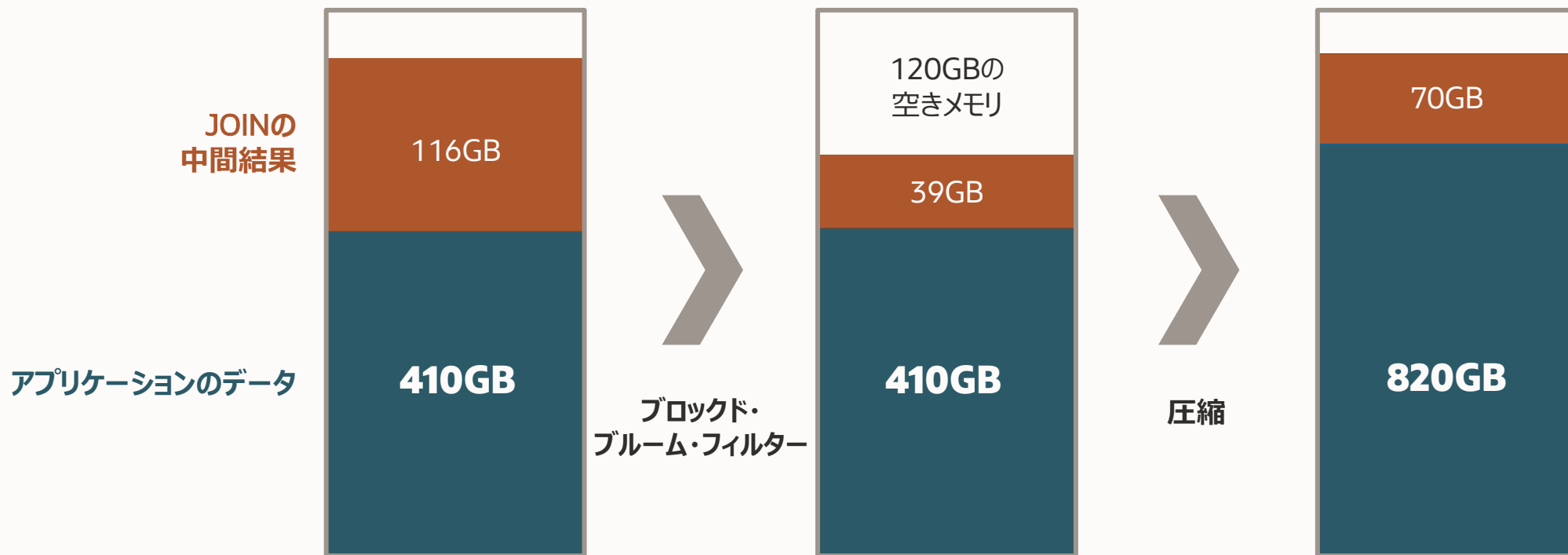
同じ

費用対効果

TPC-H 4TB	2021年8月 10 HeatWave, 1 MDSノード	最新のテスト 5 HeatWave, 1 MDSノード
	ノードあたりのデータ量	820G
	410G	
	パフォーマンス	11.5秒
	6.35秒	
	年間コスト	\$18,585
	\$34,073	
	コストパフォーマンス	\$0.007
	\$0.007	

ノードあたりの2倍のデータ量

ブルームフィルターを活用して中間結果のメモリ利用量を削減



TPC-DS benchmark

より複雑なスキーマ、
複雑なクエリ、データの偏り

TPC-DSの特性

多数のテーブルとカラム

複雑なスキーマ: 複数の多次元データベース、
第3正規形とスタースキーマのハイブリッド

複雑なクエリ

データ分散の偏り

現実に沿ったデータ量の増加

より多くのクエリ

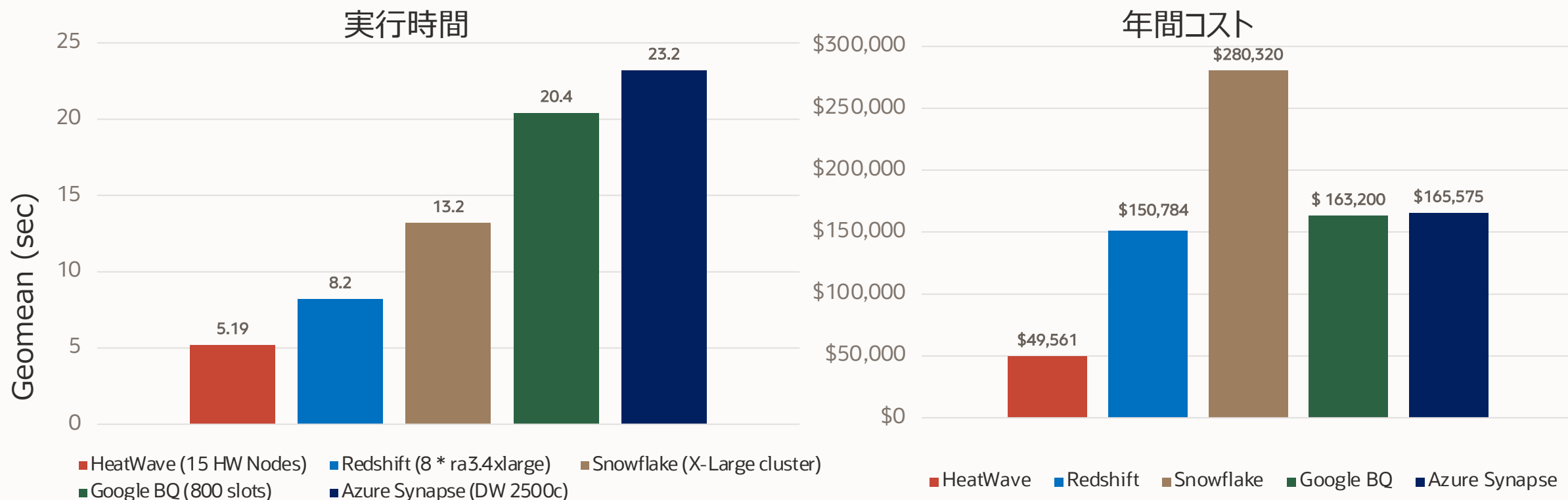
複数の大きなファクトテーブルのJOIN

クエリの最適化が複雑

費用対効果の比較:

HeatWave, Redshift, Snowflake, Google Big Query, Azure Synapse

10TB TPC-DS: 対Redshift 4.8倍, 対Snowflake 14.4倍, 対BigQuery 12.9倍, 対Synapse 14.9倍



- Redshiftの価格はリザーブドインスタンスの1年分先払い、Snowflakeは Standard Edition で計算
- Google Big Queryは定額料金、月額コミットメント、For Azure Synapseは1年間のリザーブドの価格で計算

3rd party numbers derived on Feb 15, 2022

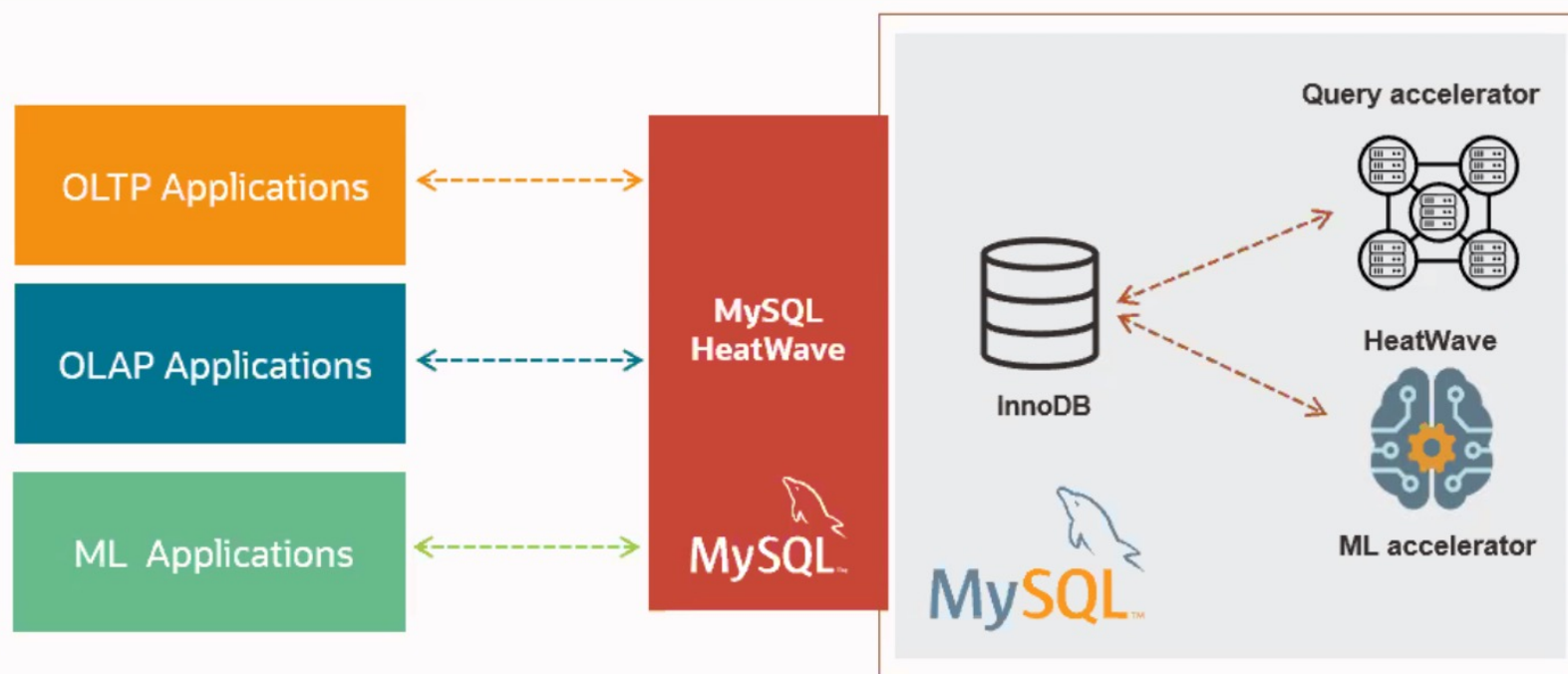
Using PAYG pricing for Snowflake. Other prices are based on 1 year pricing

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



HeatWave ML – MySQL HeatWaveの内部で機械学習エンジンが動作

追加費用不要でトランザクション処理、分析処理、予測処理を一つのデータベースで



- MySQL HeatWave内のデータに対して機械学習
→ ETL不要 & セキュリティの向上
- オラクルのAutoMLによって機械学習の工程を自動化
→ 作業効率の向上
- 少数のSQL関数の実行のみで機械学習の各工程を実行可能
→ DB利用者との親和性
- Amazon Redshift MLより高い予測精度をもち「トレーニング」は25倍以上高速かつ1%のコスト



Thank you