

ORACLE
DevLive

Level Up

MySQL Summit

MySQL Autopilot

Automating Application Development with MySQL
HeatWave

Nipun Agarwal

Senior Vice President

MySQL HeatWave

Oracle

Seema Sundara

Architect

MySQL HeatWave

Oracle



Table of Contents

MySQL Autopilot Overview 3

Autopilot for OLTP 7

Autopilot for OLAP 15

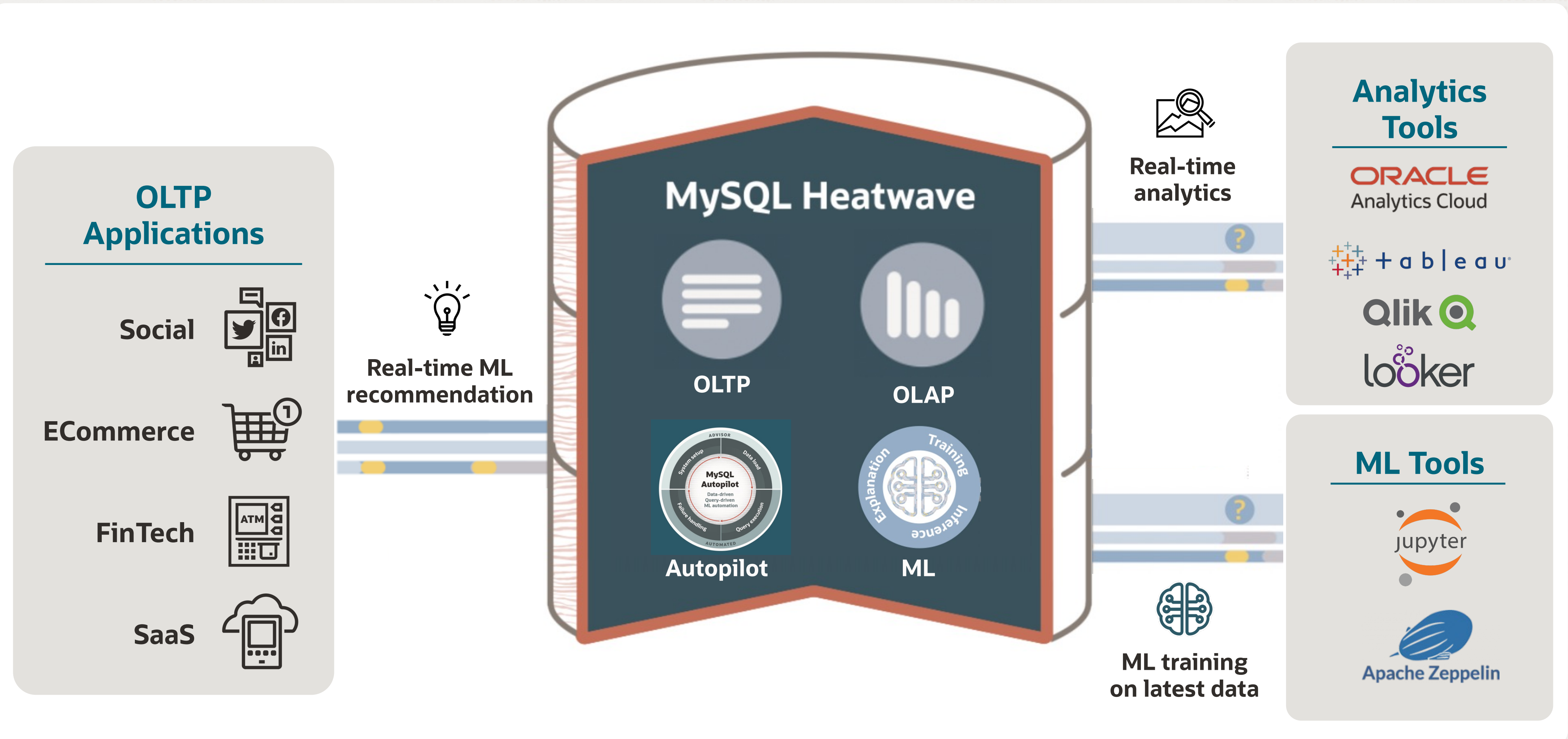
Autopilot for Lakehouse 22

Summary 28

MySQL Autopilot Overview

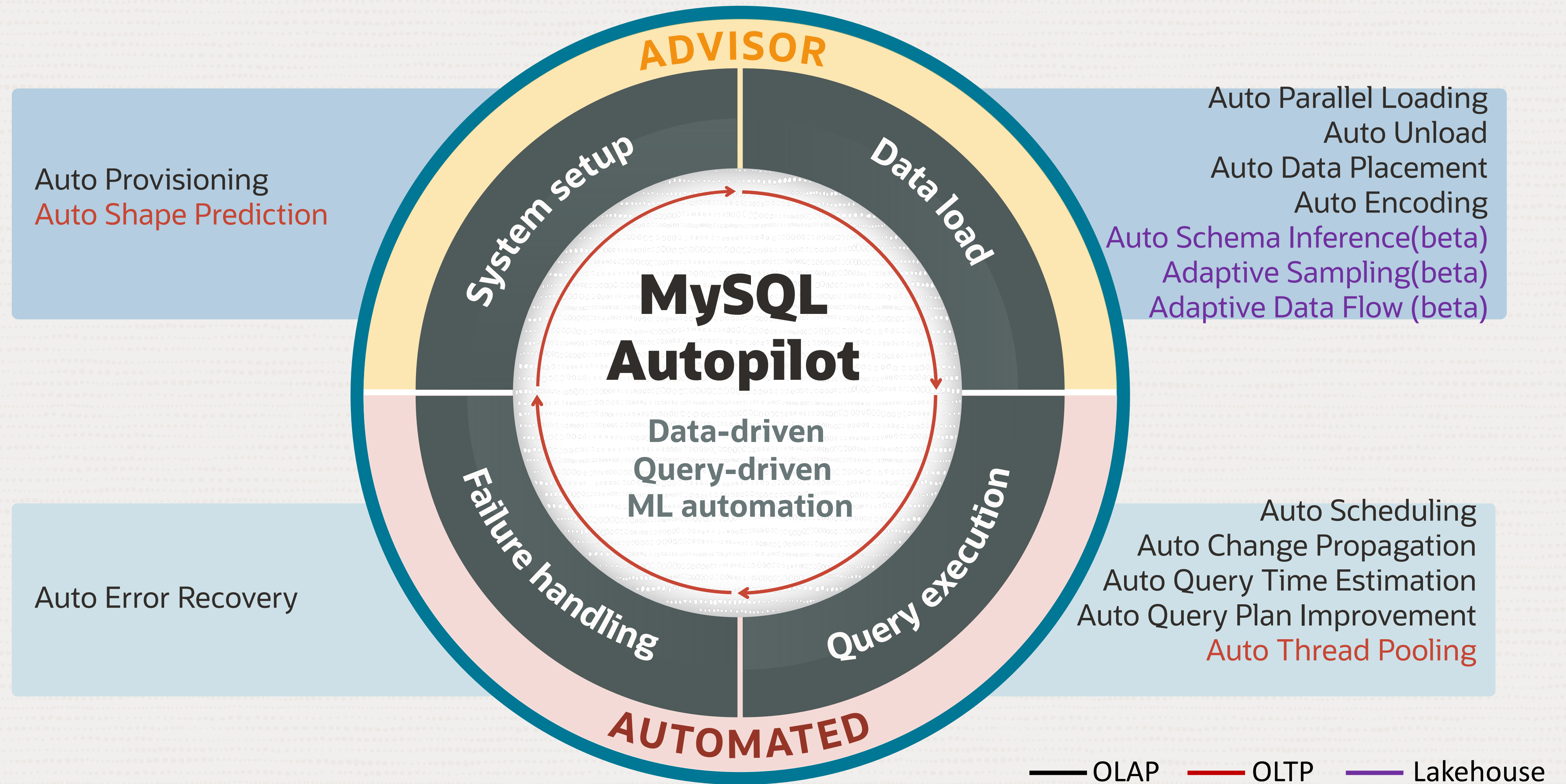


MySQL HeatWave – One Database for OLTP, OLAP, ML

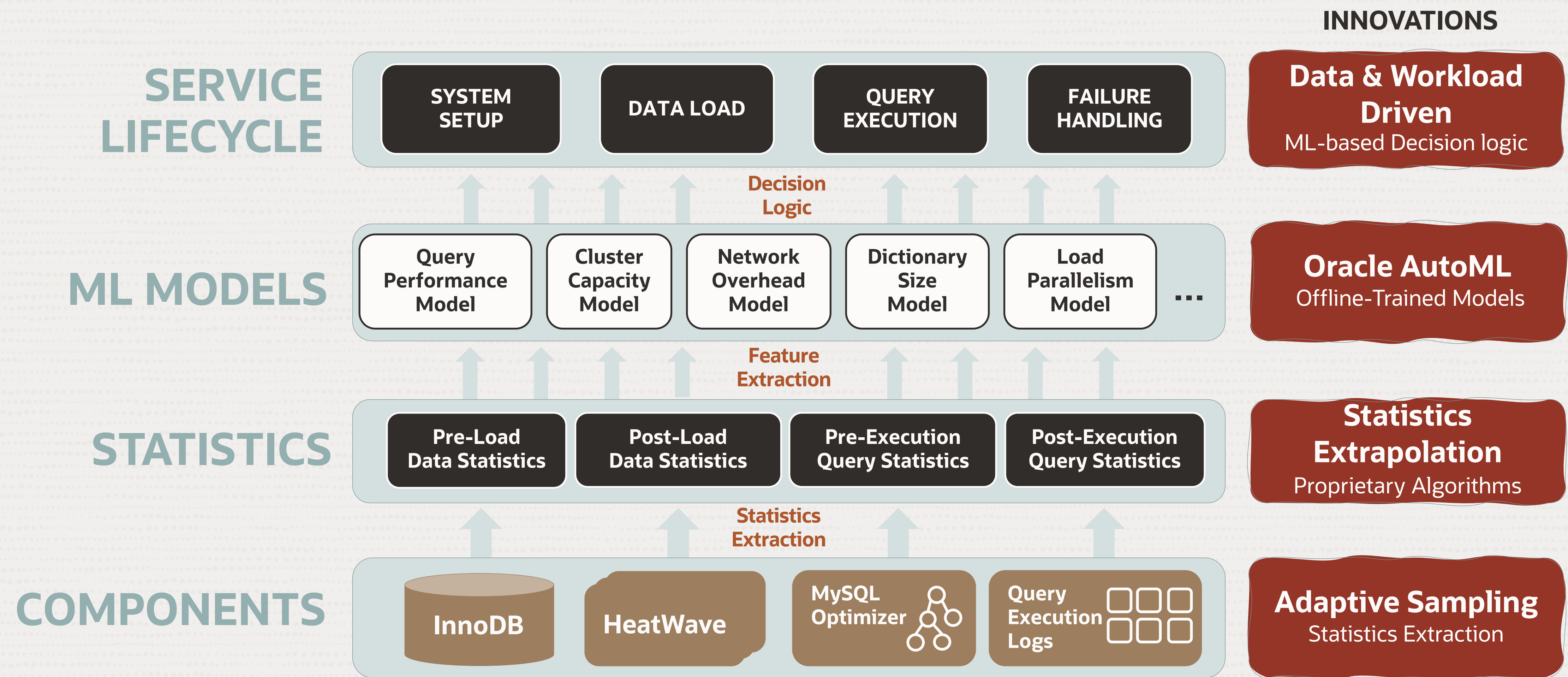


MySQL Autopilot

ML based automation to improve system performance and scalability



MySQL Autopilot Architecture

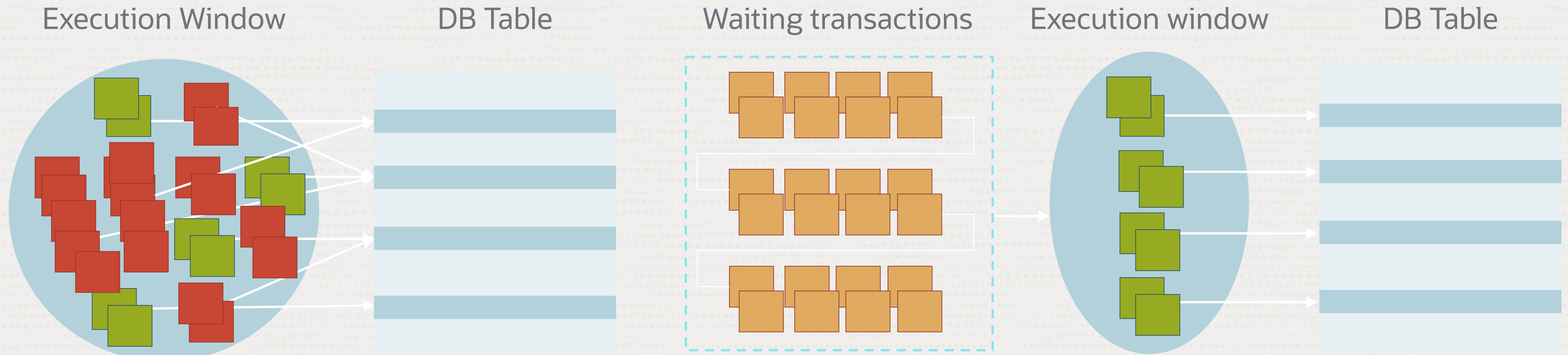


MySQL Autopilot for OLTP



Auto Thread Pooling

Workload aware admission control



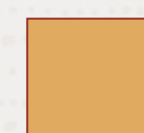
Admission control

✗ Blocked TXs create resource contention on executing TXs

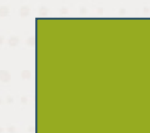
✓ Allowing fewer TXs reduces resource contention



Blocked TX



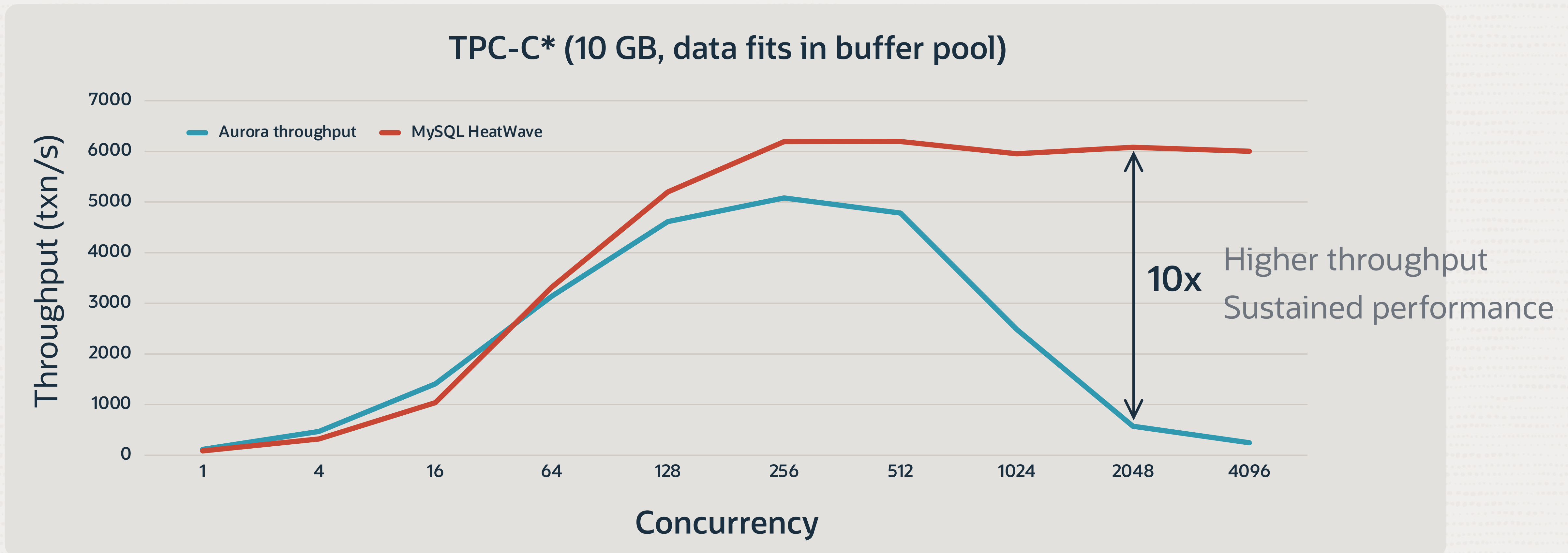
Waiting TX



Executing TX

MySQL HeatWave is 10x better than Aurora for OLTP

Auto Thread Pooling

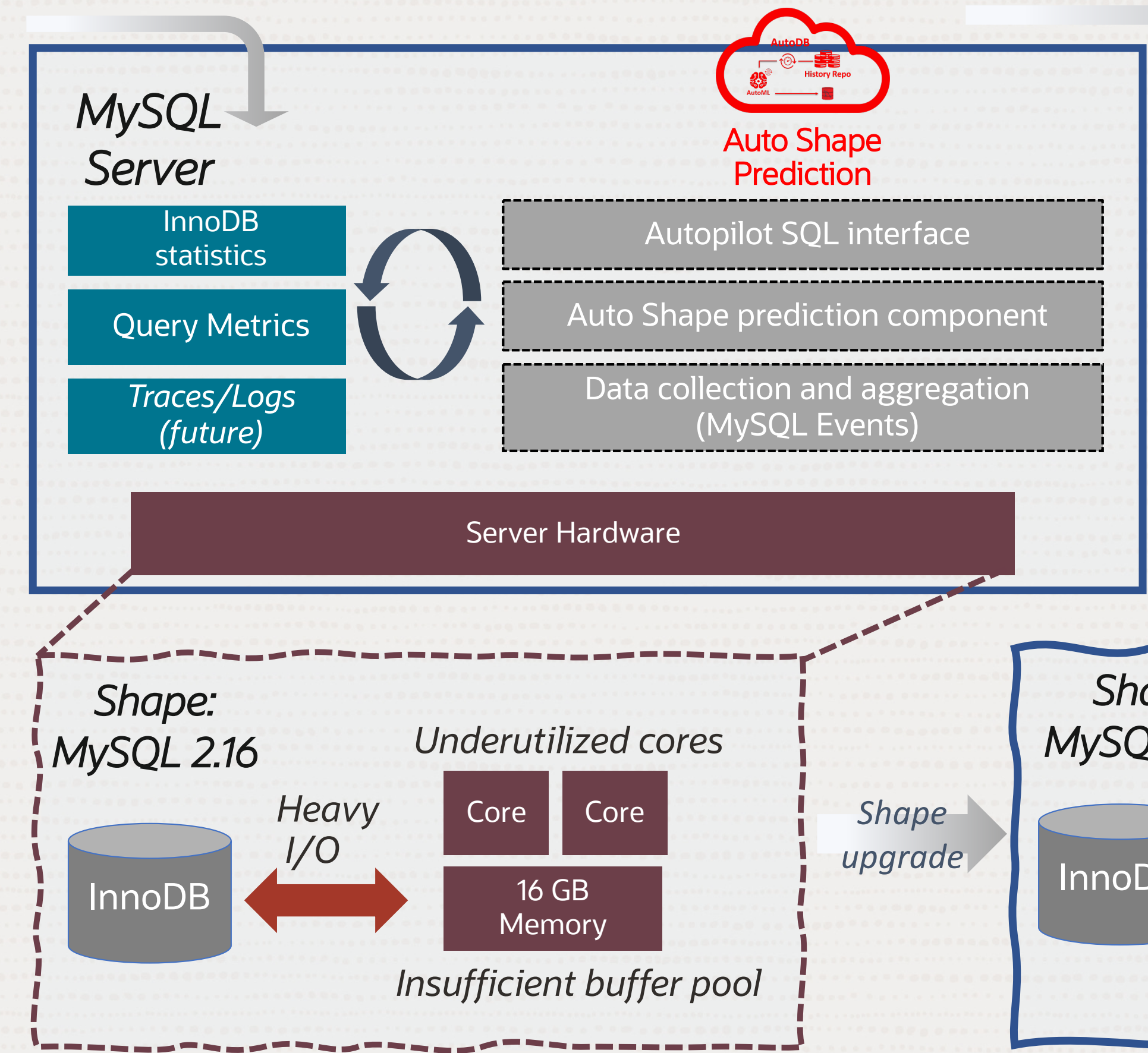


*Benchmark queries are derived from the TPC-C benchmarks, but results are not comparable to published TPC-C benchmark results since these do not comply with the TPC-C specifications.

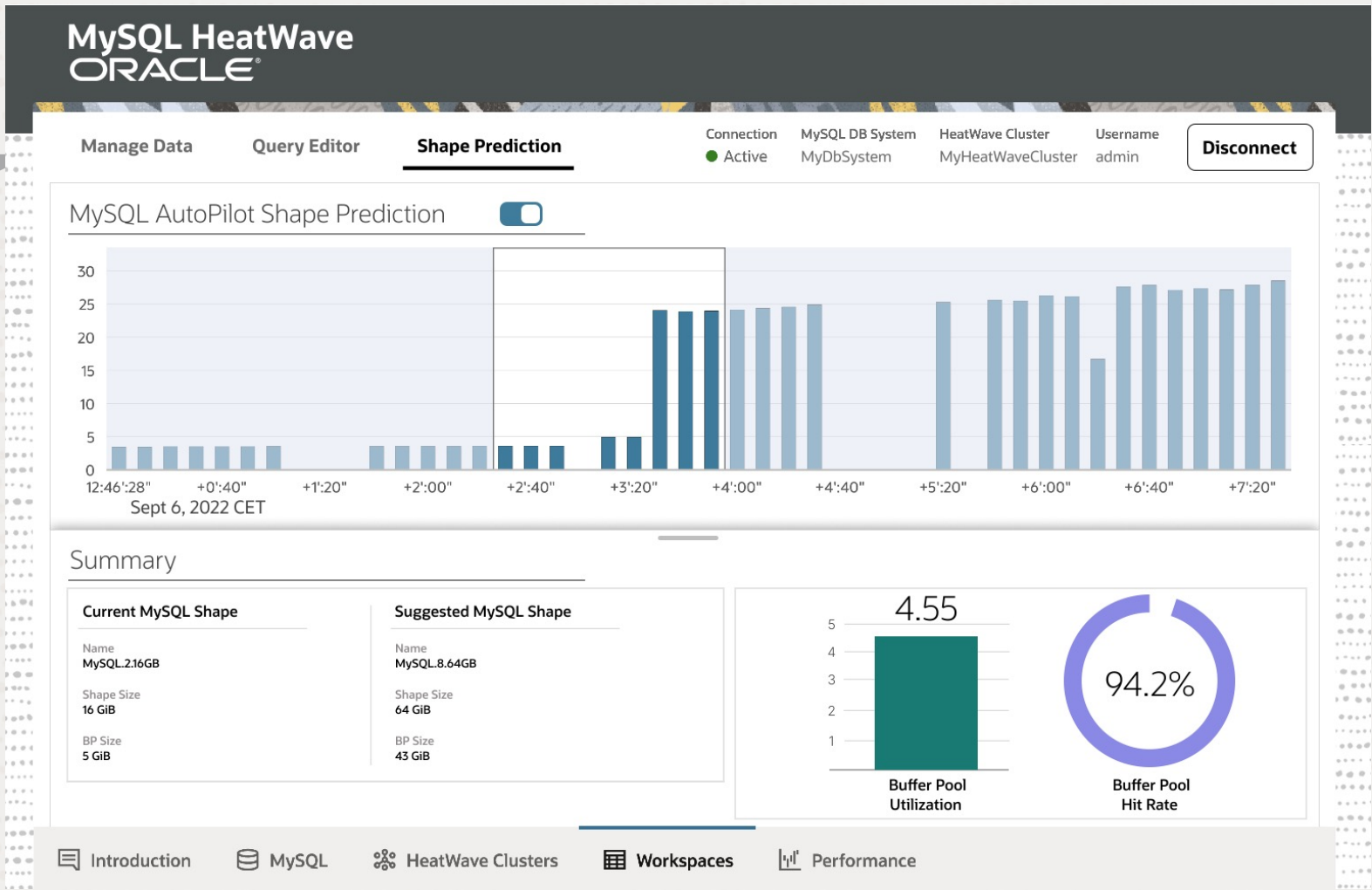
Auto Shape Prediction

Determine the optimal instance size

OLTP Workload
MySQL Events & MySQL Component
Constant workload monitoring & prediction



**Timeseries interface:
Console & SQL**
Showing predictions over time



Auto Shape Prediction for best OLTP Performance

Upsize/downsize between small shape MySQL.2.16GB and large shape MySQL8.64GB

Benchmark	BP Size (GB)	Buffer pool Hit Rate	Outcome		Throughput Improvement	Buffer pool Hit Rate	
Twitter	5	89%	UPSIZE	<div>MySQL.8.64GB ↑ MySQL.2.16GB</div> <div>→</div>	6.2x	100%	<div>• Improved performance</div> <div>• Better user experience</div>
Epinions	5	94%	UPSIZE		2x	99.6%	
Smallbank	5	95%	UPSIZE		1.5x	100%	
Sysbench TPC-C	5	93%	UPSIZE		1.8x	99.4%	
Twitter	48	100%	DOWNSIZE	<div>MySQL.8.64GB ↓ MySQL.4.32GB</div>	1.02x	100%	<div>• Cost saving with smaller instance</div> <div>• Performance remains the same</div>
YCSB	48	100%	DOWNSIZE		1.04x	100%	
Smallbank	48	100%	DOWNSIZE		0.98x	99.2%	
Sysbench TPC-C	48	100%	DOWNSIZE		0.96x	100%	

New functionality in Shape Prediction

- **HCS console integration:**
 - Users can see the recommended shape from the console
 - Users can see with historical predictions (up to one week)
 - Users can interact with predictions
- **Continuous monitoring:**
 - The data collection and predictions will be enabled upon DBsystem start by default
 - Therefore the users do not need to enable it themselves
- **Explanations:**
 - Buffer pool statistics: Users can see related buffer pool statistics that explain the suggestion.
 - **New:** Workload detection: Users can see when an active workload is detected

HCS Console: Auto Shape Prediction Summary view

Showing only recommended shape with no additional explanations

MySQL HeatWave
ORACLE

hcsaws-test-user_grp@oracle.com

ClusterWorkloadAutopilot Shape Advisor

MySQL DB System + HeatWave Cluster
4.32-MySQL-only-config

Summary

DB System Name
4.32-MySQL-only-config

DB System State
Active

Cluster Name
-

HeatWave State
-

Cluster Shape
-

Cluster Size
-

Time Periods

Selected

Mon, 20 Mar 2023 09:04:06 GMT

Mon, 20 Mar 2023 14:09:06 GMT

Available

Fri, 17 Mar 2023 16:59:06 GMT

Mon, 20 Mar 2023 14:59:06 GMT

MySQL Shape Prediction

	Current	Recommended
Shape Name	MySQL.4.32GB	MySQL.8.64GB
Memory Size	32 GiB	64 GiB
Buffer Pool Size	15 GiB	43 GiB

Recommended Action

Upsize

1. Backup the database system

2. Restore the backup, select shape: MySQL.8.64GB and default MySQL configuration

See the documentation for more information.

> Buffer Pool Insights

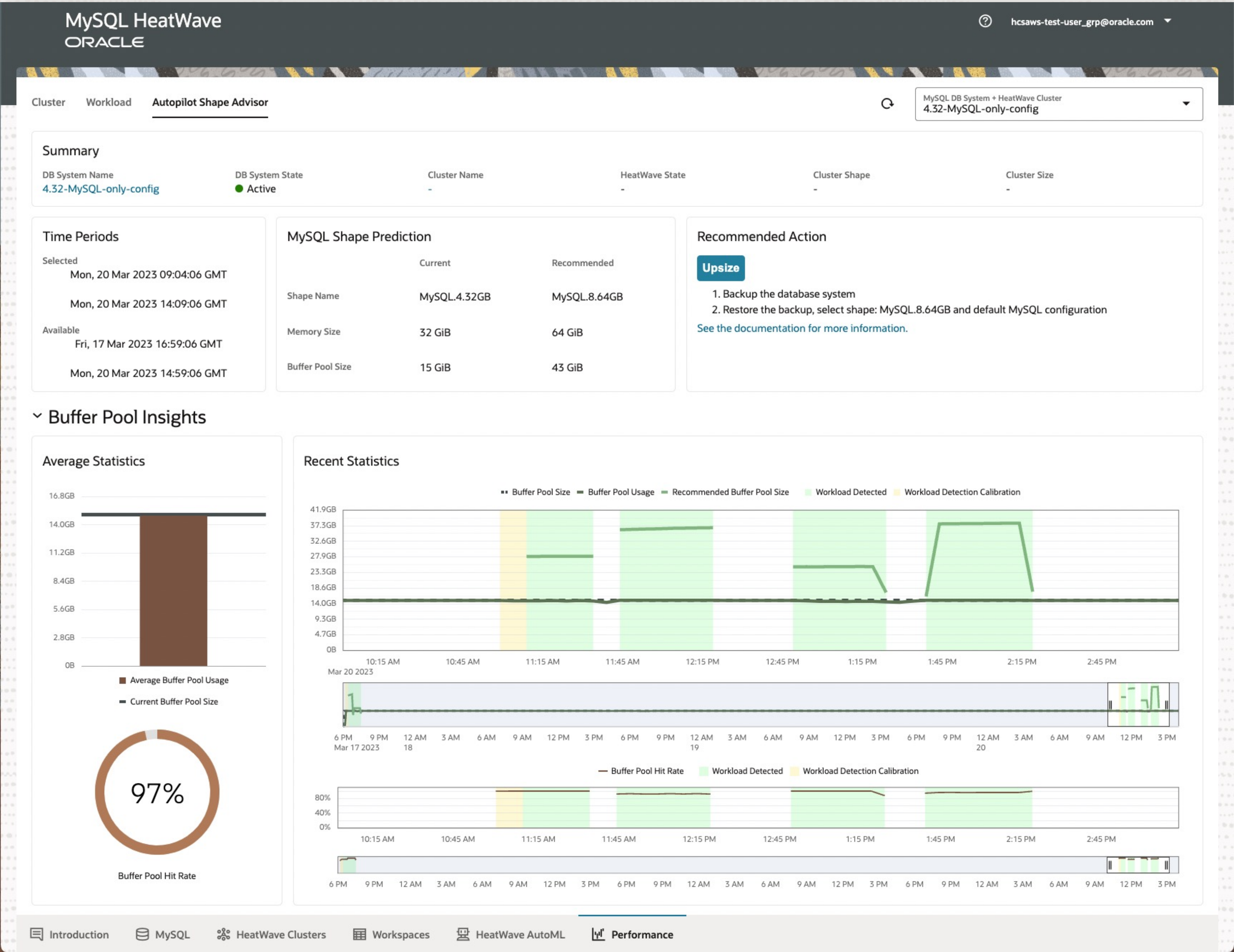
IntroductionMySQLHeatWave ClustersWorkspacesHeatWave AutoMLPerformance



Complete view

Showing explanations:

- 1. Buffer pool hit rate
- 2. Buffer pool utilization
- 3. Workload detection markers
 - *Green regions indicate when an active workload is detected*
 - *Yellow regions indicate when detection is in progress*
 - *White regions indicate idle server*

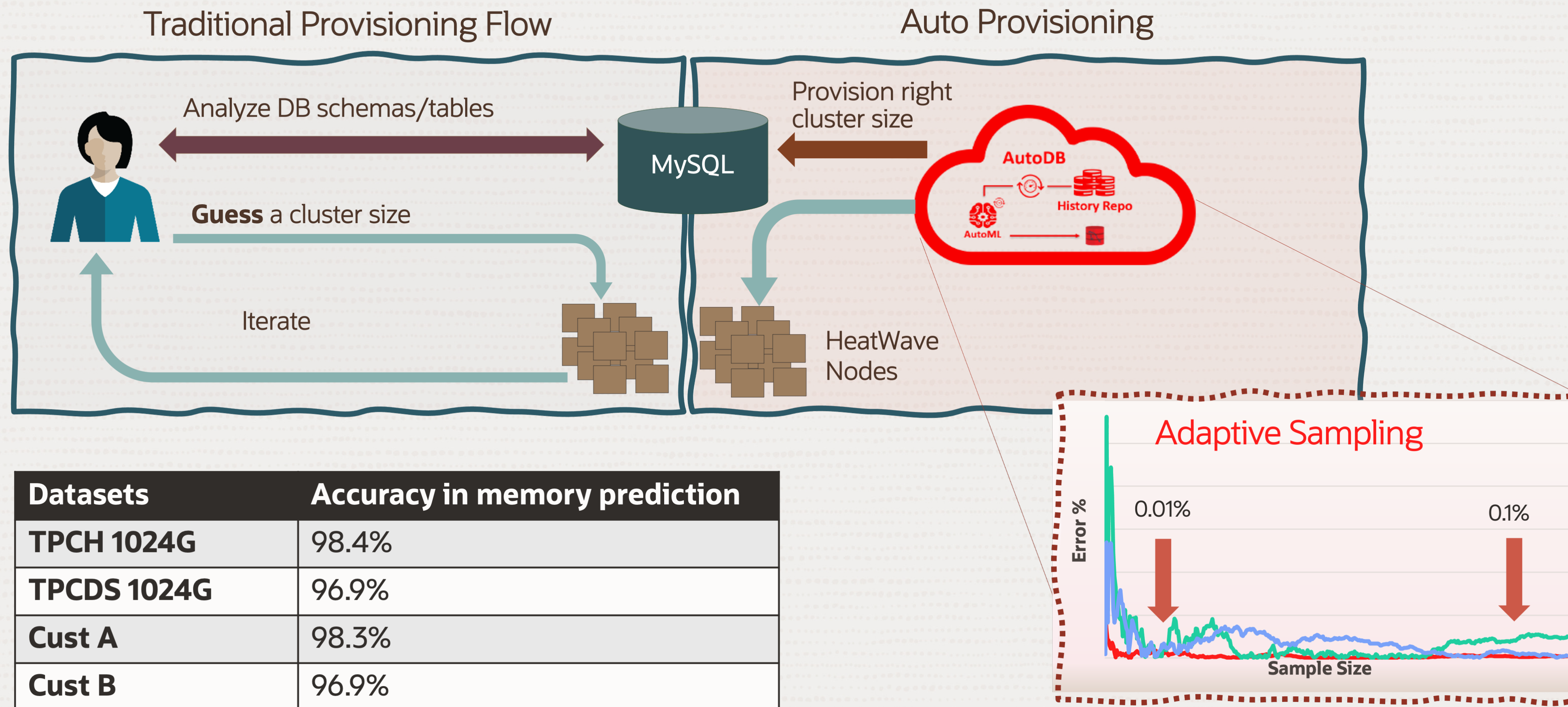


MySQL Autopilot for OLAP



Auto Provisioning

Machine learning prediction of memory usage to estimate cluster size



Auto Parallel Load/Unload

- Ease of use for customers to load/unload many tables in a single command
- Automatically decides the best degree of parallelism for load of individual tables

```
CALL sys.heatwave_load(db_list, [NULL | JSON_OBJECT(options)])
```

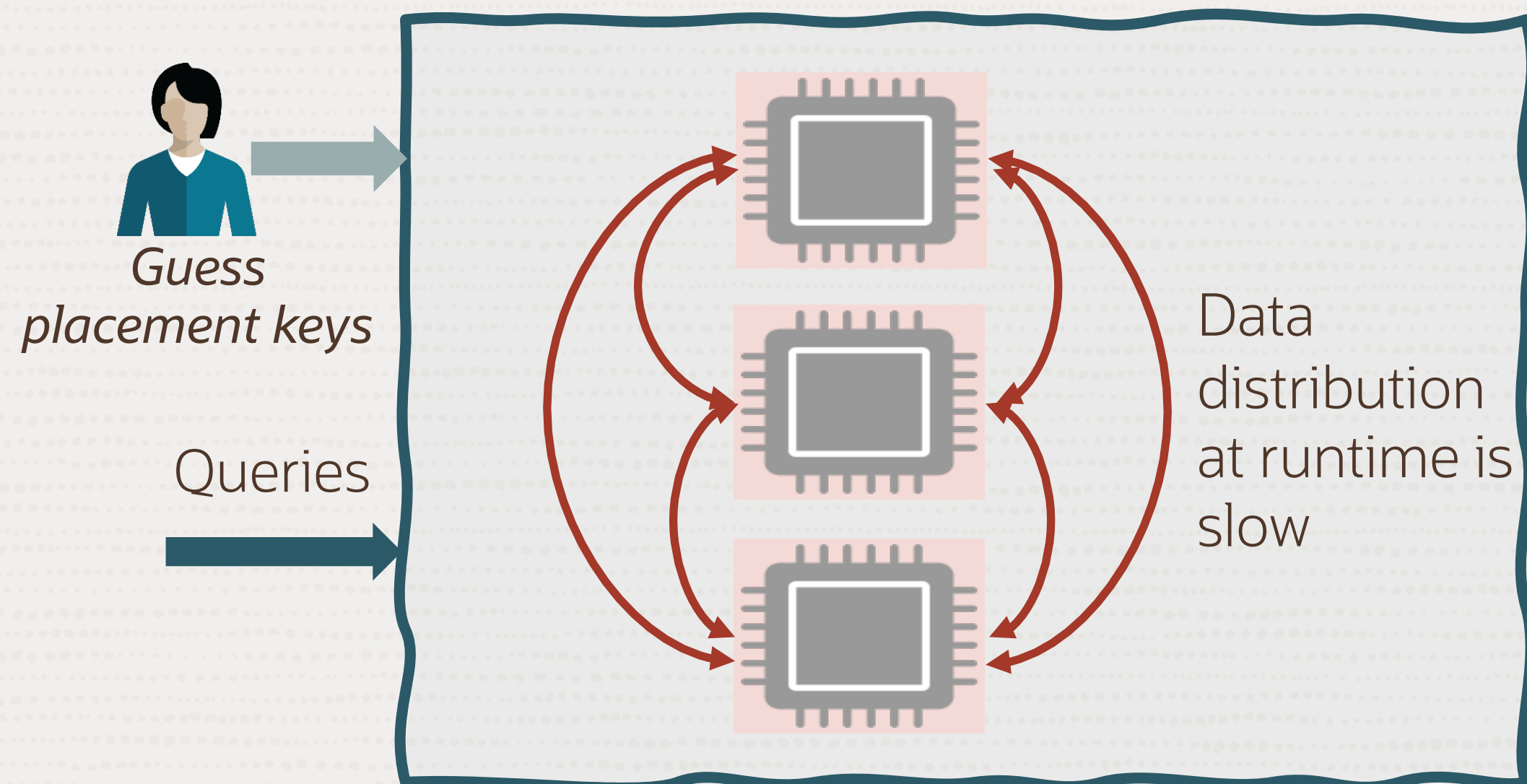
```
CALL sys.heatwave_unload (db_list, [NULL | JSON_OBJECT (options)])
```

- Auto Unload Advisor recommends unloading unused tables based on recent query activity
 - Frees up space that can be used for query execution
 - Potentially other “hot” tables can be loaded

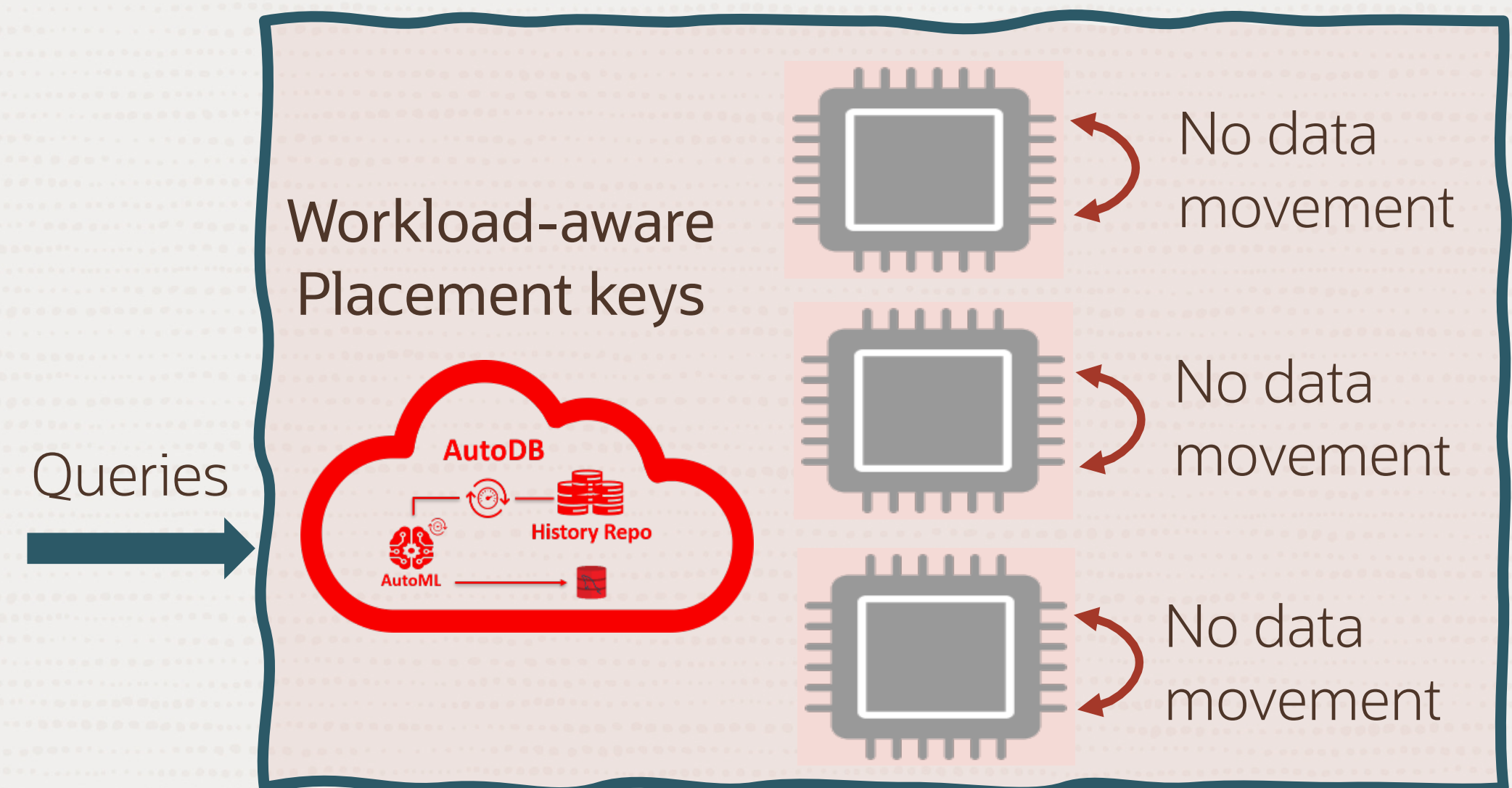
Auto Data Placement

Machine learning prediction of optimal in-memory partitioning column

Manual Data Placement



Auto Data Placement

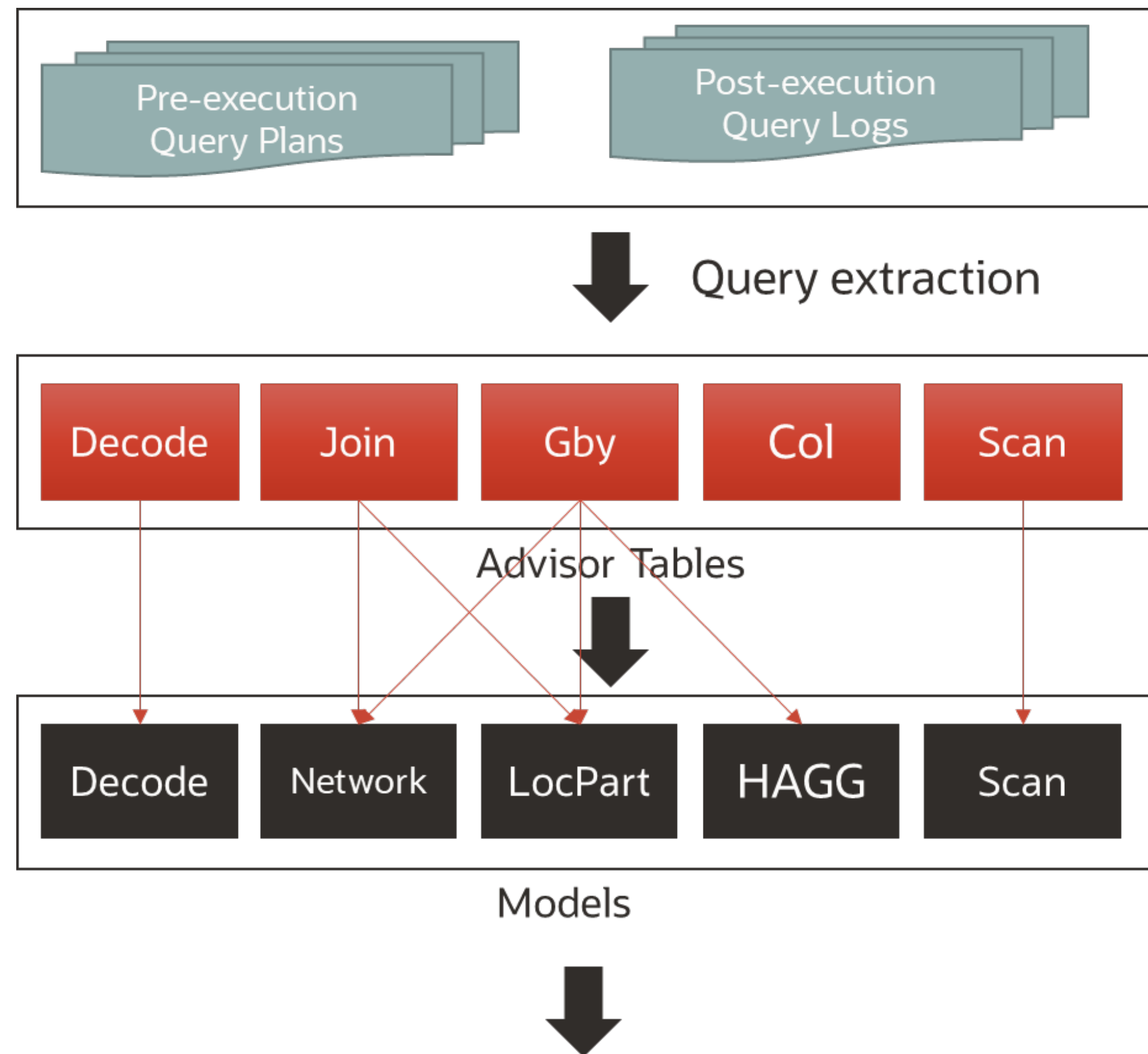


Dataset	Time with Primary Key placement	Predicted Improvement with Autopilot	Actual Improvement from Autopilot
TPCH 1024	332 sec	26%	37%
TPCH 4096	373 sec	20%	25%

- Predicts optimal columns to partition data in-memory
- Predicts runtime improvement

Auto Encoding

Incorporates run time in addition to memory usage

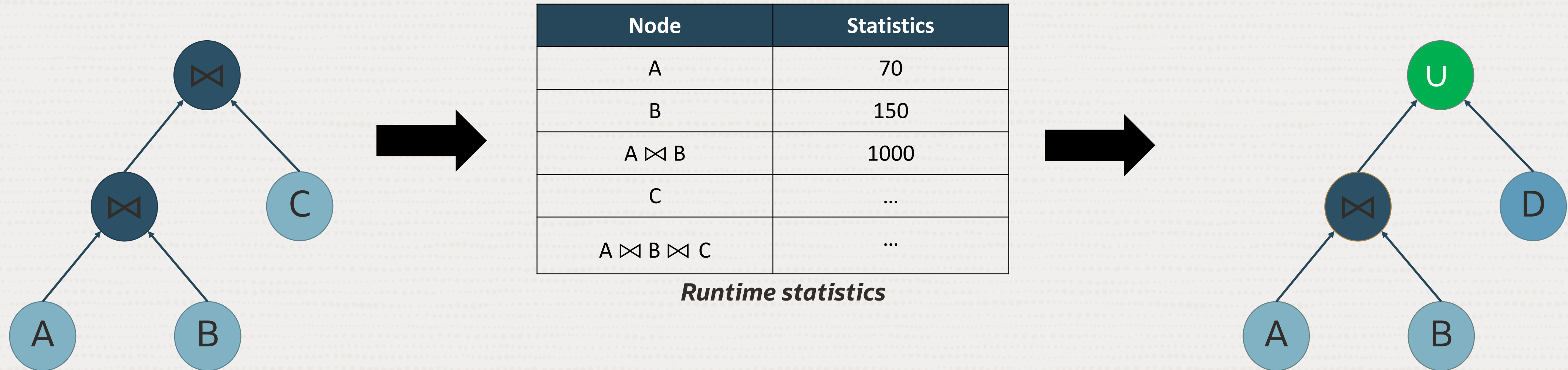


TPC-H 30 T	Runtime	
Memory optimized encoding	1187s	} 2x
New encoding	597s	

- Estimate decode time
- Predicts runtime improvement

Auto Query Plan Improvement

Optimizer learns and improves query plan based on queries executed earlier

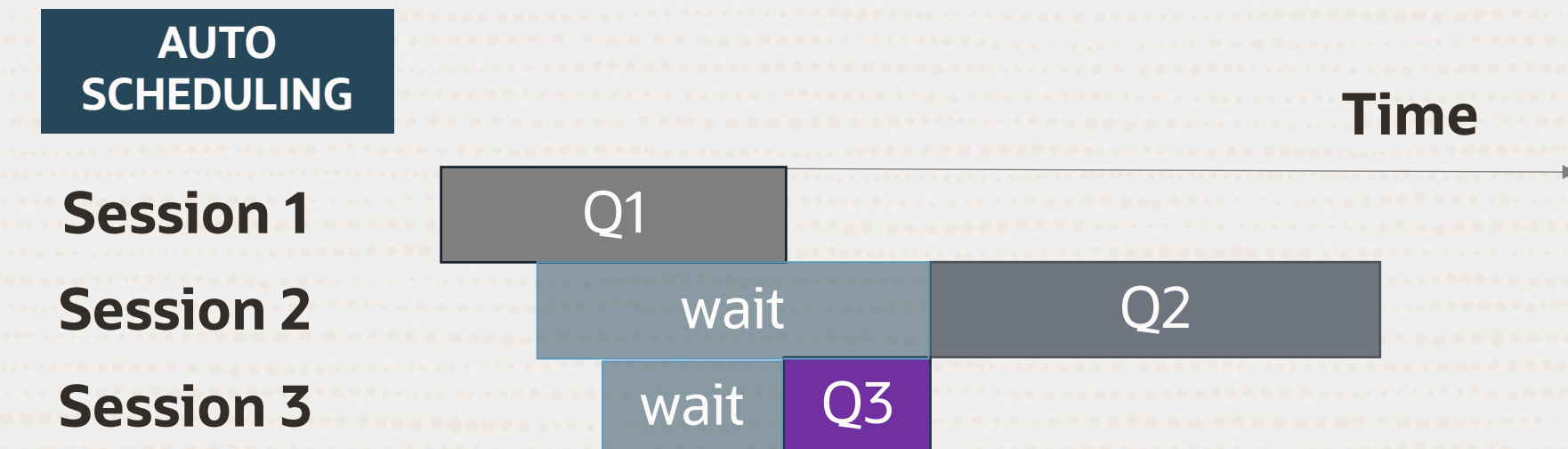
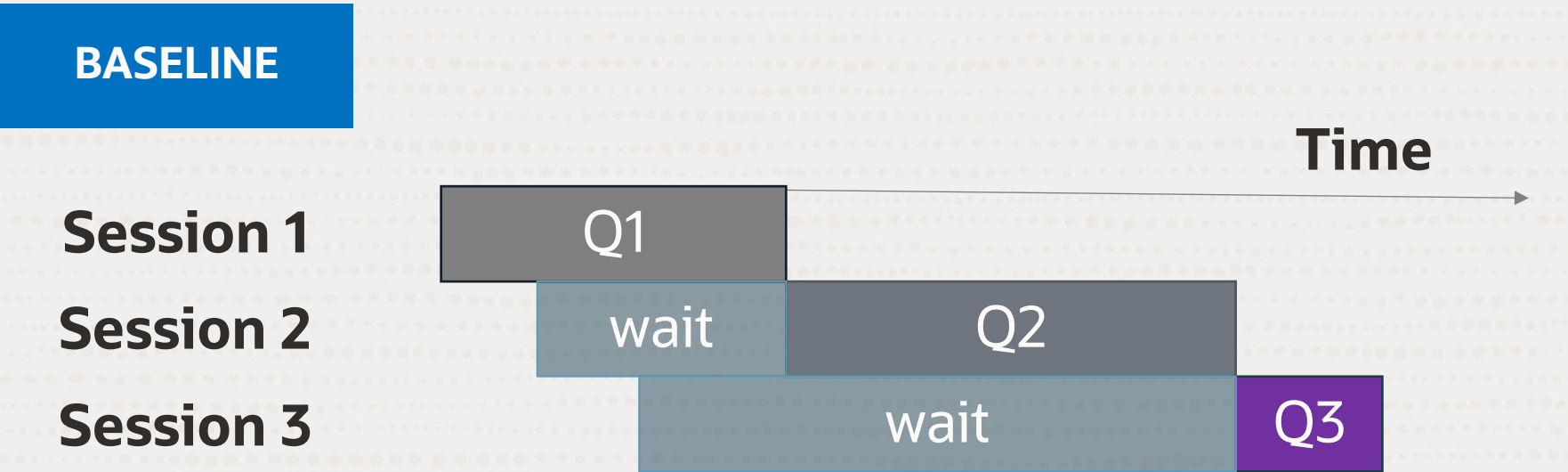


- Traditional caching techniques are not intelligent
- With Autopilot, system gets better as more queries are run
- 24TB TPC-H, TPC-DS performance improved by 40%

Auto Scheduling

Reduces wait time for mixed (OLTP + OLAP) workloads

- Analytic queries runs longer than OLTP queries
- HeatWave predicts execution time of each query
- Short queries are prioritized over long running queries
- System reduces wait time for shorter queries without changing total execution time



Short queries in workload			
Query	Baseline (sec)	Auto-Scheduling (sec)	Elapse Time Improvement
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	
Geomean	20.89	11.43	45.29%

Long queries in workload			
Query	Baseline (sec)	Auto-Scheduling (sec)	Elapse Time Increase
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	
Geomean	37.34	39.72	6.36%

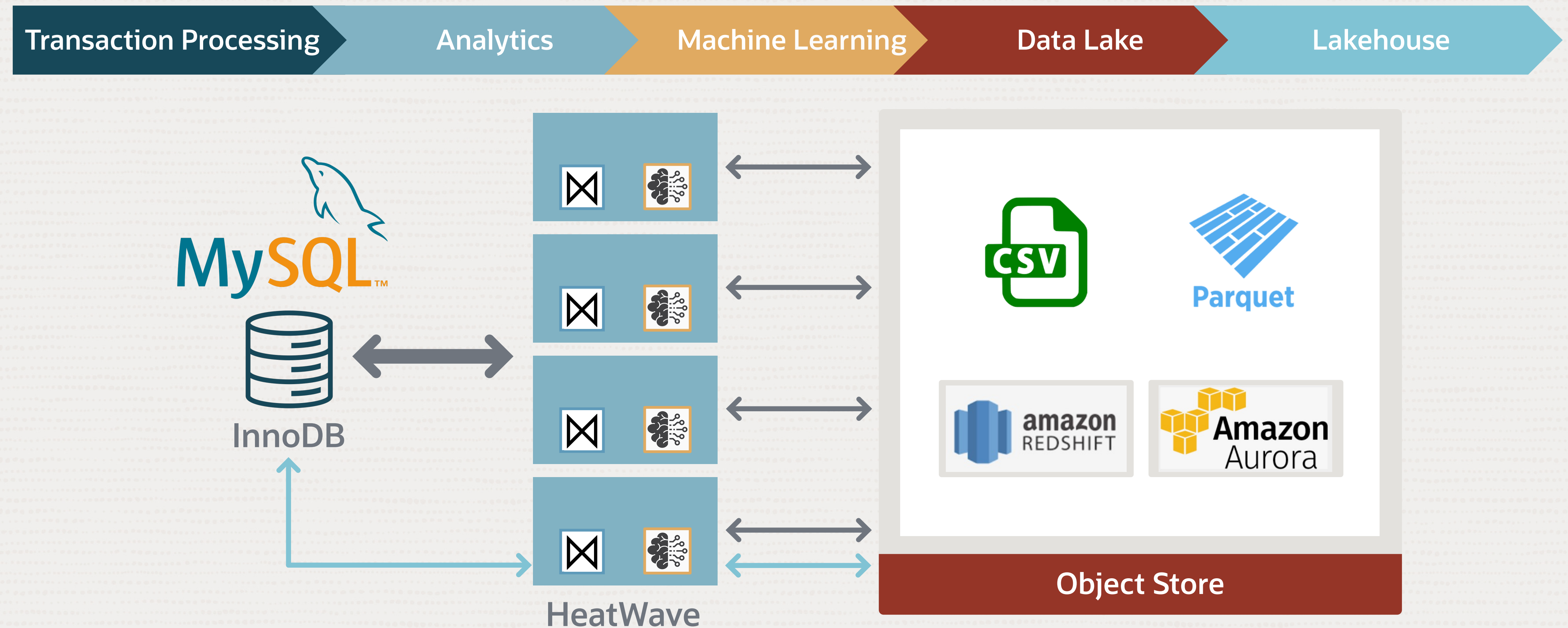


MySQL Autopilot for Lakehouse



MySQL HeatWave Lakehouse

Supports data on object store, multiple file formats, large data size



MySQL Autopilot for Lakehouse

1. Auto Schema Inference

- Sample raw files to infer column data types
- Generate DDLs to create tables

2. Adaptive Data Flow

- System adapts to the performance of object store
- Improves system performance and reliability

3. Adaptive Sampling

- Adaptively samples a fraction of files to collect stats
- Use collected stats for various Autopilot features

4. Auto Provisioning

- Adaptively sample raw files and collect statistics
- Estimate memory footprint of the data to be loaded

5. Auto Load

- Predict load time
- Load script generation

6. Auto Query Plan Improvement

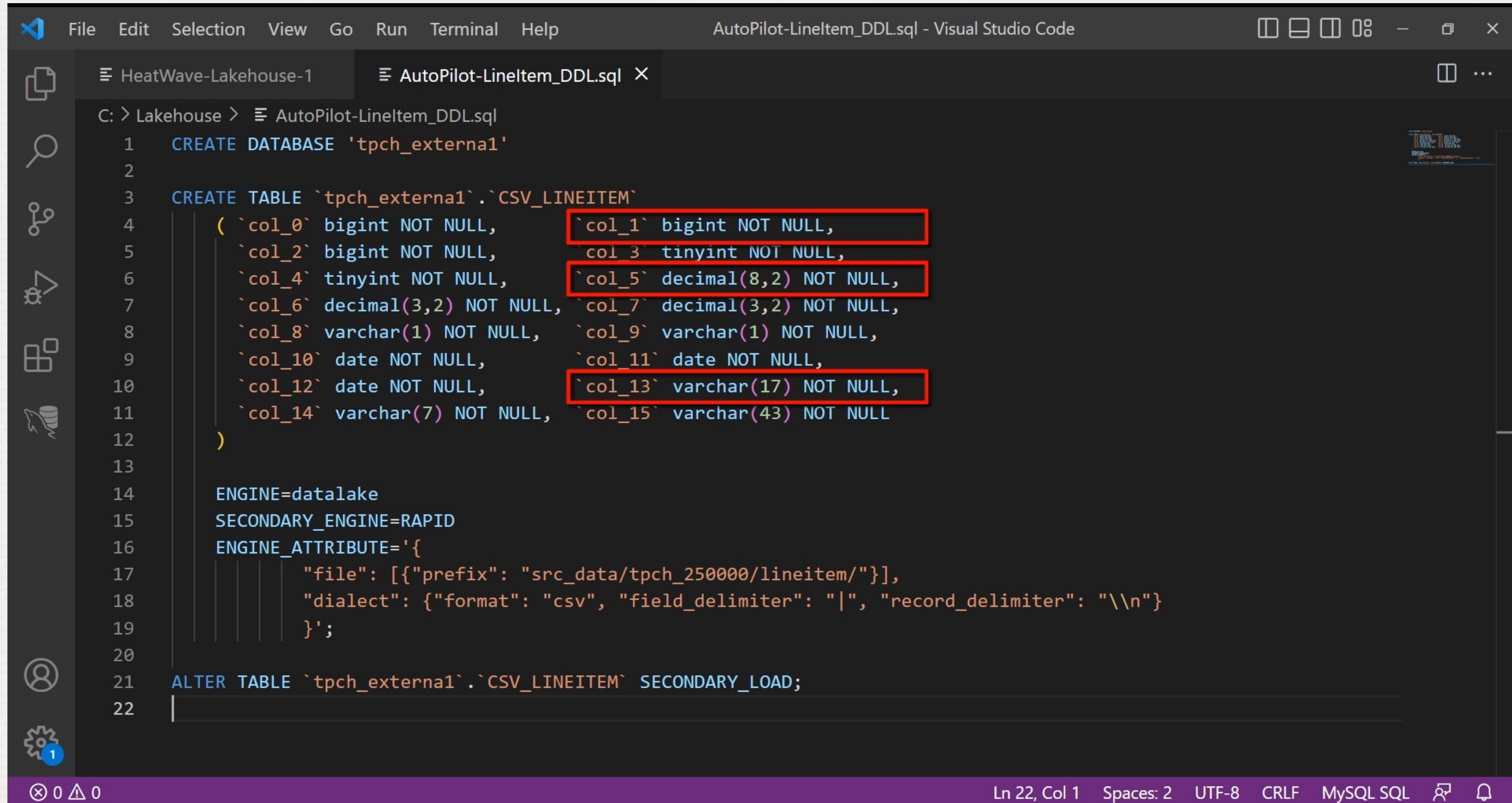
- Continuously collect statistics while running queries
- Enhance future execution plans

Auto Schema Inference

Infer table definitions for raw data stored in object storage

- Infer column data types in parallel
- Adaptive data sampling for stats and type inference & other pre-load estimations
- Integrated with Auto Load
 - Generate a complete load script based on the inferred schema
 - Predict memory footprint and load time

Auto Schema Inference



```
File Edit Selection View Go Run Terminal Help AutoPilot-LineItem_DDL.sql - Visual Studio Code

HeatWave-Lakehouse-1 AutoPilot-LineItem_DDL.sql X

C: \> Lakehouse \> AutoPilot-LineItem_DDL.sql

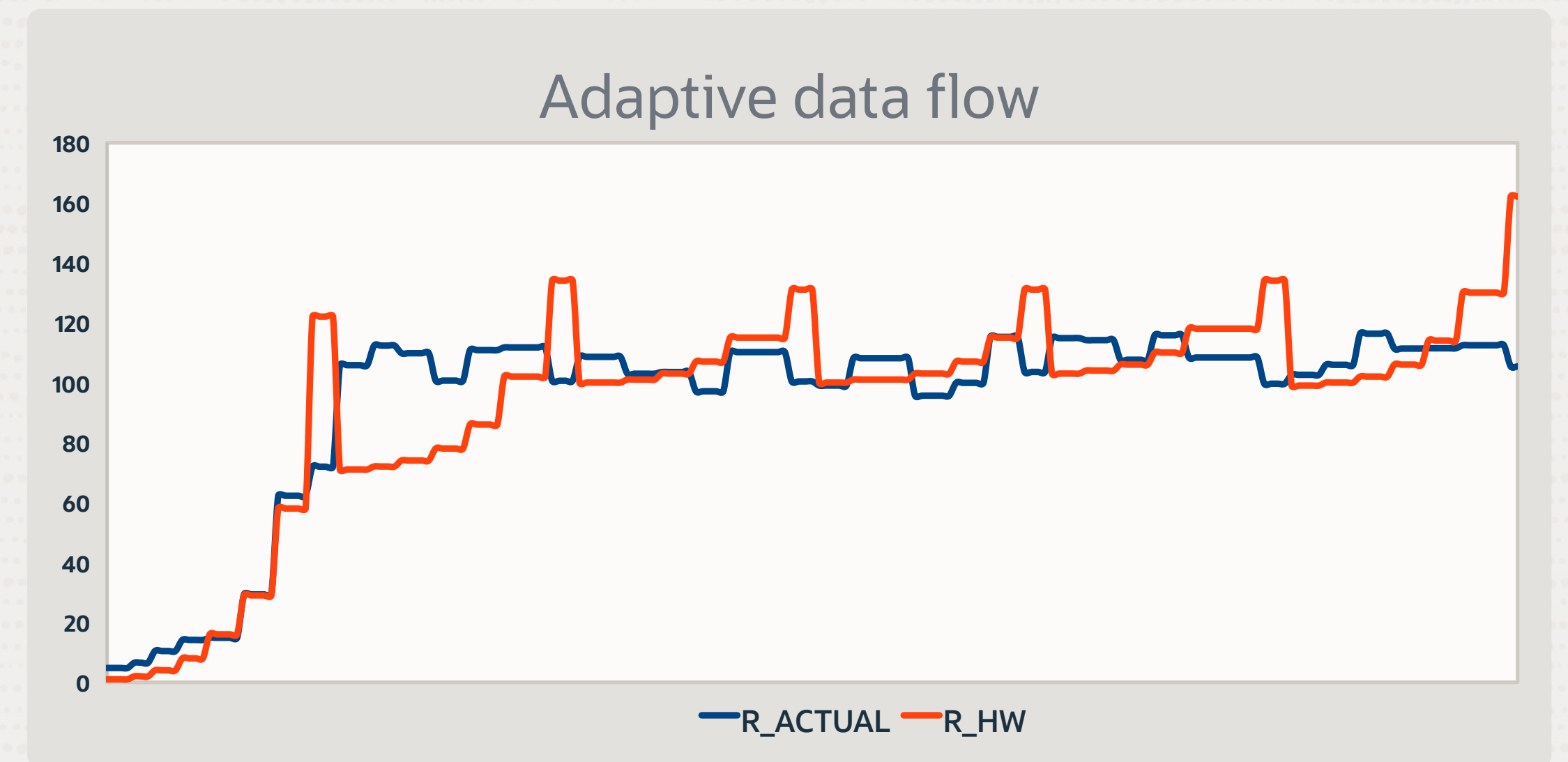
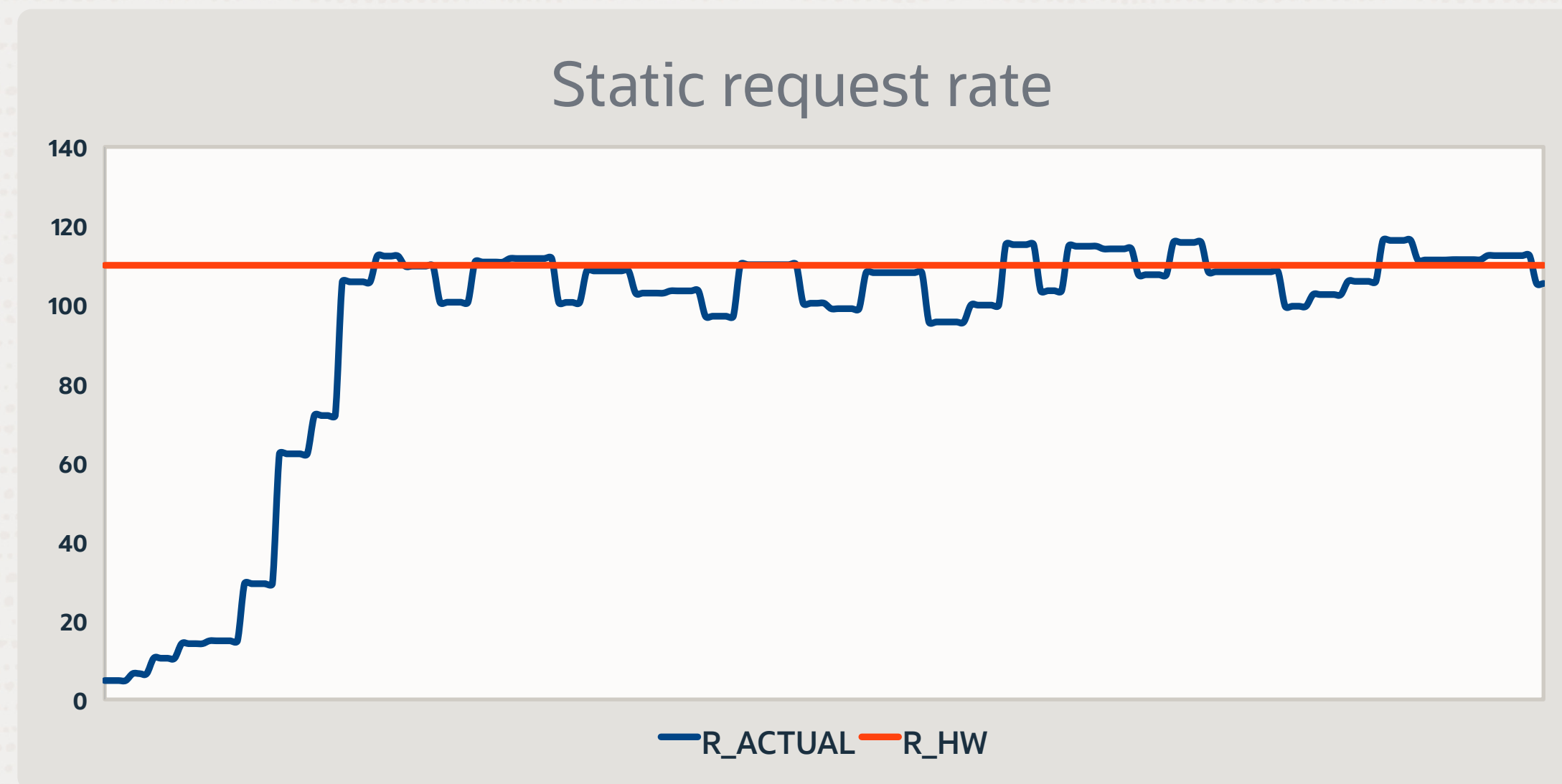
1 CREATE DATABASE 'tpch_externa1'
2
3 CREATE TABLE `tpch_externa1`.`CSV_LINEITEM`
4     ( `col_0` bigint NOT NULL, `col_1` bigint NOT NULL,
5       `col_2` bigint NOT NULL, `col_3` tinyint NOT NULL,
6       `col_4` tinyint NOT NULL, `col_5` decimal(8,2) NOT NULL,
7       `col_6` decimal(3,2) NOT NULL, `col_7` decimal(3,2) NOT NULL,
8       `col_8` varchar(1) NOT NULL, `col_9` varchar(1) NOT NULL,
9       `col_10` date NOT NULL, `col_11` date NOT NULL,
10      `col_12` date NOT NULL, `col_13` varchar(17) NOT NULL,
11      `col_14` varchar(7) NOT NULL, `col_15` varchar(43) NOT NULL
12    )
13
14    ENGINE=datalake
15    SECONDARY_ENGINE=RAPID
16    ENGINE_ATTRIBUTE='{
17      "file": [{"prefix": "src_data/tpch_250000/lineitem/"}],
18      "dialect": {"format": "csv", "field_delimiter": "|", "record_delimiter": "\\n\\n"}
19    }';
20
21 ALTER TABLE `tpch_externa1`.`CSV_LINEITEM` SECONDARY_LOAD;
22
```

Ln 22, Col 1 Spaces: 2 UTF-8 CRLF MySQL SQL

Adaptive Data Flow

HeatWave learns and adapts to object store performance

- Better performance
 - data loading from object storage is optimized for best available bandwidth
- Improved availability
 - Service errors are reduced

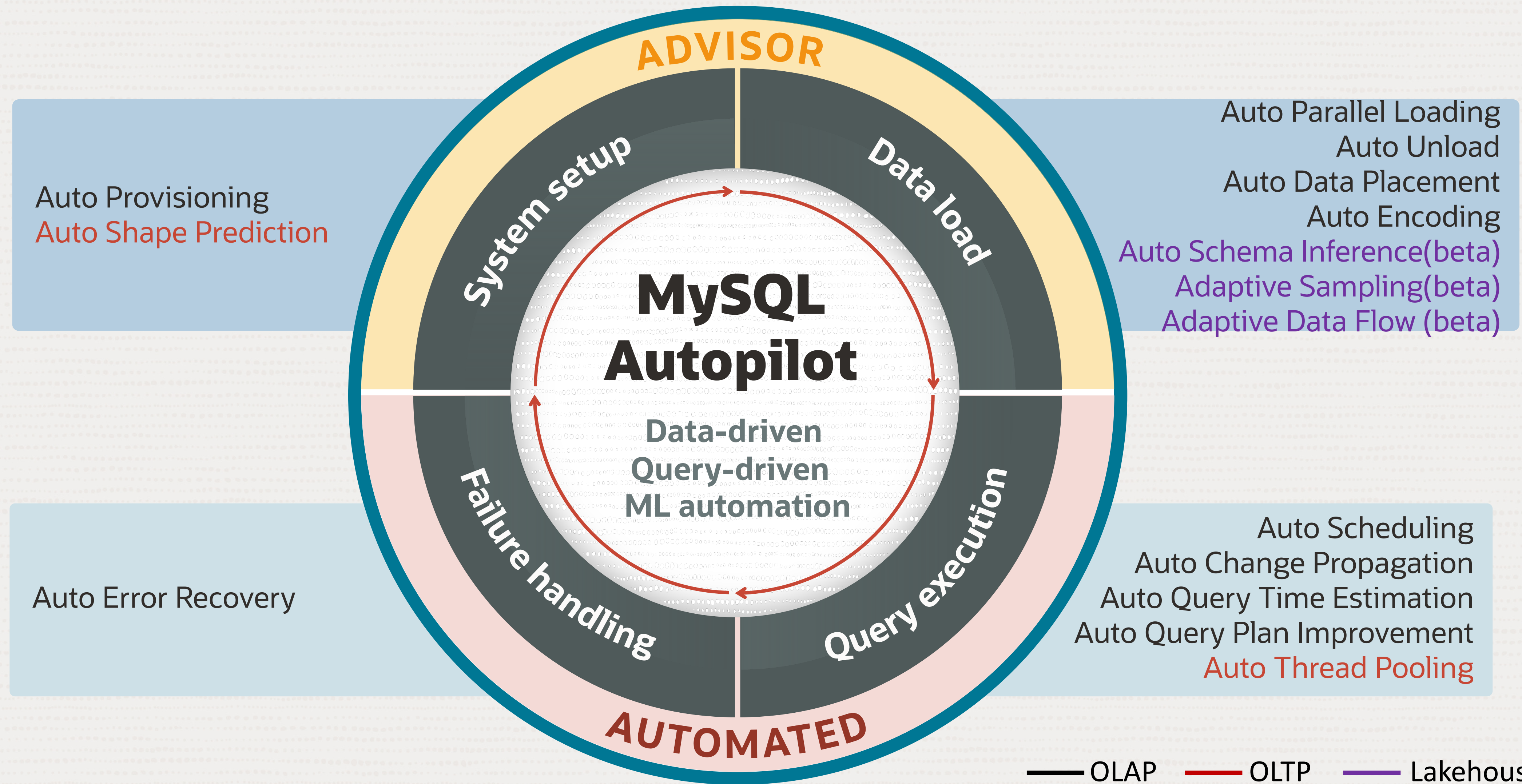


Summary



MySQL Autopilot

ML based automation to improve system performance and scalability



ORACLE
DevLive

Level Up

Please rate this session.

Session ID – MS17

