Oracle TimesTen Scaleout Developing Apps for OLTP and IoT

Doug Hood
Cloud Product Manager
Oracle Development

Yeung Wong
General Manager
Pingan Technology Shenzhen Co Ltd

October 24, 2018

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

Most Widely Used Relational In-Memory Database

Deployed by Thousands of Companies

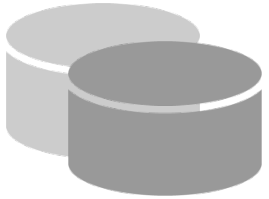
Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

Oracle TimesTen In-Memory Database

Relational Database

- Pure in-memory
- ACID compliant
- Standard SQL
- Entire database in DRAM

Persistent and Recoverable

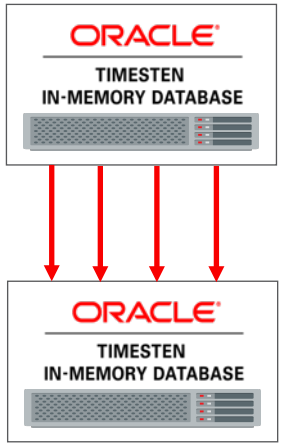


- Database and Transaction logs persisted on local disk or flash storage

Extremely Fast

- Microseconds response time
- Very high throughput

Highly Available



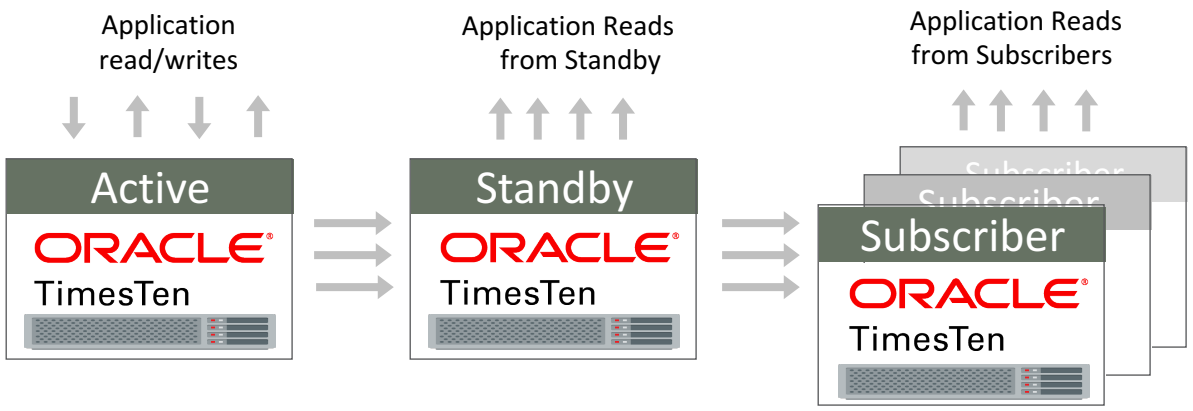
- Active-Standby and multi-master replication
- Very high performance parallel replication

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

Oracle TimesTen In-Memory Database

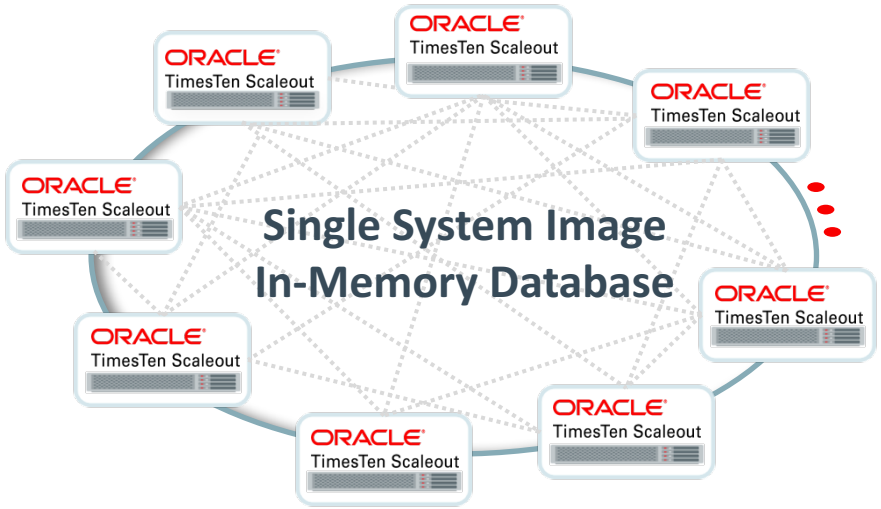
One product, two deployment modes

TimesTen Classic



- Replicated In-Memory Relational Database
- Highly Available
- **Extremely low latency** reads and writes
- Read scaling across multiple hosts

TimesTen Scaleout

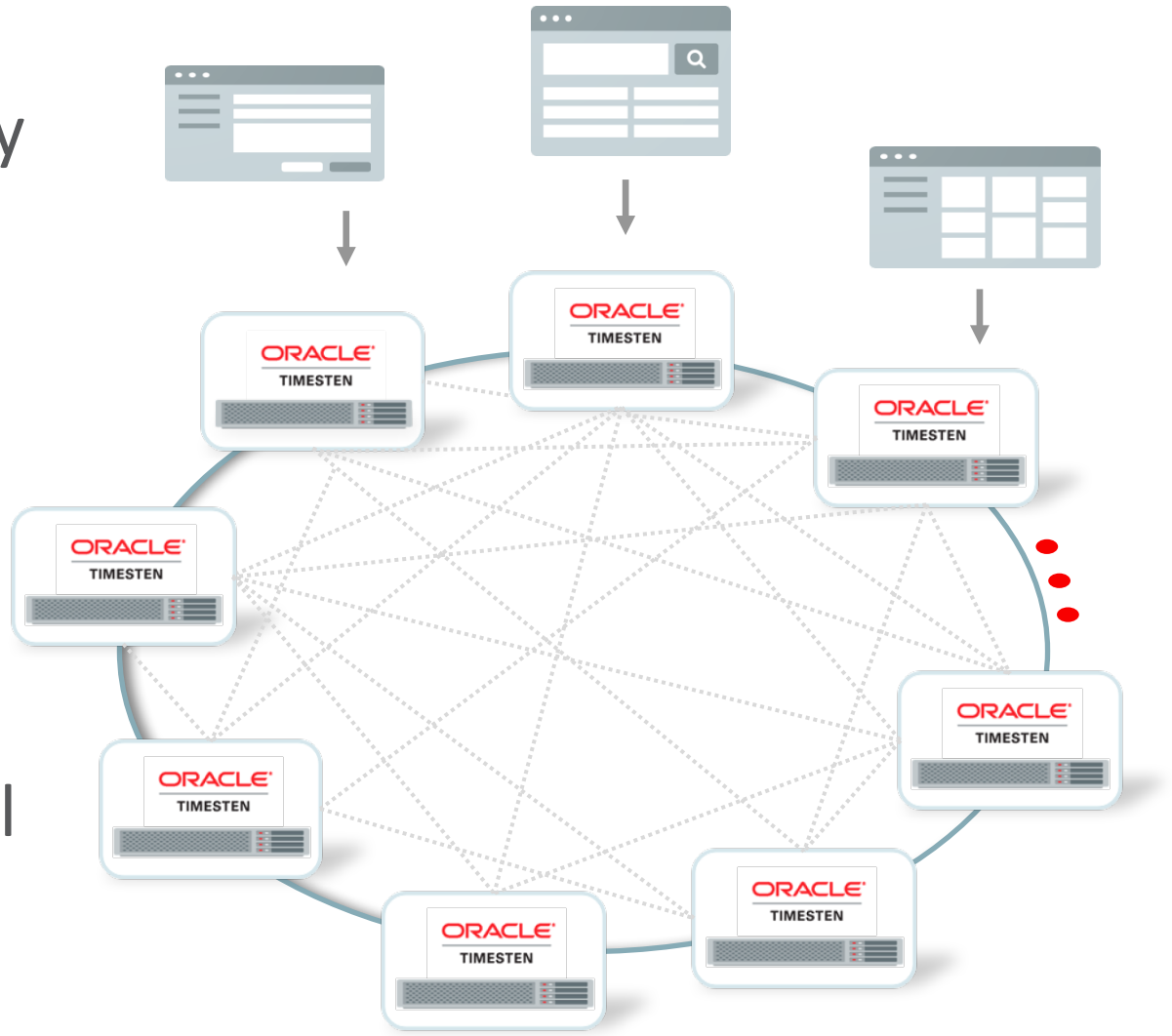


- Scale-Out In-Memory Relational Database
- Highly Available
- **Extremely high throughput** reads and writes
- **Scales both reads and writes**

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

Single Database Image

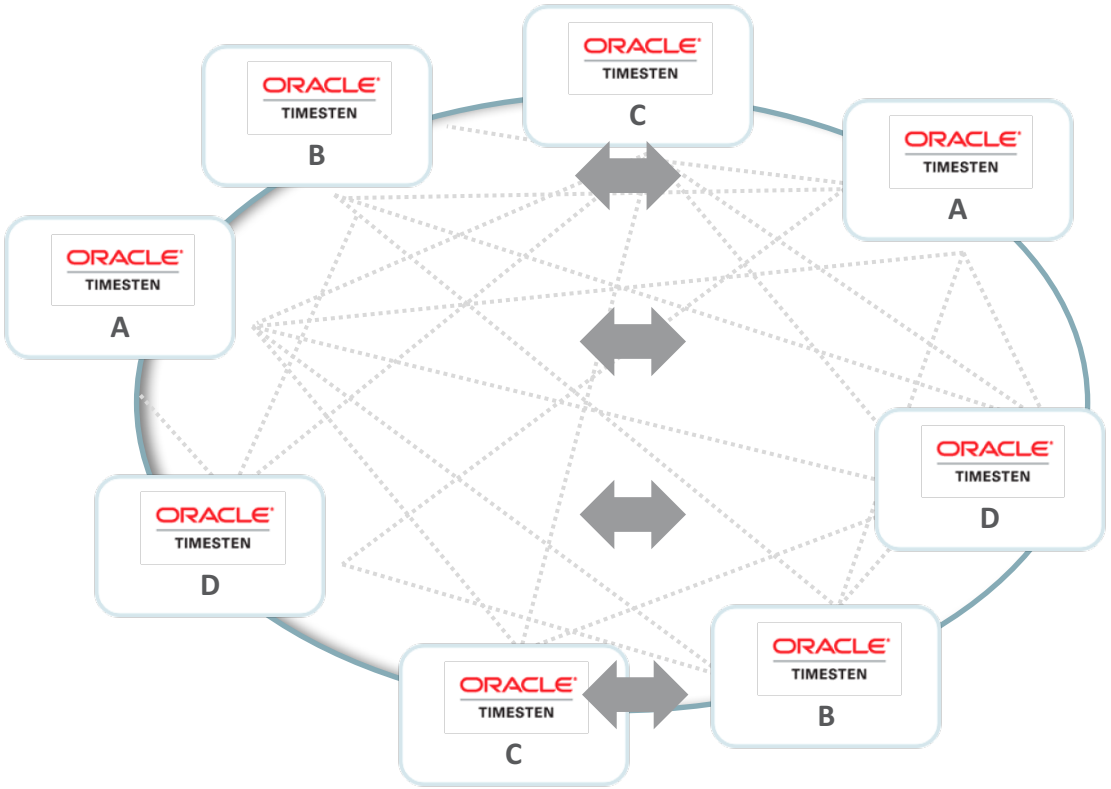
- Database size not limited by memory
- Table data distributed across all elements
  - All elements are equal
- Connect to **any** element and access **all** data
  - Distributed queries, joins & transactions
- No need to de-normalize data model



Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

High Availability and Maximum Throughput

K-Safety, All Active



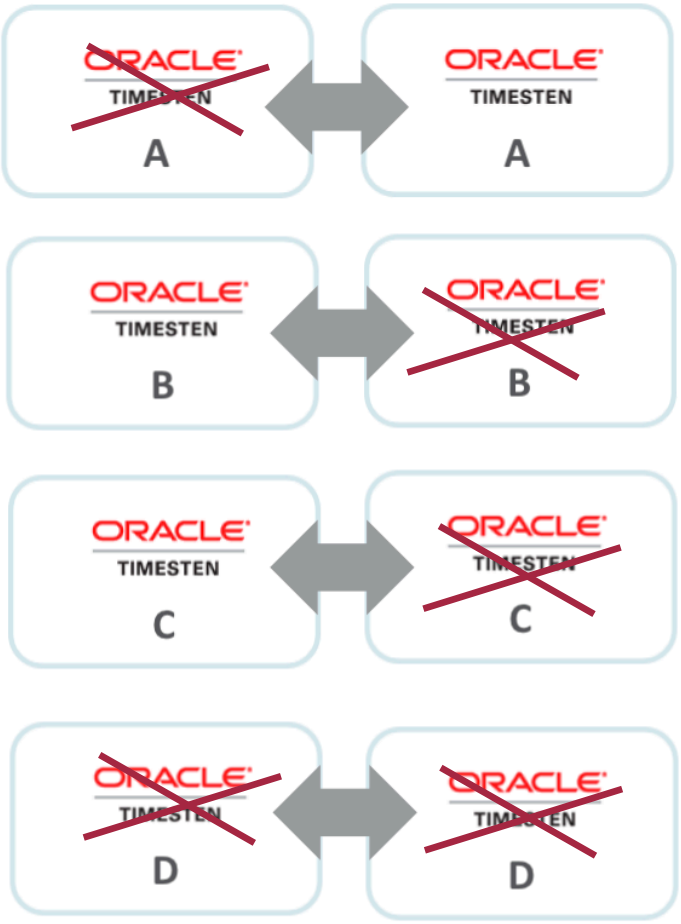
- Built-in HA via multiple copies of the data (K-safety)
  - Automatically kept in sync
- **All** replicas are **active** for **reads** and **writes**
  - Double the compute capacity
- Transactions can be initiated from and executed on any replica

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

Database Fault Tolerance – No Application Down Time

Provided one entire copy of the database is available

- If multiple elements fail, applications will continue provided there is one complete copy of the database
- Recovery after failure is automatic
- If an entire replica set is down, application can **explicitly** choose to accept partial results



Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

Using TimesTen Scaleout for OLTP

TimesTen Scaleout: The World's Fastest OLTP Database

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

What is the YCSB Workload?

- YCSB : **Y**ahoo **C**loud **S**erving **B**enchmark
  - Developed at Yahoo for Cloud Scale workloads
  - Widely used to compare scale-out databases, NoSQL databases, and (non-durable) in-memory data grids
- A series of workload types are defined:
  - Workload A: 50% reads, 50% Updates
  - Workload B: 95% reads, 5% Updates
  - Workload C: 100% reads
- The YCSB Client cannot be changed
  - DB Vendors implement the DB Client interface in Java
  - The version and exact configuration matters

Surveyed YCSB (Workload B) Results*

Product	Type	Nodes	Ops/Sec
cassandra	NoSQL DB	32	227 K
mongoDB	NoSQL DB	2	275 K
SCYLLA.	NoSQL DB	3	715 K
VOLTDB	Scale-Out RDBMS	6	1.6 M
AEROSPIKE	NoSQL DB	8	1.6 M

** There is no official repository of YCSB results*
These were the largest results we found online

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

What is the YCSB Workload?

- YCSB : **Y**ahoo **C**loud **S**erving **B**enchmark
  - Developed at Yahoo for Cloud Scale workloads
  - Widely used to compare scale-out databases, NoSQL databases, and (non-durable) in-memory data grids
- A series of workload types are defined:
  - Workload A: 50% reads, 50% Updates
  - Workload B: 95% reads, 5% Updates
  - Workload C: 100% reads
- The YCSB Client cannot be changed
  - DB Vendors implement the DB Client interface in Java
  - The version and exact configuration matters

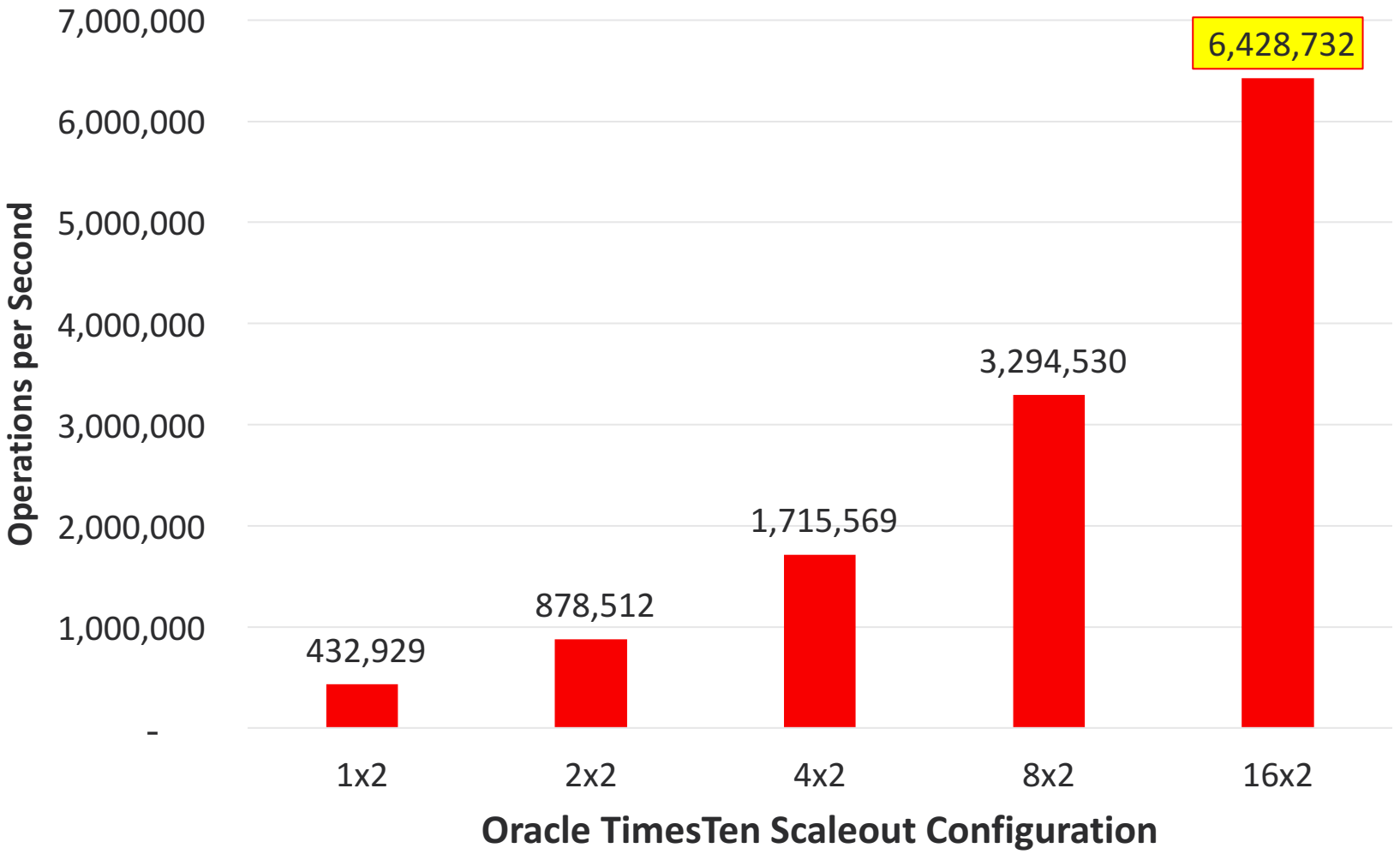
Surveyed YCSB (Workload B) Results*

Product	Type	Nodes	Ops/Sec
cassandra	NoSQL DB	32	227 K
mongoDB	NoSQL DB	2	275 K
SCYLLA.	NoSQL DB	3	715 K
VOLTDB	Scale-Out RDBMS	6	1.6 M
AEROSPIKE	NoSQL DB	8	1.6 M

** There is no official repository of YCSB results*
These were the largest results we found online

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

YCSB Workload A (50% Read 50% Update): **6.4 Million** Ops/Sec



YCSB version 0.15.0

- 1KB record
  (100-byte x 10 Fields)
- 100M records / Replica Set
- Uniform Distribution

TimesTen Scaleout

- 1 to 16 replica sets
- 2 synchronous replicas per replica set

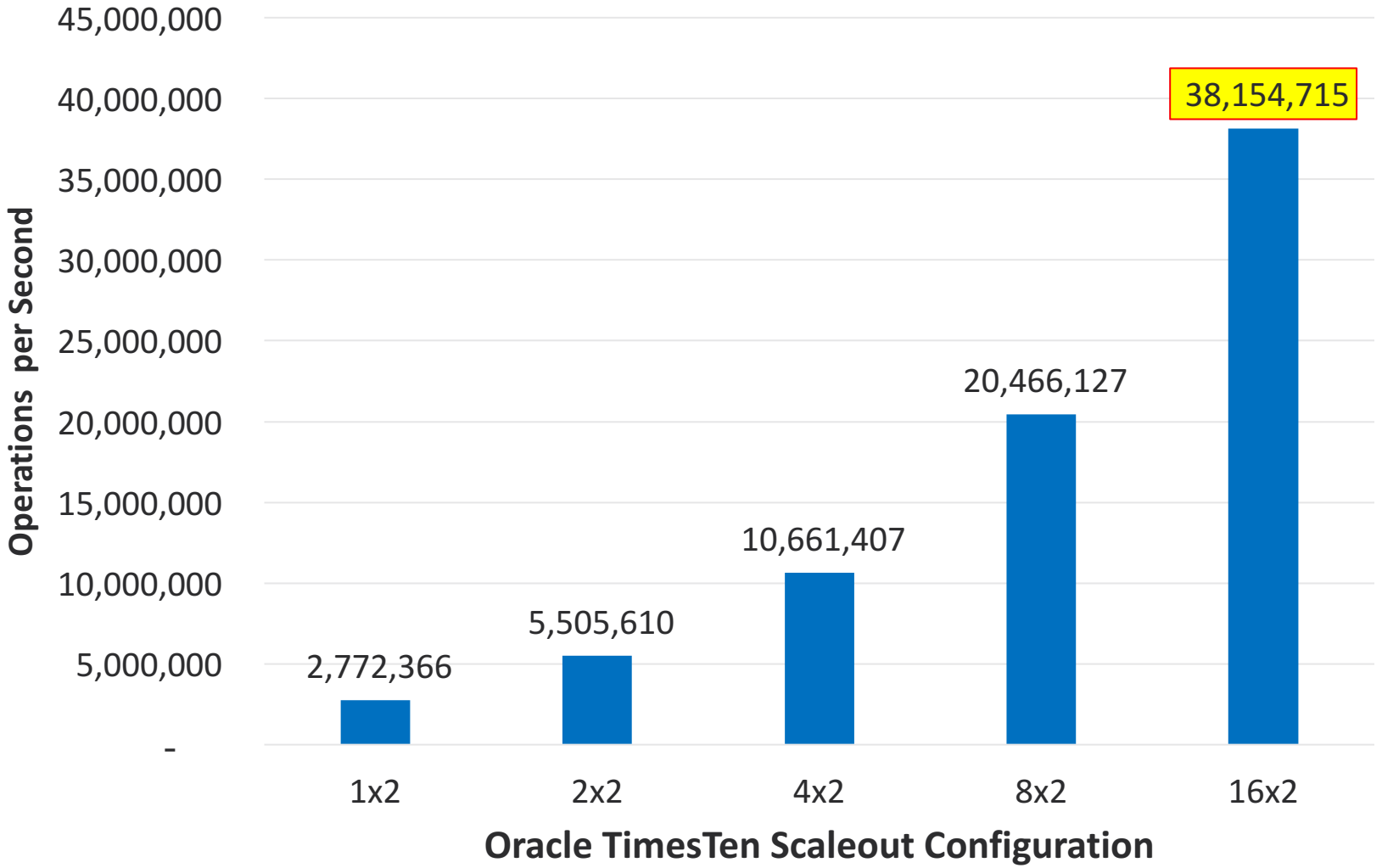
Oracle Cloud Infrastructure

- 32 * BM.DenseIO2.52

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

YCSB Workload B (95% Read 5% Update): **38 Million** Ops/Sec

Reminder: The best YCSB-B result found in our survey was **1.6 Million** Ops/Sec



YCSB version 0.15.0

- 1KB record (100-byte x 10 Fields)
- 100M records / Replica Set
- Uniform Distribution

TimesTen Scaleout

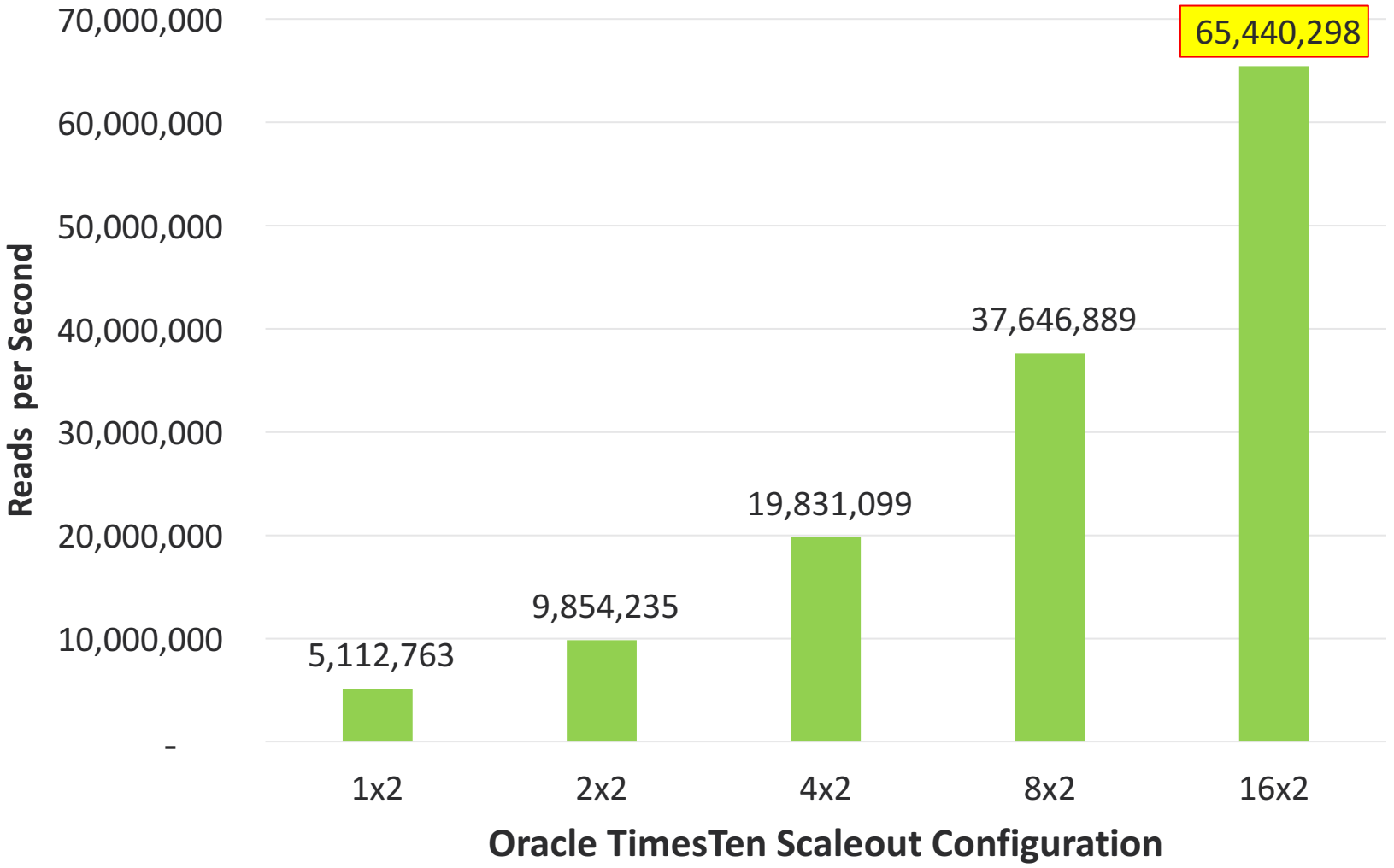
- 1 to 16 replica sets
- 2 synchronous replicas per replica set

Oracle Cloud Infrastructure

- 32 * BM.DenseIO2.52

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

YCSB Workload C (100% Read): **65 Million** Reads/Sec



YCSB version 0.15.0

- 1KB record
  (100-byte x 10 Fields)
- 100M records / Replica Set
- Uniform Distribution

TimesTen Scaleout

- 1 to 16 replica sets
- 2 synchronous replicas per replica set

Oracle Cloud Infrastructure

- 32 * BM.DenseIO2.52

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

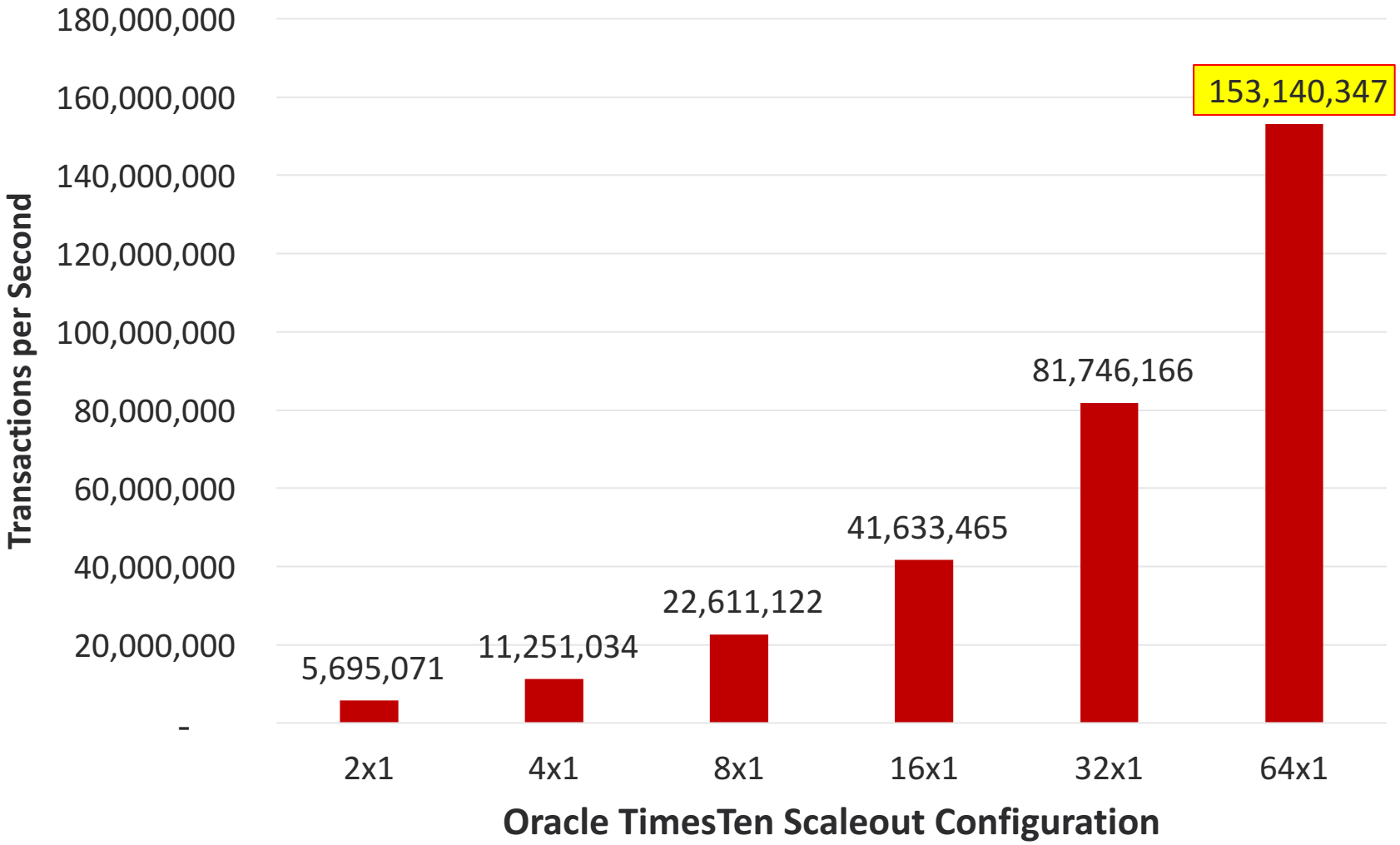
What is the TPTBM Workload?

ORACLE TimesTen

- TPTBM : **T**elecom **P**rovider **T**hroughput **B**ench**M**ark
  - A benchmark originally developed by the TimesTen team
  - Represents common operations on a Telecom Subscriber database
  - Uses *standard* SQL and *standard* database APIs
  - Shipped with Oracle TimesTen as C and Java source code for the past 15 years
  - Quickly demonstrates the performance of user's hardware
- Common workload mixes:
  - 80% Reads, 20% Updates
  - 100% Reads
- The version and exact configuration matters

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

TPTBM 80% Read 20% Update: **153 Million** Transactions/Sec



TPTBM Configuration

- 128-byte record
- 100M records / Replica Set
- Uniform Distribution

TimesTen Scaleout

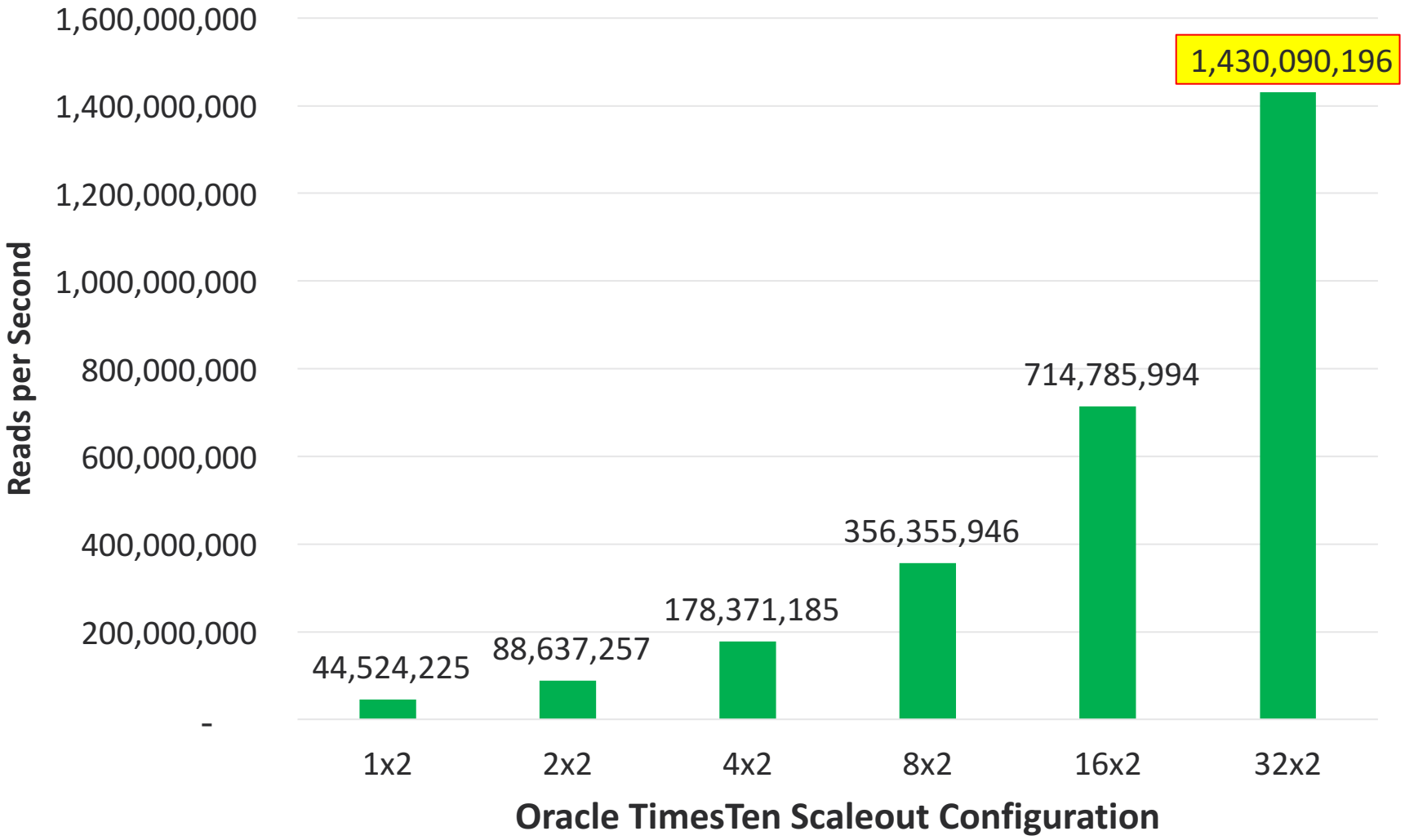
- 1 to 64 replica sets
- 1 replica per replica set

Oracle Cloud Infrastructure

- 32 * BM.DenseIO2.52
- Two TimesTen instances per compute node

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

TPTBM 100% Read: **1.4 Billion Reads** Per Second!!



TPTBM Configuration

- 128-byte record
- 100M records / Replica Set
- Uniform Distribution

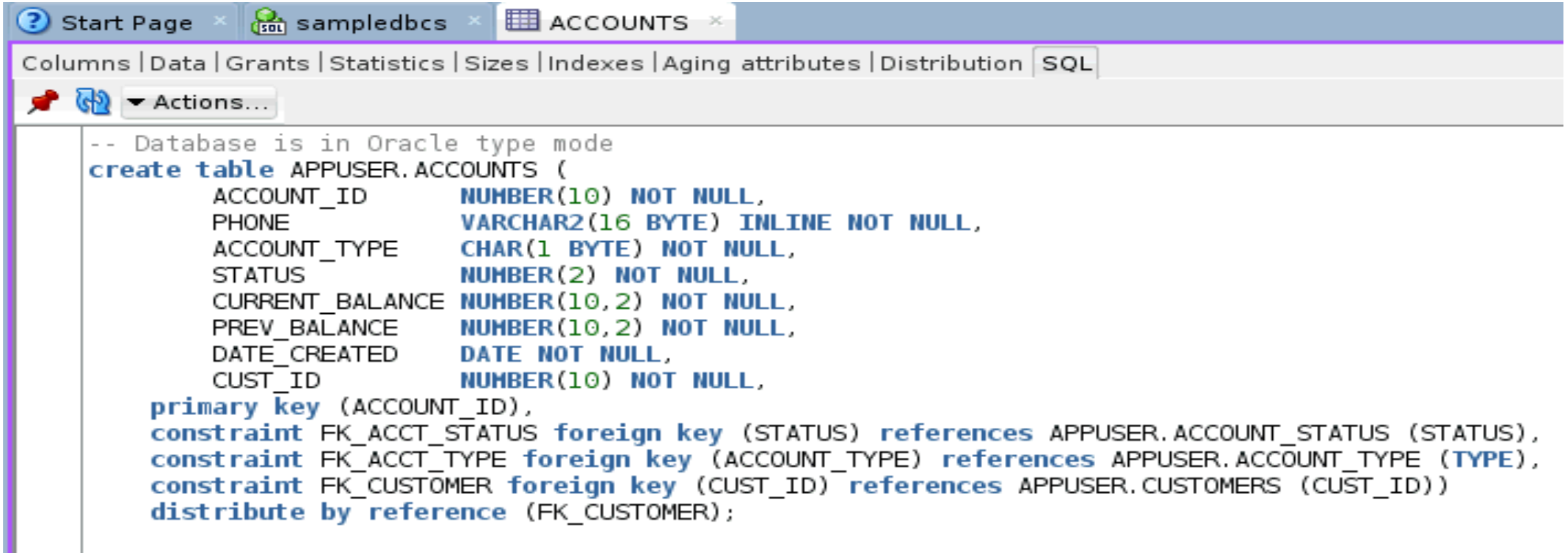
TimesTen Scaleout

- 1 to 32 replica sets
- 2 synchronous replicas per replica set

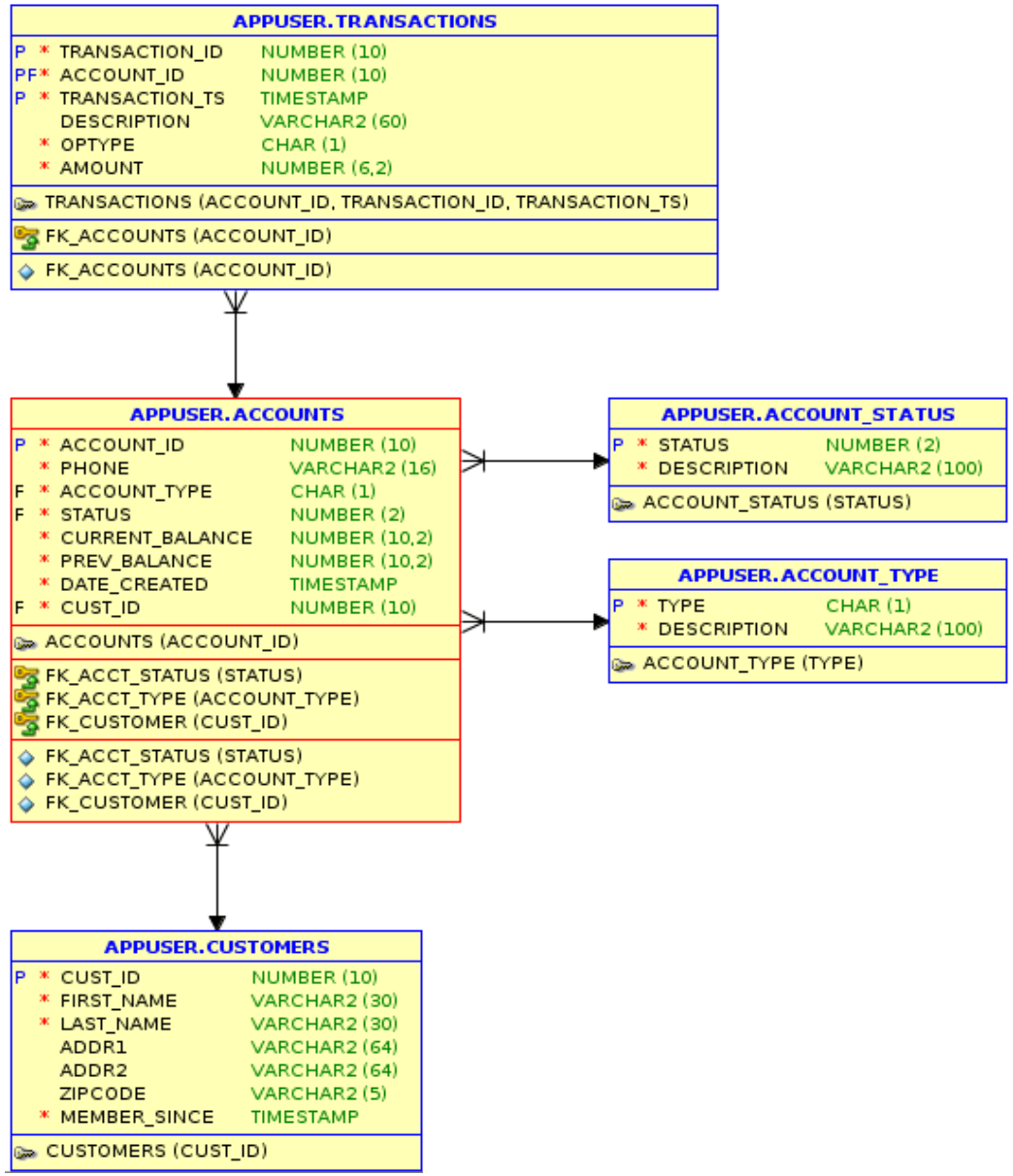
Oracle Cloud Infrastructure

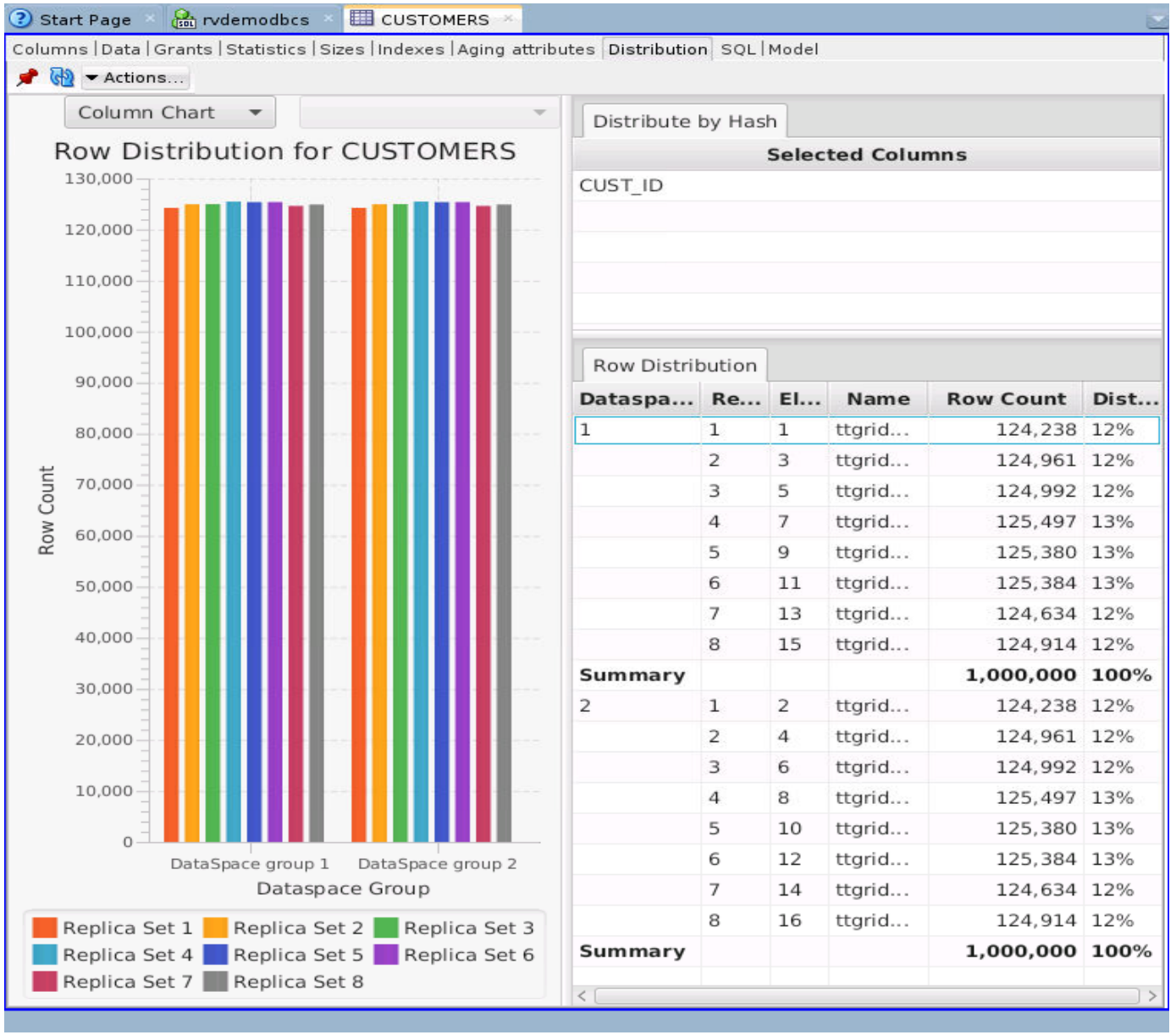
- 32 * BM.DenseIO2.52
- Two TimesTen instances per compute node

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |



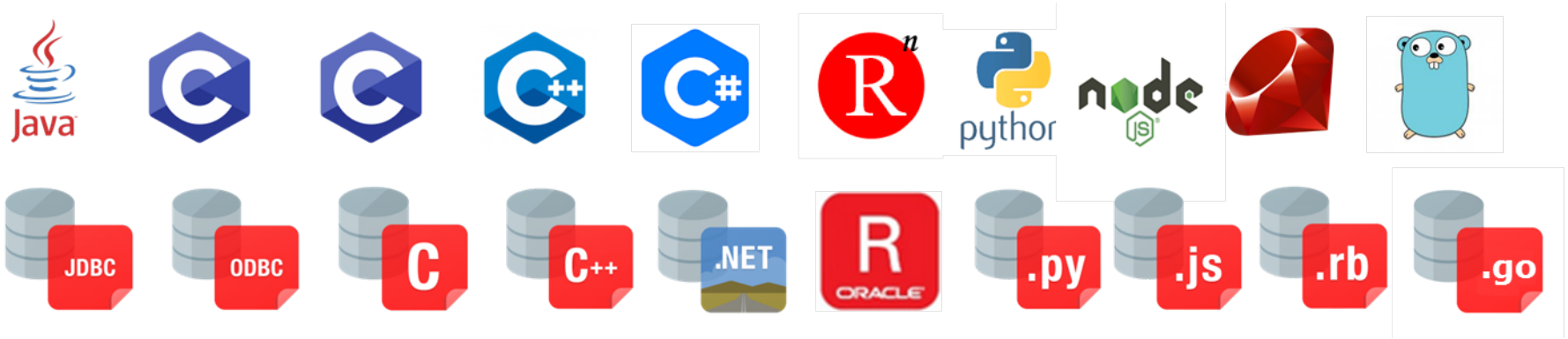
Copyright © 2018, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted





Copyright © 2018, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

TimesTen Scaleout SQL APIs

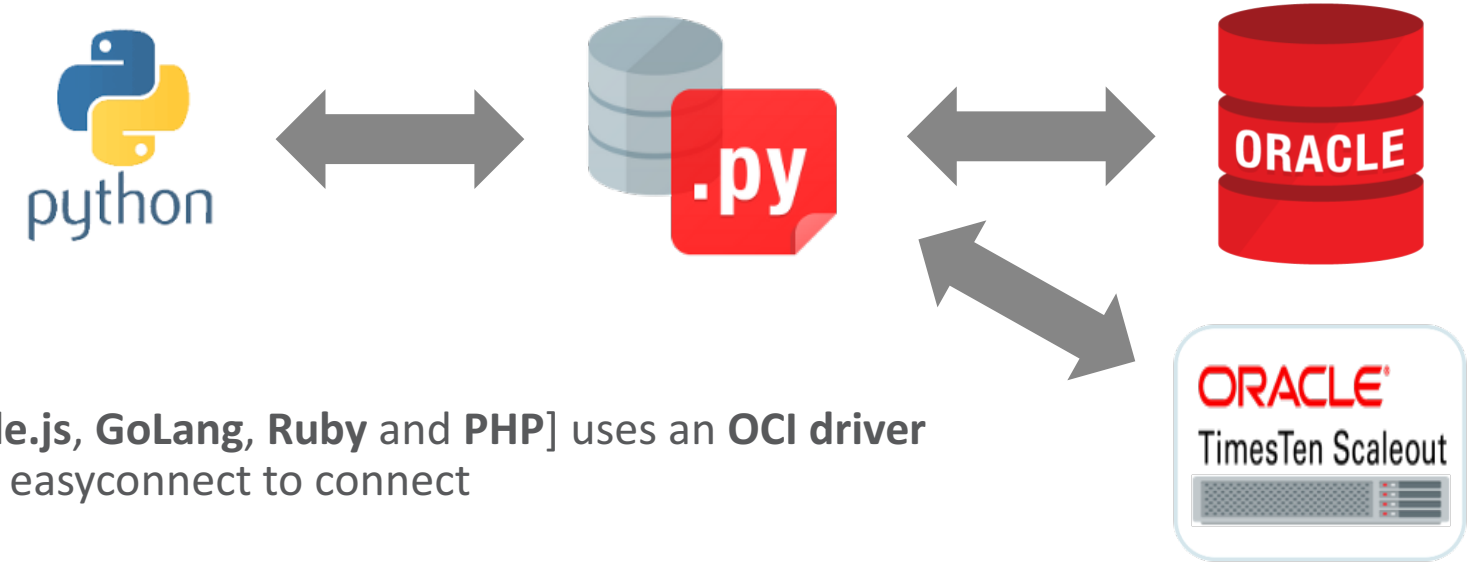


API	Comment
JDBC	The same (JDBC 4.3)
ODBC	The same (ODBC 3.5.2)
OCI	The same (OCI 11.2.0.4.+)
R-Oracle	The same (OCI 11.2.0.4.+)
ODP.Net	The same (OCI 11.2.0.4.+)
Pro*C	The same (OCI 11.2.0.4.+)
Python	The same (cx_Oracle, ODPI-C)
Ruby	The same (Ruby-ODPI, ODPI-C)
GoLang	The same (go-goracle, ODPI-C)

ORACLE

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

Using Oracle cx_Python with TimesTen Scaleout



Python [and **Node.js**, **GoLang**, **Ruby** and **PHP**] uses an **OCI driver**
Use tnsnames or easyconnect to connect

tnsnames.ora :

sampledb_1811 =(DESCRIPTION=(CONNECT_DATA = (SERVICE_NAME = **sampledb_1811**)(SERVER = **timesten_direct**)))
sampledbCS_1811 =(DESCRIPTION=(CONNECT_DATA = (SERVICE_NAME = **sampledbCS_1811**)(SERVER = **timesten_client**)))

TimesTen ODBC DSN

Client/Server or Direct Linked

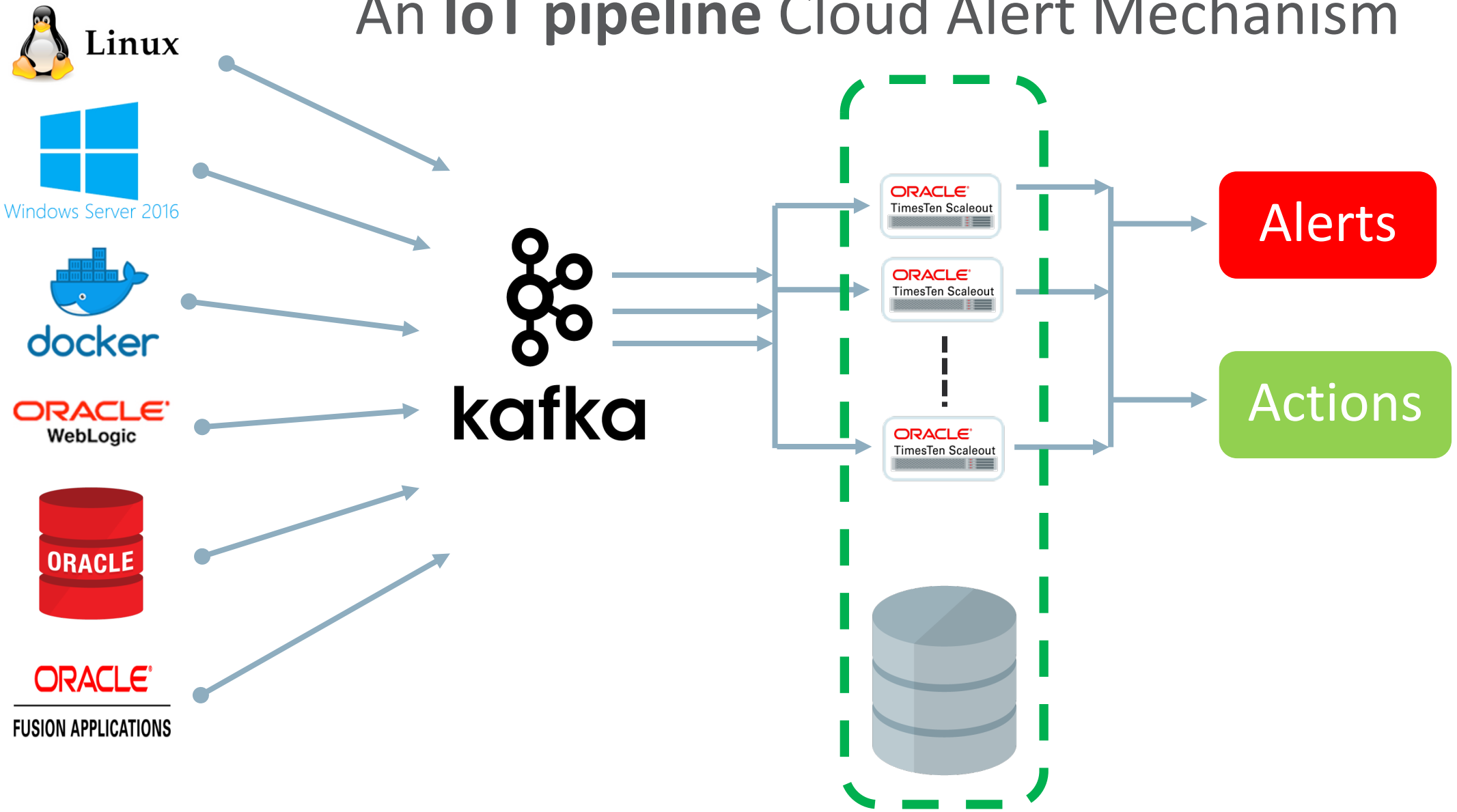
Copyright © 2018, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

Using TimesTen Scaleout for IoT

TimesTen Scaleout: The World's Fastest OLTP Database

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

An **IoT pipeline** Cloud Alert Mechanism



Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

Cloud Alerts

Each transaction included:
- Enqueues
  - insert + commit
- **Concurrent analytics rules**
- Dequeues 5-20 min later
  - Deletes + commit

Hundreds of rules. eg
map state over time to bitmaps:
000000 = Ignore
000001 = Interesting
000010 = Ignore
…
111111 = Ignore

Window from **100M** to **1B** records per node

Example Rule

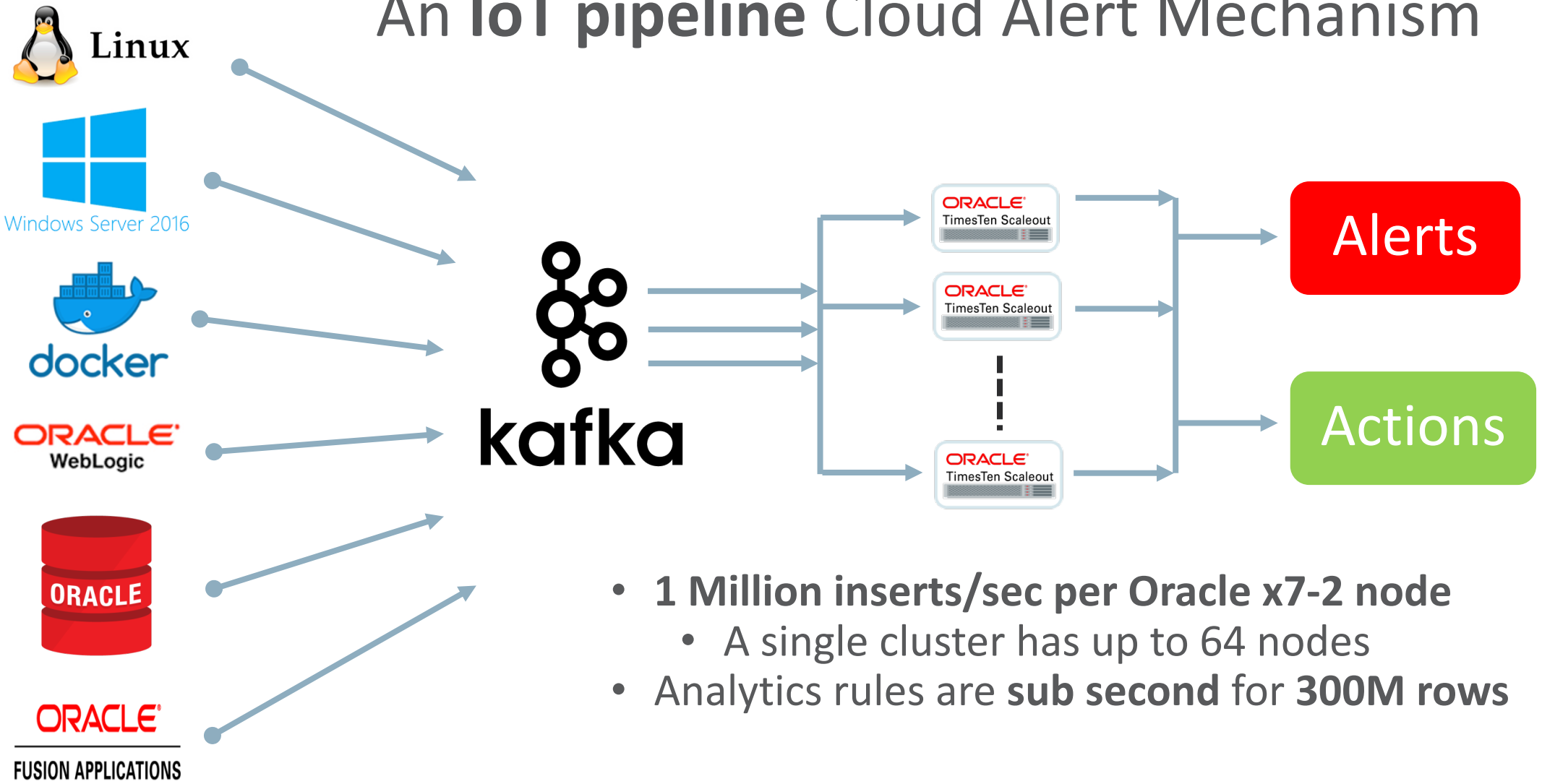
```
insert into login_results
select * from
(
select id,
      ruleid,
      evalcycle,
      target,
      userid,
      ts,
      nodeid,
      status,
      sum(loginstatus)
       over (partition by id, ruleid, evalcycle, target, userid
           order by ts
           rows between 5 preceding and 0 following) alias1
from login_table
group by id, ruleid, evalcycle, target, userid, ts, nodeid, loginstatus
order by ts asc
) where alias1 = 1
and loginstatus = 1;
```

```
create table login_table (
       ruleid      varchar2(256),
       tenantid    number,
       target      varchar2(32),
       evalcycle   number,
       nodeid      number,
       userid      varchar2(16),
       loginstatus number,
       ts          timestamp(9),
       ttl         timestamp);
```

```
create table login_results (
       tenantid        number,
       ruleid          varchar2(256),
       evalcycle       number,
       target          varchar2(32),
       userid          varchar2(16),
       ts              timestamp(9),
       nodeid          number,
       loginstatus     number,
       sum_loginStatus number,
       processed       number default 0);
```

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

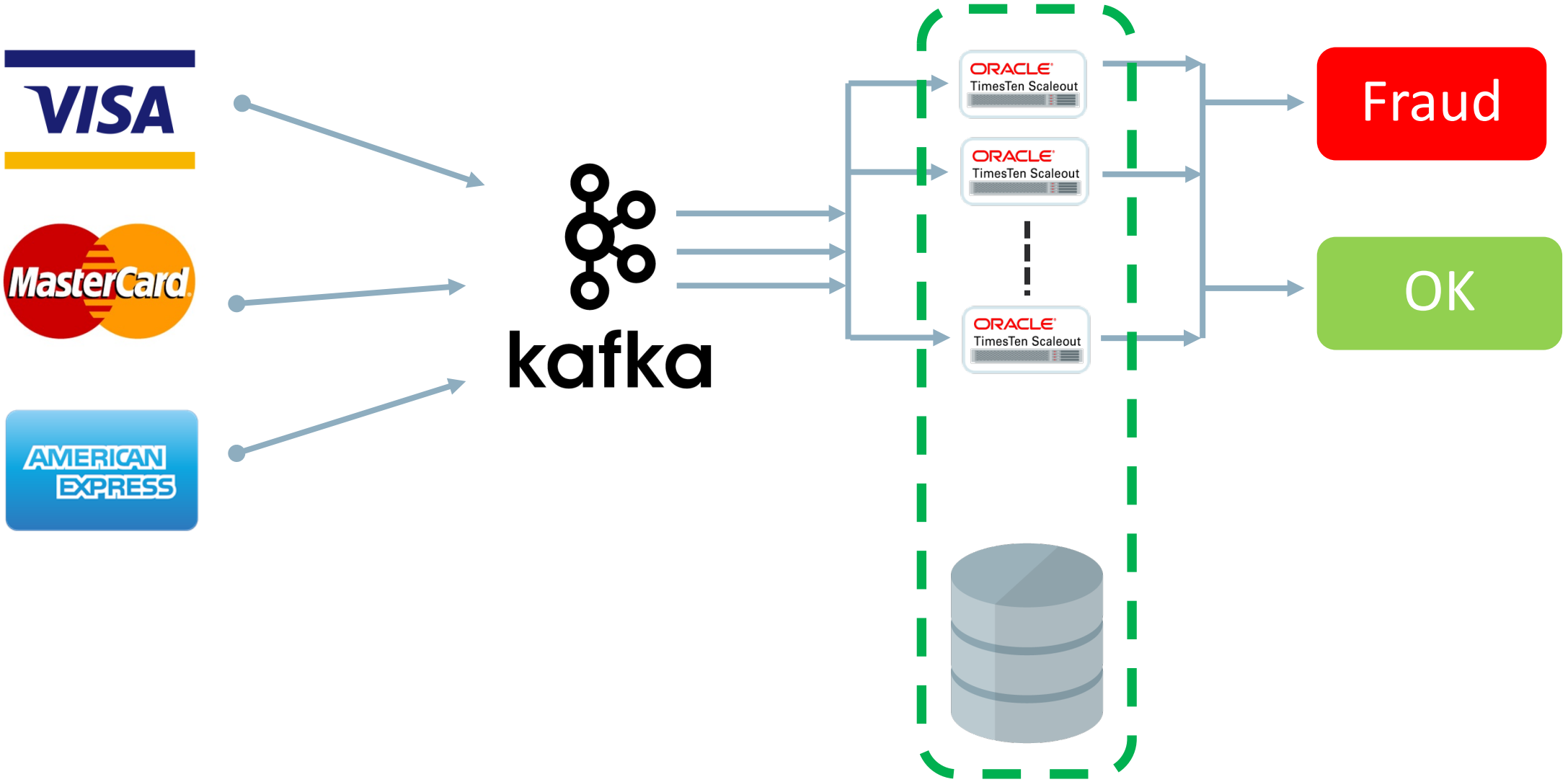
An **IoT pipeline** Cloud Alert Mechanism



- **1 Million inserts/sec per Oracle x7-2 node**
  - A single cluster has up to 64 nodes
- Analytics rules are **sub second** for **300M rows**

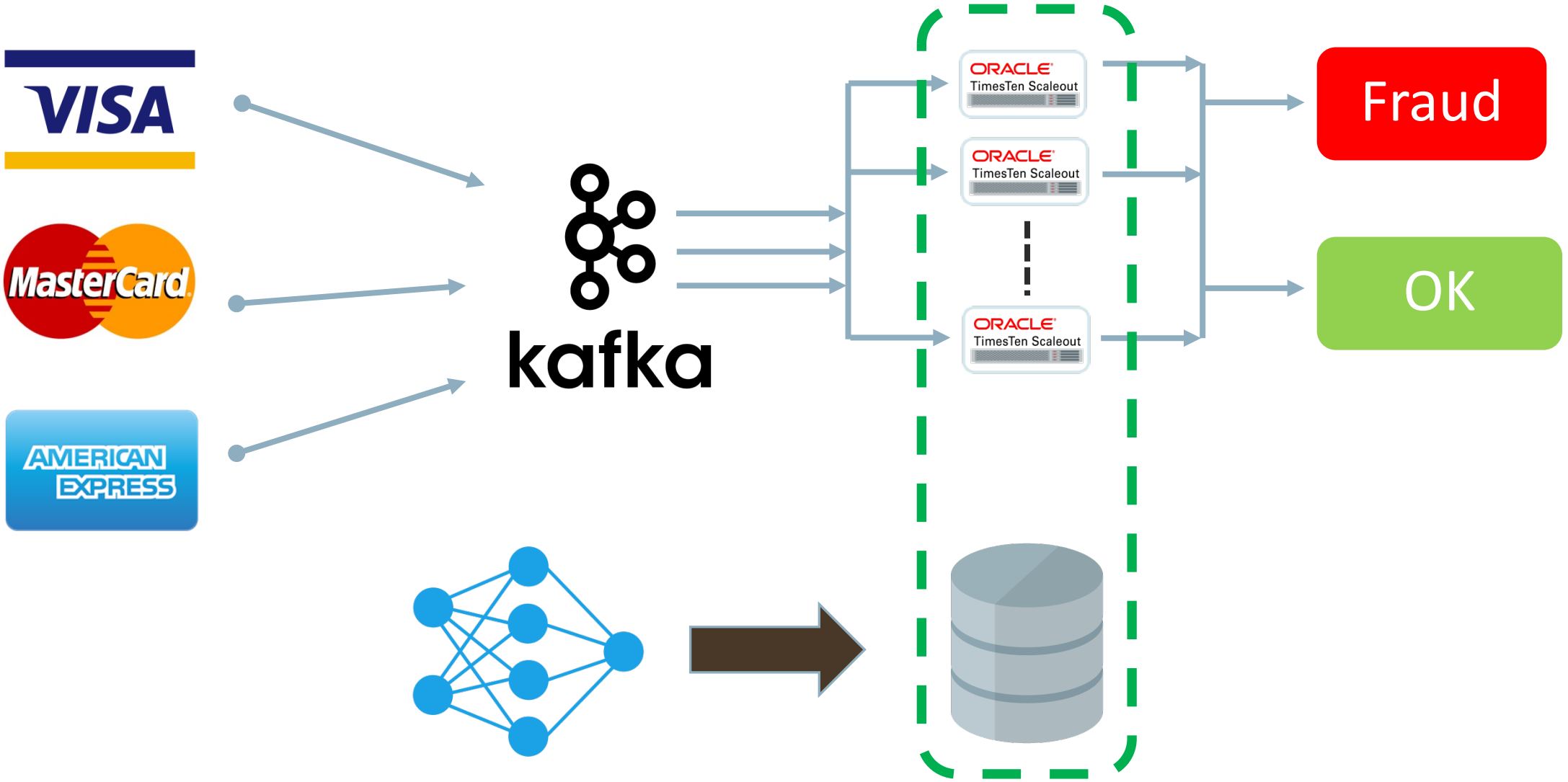
Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

An **IoT pipeline** for Real Time Credit Card Fraud Detection



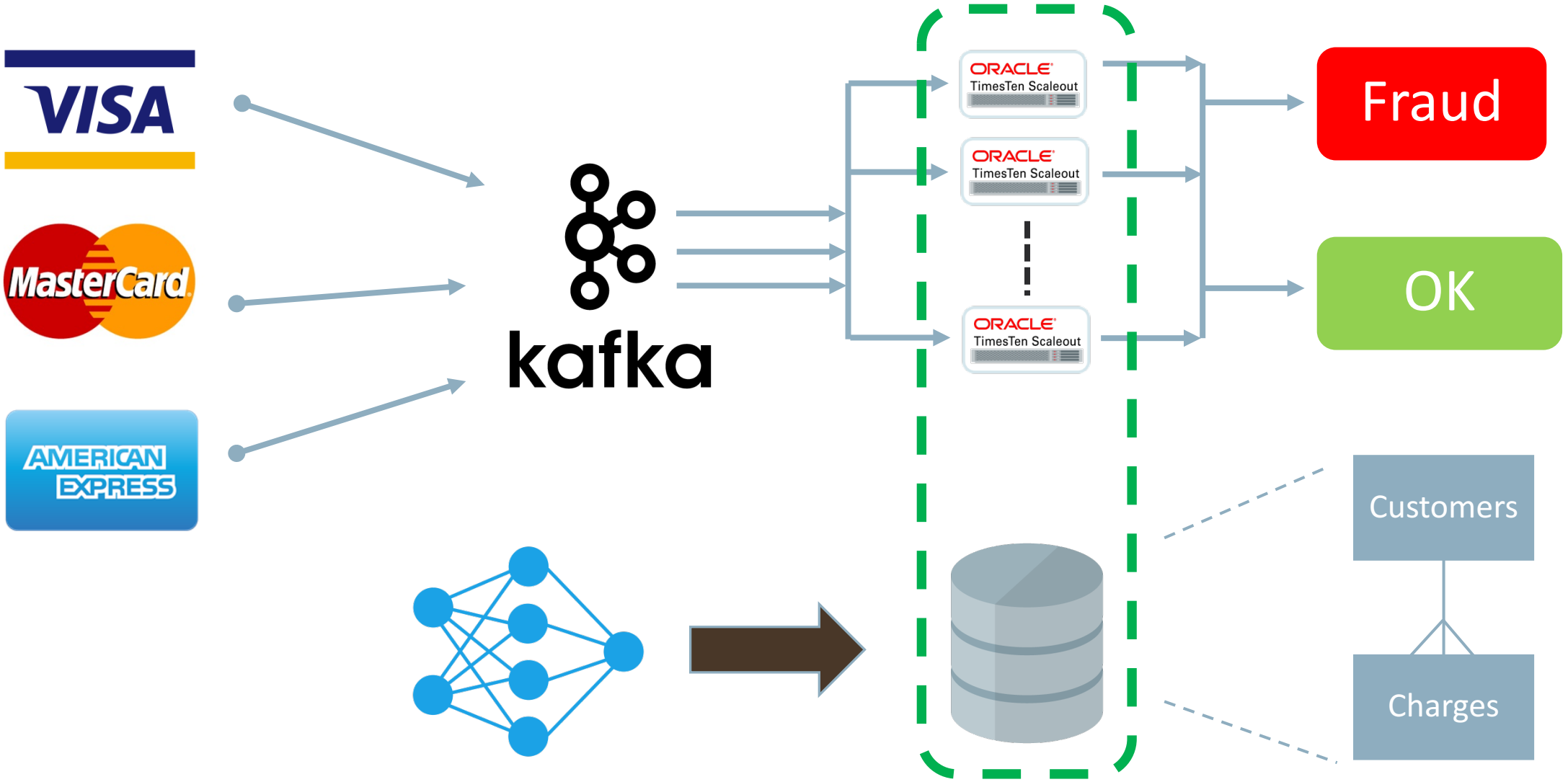
Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

An **IoT pipeline** for Real Time Credit Card Fraud Detection



Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

An **IoT pipeline** for Real Time Credit Card Fraud Detection



Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

TimesTen Scaleout Deployment

TimesTen Scaleout: The World's Fastest OLTP Database

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

TimesTen in On Premise

- TimesTen Velocity Scale requires :
  - Linux x8664 (glibc 2.12+)
    - Oracle Linux / Red Hat / CentOS 6.4+, 7+
    - Ubuntu 14.04+
    - SuSE 12+
  - JDK 8+
  - TCP/IP or IPoIB
  - A file system [eg ext4, not ext2 or ext3]
  - Enough RAM for the DB

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

TimesTen Scaleout on OCI, AWS, Azure, Google & OpenStack

1. Create your network, VMs, security configuration
2. Download TimesTen from OTN
3. Download Java
4. Download Apache Zookeeper
5. Unzip, configure and deploy

Amazon Linux

redhat

SUSE

ubuntu

CentOS

ORACLE LINUX

ORACLE TimesTen Scaleout

Java

Apache Zookeeper

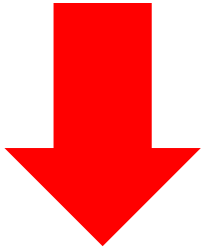
ZIP

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

TimesTen Scaleout
Oracle Cloud provisioning with TerraForm *[available soon]*

1. Download provisioning script [101KB] & TimesTen 18.1.1.3.0 [389 MB]
2. ./ScaleOutRollOut.py
   ==> Provide Region, Compartment, Shape, OS, Number of Instances

Terraform



ORACLE
LINUX

ORACLE
TimesTen Scaleout

Apache
Zookeeper

ORACLE

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

中国平安 PING AN
保险·银行·投资

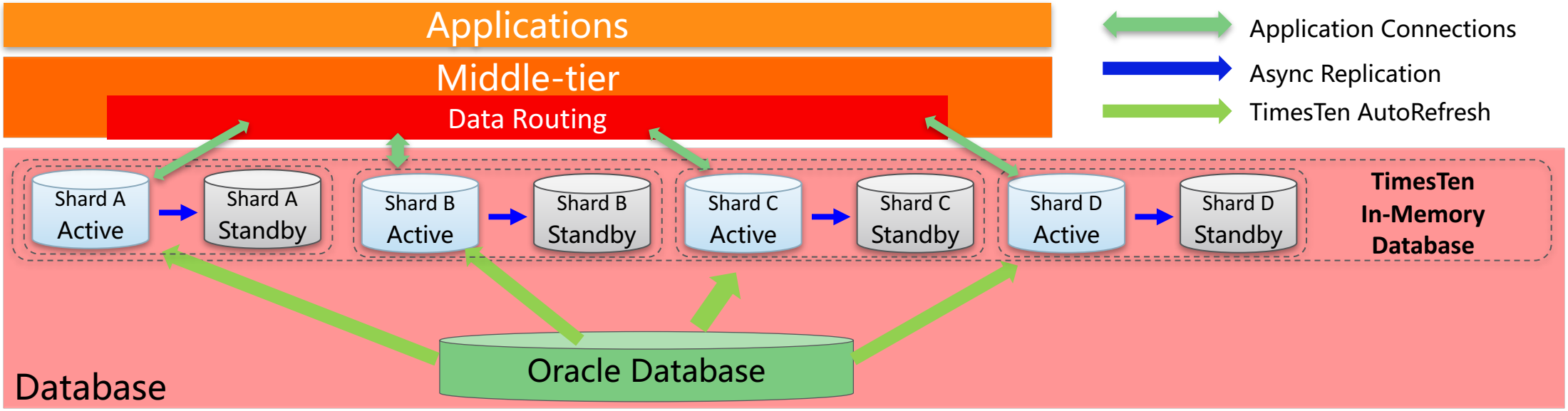
AODI
Agile Operation,Dynamic Infrastructure
运营总动员

TimesTen In-Memory Database

— Pingan Usecase

Sean Wong
General Manager of Database & Storage Product Departments
Pingan Technology (Shenzhen) Ltd

TimesTen Classic Challenges



- Applications must deal with data routing function.
- Very complicated to manage, so many shard instances.
  - You need to log into every host in order to perform management activities (such as install/deploy/start/stop).
- Have trouble in redefining tables:
  - You must recreate cache group while redefine tables.
  - You must recreate ASP (Active Standby Pair) while redefine tables.
- Difficult to scale out. Need to manually add shards and rebalance the data among the shards.
- Data distribution skew. Application-managed sharding mechanism is hard to distribute the data evenly.

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

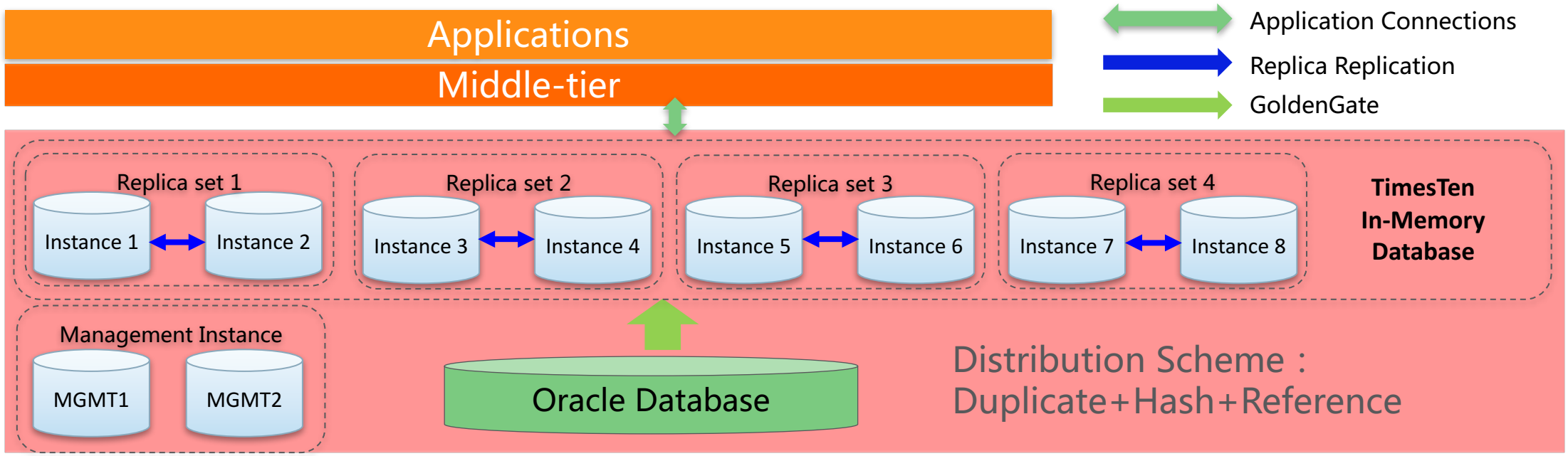
Pingan Current Challenges

- Huge data volume, tens of millions row in a single table
- Several levels of reference relationships between tables
- Complicated business logic and lengthy SQLs, but high-performance requirements
- Fast data growth rate, periodic capacity increase
- Critical business function, high availability requirements

ORACLE

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

TimesTen Scaleout (Pingan Tech.)



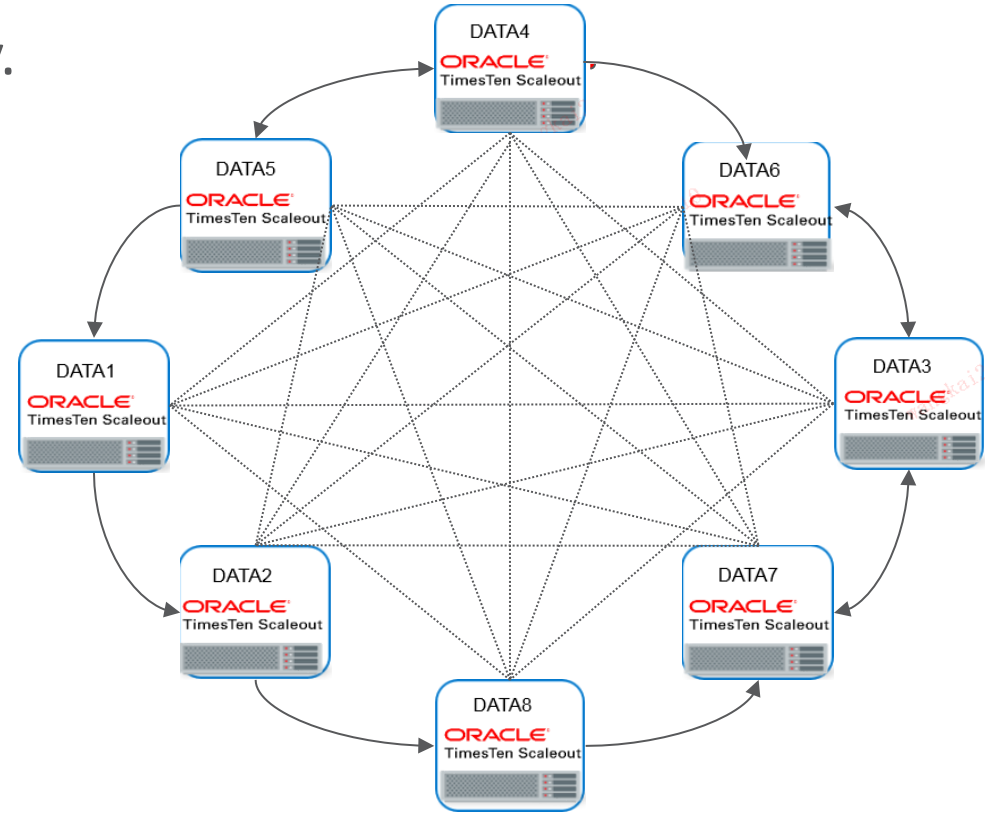
- Partition strategy: Duplicate + Hash + Reference. Hash is to solve the data distribution problem
- Allocated memory size is less than 150GB for a single instance
- Two Data Space Groups, each on difference server
- High Availability
  - Peer-to-peer design, K-safety data copies, application can connect to any node to access to the data, no matter where the data is

- Application does not need to deal with data routing function.
- Scalability:
  - TimesTen Scaleout enables you to add or remove instances in order to control both performance and the storage capacity of your database
- Centralized management:
  - You do not need to log onto every host within a grid in order to perform management activities. Instead, you conduct all management activity from a single instance using the ttGridAdmin utility.

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

TimesTen Scaleout High Availability

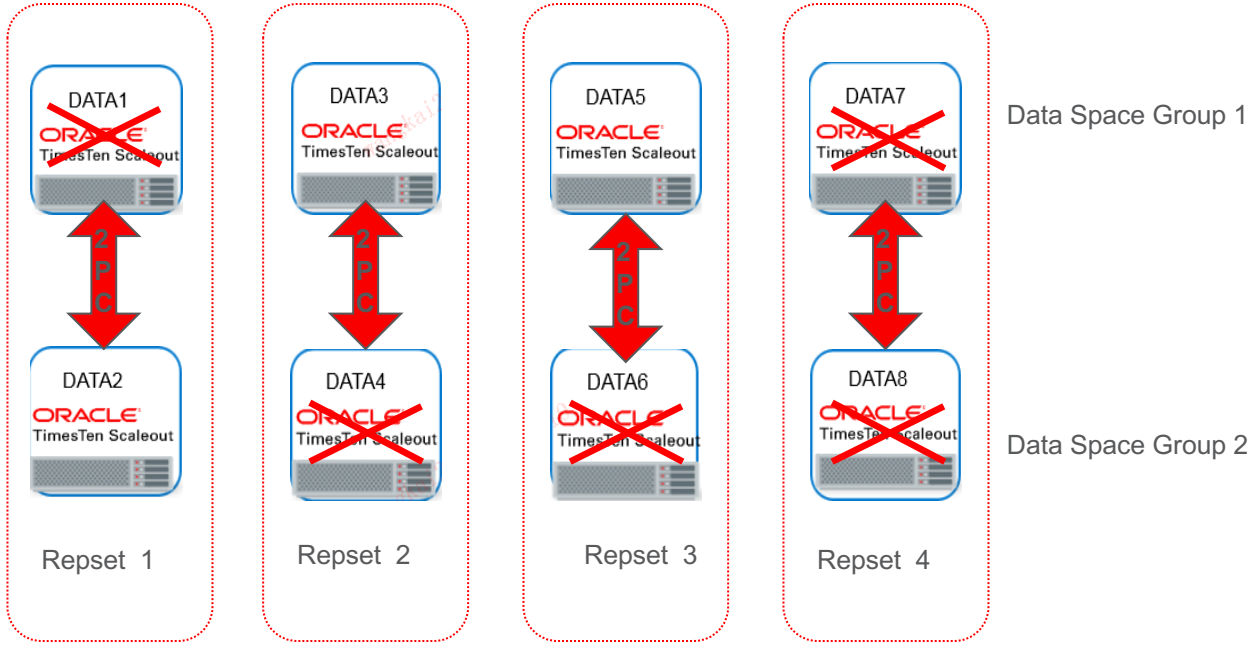
- The number of data copies is decided by K-safety. Currently K-safety 1 and 2 are supported. We set K-safety to 2
- Application can connect to any node to read or write data, no matter where the data is
- Queries on one node can be broadcasted to any other node to execute



Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

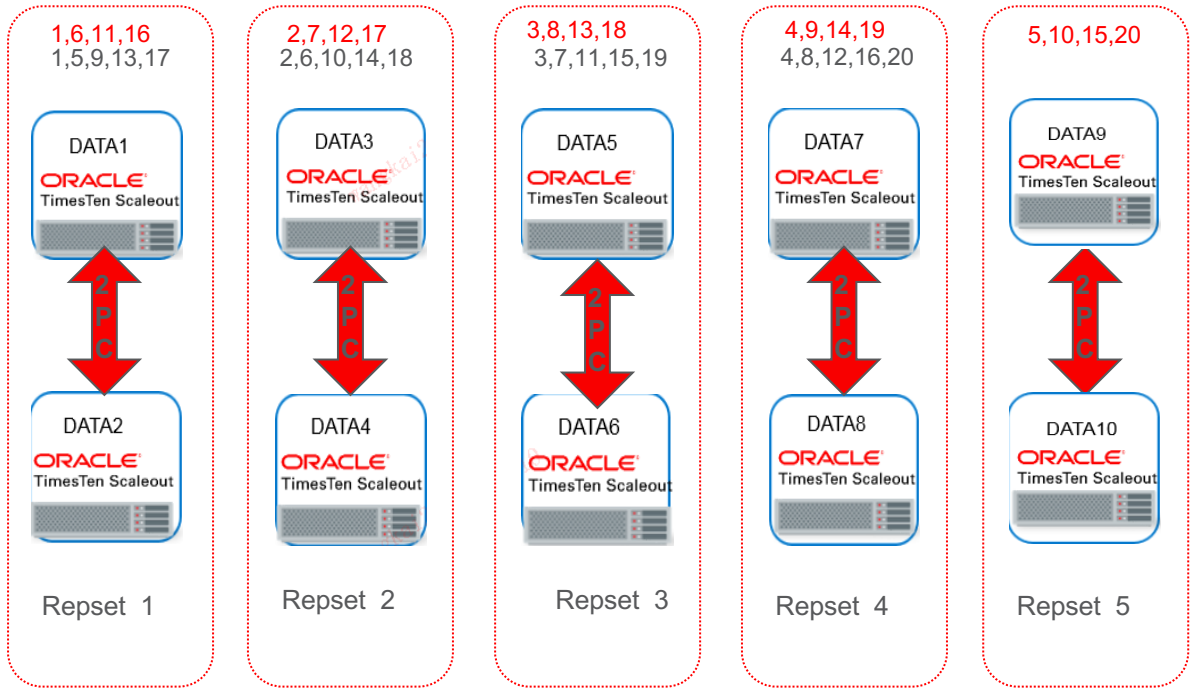
Pingan High Availability scenarios

- While multiple data nodes crashed in a cluster, as long as one data copy remains available, no impact to applications
- If the entire replicate set is not available, application still can access the data in the remaining replicate set if explicitly configured



Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

TimesTen ScaleOut



The data in Pingan Tele-sales system grows rapidly, we need to increase the capacity periodically. The scalability of TimesTen Scaleout is much more stronger than TimesTen classic.

In TimesTen Classic, it is cumbersome to scale out or down:

- Add nodes
- Redefine the data distribution mechanism and modify the distribution rule in the application
- Manually migrate the data following the new rule

It is easy to use TimesTen Scale to scale out

- Add nodes
- Manually kick start the re-distribution

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

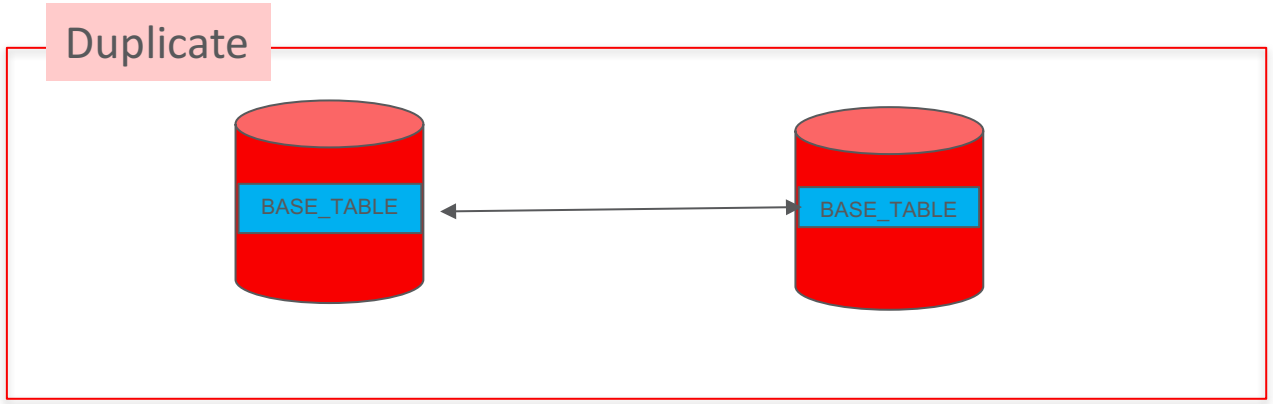
3 Partitioning Methods

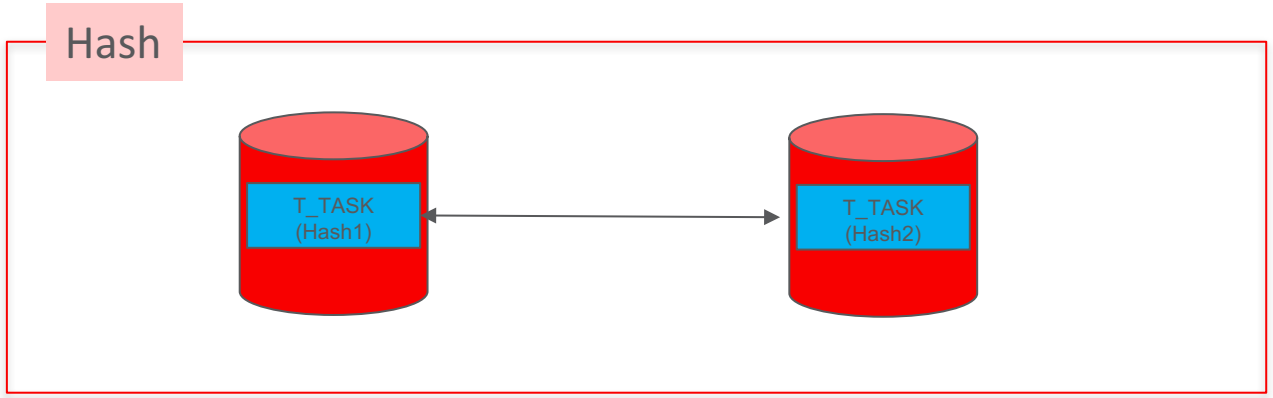
Distribution challenges:

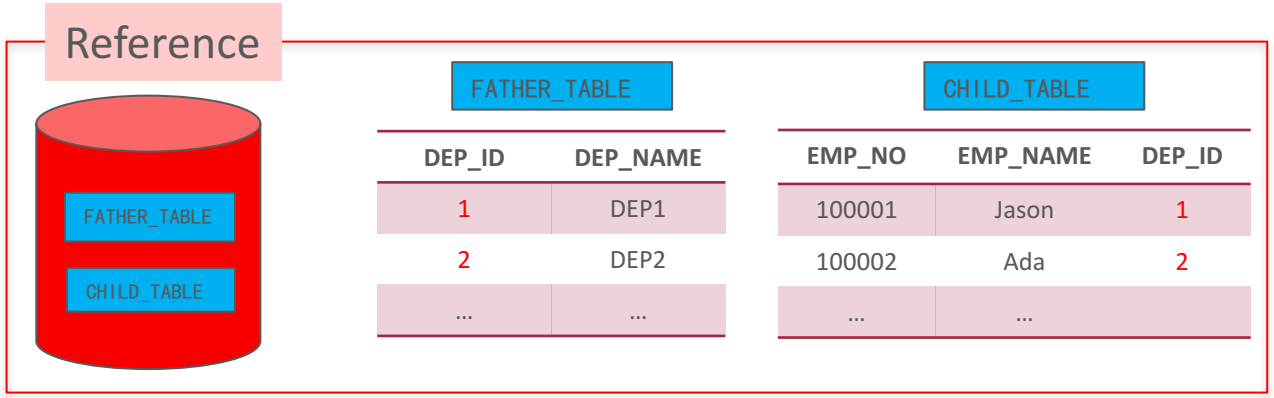
- Tens of million rows in a single table
- Many reference relationships
- Base tables with small data volume

3 Partitions Methods in Scaleout

- Duplicate：Base tables
- Hash：Large tables
- Reference：Eliminate the cross-shards query
- Hash and Reference work together







Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

Comparison between TimesTen Classic and TimesTen Scaleout

Pros

- Support real-time, high throughput OLTP applications
- Complicated SQLs can be run in parallel
- SQL-Compliant, Distributed In-Memory RDBMS
- Multiple data replicas designed for High Availability
- Highly compatible with Oracle, easy migration to Scaleout
- Checkpoint and Transaction Log for durability on every node
- Multiple Partitioning Methods (Hash, Reference, Duplicate)

Cons

- Performance drops a bit for multi-values and range scan after sharding the data

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

What we expect for TimesTen Scaleout

- Provide support for PLSQL packages, procedures and functions
- Support Cache Group like in TimesTen Classic, easy to sync data between TimesTen Scaleout and Oracle
- Support other distribution methods like Range and List
  - Currently Duplicate, Hash and Reference are supported

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

Q & A

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |