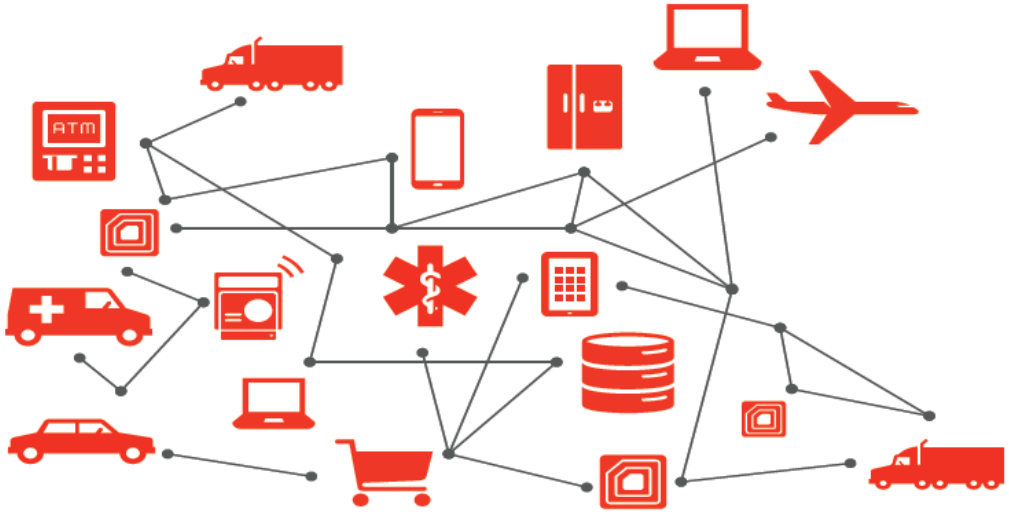
Unlocking New Big Data Insights with Hadoop & MySQL

Ricky Setyawan
MySQL Principal Consultant - ASEAN

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

An Avalanche of Data



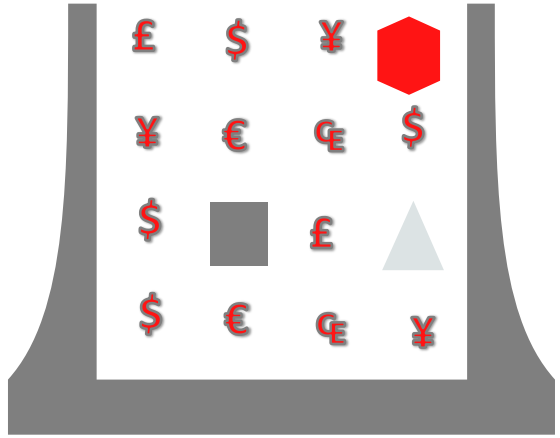
Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

Big Data What It Is, What it Means

Volume

Variety

Velocity



Create Value

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

What's Changed?

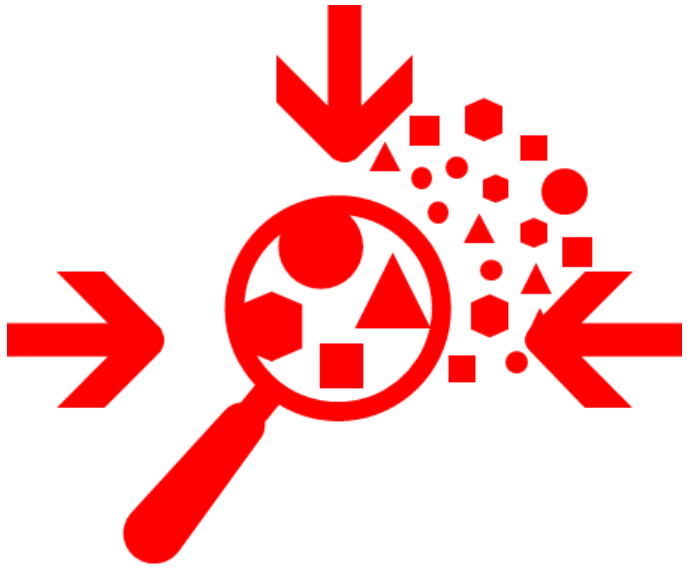
- Enablers
  - Digitization – nearly everything has a digital heartbeat
  - Ability to store much larger data volumes (distributed file systems)
  - Ability to process much larger data volumes (parallel processing)
- Why is this different from BI/DW?
  - Business formulated questions to ask upfront
  - Drove what was data collected, data model, query design

➢Big Data Enables what-if analysis, real-time discovery

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

Big Data Adoption

- Web Recommendations
- Sentiment Analysis
- Marketing Campaign Analysis
- Customer Churn Modeling
- Fraud Detection
- Research and Development
- Risk Modeling
- Machine Learning



Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

Leading Use-Case, On-Line Retail

Browsing

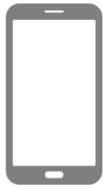
Recommendations

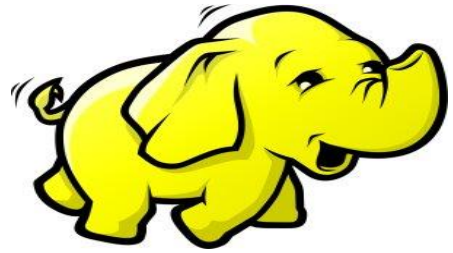
Profile,
Purchase
History

Recommendations

Social media updates
Preferences
Brands “Liked”

Web Logs:
Pages Viewed
Comments Posted



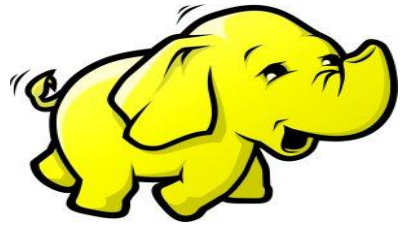


ORACLE

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

Why Hadoop?

- Scales to thousands of nodes, PB of structured and unstructured data
  - Combines data from multiple sources, schema-less
  - Run queries against all of the data
- Runs on commodity servers, handle storage and processing
- Data replicated, self-healing
- Initially just batch (Map/Reduce) processing
  - Extending with interactive querying, via Apache Drill, Cloudera Impala, Stinger etc.

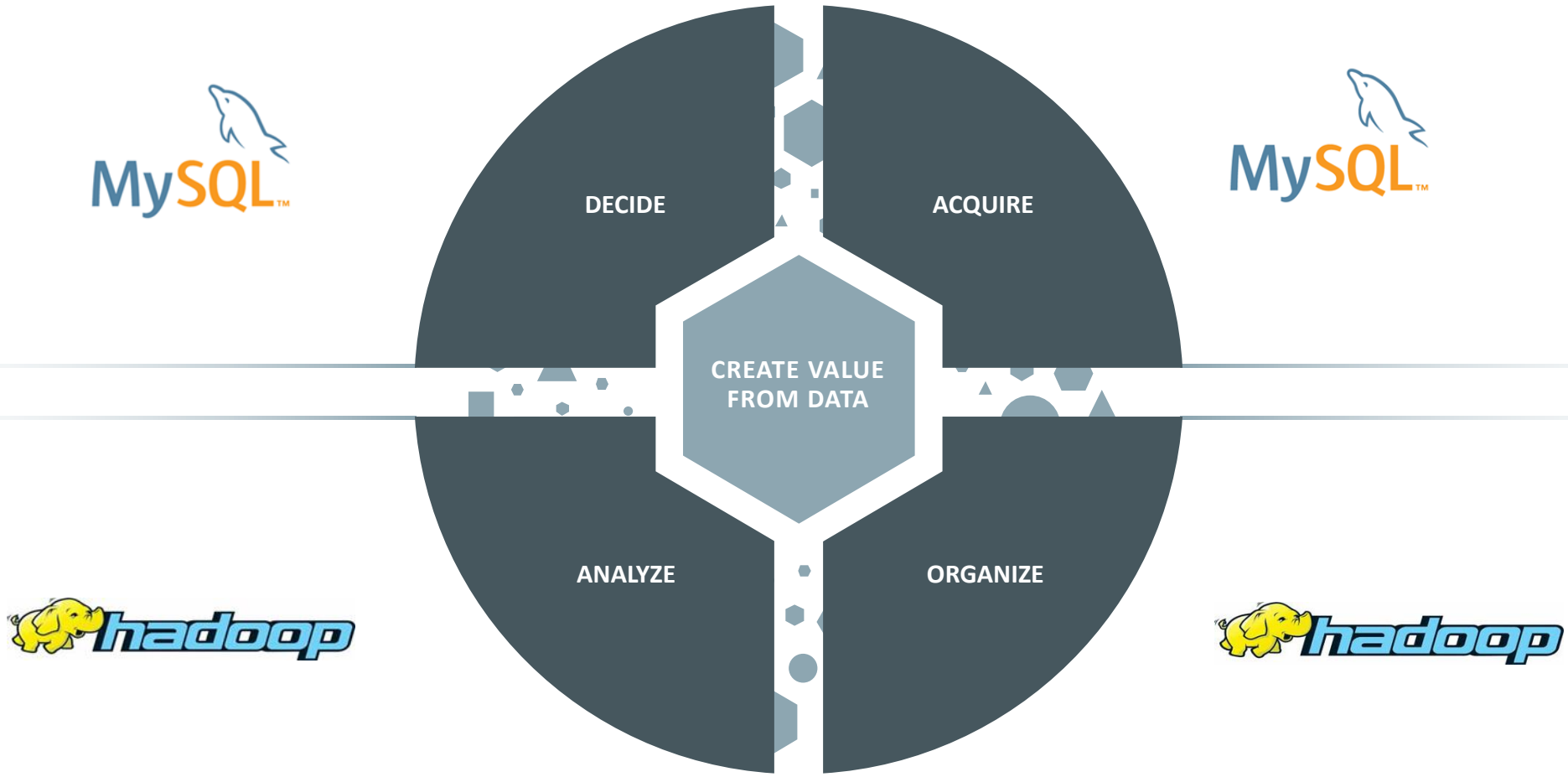


Copyright © 2014, Oracle and/or its affiliates. All rights reserved.

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

Big Data Lifecycle

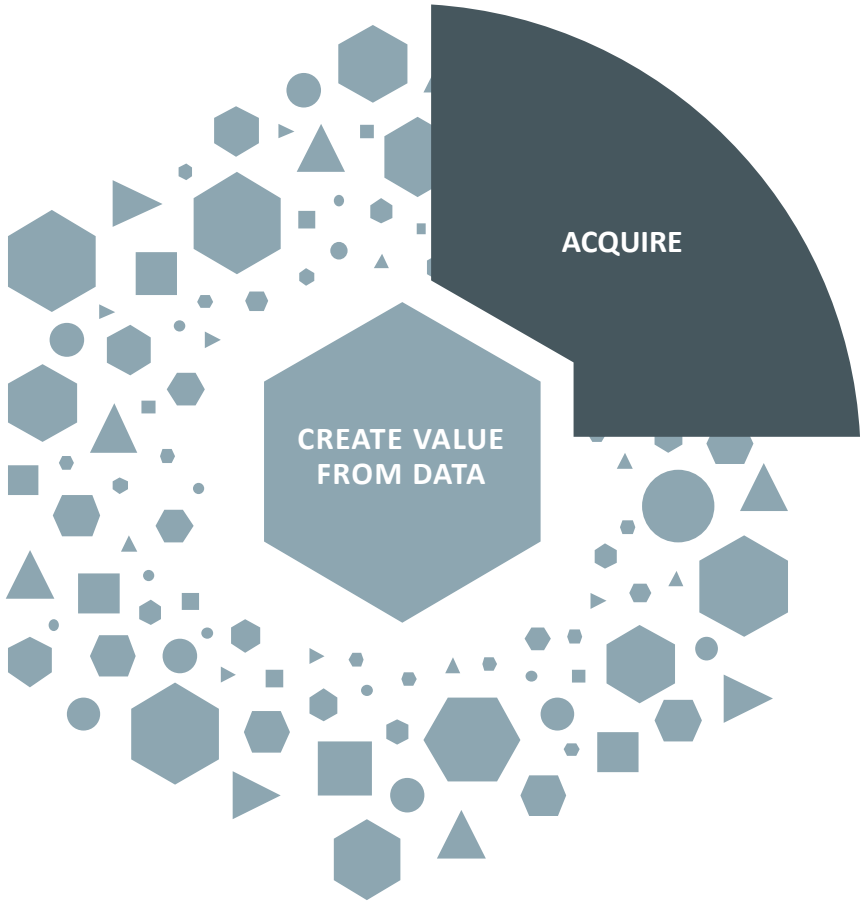
Better Decisions Using Big Data



Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

Big Data Lifecycle

Better Decisions Using Big Data



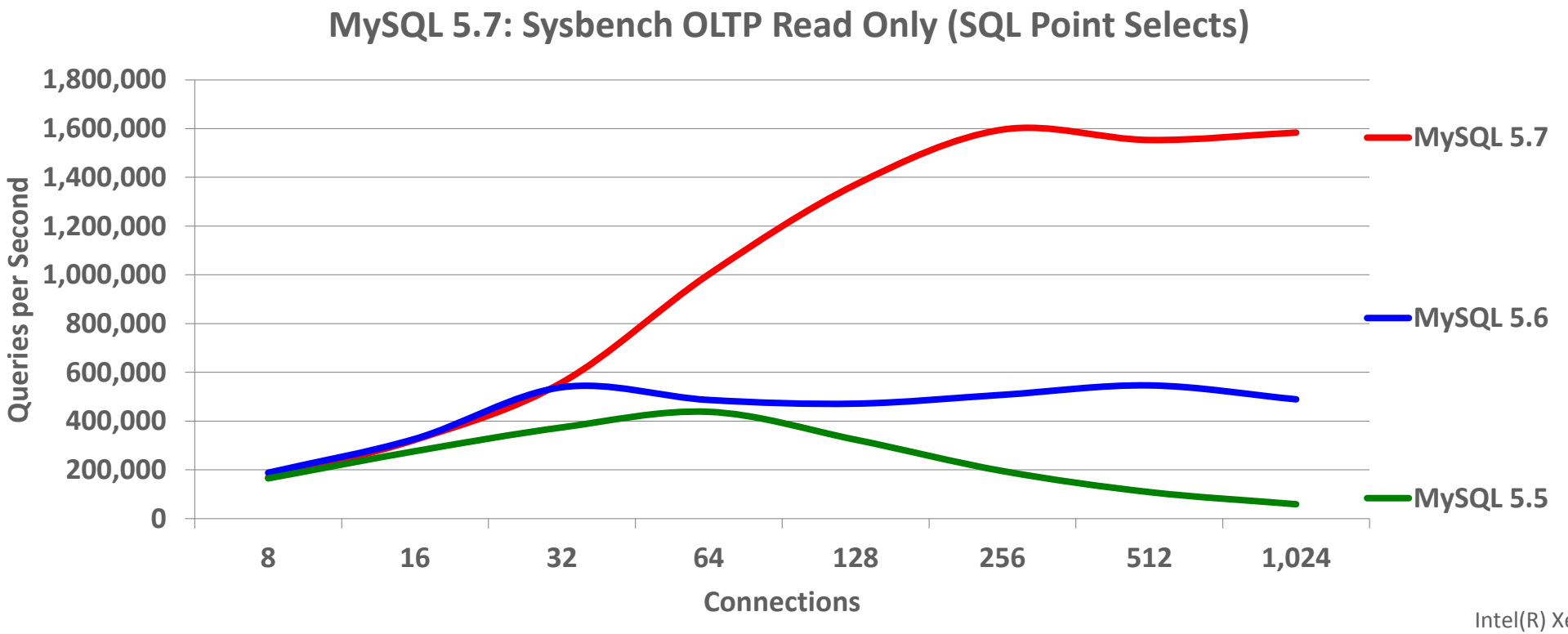
MySQL Database
MySQL Cluster
JSON Support
NoSQL Interfaces
MySQL Fabric

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

MySQL 5.7 Sysbench Benchmark: SQL Point Selects

3x Faster than MySQL 5.6

1,600,000 QPS



Intel(R) Xeon(R) CPU E7-8890 v3
4 sockets x 18 cores-HT (144 CPU threads)
2.5 Ghz, 512GB RAM
Linux kernel 3.16

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

Hybrid Database: Rock Solid Reliability + Flexibility

Modern Applications

Require **agile** development and operations with **robust data protection** and **security**

Traditional RDBMS

Proven, **transactional**, **secure**
Complex JOINs and queries
Extensive **operational tools**

NoSQL Solutions

Flexible. Easy-to-use.
Schema-less document storage

MySQL 5.7
JSON Support

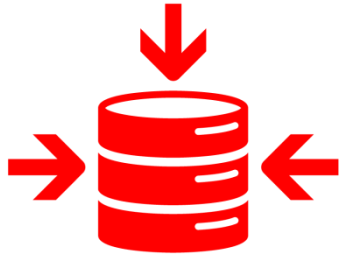
Hybrid Database

No trade-offs, best of both worlds. ACID properties & reliability of RDMS + flexible document management

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

MySQL NoSQL Interfaces: Fast, Flexible, Safe

Blazing Fast Key/Value Queries



Fully Transactional/ ACID

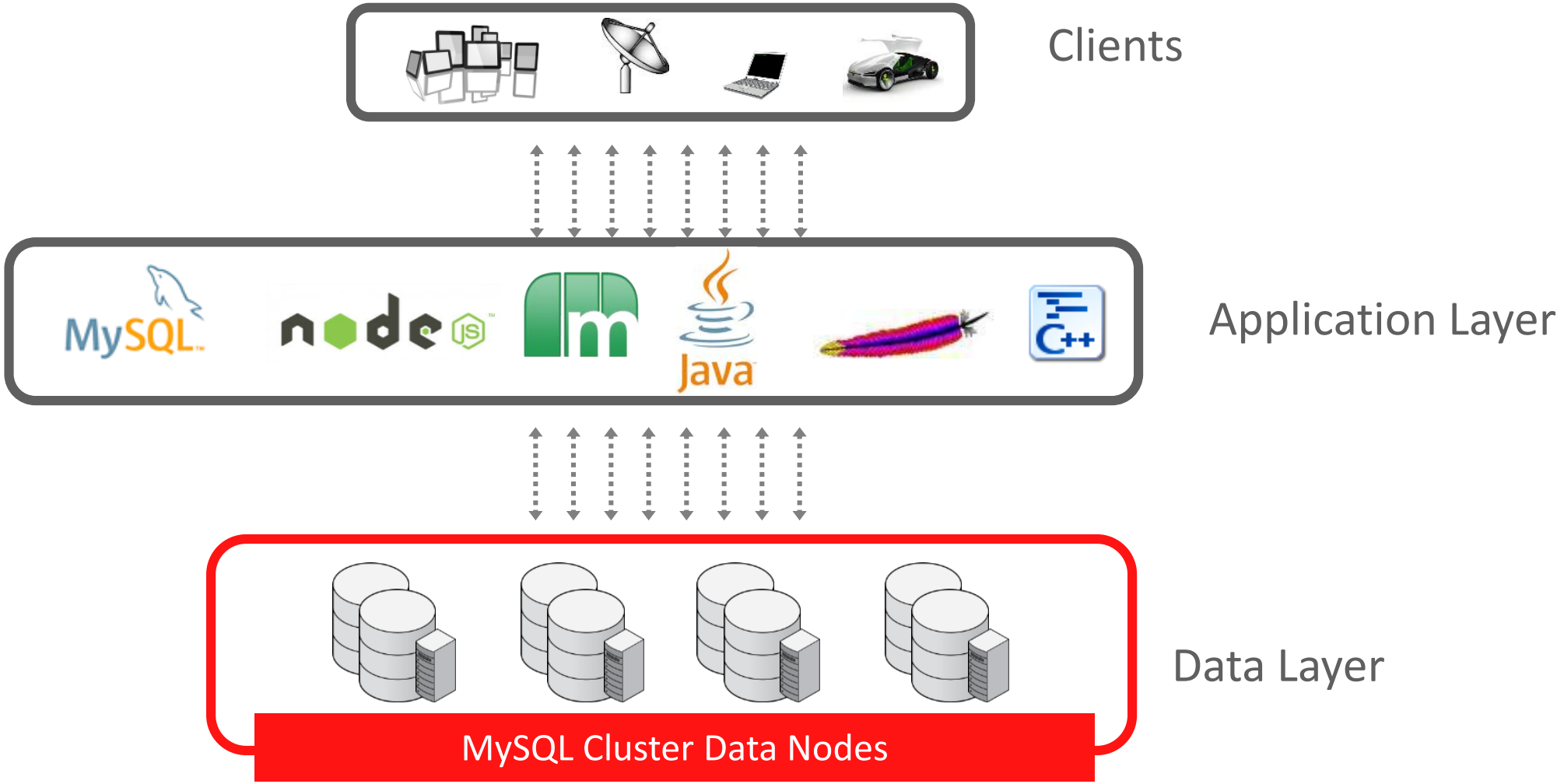
NoSQL And SQL Across the same data Set

Combined with Schema Flexibility: Online DDL

ORACLE

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

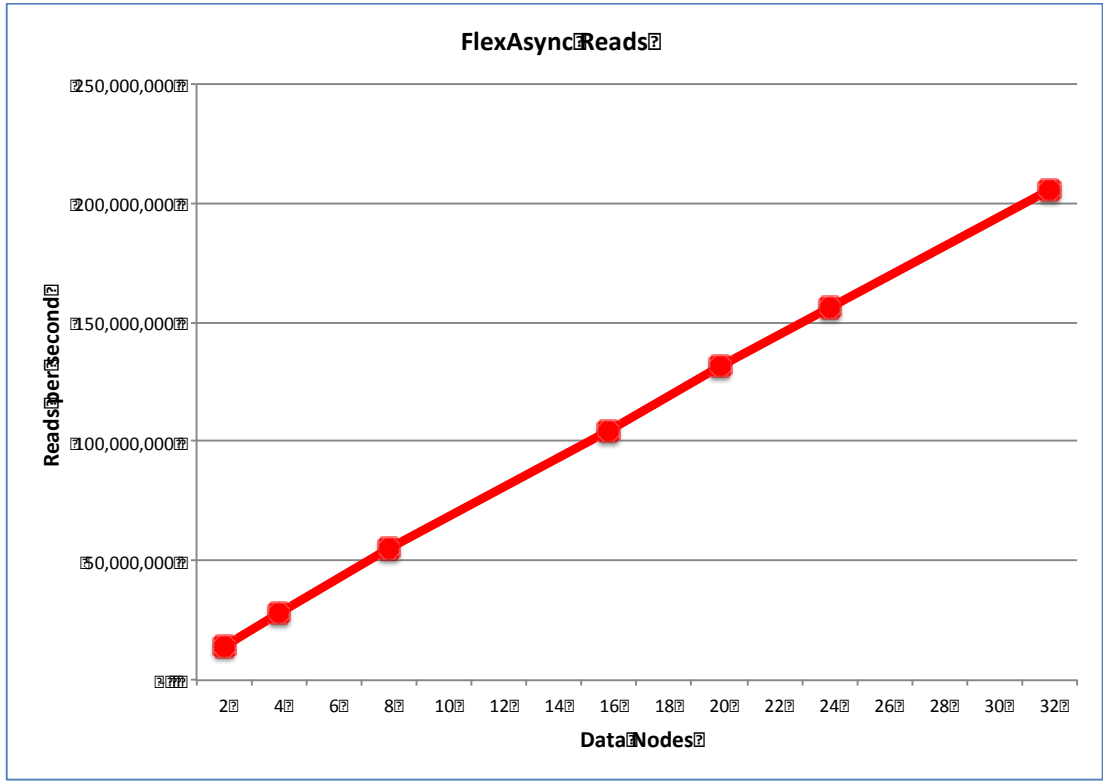
NoSQL Interfaces to MySQL Cluster



Copyright © 2015, Oracle and/or its affiliates. All rights reserved. | Copyright 2015, oracle and/or its affiliates. All rights reserved

MySQL Cluster 7.4 NoSQL Performance
200 Million NoSQL Reads/Second

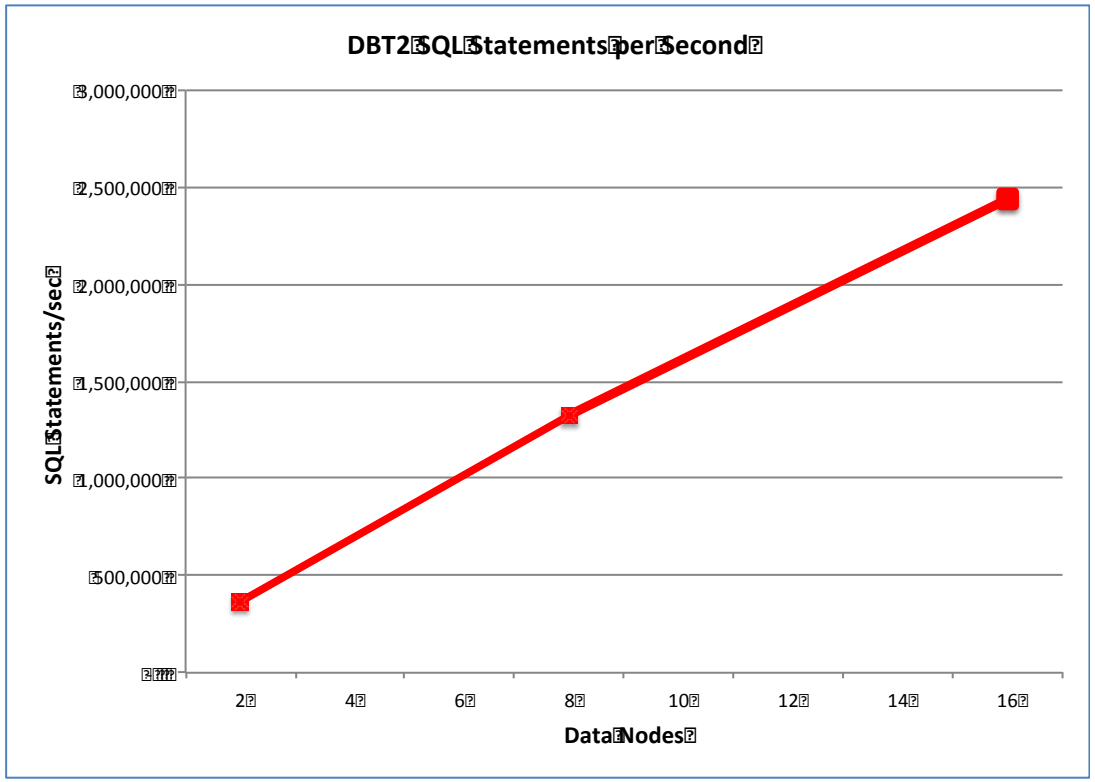
- Memory optimized tables
  - Durable
  - Mix with disk-based tables
- Massively concurrent OLTP
- Distributed Joins for analytics
- Parallel table scans for non-indexed searches
- MySQL Cluster 7.4 FlexAsych
  - 200M NoSQL Reads/Second



Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

MySQL Cluster 7.4 SQL Performance
2.5M SQL Statements/Second

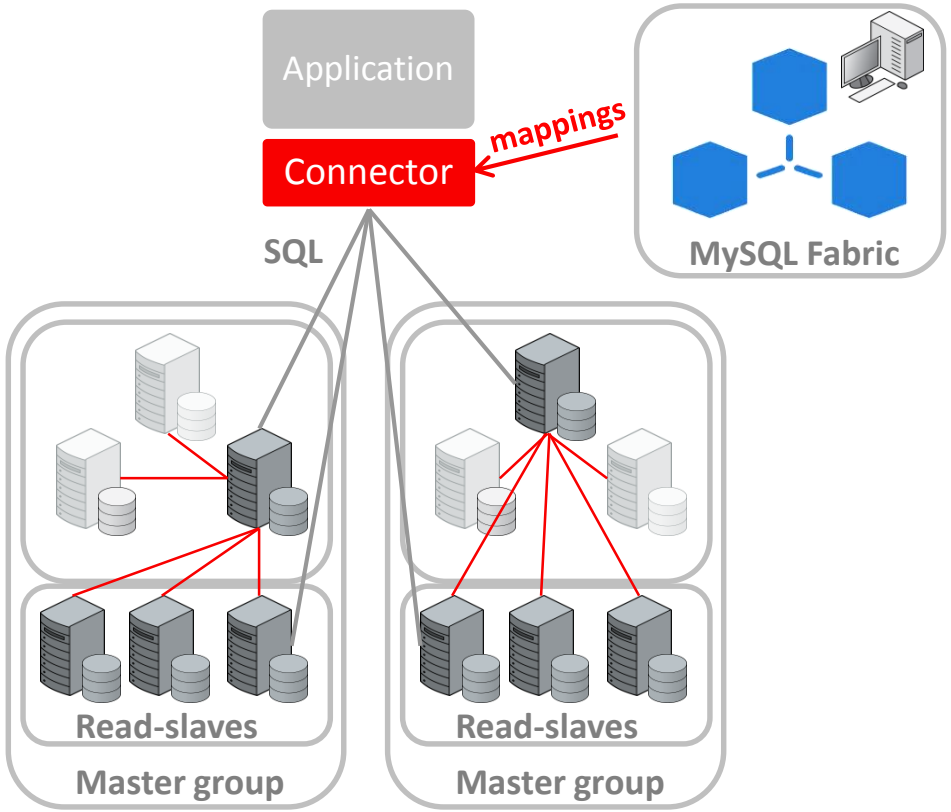
- Memory optimized tables
  - Durable
  - Mix with disk-based tables
- Massively concurrent OLTP
- Distributed Joins for analytics
- Parallel table scans for non-indexed searches
- MySQL Cluster 7.4 DBT2 BM
  - 2.5M SQL Statements/Second



Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

MySQL Fabric

Scale out with Data Sharding + High Availability

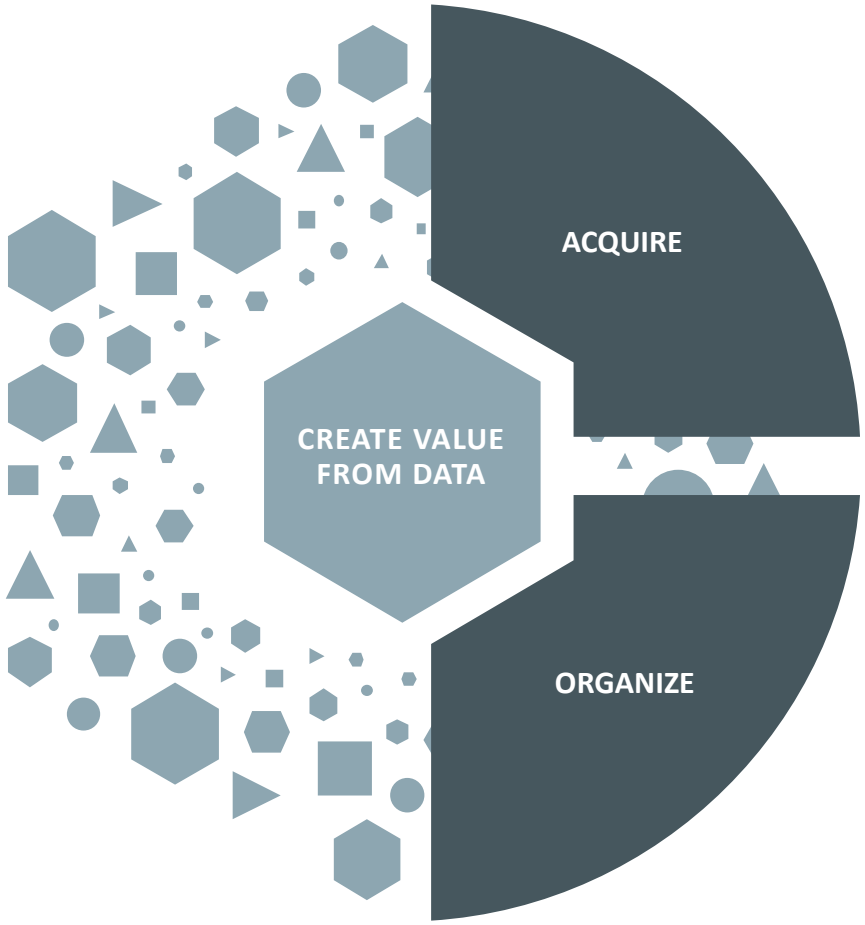


- Scale-out through sharding
  - Read AND Write
  - Standard framework, no more custom solutions
- HA out of the box
  - On top of Replication
  - Automatic failover
  - Automatic routing

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

Big Data Lifecycle

Better Decisions Using Big Data



Import Data

Apache Sqoop

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

Apache Sqoop

- Apache TLP, part of Hadoop project
  - Originally developed by Cloudera
- Bulk data import and export
  - Between Hadoop (HDFS) and external data stores
- JDBC Connector architecture
  - Supports plug-ins for specific functionality
- “Fast Path” Connector developed for MySQL

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

labs.mysql.com

MySQL Applier for Hadoop

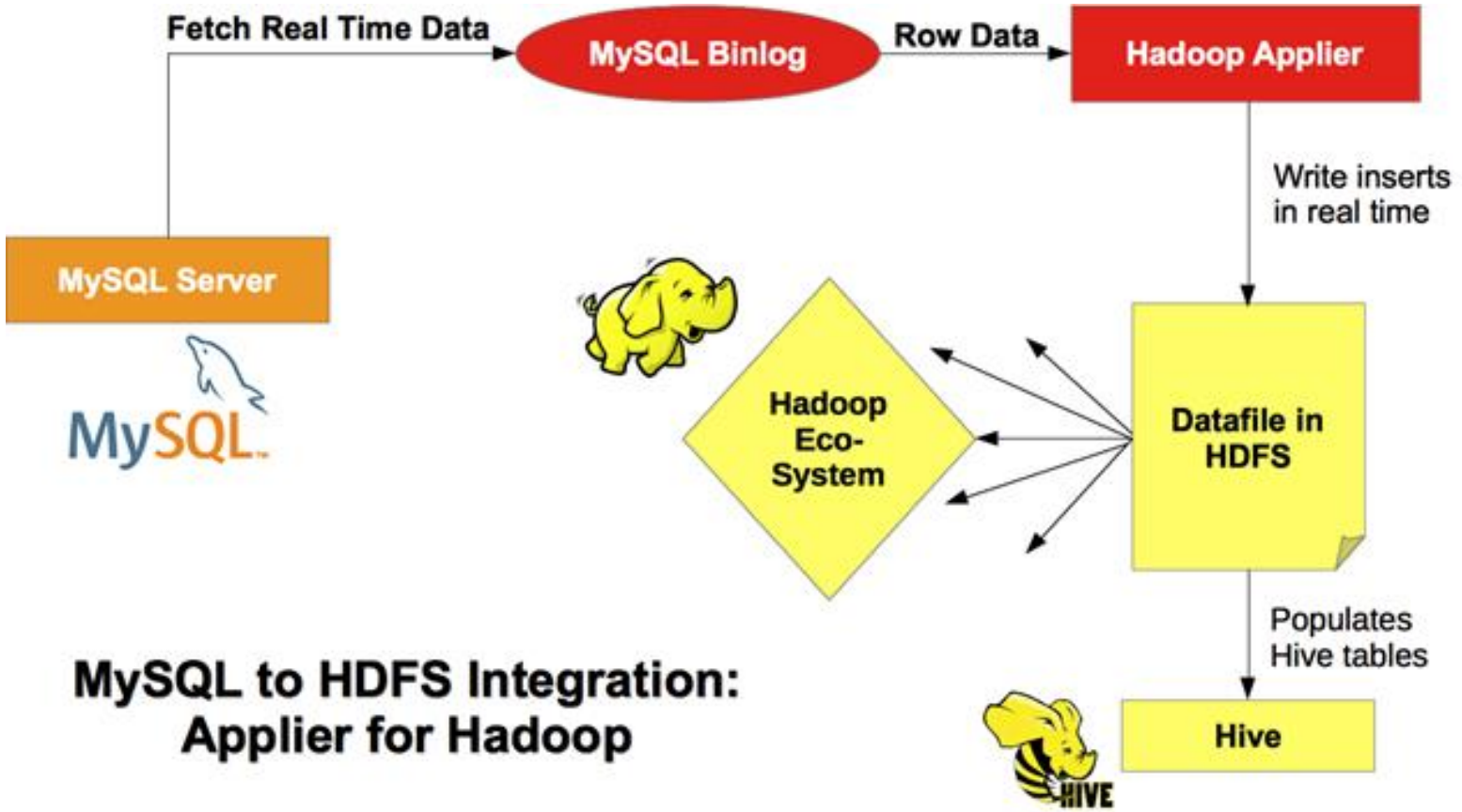
- Real-time streaming of events from MySQL to Hadoop
  - ➢Supports move towards “Speed of Thought” analytics
- Connects to the binary log, writes events to HDFS via libhdfs library
- Each database table mapped to a Hive data warehouse directory
- Enables eco-system of Hadoop tools to integrate with MySQL data
- Available for download now: labs.mysql.com

ORACLE

Copyright © 2014 Oracle and/or its affiliates. All rights reserved. | Copyright 2014, Oracle and/or its affiliates. All rights reserved.

labs.mysql.com

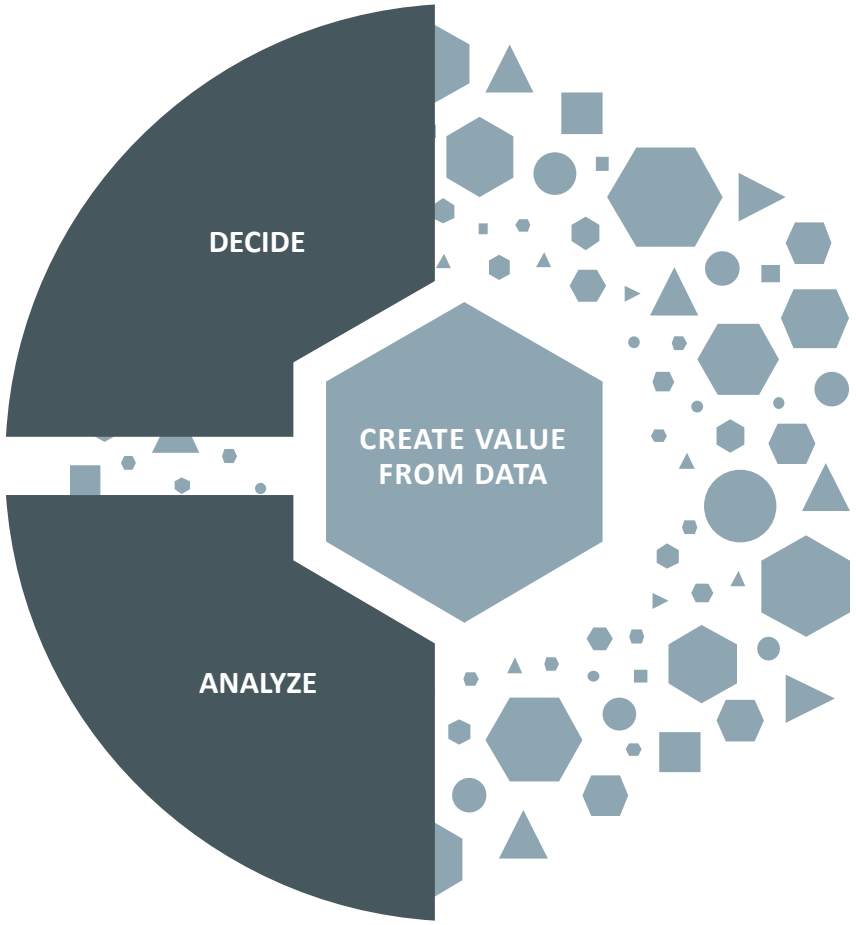
MySQL Applier for Hadoop



Copyright © 2014 Oracle and/or its affiliates. All rights reserved. |

Big Data Lifecycle

Better Decisions Using Big Data



Analyze
Export Data
Decide

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

Analyze Big Data in Hadoop

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

MySQL Reporting Database for BI

pentaho™
open source business intelligence™

MySQL™

talend*
*open data solutions

JASPERSOFT

Copyright © 2014, Oracle and/or its affiliates. All rights reserved.

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

Summary

- Create value from Big Data with MySQL
- MySQL + Hadoop: widely deployed solution (80% of Hadoop project)
- “Best of both worlds”: SQL + NoSQL Access; Schema-less data management
- Scale Out & data sharding with MySQL Fabric
- Tools and expertise to support you

ORACLE

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. |

ORACLE®