


# ORACLE CLOUDWORLD DEVELOPER



# Enabling Big Data Analytics on the Oracle Cloud



Sachin Arora  
Lead Architect, Big Data, India  
April 08, 2016

# 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, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.



# Who is this?



- Male, born in 1948
- Grew up in England
- Married twice, children
- Successful, wealthy, celebrity
- Loves dogs and the Alps

**What are  
we missing ?**





# Oracle Brings Enterprise Productivity To Big Data

**Discover** and  
Predict – Fast



**Simplify** Access to  
All Data



**Govern** and Secure  
All Data



**Enterprise-Grade** Capabilities In-Cloud and On-Premise

Performance

Integration

Availability

Scalability

Manageability

# Oracle Big Data Appliance\* 2X Faster than Do-It-Yourself<sup>1</sup>



Faster

Cost Savings

Faster Time  
to Value

• Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit <http://www.intel.com/performance>. Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit <http://www.intel.com/performance>. 1 - Configurations were compared by using the Big Data Benchmark for BigBench. Oracle\* Big Data Appliance configuration included 6 nodes comprised of: Intel® Xeon® CPU E5-2699 v3 (HT enabled) with 128 GB DDR4, 12 X 4TB HDD, Infiniband network (1 connection) observed max throughput 24 Gb/sec, Oracle\* Linux Enterprise 6, and CDH\* 5.4.4 with modified configuration. DIY cluster configuration included 6 nodes comprised of: Intel® Xeon® CPU E5-2699 v3 (HT enabled) with 128 GB DDR4, 1 x 64GB SSD for OS, 12 X 4TB HDD, 10Gb network (1 connection), CentOS\* 6.6, CDH\* 5.3.3 with minimal changes. 21% Cost Savings and 30% Faster Time to Value figures based on [ESG Report: Getting real about Big Data: Build versus Buy](#)

# Available Now: Big Data Cloud Service



## Key Features

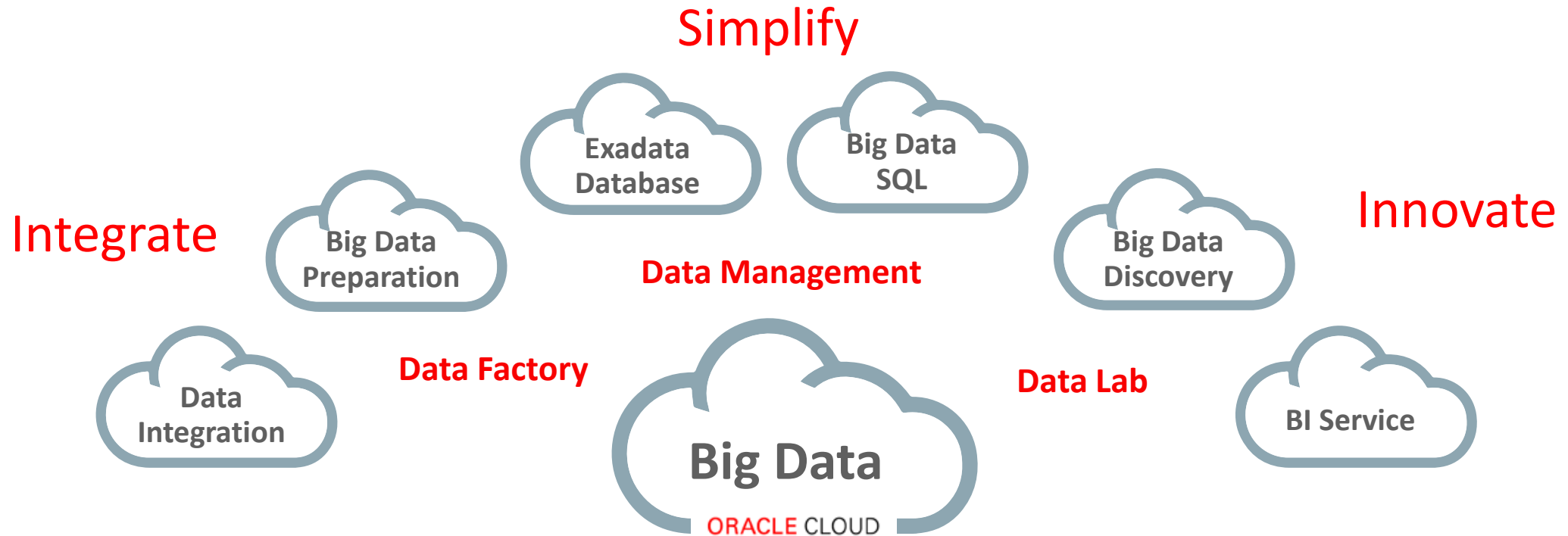
- Big Data (Hadoop, Spark) as a Service
  - Cloudera Enterprise – Data Hub Edition 5.4
  - Oracle Big Data Connectors
  - Oracle Big Data Spatial and Graph
  - Oracle Data Integrator with Advanced Big Data Option
  - Database Cloud Service integration (via Connectors)
- Big Data SQL service add-on
  - Unified query across Big Data and Exadata Cloud Services
- Platform for new Big Data Services
  - Big Data Discovery

## Benefits

- Consistently high performance
- Integration with Exadata Cloud Service means one fast SQL query on all your data



# Oracle Big Data in Oracle Public Cloud



# Core Design Principles for Oracle Big Data

**Operational Simplicity**

**Simplify Access to ALL Data**

# Core Design Principles for Oracle Big Data

**Operational Simplicity**

**Simplify Access to ALL Data**

# Core Design Principles for Oracle Big Data

## Operational Simplicity

- No Bottlenecks
- Full Stack Install and Upgrades
- Simplified Management
  - **Simplify Access to ALL Data**
  - Cluster Growth
  - Critical Node Migration
- Always Highly Available
- Always Secure
- Very Competitive Price Point



# Core Design Principles for Oracle Big Data

## Operational Simplicity

“Oracle Big Data Appliance is an excellent choice for customers looking to work with the full suite of Cloudera’s leading Hadoop-based technology. It’s more cost-effective and quicker to deploy than a DIY cluster.”

- Mike Olson, Cloudera founder, Chief Strategy Officer, and Chairman of the Board

**30%** Quicker to Deploy

**21%** Cheaper to Buy

Source: [ESG White Paper](#)



# Core Design Principles for Oracle Big Data

- Oracle Big Data SQL
  - Oracle SQL on ALL your data
  - All Native Oracle SQL Operators
  - Smart Scan for Optimized Performance
  - Govern all Data through a Single Set of Security Policies

**Operational Simplicity**

**Simplify Access to ALL Data**

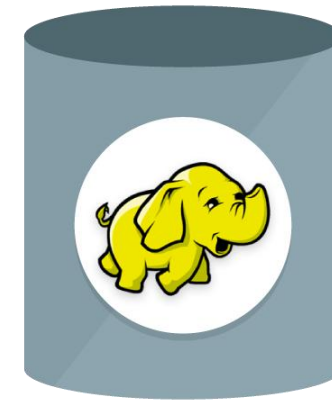
# Data Analytics Challenge

Separate data access interfaces



# Data Analytics Challenge

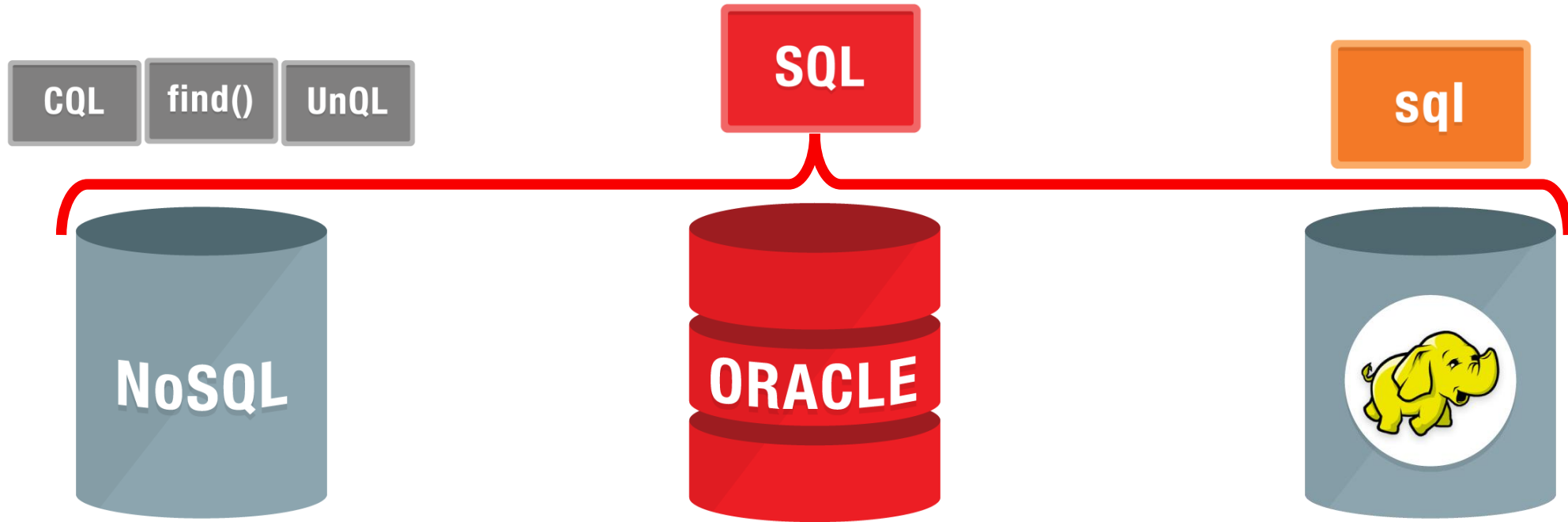
No comprehensive SQL interface across Oracle, Hadoop and NoSQL





# Oracle Big Data Management System

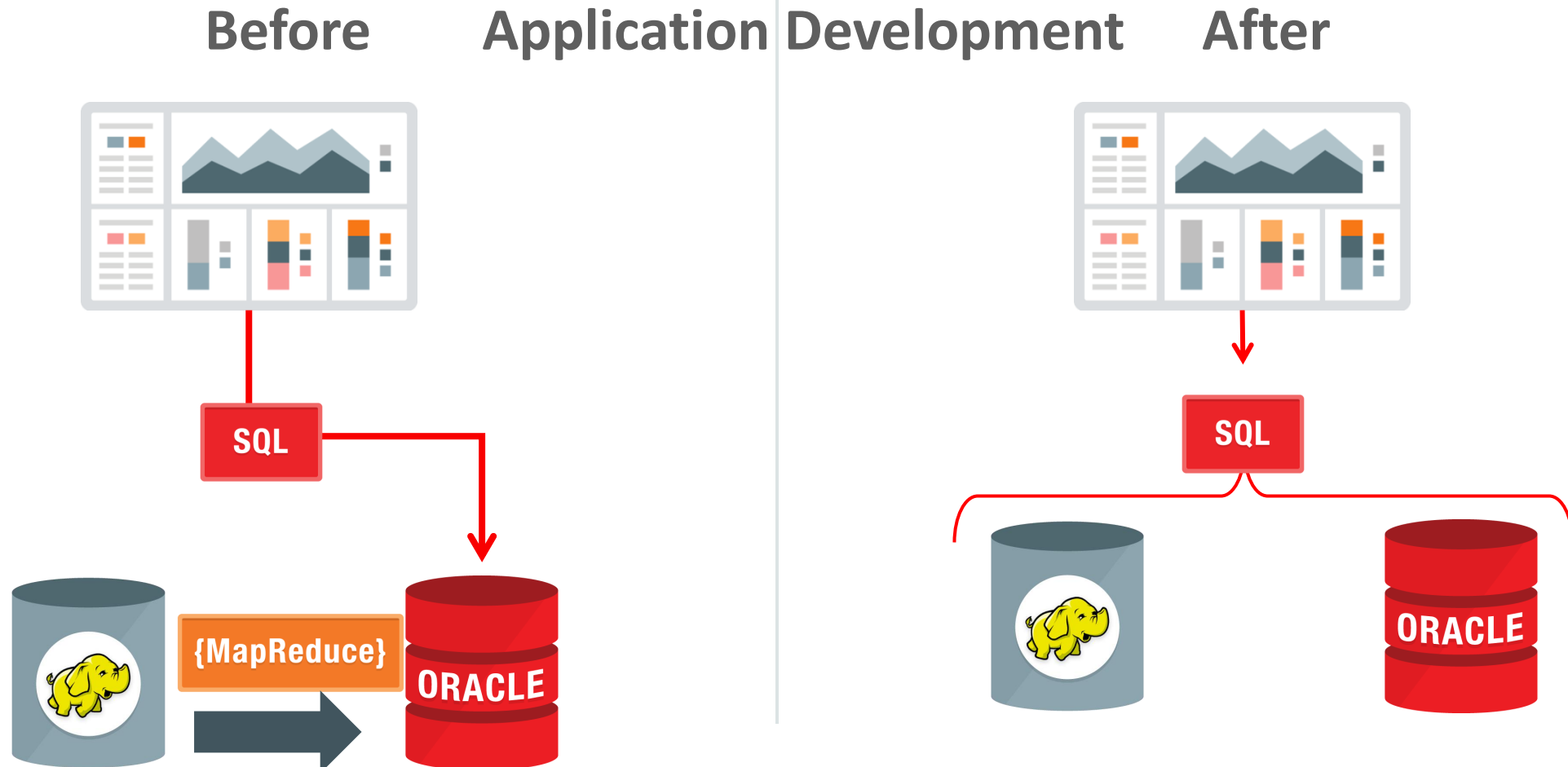
Rich, comprehensive SQL access to all enterprise data



# What Does Unified Query Mean for You?

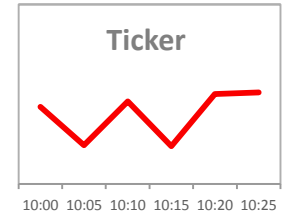


# What Does Unified Query Mean for You?



# Pattern Matching With Oracle SQL

## Compare To Penalty Paid With Limited SQL Support



## Finding Patterns in Stock Market Data - Double Bottom (W)

```
        return state;
    }

    return state;
}

private boolean eq(String a, String b) {
    if (a.isEmpty() || b.isEmpty()) {
        return false;
    }
    return a.equals(b);
}

private boolean gt(String a, String b) {
    if (a.isEmpty() || b.isEmpty()) {
        return false;
    }
    return Double.parseDouble(a) > Double.parseDouble(b);
}

private boolean lt(String a, String b) {
    if (a.isEmpty() || b.isEmpty()) {
        return false;
    }
    return Double.parseDouble(a) < Double.parseDouble(b);
}

public String getState() {
    return this.state;
}

}

BagFactory bagFactory = BagFactory.getInstance();

@Override
public Tuple exec(Tuple input) throws IOException {
    long c = 0;
```

250+ Lines of Java UDF

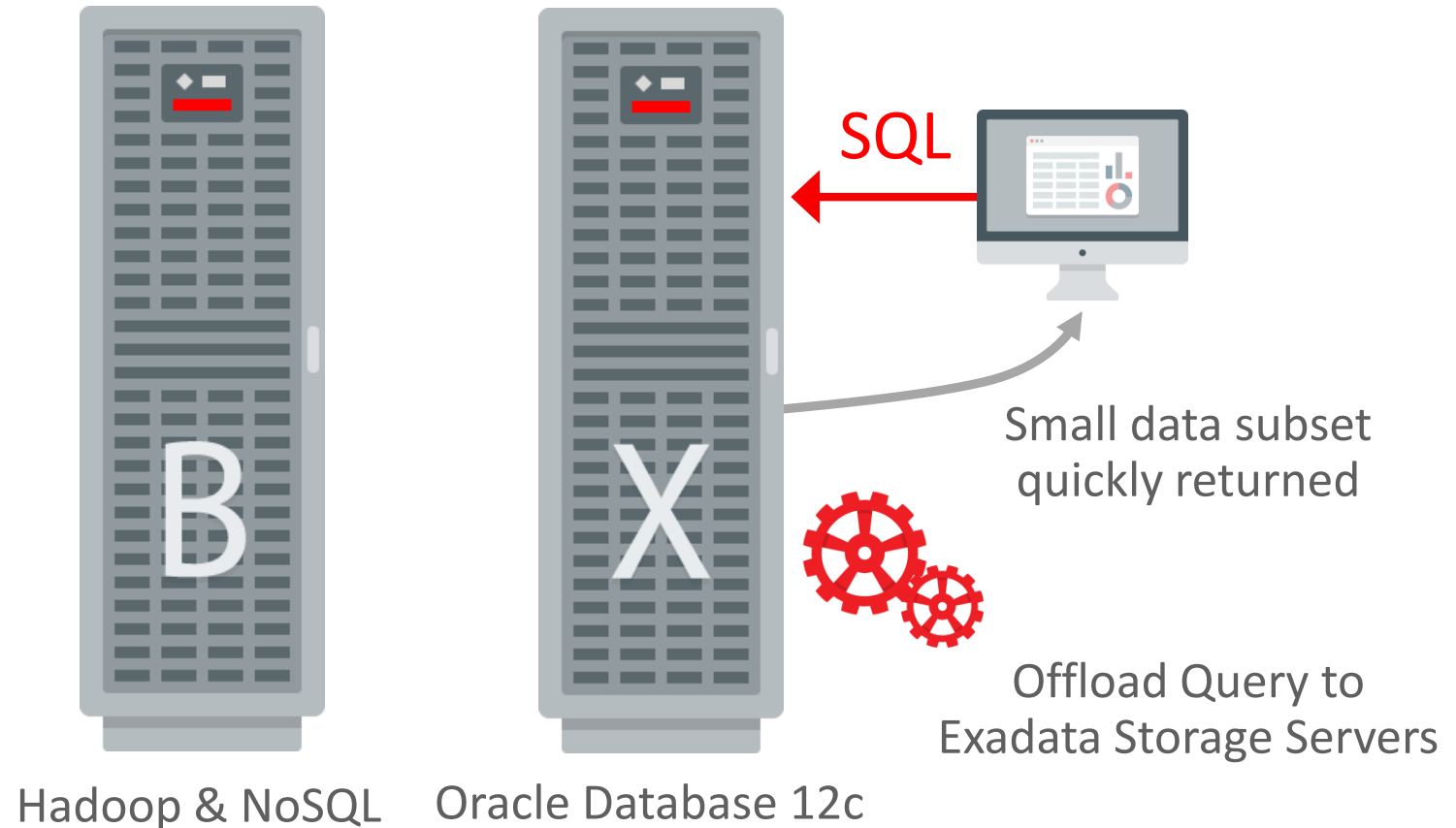
```
SELECT first_x, last_z
FROM ticker MATCH_RECOGNIZE (
    PARTITION BY name ORDER BY time
    MEASURES FIRST(x.time) AS first_x,
              LAST(z.time)  AS last_z
    ONE ROW PER MATCH
    PATTERN (X+ Y+ W+ Z+)
    DEFINE X AS (price < PREV(price)),
           Y AS (price > PREV(price)),
           W AS (price < PREV(price)),
           Z AS (price > PREV(price) AND
                z.time - FIRST(x.time) <= 7 ))
```

12 Lines of SQL

20x less code

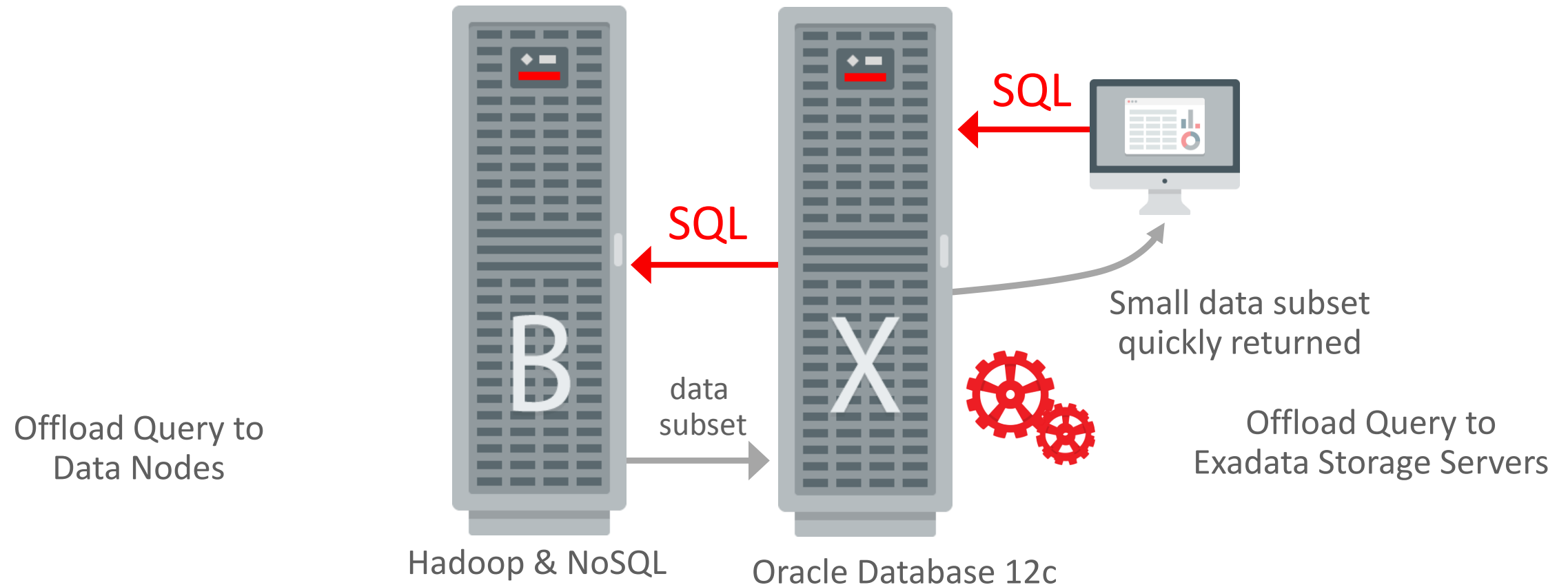


# What Gives Exadata Extreme Performance?

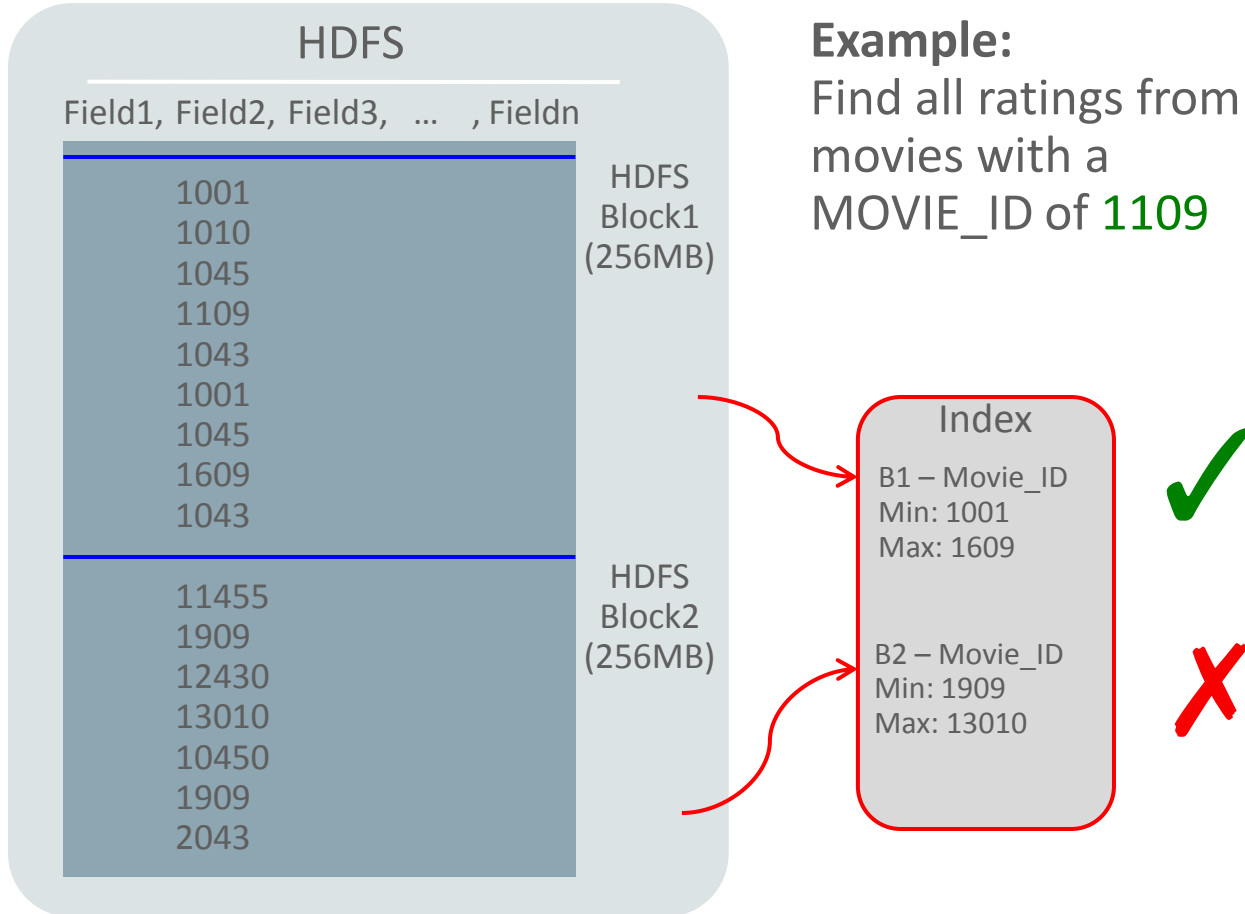


# Introducing Oracle Big Data SQL

## Massively Parallel SQL Query across Oracle, Hadoop and NoSQL



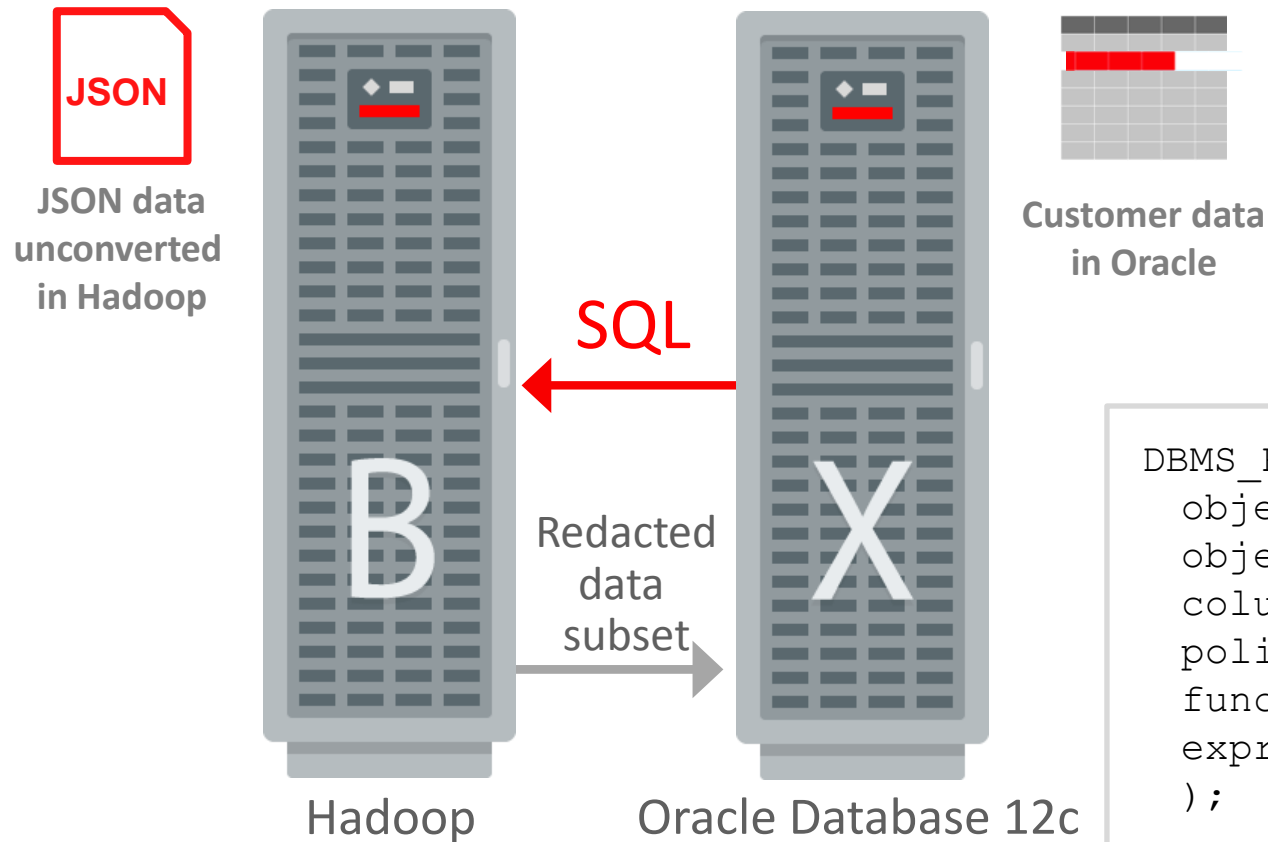
# Oracle Big Data SQL Storage Index



- Storage index provides query speed-up through transparent IO elimination of HDFS Blocks
- Columns in SQL are mapped to fields in the HDFS file via External Table Definitions
- Min / max value is recorded for each HDFS Block in a storage index

# Apply Advanced Security on Hadoop & NoSQL

**Same security policies apply to Hadoop & Relational**



- Redaction
- Virtual Private Database
- Fine-grain Access Control

```
DBMS_REDACT.ADD_POLICY(  
  object_schema => 'hr',  
  object_name => 'customer_address_ext',  
  column_name => 'ca_street_name',  
  policy_name => 'customer_address_redaction',  
  function_type => DBMS_REDACT.RANDOM,  
  expression => '1=1'  
);
```





# Integrated Cloud

## Applications & Platform Services

ORACLE®