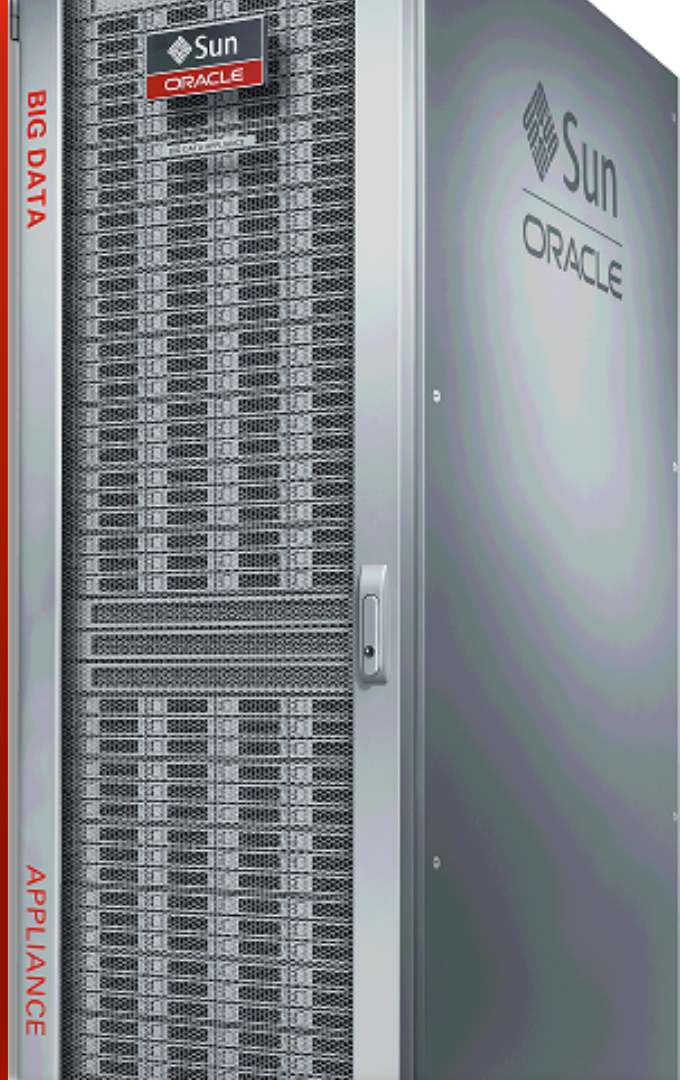
ORACLE

Hadoop Appliances: Engineered for the Enterprise

Dan McClary
Principal Product Manager, Big Data and Hadoop
@dan_mcclary



The preceding 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.

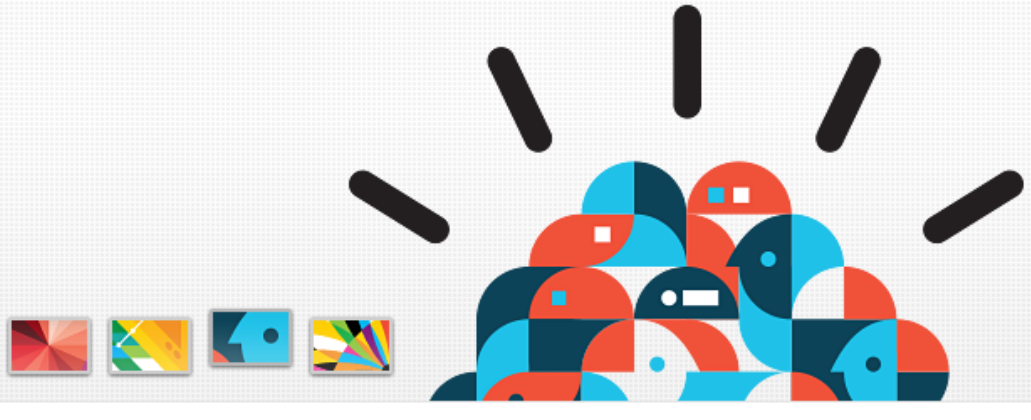
A Smarter Planet

Smarter Planet

IBM SmartCloud

Achieve new levels of innovation and efficiency

→ Learn more



Hardware and Software

ORACLE

Engineered to Work Together

Does that apply to Hadoop?

What Do Enterprises Want?

Benefits of Hadoop

Big Data is meaningful to **everybody**.

Protected Investments

So is everything else in the data center

Minimal Headaches

Focus IT resources on meaningful work

Hardware and Software

ORACLE

Engineered to Work Together

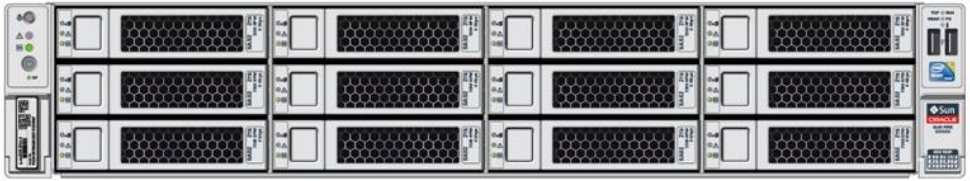
A First Hadoop Cluster

“Commodity” only works in the lab

- Difficult to scale
- Difficult to service
- Only optimizes equipment cost

Start With a Server

- Storage Dense
- Lots of Network
- Enterprise Class



Why Enterprise Disks?

- Imagine a dark, stormy, and I/O-heavy overnight job
  - It's critical to tomorrow's business
- Our cluster has been operational for a long time
  - A couple of disks fail

- Nagios wakes **you** up

Ain’t Nobody Got Time for That.

Enterprise Disks

- Enterprise disks monitor block corruption
  - Predicatively flag failing disks
- Your cluster files an automated service request
  - Oracle comes and changes the drives

- You sleep in

OK, I'll just buy enterprise servers.

Not so fast. What about ·

- Power
- Networking
- Operating System
- **Redundancy**

What “Availability” Really Means

Redundant Everywhere

Automated Failover

You’ll need two of everything

- OS Disks
- Network Switches
- PDUs and PSUs
- Bonded Cables

And it must failover seamlessly

And send a service request

What goes into 2 of everything?

Let's Look at the OS

- Before the factory
  - Install Linux
  - Setup RAID
  - Cut an image
- At the factory
  - For each server in a rack
    - Install the image
    - Rigorously test the server

So what, it's just a Linux install.

Hadoop Ate My Heap

More than just an install

- We need to support real workloads
- On a default Linux 5 install, we set up a test
  - Slightly oversubscribed cluster
  - Memory-heavy jobs
- Servers started falling over
  - There was plenty of free memory
  - Hadoop jobs couldn't get any heap
- **Why?**

Don’t Forget the File Cache

More than just an install

- Very fast network interfaces → fast allocation
- kswapd wasn’t freeing enough memory
- Defaults couldn’t keep up with requests

- How familiar are you with `/proc/sys/vm`?

Let's Look at the Network

- Three networks
  - Admin (1GbE)
  - Client (10GbE)
  - Internal (40Gb3)
- Automated install
  - Given a set of IPs
    - Setup TCP/IP/Infiniband, VNICs, host mappings
- Ensure Hadoop can multihome
  - Work with the community to make it a requirement

Hardware and Software

ORACLE

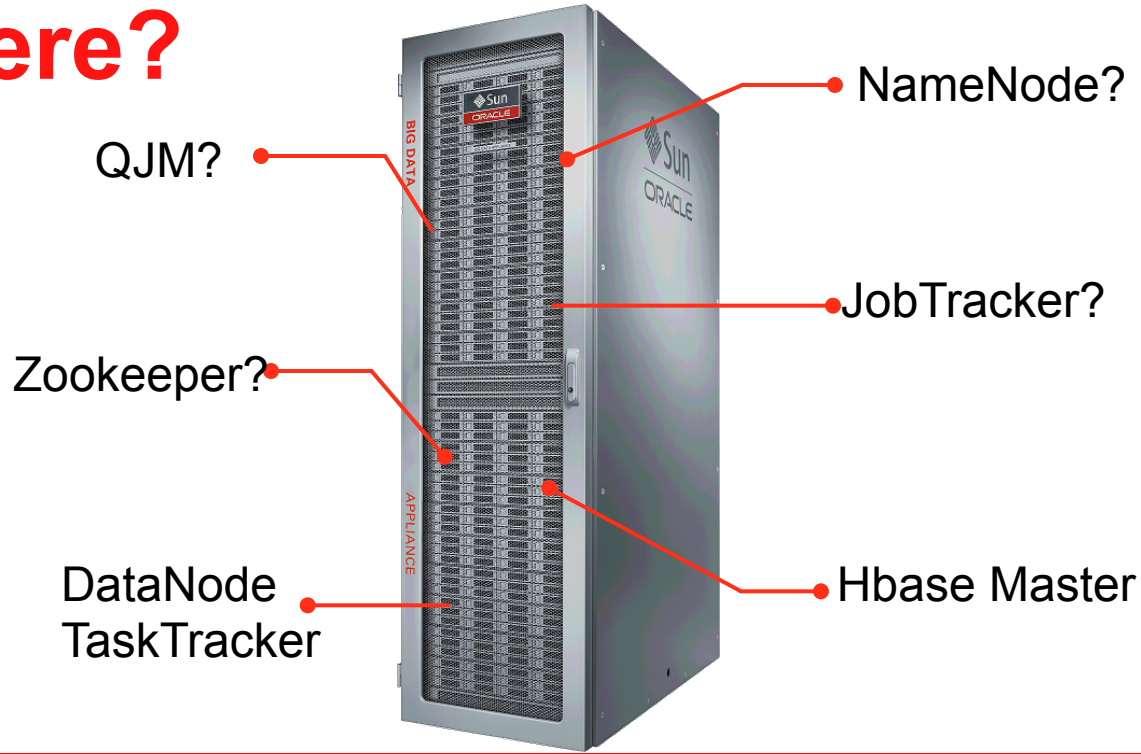
Engineered to Work Together

After the Hardware

What Goes Where?

Don’t Forget

- Multiple racks/cluster
- Multiple clusters/rack
- Clusters grow over time



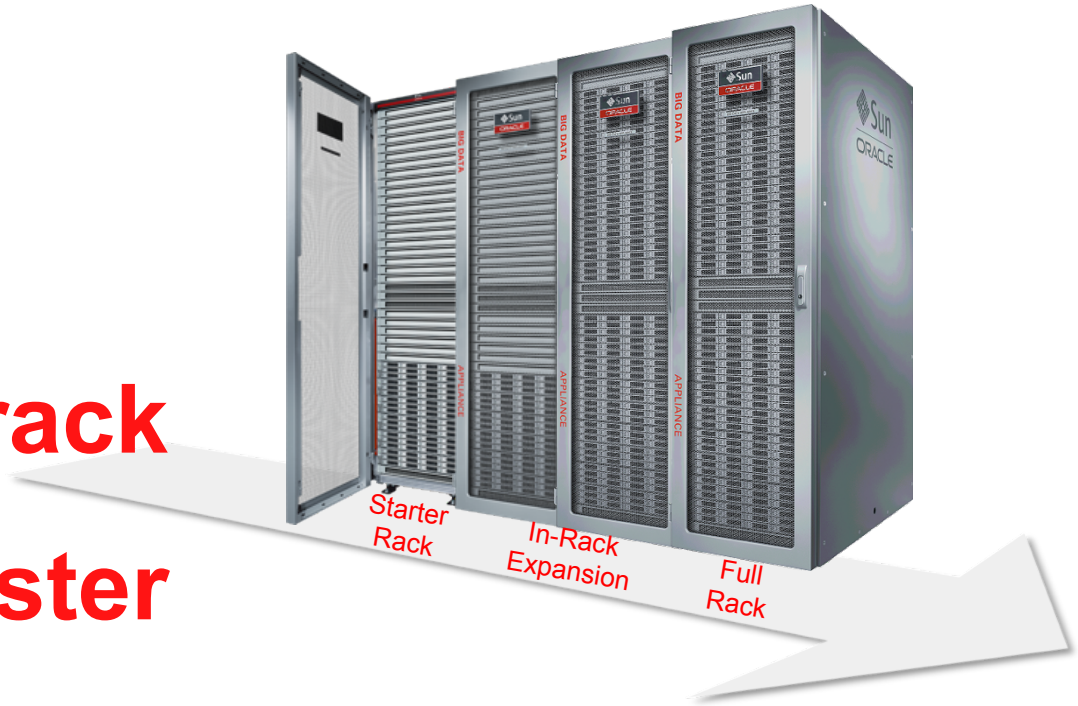
Assumed Growth Path

Some start small

Some start large

Multiple clusters/rack

Multiple racks/cluster



Specialization for Flexibility

How We Expand

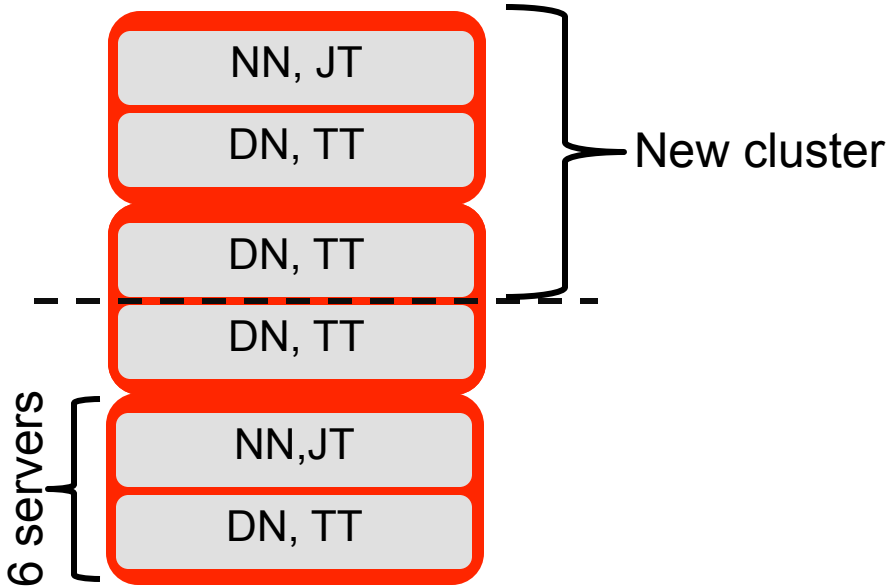
- Minimize role redefinition to simplify growth
- On failure, “general” nodes can be made “special”

4 Nodes of “Management”

DataNodes, TaskTrackers, RegionServers, etc.

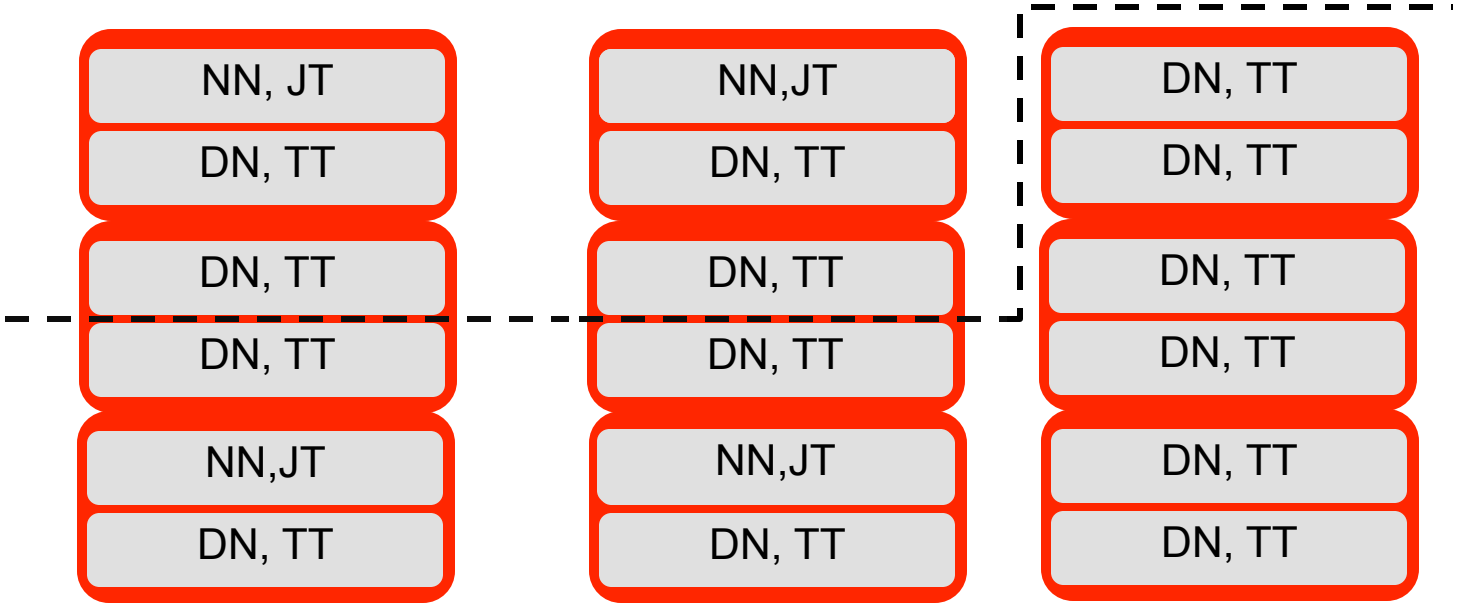
How We Expand

Grow rack



How We Expand

Add a rack and balance HA services



Making Expansion Simple

- First, define roles in Puppet
  - Clear definition of server roles
  - Allows for new roles over time
- Next, add HA where there wasn't any
  - MySQL
  - Internal KDC
- But distributing bits is the easy part

Rules for Enterprise Upgrades

“One-click” upgrades

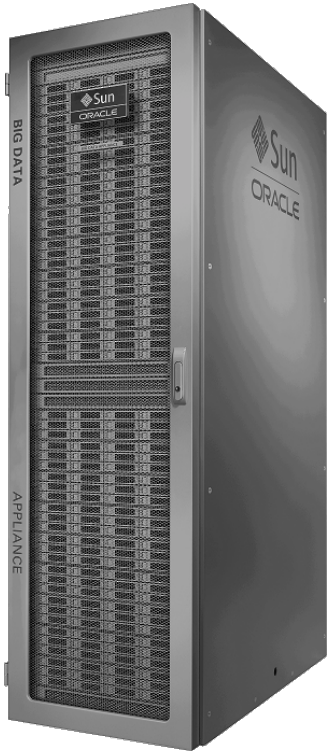
Multiple histories

A patch or upgrade needs to
- Add new features
- Upgrade applications
- Upgrade OS
- Upgrade firmware

And not all customers will be at the same starting point

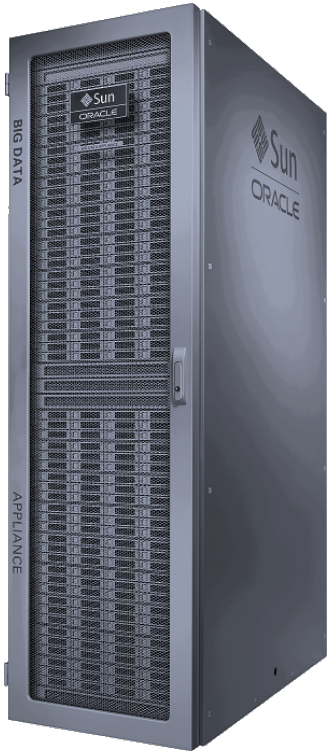
And topologies will change

Different Starting Points



Customer A

- CDH 4.1
- No HA Services



Customer B

- CDH 4.3
- HA NameNode

One Outcome

Customer A
- CDH 4.1
- No HA Services
- 3 Mgmt Nodes

Customer B
- CDH 4.3
- HA NameNode
- 4 Mgmt Nodes

Final State
- CDH 4.4
- HA NameNode
- HA JobTracker
- Kerberos Enabled
- 4 Mgmt Nodes

Scripted Upgrades

- Discover cluster state
  - Compare end state to cluster state
  - Determine extra operations
    - Add NameNode and JobTracker
    - Setup Failover
    - Install KDCs, Setup keytabs
    - Enable Kerberos
- Install upgraded packages
- Configure and (re)start services

Hardware and Software

ORACLE

Engineered to Work Together

Beyond Hadoop

What Do Enterprises Want?

Benefits of Hadoop ✓

Protected Investments How do we integrate Hadoop with all the things?

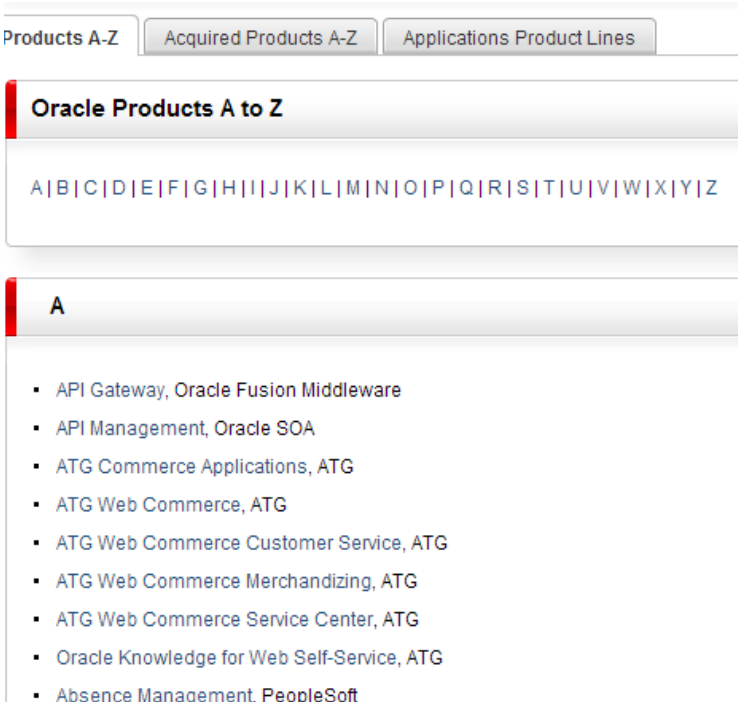
Minimal Headaches ✓

Integrate Everything

We make a lot of stuff

918 products

How does Hadoop fit?



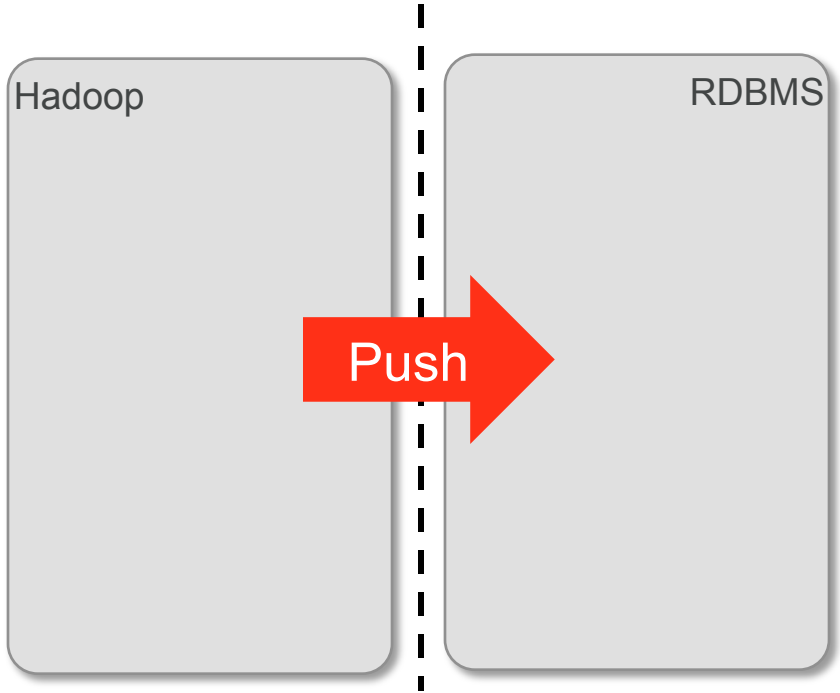
So where do we start?

Maybe here

Load to Oracle

Hadoop produces dataset

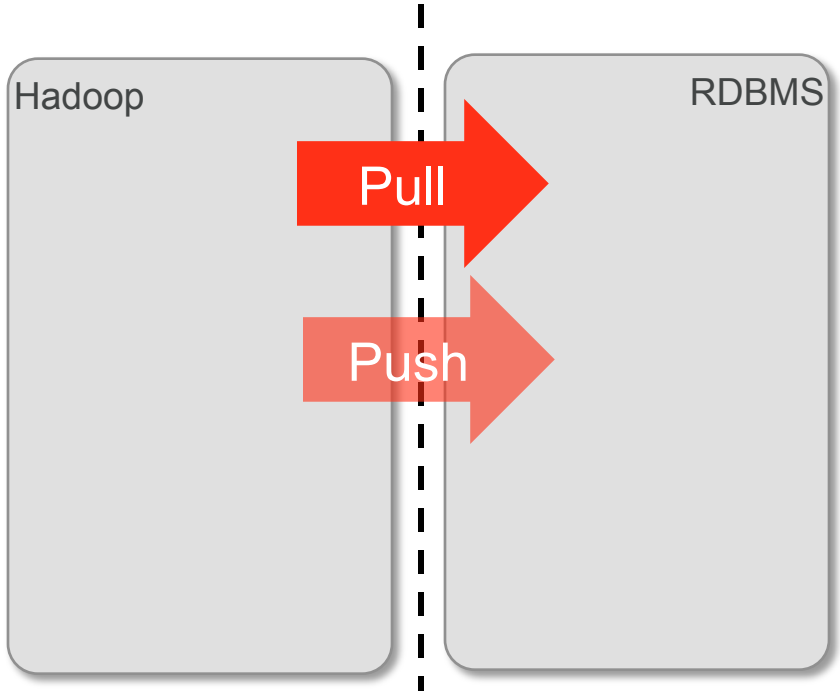
Oracle publishes dataset



Stream from Hadoop

Hadoop for big analytics

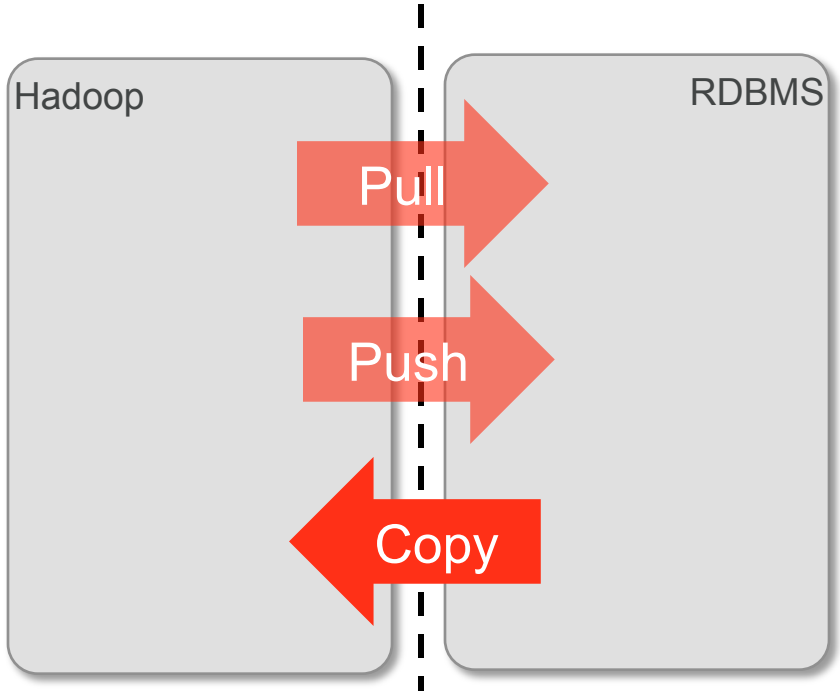
Oracle for sensitive data



Copy to Hadoop

Oracle for mission critical

Hadoop for unified data



Even More Integrations

Fast Data

- Oracle Event Processing

Data Integration

- Oracle Data Integrator
- Oracle Golden Gate

Security

- Oracle Audit Vault and Database Firewall

Hardware and Software

ORACLE

Engineered to Work Together

Q&A

Hardware and Software

ORACLE

Engineered to Work Together


ORACLE®

Hardware and Software

ORACLE

Engineered to Work Together

Optimized and Flexible

- Optimized roles for Hadoop usage
- Simplified elasticity and management

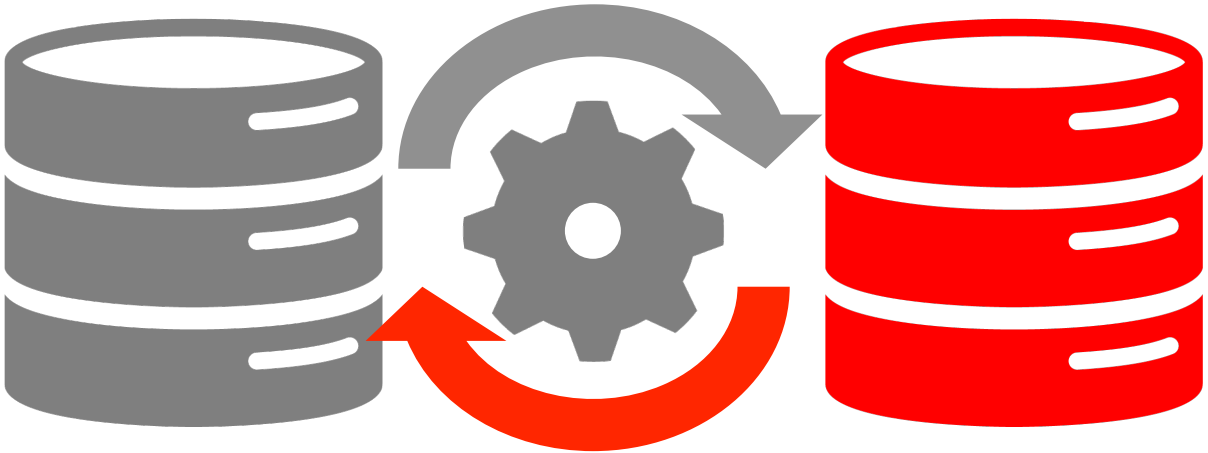
Hardware and Software

ORACLE

Engineered to Work Together

Integrated With Everything

- Integration with core products
- More integrations every day



Hardware and Software

ORACLE

Engineered to Work Together

We're not marketing geniuses

Hardware and Software

ORACLE

Engineered to Work Together

Does that apply to Hadoop?