Merkle Big Data Journey

March 2016

Who we are – Our Brand

Our Work

Data-Driven, Technology-Enabled Performance Marketing

Our Services

Strategy | Analytics | Technology | Data | Creative | Media

Our Value

Acquiring, Retaining & Maximizing The Value Of Your Most Profitable Customers

Our Differentiation

Our heritage in data, technology and analytics gives us the ability to gain unique insights that when combined with our strength in performance media, fuels our ability to develop and execute addressable experiences that drive customer engagement, loyalty and shareholder value.

Our Passion

Helping the best brands in the world create competitive advantage through people-based marketing.

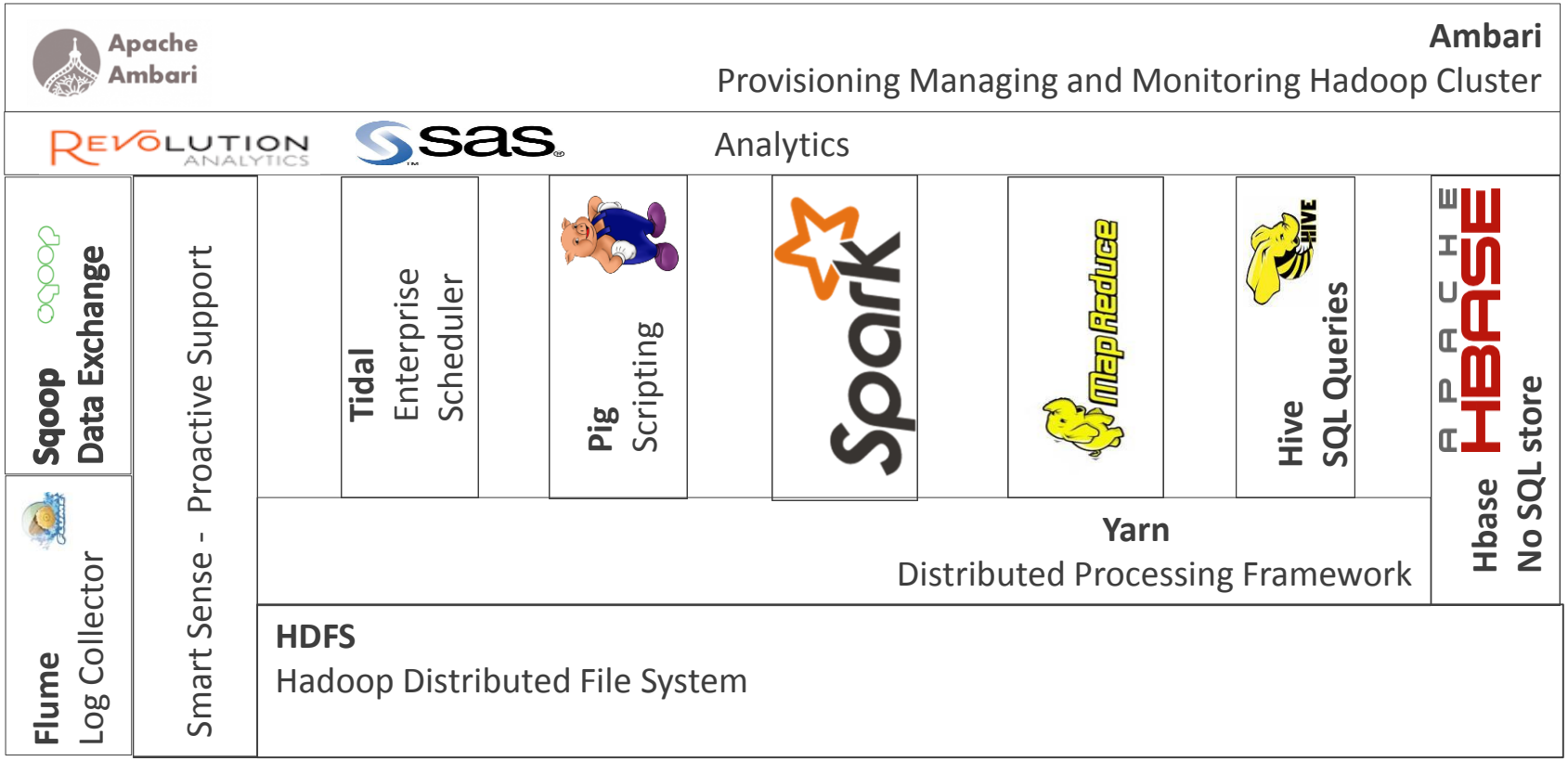
Hadoop Adoption Facts

- Survey performed with 284 companies in May 2015
  - 26% of companies are deploying or experimenting with Hadoop
  - 11% plan to invest within 12 months
  - 7% will invest within 2 years

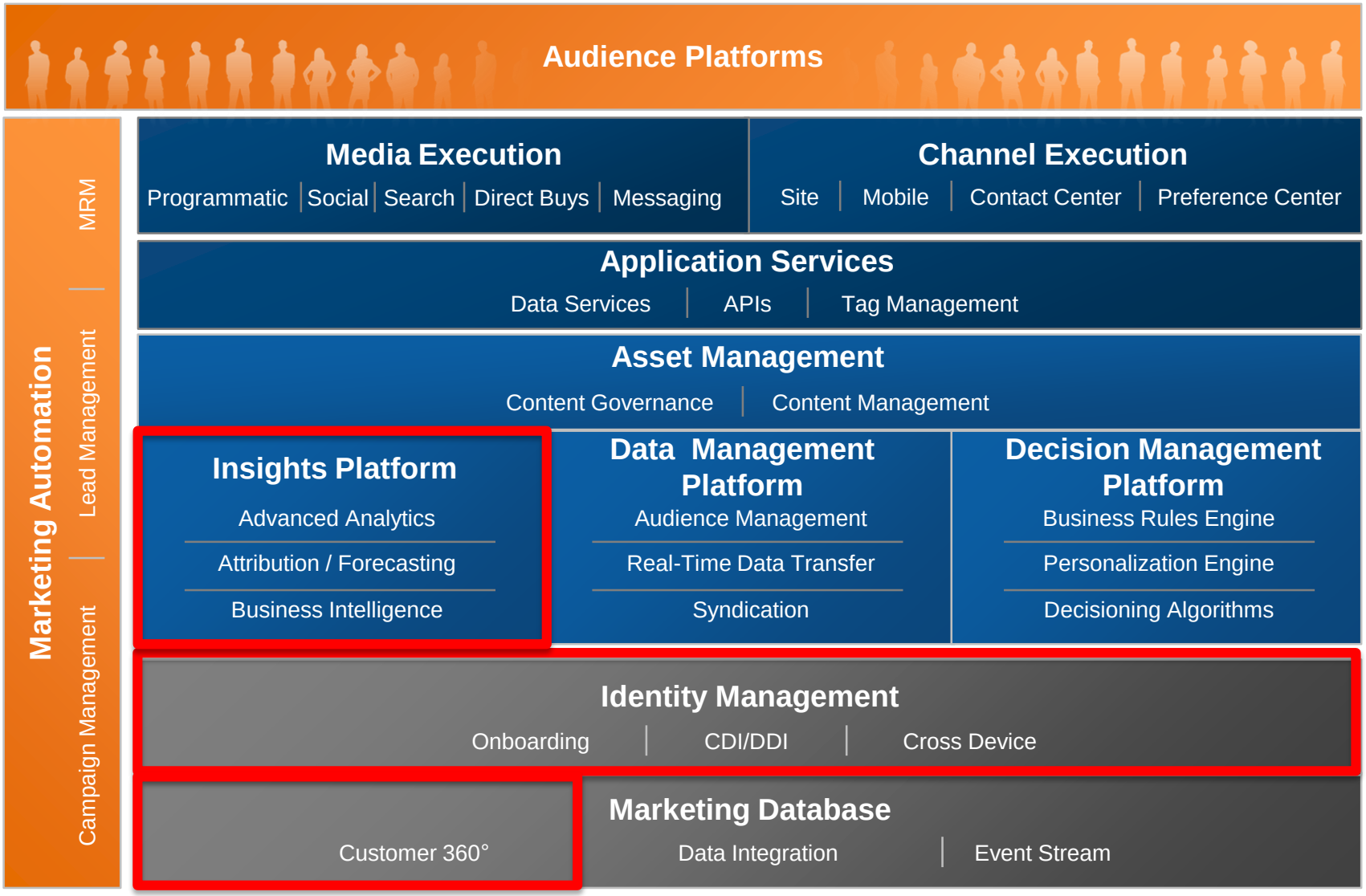
- Adoption inhibitors
  - Skills Gap – 57%
  - Determining how to get value from Hadoop – 49%

Gartner Survey Highlights Challenges to Hadoop Adoption (2015)

Merkle Hadoop Ecosystem



Merkle Connected CRM Framework



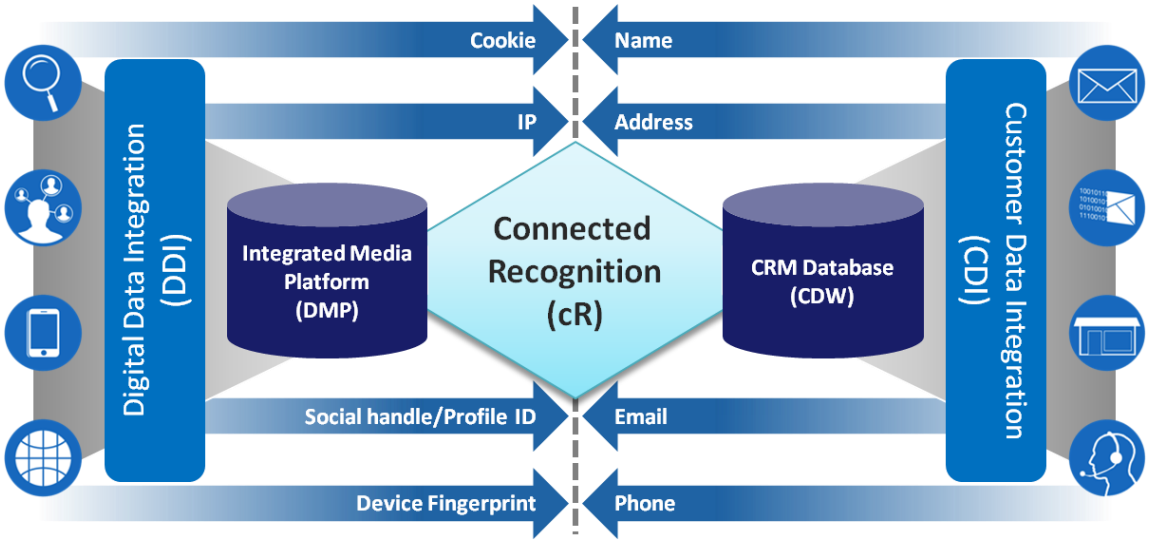
Identity Management

Key Business Benefits:

- Configurable Customer Data Integration (CDI) rules
- Digital Data Integration (DDI) processes tags, cookie, mobile IDs, and social handles
- Session and Event Level aggregates and processes – Rationalizes the sea of digital data into meaningful information linked to your customer database.
- Global Address Standardization & Hygiene – currently supports more than 250 countries today including Unicode processes

Advanced Technology:

- Merkle's fastest and most advanced CDI engine
- Expandable DDI architecture can process a wide array of event types and identity links
- Gateway to Omni-Channel Data Enhancement and Match Services
- Supports DMP integration and on-boarding



Insights Platform components

Sources & Connectors

- APIs and Adapters
- CRM DB
- Media Data
- 3rd party data
- Transaction data

Data Constructs

Reporting Marts

Analytical Marts & Cubes

Big Data Environment

Insights Portal

Portal Interface

Standard & Custom Dashboards

Ad-hoc Reporting

Module Interfaces

Retail *IWM* *B2B* *Banking* *TME* *Life Sciences* *Non Profit*

Insights Engines

Attribution Engine

$$y_t = \alpha + \sum_{i=1}^{p} \beta_i x_{it}^{p_i} + \varepsilon_t$$

Scenario Planning & Optimization

Alerts

Profile Generator

Insights Playground

Detailed Data

Tools

Advanced Analytics

- Predictive modeling
- Clustering
- Simulation

MERKLE

Benefits of Big Data

- Performance – 10x performance improvement
- Cost – 60% TCO reduction
- Scalability – 80% reduction in time to scale

We won the fight, but it wasn't pretty...

Big Data Challenges

- Security
- Architecture
- Keeping pace with the rate of change
- Talent
- Managing workload/stability

Security Challenge

- **Assumption** – Security was built-in like any modern platform/application
- **Lesson Learned** – Minimal security by default and no planning for multi-tenancy

Security Challenge

Out of the box, Hadoop has no native authentication controls and does not challenge user identities

Example

- **Shawn to Hadoop** - Hello Hadoop – I'm Tim Talbot, please delete all of the data
- **Hadoop to Shawn** – Hello Tim, all data has been deleted

Security Challenge

Kerberos Enables Authentication

- **Shawn to Hadoop** - Hello Hadoop – I'm Tim Talbot, please delete all of the data
- **Hadoop to Shawn** – Hello Tim, I'm checking my Kerberos server to validate your credentials
- **Hadoop to Shawn** – You are not Tim Talbot - request denied

Security Challenge

- Multi-tenant challenge
  - Lack of metadata restriction
- Application Security Limitations
  - Some applications require access to OS level commands
    - Allow users to run netstat, ping, host commands, etc.
  - Some applications are required to run as a single common system account and do not support user impersonation

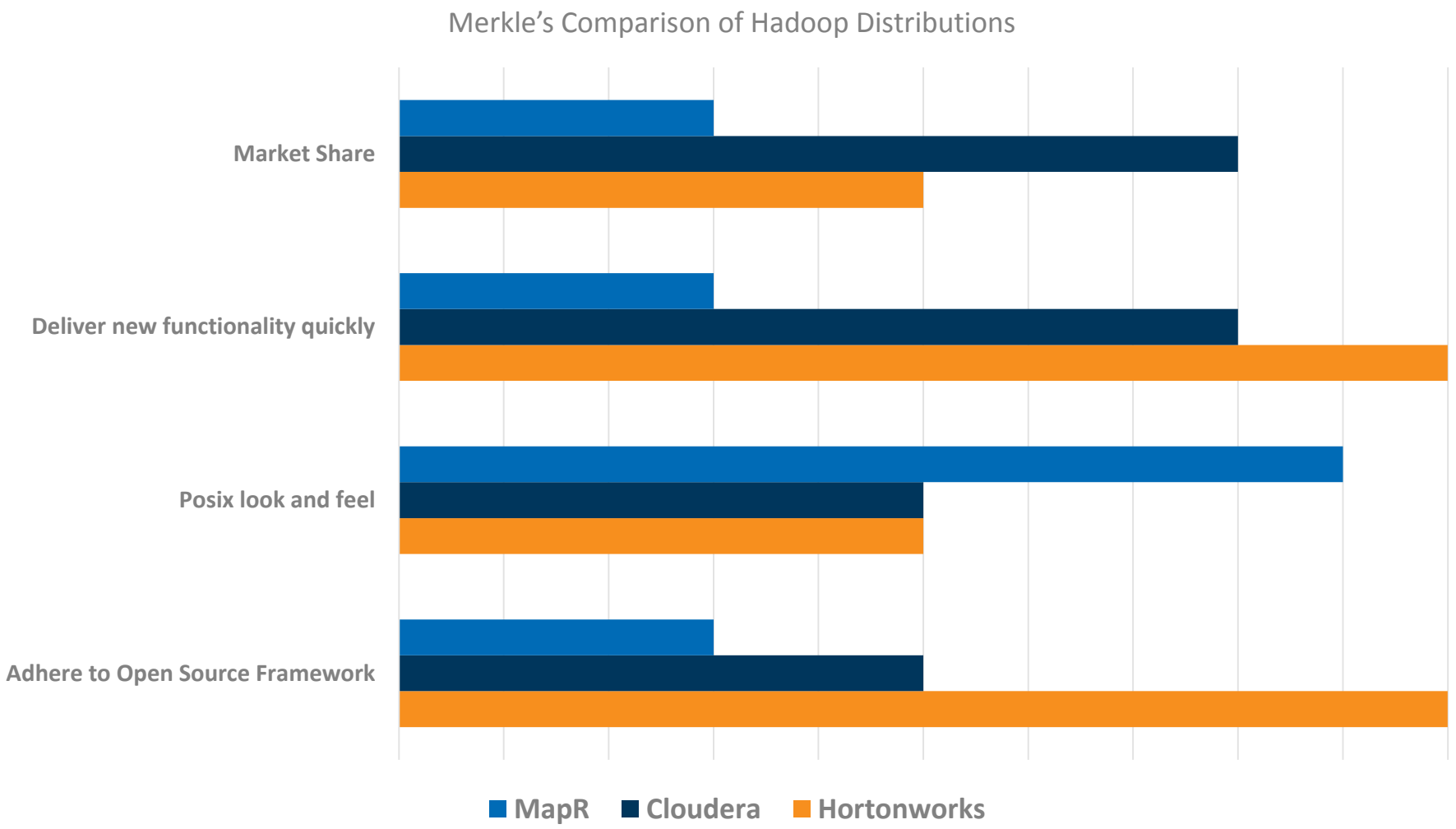
Architecture Challenge

- **Assumption** – Focus on performance when architecting the solution
- **Lesson Learned** – Performance is one area to evaluate, but there are many

Architectural Challenge

- Establish the fundamental use case and stick to it
  - Define methods of user access and data manipulation
  - Clearly Define Use Cases
    - Multi-tenant vs. single workload
    - Batch vs transactional
    - Random user generated queries
  - Workload Evaluation
    - Size of incoming data
    - Execution timeframe requirements
    - Dynamic nature of workloads
    - Concurrency
    - Direct query access

Architectural Challenge



Architectural Challenge

- Thoroughly evaluate the integration/segregation of Compute and Storage
  - Integration of compute and storage
    - Data locality - faster processing
  - Segregate of compute and storage
    - More efficient use of resources
    - Release management flexibility
- Cloud and/or On-prem deployment
- Hadoop is not the answer to every question
  - Transactional workloads with millisecond response times
  - Small data

Rate of Change Challenge

- Assumption –
  - We will manage Hadoop like any other platform and update when needed
- Lesson Learned –
  - Hadoop is an evolving platform with accelerating changes – changing faster than anything that we have experienced

Rate of Change Challenge

- Establish a governance model that can set the direction for the platform
  - Include architects, developers, administrators and business owners

  <u>High Level topics</u>
  - Capacity
  - Incidents
  - Roadmap

- Pick a release schedule and stick to it
  - With new releases every month it is easy to get half way finished evaluating a release and then start looking at the new one

Talent Challenge

- Assumption – We would leverage vendor for consulting and train our staff
- Lesson Learned – It's both an acquisition and retention issue

Talent Challenge

- Acquisition
  - Use a consultant for the first install and train staff
- Retention
  - Hadoop administrators are in strong demand and are likely being recruited
- Our approach to both acquisition and retention is to train all tiers of support on Hadoop – make administration and development a core competency

Managing Workload Challenge

- Assumption –
  - Hadoop would manage workload through the default scheduler
  - We could determine cause and effect of configuration changes on existing workload

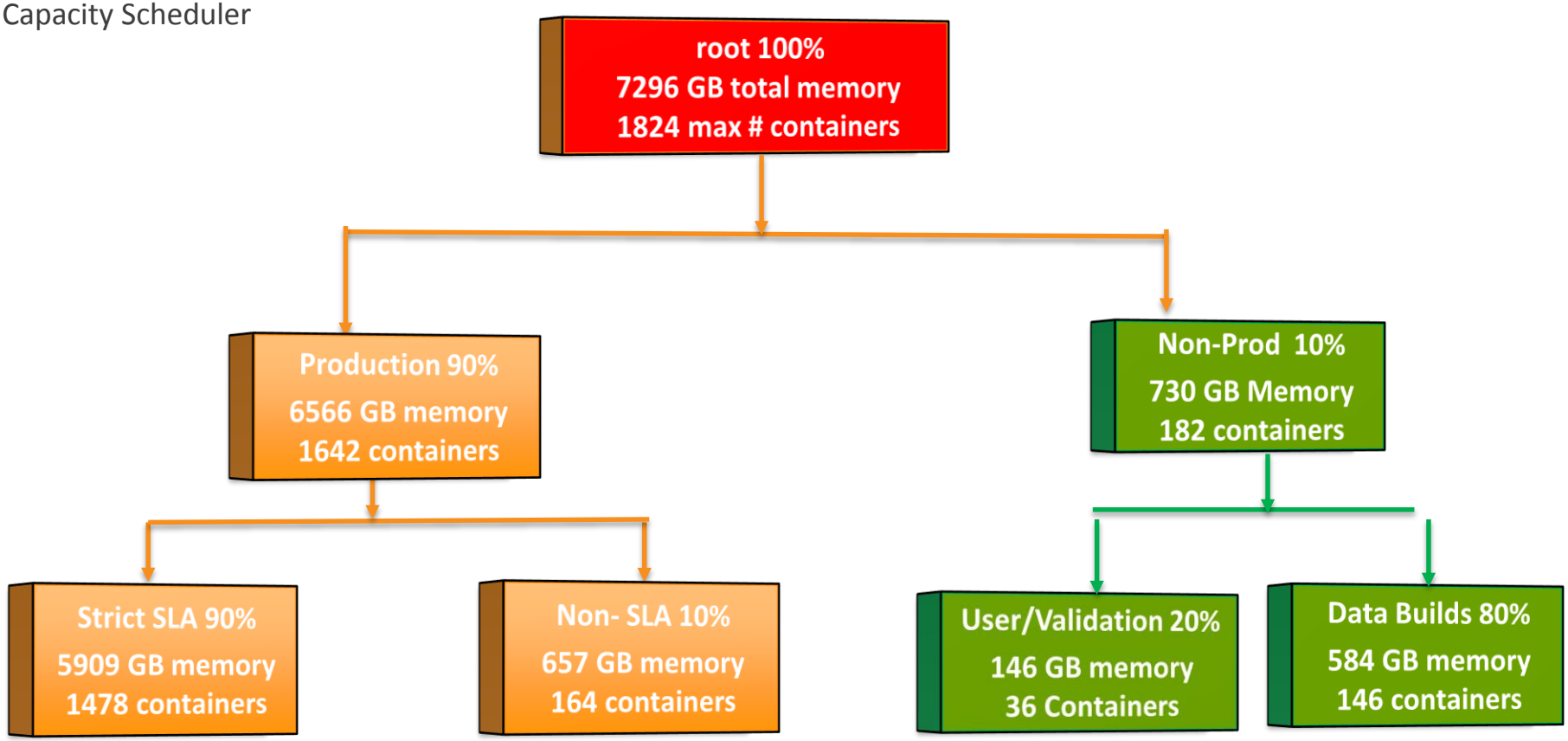
- Lesson Learned –
  - Establish a role that is responsible for performance and capacity management. A comprehensive capacity management plan must be developed early and routinely evaluated throughout the lifecycle
  - There are an infinite number of configuration variables. Have to make a change and evaluate.

Managing Workload Challenge

- Requires management of resources at a very granular level
  - CPU – C groups
  - Memory - Yarn Queues
  - Storage – HDFS quotas
  - JVM management
  - I/O management
- Manage background processes
  - Hbase compaction
- Maintain solution ratios
  - CPU to spindle ratio
  - Mapper to reducer ratio
  - Hbase region file – region server

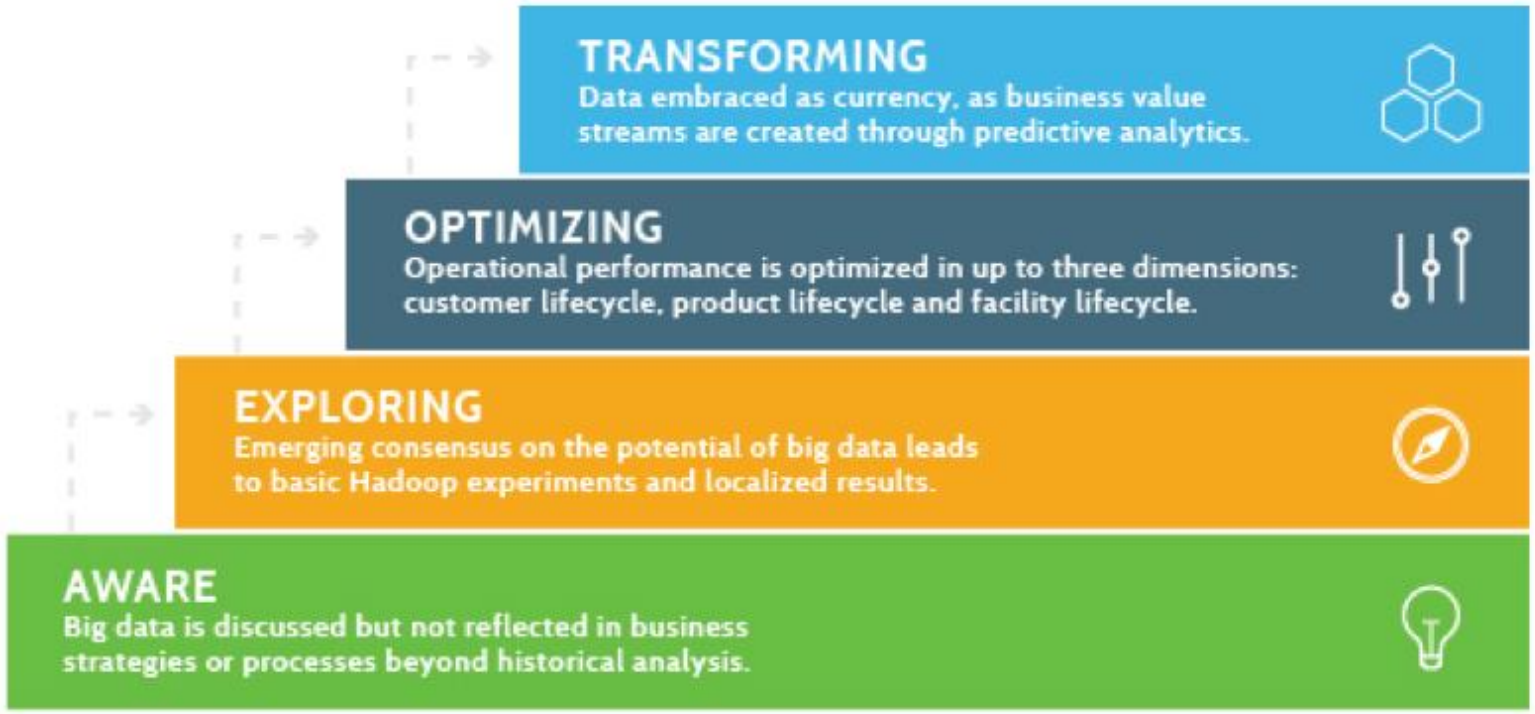
Managing Workload Challenge

Capacity Scheduler



Getting Started with Big Data

Hortonworks Big Data Maturity Model



Starting your Big Data Journey

Thank You!

merkleinc.com