



EMC Big Data Strategy

George Demarest

Senior Director CIO and Big Data Marketing
EMC Corporation

EMC²

The Digital Universe



Data Volumes

2000: 2 Exabytes a year
2011: 2 Exabytes a day



By 2020, the Internet will
connect **7.6B** people
and **200B** things
(sensors, machines, cars, appliances...)



Less than **1%** of
the World's Data
is Analyzed



EMC²

“Software is Eating the World”



Software is Changing Industries



\$3.5B valuation
Financial Services



\$3.5B valuation
Transportation



\$20B valuation
Entertainment



\$3.5B valuation
Travel & Hospitality



\$3.2B Acquisition by Google
Home Automation



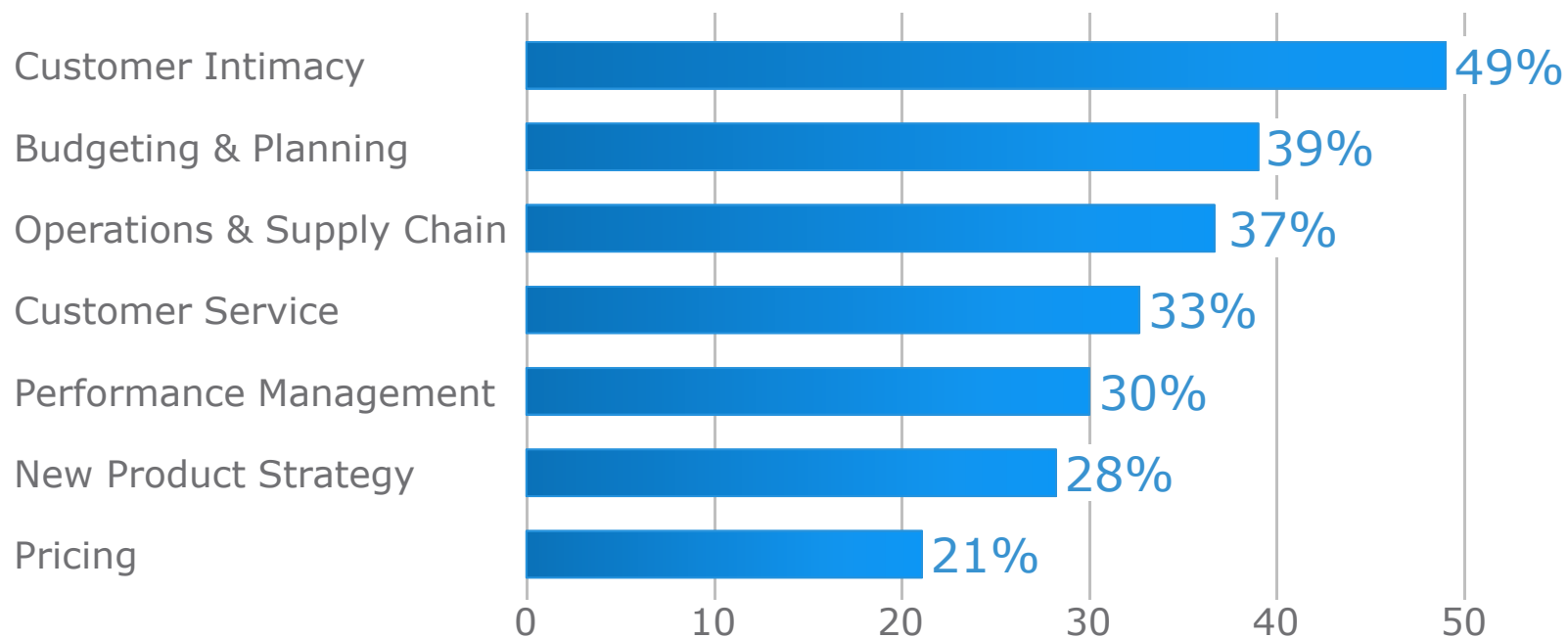
\$1.1B acquisition
Agriculture

Location & Types Of Big (& Fast!) Data



How Companies Are Using Big Data Analytics

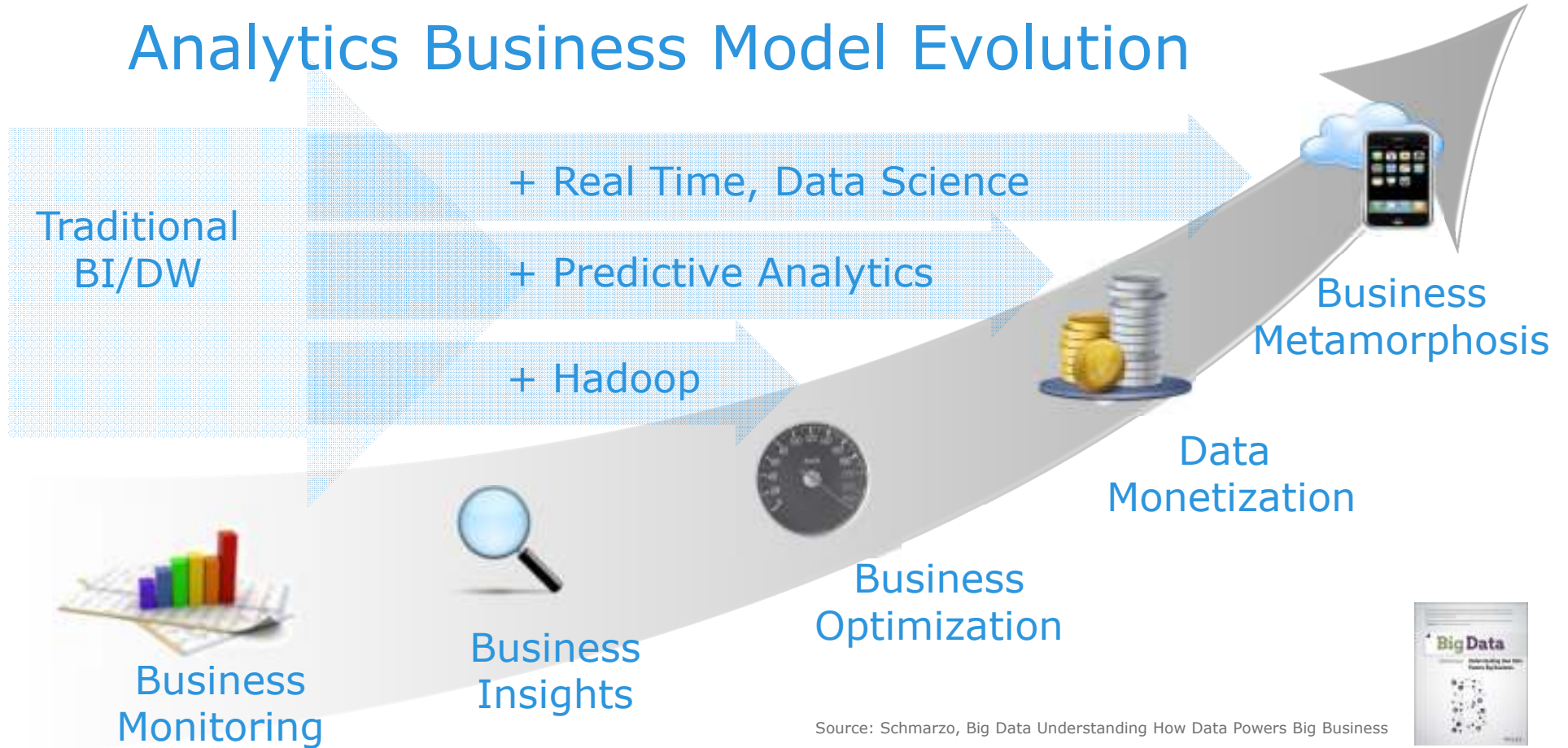
Functional Areas Where Companies Are Using Big Data



McKinsey Global Survey of 1,469 C-level executive respondents at a range of industries and company sizes, "Minding Your Digital Business," 2012.

EMC²

Analytics Business Model Evolution



Source: Schmarzo, Big Data Understanding How Data Powers Big Business

EMC²

The Rise of Data Science

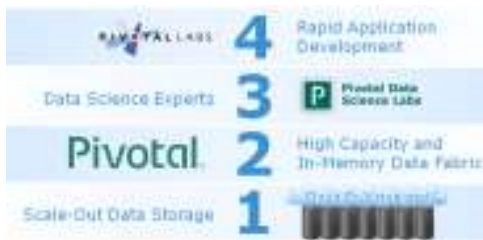


Data Scientist:
The Sexiest Job of the 21st Century



"...continued
shortage of Data
Scientists"

EMC in the Big Data Landscape



Product/Service Portfolio



Thought Leadership and Real World Implementation



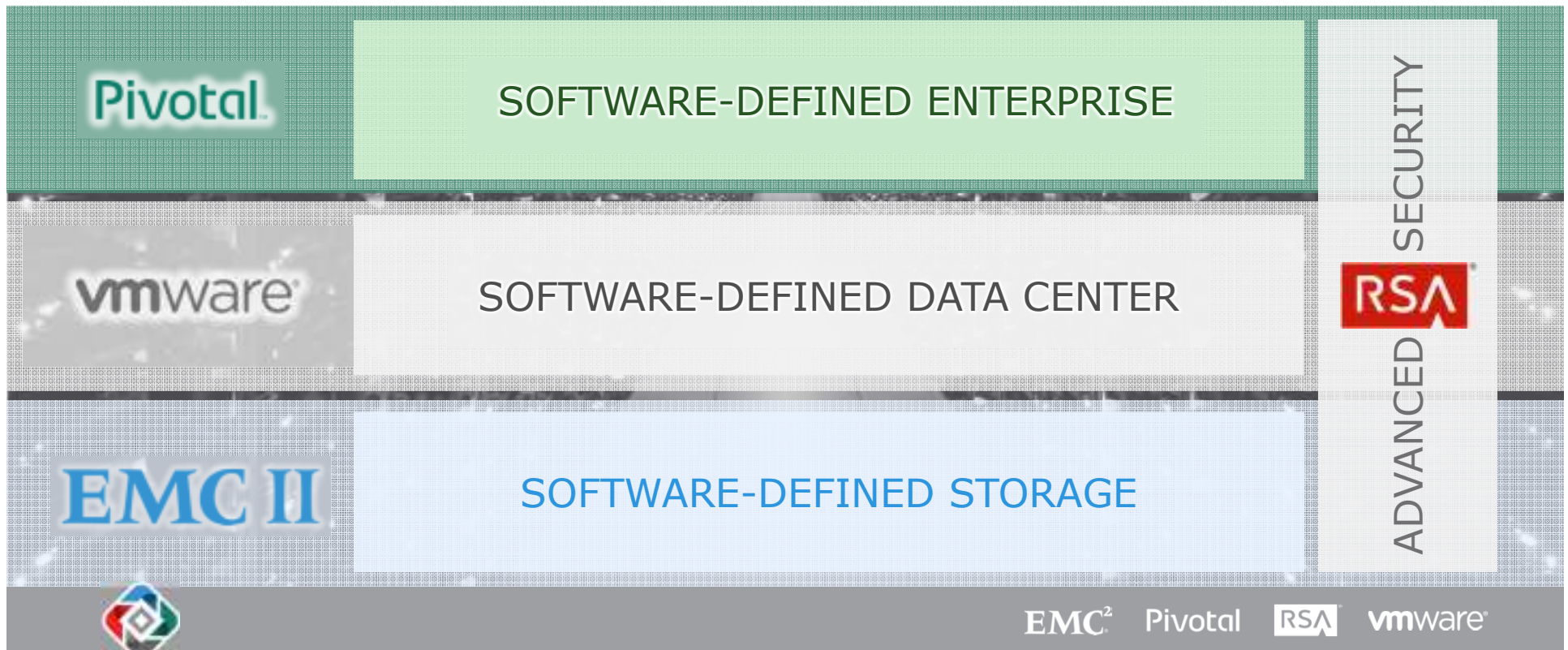
Data Science Curriculum



EMC is a Big Data Driven Business

A Unique Federation Of Companies

Best Of Breed, Architected Horizontally, Unparalleled Choice



EMC Federation Big Data Ecosystem

Pivotal



- Multi-cloud PaaS
- Enterprise Class Hadoop/Analytics
- Agile Development

vmware



- Rapid Provisioning
- Elastic Scale
- Optimal Utilization
- Pervasive Automation

EMC II



- Industry Leading Storage Portfolio
- Software-defined Storage
- Scale-up, Scale-out Data Services
- Data Protection Architecture

RSA

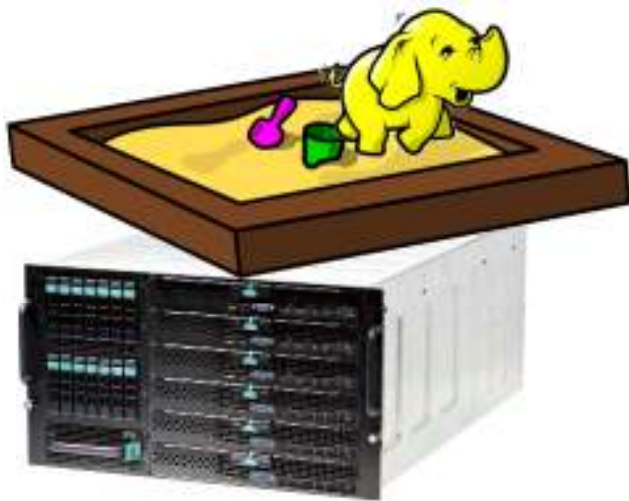


- Risk Intelligence
- Advanced Analytics
- Adaptive Controls
- Intelligence Sharing

EMC²

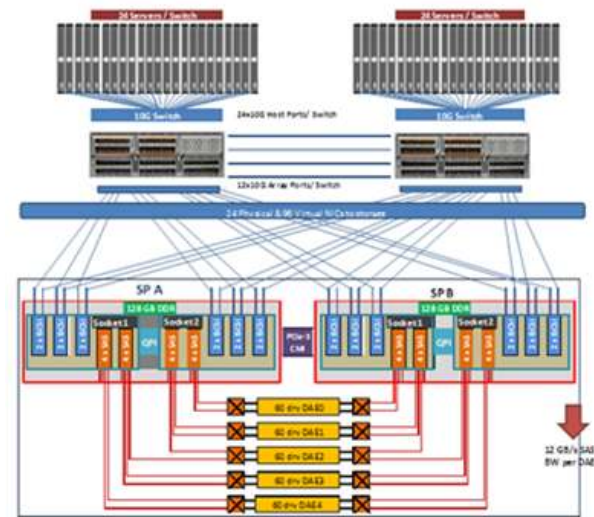
Hadoop: Moving out of the Sandbox

Dev/test sandbox



Commodity server, storage,
community support

Enterprise Production/Mission Critical



Server pool, tiered storage,
commercial support

Mission Critical Hadoop

Early Adopters Pushing the Bounds of the Platform



YAHOO!

Over 40,000
Hadoop Nodes



facebook

21 PB of storage in a
single [HDFS](#) cluster



qu^xntcast

Processes 20 PB/day



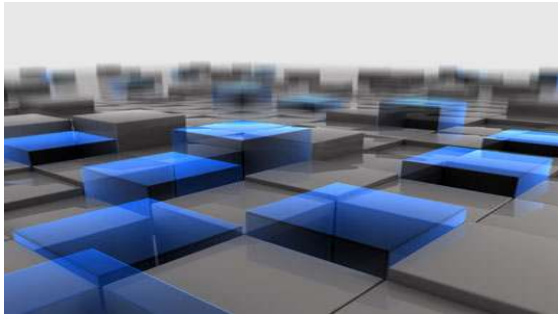
ebay™

40,000+ cores process
350+ million searches/day



EMC²

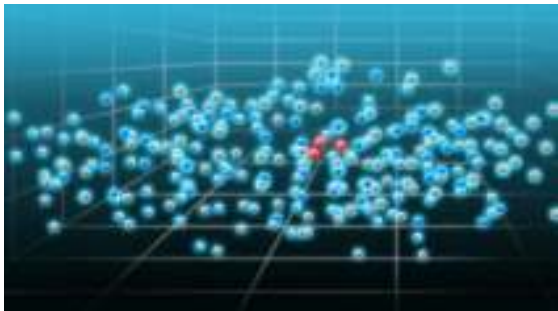
Applying Enterprise Techniques to Hadoop



Virtualization



Compression, deduplication



Advanced Security



Flash

VMware vSphere and Big Data Extensions

Data Managed by Hadoop Needs Private Cloud Operations



**Deploy a Hadoop Cluster
in under 30 Minutes**



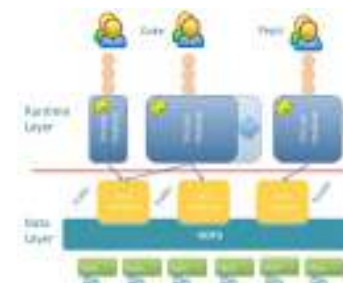
vSphere High Availability (HA)



vSphere Fault Tolerance



Live Machine Migration



VM level Multi-tenancy

EMC ViPR, Software-Defined Storage

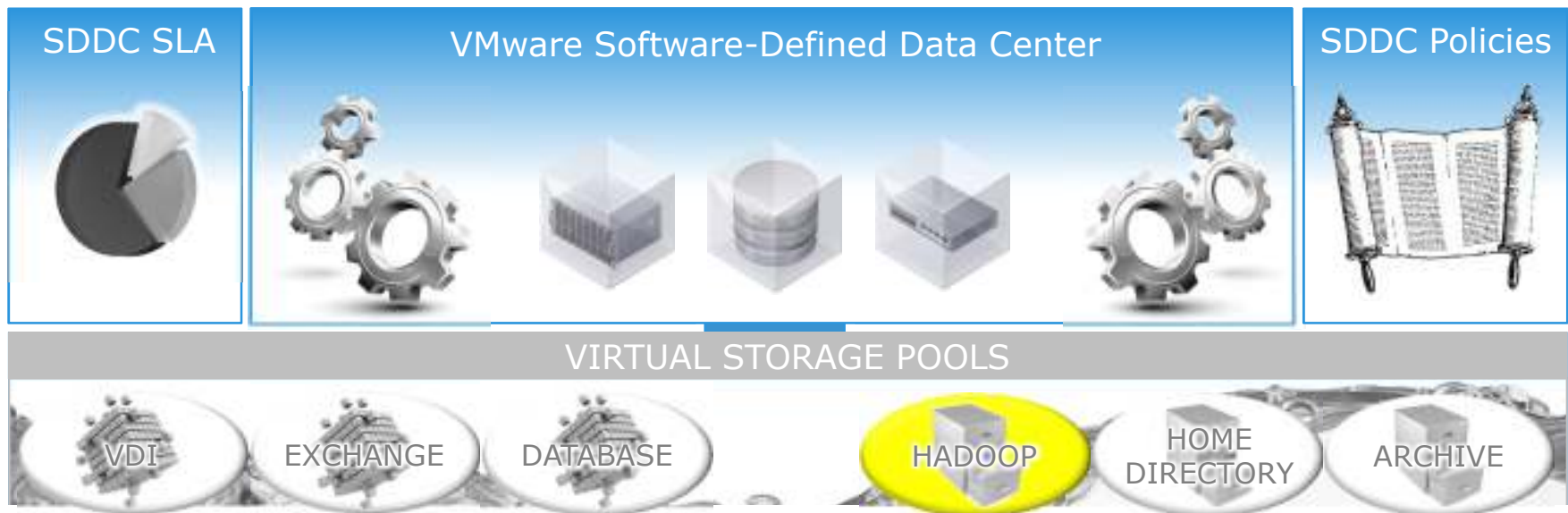
Software Control Plane and Data Plane



EMC²

ViPR Controller : Integrates With VMware

Software-Defined Storage For The Software-Defined Data Center



Isilon: Scale-Out Storage for Big Data



Massive Scalability

20 PB in a single file system

Enterprise Hadoop Ready

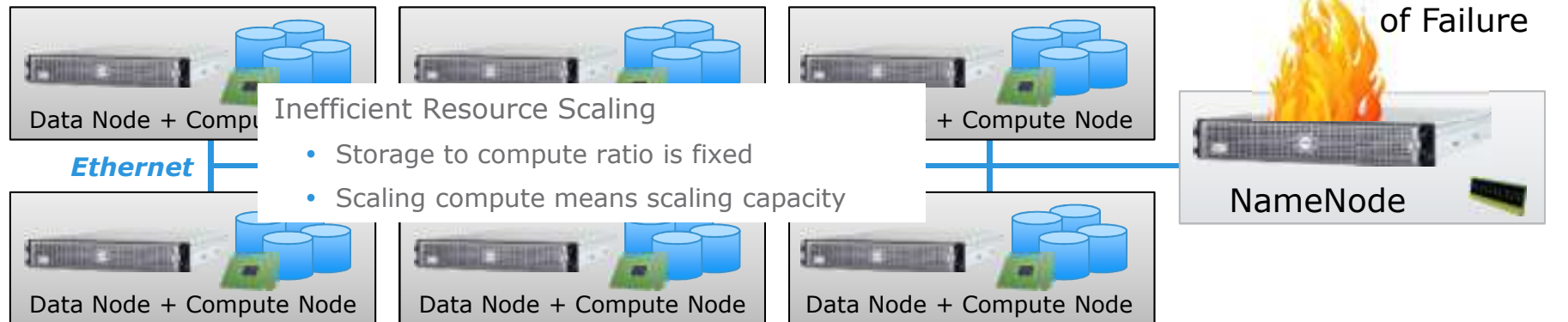
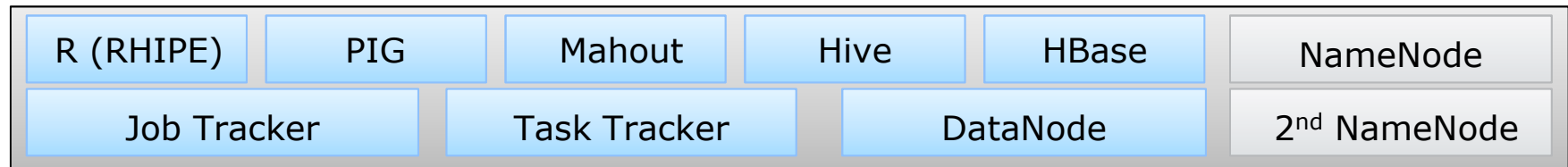
HDFS is built-in to the OneFS operating system

Management Simplicity

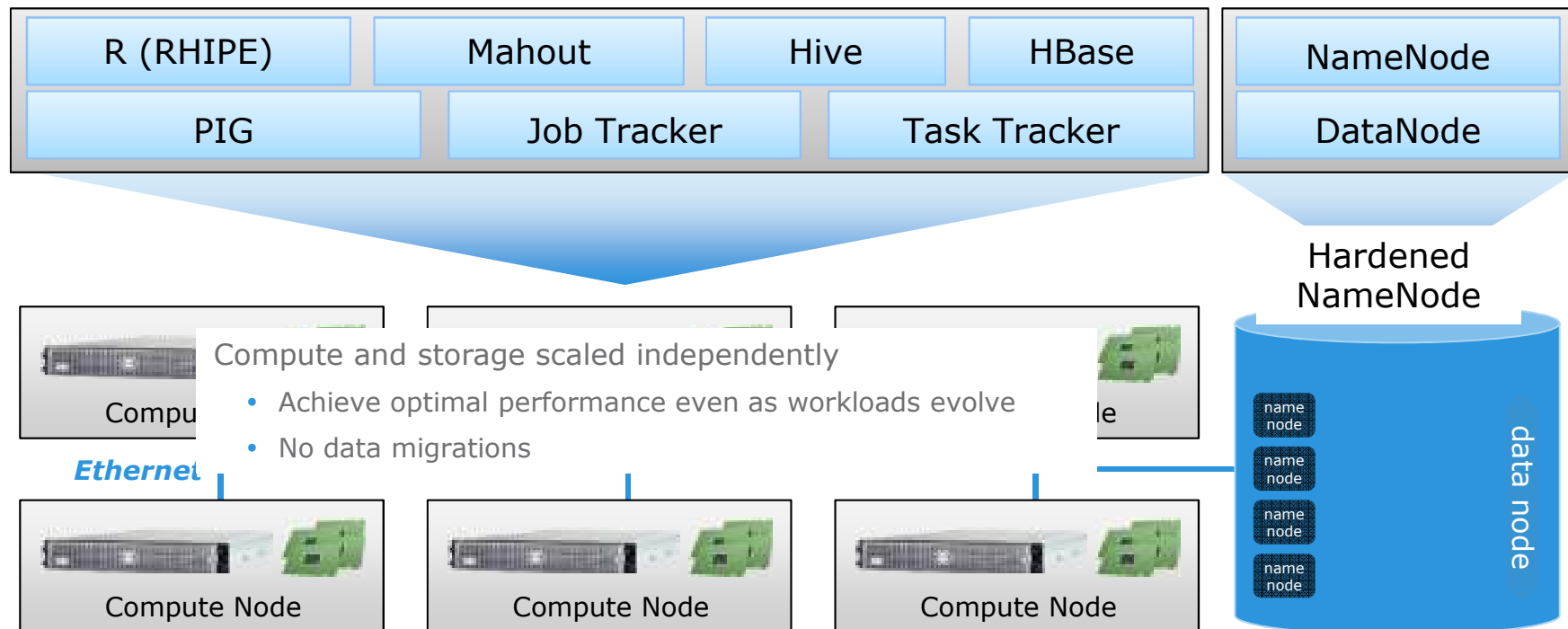
Manage PBs just like you would TBs

EMC²

Traditional Hadoop Architecture



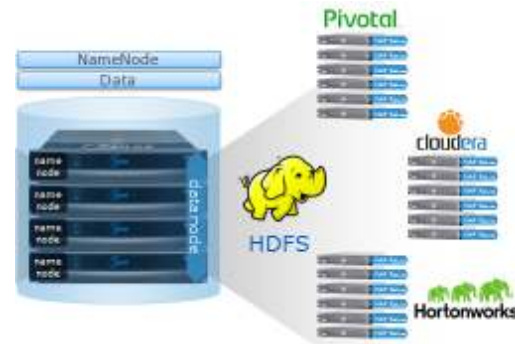
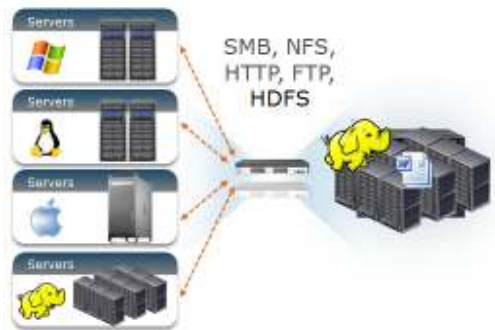
Hadoop Architecture with Isilon



Unique to Isilon Hadoop Deployments

Avoid Big Data Silos

- Native HDFS support
- Native multi-protocol support



No lock-in

- Supports Multiple Distributions

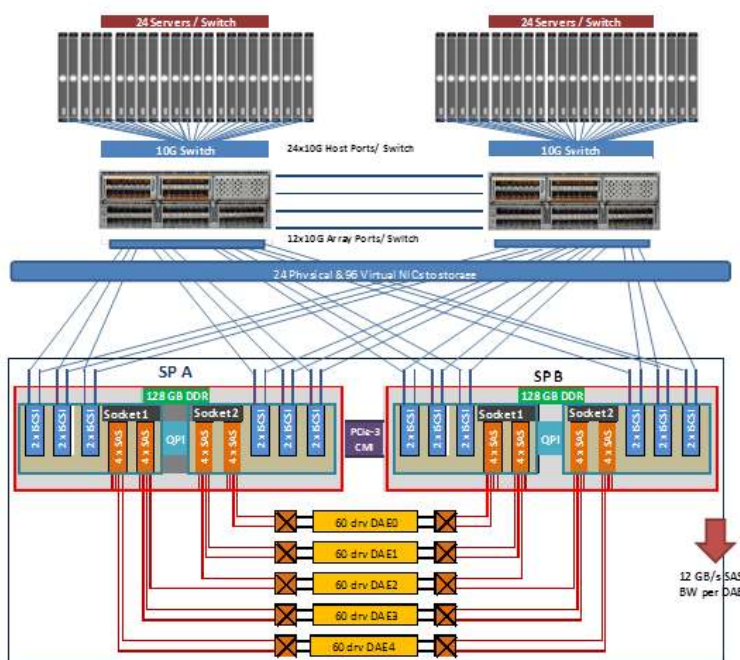


Enterprise Capabilities

- Multi-node replication
- Snapshots
- Security, DR, Backup

Enterprise Hadoop in Production

E-commerce POC Using EMC VNX-8000



Hadoop Implementations Will Multiply and Vary

Enterprise Considerations: Density, Performance, Latency, RAS

- Servers – 48 x Cisco UCS C-240 M3
 - 2 x Intel E5-2665
 - 256 GB Memory
 - 2 x 10 Gbps Ethernet
- Network
 - 2 x Cisco 6296UP Fabric interconnects
 - 6 x Cisco 2232 Fabric extenders
- Storage
 - 1 x EMC VNX2 (Prometheus)
 - 6 x Dual-Port 10 Gbps Ethernet
 - 4 x 6 Gbps SAS
 - 5 x Voyager Dense DAEs
 - 300 x 3 TB NL-SAS drives
 - RAID 6 (10+2 Configuration)

Pivotal

A NEW PLATFORM FOR A NEW ERA

Pivotal

BIG DATA

FAST DATA

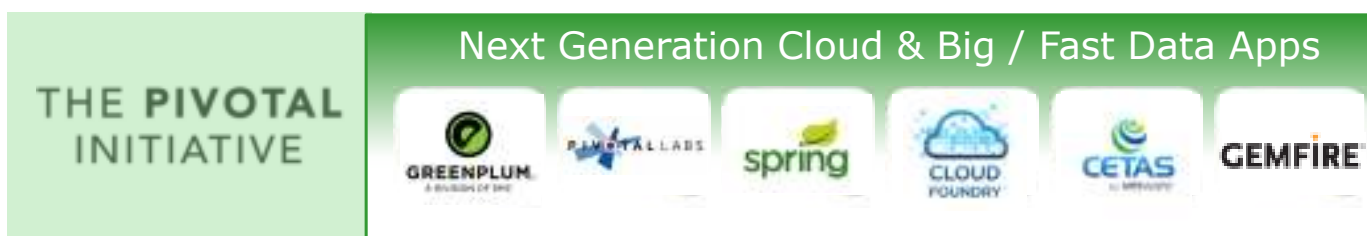
APPS

CLOUD INDEPENDENCE

A NEW PLATFORM **FOR A NEW ERA**

Pivotal

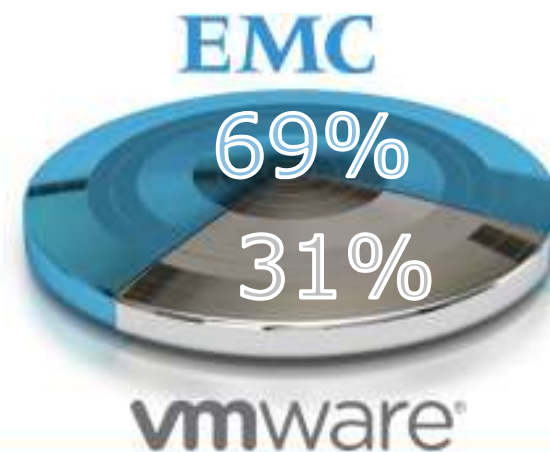
The Pivotal Initiative: 3rd Family Member



1250
EMPLOYEES

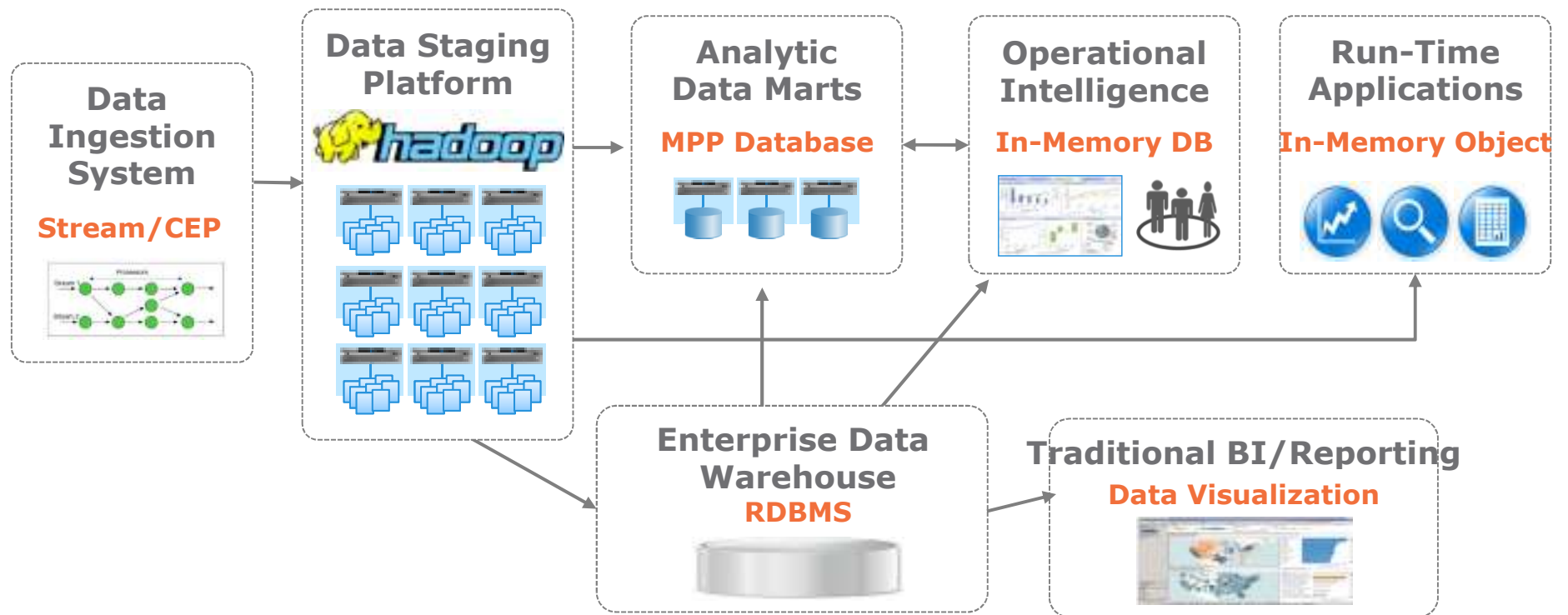


~\$300M
IN 2013 REVENUE

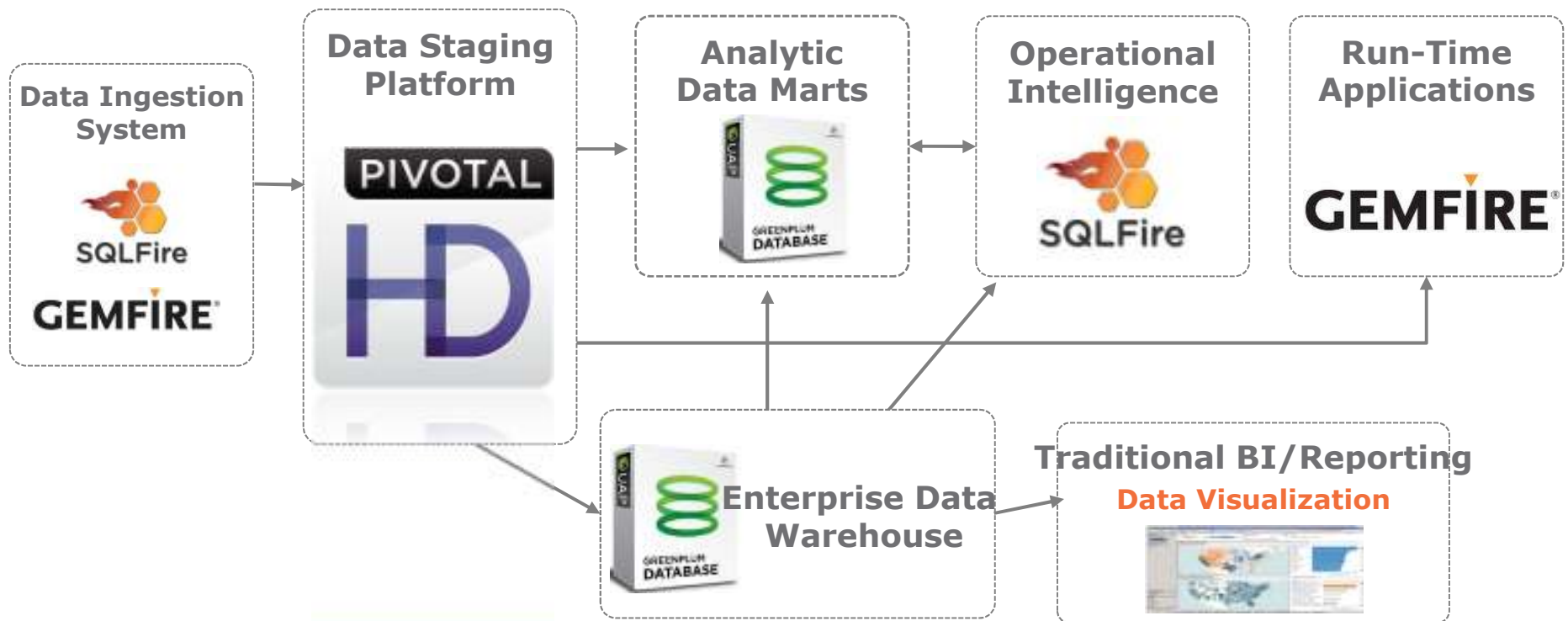


EMC²

Emerging Big Data Platform Architecture



Big Data Platform Product Portfolio Today



Pivotal Solution Areas



Big Data / Hadoop

Enterprise Data Lake
(EDL)

Pivotal One Data Suite

DS Lab, PHD, Hawk, Gemfire
XD, Spring XD, PDD, GPDB,
Chorus

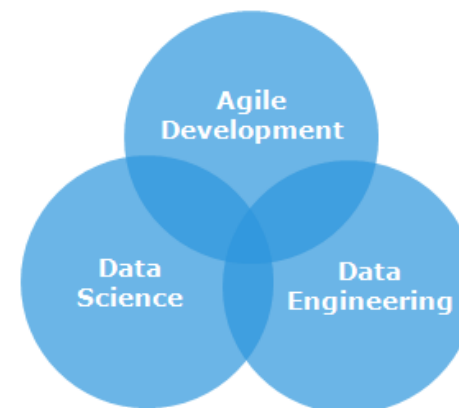


PaaS / Cloud platforms

Enterprise PaaS

Pivotal CF

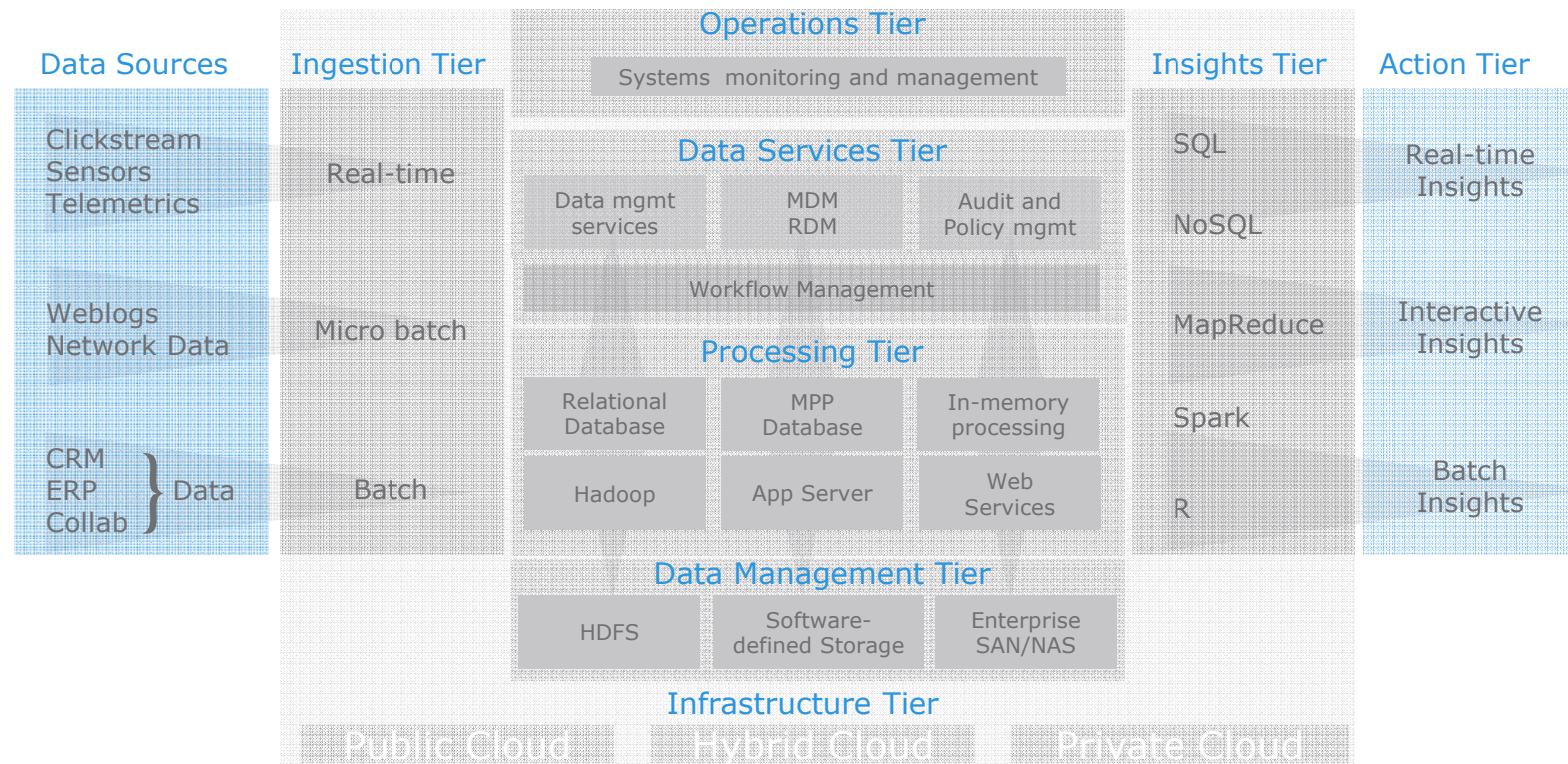
Add on Pivotal One Services



EMC²

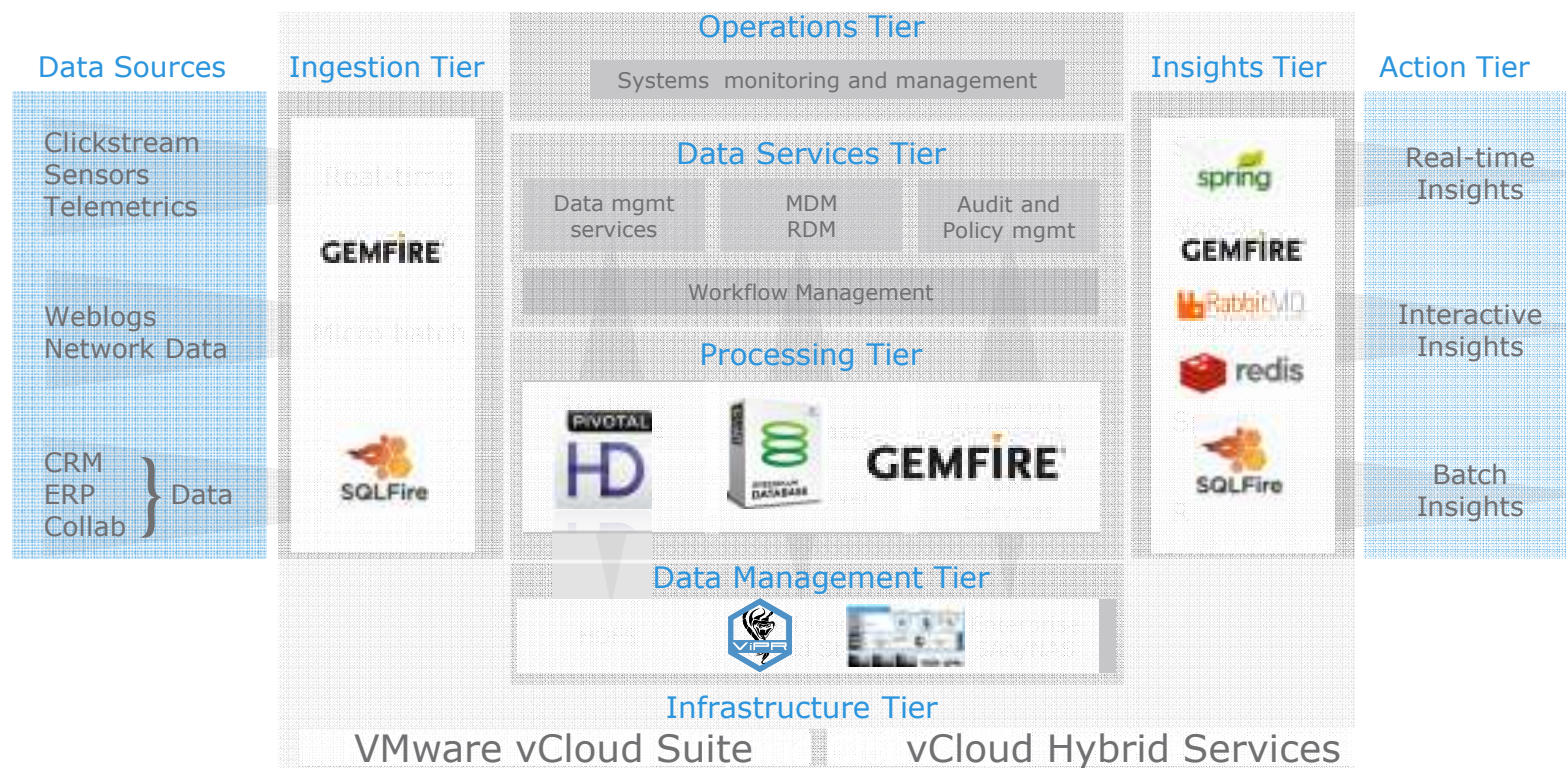
The Emerging Data Platform Ecosystem

Business Data Lake



Business Data Lake

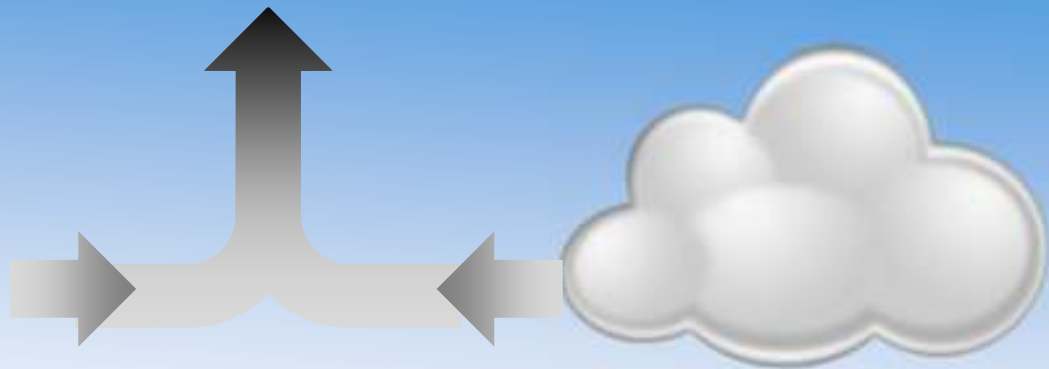
EMC Federation Solutions



Enterprise Platform-as-a-Service (PaaS)



Agile Software Methodology



Cloud Delivery Architecture

Enterprise PaaS with Pivotal CF

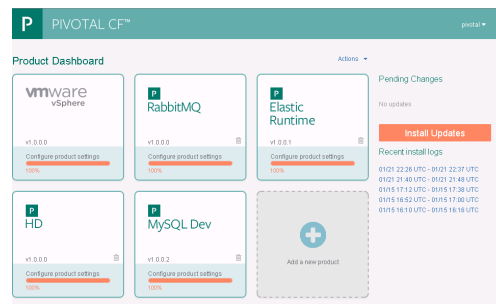


PaaS / Cloud platforms

Enterprise PaaS

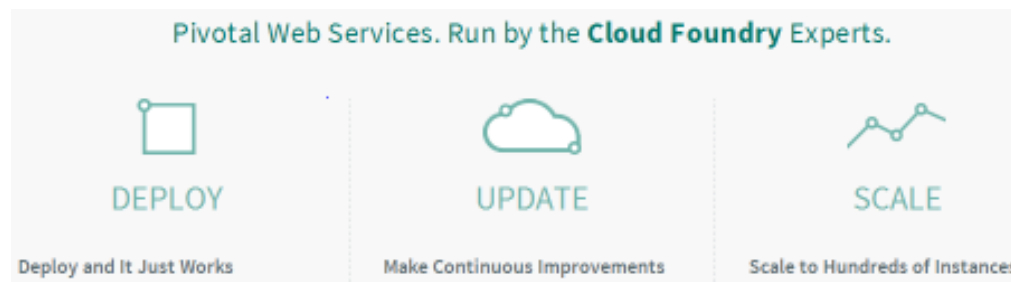
Pivotal CF

Add on Pivotal One Services



Pivotal CF: the only enterprise distribution of Cloud Foundry

Pivotal Web Services



EMC²



Cloud Foundry is becoming
the de factor standard in
PaaS

"IaaS without PaaS
capabilities becoming
a dying breed"



Enterprise Platform as a Service (PaaS)

The screenshot displays the Pivotal CF Product Dashboard. The header bar is teal with the Pivotal logo and 'PIVOTAL CF™' on the left, and a 'pivotal' dropdown on the right. The main content area is light purple and titled 'Product Dashboard' with an 'Actions' dropdown. It features five product tiles: VMware vSphere (v1.0.0.0), RabbitMQ (v1.0.0.0), Elastic Runtime (v1.0.0.1), HD (v1.0.0.0), and MySQL Dev (v1.0.0.2). Each tile includes a 'Configure product settings' button and a 100% progress bar. A sixth tile with a plus icon is labeled 'Add a new product'. To the right, the 'Pending Changes' section shows 'No updates' and an 'Install Updates' button. Below this, the 'Recent install logs' section lists several log entries with timestamps.

PIVOTAL CF™

Product Dashboard

vmware vSphere

v1.0.0.0

Configure product settings

100%

RabbitMQ

v1.0.0.0

Configure product settings

100%

Elastic Runtime

v1.0.0.1

Configure product settings

100%

HD

v1.0.0.0

Configure product settings

100%

MySQL Dev

v1.0.0.2

Configure product settings

100%

Add a new product

Pending Changes

No updates

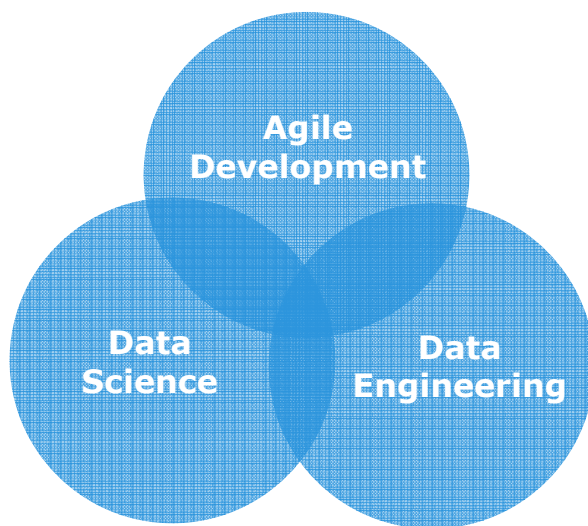
Install Updates

Recent install logs

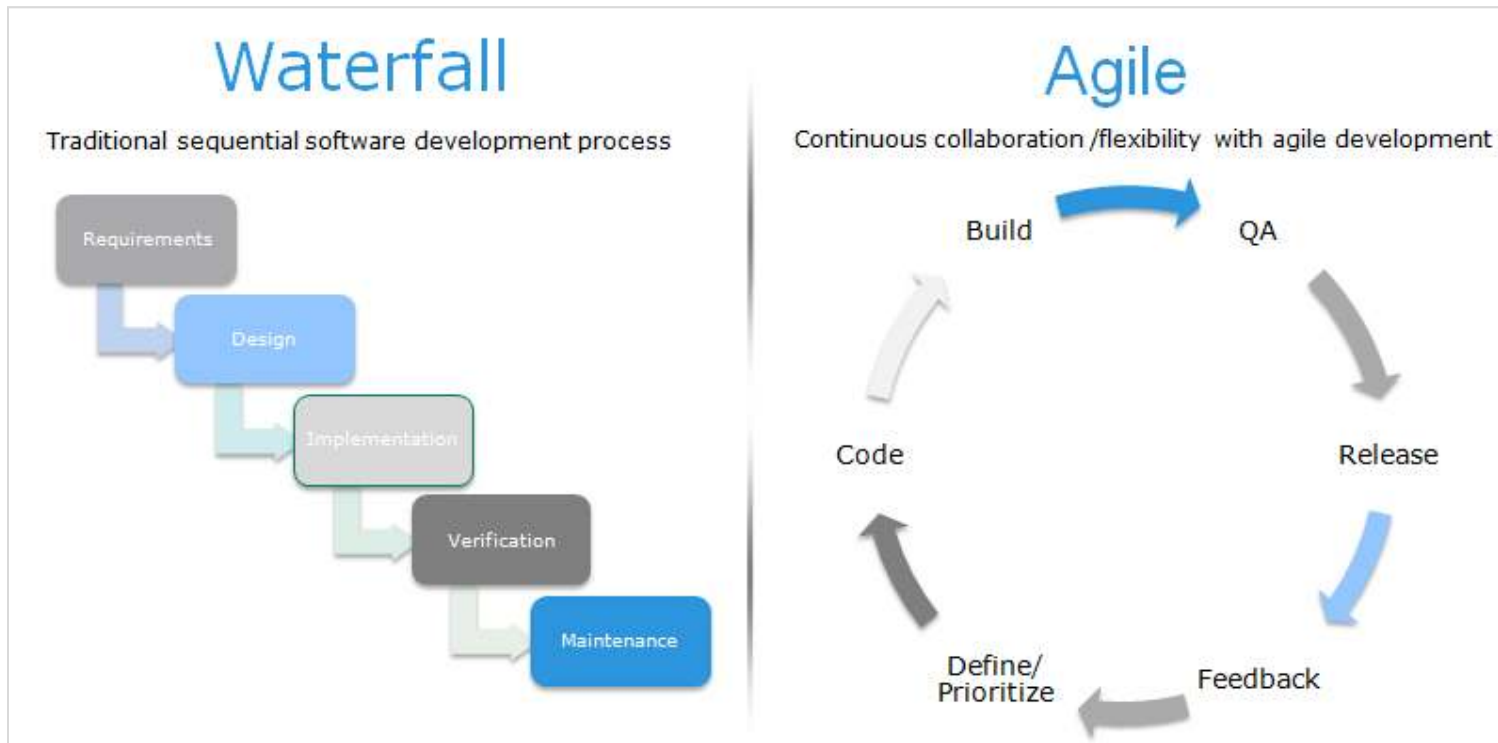
- 01/21 22:26 UTC - 01/21 22:37 UTC
- 01/21 21:40 UTC - 01/21 21:48 UTC
- 01/15 17:12 UTC - 01/15 17:38 UTC
- 01/15 16:52 UTC - 01/15 17:00 UTC
- 01/15 16:10 UTC - 01/15 16:16 UTC

Pivotal Data Labs + Pivotal Labs

The Magic in the Middle



Agile Development Becoming the Norm



Agile Development Becoming the Norm

PAST



Acceptable to delay
updates and releases

CURRENT & FUTURE



Immediate
updates

Enabling the Predictive Enterprise

EMC Consulting Big Data Services

1

ANALYTICS & INSIGHTS

- Big Data Vision Workshops
- Big Data Advisory Service

2

INFRASTRUCTURE & ARCHITECTURE

- DW to Pivotal Migration
- Big Fast Data Architecture

3

ORGANIZATIONAL TRANSFORMATION

- Data Science Education Services
- Industry Services
 - Financial Services
 - Healthcare
 - Telco

EMC²

EMC Education Big Data Curriculum

EMC Academic Alliance and Big Data Courses



For more information, visit <http://education.EMC.com/datascience> or http://education.EMC.com/Business_Transformation

EMC²

EMC Big Data Projects in Production

Global Sales

Portfolio cross-selling

- Prioritizing cross-sell and competitive takeout within existing accounts
- 30TB of structured data from 10+ sources in Greenplum DCA
- Higher share of wallet and new lead generation
- Increased sales productivity

Contract renewals

- Predicting customers' likelihood to renew maintenance contracts
- 100M+ rows of account data and 35+ predictive factors processed weekly
- Increased inside-sales and quoting productivity
- Higher renewals rate

Total Customer Experience

Customer quality and prioritization

- Identifying emerging product, service, deployment and sales quality issues
- 12TB of supplier, shipping, account, sales, ..., data analyzed daily
- Product, service and process improvements; lower support costs
- Higher customer satisfaction and loyalty

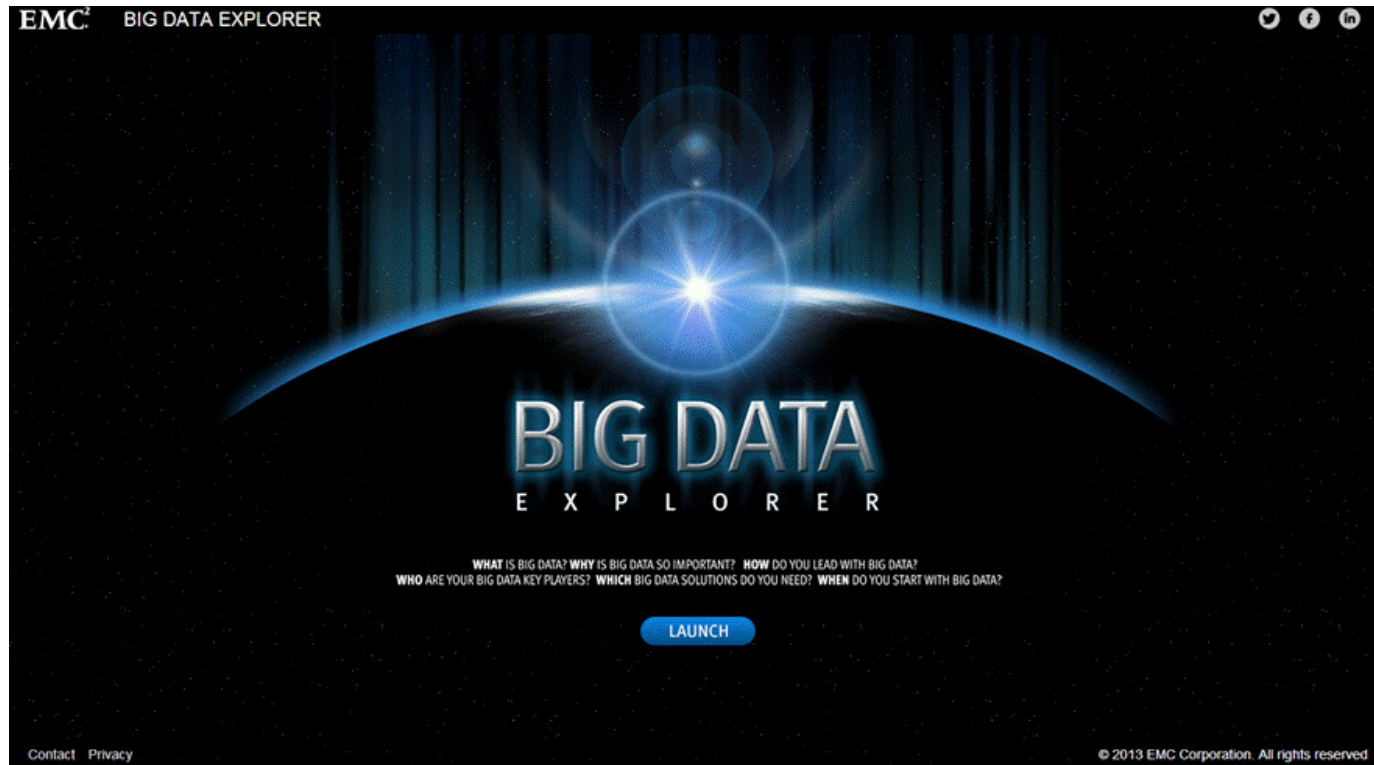
EMC IT

Incident prediction

- Predicting IT incidents for mission critical applications
- 150k alerts per day; 6k KPIs/sys logs processed every 2 mins
- Reduced downtime of mission critical applications
- Increased IT staff effectiveness

Explore the Big Data Universe

<http://www.emc.com/microsites/big-data-explorer/>



EMC²

EMC Federation Big Data Ecosystem

Openness and Interoperability

Pivotal



- Multi-cloud PaaS
- Enterprise Class Hadoop/Analytics
- Agile Development

vmware



- Rapid Provisioning
- Elastic Scale
- Optimal Utilization
- Pervasive Automation

EMC II



- Industry Leading Storage Portfolio
- Software-defined Storage
- Scale-up, Scale-out Data Services
- Data Protection Architecture

RSA



- Risk Intelligence
- Advanced Analytics
- Adaptive Controls
- Intelligence Sharing

EMC²

EMC²®