

Big Data Analytics on Amazon Web Services



Get Started with Big-Data-as-a-Service



Today's Speakers



Tom Phelan

Co-Founder and Chief Architect
BlueData Software
 @tapbluedata



Anant Chintamaneni

Vice President of Products
BlueData Software
 @anantcman



Agenda

- Big-Data-as-a-Service (BDaaS) Defined
- BDaaS on Amazon Web Services
- Challenges
- Solution Overview
- Customer Examples
- Key Takeaways
- Q & A < Enter questions at any time in your BrightTalk web client

Big-Data-as-a-Service Defined



**On-Demand, Self-Service, Elastic
Big Data Infrastructure, Applications, Analytics**

“BDaaS basically provides a cloud-based framework that offers end-to-end big data solutions to business organizations.”

Source: <http://www.marketsandmarkets.com/Market-Reports/big-data-as-a-service-market-4129107.html>



On-Premises



Public Cloud

Why BDaaS on a Public Cloud?



- Big Data infrastructure has become too complex
 - Few organizations have the skills & resources to manage Big Data infra
 - “Big Data” is no longer just Hadoop. It’s a community of interoperating frameworks, each requiring its own cluster
- Increasing importance of agility, flexibility, and expandability
 - The rapid pace of evolution in the Big Data space renders traditional infra platforms obsolete before they can be fully implemented in the enterprise
 - IoT (machine data) is increasingly being stored in the public cloud
 - Data Science workflows demand flexibility and bursts in computing power
- Constant need to drive down IT costs

Why BDaaS on AWS?

On-Demand
Access

Flexible Compute &
Local Storage



Lower Upfront
Costs

Shared, Cost-
Effective Big Data
Storage (S3)

BDaaS Landscape on AWS

Amazon Services



Analytics

Athena

EMR

CloudSearch

Elasticsearch Service

Kinesis

Data Pipeline

QuickSight

Big Data Products (e.g. AWS Marketplace)



Zoomdata

★★★★★ (1) Version 2.3.4 | Sold by Zoomdata

Starting from \$2.7K/yr or from \$21,700.00/yr (up to 10% savings for software + AWS usage fees)
Zoomdata is built on AWS Lambda for fast, real-time Analytics for Big Data and includes smart connectors for Redshift, S3, Kinesis, Apache Spark, Cloudwatch, Hadoop, MapR, Elastic,...

Hortonworks Data Cloud - HDP Services

★★★★★ (1) Version 1.1.1 | Sold by Hortonworks

Starting from \$8.5K/yr or from \$79K.00/yr (10% savings for software + AWS usage fees)
Hortonworks Data Cloud for AWS is a platform for analyzing and processing data. Hortonworks Data Cloud enables you to quickly launch Apache Hive and Apache Spark clusters...
Ubuntu/Ubuntu, Amazon Linux 2016.03 - 64-bit Amazon Machine Image (AMI)

Hortonworks Data Cloud - Controller Service

★★★★★ (1) Version 1.8.1 | Sold by Hortonworks

Starting from \$8.5K/yr or from \$79K.00/yr (10% savings for software + AWS usage fees)
Hortonworks Data Cloud for AWS is a platform for analyzing and processing data. Hortonworks Data Cloud enables you to quickly launch Apache Hive and Apache Spark clusters...
Ubuntu/Ubuntu, Amazon Linux 2016.03 - 64-bit Amazon Machine Image (AMI)

Apache Cassandra

★★★★★ (1) Sold by installcassandra

Installcassandra lets companies effortlessly run and scale the world's most powerful NoSQL on AWS. Installcassandra manages and monitors Cassandra for you, letting you focus your resources...

MapR Converged Enterprise Edition Plus

★★★★★ (1) Version 5.2 | Sold by MapR Technologies Inc.

Starting from \$6.1K/yr or from \$64K.00/yr (up to 4% savings for software + AWS usage fees)
MapR Converged Enterprise Edition Plus includes 24/7 support for the MapR Converged

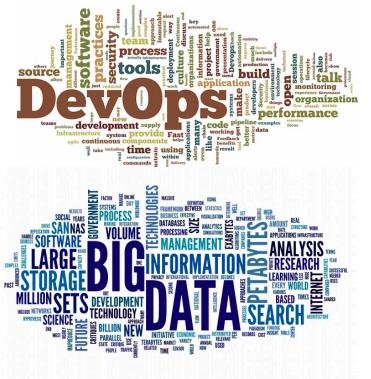
Managed SaaS



The Big Data SaaS Company



BIG DATA AS A SERVICE



Simple use case

Product Choices

Product Evaluation

Silo'ed user experience

Support

Hadoop/Spark Only

Flexibility & Choice

Skills, UX, Lock-In

Gap: Bring your own Big Data application with minimal DevOps overhead & intuitive UX

Challenges with BDaaS on AWS

- Flexibility
- Complexity
- Administration
- Data Gravity
- Security and Auditing
- Multi-Cloud

Big Data Software Flexibility

BI/Integration Tools



Analytics & Machine Learning



App & Workflow Development



Distributed Data Platforms



Challenges with BDaaS on AWS

- Flexibility
- Complexity
- Administration
- Data Gravity
- Security and Auditing
- Multi-Cloud

Onboarding Complexity

- Migrating platforms to the Public Cloud (e.g. AWS) is not easy
 - Can all the security and monitoring rules put in place in the data center be maintained in the public cloud?
 - Where is my data?
- Starting in the Public Cloud (e.g. AWS) is not so easy either
 - Why do I need a CloudFormation template?
 - Why do I need to run SSH tunnels and FoxyProxy etc?
 - VP(C, E, N, etc.) what?

Deployment Complexity

- Deploying any specific distributed platform requires specialized skills
- 100's of steps for a “single” data platform. Spark/Hadoop example:
 - Select AMI, instance type, # of instances
 - Add storage, security group, launch instances
 - Set up keys, connect to instance, configure network
 - Copy software, install dependencies, install software
 - Repeat for each instance OR use management software (Cloudera Manager, Ambari)
 - Setup KDC and enforce Kerberos-based authentication, control access to cloud storage
 - Automate using Chef/Puppet or AWS-specific tools
 - Build additional management UI for access control, tools to capture costs, etc.
- Repeat for other data platforms & tools required in your pipeline

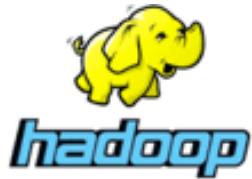
Challenges with BDaaS on AWS

- Flexibility
- Complexity
- Administration
- Data Gravity
- Security and Auditing
- Multi-Cloud

Administration

- More than powering on AWS EC2 instances
- More than installing the software bits
- Need to configure the right services on the right instances so the whole cluster works without further human intervention
- Need to manage users, user access, and data sources
- Need to control costs and cost assignment in AWS
- And more ...

Software Administration



Provide access to management consoles such as Cloudera Manager, Ambari, MapR Control Center, OR Resource Manager, YARN Job History Server



Spark Master UI ...



Standard Monitoring Tool (e.g. Ops Center)

Gap: Need to enable easy access to all management consoles from a centralized UI

System Administration

Building a cluster for one or more data platforms is just the start ...



User access
(AWS UI or purpose-built UI)

Multi-tenancy
(teams, environments)

Data sources
(S3 storage access)



App/Data Integration
(existing apps, other data platforms)

End User Tools
(gateway/edge node configuration)

Monitoring & Mgmt
(services, log collection)



Elasticity
(scaling clusters up/down)

Cost controls
(resource policies, stop instances)

Cost visibility
(by team, cluster, month, day)

Challenges with BDaaS on AWS

- Flexibility
- Complexity
- Administration
- **Data Gravity**
- Security and Auditing
- Multi-Cloud

Where Is Your Data?

- You can't run your Big Data jobs without the data
- On-premises storage
 - Requires migration/ingestion into a local HDFS file system
- In Amazon S3
 - Requires access setup
- In a database (e.g. RDS)
 - Requires network connectivity
- Somewhere else
 - ?



Challenges with BDaaS on AWS

- Flexibility
- Complexity
- Administration
- Data Gravity
- Security and Auditing
- Multi-Cloud

Security and Auditing

- LDAP/AD
 - Authorization and authentication
- Kerberos
 - Secure HDFS file-level data access
- Amazon S3/Public Object Store
 - Authorization and ACLs (requires cloud-specific expertise)
- Knox, Ranger
 - Secure column level data access, data access auditing
- Logging



Challenges with BDaaS on AWS

- Flexibility
- Complexity
- Administration
- Data Gravity
- Security and Auditing
- Multi-Cloud

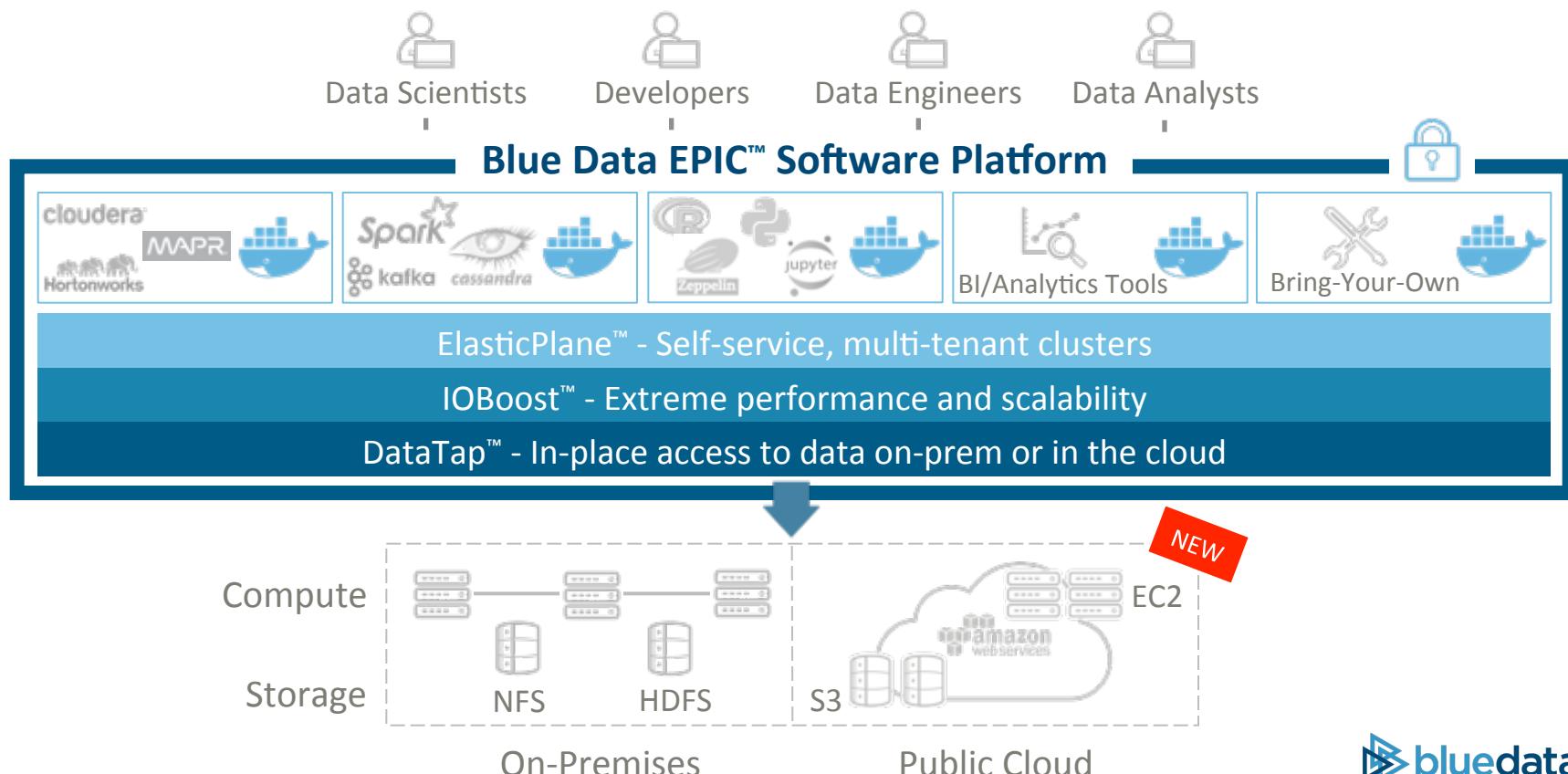
AWS + 1 and Multi-Cloud Strategies



Google Cloud Platform



BlueData EPIC: On-Prem or on AWS



Introducing BlueData EPIC on AWS

BDaaS software platform deployed on a single Amazon EC2 instance that delivers ...



on



Introducing BlueData EPIC on AWS



Tenants



Users



Hosts



Nodes

Site Admin
John / Site Admin



User Guide

App Store

Images

Add-On Images

cloudera



CDH 5.7 with Clo

✓ Insta

MAPR



Deploy multiple Big Data platforms & tools in minutes

With No DevOps Skills Required



DataStax Cassandra 3.9

✓ Installed



Kafka 0.9.0.1

✓ Installed



Spark 1.6.1 with Jupyter

✓ Installed



Splunk Enterprise 6.3

✓ Installed

Dashboard

Tenants

Users

Hosts

Nodes

Installation

App Store

Settings

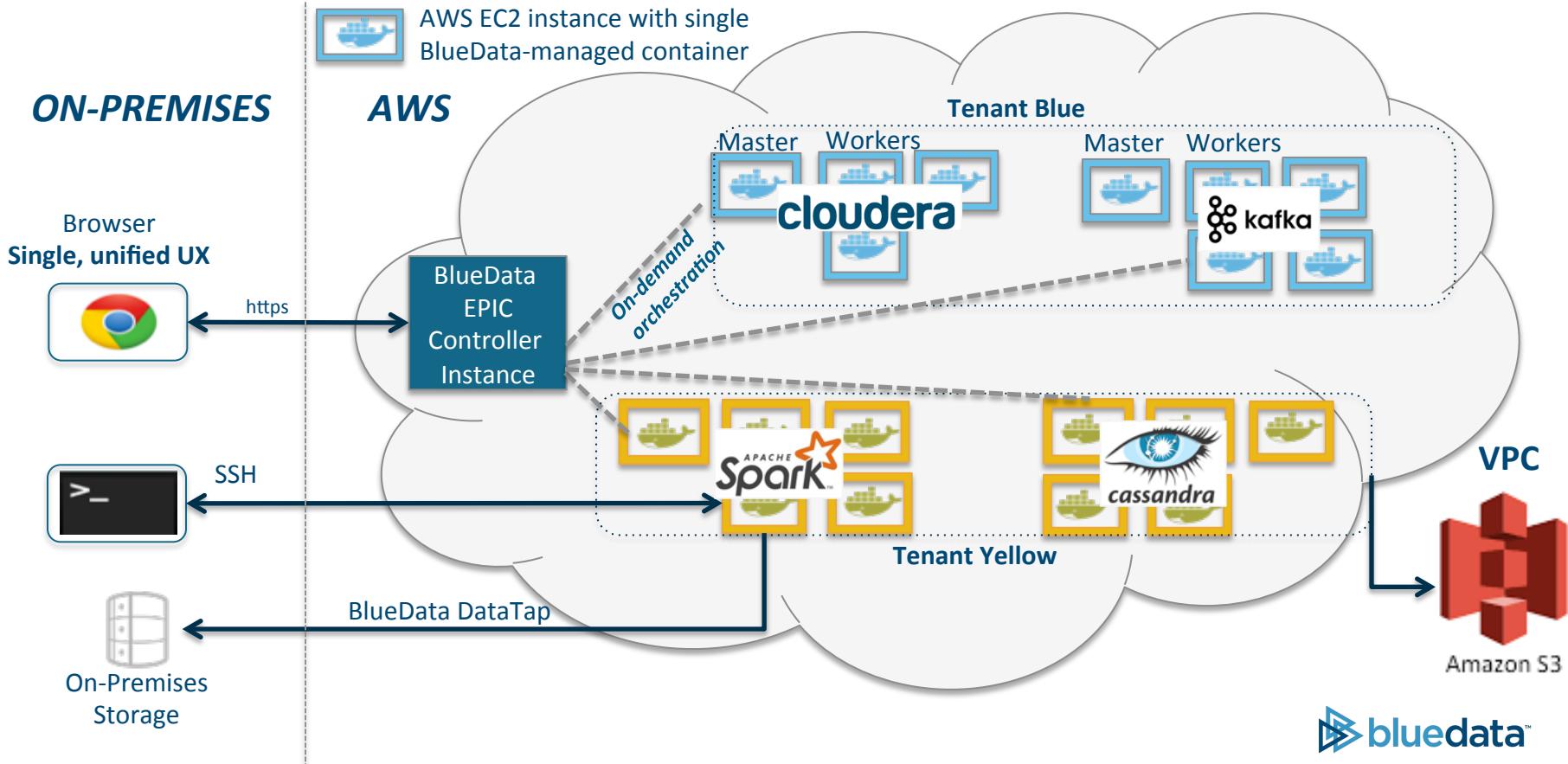
Support



Introducing BlueData EPIC on AWS

- **Simplified UI** abstracts users from AWS-specific details
- **Reduces AWS onboarding** / development (DevOps) costs
- **Flexibility and choice** for Big Data apps, distributions, and tools
- **Reduces AWS costs** with fine-grained resource quotas, start/stop controls, and cost reporting integrations (via tags)
- **Faster time to insight** with in-place analytics against on-premises data via DataTap
- **Improves data governance** and lowers risk with pre-built integrations to S3 and VPC (incl. site-to-site VPN)

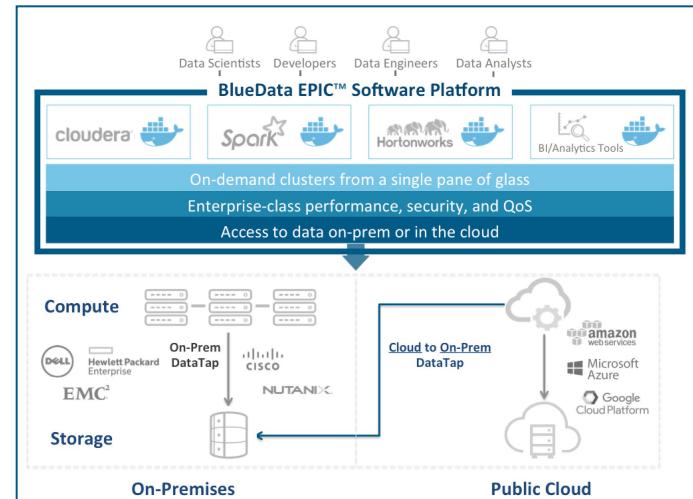
BlueData EPIC on AWS Architecture



BlueData for Big-Data-as-a-Service

Unified BDaaS software platform for Big Data
whether on-prem or in the public cloud

- Self-service creation of Big Data clusters running on Docker containers
- Ultimate flexibility and choice for Big Data frameworks, distributions, applications
- Tap into both on-prem and cloud storage
- Any server, any storage, any public cloud
- Enterprise-class security and control



Deploy BlueData EPIC Controller in AWS

Obtain BlueData EPIC CloudFormation JSON File & Follow Wizard → 20 minutes

The screenshot shows the AWS CloudFormation 'Create stack' wizard in progress. The current step is 'Select Template'. On the left, a sidebar lists steps: 'Select Template' (highlighted in orange), 'Specify Details', 'Options', and 'Review'. The main area has two sections: 'Design a template' (with a 'Design template' button) and 'Choose a template' (with three options: 'Select a sample template', 'Upload a template to Amazon S3' (selected), and 'Specify an Amazon S3 template URL'). The 'Upload a template to Amazon S3' section contains a 'Choose File' input field with the path 'epic-cfn-by...5-8755.json'. A large blue callout box covers this section, with an arrow pointing from the 'Choose File' field to a blue button labeled 'Upload JSON File'.

Deploy BlueData EPIC Controller in AWS

Obtain BlueData EPIC CloudFormation JSON File & Follow Wizard → 20 minutes

The screenshot shows the 'Specify Details' step of the AWS CloudFormation wizard. It includes sections for 'Parameters', 'VPC configuration', 'Instance configuration', and 'Security configuration'. A blue box highlights the 'VPC Subnet ID' field, which is set to 'subnet-00000000'. A blue arrow points from this field to a callout box containing the text: 'Deploy into existing VPC or allow BlueData to create new VPC'. Another blue box highlights the 'Controller Instance Type' dropdown, which is set to 'c4.4xlarge'. A blue arrow points from this dropdown to a callout box containing the text: 'Specify BlueData EPIC Controller Instance (defaults are perfect!)'. The 'Key Pair' field at the bottom is set to 'anant'.

Specify Details

Stack name: BLUEDATA-EPIC-FOR-AWS

Parameters

VPC configuration

VPC Subnet ID: subnet-00000000

Assign Public IPs for all virtual nodes: true

Instance configuration

Controller Instance Type: c4.4xlarge

Controller Instance root disk size: 80

Timezone: UTC

Enable CloudWatch for Controller Instance: Yes

Security configuration

Key Pair: anant

Deploy into existing VPC or allow BlueData to create new VPC

Specify BlueData EPIC Controller Instance (defaults are perfect!)

Deploy BlueData EPIC Controller in AWS

Obtain BlueData EPIC CloudFormation JSON File & Follow Wizard → 20 minutes

The screenshot shows the AWS CloudFormation console with the following details:

- Stack Name:** BLUEDATA-EPIC-FOR-AWS
- Created Time:** 2016-12-12 17:10:34 UTC-0800
- Status:** CREATE_COMPLETE
- Description:** EPIC Enterprise Controller (2.5-8755).

The **Outputs** tab is selected, showing the following output data:

Key	Value	Description
WebAccessPrivate	http://10.50.1.60	EPIC Controller web login
ControllerPrivateIp	10.50.1.60	EPIC Controller Private IP (Within VPC)
Credentials	User: admin, Password: admin123	Admin credentials for EPIC web login

Ready-to-use multi-tenant BDaaS App
(Users never see AWS UI from here on ...)

Simplified UI for End Users

AWS UI abstracted from end users (incl. admins) and is “under the hood”

The screenshot shows the BlueData Console interface, which is a simplified version of the AWS Management Console. It features a dark blue header with the BlueData logo and navigation links for EC2 Management Console, BlueData Console, Site Admin, Admin / Site Admin, User Guide, and Help.

The main dashboard has a sidebar with icons for Dashboard, Tenants, Users (highlighted in blue), Nodes, App Store, Settings, Support, and a back arrow. The main content area includes a "Dashboard" section with tabs for Usage and Metrics, and a "Create tenants & assign users" callout. Below this are sections for Cores Used, Node Storage Used (GB), and Memory Used (GB). A "Manage and onboard users independent of AWS (through integration with on-premises/cloud AD)" callout is overlaid on the user management section. To the right is a "Usage by Tenants" donut chart with segments for DataScience (orange) and Data Engg2 (light blue).

EC2 Management Console

BlueData Console

Site Admin
Admin / Site Admin

User Guide

Help

bluedata™

Dashboard

Tenants

Users

Nodes

App Store

Settings

Support

Back

Dashboard

Usage

Create tenants & assign users

Manage and onboard users independent of AWS (through integration with on-premises/cloud AD)

Cores Used

Node Storage Used (GB)

Memory Used (GB)

Data Engg2

DataScience

Usage by Tenants

Simplified Cloud Onboarding

Eliminate extensive “DevOps” FTE investment



DOCKER-BASED IMAGES OF YOUR CHOICE – SAME FOR ON-PREM, AWS, OR ANY CLOUD
(Use BlueData App Workbench to update existing images for adding newer versions or net new apps)

CDH 5.8.2 with Cloudera Manager	MySQL	Cassandra 2.1.10	Spark 2.0.1	Spark 2.0.0	Spark 2.0.1 on CentOS7
Installed	Installed	Installed	Installed	Installed	
Tools for R users Version: 1.0 Root Disk: 30 GB 	redhat® RHEL 6.7	jupyter Jupyter	neo4j	CentOS CentOS 6.7	Spark Spark 1.6.1 with R
Installed	Installed	Installed	Installed	Installed	Installed
kafka Apache Kafka 0.9.0.1	Spark Spark 1.5.2	Spark Spark 2.0.1 with Zeppelin Jupyter	Hortonworks HDP 2.3 with Ambari 2.2	Spark Spark 1.6.0	
Installed		Installed	Installed		

Simplified Cloud Onboarding

Eliminate extensive “DevOps” FTE investment

POINT & CLICK CLUSTER CREATION – DATA-DRIVEN BASED ON IMAGE DEFINITION

Fully automated orchestration, AWS resource controls, security, and elasticity (scale up/down)

Create New Cluster

Cluster Name ? Cassandra
Select Cluster Type ? **Hadoop** DataScience
HBase
Kafka
Spark
Splunk
Utility
databases

Distribution ? AM, 1

Master Node Flavor ? 1

Worker Count ? ex-small - 2 VCPU, 6144 MB RAM

Worker Node Flavor ? Cluster HA ?

Pig, Hive, & Oozie ?

Submit

Resource and Cost Controls

Onboard multiple users/groups with fine-grained quotas & visibility

Edit Tenant

Tenant Name ? Data Engg

Tenant Description ? Data Engg

Tenant Keypair Visibility ? All

Quotas **Kerberos**

Maximum Cores ? 100

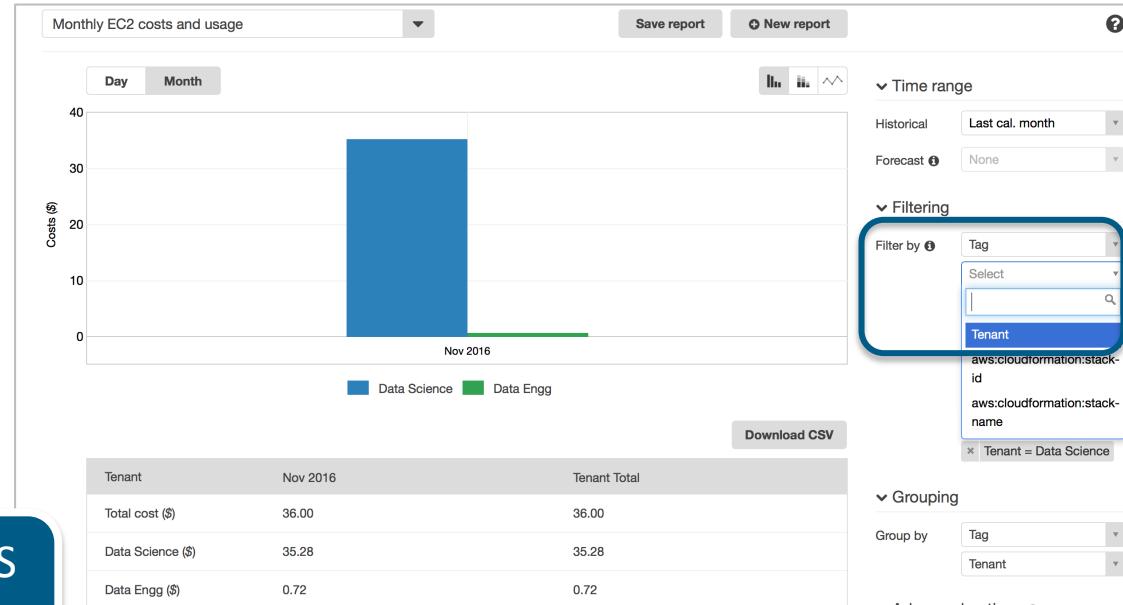
Maximum Memory (GB) ? 200

Maximum Node Storage (GB) ? 600

AWS Access Key ?
AWS Secret Key ?

MULTI-TENANT RESOURCE CONTROLS
*# of EC2 instances that can be created by a user
are limited by the quotas, thereby reducing costs*

COST VISIBILITY
Admins can view costs by tenants



Resource and Cost Controls

Stop clusters & re-start with full state retained

CLUSTER LEVEL ACTIONS TO SAVE EC2 COSTS

Initiated by end user without ever logging into AWS Management Console

Cluster List							
Actions		Tenant KeyPair		Actions			
Display 10 records		Search		Start			
#	Name	Type	Distribution	Flavor Details	Workers	Details	Status
<input type="checkbox"/>	HDP	Hadoop	HDP 2.4 with Ambari 2.2 / YARN	Master: Medium - 4 VCPU, 16384 MB RAM, 40 GB root disk Worker: Medium - 4 VCPU, 16384 MB RAM, 40 GB root disk	2	Apps Installed: (Pig/Hive/Hue)	<input type="button" value="Stop"/> <input type="button" value="Reboot"/> <input type="button" value="Delete"/>
<input checked="" type="checkbox"/>	Spark 2.0.1	Spark	Spark 2.0.1	Master: Medium - 4 VCPU, 16384 MB RAM, 40 GB root disk Worker: Medium - 4 VCPU, 16384 MB RAM, 40 GB root disk	1	ready	
<input type="checkbox"/>	CDH57	Hadoop	CDH 5.7.0 with Cloudera Manager / YARN	Master: Medium - 4 VCPU, 16384 MB RAM, 40 GB root disk Worker: Small - 2 VCPU, 8192 MB RAM, 40 GB root disk	2	Apps Installed: (Pig/Hive/Impala/Hue)	<input type="button" value="stopped"/>

Showing 1 to 3 of 3 entries

Pre-built Security Integrations

LDAP/AD integration to BlueData EPIC UI and tenants

System Settings

User Authentication Flavor License

Authentication Type ? Active Directory

Host ? 10.1.10.118

Port ? 389

Bind Type ? Search Bind

User Attribute ? cn

Base DN ? dc=qatest,dc=com

Bind DN ? dev@qatest.com
(Optional)

Centralized Active
Directory settings

CA Certificate File ? QA-AD.crt

Tenant-level AD/LDAP group
mappings

Edit Tenant

Tenant Name ? Data Science

Tenant Description ? Data Science

Tenant Keypair Visibility ? Tenant Admin Only

Quotas Groups Kerberos

External User Groups ?
(Optional)

CN=Eng,OU=Engineering,DC=qatest	Member	
CN=QA,OU=Engineering,DC=qatest	Admin	
CN=Sales,OU=Engineering,DC=qatest	Admin	

Pre-built Security Integrations

LDAP/AD integration to BlueData EPIC UI and Big Data nodes

The screenshot shows the BlueData EPIC web-based UI interface. On the left is a sidebar with icons for Dashboard, Jobs, Clusters (selected), DataTaps, and Nodes. The main area displays a "Spark 2.0.1" cluster's Node List. A modal window titled "Node List" shows two nodes: "ip-10-50-1-146.ec2.internal" (Spark 2.0.1, master, IP 10.50.1.146) and "ip-10-50-1-89.ec2.internal". A callout box highlights the "Data Science Joel Baxter / Member" user information in the top right corner of the UI.

BlueData web-based UI integrated with AD authentication (e.g. user = Joel Baxter)

Name	Distribution	Role	Instance IP	Process List
ip-10-50-1-146.ec2.internal	Spark 2.0.1	master	10.50.1.146	Spark master , Spark worker , Zeppelin server
ip-10-50-1-89.ec2.internal				

Anants-MacBook-Pro:Downloads anant\$ ssh "Joel Baxter"@10.50.1.146

Joel Baxter@10.50.1.146's password:

Last login: Tue Dec 6 16:32:53 2016 from ip-10-1-10-56.ec2.internal
[Joel Baxter@ip-10-50-1-146 ~]\$

Login to nodes access controlled by AD

Pre-built Security Integrations

Automated S3a integration for tenant-level credentials

Edit Tenant

Tenant Name ?

Data Science

Tenant Description ?

Data Science

Tenant Keypair Visibility ?

Tenant Admin Only

Quotas

Groups

Kerberos

Maximum Cores ?

100

Maximum Memory (GB) ?

200

Maximum Node Storage (GB) ?

500

AWS Access Key ?

.....

AWS Secret Key ?

.....

AWS Region ?

us-east-1

Create New Job

Job Name ? Analytics Job against S3 storage

Job Type ? Spark - Scala Jar

Jar File ? bluedata-spark-scala-sample_2.10-1.0.jar

Change

Job Dependencies ? Choose a file

Choose

App Name ? SimpleApp

Cluster Type ?

Persistent

Cluster ? Spark 2.0.1

Spark Arguments ?

App Arguments ? s3a://anant-testing/sample/sample_text.txt

Insert

✓ Submit

Injected into every Spark
and Hadoop cluster.
Access via s3a://....



Pre-built Security Integrations

Automated Kerberos

Edit Tenant

Tenant Name: Data Science

Tenant Description: Data Science

Tenant Keypair Visibility: Tenant Admin Only

Kerberos tab selected

KDC Type: Active Directory

KDC Host: 10.1.10.118

Kerberos Security Realm: QATEST.COM

Kerberos Encryption Types: (Optional)

KDC Username: Administrator@QATEST.COM

AD Suffix: ou=Engineering,DC=qatest,DC=c

AD LDAPS Port: 636 (Optional)

Centralized Kerberos credentials by Tenant

Create New Cluster

Cluster Name: Hadoop

Select Cluster Type: Hadoop

Distribution: CDH 5.7.0 with Cloudera Manager

MR Type: YARN

Master Node Flavor: Small - 2 VCPU, 8192 MB RAM

Worker Count: 3 Available licensed: 22

Worker Node Flavor: Small - 2 VCPU, 8192 MB RAM

Cluster HA:

Pig, Hive, Oozie, Impala & Hue:

Spark:

Enable Kerberos:

Submit button

```
graph LR; A[Edit Tenant] --> B[Create New Cluster]
```

On-Prem Data Connectivity

Accelerate time-to-insight with security and control

CONNECT TO ANY ON-PREM HDFS OR NFS SYSTEM

* Site to site VPN and/or AWS Direct Connect recommended

Edit DataTap

Name

Description

Select Type

Host

Standby NameNode Host

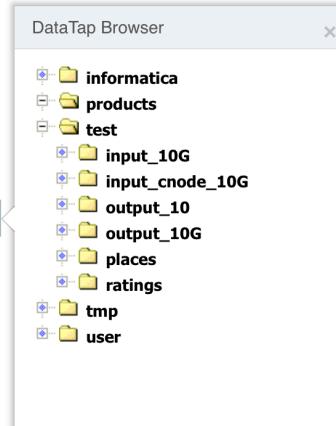
Port

Path (Optional)

Kerberos Protected (Optional)

Username (Optional)

Read Only (Optional)



New Job

Job Name

Job Type

Jar File

Job Dependencies

App Name

Cluster Type Transient Persistent

Cluster

Target remote HDFS using dtap://..

App Arguments

Customer Example: Fintech

Spark with Jupyter Notebooks and Cassandra on AWS

- Financial firm with proprietary research and trading technology
- Multiple developer and data science teams
- Custom data analytics pipelines
- App serving with Cassandra as backend
- Dev, Test, and Production environments
- Fully self-sufficient without AWS expertise

“ With BlueData, we can improve the productivity of our developers, shorten our development cycles, and control AWS costs. ”

Customer Example: ISV

Cloudera, Hortonworks, & MapR on AWS

- Data integration ISV testing their product with multiple Hadoop distributions on AWS
- Multiple QA engineers & teams
- Quick self-service access to Hadoop clusters
- Allow DevOps teams to focus on other core competencies
- Create images of new versions of Hadoop with product artifacts as necessary

“ With BlueData, we can be self-sufficient and accelerate our testing cycles by using reproducible Hadoop clusters integrated with our product. ”

BlueData EPIC on AWS

Big Data App Choice and Flexibility
(vs. opinionated stack with other BDaaS)

Simplifies Big Data DevOps on AWS
(vs. DIY initiatives on AWS)



**Speed
& agility**



**Security
& control**



**Efficiency
& lower cost**

Why BlueData for BDaaS?

Flexible software solution (Your BDaaS) vs. rigid managed service (Their BDaaS)

- Bring your own distro/application, provide application and cloud flexibility

In-place analytics against on-premises data and/or cloud storage

- Leverage DataTap to connect Hadoop/Spark compute with remote storage

Vendor-agnostic, Docker container orchestration provides maximum flexibility

- Same Docker images for on-premises and cloud deployments

Unique hybrid and multi-cloud model for BDaaS [*Next: Azure and GCP support*]

- Start on-prem and extend to cloud, or focus on cloud-first and multi-cloud

Key Takeaways: Big Data on AWS

- Know your “use case” and map to the right type of BDaaS
 - AWS EMR may be good fit if the use case is singular (e.g. Hadoop/Spark)
 - Fully-managed SaaS if you care more about support than choice & flexibility
 - Flexible platform with choice of frameworks, apps, versions for DevOps agility
- Big-Data-as-a-Service is more than Hadoop or Spark-as-a-Service
 - Modern analytics include real-time streams, NoSQL, ML runtimes, notebooks, etc.
 - Streamlined interface with common tool chain is key for faster time-to-value
 - Future-proof your implementations to deploy on-prem and/or multiple clouds
- DIY on AWS will be DOA ... leverage a purpose-built platform
 - Focus on your Big Data applications, not the AWS infrastructure and plumbing
 - BlueData EPIC is a turnkey, flexible BDaaS platform – for on-prem and cloud



Tom Phelan

 @tapbluedata

Anant Chintamaneni

 @anantcman

Thank You

For more information:

www.bluedata.com

sales@bluedata.com

www.bluedata.com/aws

TRY BLUEDATA EPIC ON AWS

