Spec Sheet

BlueData EPIC™ on AWS

Ultimate flexibility and choice for Big-Data-as-a-Service on Amazon Web Services

BENEFITS

▷ **Simplified user experience** for both administrators and data science teams, abstracting the AWS-specific infrastructure so they can focus on their Big Data needs.

▷ **Faster AWS onboarding** for multiple teams and Big Data workloads, eliminating the need for DevOps expertise and reducing the cost and time involved.

▷ **Greater agility and flexibility**, with self-service clusters pre-configured on Amazon EC2 for Spark, Hadoop, Kafka, Cassandra, and other Big Data applications.

▷ **Reduced AWS costs** through the use of fine-grained resource quotas, start/stop controls, and cost reporting in a multi-tenant environment.

▷ **Faster time to insights** with pre-built cluster integrations to Amazon S3 and in-place analytics against on-premises data.

▷ **Improved data governance** with integrations to Amazon VPC (including site-to-site VPN), Active Directory, and Kerberos for authentication.

Big Data Analytics on AWS

The BlueData EPIC (Elastic Private Instant Clusters) software platform makes deployment of Big Data infrastructure and applications easier, faster, and more cost-effective—whether on-premises or on the public cloud.

With BlueData EPIC on AWS, you can quickly and easily deploy your preferred Big Data applications, distributions, and tools; leverage enterprise-class security and cost controls for multi-tenant deployments on the Amazon cloud; and tap into both Amazon S3 and on-premises storage for your Big Data analytics.

Key Features

The BlueData EPIC software platform enables simplified, consistent, and repeatable deployment of multiple Big Data applications and tools in a multi-tenant architecture.

BlueData EPIC on AWS provides the ability to spin up your own unique Big Data environments for a range of different data science and analytics use cases within minutes—without requiring DevOps or AWS-specific skills. BlueData EPIC delivers a single unified Big-Data-as-a-Service platform to build, run, and manage a variety of different Big Data frameworks including Hadoop, Spark, Kafka, Cassandra, and more on AWS—instead of managing each deployment with different processes and tools.

This common platform minimizes the refactoring work required for Big Data developers and data science teams to get up and running with their preferred Big Data applications and tools on AWS— along with the flexibility to quickly and easily upgrade to the latest releases and new versions.

DevOps Agility for Big Data

BlueData EPIC software is deployed into your AWS account with a streamlined, wizard-based process that takes only 20-25 minutes. The EPIC software is automatically installed in a single AWS controller instance within a private Amazon VPC subnet or (optionally) a new VPC subnet.

Once installed, BlueData EPIC provides a simple and easy-to-use interface that abstracts both administrators and end users from the inner-workings of AWS. As an administrator, you can quickly and easily onboard one or more user groups (i.e. tenants) and assign them different AWS usage quotas in a multi-tenant environment.

Your end users can then log into BlueData EPIC's self-service interface to spin up instant clusters on Amazon EC2 and scale up or scale down in just a few mouse click—with built-in elasticity for all cluster types. Amazon S3 connectivity is pre-integrated into all clusters, using tenant-level access credentials.

Application Flexibility and Choice

BlueData EPIC for AWS offers a pre-populated App Store with Docker-based application images for multiple Big Data frameworks and tools. You can install these images within minutes and make them available to your end users to spin up their own clusters—eliminating the complexity and time associated with configuring and tuning each of these applications to run on AWS. Each cluster runs in an embedded and fully-managed Docker container, providing abstraction from the underlying AWS infrastructure.

The standard App Store includes the following ready-to-run images:

- Cloudera CDH 5.7 with Cloudera Manager (CDH 5.8 and 5.9 are available upon request)
- Hortonworks HDP 2.4 with Ambari 2.2
- MapR 5.1 with MapR Control System
- Apache Spark 2.0.1 with Zeppelin Notebook
- Apache Spark 1.6.1 with Job Server and Zeppelin Notebook
- DataStax Distribution of Apache Cassandra 3.9
- DataStax Distribution of Apache Cassandra 2.1.10
- Apache Kafka 0.9.0.1
- Ubuntu 16.04 (Xenial Xerus)
- CentOS 7 / CentOS 6
- RHEL 7 / RHEL 6

These same images can be used on AWS or for on-premises deployments of the BlueData EPIC platform, providing portability and a common user experience in a hybrid architecture. In addition to the above, pre-configured images for other Big Data applications (e.g. Splunk) and tools (e.g. Jupyter Notebook) may be available upon request.

BlueData EPIC also provides an App Workbench that allows your administers to easily modify and update these images (e.g. to upgrade to the latest version) within your own App Store—or create new Docker-based application images for other Big Data frameworks and tools preferred by your data scientists, developers, and analysts.

Cost Savings, Visibility, and Control

Administrators can segregate different teams or user groups as separate tenants, with visibility into their usage and costs. Each tenant can be assigned resource quotas to limit their CPU, memory, and storage usage; you can also set restrictions on which Amazon EC2 instances are available for these users.

To control AWS costs, your users can stop a cluster when it's no longer in use; they can easily restart the cluster later with the same metadata, data, and networking configurations.

You can also drill down into the usage and billing by tenant, using standard AWS Cost Explorer reports; all Amazon EC2 instances launched by BlueData EPIC are tagged with the tenant name and tenant ID.

Authentication, Security, and User Management

You can deploy BlueData EPIC into your existing Amazon VPC, ensuring logical isolation for your EC2 instances from other virtual networks in the AWS cloud. And you can easily integrate with your existing enterprise security model through Active Directory/LDAP integration available within BlueData EPIC.

You can also automatically associate an AWS user account for each tenant. This eliminates the time and potential errors of manually configuring this for every cluster in the tenant, and controls their access to data in Amazon S3.

And you automate the Kerberos configuration setup for each cluster, ensuring security and data governance.

Access to Amazon S3 and On-Premises Storage

With BlueData EPIC, you can tap into data stored on your on-premises HDFS and network file systems—using our proprietary DataTap™ functionality. This means users can analyze data either in Amazon S3 or on-premises, avoiding the time and cost of moving data to AWS. BlueData is the only platform for Big-Data-as-a-Service that supports access to both S3 and cloud storage for Big Data analytics.

Minimum Requirements
An AWS account is required for BlueData EPIC on AWS.

The one time setup of the BlueData EPIC controller requires an AWS IAM user account with the following AWS permissions:

- AmazonEC2FullAccess
- AmazonS3FullAccess
- AmazonVPCFullAccess
- AWS CloudFormation (ability to "Create Stack", "Describe", "List", "Get", "Delete", "Validate")

An individual "root" (i.e. personal AWS user account) will need to have each of the above privileges. Larger enterprises will typically have multiple AWS IAM user accounts for different business groups.

The BlueData EPIC controller must be activated via a license key. The unique controller ID (available in the UI) must be provided to BlueData Support in order to obtain the license key.

If deploying into an existing Amazon VPC:

- The subnet for site-to-site VPN (if used) must be provided;
- The security group for the subnet must permit the BlueData EPIC controller instance to access the internet temporarily, in order to finish deployment of the stack.
- Amazon S3 endpoints must be configured for the subnet so S3 buckets may be accessed by BlueData-provisioned clusters.

To ensure optimal performance when tapping into on-premises storage, Amazon Direct Connect and/or a site-to-site VPN between the Amazon VPC and your site is recommended.

BlueData EPIC is currently available in all North America AWS regions; other regions may be available upon request. The BlueData EPIC controller cannot manage instances across different AWS regions.

Pricing
BlueData EPIC on AWS starts at $499 per month for 15 managed Amazon EC2 instances, including standard support and training. Premium support is available at 20% of the monthly license.

Sign up for a free two-week trial at bluedata.com/aws