DigitalEdge™

Getting Started

on a Public Cloud

Version 1.3

September 2014

leidos

DISCLAIMER OF WARRANTY AND LIMITATION OF LIABILITY

The Software accompanying this Documentation is provided with the Limited Warranty contained in the License Agreement for that Software. Leidos, its affiliates and suppliers, disclaim all warranties that the Software will perform as expected or desired on any machine or in any environment. Leidos, its affiliates and suppliers, further disclaim any warranties that this Documentation is complete, accurate, or error-free. Both the Software and the Documentation are subject to updates or changes at Leidos' sole discretion. LEIDOS, ITS LICENSORS AND SUPPLIERS MAKE NO OTHER WARRANTIES, WRITTEN OR ORAL, EXPRESS OR IMPLIED RELATING TO THE PRODUCTS, SOFTWARE, AND DOCUMENTATION. LEIDOS, ITS LICENSORS AND SUPPLIERS DISCLAIM ALL IMPLIED WARRANTIES, INCLUDING WITHOUT LIMITATION ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, USE, TITLE, AND NON-INFRINGEMENT OF THIRD PARTY RIGHTS. In no event shall Leidos, its affiliates or suppliers, be liable to the End User for any consequential, incidental, indirect, exemplary, punitive, or special damages (including lost profits, lost data, or cost of substitute goods or services) related to or arising out of the use of this Software and Documentation however caused and whether such damages are based in tort (including negligence), contract, or otherwise, and regardless of whether Leidos, its affiliates or suppliers, has been advised of the possibility of such damages in advance. TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAWS, END USER ACKNOWLEDGES AND AGREES THAT LEIDOS AND ITS AFFILIATES AND SUPPLIERS IN NO EVENT SHALL BE RESPONSIBLE OR LIABLE TO THE END USER FOR ANY AMOUNTS IN EXCESS OF THE FEES PAID BY THE END USER TO LEIDOS. LEIDOS SHALL NOT BE RESPONSIBLE FOR ANY MATTER BEYOND ITS REASONABLE CONTROL.

LEIDOS PROPRIETARY INFORMATION

This document contains Leidos Proprietary Information. It may be used by recipient only for the purpose for which it was transmitted and will be returned or destroyed upon request or when no longer needed by recipient. It may not be copied or communicated without the advance written consent of Leidos. This document contains trade secrets and commercial or financial information which are privileged and confidential and exempt from disclosure under the Freedom of Information Act, 5 U.S.C. § 552.

TRADEMARKS AND ACKNOWLEDGMENTS

Private installations of DigitalEdge are powered by Eucalyptus®.

Public cloud installations of DigitalEdge are powered by Amazon Web Services™.

The following list includes all trademarks that are referenced throughout the DigitalEdge documentation suite.

Adobe, Flash, PDF, and Shockwave are either registered trademarks or trademarks of Adobe Systems Incorporate in the United States and/or other countries.

Amazon Web Services, AWS, Amazon Elastic Compute Cloud, Amazon EC2, EC2, Amazon Simple Storage Service, Amazon S3, Amazon VPC, Amazon DynamoDB, Amazon Route 53, the "Powered by Amazon Web Services" logo, are trademarks of Amazon.com, Inc. or its affiliates in the United States and/or other countries.

Apache, Archiva, Cassandra, Hadoop, Hive, HBase, Hue, Lucene, Maven, Apache Phoenix, Solr, Zoie, ActiveMQ are all trademarks of The Apache Software Foundation.

ArcSight is a registered trademark of ArcSight, Inc.

CAS is copyright 2007, JA-SIG, Inc.

CentOS is a trademark of the CentOS Project.

Cloudera is a registered trademark of Cloudera, Inc.

CloudShield is a registered trademark of CloudShield Technologies, Inc. in the U.S. and/or other countries.

CTools are open-source tools produced and managed by Webdetails Consulting Company in Portugal.

Drupal is a registered trademark of Dries Buytaert.

Elasticsearch is a trademark of Elasticsearch BV, registered in the U.S. and in other countries.

Eucalyptus and Walrus are registered trademarks of Eucalyptus Systems, Inc.

CONTACT INFORMATION

Leidos Franklin Center
6841 Benjamin Franklin Drive
Columbia, Maryland 21046

Email: DigitalEdgeSupport@Leidos.com

DigitalEdge Technical Support: 443-367-7770

DigitalEdge Sales Support: 443-367-7800

To submit ideas or feedback: https://www9.v1ideas.com/digitaledge/welcome

Contents

Chapter 1: Introduction

DigitalEdge is a highly configurable software platform providing real-time analytics of big data in motion for cyber-security. This *Getting Started Guide* helps you with:

- Understanding the high level platform architecture
- Installing the operating system and cloud software
- Setting up DigitalEdge
- Working with tenant accounts

Before plunging in, be sure to read the *Overview Guide*; it describes the DigitalEdge architecture, concepts, and terminology. These instructions assume you are familiar with those concepts.

This *Getting Started Guide* helps you configure the servers and network interfaces required for the DigitalEdge platform.

Once you have the platform set up and installed, use the *Configuration Guide* to plan, configure, and build the DigitalEdge data models, work flow, and plug-in components. Then, use the *Operations Guide* to manage and maintain DigitalEdge.

DigitalEdge Platform

DigitalEdge is an advanced, customizable software platform that enables a full lifecycle from data discovery to actionable intelligence. It provides enhanced ingestion of structured and unstructured data into customized work flows for real-time situational awareness. The processing pipeline can be built with custom plug-ins to transport source data, parse and enrich data, and load data into big data repositories or enterprise systems. Data can be queried, analyzed, or reported in near real-time. In short, DigitalEdge integrates ETL (extract, transform, and load) real-time stream processing, and big data NoSQL ("Not Only SQL") stores into a high performance analytic system. Instead of waiting hours for actionable data and reports, analysts can now achieve near real-time situational awareness. DigitalEdge provides a rapid response to changing environments.

DigitalEdge is Java based and provides an extensible architecture, APIs, and development kits. The platform provides plug-in architecture for sharing components in a problem domain. DigitalEdge can be provisioned as a fully integrated solution on x86 hardware, or off-premise on a public cloud. DigitalEdge runs as a platform-as-a-service (PaaS) in:

- The Amazon AWS™ public Elastic Compute Cloud™ (EC2), using the Virtual Private Cloud™ (VPC™) environment for security
- A private cloud using Eucalyptus® in your own data center

Cloud computing provides on-demand network access to a shared pool of configurable resources such as servers, storage, applications, and services. Resources are automatically provisioned with minimal intervention, scaling up during peak times and scaling back as needs decrease.

Product documentation

DigitalEdge is a complex big data platform. The system comes with a complete set of documentation in PDF and HTML5 formats to help you master DigitalEdge:

Document	Use	Audience
Overview Guide	Basic information about the DigitalEdge platform, including architecture, concepts, and terminology; a must-read before working with any aspect of DigitalEdge	Anyone working with DigitalEdge in any capacity
Configuration Guide	Instructions for defining data models and building processing pipelines	Data Specialists, DigitalEdge Administrators
Operations Guide	Daily administration information, covering monitoring, managing, and modifying the platform	DigitalEdge Administrators
Cookbook	Guidelines and procedures for many common tasks in a DigitalEdge system	DigitalEdge Administrators
DigitalEdge SDK Guide	Reference for building custom plug-in components	Developers
Alerts API Guide	Reference for specifying data triggers and notifications for an alerting capability	Developers
Search API Guide	Reference for providing search services on a Lucene data sink node	Developers

Typographical conventions

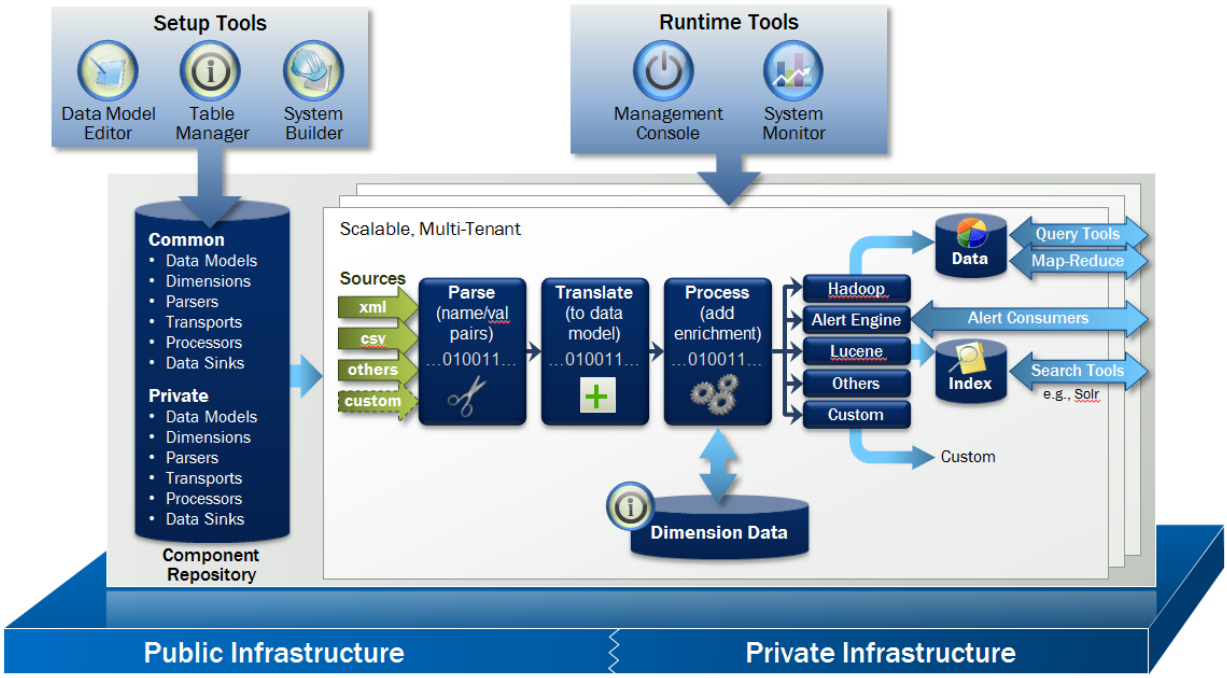
The following style conventions are used throughout this documentation:

Type of Information	Style in Documentation
Code, commands, filenames	`code`
Cross references	Click to see this topic
Emphasis	*important point*
Hyperlinks	Click to go to this site
Notes, warnings, tips	
References to other documents	*Document Title*
Sample code blocks	`code`

Type of Information	Style in Documentation
Troubleshooting issue or problem	
Troubleshooting solution	
User input	*Italics*
User interface labels and controls	**Bold**
Variables	<*change-this-name*>

Chapter 2: DigitalEdge Architecture

DigitalEdge is highly configurable, with a plug-in architecture that lets you swap components in and out. Plug-in components are stored in the Component Repository. The DigitalEdge system architecture is designed as follows:



Data moves through system processors which are configured and customized with the DigitalEdge Setup Tools. System Builder builds and assembles the components into a processing pipeline. The processing pipeline is completely configurable with the Setup Tools.The data flow includes these steps:

1. Transports grab data from data sources and feed the data into DigitalEdge.
2. Data is extracted by parsers.
3. The fusion engine translates and normalizes the data to the DigitalEdge input model.
4. The enrichment engine adds dimensional data and algorithmic enrichments to provide context and meaning to data, resulting in all relevant data being integrated into one record
5. Data is processed and stored in persistent data sinks managed by DigitalEdge (Hadoop, Hive, HBase, MongoDB, etc.) or sent to other data sinks for post-processing (indexing, alerting, etc.). Data can also be sent to systems outside of DigitalEdge.
6. Various web apps makes the data accessible in several ways:
    - Indexed data is searchable through the Search API or the Search app.
    - Configurable situational information is sent to users by the alerting engine.
    - Data can be viewed in dashboards or other external applications.

Tenant Management System (TMS)

The Tenant Management System is the DigitalEdge application for creating and managing tenant accounts.

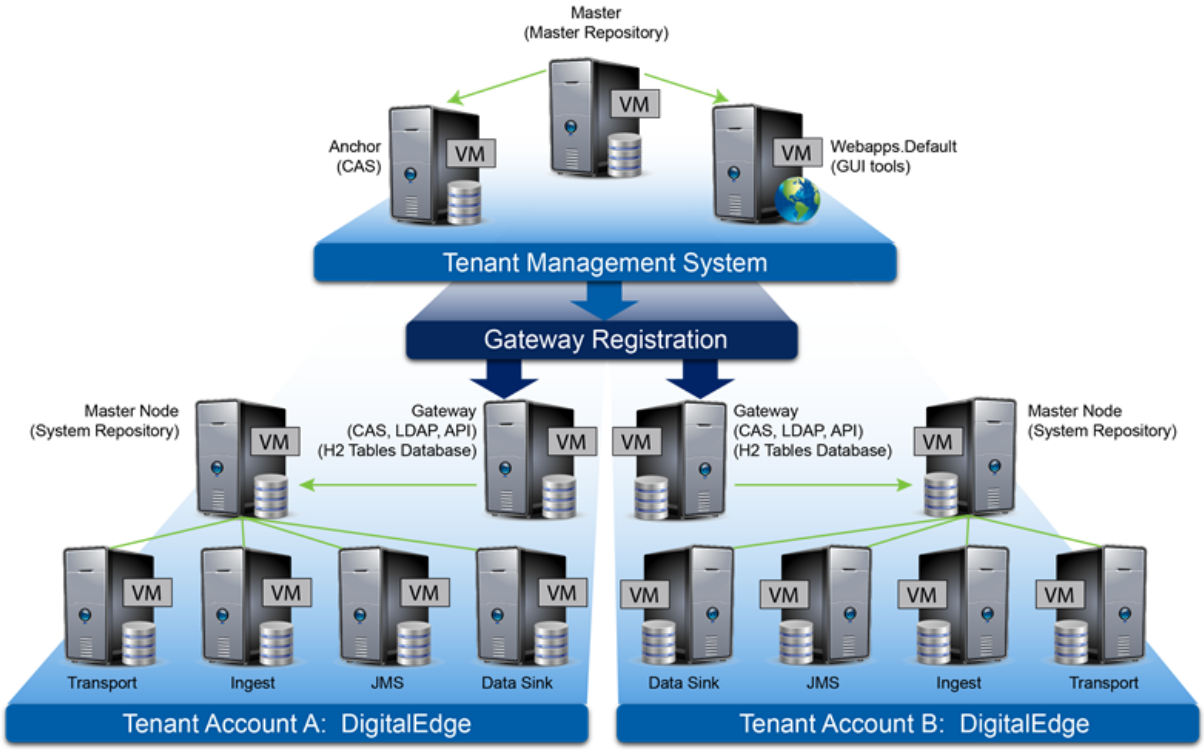
In the AWS™ public cloud, Leidos owns and controls TMS, and each tenant account represents an organization that has contracted with Leidos to implement one or more DigitalEdge systems.

A *primary tenant* is the first tenant (user) created in an account. The primary tenant owns all the resources: the system repository, LDAP, the tenant database, etc. One or more *secondary tenants* may be created in an account. All secondary tenants share the account resources that are owned by their primary tenant (system repository, LDAP, etc.), share and see all systems created under an account, and have the same privileges as the primary tenant. But secondary tenants have different logon credentials for security purposes. Leidos creates all tenant accounts, primary tenants, and secondary tenants for your system.

Logically, TMS is above the tenant accounts. TMS provides administrative services at an oversight level through the Management Console to:

- Create new tenant accounts
- Manage, set up, and start tenant applications
- Manage user identities
- Store and manage the DigitalEdge private components
- Navigate to other DigitalEdge tools
- View system logs
- Provide an additional level of security

From a high level perspective, TMS and tenant accounts interact as follows:



- TMS is launched at the AWS cloud level.
- The TMS Master node sets up and launches all the TMS nodes.
- Tthe DigitalEdge Administrator registers with DigitalEdge to configure a new tenant account the DigitalEdge Gateway.
- The DigitalEdge Administrator builds and starts up DigitalEdge systems.
- The tenant's Gateway node spawns the tenant's Master node for new systems.
- The tenant's Master node launches and manages all other nodes for DigitalEdge systems in the tenant account.

This diagram represents the initialization of a basic DigitalEdge system. Depending on the needs of a tenant's system, the DigitalEdge Administrator may also configure and start up:

- Multiple instances of transport, ingest, JMS, or data sink nodes
- Alerting node(s)
- Search node(s)
- User applications

Chapter 3: Prerequisites

The following components are required before you start to install the system and DigitalEdge. Check with Leidos for any items that you do not have in hand.

Hardware components

- Client PC to run and to access DigitalEdge

Software components

Before getting started, you will need the following software:

- Web browser recommendations: the latest version of Firefox (minimally, Extended Support Release version 17 or 24) or Chrome (minimally, version 21)
- Shockwave® (Flash®) Player: for DigitalEdge web applications
- AWS tenant account

Chapter 4: Installation

Getting started with DigitalEdge on a public cloud involves the following self-service steps. If at any time you need assistance, contact Leidos Technical Support.

See "Getting an Amazon account" on page 10

See "Signing up for DigitalEdge" on page 10

See "Getting the DigitalEdge image" on page 11

See "Registering with DigitalEdge" on page 11

See "Logging in" on page 13

Types of AWS environments

Amazon has offered several different Amazon Web Services™ (AWS) accounts which you can use as the cloud platform for DigitalEdge:

- **Amazon EC2-VPC account**: Any basic AWS account created after December 4, 2013 is an EC2-VPC account. The Elastic Compute Cloud™ (EC2™) is Amazon's primary web service where you "rent" virtual computers (in Amazon's data centers) for running DigitalEdge (or other applications). By default, all EC2 accounts are created in Amazon's Virtual Private Cloud™ (VPC™), an isolated environment within the AWS cloud where you can launch applications in a more secure, virtual network.
- **Amazon EC2-Classic account**: Prior to December 4, 2013, AWS accounts were created as EC2-Classic (Elastic Compute Cloud™) accounts, running in Amazon's public cloud. EC2-Classic accounts include the option to create a VPC system if it is needed; you can use the DigitalEdge System Builder tool to configure a VPC in these accounts. These accounts may also support EC2-VPC accounts in regions that you haven't yet used. See Amazon's explanation of these accounts for more specifics.

Please contact Leidos if you need guidance when signing up for an Amazon AWS account.

About VPC

When you sign up for an Amazon AWS account, your account is automatically created in the VPC (virtual private cloud) environment as an Amazon EC2-VPC account for a more secure system. Amazon VPC accounts provide security features such as security groups, network access control lists, filtering traffic at the instance and subnet levels, and restricting access to stored data in S3.

If you created an AWS account prior to 2014, those EC2-Classic accounts were created as non-VPC accounts by Amazon. EC2-Classic accounts include the option to create a VPC system if it is needed, by configuring the VPC with the DigitalEdge System Builder tool. See the DigitalEdge *Configuration Guide* for details.

Facts about VPC and DigitalEdge:

- To help protect sensitive data, a VPC system is isolated from your EC2 environment by a NAT (Network Address Translation) in the public subnet in each Availability Zone, which serves as a firewall into the private VPC subnet. Amazon also creates an Internet gateway and connects it to your VPC. The CIDR block for your VPC is 172.31.0.0/16. See Amazon's VPC documentation for more detailed specifications.
- A DigitalEdge system built in the VPC will run securely in the private subnet. The NAT will allow outbound traffic to access an external instance and will block inbound traffic from the VPC system.
- VPC only allows access via the following designated ports:
    - 443: routed to webapps.main
    - 8443, 5555, 1098, and 61515: routed to the Master node
- VPC uses its own security groups (identified with a vpc preface, such as vpc.internal.default, vpc.webapp.default, etc.)
- VPC has the potential of hosting 10 DigitalEdge systems, each with 1000 IPs
- NAT uses a Route Table to access the Internet; Amazon allows 10 entries in the Route Table

- Amazon provides 5 elastic IPs to start a VPC; you can request more elastic IPs from Amazon (use this request form) to increase the VPC's capacity to host 10 systems
- Each NAT is assigned an elastic IP which is publicly addressable, but DigitalEdge binds a system domain name to the elastic IP, allowing your users to access DigitalEdge by name
- You do not have to assign IP addresses in the VPC; IPs are assigned automatically by Amazon
- You will have one DigitalEdge Gateway for your Amazon account.

Getting an Amazon account

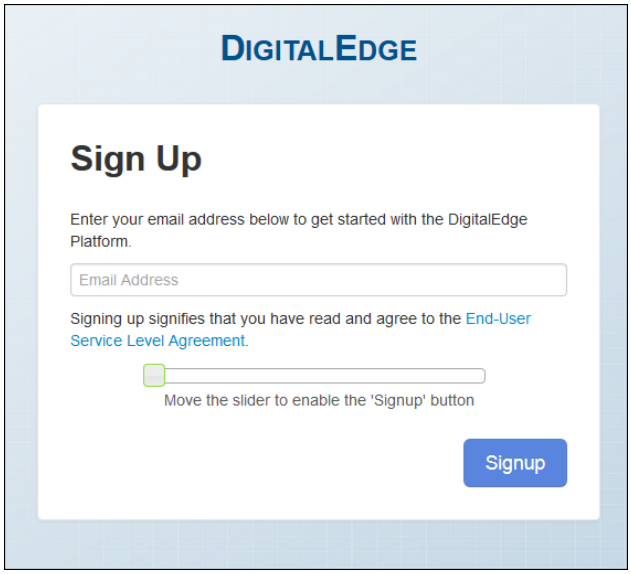
To get started with DigitalEdge in the public cloud, you need to obtain an Amazon Web Services™ (AWS) account. Follow these steps to:

1. Go to http://aws.amazon.com
2. Follow the Amazon instructions to **Sign Up** for a new account.

Signing up for DigitalEdge

Once you have an Amazon account, you can get DigitalEdge. Follow these steps to initiate the process:

1. Go to https://default.tms-dev.deleidos.com/signup/. The DigitalEdge **Sign Up** page appears.



2. Provide your email address so that Leidos can contact you with registration instructions.
3. Read the **End-User Service Level Agreement** and move the slider to the right.
4. Click **Signup**. The information is sent to Leidos. You will receive an email message with **DigitalEdge Registration Instructions**.

Getting the DigitalEdge image

Once you have signed up for DigitalEdge, you must acquire the DigitalEdge registration image from the AWS Marketplace:

1. Sign in to https://aws.amazon.com/marketplace/ with your Amazon account credentials.
2. Search for the DigitalEdge registration image.
3. Follow the Amazon instructions to launch the image in your account. When prompted, provide the requested user data listed in your **DigitalEdge Registration Instructions** email message (e.g., access key, account number, etc.).

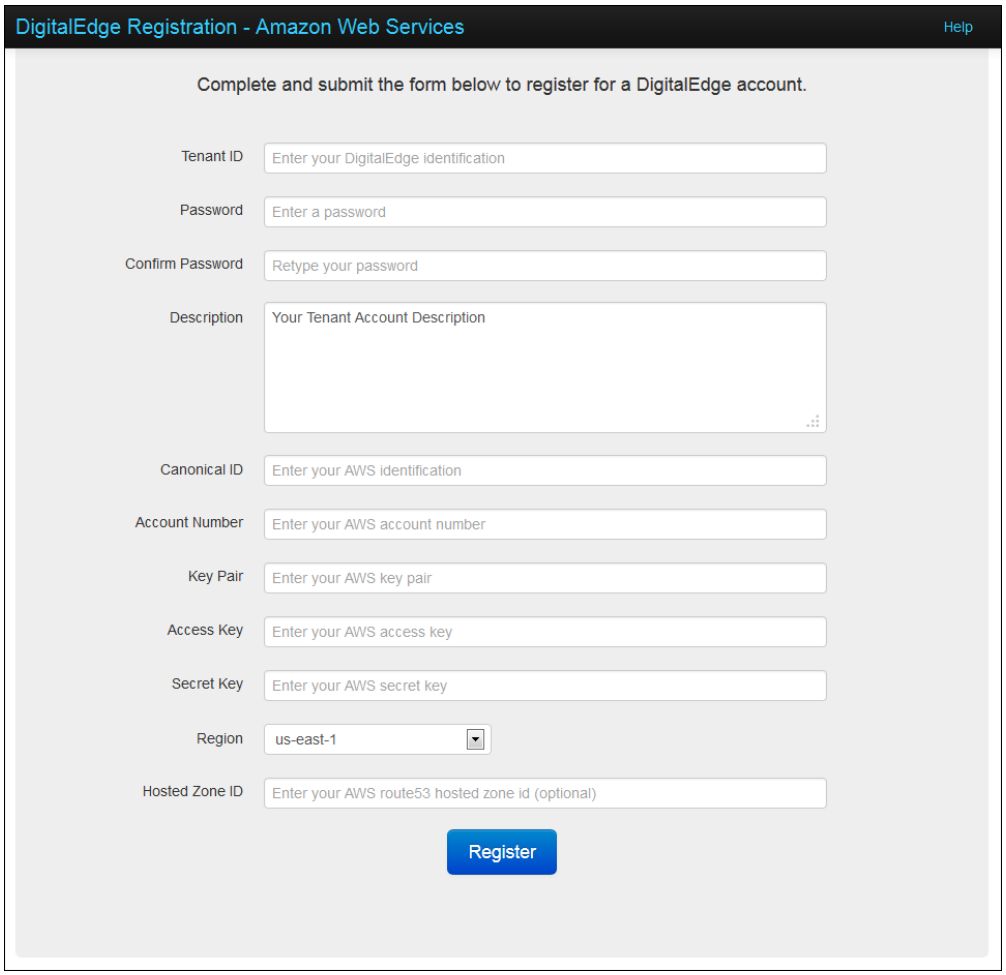
Registering with DigitalEdge

When you have launched the DigitalEdge image, follow these steps to register and to get started. This is primarily a self-service process:

1. Locate the instance's public DNS.
2. In a browser, access the following address, substituting your public DNS for `<digitaledge-registration-instance-public-dns>`:

   `https://<digitaledge-registration-instance-public-dns>/register`

3. Enter your UUID, which was provided in the DigitalEdge Registration Instructions email and click **Enter**. The DigitalEdge Registration Form appears.



4. Gather the following information, then fill in the form:

   Tenant ID: Your primary tenant log-on for the DigitalEdge **Management Console**

   Password: Your credentials for logging on to the DigitalEdge **Management Console**

   Description: Optionally provide a brief description of your DigitalEdge project.

   Canonical ID: Your AWS™ identification

   Account Number: Your AWS™ account, always a 12 digit number, such as: `197679631704`

   Key Pair: The key pair for your AWS™ account

   Access Key: The access key for your AWS™ account; part of your credentials for a new certificate (for SSH)

   Secret Key: The secret key for your AWS™ account; part of your credentials for a new certificate (for SSH)

   Region: The Amazon service region (data center locale) you are using

Hosted Zone ID: Optionally supply your Amazon Route 53™ hosted zone ID. DigitalEdge will use it to determine your domain name for all systems built in that account. If you do not enter an ID, the domain name will default to a Leidos domain name.

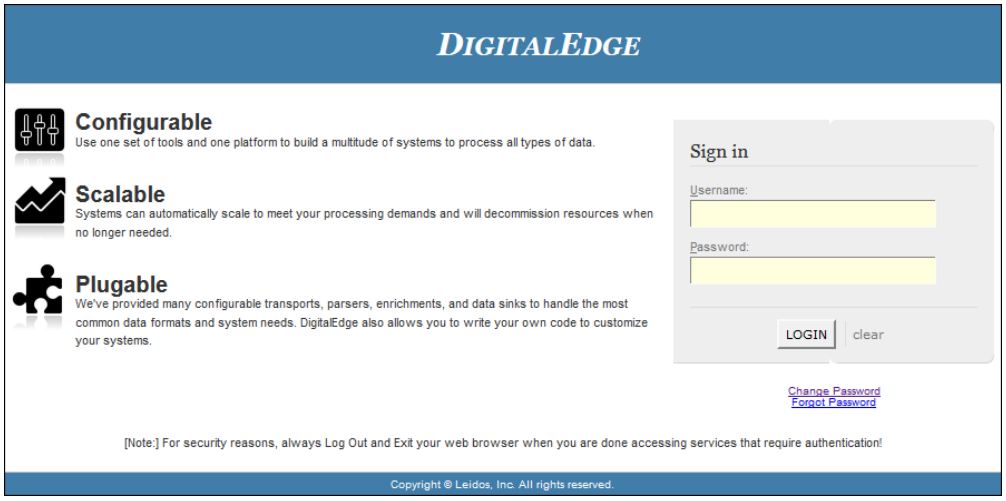
Credentials and keys that you enter on the Registration form stay in *your* AWS account; they are not copied outside your account or into a Leidos account.

8. Click Register. You will see a progress bar until registration is complete. Turnaround on a DigitalEdge registration and image building can take 30-45 minutes. If a problem occurs, contact DigitalEdge Technical Support.

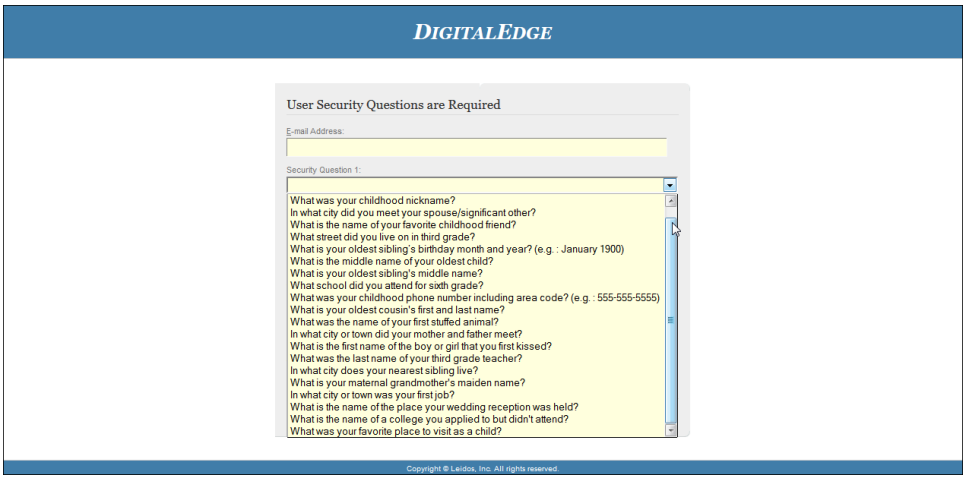
Logging in

Use this procedure to log on to the DigitalEdge **Management Console**.

1. In a web browser, go to `https://default.<system_domain_name>/tenantconsole`
2. Enter your **Username** and **Password**.
3. Click **LOGIN**.



4. The first time you log on to DigitalEdge, you must complete your registration:
   a. Supply your email address, and three security questions and answers to be used when confirming your identity:



b. Click **Continue**.

c. Read the DigitalEdge **License and Support Agreement** and click **ACCEPT**.

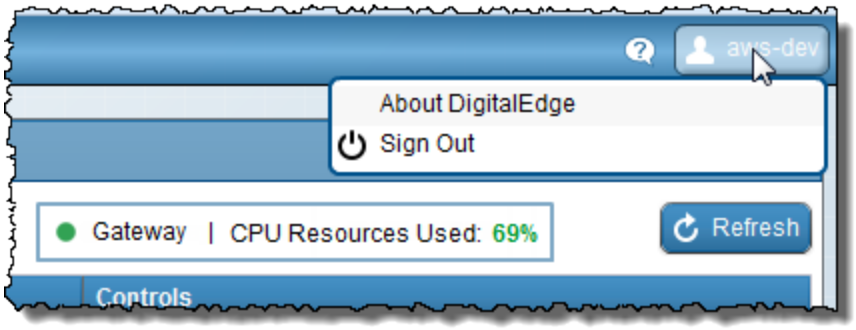
5. You can access all the DigitalEdge Setup and Runtime tools from the **Management Console**.

You cannot access the **Management Console** with an expired DigitalEdge license; contact Support for a new license.

Logging out

Use this procedure to log out of DigitalEdge.

1. Go to the **Management Console**.
2. Click the user icon in the upper right corner and select **Sign Out**.



Use the same procedure to log out of any Setup or Runtime UI tool.

When you **Sign Out** of one tool, all open tools are automatically signed out.

Chapter 5: Installing a Software Upgrade

As a DigitalEdge Administrator, you can install a new version of DigitalEdge when needed. This process, also known as a gateway restart, can be used to install a major release, minor release, or patches.

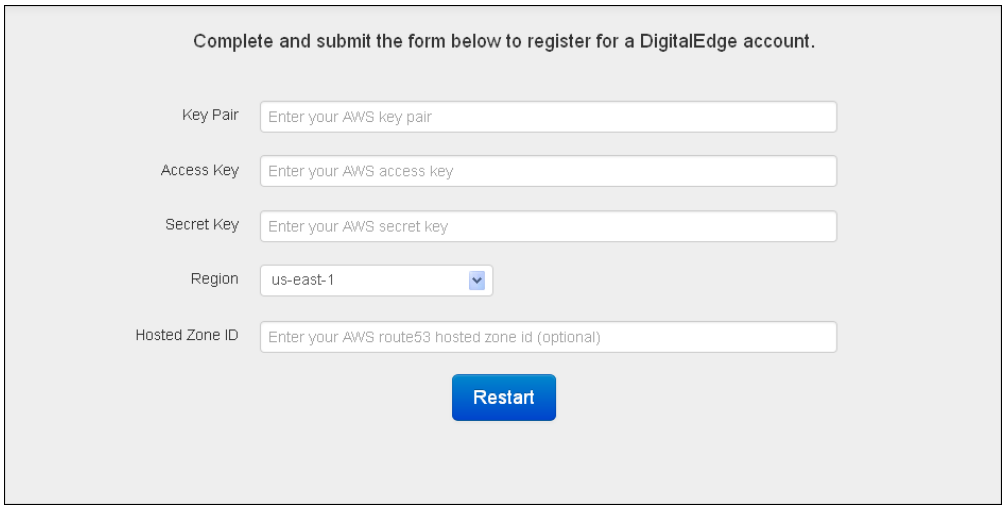
Prerequisites

- DigitalEdge is registered and in use
- DigitalEdge software upgrade notice from Leidos
- Your DigitalEdge system(s) can be up and running for most upgrades

When Leidos pushes a new software release to the DigitalEdge Amazon bucket, your version will be upgraded seamlessly. However, there are times when the Gateway is upgraded that you may want to **Stop** your systems in the **Management Console** temporarily if you are not worried about interrupting the data flow during a software upgrade. If you want to keep data flowing, you should stop the ingest node(s) temporarily. For each software update type, Leidos will provide detailed information about stopping and restarting systems.

Follow these steps to install a DigitalEdge upgrade:

1. Go to the Amazon Marketplace and locate the DigitalEdge application. The URL for a software upgrade is provided for you in the upgrade notice from Leidos.
2. Enter your UUID, which was provided in the DigitalEdge upgrade notice and click **Enter**. The DigitalEdge Restart Gateway form appears.



3. Complete the registration form with your current information:

   Key Pair: The key pair for your AWS™ account

   Access Key: The access key for your AWS™ account

   Secret Key: The secret key for your AWS™ account

Region: The Amazon service region (data center locale) you are using

Hosted Zone ID: Your hosted zone ID from Amazon's Route 53™ DNS web service

Credentials and keys that you enter on the Restart Gateway form stay in *your* AWS account; they are not copied outside your account or into a Leidos account.

4. Click Restart. You will see a progress bar until the restart is complete. The upgrade process can take 30-45 minutes. If a problem occurs, contact DigitalEdge Technical Support.
5. When the restart is complete, access the DigitalEdge **Management Console** to verify that the Gateway status is green.
6. Access **System Builder**. For each system you have built, enter the new **Software Version** number and click **Build** to rebuild the system with the new software.
7. Go to the **Management Console**. For each system, **Stop** and **Start** each system.

Appendix A: Terminology

Term	Definition
Alert	Email notification to a user that a potential fraud has been detected by DigitalEdge
AMI (Amazon Machine Image)	A bootable server that is a special type of pre-configured virtual machine in the cloud; AMIs serve as basic units of service deployment
Anchor node	In TMS, the anchor node hosts CAS, LDAP, and the TMS database
AWS™ (Amazon Web Services™)	Remote web services that comprise the cloud computing platform offered by Amazon and implemented in the Eucalyptus platform
Data sink	A queue, server, or database that can receive pipeline-processed JSON data to store or post-process for other uses
Dead letter queue	If an incoming record cannot be parsed for any reason, rather than ignoring it and dropping it out of the system, DigitalEdge saves the record in the dead letter queue where you can examine it and correct it.
EC2™ (Amazon Elastic Compute Cloud™)	A key part of Amazon's AWS™ public cloud computing platform, providing users the ability to create, launch, and end virtual server instances in a scalable deployment of applications
Elastic IP	A static IP address designed for cloud computing, associated with your Amazon account, not an instance; EC2 lets you mask problems by remapping an elastic IP address to a replacement instance
EMI (Eucalyptus Machine Image)	A pre-configured virtual machine, including the operating system and virtual application software, that can be used to create an instance in a Eucalyptus environment
Gateway node	A node in a tenant system that hosts CAS for single sign on permissions and LDAP for user account credentials. The Gateway node starts and stops systems, creates and deletes systems and security groups, and synchronizes components.
Hybridfox	An optional Firefox add-on that provides an interface to cloud accounts, including AWS and Eucalyptus, to help you manage images, instances, security groups, key pairs, elastic IPs, and storage.

Term	Definition
IAAS	The Infrastructure As a Service (IAAS) number is a tenant ID assigned by the facility providing your cloud services (e.g., Amazon or Eucalyptus). In the Eucalyptus Console, it is identified under **Identity Management** > **Accounts** > **ID#**. It is always a 12 digit number. In the DigitalEdge Management Console, it is listed as the account ID.
Master node	A VM (virtual machine) that launches all other nodes in a system. The master node handles auto-scaling, internal monitoring, starting and stopping for all instances. In TMS, the master node includes the Master Repository. In a tenant account, the master node includes the System Repository.
Master Repository	The Master Repository resides in TMS. It is the storage location for all common plug-in components provided with DigitalEdge, and private plug-in components used by each tenant account.
NAT (Network Address Translation)	An instance which is configured to perform network address translation and to serve as a firewall into the private Amazon VPC subnet
POC	Point of Contact information
Private IP	An internal RFC 1918 address that is only routable within the EC2 Cloud; traffic outside your EC2 network cannot access this IP
Public IP	An Internet routable IP address assigned by the system for all instances; Traffic routed to a public IP is translated via NAT and forwarded to an instance's private IP address
Repository	The storage location for all the plug-in components; the System Repository stores private components used in a tenant's DigitalEdge account, the Master Repository resides at the TMS level and stores all common and private components
S3™ (Amazon Simple Storage Service™)	The online storage web service provided with AWS™ and used as a data source for public cloud instantiations
Splitter	Each transport works with a specific incoming record type (JSON, XML, PCAP, etc.); the tranport's record-format parameter uses a splitter to define record boundaries when the input data includes multiple records
Tenant account	A tenant is an account on a cloud platform. In the public cloud, a tenant account typically represents an organization that is building an AWS application. On a private cloud,

Term	Definition
Primary tenant Secondary tenant	internal to an organization, a tenant account is usually a project or a department that runs its own secure applications. A *primary* tenant is the first tenant created in a DigitalEdge account (via the Installation program on a Eucalyptus system, via Registration on AWS systems). The primary tenant owns all the DigitalEdge resources: the system repository, LDAP, the tenant database, etc. and does not share data with other tenants. One or more *secondary* tenants may be created in an account. A secondary tenant is created by a TMS Administrator in the Management Console. All secondary tenants share the account resources that are owned by their primary tenant (system repository, LDAP, etc.), share and see all systems created under an account, and have the same privileges as the primary tenant. But secondary tenants have different logon credentials for security purposes.
TMS (Tenant Management System)	The Tenant Management System is a behind-the-scenes infrastructure for DigitalEdge to create and manage tenant accounts. TMS provides services to create new accounts, to monitor tenant applications, to calculate tenant usage activity and charges, to manage user identities and permissions, to manage the DigitalEdge GUI tools and plug-in components, and to provide security.
VPC™ (Amazon Virtual Private Cloud™)	An isolated environment within the AWS cloud where you can launch applications in a more secure, virtual network

Appendix B: What DigitalEdge Registration Does

DigitalEdge Registration does most work behind the scenes with minimal manual intervention. Here is what it does:

- Locates your tenant account information in TMS
    - Creates a DigitalEdge tenant account in TMS
    - Creates a primary tenant login account in TMS
- Creates EC2 security groups
- Creates the initial VPC environment, including the VPC and the public subnet
    - Creates the VPC security groups
- Creates a new AMI and registers it with TMS
    - Downloads several required third-party components
- Builds a tenant Gateway system configuration
    - Uploads it to the TMS S3 bucket
    - Grants read permission to your tenant account for this configuration
- Launches a new Gateway instance with the latest software release

Appendix C: What a Software Upgrade Does

The DigitalEdge software upgrade does most work behind the scenes with minimal manual intervention. Here is what it does:

- Locates your tenant account information in TMS
- Creates a new AMI (if needed) and registers it with TMS
    - Downloads several required third-party components
- Creates any missing security group(s)
- Builds a tenant Gateway system configuration
    - Uploads it to the TMS S3 bucket
    - Grants read permission to your tenant account for this configuration
- Terminates your Gateway instance
- Launches a new Gateway instance with the latest software release

Appendix D: What Each Node Does

Each node in DigitalEdge is a virtual machine and an instance of a process group, most of which are auto-scaling. To help size a system, the following table provides details about what each node does.

Node	Content
webapps.main (on TMS)	Home to all the DigitalEdge APIs, Setup tools, and Runtime tools, including: • Management Console • Data Modeler • Table Manager • System Builder • System Monitor
anchor (on TMS)	Security and authentication node, housing: • CAS (JA-SIG Central Authentication Service) • LDAP • TMS database • TMS Gateway
gateway	The node that controls a DigitalEdge system, including: • Launching the master node • Starting and stopping systems • Creating and deleting systems and security groups • Synchronizing components and repositories • Housing: ◦ CAS for single sign-on permissions ◦ LDAP for user account credentials ◦ APIs ◦ Tenant database
master	The management node of DigitalEdge, controlling: • Starting and stopping all instances • Monitoring for auto-scaling • Adding and removing nodes based on load and storage utilization • Handling virtual storage allocations

Node	Content
	• Gathering metrics for auto-scaling decisions • Housing the System Repository
transport	Controlling all transports through the Transport API
jms.external	First entry point into DigitalEdge, and a staging area for incoming data to: • Receive data pushed into the jms.external queue by other clients • Feed data directly into DigitalEdge • Manage the parsing queue • Receive processed alerts from the datasink.alert that match alerting criteria, and place a message in this queue for notifications
ingest.all	Ingest node to handle processing pipeline tasks, including: • Parsing • Enrichment
jms.internal	Internal staging area for the next steps in the processing pipeline; a buffer for records queued up waiting for the next phase of processing: • Post-enrichment record holding • Temporary record storage
datasink	Each type of data sink has its own node and processes data for specialized uses; for example: • datasink.alert - filtering records against alert criteria, sending alert messages to the configured recipient (such as a topic on the jms.external node, an email message, etc.) • datasink.hbase - storing records to the Hadoop Distributed File System (HDFS) • datasink.hive - storing records to HDFS • datasink.lucene - indexing records for searching • datasink.mongodb - storing JSON-based records and providing a query interface Some data sinks automatically add additional nodes when they are spawned; for example, HBase and Hive add nodes (such as zookeeper) that are needed for a complete HBase ecosystem

Node	Content
webapps.main (on tenant)	Home to all webapps and REST APIs, including: • Search app • Metrics API

Index