

1. Wandisco client on-prem install or create wandisco VM with test hadoop cluster if no on prem
2. Request wandisco support and Microsoft to whitelist subscription ID
3. Wandisco VM / RG / NIC / P.Endpoint / Public IP / Live Data migrator approval
4. Deploy live data migrator
 - a. Host VM on Azure
 - b. Host VM on-premise
 - i. wget <https://wandisco.com/downloads/livedata-migrator.sh>
 - ii. sudo su -
 - iii. chmod +x livedata-migrator.sh && ./livedata-migrator.sh
 - iv. systemctl status livedata-migrator
 - v. systemctl edit livedata-migrator
 - vi. [Service]

LimitNOFILE=64000
 - vii. systemctl restart livedata-migrator
 - viii. /etc/systemd/system/livedata-migrator.service.d/override.conf
5. Wandisco:8081 URL to be whitelisted
6. Source (azure blob) / target (hdfs – onprem) configuration

Component	Default system user	Default system user group
Data Migrator	hdfs	hdfs
UI	hdfs	hdfs
Hive Migrator	hive	hadoop

[Home](#) > [Marketplace](#) > [Data Migrator for Azure](#) >

Create LiveData Migrator ...

Basics Tags Review + create

Project details

Select the subscription to manage deployed resources and costs. Use resource group to manage all your resources.

Subscription * ⓘ

Pay-As-You-Go

Resource group * ⓘ

NetworkWatcherRG

[Create new](#)

Instance details

Name *

wandisco

Region * ⓘ

East US

Do you need a Hadoop test cluster? * ⓘ

☒ Yes

☐ No

Cluster Admin Username *

narenwandisco

Cluster Admin Password * ⓘ

.....

Confirm password * ⓘ

.....



wandisco.ldma-20240525153037 | Overview

Deployment



Delete



Cancel



Redeploy



Download



Refresh



Overview



Inputs



Outputs



Template



Subscription ID not in allowed list. Click here for details



Deployment name : wandisco.ldma-20240525153037

Subscription : Pay-As-You-Go

Resource group : NetworkWatcherRG

Start time : 5/25/2024, 3:34:09 PM

Correlation ID : 900515fe-4937-4533-a



Deployment details

Resource	Type	Status
wandisco-kkdhhkgsie5qg-vm	Virtual machine	OK
wandisco-kkdhhkgsie5qg-ni	Network interface	Created
wandisco	LiveData Migrator	BadRequest (Error details)
wandisco-kkdhhkgsie5qg-nsg	Network security group	OK
wandisco-kkdhhkgsie5qg-pubip	Public IP address	OK
wandisco-kkdhhkgsie5qg-vnet	Virtual network	OK

<input type="checkbox"/> wandisco-kkdhhkgsie5qg-vnet	Virtual network	East US
<input type="checkbox"/> wandisco-kkdhhkgsie5qg-vm	Virtual machine	East US
<input type="checkbox"/> wandisco	Storage account	East US
<input type="checkbox"/> wandisco-kkdhhkgsie5qg-pubip	Public IP address	East US
<input type="checkbox"/> NetworkWatcher_eastus	Network Watcher	East US
<input type="checkbox"/> wandisco-kkdhhkgsie5qg-nsg	Network security group	East US
<input type="checkbox"/> wandisco-kkdhhkgsie5qg-ni	Network Interface	East US
<input type="checkbox"/> wandisco-kkdhhkgsie5qg-vm_OsDisk_1_974cf75d32634605bd...	Disk	East US

Subject: Request for Subscription Whitelisting for WANdisco Service on Azure

Dear WANdisco Support Team,

I am writing to request that my Azure subscription ID be added to the allowed list for accessing the WANdisco service on Azure. I am currently encountering the following error when attempting to create a WANdisco resource: "Subscription ID not in allowed list (Code: ValidationFailure)".

Here are my subscription details:

- Subscription ID: [Your Subscription ID]
- Subscription Name: [Your Subscription Name]
- Contact Email: [Your Contact Email]

I would appreciate your assistance in adding my subscription ID to the allowed list or providing any additional steps I need to take to gain access.

Thank you for your support.

Best regards,

[Your Name]

[Your Contact Information]

wandisco-kkdhkgsie5qg-vm ✨ ☆ ...
Virtual machine

Search 🔍 ⌵ ⏪

Connect 🔗 ⌵ Start ▶ Restart ↺ Stop □ Hibernate ⌚ Capture 📷 Delete 🗑️ Refresh ↻ O

Overview

- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems
- Connect
 - Connect 🔗
 - Bastion 🔗
- Networking
- Settings
- Availability + scale

Essentials

Resource group (move) NetworkWatcherRG	Operating system Linux (ubuntu 18.04)
Status Running	Size Standard D8 v3 (8 vcpus, 32 GiB memory)
Location East US	Public IP address 52.170.239.206
Subscription (move) Pay-As-You-Go	Virtual network/subnet wandisco-kkdhkgsie5qg-vnet/default
Subscription ID 31ff243b-0053-4e01-b5d9-186e31b8a4e4	DNS name Not configured
	Health state -
	Time created 5/25/2024, 10:04 AM UTC

```
narenwandisco@wandisco-ldma:~$ ls -lrt
total 16
-rwxrwxr-x 1 narenwandisco narenwandisco 15602 May 25 17:56 livedata-migrator.sh
narenwandisco@wandisco-ldma:~$ df -h
Filesystem      Size  Used Avail Use% Mounted on
udev            2.0G   0    2.0G   0% /dev
tmpfs           394M  1.1M  393M   1% /run
/dev/sda1       248G  21G   227G   9% /
tmpfs           2.0G  16K   2.0G   1% /dev/shm
tmpfs           5.0M   0   5.0M   0% /run/lock
tmpfs           2.0G   0   2.0G   0% /sys/fs/cgroup
/dev/sda15      105M  6.7M   98M   7% /boot/efi
/dev/sdb1       7.9G  36M   7.4G   1% /mnt
overlay         248G  21G   227G   9% /var/lib/docker/overlay2/4b492c3889bd398523a59ebf
tmpfs           394M   0   394M   0% /run/user/1012
tmpfs           394M   0   394M   0% /run/user/1004
tmpfs           394M   0   394M   0% /run/user/1001
tmpfs           394M   0   394M   0% /run/user/113
```

[Install Data Migrator | Data Migrator \(wandisco.com\)](#)

```
ator.shndisco@wandisco-ldma:~$ wget https://wandisco.com/downloads/livedata-migra
--2024-05-25 17:56:33-- https://wandisco.com/downloads/livedata-migrator.sh
Resolving wandisco.com (wandisco.com)... 45.79.162.226
Connecting to wandisco.com (wandisco.com)|45.79.162.226|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 15602 (15K) [application/file]
Saving to: 'livedata-migrator.sh'

livedata-migrator.s 100%[=====>] 15.24K --.-KB/s in 0s

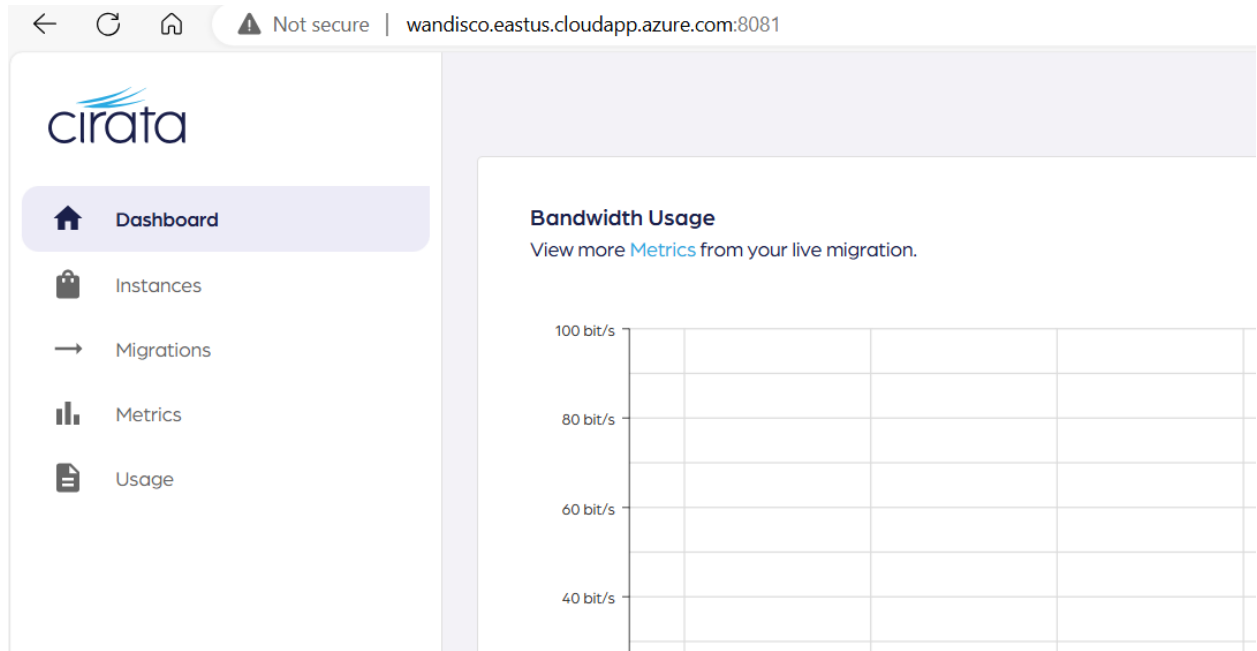
2024-05-25 17:56:33 (189 MB/s) - 'livedata-migrator.sh' saved [15602/15602]
```





```
ata-migrator.shdma:/home/narenwandisco# chmod +x livedata-migrator.sh && ./liveda
A supported version of Java was detected on this host, installation will continue.
OK
Hit:1 http://azure.archive.ubuntu.com/ubuntu bionic InRelease
Hit:2 http://azure.archive.ubuntu.com/ubuntu bionic-updates InRelease
Hit:3 http://azure.archive.ubuntu.com/ubuntu bionic-backports InRelease
Hit:4 http://security.ubuntu.com/ubuntu bionic-security InRelease
Hit:5 https://packages.microsoft.com/repos/microsoft-ubuntu-bionic-prod bionic InRelease
Ign:6 http://pkg.wandisco.com/apt apt InRelease
Get:7 http://pkg.wandisco.com/apt apt Release [2090 B]
Get:8 http://pkg.wandisco.com/apt apt Release.gpg [659 B]
Get:9 http://pkg.wandisco.com/apt apt/main amd64 Packages [18.9 kB]
Fetched 21.7 kB in 1s (30.0 kB/s)
Reading package lists... Done
Reading package lists... Done
Building dependency tree
Reading state information... Done
Selected version '1.28.4-5064' (Official WANdisco APT Repository:pkg.wandisco.com [all]) for 'livedata-migrator'
Selected version '1.20.2-545' (Official WANdisco APT Repository:pkg.wandisco.com [all]) for 'livedata-migrator-cli'
Selected version '2.0.2-1972' (Official WANdisco APT Repository:pkg.wandisco.com [all]) for 'hivemigrator'
Selected version '13.11.2-4349' (Official WANdisco APT Repository:pkg.wandisco.com [all]) for 'livedata-ui'
The following NEW packages will be installed:
  hivemigrator livedata-migrator livedata-migrator-cli livedata-ui
0 upgraded, 4 newly installed, 0 to remove and 207 not upgraded.
Need to get 1550 MB of archives.
After this operation, 1698 MB of additional disk space will be used.
Get:1 http://pkg.wandisco.com/apt apt/main amd64 hivemigrator all 2.0.2-1972 [599 MB]
Get:2 http://pkg.wandisco.com/apt apt/main amd64 livedata-migrator all 1.28.4-5064 [795 MB]
Get:3 http://pkg.wandisco.com/apt apt/main amd64 livedata-migrator-cli all 1.20.2-545 [41.2 MB]
Get:4 http://pkg.wandisco.com/apt apt/main amd64 livedata-ui all 13.11.2-4349 [114 MB]
Fetched 1550 MB in 40s (38.7 MB/s)
```

```
-account-name wandisco --subscription 31ff243b-0053-4e01-b5d9-186e31b8a4e4show --
Subscription '31ff243b-0053-4e01-b5d9-186e31b8a4e4' not recognized.
Please run 'az login' to setup account. hdfs dfs -ls wasbs://wandisco@wandisco.blob.core.windows.net/ wa
24/05/25 17:46:14 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.propertiesows.net/
24/05/25 17:46:14 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
24/05/25 17:46:14 INFO impl.MetricsSystemImpl: azure-file-system metrics system startedt-name wandisco
24/05/25 17:46:15 INFO impl.MetricsSystemImpl: Stopping azure-file-system metrics system...
24/05/25 17:46:15 INFO impl.MetricsSystemImpl: azure-file-system metrics system stopped.
24/05/25 17:46:15 INFO impl.MetricsSystemImpl: azure-file-system metrics system shutdown complete.
windows.net/o@wandisco-ldma:~$ hdfs dfs -ls wasbs://wandisco@wandisco.blob.core.w
24/05/25 17:52:02 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
24/05/25 17:52:03 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
24/05/25 17:52:03 INFO impl.MetricsSystemImpl: azure-file-system metrics system started
Found 1 items
-rwxrwxrwx 1 56850 2024-05-25 17:51 wasbs://wandisco@wandisco.blob.core.windows.net/jenkins-plugin.png
24/05/25 17:52:03 INFO impl.MetricsSystemImpl: Stopping azure-file-system metrics system...
24/05/25 17:52:03 INFO impl.MetricsSystemImpl: azure-file-system metrics system stopped.
24/05/25 17:52:03 INFO impl.MetricsSystemImpl: azure-file-system metrics system shutdown complete.
narenwandisco@wandisco-ldma:~$
```



Essentials


Resource group (move)	: NetworkWatcherRG	SKU	: Basic
Location (move)	: East US	Tier	: Regional
Subscription (move)	: Pay-As-You-Go	IP address	: 52.170.239.206
Subscription ID	: 31ff243b-0053-4e01-b5d9-186e31b8a4e4	DNS name	: wandisco.eastus.cloudapp.azure.com
		Domain name label sco...	: -



 0    naren
narenguy@gmail.com
Admin

Filesystems

 Couldn't detect your filesystem. Configure a [filesystem](#) now.

Name	Type	Status	Location
------	------	--------	----------

0


N

naren
narenguy@gmail.com
Admin

Source Filesystem Configuration

* Required fields

Filesystem Type

 Hadoop Distributed File System (HDFS)

Display Name *

hdfs

Enter a display name for this filesystem.

Default Filesystem *

hdfs://narenwandisco1:8020

Enter the filesystem address. For example, hdfs://nameservice:8020.

Kerberos Configuration

If your filesystem uses Kerberos, enter the following:
1. Your Kerberos principal and keytab.
2. The path(s) to your HDFS configuration files in the **Configuration Property File Paths** field under **Additional Configuration**.
Note: Ensure the path(s) contains the core-site.xml and hdfs-site.xml files for this Hadoop cluster. [Learn more](#).

Kerberos Principal

narenwandisco@cloudapp.azure.com

For example, username@EXAMPLE.COM.

Kerberos Keytab Location

/etc/security/keytabs/hdfs.headless.keytab.

I - localhost


Overview

Metrics

Diagnostics

Path Status

figuration ▾

ystems & Agents  ▴

Filesystems

Metastore Agents

Data Transfer Agents

ctions ▾

Target Filesystem Configuration

* Required fields

Filesystem Type

 Azure Data Lake Storage (ADLS) Gen2

Display Name *

blob

Enter a display name for this filesystem.

Data Lake Storage Endpoint

wandisco.dfs.core.windows.net

Overwrite the default endpoint by entering another here. [Learn more](#).

Authentication Type

Service Principal (OAuth2)



Enter your Azure storage account details. [Learn more](#).

Account Name *

wandisco

Container Name *

wandiscoc1

Client ID *

Secret *

OAuth2 Endpoint *

https://login.microsoftonline.com/30c1a35d05aa4afabfc1706f6e29d24b/oauth2/v2.0/token

For example, https://login.microsoftonline.com/TENANT/oauth2/v2.0/token. [Learn more](#).



Azure Data Lake Storage (ADLS) Gen2

Display Name *

blob

Enter a display name for this filesystem.

Data Lake Storage Endpoint

databricks-rg-workspace1-pnow4gmllly6ds.dfs.core.windows.net

Overwrite the default endpoint by entering another here. [Learn more.](#)

Authentication Type

Shared Key



Enter your Azure storage account details. [Learn more.](#)

Account Name *

wandisco

Access Key *

.....

Container Name *

wandiscoc1



Use Secure Protocol

SOURCE HDFS ON-PREM

Target	<u>One-time migration</u>	<u>Live migration</u>	<u>Recurring migration</u>
Alibaba Cloud Object Storage Service	✓	✓	✓
Amazon S3	✓	✓	✓
Azure Data Lake Storage Gen2	✓	✓	✓
Google Cloud Storage	✓	✓	✓
Hadoop Distributed File System*	✓	✓	✓
IBM Cloud Object Storage	✓	✓	✓
Local Filesystem	✓	✓	✓
Oracle Object Storage	✓	✓	✓
S3	✓	✓	✓

SOURCE: ADLS

Target	<u>One-time migration</u>	<u>Live migration*</u>	<u>Recurring migration</u>
Alibaba Cloud Object Storage Service	✓		✓
Amazon S3	✓		✓
Azure Data Lake Storage Gen2	✓	✓	✓
Google Cloud Storage	✓		✓
Hadoop Distributed File System*	✓		✓
IBM Cloud Object Storage	✓		✓
Local Filesystem	✓		✓
Oracle Object Storage	✓		✓
S3	✓		✓

Diagnostics

Path Status

Configuration

Filesystems & Agents

Filesystems

Metastore Agents

Data Transfer Agents

Migrations

Data

Data Migrations

Retries

Path Mappings

Exclusion Templates

Metadata

Metadata Rules

Metadata Migrations

Migration Status

Live & Recurring: 0

In progress & Complete: 0

Stopped: 0

Inactive: 0

Migration Priority

High: 0

Normal: 0

Low: 0

Running Migrations

Select how many migrations can run simultaneously, excluding the ones in a live or recurring state.

Running Migration Limit

2

0 indicates no limit.

Data Migrations

+ Create migration

Bulk Action

Filter & Search

Supported migrations for the selected action are shown below.

Start

Filter by Source

Filter by Target

Filter by Group

Search

To create a migration, add a source and target filesystem.

Filesystem Options

☐ Live Migration

Changes made to the source filesystem will be migrated in real time using the notification system defined for this storage. [Learn more.](#)

☒ One-time Migration

Existing data on the source will be moved to the target, after which the migration will be complete. [Learn more.](#)