

How to copy data from one HDFS cluster to another HDFS cluster?

Using **hadoop distcp** command we can copy data from one HDFS Cluster to another HDFS cluster

Go to prod1 env NameNode VM (Ex: 10.0.0.6) run this bellow command to copy from one HDFS Cluster to another HDFS cluster

Syntax:

```
sudo -u hdfs hadoop distcp <Source NameNode metadata service/Source Path> <Destination NameNode metadata service/Destination Path>
```

Example:

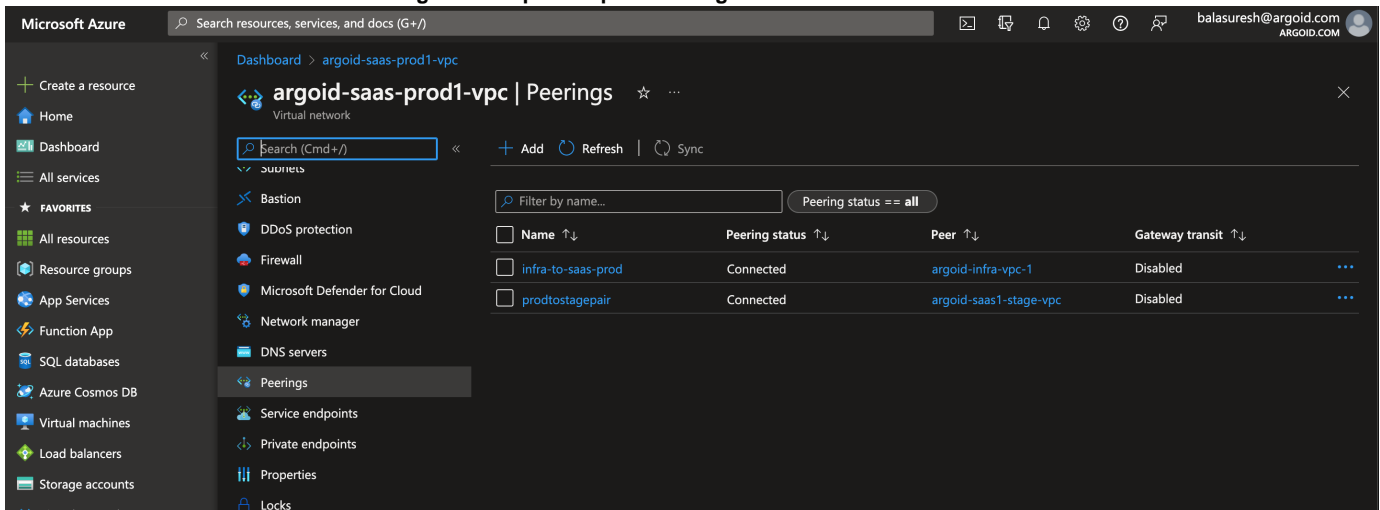
```
sudo -u hdfs hadoop distcp hdfs://10.0.0.6:8020/data/azadea/prod/kw/ingestion/entities/users/version=1/date=2022-07-06/hour=06 hdfs://192.0.1.6:8020/tmp/amith/data/azadea/prod/kw/ingestion/entities/users/version=1/date=2022-07-06/hour=06
```

Note:

To Copy data from one HDFS cluster to another HDFS cluster connectivity should be there (Both clusters should be accessible)

In our case we have enabled VPC pairing so both clusters can communicate with each other

Azure Dashboard > Virtual Networks > argoid-saas-prod1-vpc > Peerings



Name	Peering status	Peer	Gateway transit
infra-to-saas-prod	Connected	argoid-infra-vpc-1	Disabled
prodtostagepair	Connected	argoid-saas1-stage-vpc	Disabled

Azure Dashboard > Virtual Networks > argoid-saas1-stage-vpc > Peerings

Microsoft Azure

Search resources, services, and docs (G+ /)

stage

1/2

lasuresh@argoid.com
ARGOID.COM

Create a resource

Home

Dashboard

All services

FAVORITES

Resource groups

App Services

Function App

SQL databases

Azure Cosmos DB

Virtual machines

Load balancers

Storage accounts

Virtual networks

Azure Active Directory

Monitor

Dashboard > Virtual networks > argoid-saas1-stage-vpc

argoid-saas1-stage-vpc | Peerings

Virtual network

Search (Cmd+ /)

+ Add Refresh Sync

Filter by name...

Peering status == all

Name ↑↓	Peering status ↑↓	Peer ↑↓	Gateway transit ↑↓	
infra-to-saas-stage	Connected	argoid-infra-vpc-1	Disabled	...
stagevpc	Connected	argoid-saas-prod1-vpc	Disabled	...

Settings

Address space

Connected devices

Subnets

Bastion

DDoS protection

Firewall

Microsoft Defender for Cloud

Network manager

DNS servers

Peerings

Service endpoints

Private endpoints