

# Data Avenue referencia architektúra az ELKH Cloud-on



Rusznák Attila SZTAKI

## Bemutatkozás

## Rusznák Attila

- SZTAKI PERL fejlesztő
- ÓE NIK oktató
- SZTE tanszéki mérnök



## Data Avenue

## Mi az a Data Avenue?



- Egy grafikus felülettel rendelkező tároló menedzsment szolgáltatás
  - Számos protokollt támogat: sFTP, Amazon S3, Apache HDFS
  - > Az adatokat segítségével egyik szerverről a másikra mozgathatjuk
- Támogatja az olyan általános műveleteket, mint:
  - könyvtárak és fájlok listázása
  - könyvtárak létrehozása
  - ► törlés
  - átnevezés
  - fájlok feltöltése és letöltése
- REST API támogatottság
- Nyílt forráskódú (https://github.com/SZTAKI-LPDS/data-avenue)

🗧 😑 🔤 Data Avenue	× +										
← → C ▲ Not Secure   192.1	68.154.51:8080/	dataavenue/u	i/			07	☆	0		A	0
Main Uploads Transfers	Settings	About									
s3://192.168.154.2:80/imi-test	1	Go	URI (e.g	∫.: s3://av	√s.amazoi	n.com	/bucl	ket/	)	Go	
Name					Name	е					
<b>i</b>											
1000file/											
<b>1111</b> /											
2/											
source/											
target/											
upload/											
README.md											
SanDiskSecureAccessV3.01_win.e	xe										
mytest.csv											
train-images-idx3-ubyte.gz											
Download Copy Move	MkDir Delet	e Rename	Upload:	Choose 1	Files No	file	cho	sen			
	/					- <u>/</u>					

ELKH Cloud



nyomonkövethető.

## A Data Avenue előnyei

- **gyors** (többszálú, Java alpú, több ezer dir bejegyzést képes kezelni)
- hatékony (third-party transzferek, nincs lemezterület igény)
- **robosztus** (évekig futtatható újraindítás vagy erőforráskiesés nélkül)
- **biztonságos** (nem tárol érzékeny adatokat, támogatja a biztonságos kapcsolatot)
- **hibatűrő** (hiba esetén automatikusan helyreáll)
- **könnyed GUI** (javascript/angular alapú felület, a legtöbb böngésző támogatja)
- > nyomonkövethező (naplózza az átviteleket)

Az 1.0-ás verzió 2020 május 7-én jelent meg!

C) Sei	arch or jump to		7 Pull	requests Issu	es Marketplace	Explore			¢
		Learn	Git and	d GitHul	without	any co	ode!		
	Using t	he Hello World	guide, you'l	I start a bran	ch, write com	ments, and	open a pul	I reques	st.
	Using t	he Hello World	guide, you'l	ll start a bran Read the	ch, write com guide	ments, and	open a pul	ll reques	st.
	Using ti	he Hello World	guide, you'l	ll start a bran Read the	ch, write com guide	ments, and	open a pul	l reques	st.
SZTAKI	Using ti -LPDS / data	ne Hello World	guide, you'	I start a bran Read the	ch, write com	onents, and © Unw	open a pul	t reques €	o

# A Data Avenue referencia architektúra bemutatása

# A Data Avenue architektúrája

## Három fő komponensből áll:

- Data Avenue szerver
  - > Apache Tomcat-ben fut, mint "dataavenue" webapp
  - kapcsolatot teremt a tárolókkal és a felhasználóval

## Adatbázis

MysqlServer: a fájlmozgatásokról és a felhasználókról tárol adatokat

## Proxy

HAProxy, mely HTTP (80) és HTTPS (443) portokon biztosít hozzáférést a felhasználónak



# Megoldás használatának lépései

## Felhasználó feladatköre:

- **0. Lépés: Előkészítés** (ELKH Cloud projekt, Üres Ubuntu VM elindítás)
- 1. Lépés: Occopus telepítés/konfigurálás a virtuális gépen
- Lépés: Leírók letöltése a virtuális gépre Occopus/ELKH Cloud weboldala
- 3. Lépés: Tűzfalszabályok létrehozása ELKH Cloud OpenStack felületén
- 4. Lépés: Leírók személyre szabása a virtuális gépen
- 5. Lépés: Occopus aktiválása \$ source ~/occopus/bin/activate
- 6. Lépés: Leírók importálása Occopus számára
   \$ occopus-import nodes/node\_definitions.yaml
- 7. Lépés: Infrastruktúra kiépítése\$ occopus-build --parallelize infra-dataavenue.yaml
- 8. Lépés: Infrastruktúra használata



## 2. lépés: A leírók letöltése a VM-re

Csatlakozás

Szolgáltatások Hírek

GYIK Projektek

Dokumentumok

Publikációk

Fórum Kapcsolat

ELKH Cloud

## Felhasználást segítő szolgáltat DataAvenue

### A rendelkezésre álló referencia architektúrák és leírásuk:

- Occopus cloud orchestrator indítása
- JupyterLab
- DataAvenue
- Cloud alkalmazásokat tan sogató portál indítása
- Flowbster Autodock Vina
- CQueue klaszter .
- Docker-Swarm klaszter kiépítése (FAssítés: ELKH Cloud Microso .
- Kubernetes klaszter
- Apache Hadoop klaszter kiépítése
- Apache Spark klaszter RStudio stack-el
- Apache Spark klaszter Python stack-el (Trissités: ELKH Cloud M.
- TensorFlow, Keras, Jupyter Notebook stark
- TensorFlow, Keras, Jupyter Notebook OPU stack (Frissités: El támogatással)
- Horovod klaszter
- Kafka klaszter
- Slurm klaszter

### Mi a Data Avenue?

A Data Avenue egy olyan tároló menedzsment szolgáltatás, amely lehetővé teszi a különböző típusú tárolási források (beleértve S3, sftp, GridFTP, iRODS, SRM szerverek) elérését egy egységes felület segítségével. A rendelkezésre álló REST API lehetővé teszi az összes tipikus tárolási műveletet, mint például mappák/bucketek készítése, fájlok/mappák átnevezése vagy törlése, fájlok feltöltése/letöltése, fájlok/mappák másolása vagy áthelyezése a különböző tárolási erőforrások között, akár egyszerűen a "curl" parancs használatával a parancssorból. A DataAvenue automatikusan lefordítja a felhasználók REST parancsát a megfelelő tárolási protokollokra, és kezeli a hosszú távú adatátvitelt a háttérben. Ebben a bemutatóban létrehozunk egy klasztert, amely két csomóponttípust tartalmaz. A DataAvenue csomóponton a DataAvenue alkalmazás fog futni, és egy előre meghatározott számú tárolási csomóponton (storage node) pedig egy S3 tároló fog futni, hogy képesek legyünk kipróbálni a DataAvenue fájlátviteli szoftvert, mint például bucketek létrehozását, a fájlok letöltését vagy másolását. Ceph és Docker komponenseket használtunk fel a klaszter felépítésére.

Használati és telepítési útmutató:

Data Avenue dokumentáció és telepítési útmutató:

https://github.com/SZTAKI-LPDS/data-avenue

Data Avenue referencia architektúra:

https://occopus.readthedocs.io/en/latest/tutorial-building-clusters.html#dataavenue-cluster

## https://science-cloud.hu/felhasznalast-segito-szolgaltatasok



# 2. lépés: A leírók letöltése a VM-re

## A DataAvenue leírása

Az Occopus weblapján:

https://occopus.readthedocs.io/en/latest/tutorialbuilding-clusters.html#dataavenue-cluster

## A GitHib-on:

https://github.com/SZTAKI-LPDS/data-avenue

### DataAvenue cluster

Data Avenue is a data storage management service that enables to access different types of storage resources (including S3, sftp, GridFTP, iRODS, SRM servers) using a uniform interface. The provided REST API allows of performing all the typical storage operations such as creating folders/buckets, renaming or deleting files/folders, uploading/downloading files, or copying/moving files/folders between different storage resources, respectively, even simply using 'curl' from command line. Data Avenue automatically translates users' REST commands to the appropriate storage protocols, and manages long-running data transfers in the background.

In this tutorial we establish a cluster with two nodes types. On the DataAvenue node the DataAvenue application will run, and an S3 storage will run, in order to be able to try DataAvenue file transfer software such as making buckets, download or copy files. We used MinIO and Docker components to build-up the cluster.

#### Features

- creating two types of nodes through contextualisation
- using the nova resource handler

#### Prerequisites

- accessing an Occopus compatible interface
- target cloud contains an Ubuntu image with cloud-init support

#### Download

You can download the example as tutorial.examples.dataavenue-cluster .



# 3. lépés: A tűzfalszabályok beállítása

## 🔲 openstack Project Compute Ove Insta Volu Im Access & Sec Network Orchestration Identity

### Access & Security

/ Manage Security Group Rules: DataAvenue (f7ff030abdcc-4d31-8a80-5416ee43e34c)

	Direction	Ether Type	IP Protocol	Port Range	Remote IP Prefix	
erview	Egress	IPv6	Any	Any	::/0	Kimenő forgalom
umes	Egress	IPv4	Any	Any	0.0.0.0/0	5
ages	Ingress	IPv4	TCP	22 (SSH)	0.0.0.0/0	SSH hozzáférés
curity	Ingress	IPv4	TCP	80 (HTTP)	0.0.0.0/0	HTTP hozzáférés
<u> </u>	Ingress	IPv4	TCP	443 (HTTPS)	0.0.0.0/0	HTTPS hozzáférés
~	Ingress	IPv4	TCP	8080	0.0.0.0/0	DataAvenue service

# 4. lépés: A leírók testreszabása

Infrastruktúra leíró fájl (dataavenue-cluster/infra-dataavenue.yaml)

Nova erőforrás szekció:

```
infra name: dataavenue-infra
user_id: somebody@somewhere.com
nodes:
  - &D
    name: dataaavenue
   type: dataaavenue_node
  - &S
   name: storage
   type: storage_node
    scaling:
      min: 1
                Skálázhatóság
      max: 1
    variables
      access key: A802WPCWAELW61RWDG08
      secret key: FWd1mccBfnw6VHa2vod98NEQktRCY1Cronxb01aQ
dependencies:
    connection: [ *S, *D ]
```

A szolgáltatás használatához hozzáférési kulcsokra van szükség, melyet itt adhatunk meg. Használhatjuk az (ábrán lévő) alapértelmezett kulcsokat is.



https://occopus.readthedocs.io/en/latest/user-doc-collecting-resources.html#openstack-horizon-nova

## 4. lépés: A leírók testreszabása

Csomópont definíciós fájl (dataavenue-cluster/nodes/node\_definition.yaml)

Nova erőforrás szekció:



resource:
 type: nova
 endpoint: https://sztaki.cloud.mta.hu:5000/v3
 project\_id: caml6db63ddf47a98045ef9c726vgdpp
 user\_domain\_name: Default
 image\_id: zgsf1dc3-b6d5-4b15-942e-61e0ef218dk
 network\_id: 3yqqqe1c-858c-4047-a48a-e2fab0nd547
 flavor\_name: 3
 key\_name: johndoe-key
 security\_groups: [DataAvenue]
 floating\_ip: yes
contextualisation:
 type: cloudinit
 context\_template: !yaml\_import

'node def:spark master node':

ELKH Cloud

- url: file://cloud\_init\_dataavenue.yaml
- | ..



# 5-6. lépés: Aktiválás és importálás

Az 5. lépésben aktiváljuk az Occopus virtualenv-et (ha még nem történt meg):

ubuntu@occopus:~\$ source \$HOME/occopus/bin/activate

A 6. lépésben importáljuk be a leíró mappájából a megfelelő fájlt:

- (occopus) ubuntu@occopus:~\$ occopus-import dataavenuecluster/nodes/node\_definitions.yaml
- Successfully imported nodes: dataavenue\_node, storage\_node

### 🧬 ubuntu@occopus: ~

(occopus) ubuntu@occopus:~\$ occopus-import dataavenue-cluster/nodes/node\_definitions.yaml Successfully imported nodes: dataaavenue\_node, storage\_node (occopus) ubuntu@occopus:~\$

### Fontos!

- Minden módosításkor újra kell importálni a leíró fájlokat a 6. lépés szerint.
- A nodes mappában lévő további fájlokat csak saját felelősségre szerkesszék.



# 7. lépés: Az infrastruktúra kiépítése

Parancs a klaszter kiépítéséhez:

ubuntu@occopus:~\$ occopus-build dataavenue-cluster/infra-dataavenue.yaml

<pre>\$ occopus-build infra-dataaven</pre>	ue.yaml
** 2021-06-13 11:13:21,391	Creating node 'dataavenue'/'f1e7a95d-ec83-472c-8384-6f98fcc91a6b'
** 2021-06-13 11:21:04,184	Health checking for node 'dataavenue'/'f1e7a95d-ec83-472c-8384-6f98fcc91a6b'
** 2021-06-13 11:21:05,360	Checking node reachability (f1e7a95d-ec83-472c-8384-6f98fcc91a6b):
** 2021-06-13 11:21:05,371	192.168.10.211 => ready
** 2021-06-13 11:21:05,371	Checking port availability (f1e7a95d-ec83-472c-8384-6f98fcc91a6b):
** 2021-06-13 11:21:05,373	8080 => ready
** 2021-06-13 11:21:05,373	Health checking result: ready
** 2021-06-13 11:21:05,376	Node 'dataavenue'/' f1e7a95d-ec83-472c-8384-6f98fcc91a6b' is ready.
** 2021-06-13 11:21:05,409	Creating node 'storage'/'64c98e6c-ad83-4b23-8c66-f247376f6a20'

Instance Name	lmage Name	IP Address	Size	Key Pair	Status	Availability Zone	Task	Power State	Time since created	Actions
occopus-dataavenue-infra- 1d14b3b3-storage-52e266df	Ubuntu 18.04 LTS Cloud image	192.168.10.210 Floating IPs: 193.224.59.201	m1.medium	attila- key	Active	nova	None	Running	2 minutes	Create Snapshot 💌
occopus-dataavenue-infra- 1d14b3b3-dataaavenue- 9fe195f4	Ubuntu 18.04 LTS Cloud image	192.168.10.209 Floating IPs: 193.224.59.182	m1.medium	attila- key	Active	nova	None	Running	10 minutes	Create Snapshot 💌

# 7. lépés: Az infrastruktúra kiépítése

## Az Occopus log üzenete kiépítés közben:

(occopus) ubuntu	@occopus:~\$ occ	opus-build dataavenue-cluster/infra-dataavenue.yaml
Using default con	nfiguration fil	e: '/home/ubuntu/.occopus/occopus_config.yaml'
Using default aut	thentication fi	le: '/home/ubuntu/.occopus/auth_data.yaml'
** 2021-06-28 12	:20:52,712	Starting up; PID = 10835
** 2021-06-28 12	:20:52,715	[SchemaCheck] WARNING: missing "scaling" parameter in node 'dataaavenue', using default scaling (single instance)
** 2021-06-28 12	:20:52,722	Submitted infrastructure: 'ldl4b3b3-b7cf-4a5d-a36e-7144f96e50f2'
** 2021-06-28 12	:20:52,723	Start maintaining the infrastructure ldl4b3b3-b7cf-4a5d-a36e-7144f96e50f2
** 2021-06-28 12	:20:52,774	Creating node 'dataaavenue'/'9fe195f4-4151-4b5e-a0db-4c536cb0a20d'
** 2021-06-28 12	:20:55,683	Waiting for node 'dataaavenue'/'9fel95f4-4151-4b5e-a0db-4c536cb0a20d' to become ready. No timeout.
** 2021-06-28 12	:21:37,530	Health checking for node 'dataaavenue'/'9fe195f4-4151-4b5e-a0db-4c536cb0a20d'
** 2021-06-28 12	:21:38,758	Checking node reachability (9fel95f4-4151-4b5e-a0db-4c536cb0a20d):
** 2021-06-28 12	:21:39,767	192.168.10.209 => pending
** 2021-06-28 12	:21:39,768	Health checking result: pending
** 2021-06-28 12	:21:39,770	Service on node 'dataaavenue'/'9fel95f4-4151-4b5e-a0db-4c536cb0a20d' is down for 0.000 seconds! (Timeout for restart: 1200s)
** 2021-06-28 12	:21:50,145	Health checking for node 'dataaavenue'/'9fe195f4-4151-4b5e-a0db-4c536cb0a20d'
** 2021-06-28 12	:21:51,259	Checking node reachability (9fe195f4-4151-4b5e-a0db-4c536cb0a20d):
** 2021-06-28 12	:21:52,268	193.224.59.182 => pending
** 2021-06-28 12	:21:52,269	Health checking result: pending
** 2021-06-28 12	:21:52,270	Service on node 'dataaavenue'/'9fe195f4-4151-4b5e-a0db-4c536cb0a20d' is down for 12.501 seconds! (Timeout for restart: 1200s
** 2021-06-28 12	:22:02,651	Health checking for node 'dataaavenue'/'9fe195f4-4151-4b5e-a0db-4c536cb0a20d'
** 2021-06-28 12	:22:03,749	Ch <mark>ecking node reachability</mark> (9fe195f4-4151-4b5e-a0db-4c536cb0a20d):
** 2021-06-28 12	:22:03,758	193.224.59.182 => ready
** 2021-06-28 12	:22:03,758	Checking port availability (9fe195f4-4151-4b5e-a0db-4c536cb0a20d):
** 2021-06-28 12	:22:03,760	8080 => pending



# 8. lépés: Az infrastruktúra használata

A DataAvenue elérhető külső IP címről is (amennyiben rendelünk hozzá):

https://<DataAvenuelP>/dataavenue/ui/

A MinIO tárhely szintén elérhető külső IP címről is:

http://<MinIOStorageIP>/minio/login



## MinIO: Bejelentkezés a storage-ra

_	Access Key	
_	Secret Key	
		S
193.224.59.201		9

A hitelesítő adatok megadását követően lehet bejelentkezni a tárolóba (alapért. kulcsok: 15. dia).







## MinIO: Fájl feltöltése a storage-ra

Upload file ①

A mappába az Upload file gombot megnyomva tudunk fájlt feltölteni.

=	≣	File 'my_image.jpg' uploaded	successfully.	×		•••
	test / 🗭					
	Name		Size	Last Modified	↓º 1	
	my_image.jpg		159.88 KB	Feb 24, 2021 12:1	6 PM •	

A feltöltött képfájl megjelenik a tárolóban.



## A Data Avenue felületének használata

	Main	Uploads	Transfers	Settings	About	
	s3://19	3.224.59.201	l:80/test/		Go 🗌	hdfs://193.224.59.182:9000/sztaki/ Go
			Name		^	Name
	☐ 1 my im	age.jpg				input/
1	Csa tár pul seg A v 80-	itlakozzi olónkra blikus IP gítségéve régére ír as ports	unk a Mi a megac cím el. juk be m számot!	nIO lott indig a	×	2 Itt pedig a Hadoop HDFS tárterületéhez kapcsolódtunk szintén a publikus IP segítségével. HDFS-nél mindig a 9000-es portot kell megadnunk.

Rename

Download

Move

Copy

MkDir

Delete

Upload:

Tallózás... Nincsenek kijelölve fájlok.



# Példa: Fájl másolása a HDFS-re



# Sikeres átmozgatás



**ELKH Cloud** 

# ELKH Cloud

# Az infrastruktúra törlése

Az Occopus a kiépítés végén fontos információkat közöl:

** 2021-06-13 11:22:45,785	Submitted infrastructure: 'd71cf691-606f-471c-a471-0b9a9f899483'
** 2021-06-13 11:22:45,812	List of nodes/instances/addresses:
** 2021-06-13 11:22:45,812	dataaavenue:
** 2021-06-13 11:22:45,813	f1e7a95d-ec83-472c-8384-6f98fcc91a6b:
** 2021-06-13 11:22:45,813	192.168.10.212
** 2021-06-13 11:22:45,813	storage:
** 2021-06-13 11:22:45,814	64c98e6c-ad83-4b23-8c66-f247376f6a20:
** 2021-06-13 11:22:45,814	192.168.10.213

Az infrastruktúra ID-ja lekérdezhető a következő Occopus utasítással is:

## occopus-maintain -1

Az infrsatruktúra törlése az egyedi ID alapján:

occopus-destroy -i d71cf691-606f-471c-a471-0b9a9f899483

## Az infrastruktúra törlése

(occopus) ubuntu@occopus:~\$ occopus-destroy -i d71cf691-606f-471c-a471-0b9a9f899483 Using default configuration file: '/home/ubuntu/.occopus/occopus\_config.yaml' Using default authentication file: '/home/ubuntu/.occopus/auth\_data.yaml' \*\* 2021-06-13 13:16:17,812 Starting up; PID = 11336 \*\* 2021-06-13 13:16:17,824 Start dropping infrastructure d71cf691-606f-471c-a471-0b9a9f899483 \*\* 2021-06-13 13:16:17,824 Dropping node 'dataaavenue'/'d43dbde8-4c60-49b2-a7f3-10ad4f0bc94e' \*\* 2021-06-13 13:16:20,749 Dropping node 'storage'/'64c98e6c-ad83-4b23-8c66-f247376f6a20' \*\* 2021-06-13 13:16:23,114 Finished dropping infrastructure d71cf691-606f-471c-a471-0b9a9f899483

Instance Name	lmage Name	IP Address	Size	Key Pair	Status	Availability Zone	Task	Power State	Time since created	Actions
occopus-dataavenue-infra- d71cf691-storage-64c98e6c	Ubuntu 18.04 LTS Cloud image	192.168.10.213 Floating IPs: 193.224.59.201	m1.medium	attila- key	Active	nova	None	Running	13 minutes	Create Snapshot 👻
occopus-dataavenue-infra- d71cf691-dataaavenue- d43dbde8	Ubuntu 18.04 LTS Cloud image	192.168.10.212 Floating IPs: 193.224.59.182	m1.medium	attila- key	Active	nova	Deleting	Running	21 minutes	Update Metadata 🔻

# Összefoglalás

- A DataVenue megismerése
- A DataAvenue referencia architektúra kiépítése
- A DataAvenue webes felület és a MinIO storage használata
- ELKH Cloud technikai támogatás: info@science-cloud.hu



# Köszönöm a figyelmet!