



Bevezetés az ELKH Cloud használatába



Dr. Kacsuk Péter
ELKH Cloud projekt
szakmai vezetője

A mai rendezvény célja

1. Megmutatni az ELKH Cloud használati módjait és lehetőségeit (bevezető előadás)
2. Megválaszolni az ATK munkatársak ELKH Clouddal kapcsolatban feltett kérdéseit (bevezető előadás)
3. Megmutatni, hogy egy ATK-ás felhasználó hogyan tudta használni az ELKH Cloud-ot (2. előadás)
4. Megmutatni, hogy az R programozók hogyan tudják kihasználni a Cloud-ban rejlő párhuzamos végrehajtás lehetőségeit (3. előadás)
5. Egy konkrét példán megmutatni a Cloud használatának előnyeit (4. előadás)
6. Ami nem cél:
 - ▶ Ez nem oktatás, nem tanítunk meg semmit, arra vannak más előadásanyagok és rendezvények

Az MTA/ELKH Cloud története dióhéjban

- ▶ 2016. okt. 1: Az MTA Cloud, mint teljes körű Infrastructure as a Service (IaaS) szolgáltatás elindul
- ▶ 2020. július 1: Elindul az ELKH Cloud projekt, melynek célja a Cloud korszerűsítése és jelentős kapacitásbővítése
- ▶ 2020. okt. 1: a cloud hivatalos átnevezése **ELKH Cloud** névre
- ▶ 2021. július 1: Az új ELKH Cloud tervezett átadása a magyar kutatói közösség számára

Kinek ajánljuk az ELKH Cloudot?

Olyan kutatóknak, akik

- ▶ hatalmas adatmennyiséggel **dolgoznak** vagy **osztják meg** egymás között
- ▶ akár **több számítógépen** használnak kutatást támogató szoftvereket
- ▶ **számítógépfürtöt, szerverparkot, adatközpontot** használnak
- ▶ **számításigényes** vagy **adatintenzív szimulációkat** és/vagy egyéb **feldolgozásokat** végeznek

Az ELKH Cloud célja a **kutatás támogatása**

- ▶ nem támogatja az irodai tevékenységeket (pl.: e-mail rendszer)
- ▶ sem a szolgáltatási tevékenységeket (pl.: webszerver)



Kutatási e-infrastruktúrák használati módjai

- ▶ **HPC** (High Performance Computing) infrastruktúra
- ▶ Cél: **egyetlen program** futtatásának felgyorsítása, hogy a lehető legrövidebb idő alatt fusson le
- ▶ Módszer: egyidejűleg **sok koordinált processzor** alkalmazása és optimalizált kommunikációs hálózat a processzorok között
- ▶ Tipikus infrastruktúra: szuperszámítógép
- ▶ Szolgáltató: **KIFÜ**
- ▶ **HTC** (High Throughput Computing) infrastruktúra
- ▶ Cél: Egységnyi idő alatt **minél több program** lefuttatása.
- ▶ Módszer: egyidejűleg **sok egymástól független processzor** alkalmazása
- ▶ Tipikus infrastruktúra: felhő
- ▶ Szolgáltató: **SZTAKI** és **WDC**

Mire jó az ELKH Cloud?

A cloud szolgáltatásokat a következő három szintre tudjuk besorolni:

- ▶ **Infrastuktúra Cloud (IaaS)**
- ▶ Platform Cloud (PaaS)
- ▶ Szoftver Cloud (SaaS)

Az IaaS lehetővé teszi a további szintek felépítését is.

Az ELKH Cloud előnyei:

- ▶ nincs hosszú és bonyodalmas beszerzési eljárás
- ▶ számítási fürtök is használhatók (pl.: docker, Spark fürtök)
- ▶ különböző típusú és méretű infrastruktúra alakítható ki
- ▶ nagyméretű tárolókapacitás igényelhető



Kutatási projekt indítása

Hagyományos környezetben	Felhő környezetben
Van egy ötlet	
Projektjavaslat írása 6 hónap várakozási idő	A kívánt infrastruktúra lefoglalása / létrehozása a felhőben (20-30 perc)
Ha sikeres, akkor: <ul style="list-style-type: none">• Számítógépek beszerzése (hónapok)• Szoftverek telepítése (hetek)	
Megkezdődik az ötlet kidolgozása	

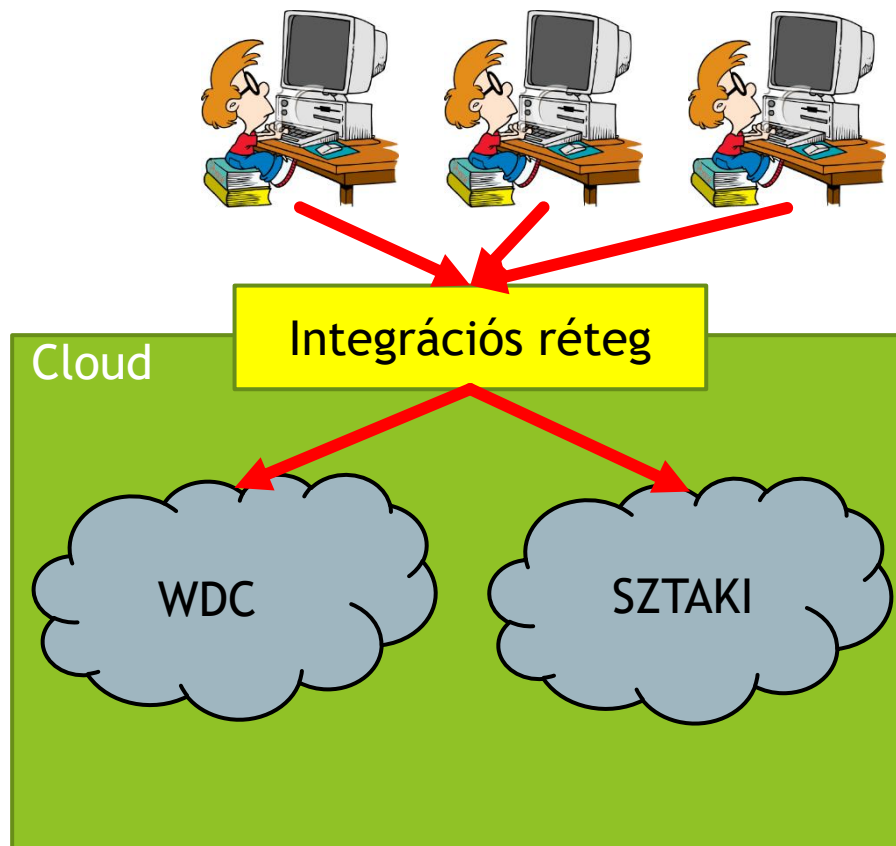


Ha elindult a kutatási projekt

Hagyományos környezetben	Felhő környezetben
<p>Az infrastruktúra esetleg alultervezett</p> <ul style="list-style-type: none">• az alkalmazások / szimulációk lassabban futnak	<p>Az infrastruktúra dinamikusan skálázható az aktuális igényeknek megfelelően</p>
<p>Az infrastruktúra esetleg túltervezett</p> <ul style="list-style-type: none">• felesleges beruházás• nem mindig van kihasználva• nem ajánlható fel másoknak	<p>Az intézeti saját infrastruktúra és a felhő kapacitás optimális arányával is lehet élni</p>



Az ELKH Cloud mint felhő föderáció



- Közös beruházási és fejlesztési tervek
- Közös cloud middleware: OpenStack
- Közös policy
- Közös kvóta kiosztás
- Közös weblap
- Közös ismeretterjesztés és disszemináció

Az MTA és ELKH Cloud kapacitásának összehasonlítása



	Jelenlegi kapacitás	Tervezett kapacitás	Megvalósuló kapacitás
vCPU (max)	1368	+2560	+5608
GPU core	12	+64	+64
vGPU (max)	12	+2048	+2548
RAM (TB)	3,25	+8	+28.2
SSD storage (TB)	0	+153	+337
HDD storage (TB)	527	+960	+951
Tensor GPU teljesítmény (PFLOPS)	~0	+7.16	+14.24
Lebegőpontos GPU teljesítmény (PFLOPS)	~0	+0.89	+1.27
Hálózati sávszélesség (Gbps)	10	100	100

MTA Cloud és ELKH Cloud összehasonlítása

MTA Cloud

Kapacitás:

- ▶ Kis kapacitású cloud

Felhasználók:

- ▶ MTA Kutatók

Nemzetközi elérés:

- ▶ Korlátozott

ELKH Cloud

Kapacitás:

- ▶ Közepes kapacitású cloud

Felhasználók:

- ▶ ELKH Kutatók
- ▶ **Egyetemi kutatók**
- ▶ Üzleti szféra kutatói

Nemzetközi elérés: Részvétel nagy európai programokban:

- ▶ EOSC
- ▶ RDA
- ▶ ESFRI

Az ELKH Cloud jelenlegi kihasználtsága



- ▶ Az elindított projektek száma: 146
- ▶ A befejezett projektek száma: 62
- ▶ Web lapon elérhető ezekről minden lényeges információ:
<https://cloud.mta.hu/projektek>

Projektek

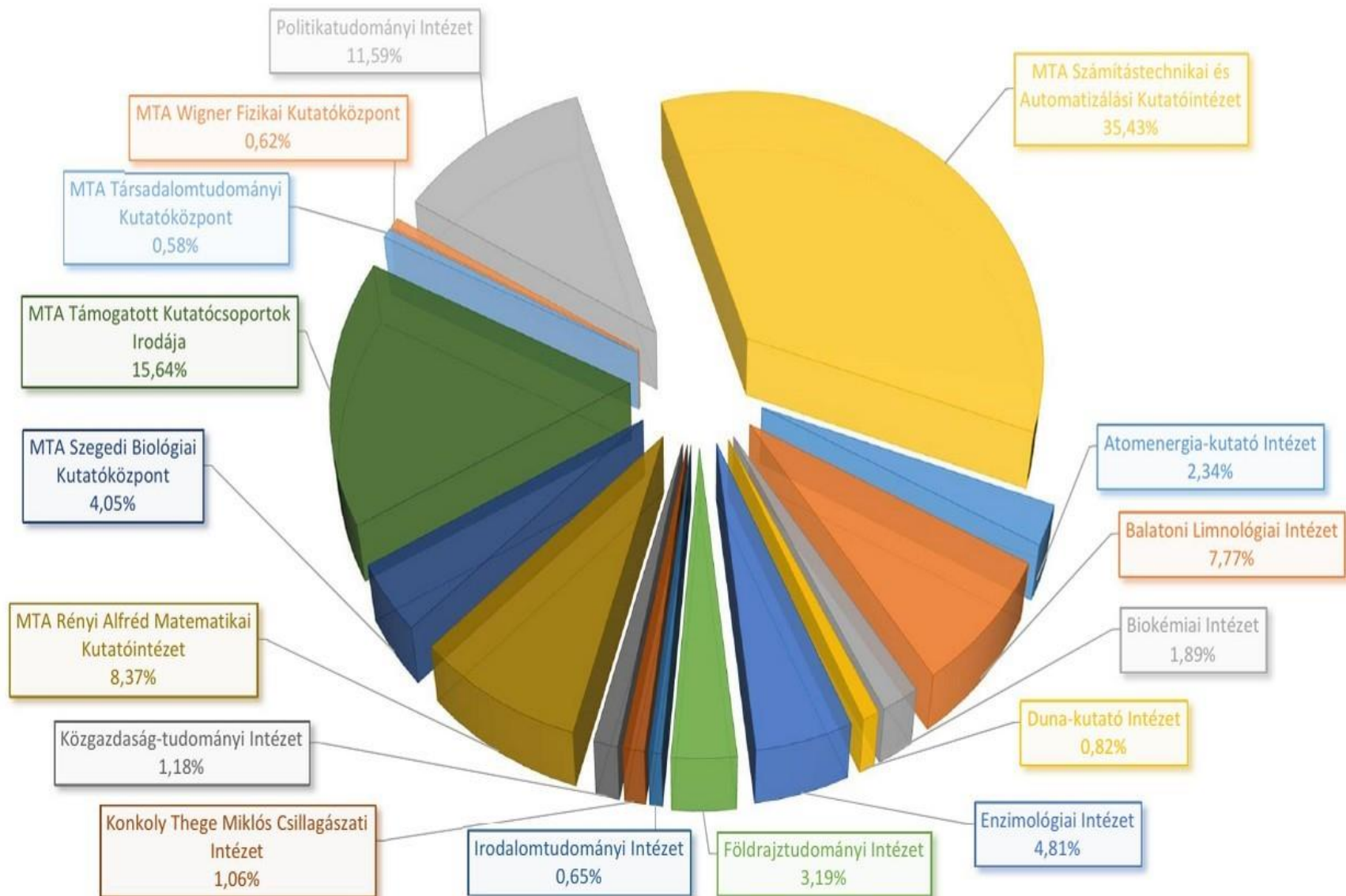
Összes Futó Befejezett

51 - 60 | 129 projekt

Projekt neve ▲	Intézmény	Vezető
FSL teszt	Támogatott Kutatócsoportok Irodája	Kincses Zsigmond Tamás
<p>Kutatócsoportunk mágneses rezonancia képanalízissel foglalkozik az FSL programcsomag segítségével. Tekintettel arra, hogy a felvételek analizésének jelentős része voxelenként, egymástól függetlenül történik az eljárások jól parallelizálhatók. Az Oxfordi Egyetem FMRIB csoportja által fejlesztett FSL softvercsomag grid engine detektálása esetén a számításokat automatikusan parallelizálva indítja el.</p> <p>Jelen pilot project célja az, hogy megvizsgáljuk, hogy az MTA Cloud alkalmas-e a fenti programcsomag és analízisek futtatására.</p>		
Gamma-sugárzó aktív galaxismagok Fermi adatainak likelihood analízise	Konkoly Thege Miklós Csillagászati Intézet	Kun Emma
<p>A projekt keretében az elektromágneses spektrum gamma tartományában is erősen sugárzó aktív galaxismagok Fermi Űrtávcsővel (https://fermi.gsfc.nasa.gov) készült méréseit szándékozzuk feldolgozni a Fermitools nevű (https://fermi.gsfc.nasa.gov/ssc/data/analysis/software/) dedikált szoftverrel és python kódokkal.</p>		
Génexpressziós korrelációs hálózat készítése	Mezőgazdasági Intézet	Majláth Imre

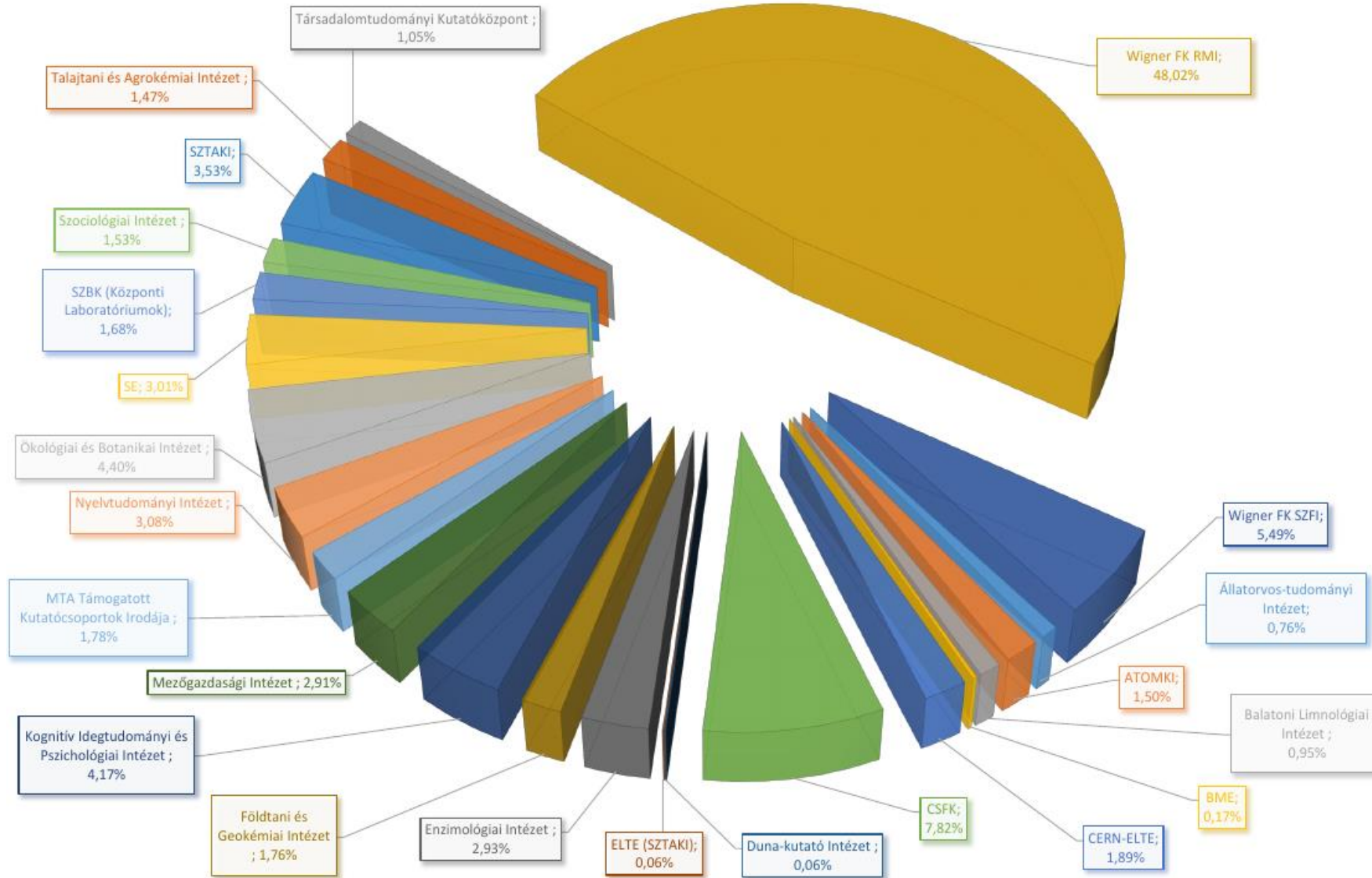
Kihasználtság

MTA CLOUD - SZTAKI ADATKÖZPONT RÉSZ KIHASZNÁLTSÁGA INTÉZMÉNYEK SZERINT 2016.10.01 - 2019.09.30 IDŐSZAKBAN



Kihasználtság

MTA CLOUD - WIGNER ADATKÖZPONT RÉSZ
KIHASZNÁLTSÁGA INTÉZMÉNYEK SZERINT
2016.10.01 - 2019.09.30 IDŐSZAKBAN



Az ELKH Cloud koncepciója

- ▶ **Az ELKH Cloud egy e-infrastruktúra keretrendszer**, ami
 - ▶ Nem felhasználó-orientált, hanem **projekt-orientált**
 - ▶ Elsődlegesen projektek regisztrálnak
 - ▶ A felhasználóknak nem a cloudra, hanem a projekthez kell regisztrálniuk
 - ▶ Egy projekt **kvótát** kap és ezen belül olyan e-infrastruktúrát épít fel a felhőben, amelyet csak akar
 - ▶ Pl. csinálhat Hadoop v. Spark klasztert, de azt csak az ő felhasználói éri el

Az ELKH Cloud koncepciója

- ▶ e-infrastruktúrát felépíteni a felhőben nem könnyű, ezért az ELKH Cloud projekt
 - ▶ A tipikus, gyakran használt e-infrastruktúrák kiépítéséhez **referencia architektúrákat** biztosít, amikből a kívánt infrastruktúra gyorsan felépíthető
 - ▶ A nem tipikus e-infrastruktúrák felépítéséhez közvetlen segítséget nyújt a felhasználóknak, majd az így létrejött e-infrastruktúrából referencia architektúrát csinál
- ▶ A referencia architektúrák tárolására és elérésére repozitóriumot biztosítunk, ahonnan minden ELKH felhasználó elérheti és használhatja ezeket

Az ELKH Cloud koncepciója

▶ **ATK kapcsolódó kérdése:**

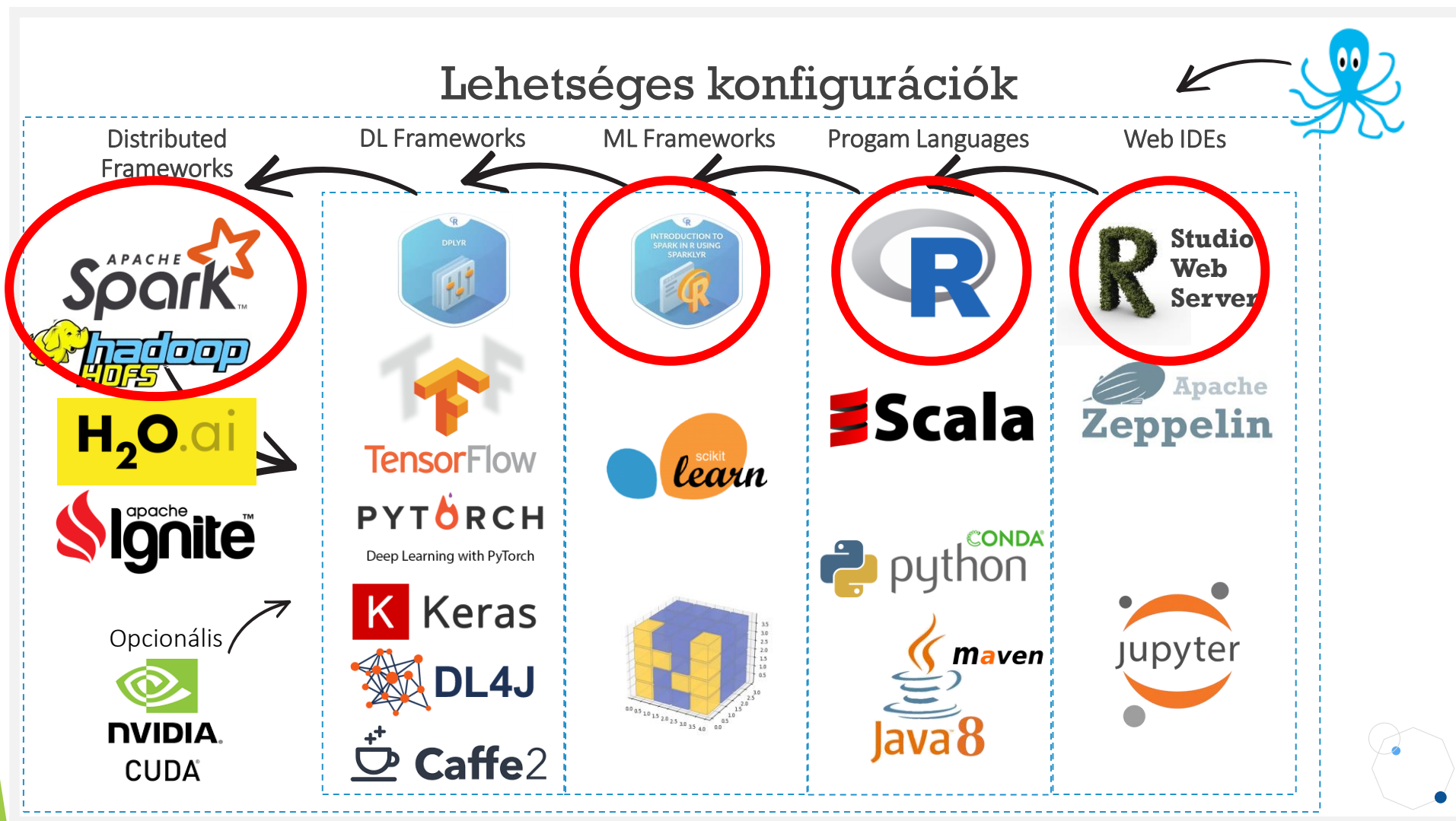
- ▶ Milyen programok használhatók az ELKH Cloud infrastruktúrában?

▶ **Válasz:**

- ▶ Bármilyen nyílt forráskódú program, ami vagy a Ref. Arch-ban van benne, vagy a felhasználó telepíti
- ▶ Bármilyen licencces program, aminek a licencéről a felhasználó gondoskodik

Miért kellenek referencia architektúrák?

Spark és Hadoop referencia architektúrák Big Data és egyszerűbb ML alkalmazásokhoz



Felhasználó

edu ID

Belépés

magyar

English

Címlap

Felhasználást segítő szolgáltatások

Az ELKH Cloud projekt koncepciójának megfelelően a felhasználást segítő szolgáltatások referencia architektúra formájában vannak megadva. A projektek ezekből építhetik fel azokat a szolgáltatásokat, amiket az alábbi listából kiválogattak.

A szolgáltatások telepítésének és elindításának lépései a következők:

0. Lépés: Előkészítés (Üres Ubuntu virtuális gép indítása a saját projekten belül)

Az 1-8 lépések leírását ld. az [Occopus cloud orchestrator indítása referencia architektúra leírásában](#)

1. Lépés: Occopus telepítés/konfigurálás a virtuális gépen

2. Lépés: Leírók letöltése a virtuális gépre

3. Lépés: Tűzfalszabályok létrehozása ELKH Cloud OpenStack felületén

4. Lépés: Occopus leírók személyre szabása a virtuális gépen

5. Lépés: Occopus aktiválása

6. Lépés: Leírók importálása Occopus számára

7. Lépés: Szolgáltatás kiépítése

8. Lépés: A kiépített szolgáltatás használata

5. Lépés: Occopus aktiválása
6. Lépés: Leírók importálása Occopus számára
7. Lépés: Szolgáltatás kiépítése
8. Lépés: A kiépített szolgáltatás használata

A rendelkezésre álló referencia architektúrák és leírásuk:

- [Occopus cloud orchestrator indítása](#)
- [JupyterLab](#)
- [DataAvenue](#)
- [Cloud alkalmazásokat támogató portál indítása](#)
- [Flowbster - Autodock Vina](#)
- [CQueue klaszter](#)
- [Docker-Swarm klaszter kiépítése](#) *(Frissítés: ELKH Cloud - Microsoft Azure hibrid felhő támogatással)*
- [Kubernetes klaszter](#)
- [Apache Hadoop klaszter kiépítése](#)
- [Apache Spark klaszter RStudio stack-el](#)
- [Apache Spark klaszter Python stack-el](#) *(Frissítés: ELKH Cloud - Microsoft Azure hibrid felhő támogatással)*
- [TensorFlow, Keras, Jupyter Notebook stack](#)
- [TensorFlow, Keras, Jupyter Notebook GPU stack](#) *(Frissítés: ELKH Cloud - Microsoft Azure hibrid felhő támogatással)*
- [Horovod klaszter](#)
- [Kafka klaszter](#)

A kvóta rendszer

▶ **ATK kapcsolódó kérdése:**

- ▶ Lehetőség van a felhő infrastruktúra alkalomszerű használatára?

▶ **Válasz:**

- ▶ Jelenleg nincs rá lehetőség
- ▶ DE: azt meg lehet tenni, hogy a kvótán belül minden VM-et leállít a projekt és ezzel felszabadítja az erőforrásokat mások számára
- ▶ Ezt a későbbiekben monitorozni és jutalmazni fogjuk
- ▶ A ref. arch-k épp ezt a használati módot tudják segíteni, mert általuk gyorsan és szinte automatikusan újraépíthető egy korábban lebontott VM struktúra

Projekt regisztráció és bejelentkezés

▶ ATK kapcsolódó kérdése:

- ▶ Mennyi időre biztosított a szolgáltatás elérhetősége a felhőn indított projektek számára?

▶ Válasz:

- ▶ Az ELKH-val most folynak a tárgyalások, hogy a következő 3 évre kössünk szerződést a felhő üzemeltetésére. (Szinte biztos, hogy ez meg fog történni.)
- ▶ Ha ez megtörténik, akkor 3 évre garantált a szolgáltatás.
- ▶ Jó esély van arra, hogy a 3 év után újabb 3 éves szerződés lesz megkötve.

Projekt regisztráció és bejelentkezés

- ▶ **ATK kapcsolódó kérdése:**
 - ▶ Hogyan lehet igényelni az ELKH Cloud infrastruktúra felhő használatát?
- ▶ **Válasz:**
 - ▶ A most következő diák, ezt magyarázzák el



Projekt regisztráció és bejelentkezés

A regisztráció menete

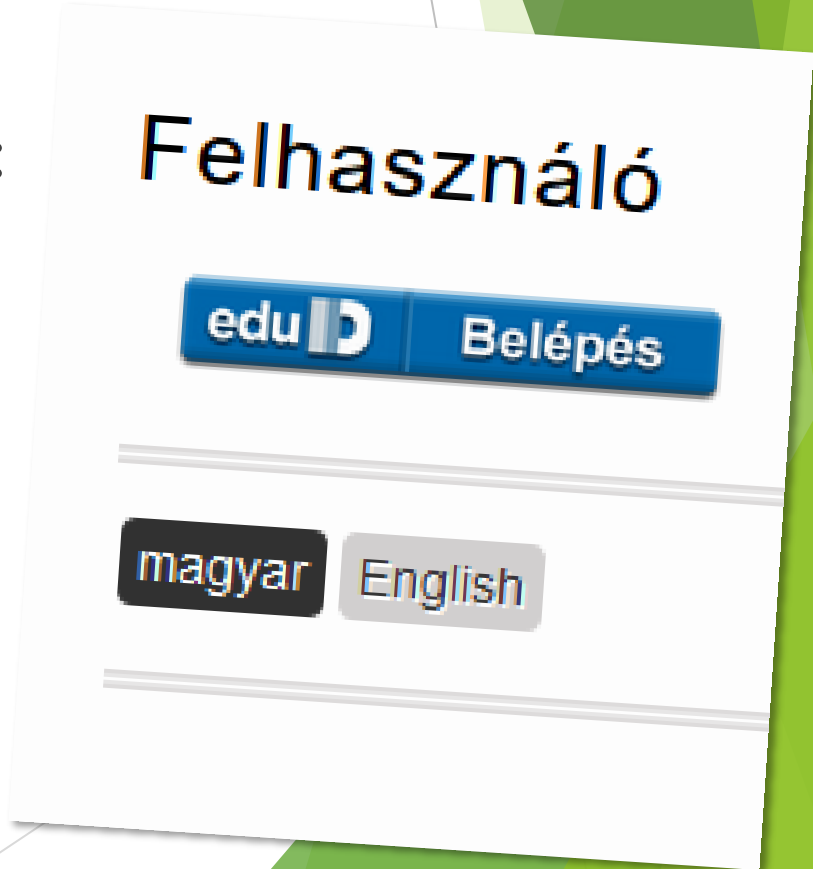
Regisztráció az ELKH Cloud oldalán:

▶ <https://science-cloud.hu/csatlakozas-az-elkh-cloudhoz>

A regisztrációs űrlap elérése két módon lehetséges:

- ▶ **eduID** segítségével
- ▶ **Akadémiai Adattár** azonosítóval

További részletek az ELKH Cloud GYIK oldalán.



Felhasználó

eduID

Belépés

magyar

English

Hogyan zajlik a regisztrációs folyamat?



1. Az ELKH Cloud web oldalon **kitöltjük a regisztrációs űrlapot.**
 - ▶ erre csak projekt regisztráció esetén van szükség
2. A rendszer nyugtázza a regisztrációt egy **automatikus üzenet kiküldésével.**
 - ▶ egy műszaki kolléga felveszi a kapcsolatot telefonon a projektvezetővel
3. Az **ELKH Cloud csapat döntést hoz** a felhők terheltsége és az igények alapján:
 - ▶ melyik felhőhöz adjon hozzáférést (SZTAKI vagy Wigner)
 - ▶ mekkora kvótával
4. A felhasználó **beléphet az ELKH Cloud felületére** és létrehozhatja a kívánt infrastruktúrát

Új projekt regisztrációja

A regisztráció során a következőket kell megadni:

- ▶ A projekt neve, leírása
- ▶ Az erőforrásokhoz hozzáférő kutatók száma
- ▶ A futtatni kívánt szoftverek listája
- ▶ A szükséges virtuális gépek száma
- ▶ A szolgáltatás igénybevételének a helye
- ▶ A teljes projektre igényelt erőforrások összessége
- ▶ A projektzárás tervezett időpontja
- ▶ A projektvezető adatai
- ▶ Egyéb megjegyzés

Projekt beküldése

Project request

Projekt neve

Project name *

Azonosító








Identifier *

Rövidített projekt név, illetve -azonosító. Jellemzően a projekt nevéből származtatott mozaikszó, mely legalább 2 és legfeljebb 15 karakterből áll. Az angol ABC nagy betűit, számokat és kötőjeleket tartalmazhat, de nem kezdődhet, és nem végződhet kötőjellel.

Short project name or project ID you would like to have. Usually the acronym of the project which contains between 2 to 15 characters. It can contain english capital letters, numbers, dash, but it can not start or end with dash.

Rövid leírás

Short Summary *

B **I**       Formátum -  Forráskód

About text formats ?

Kérjük, a projekt leírásában szerepeltessék a projekthez kapcsolódó URL-eket is.

Please, include the related URLs in the project description.

További felhasználó hozzáadása meglévő projekthez

- ▶ A kutató jelentkezik a projektvezetőjénél.
- ▶ A projektvezető értesíti a szolgáltatót (SZTAKI vagy WIGNER)
- ▶ WIGNER esetén:
 - ▶ Az adminisztrációt egy cloudos kolléga fogja elvégezni
- ▶ SZTAKI esetén:
 - ▶ A projektvezető a HEXAA rendszerben regisztrálja be az új projekt tagot

Projekthez történő csatlakozás esetén nem kell regisztrációs űrlapot kitölteni!

További felhasználó hozzáadása meglévő projekthez

▶ **ATK kapcsolódó kérdése:**

- ▶ Hány kutató kaphat hozzáférést az ELKH Cloudban igényelt projekthez?

▶ **Válasz:**

- ▶ Amennyit a projektvezető engedélyez
- ▶ Az ELKH Cloud csapat ebbe nem szól bele

A kvóta rendszer

Célja, hogy a **legtöbb felhasználó** hozzáférhessen a számára szükséges erőforrásokhoz úgy, hogy ezzel másoktól szükségtelenül ne vegyen el erőforrásokat.

A kvótarendszerrel kapcsolatban néhány fontos információ:

- ▶ regisztrációkor minden projekt kap egy **alapkvótát**
- ▶ ha **több erőforrásra** van szükség, bármikor **igényelhető**
 - ▶ ennek mértéke függ a projekt fontosságától és a terheltségtől



A kvóta rendszer

▶ **ATK kapcsolódó kérdése:**

- ▶ Milyen számítási kapacitás érhető el az ELKH Cloud-ban?

▶ **Válasz:**

- ▶ Amilyet a felhasználó a projekt igénylésébe beír és amit az ELKH Cloud csapat reálisnak tart
- ▶ Ha úgy látjuk, hogy irreálisan nagyok az igények, akkor egy tárgyalási folyamat kezdődik, ahol megpróbáljuk a valós igényeket átbeszélni
- ▶ A cloud nem szuperszámítógép, az olyan igényekkel a KIFÜ-höz érdemes menni

A kvóta rendszer

▶ ATK kapcsolódó kérdése:

- ▶ Mekkora tárhely igényelhető kutatónként?

▶ Válasz:

- ▶ Nem kutatónként kell igényelni az erőforrásokat, hanem projektként
- ▶ A projekten belül a tagok egymásközt osztják el a felhasználható erőforrásokat (pl. tárhelyet)

A kvóta rendszer

▶ **ATK kapcsolódó kérdése:**

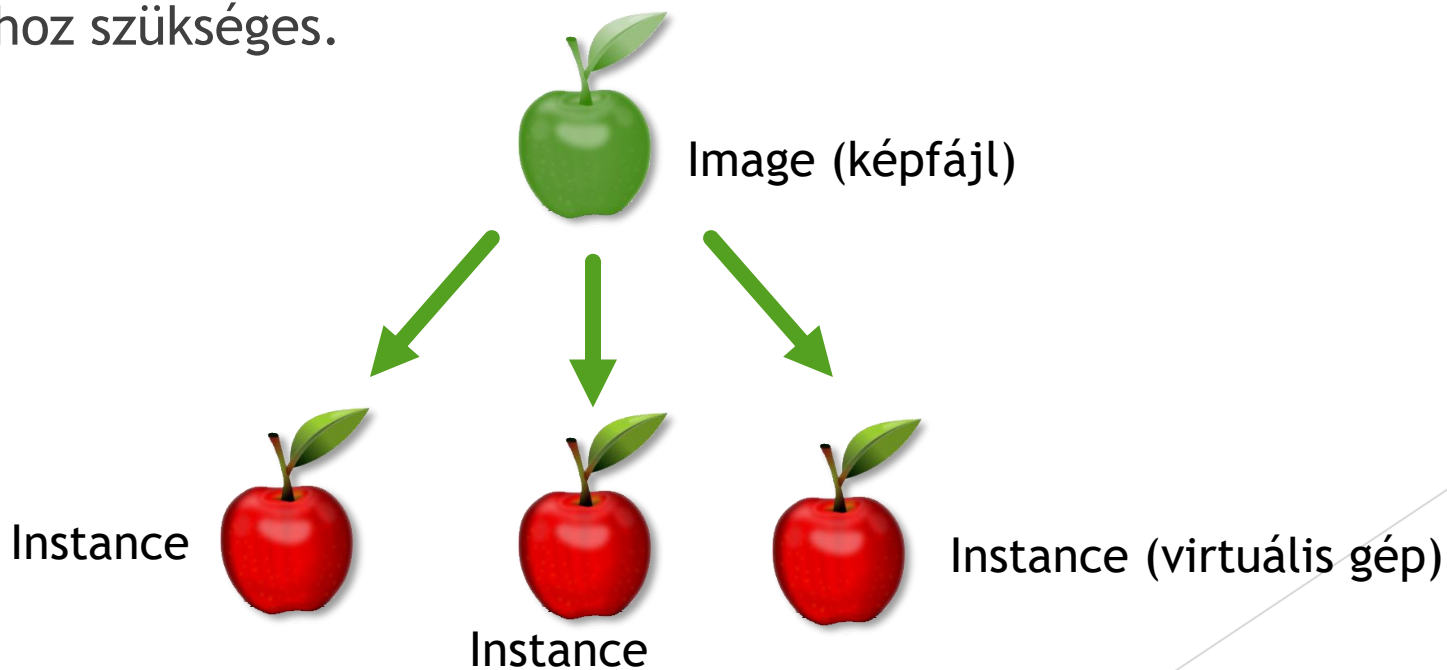
- ▶ A projektben résztvevő kutatók hogyan tudják együttműködve használni a szolgáltatást?

▶ **Válasz:**

- ▶ A projekt kvótája és minden ami ezen a területen történik, az közös minden projekt tag számára, azaz mindenki lát és használhat mindent a területen belül (hiszen azonos projekten dolgoznak)
- ▶ Ez persze felelősség is, mert egy projekt tag mindent törölhet, vagy felülírhat. Ezért a projekt vezető felelőssége és döntése, hogy kit enged be projekt tagnak

Az instance és az image

- ▶ **Image:** a képfájl valójában egy statikus adathalmaz, ami magát a szoftvert tartalmazza, azaz magát az operációs rendszert és esetleg azokat a további szoftvereket amit már a virtuális gép indításakor feltelepítve találunk a gépen.
- ▶ **Instance:** a példány pedig egy futtatható virtuális gép. Minden esetben egy képfájl alapján kerül létrehozásra és rendelkezik azokkal képességekkel, ami egy operációs rendszer és a rajta lévő szoftverek futtatásához szükséges.



Rendelkezésre álló VM típusok

Az ELKH Cloud ötféle alaptípust (ún. flavort) támogat:

Név	VCPU	RAM	Merevlemez
m1.tiny	1	512 MB	1 GB
m1.small	1	2 GB	20 GB
m1.medium	2	4 GB	40 GB
m1.large	4	8 GB	80 GB
m1.xlarge	8	16 GB	160 GB

Az alapkvóta, amit minden projekt megkap: **m1.xlarge**

Ez az erőforrásmennyiség **felhasználható:**

- ▶ egyetlen nagy VM-ként
- ▶ több kisebb VM-ként (pl.: 8db. small indításával)



Virtuális gépek

▶ **ATK kapcsolódó kérdése:**

- ▶ Egy projekthez több virtuális gép is igényelhető?

▶ **Válasz:**

- ▶ Nem virtuális gépet, hanem kvótát kell igényelni
- ▶ A kvóta határon belül minden projekt annyi virtuális gépet indít magának, amennyire szüksége van
- ▶ Virtuális gépet lehet indítani
 - ▶ az előre elkészített képfájlokból (pl. Ubuntu gép, Windows gép)
 - ▶ Referencia architektúrákból (tipikusan több együttműködő virtuális gép jön létre)

Virtuális gépek

▶ **ATK kapcsolódó kérdése:**

- ▶ A projekt során változtatható-e az (igényelt) használt virtuális gép(ek) paraméterei - CPU, RAM, tárhely?

▶ **Válasz:**

- ▶ Igen, a VM indításakor más VM alaptípust (flavort) kell választani.

Képfájlok

▶ Kapcsolódó kérdés:

- ▶ Egy projekt hozhat-e létre új képfájlt?

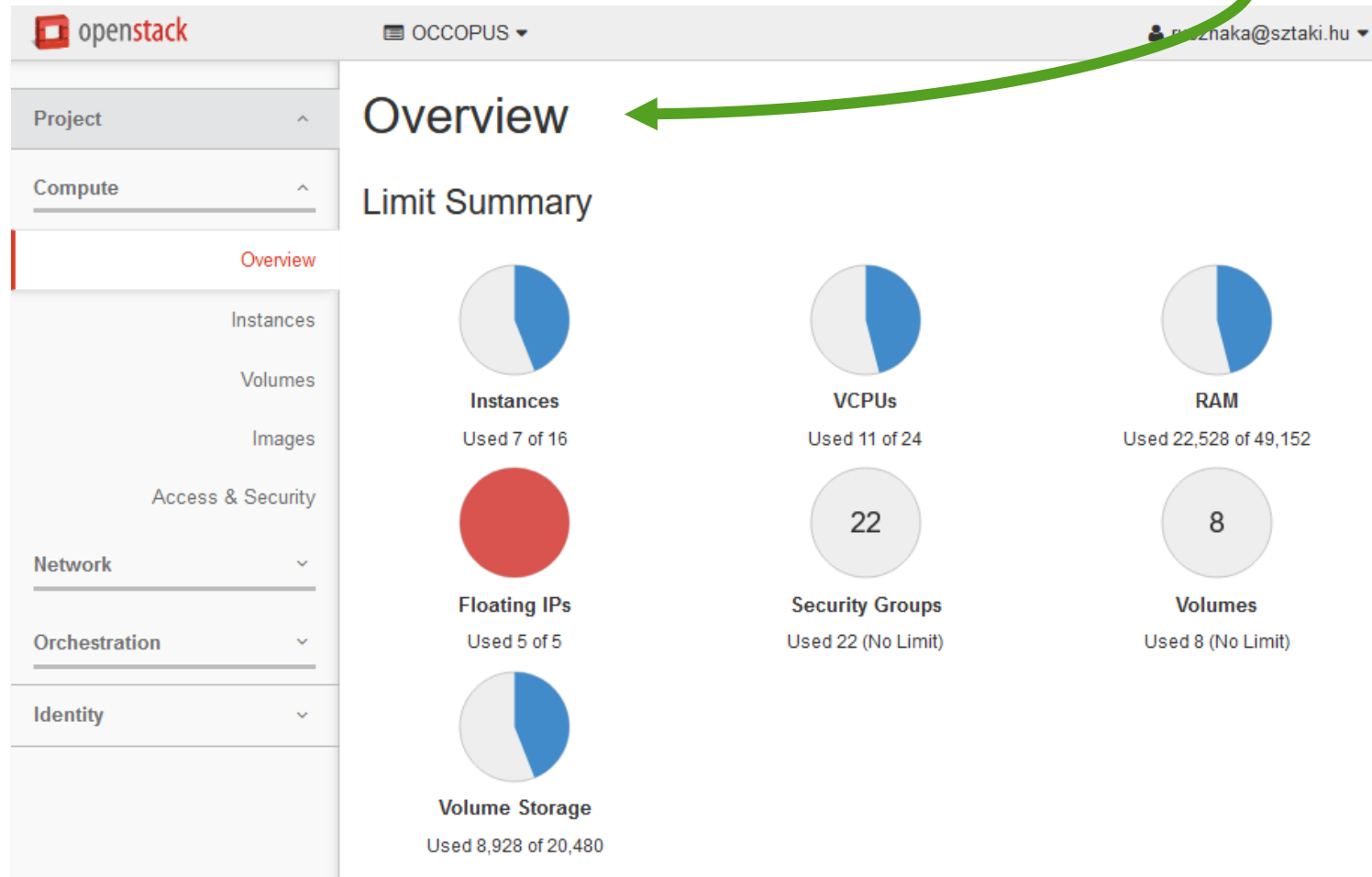
▶ Válasz:

- ▶ Biztonsági okokból nem
- ▶ Ha mégis van erre igény, akkor azt külön kérvényezni kell a info@science-cloud.hu címre írt e-mailben
- ▶ Indokolt esetben a cloud fejlesztő csapat segít a szükséges képfájl létrehozásában és letesztelésében, hogy garantáltan jól és biztonságosan működjenek az abból létrehozott VM-ek

ELKH Cloud kezdőlap, amit a bejelentkezés után látunk

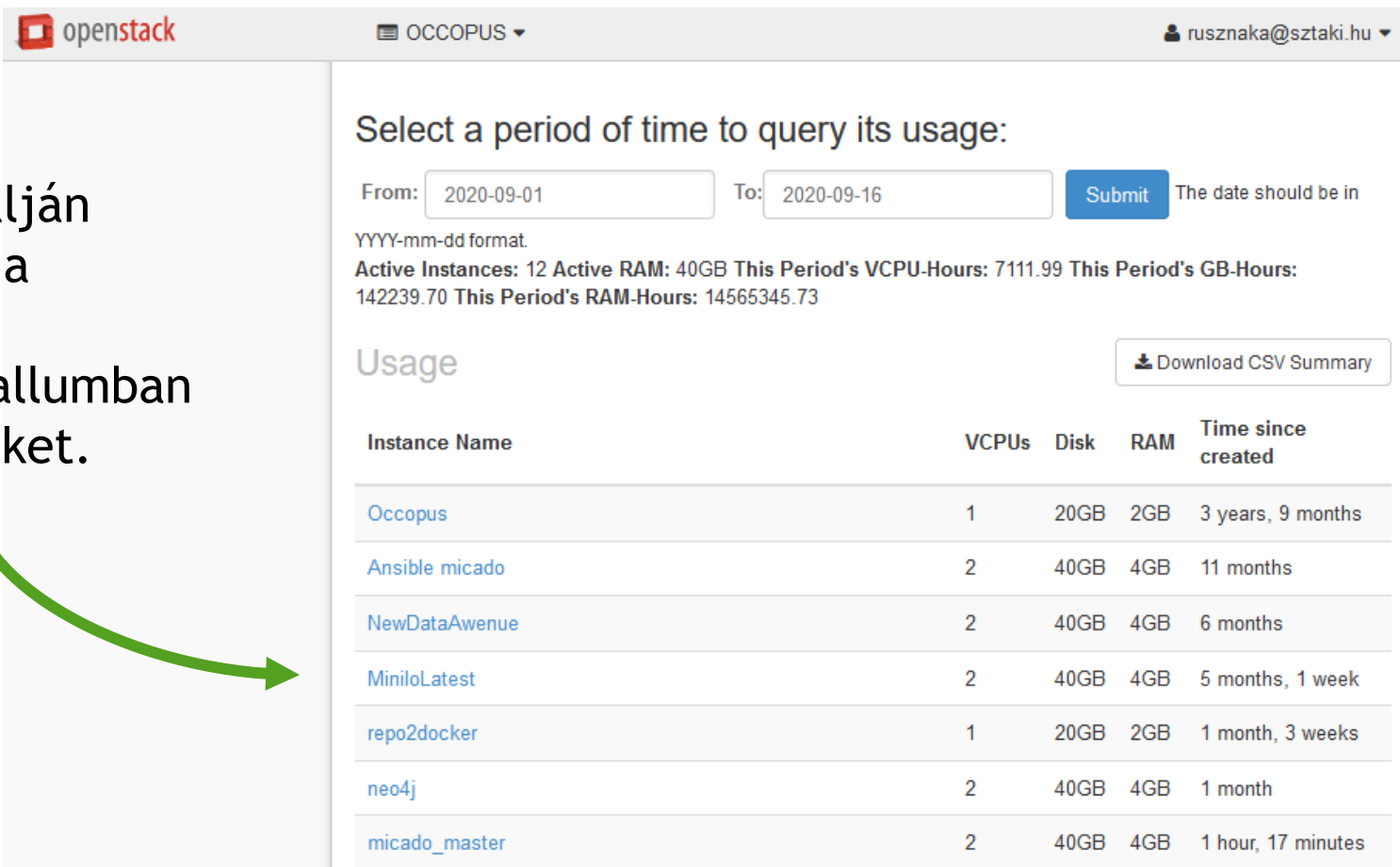


Az ELKH Cloud áttekintő oldala



ELKH Cloud kezdőlap

Az oldal alján láthatjuk a megadott időintervallumban futó VM-eket.



The screenshot shows the OpenStack dashboard for the OCCOPUS project. The user is rusznaka@sztaki.hu. The page displays usage statistics for a selected period from 2020-09-01 to 2020-09-16. The usage summary includes 12 active instances, 40GB active RAM, 7111.99 VCPU-hours, 142239.70 GB-hours, and 14565345.73 RAM-hours. A table lists the running VM instances with their names, VCPUs, disk space, RAM, and time since creation.

openstack OCCOPUS rusznaka@sztaki.hu

Select a period of time to query its usage:

From: 2020-09-01 To: 2020-09-16 Submit The date should be in YYYY-mm-dd format.

Active Instances: 12 Active RAM: 40GB This Period's VCPU-Hours: 7111.99 This Period's GB-Hours: 142239.70 This Period's RAM-Hours: 14565345.73

Usage [Download CSV Summary](#)

Instance Name	VCPUs	Disk	RAM	Time since created
Occopus	1	20GB	2GB	3 years, 9 months
Ansible micado	2	40GB	4GB	11 months
NewDataAvenue	2	40GB	4GB	6 months
MiniloLatest	2	40GB	4GB	5 months, 1 week
repo2docker	1	20GB	2GB	1 month, 3 weeks
neo4j	2	40GB	4GB	1 month
micado_master	2	40GB	4GB	1 hour, 17 minutes

Kötetek



Project ^

Compute ^

Overview

Instances

Volumes

Images

Access & Security

Network v

Orchestration v

Identity v

Volumes

Volumes | Volume Snapshots | Volume Consistency Groups

Filter

<input type="checkbox"/>	Name	Description	Size	Status	Type	Attached To	Availability Zone	Bootable	Encrypted	Actions
<input type="checkbox"/>	attila-test	-	1GiB	Available	-		nova	No	No	<input type="button" value="Edit Volume"/> <input type="button" value="v"/>
<input type="checkbox"/>	DataAvenueTest	-	4GiB	Available	-		nova	Yes	No	<input type="button" value="Edit Volume"/> <input type="button" value="v"/>
<input type="checkbox"/>	gabor_spark_master	-	250GiB	Available	-		nova	No	No	<input type="button" value="Edit Volume"/> <input type="button" value="v"/>
<input type="checkbox"/>	volume_for_hdfs_server	-	500GiB	Available	-		nova	No	No	<input type="button" value="Edit Volume"/> <input type="button" value="v"/>
<input type="checkbox"/>	volume_for_storage_new	-	2048GiB	Available	-		nova	No	No	<input type="button" value="Edit Volume"/> <input type="button" value="v"/>
<input type="checkbox"/>	volume_for_storage2	-	2048GiB	Available	-		nova	No	No	<input type="button" value="Edit Volume"/> <input type="button" value="v"/>

Képfájlok

Project ^

Compute ^

Overview

Instances

Volumes

Images

Access & Security

Network v

Orchestration v

Identity v

Images

Project (6) Shared with Me (0) Public (12) [+ Create Image](#) [Delete Images](#)

<input type="checkbox"/>	Image Name	Type	Status	Public	Protected	Format	Size	Actions
<input type="checkbox"/>	CentOS 7 Cloud image	Image	Active	Yes	No	Raw	8.0 GB	Launch v
<input type="checkbox"/>	DCI Bridge	Image	Active	Yes	No	Raw	2.2 GB	Launch v
<input type="checkbox"/>	DCI Bridge + Python	Image	Active	Yes	No	Raw	2.2 GB	Launch v
<input type="checkbox"/>	gUSE	Image	Active	Yes	No	Raw	4.2 GB	Launch v
<input type="checkbox"/>	gUSE-10	Image	Active	Yes	No	Raw	4.2 GB	Launch v
<input type="checkbox"/>	Ubuntu 14.04 LTS Cloud image	Image	Active	Yes	No	Raw	2.2 GB	Launch v
<input type="checkbox"/>	Ubuntu 16.04 LTS Cloud image	Image	Active	Yes	No	Raw	2.2 GB	Launch v
<input type="checkbox"/>	Ubuntu 16.04 LTS for Heat	Image	Active	Yes	No	Raw	2.6 GB	Launch v

Virtuális gép létrehozása

Kattintsunk a **Launch Instance** gombra.

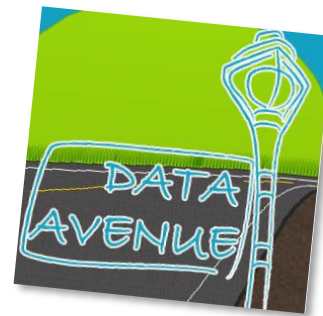


The screenshot shows the OpenStack dashboard interface. The top left corner displays the "openstack" logo and the user "oktatas". The main heading is "Instances". Below the heading, there is a search bar labeled "Instance Name =" and a "Filter" button. To the right of the search bar are three buttons: "Launch Instance" (with a cloud icon), "Delete Instances" (with a trash icon), and "More Actions". Below these buttons is a table with the following columns: Instance Name, Image Name, IP Address, Size, Key Pair, Status, Availability Zone, Task, Power State, Time since created, and Actions. The table currently contains no data, with the text "No items to display." centered below the column headers. On the left side of the dashboard, there is a navigation menu with the following items: Project, Compute, Overview, Instances (highlighted in red), Volumes, Images, Access & Security, and Network.

Adatmozgatás

- ▶ **ATK kapcsolódó kérdése:**
 - ▶ Hogyan lehet adatokat mozgatni a virtuális gépek és a kutatók tárhelyei között?
- ▶ **Válasz:**
 - ▶ A **Data Avenue ref. arch.** használatával
 - ▶ A következő diák mutatják ezt a lehetőséget

Mi az a Data Avenue?



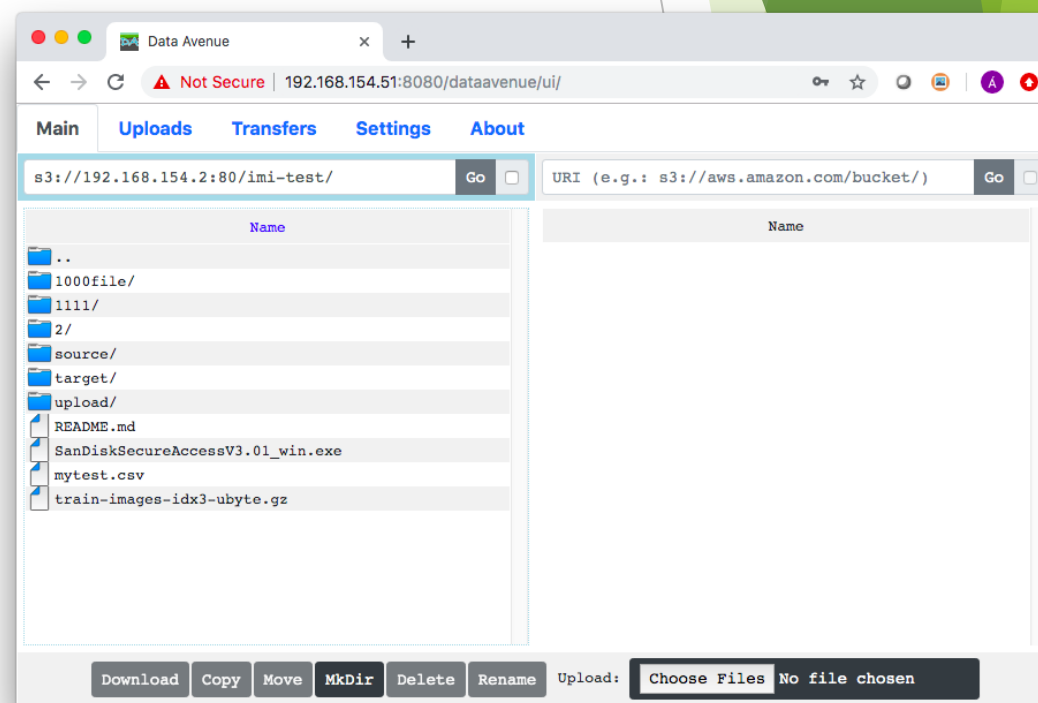
- ▶ Egy grafikus felülettel rendelkező tároló menedzsment szolgáltatás
 - ▶ Számos protokollt támogat: sFTP, Amazon S3, Apache HDFS
 - ▶ Az adatokat segítségével egyik szerverről a másikra mozgathatjuk

- ▶ Támogatja az olyan általános műveleteket, mint:

- ▶ könyvtárak és fájlok listázása
- ▶ könyvtárak létrehozása
- ▶ törlés
- ▶ átnevezés
- ▶ fájlok feltöltése és letöltése

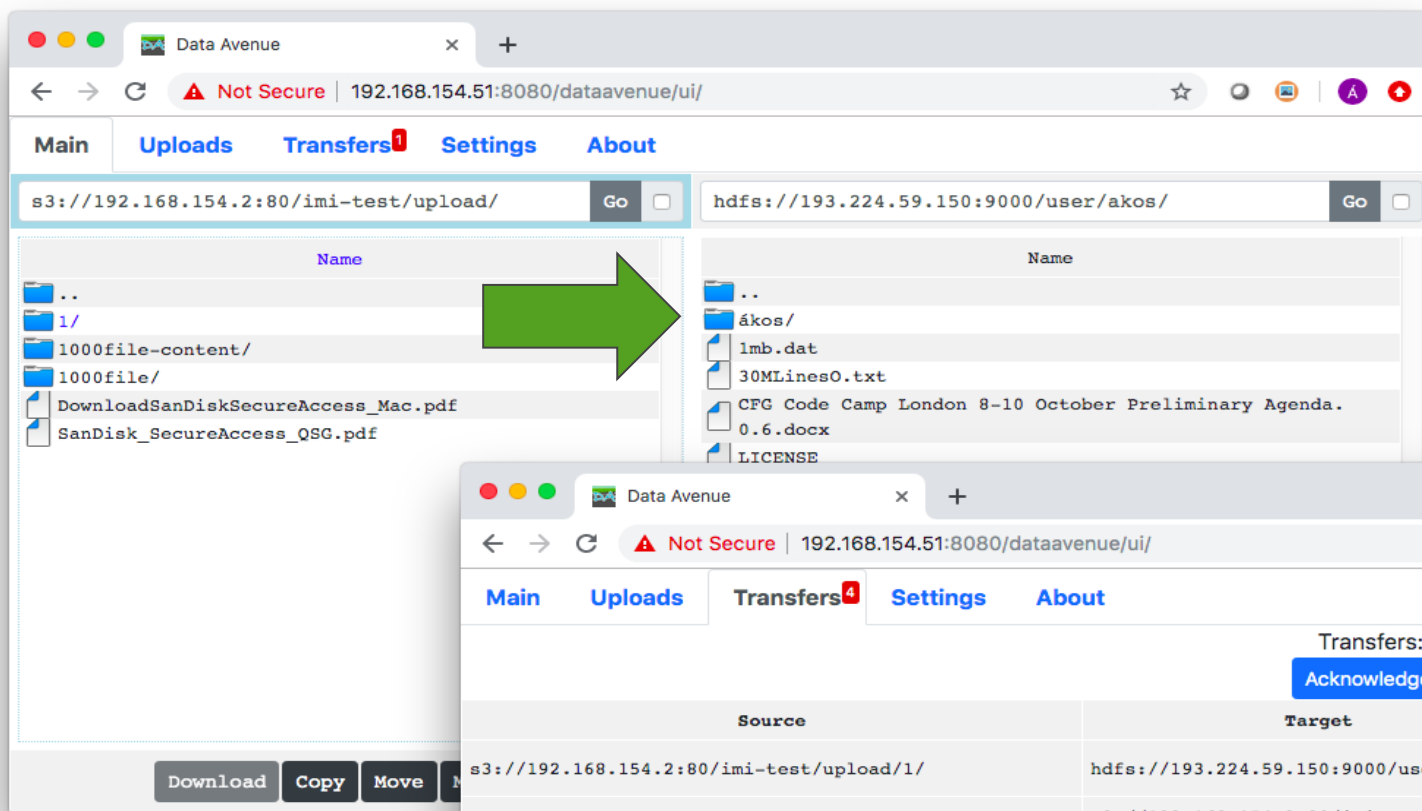
- ▶ REST API van és ezért programból is használható

- ▶ Nyílt forráskódú (<https://github.com/SZTAKI-LPDS/data-avenue>)



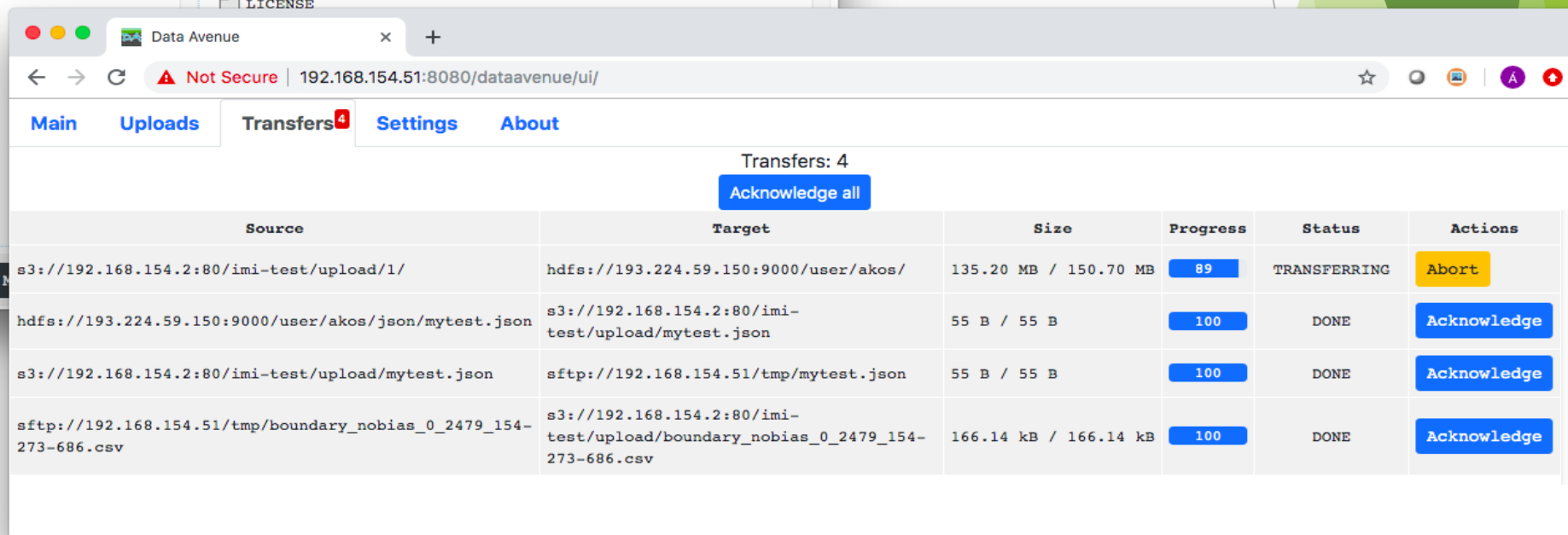
Adatok mozgatása

Fájlok és könyvtárak áthelyezése vagy másolása egyik helyről a másikra.



The screenshot shows the Data Avenue web interface with two file browser panes. The left pane shows the source directory: `s3://192.168.154.2:80/imi-test/upload/`. The right pane shows the target directory: `hdfs://193.224.59.150:9000/user/akos/`. A large green arrow points from the source directory to the target directory, indicating the direction of the file transfer.

Minden átvitelt monitoroz a szerver a háttérben, ami szintén nyomonkövethető.



The screenshot shows the Data Avenue web interface displaying a list of active transfers. The interface includes a 'Transfers: 4' indicator and an 'Acknowledge all' button. The table below lists the details of each transfer.

Source	Target	Size	Progress	Status	Actions
<code>s3://192.168.154.2:80/imi-test/upload/1/</code>	<code>hdfs://193.224.59.150:9000/user/akos/</code>	135.20 MB / 150.70 MB	89	TRANSFERRING	Abort
<code>hdfs://193.224.59.150:9000/user/akos/json/mytest.json</code>	<code>s3://192.168.154.2:80/imi-test/upload/mytest.json</code>	55 B / 55 B	100	DONE	Acknowledge
<code>s3://192.168.154.2:80/imi-test/upload/mytest.json</code>	<code>sftp://192.168.154.51/tmp/mytest.json</code>	55 B / 55 B	100	DONE	Acknowledge
<code>sftp://192.168.154.51/tmp/boundary_nobias_0_2479_154-273-686.csv</code>	<code>s3://192.168.154.2:80/imi-test/upload/boundary_nobias_0_2479_154-273-686.csv</code>	166.14 kB / 166.14 kB	100	DONE	Acknowledge

Data Avenue használati módok

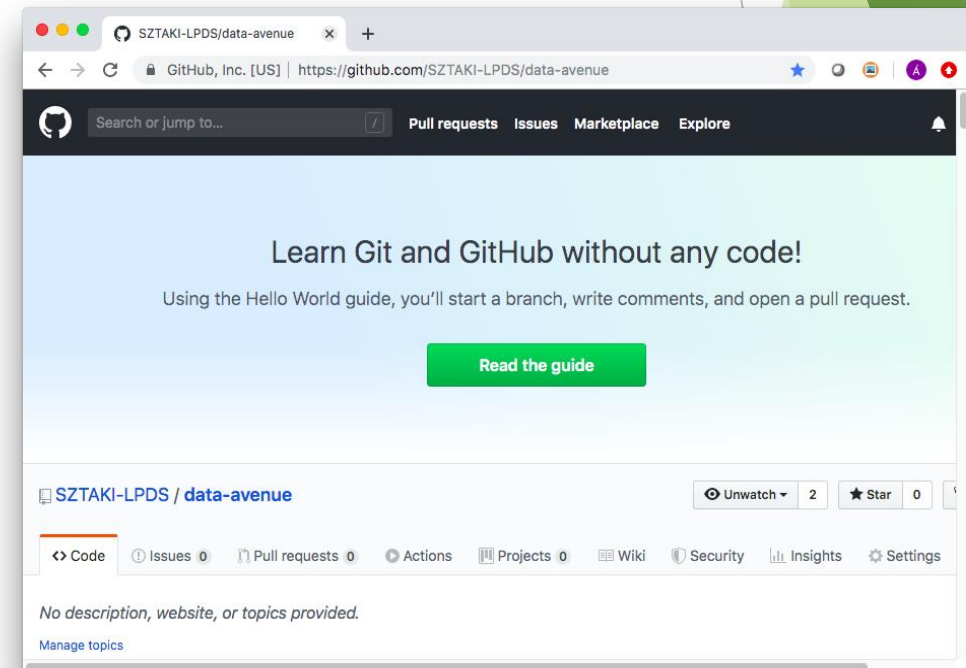
- ▶ Adat (fájl, könyvtár) feltöltése a felhasználó gépéről a felhőbe
- ▶ Adat (fájl, könyvtár) letöltése a felhőből a felhasználó gépére
- ▶ Adat (fájl, könyvtár) áttöltése más felhőből az ELKH Cloud-ba
- ▶ Adat (fájl, könyvtár) áttöltése ELKH Cloud-ból más felhőbe
- ▶ Adat (fájl, könyvtár) áttöltése külső tárolóból az ELKH Cloud-ba
- ▶ Adat (fájl, könyvtár) áttöltése ELKH Cloud-ból külső tárolóba

A Data Avenue előnyei



- ▶ **gyors** (többszálú, Java alapú, 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)
- ▶ **könnyű GUI** (javascript/angular alapú felület, a legtöbb böngésző támogatja)
- ▶ **nyomonkövethető** (naplózza az átviteleket)

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



Adatelérés

▶ ATK kapcsolódó kérdése:

- ▶ Biztonságban vannak a számítási felhőbe feltöltött kutatási adataim? Védve vannak-e az illetéktelen hozzáféréstől?

▶ Válasz:

- ▶ Igen, más projektekből nem lehet elérni őket.
- ▶ A projekten belül viszont egy másik projekt tag törölheti, vagy módosíthatja őket. De itt pont ez a cél, hogy az egy projekten dolgozók képesek legyenek együtt dolgozni.

Adatmozgatás

- ▶ **ATK kapcsolódó kérdése:**
 - ▶ Az adattárolás megfelel GDPR és a magyar info törvényben megfogalmazott szempontoknak? Mi erre a biztosíték?
- ▶ **Válasz:**
 - ▶ Igen, a cloud csapat semmiféle módon nem használja a projekteken tárolt adatokat.
 - ▶ A projekten belül viszont a projekt feladata a GDPR szabályok betartása. Ezt a felhasználóknak el is kell fogadniuk a projekt igénylésnél.

Cloud oktatás



- ▶ Tematika alapján kétféle tanfolyamot szervezünk:
 - ▶ Bevezető tanfolyam kezdőknek (2 havonta)
 - ▶ Haladó tanfolyamok különböző célcsoportoknak (2 havonta)
 - ▶ Mesterséges intelligencia tanfolyam (Tensorflow, Keras)
 - ▶ Big Data tanfolyam (Spark)
 - ▶ Adat kezelés és adat repositórium tanfolyam
 - ▶ Adat kezelési terv készítési tanfolyam
 - ▶ HTC (High throughput computing) tanfolyam v. emeltszintű felhasználói támogatás tanfolyam - előkészítés alatt
- ▶ Az előadás anyagok elérhetők az ELKH Cloud weblapján:
 - ▶ [Prezentációk | ELKH Cloud \(science-cloud.hu\)](https://science-cloud.hu)
- ▶ Az előadások videóra vett anyaga szintén elérhető lesz az ELKH Cloud új weblapján

További információk

- ▶ Web: <https://science-cloud.hu/>
- ▶ GYIK: <https://science-cloud.hu/gyakran-ismetelt-kerdesek>
- ▶ E-mail: info@science-cloud.hu



Köszönöm a figyelmet

Kérdések?