

<b>Nombre:</b> Jesús Alejandro Valencia Valadez		<b>Matrícula:</b> 3013480
<ul style="list-style-type: none"> <li><b>Nombre del curso:</b> Infraestructura para Big Data</li> </ul>		
<b>Módulo:</b> 1		<b>Actividad:</b> Avance 2 de Evidencia 1

## 1- Entrega de captura de pantalla de su máquina en Azure (o local)

Inicio >


**Máquinas virtuales** ✨ ...

Universidad Tecmilenio (utm.edu.mx)

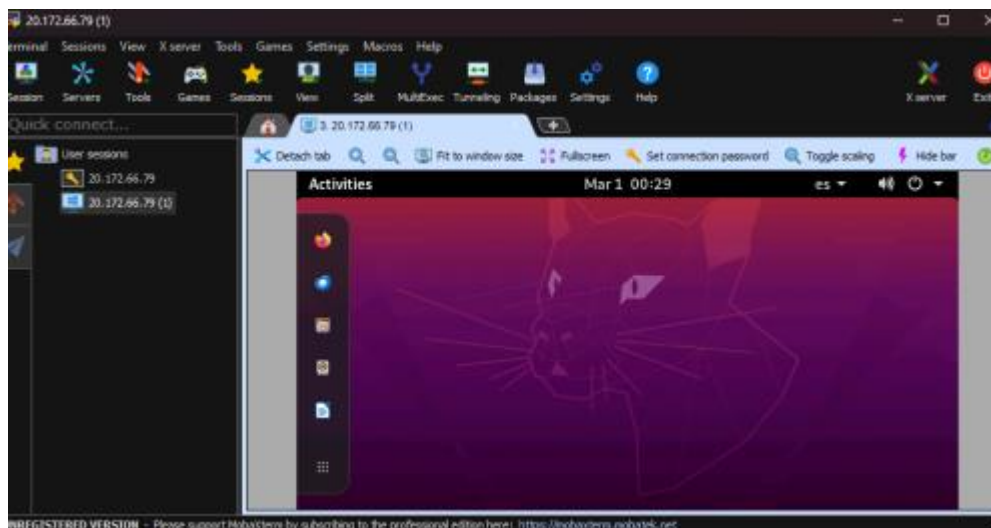
+ Crear ▾ Cambiar al modo clásico Reservas ▾ Administrar vista ▾ Actualizar ▾ Exportar a CSV Abrir consulta | Asignar etiquetas ▶ Iniciar ↺ Reiniciar □ Detener 🗑 Eliminar ☰

Filtrar por cualquier campo Suscripción es igual a **todo** Tipo es igual a **todo** Grupo de recursos es igual a **todo** Ubicación es igual a **todo** X Agregar filtro

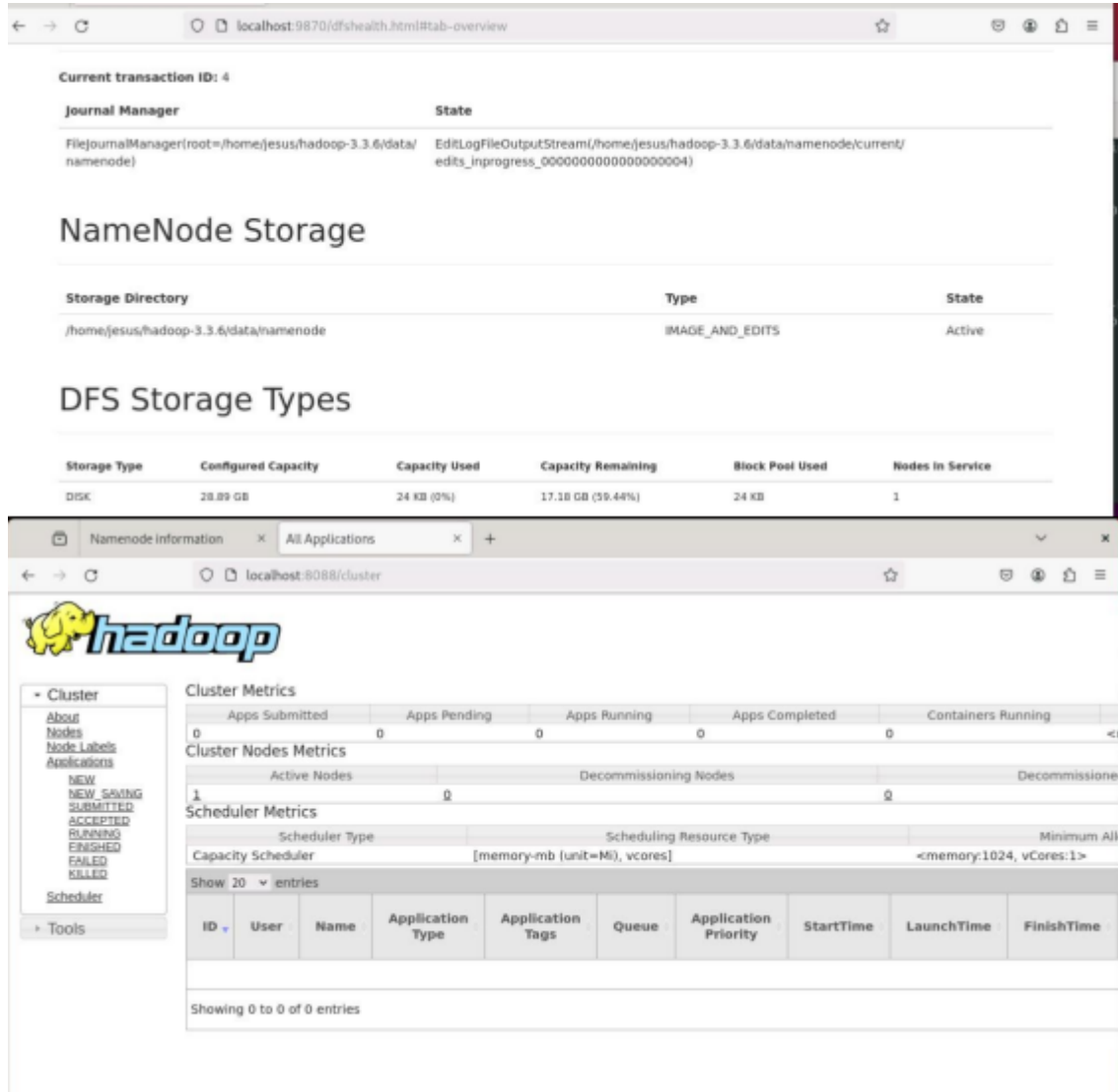
Mostrando de 1 a 1 de 1 registros.

<input type="checkbox"/> Nombre ↑↓	Tipo ↑↓	Suscripción ↑↓	Grupo de recursos ↑↓	Ubicación ↑↓	Estado ↑↓	Sistema operativo ↑↓	Tam
<input type="checkbox"/>  <b>jesus</b>	Máquina virtual	Azure for Students	jesus_group_02061455	westus3	Detenido (desasignado)	Linux	Stan

## 2- Entrega de captura de pantalla de ejecución de Ubuntu



### 3-Entrega de captura de pantalla de su clúster (Hadoop para quienes lo hicieron local)



The image shows two screenshots of Hadoop web interfaces. The top screenshot is the 'DFS Health' page, and the bottom screenshot is the 'Cluster Overview' page.

**DFS Health Page (Top):**

- Current transaction ID: 4
- Journal Manager**

Journal Manager	State
FileJournalManager(root=/home/jesus/hadoop-3.3.6/data/namenode)	EditLogFileOutputStream(/home/jesus/hadoop-3.3.6/data/namenode/current/edits_inprogress_0000000000000000004)
- NameNode Storage**

Storage Directory	Type	State
/home/jesus/hadoop-3.3.6/data/namenode	IMAGE_AND_EDITS	Active
- DFS Storage Types**

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes in Service
DISK	28.89 GB	24 KB (0%)	17.18 GB (59.44%)	24 KB	1

**Cluster Overview Page (Bottom):**

- Browser tabs: Namenode Information, All Applications
- URL: localhost:8088/cluster
- Cluster Metrics**

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running
0	0	0	0	0
- Cluster Nodes Metrics**

Active Nodes	Decommissioning Nodes	Decommissioned Nodes
1	0	0
- Scheduler Metrics**

Scheduler Type	Scheduling Resource Type	Minimum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>
- Applications Table**

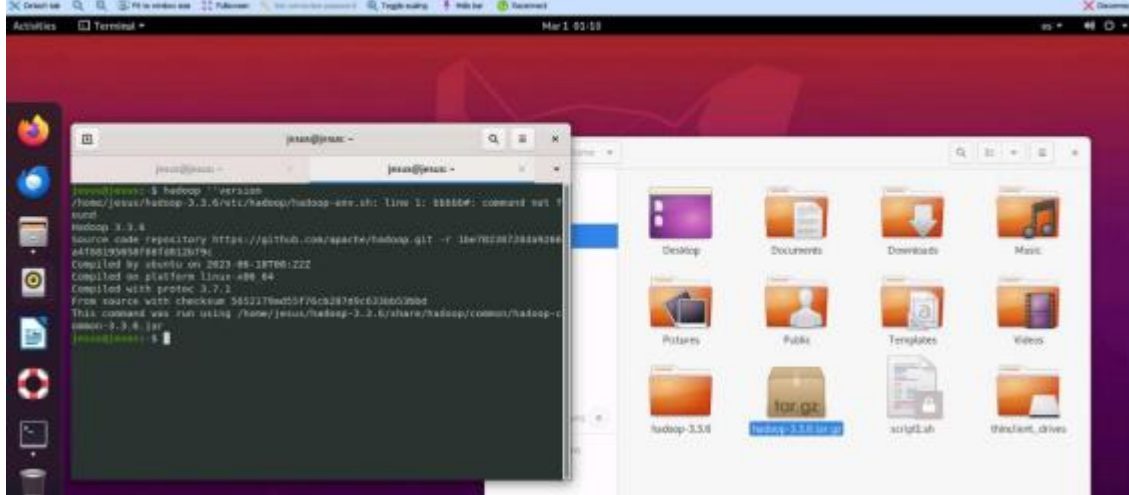
Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime
Showing 0 to 0 of 0 entries									

```

2024-03-01 02:20:14,021 INFO namenode.FSImage: FSImageSaver clean checkpoint:
txid=0 when meet shutdown.
2024-03-01 02:20:14,021 INFO namenode.NameNode: SHUTDOWN_MSG:
*****
SHUTDOWN_MSG: Shutting down NameNode at jesu.internal.cloudapp.net/10.1.0.4
*****
jesu@jesu:~$ /home/jesu/hadoop-3.3.6/sbin/start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [jesu]
2024-03-01 02:20:46,581 WARN util.NativeCodeLoader: Unable to load native-had
oop library for your platform... using builtin-java classes where applicable
jesu@jesu:~$ /home/jesu/hadoop-3.3.6/sbin/start-yarn.sh
Starting resourcemanager
Starting nodemanagers
jesu@jesu:~$ jps
21744 NodeManager
22099 Jps
21412 SecondaryNameNode
21589 ResourceManager
21158 DataNode
20998 NameNode
jesu@jesu:~$ /home/jesu/hadoop-3.3.6/sbin/stop-dfs.sh

```



#### 4- Investigación de ¿Qué es Ambari View? (Máximo una cuartilla, mínimo media cuartilla)

**Apache Ambari** es un proyecto de **software** desarrollado por la **Fundación Apache**. Su función principal es **facilitar la administración y supervisión de clústeres de Hadoop**, permitiendo a los usuarios integrar Hadoop con su infraestructura empresarial existente. Anteriormente, Ambari era un subproyecto, pero actualmente forma parte del nivel superior de proyectos de Apache.

Empresas como **IBM, Hortonworks, Spotify y eBay** utilizan Ambari para simplificar la operación de clústeres Hadoop. Ofrece una amplia variedad de funciones para la administración, incluyendo el monitoreo y la instalación de clústeres.

En cuanto al **costo**, este puede variar según varios factores:

1. **Hardware:** Para clústeres pequeños de hasta 5 nodos, el costo de hardware puede rondar los **\$10,000**. Sin embargo, si se desea crear un clúster con más de 100 nodos, el costo podría superar los **\$100,000**.

2. **Software:** La ejecución de un clúster Hadoop también implica costos de software. Por ejemplo, Hadoop Enterprise tiene un precio de **\$2,000 por nodo**.

Ambari está diseñado principalmente para clústeres grandes y complejos debido a estos costos. No obstante, también puede ser útil para implementaciones más pequeñas si se prefiere un enfoque manual.

En cuanto a los **servicios en los que se puede implementar**, Ambari es versátil:

- **Almacenamiento y procesamiento de grandes cantidades de datos.**
- **Educación o entornos escolares:** Puede llevar un registro completo de la escuela y los alumnos.
- **Gobierno:** Debido a su alta seguridad, es ideal para proteger datos sensibles.

En resumen, Ambari es una herramienta valiosa para la administración de clústeres Hadoop, independientemente de su tamaño o complejidad.

## 5- Captura de pantalla del ejercicio en mongo código en Python e inserción en Mongo compass

```
[1] !pip install pymongo

Collecting pymongo
  Downloading pymongo-4.7.0-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (670 kB)
    670.1/670.1 kB 4.0 MB/s eta 0:00:00
Collecting dnspython<3.0.0, >=1.16.0 (from pymongo)
  Downloading dnspython-2.6.1-py3-none-any.whl (307 kB)
    307.7/307.7 kB 23.6 MB/s eta 0:00:00
Installing collected packages: dnspython, pymongo
Successfully installed dnspython-2.6.1 pymongo-4.7.0

[2] import pymongo

[3] # To connect to MongoDB server:
client = pymongo.MongoClient('localhost', 27017)

# Create Database:
db = client.mongodb_workshop

# Create Collection:
collection = db.collection["coach_details"]

Data to store/insert into MongoDB Database:

[4] # One document:
doc_1 = {'_id': 1, "first_name": "Roldoro", "gender": "M", "shift": ["AM", "PM"], "ph_num": 456378903, "trains_for": ["Valleyball", "basketball"], "manager": "Jack"}

[5] # Multiple documents:
docs = [
    {"first_name": "Miguel", "id": 1, "shift": "AM", "ph_num": 234564900, "trains_for": ["tennis", "badminton"], "manager": "Jack"},
    {"first_name": "Victor", "gender": "M", "shift": "AM", "ph_num": 907654321, "trains_for": ["Valleyball", "basketball"], "manager": "Amin"},
    {"first_name": "Rossi", "gender": "M", "shift": "PM", "ph_num": 456234900, "trains_for": "badminton", "manager": "Amin"},
    {"first_name": "alberto", "gender": "M", "shift": "AM", "ph_num": 123456789, "trains_for": "tennis", "manager": "Jack"},
    {"first_name": "Emma", "gender": "F", "shift": "PM", "ph_num": 776655443, "trains_for": "indoor archery", "manager": "Amin"},
    {"first_name": "Hira", "gender": "F", "shift": ["AM", "PM"], "ph_num": 246813579, "trains_for": ["Valleyball", "basketball"], "manager": "Jack"},
    {"first_name": "Isha", "gender": "F", "shift": "PM", "ph_num": 135792468, "trains_for": "indoor archery", "manager": "Amin"},
    {"first_name": "Alice", "gender": "M", "shift": "AM", "ph_num": 2244668800, "trains_for": "badminton", "manager": "Jack"},
    {"first_name": "Deepali", "gender": "F", "shift": ["AM", "PM"], "ph_num": 113355779, "trains_for": ["tennis", "badminton"], "manager": "Amin"}
]
```