



DIGITAL
TALENT
SCHOLARSHIP

DIGITAL TALENT SCHOLARSHIP 2019

Big Data Analytics



EMR Lab: Sqoop, Oozie

Oleh: Imam Cholissodin | imamcs@ub.ac.id, Putra Pandu Adikara, Sufia Adha Putri

Asisten: Guedho, Sukma, Anshori, Aang dan Gusti

Fakultas Ilmu Komputer (Filkom) Universitas Brawijaya (UB)

Pokok Pembahasan

- Menggunakan Sqoop pada EMR
- Menggunakan Oozie pada EMR
- Tugas



TERBUKA
UNTUK
DISABILITAS

BREAK
YOUR
LIMITS!

Lab Sqoop

S Q Q O O P

Install Sqoop

- Requirements : Hadoop, Java, Database Connector
- Install Sqoop -> <https://www.apache.org/dist/sqoop/>



Sqoop di EMR

- Buat cluster pada layanan EMR, beri centang seperti berikut :

The screenshot shows the 'Create Cluster - Advanced Options' page in the AWS Management Console. The 'Software Configuration' section is highlighted, showing various software packages that can be included in the cluster. Arrows point to the checked boxes for Hadoop 2.8.5, Sqoop 1.4.7, and Oozie 5.1.0, indicating they should be selected. Other packages like Zeppelin, Tez, and Flink are also listed.

Step 1: Software and Steps

Software Configuration

Release: emr-5.27.0

Selected	Software
<input checked="" type="checkbox"/>	Hadoop 2.8.5
<input type="checkbox"/>	JupyterHub 1.0.0
<input type="checkbox"/>	Ganglia 3.7.2
<input type="checkbox"/>	Hive 2.3.5
<input type="checkbox"/>	MXNet 1.4.0
<input type="checkbox"/>	Hue 4.4.0
<input type="checkbox"/>	Spark 2.4.4
<input checked="" type="checkbox"/>	Sqoop 1.4.7
<input type="checkbox"/>	Zeppelin 0.8.1
<input type="checkbox"/>	Tez 0.9.2
<input type="checkbox"/>	HBase 1.4.10
<input type="checkbox"/>	Presto 0.224
<input type="checkbox"/>	Phoenix 4.14.2
<input type="checkbox"/>	HCatalog 2.3.5
<input type="checkbox"/>	Livy 0.6.0
<input type="checkbox"/>	Flink 1.8.1
<input checked="" type="checkbox"/>	Pig 0.17.0
<input type="checkbox"/>	ZooKeeper 3.4.14
<input type="checkbox"/>	Mahout 0.13.0
<input checked="" type="checkbox"/>	Oozie 5.1.0
<input type="checkbox"/>	TensorFlow 1.14.0

Multi-master support: Enable multi-master support

Edit software settings: Enter configuration Load JSON from S3

```
classification=config-file-name,properties=[myKey1=myValue1,myKey2=myValue2]
```

Add steps (optional)

Step type: Select a step | Configure

Feedback English (US) © 2008 - 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

- Kemudian next

Sqoop di EMR

- Berikan mask Spot untuk master dan core node
- Ubah instance type menjadi m3.xlarge

The screenshot shows the AWS EMR Instance Configuration page. It lists three node types: Master, Core, and Task. For each node type, there is a dropdown for 'Instance type' and a section for 'Purchasing option'. Arrows point from the 'Instance type' dropdowns to the 'Spot' radio button in the 'Purchasing option' section for both the Master and Core nodes.

Node type	Instance type	Instance count	Purchasing option	Auto Scaling
Master Master - 1	m3.xlarge 8 vCore, 15 GiB memory, 80 SSD GB storage EBS Storage: none Add configuration settings	1 Instances	<input type="radio"/> On-demand <input checked="" type="radio"/> Spot Use on-demand as max price	Not available for Master
Core Core - 2	m3.xlarge 8 vCore, 15 GiB memory, 80 SSD GB storage EBS Storage: none Add configuration settings	2 Instances	<input type="radio"/> On-demand <input checked="" type="radio"/> Spot Use on-demand as max price	Not enabled
Task Task - 3	m5.xlarge 4 vCore, 16 GiB memory, EBS only storage EBS Storage: 64 GiB Add configuration settings	0 Instances	<input type="radio"/> On-demand <input type="radio"/> Spot Use on-demand as max price	Not enabled

Choose the instance type, number of instances, and a purchasing option. You can choose to use On-Demand Instances, Spot Instances, or both. The instance type and purchasing option apply to all EC2 instances in each instance group, and you can only specify these options for an instance group when you create it. [Learn more about instance purchasing options](#)

Feedback English (US) © 2008 - 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

- Kemudian klik Next

Sqoop di EMR

- Klik Next terus hingga muncul tampilan seperti berikut :
- Pilih EC2 key pair yang pernah dibuat sebelumnya

Create Cluster - Advanced Options [Go to quick options](#)

Step 1: Software and Steps
Step 2: Hardware
Step 3: General Cluster Settings
Step 4: Security

Security Options

EC2 key pair **golobok** [i](#)

Cluster visible to all IAM users in account [i](#)

Permissions [i](#)

Default Custom

Use default IAM roles. If roles are not present, they will be automatically created for you with managed policies for automatic policy updates.

EMR role [EMR_DefaultRole](#) [i](#)

EC2 instance profile [EMR_EC2_DefaultRole](#) [i](#)

Auto Scaling role [EMR_AutoScaling_DefaultRole](#) [i](#)

► Security Configuration
► EC2 security groups

[Cancel](#) [Previous](#) **Create cluster**

[Feedback](#) [English \(US\)](#)

© 2008 - 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved. [Privacy Policy](#) [Terms of Use](#)

- Kemudian klik Create Cluster

Sqoop di EMR

- Tunggu status **Starting** menjadi **Waiting**.

Cluster: My cluster **Starting** Configuring cluster software

Cluster: My cluster **Waiting** Cluster ready after last step completed.

Summary	Configuration details
ID: j-259E7HFBH74K8	Release label: emr-5.25.0
Creation date: 2019-07-30 03:14 (UTC+7)	Hadoop distribution: Amazon 2.8.5
Elapsed time: 17 minutes	Applications: Hive 2.3.5, Pig 0.17.0, Hue 4.4.0, Spark 2.4.3, HBase 1.4.9, ZooKeeper 3.4.14
Auto-terminate: No	Log URI: s3://aws-logs-146080089880-us-east-1/elasticmapreduce/
Termination On Change protection:	EMRFS consistent view: Disabled
	Custom AMI ID: --

Sqoop di EMR

- Sebelum melakukan remote ssh ke EMR Cluster, lakukan perubahan **Security Group** pada Master node

The screenshot shows the AWS Management Console interface for an EMR cluster named 'TERRIKA'. The left sidebar lists various cluster management options like Clusters, Security configurations, and VPC subnets. The main content area displays the cluster's summary, including its ID (j-2UP80WFK03UAO), creation date (2019-10-05 01:05 UTC+7), and auto-termination settings. The 'Network and hardware' section details the availability zone (us-east-1a), subnet ID (subnet-0434d549), and task node configurations. The 'Security and access' section shows the EC2 instance profile (EMR_EC2_DefaultRole), EMR role (EMR_DefaultRole), Auto Scaling role (EMR_AutoScaling_DefaultRole), and visible users (All). The 'Security groups' section is highlighted with a red arrow, showing two security groups assigned to the master node: sg-04db114224bcbbf84 and sg-095c37789ee453237. The master node's role is listed as (ElasticMapReduce-master) under the second security group.

AWS Services Resource Groups Summary Application history Monitoring Hardware Configurations Events Steps Bootstrap actions

vocstartsoft/user3

Amazon EMR

Clusters Security configurations Block public access VPC subnets Events Notebooks Help What's new

Connections: Enable Web Connection – Resource Manager ... (View All)

Master public DNS: ec2-3-95-180-214.compute-1.amazonaws.com SSH

Tags: -- View All / Edit

Summary

ID: j-2UP80WFK03UAO
Creation date: 2019-10-05 01:05 (UTC+7)
Elapsed time: 7 minutes
Auto-terminate: No
Termination On Change protection:

Configuration details

Release label: emr-5.27.0
Hadoop distribution: Amazon 2.8.5
Applications: Pig 0.17.0, Sqoop 1.4.7, Oozie 5.1.0
Log URI: s3://aws-logs-056823078307-us-east-1/elasticmapreduce/

EMRFS consistent view: Disabled

Custom AMI ID: --

Network and hardware

Availability zone: us-east-1a
Subnet ID: subnet-0434d549
Master: Bootstrapping 1 m3.xlarge Spot (max on-demand)
Core: Provisioning 2 m3.xlarge Spot (max on-demand)
Task: --

Security and access

Key name: golobok
EC2 instance profile: EMR_EC2_DefaultRole
EMR role: EMR_DefaultRole
Auto Scaling role: EMR_AutoScaling_DefaultRole
Visible to all users: All Change

Security groups for sg-04db114224bcbbf84
Master: (ElasticMapReduce-master)

Security groups for sg-095c37789ee453237
Core & Task: (ElasticMapReduce-slave)

Sqoop di EMR

- Klik **inbound**, kemudian pilih **edit**

The screenshot shows the AWS Management Console interface for managing security groups. The left sidebar lists various AWS services. The main area displays a table of security groups with columns for Name, Group ID, Group Name, VPC ID, Owner, and Description. Two groups are listed: 'ElasticMapReduce-master' and 'ElasticMapReduce-slave'. Below the table, a specific security group ('sg-04db114224bcbbf84') is selected. The 'Inbound' tab is active, and an arrow points to the 'Edit' button. Another arrow points to the 'Inbound' tab itself. The bottom section shows a detailed list of inbound rules with columns for Type, Protocol, Port Range, Source, and Description.

Name	Group ID	Group Name	VPC ID	Owner	Description
sg-04db114224bcbbf84	sg-04db114224bcbbf84	ElasticMapReduce-master	vpc-580c6122	056823078307	Master group for Elastic MapReduce cre
sg-095c37789ee453237	sg-095c37789ee453237	ElasticMapReduce-slave	vpc-580c6122	056823078307	Slave group for Elastic MapReduce crea

Security Group: sg-04db114224bcbbf84

Inbound

Edit

Type	Protocol	Port Range	Source	Description
All TCP	TCP	0 - 65535	sg-04db114224bcbbf84 (ElasticM	
All TCP	TCP	0 - 65535	sg-095c37789ee453237 (ElasticM	
SSH	TCP	22	175.45.188.252/32	
SSH	TCP	22	::/0	
Custom TCP Rule	TCP	8443	207.171.167.101/32	

Sqoop di EMR

- Add Rule, tambahkan SSH dengan source Anywhere dan Save

The screenshot shows the AWS EC2 Security Groups interface. On the left, there's a sidebar with various EC2-related options like Dashboard, Events, Tags, Reports, Limits, Instances, Launch Templates, Spot Requests, Reserved Instances, Dedicated Hosts, Scheduled Instances, Capacity Reservations, AMIs, Bundle Tasks, and Volumes. The main area displays a list of existing security group rules. A new rule is being added at the bottom:

Protocol & Port Range	Source	Description
Custom TCP F	TCP	8443
Custom TCP F	TCP	8443
Custom TCP F	TCP	8443
Custom TCP F	TCP	8443
Custom TCP F	TCP	8443
Custom TCP F	TCP	8443
Custom TCP F	TCP	8443
Custom TCP F	TCP	8443
All UDP	UDP	0 - 65535
All UDP	UDP	0 - 65535
All ICMP - IPv4	ICMP	0 - 65535
All ICMP - IPv4	ICMP	0 - 65535
SSH	TCP	22

The 'Protocol & Port Range' row for the new rule has 'SSH' selected from a dropdown. The 'Source' row has 'Anywhere' selected from a dropdown. A cursor arrow points to the 'Save' button at the bottom right of the form.

NOTE: Any edits made on existing rules will result in the edited rule being deleted and a new rule created with the new details. This will cause traffic that depends on that rule to be dropped for a very brief period of time until the new rule can be created.

Cancel Save

Sqoop di EMR

- Kembali ke EMR Cluster, kemudian pilih SSH

Connections: Enable Web Connection – Resource Manager ... (View All)

Master public DNS: ec2-3-95-180-214.compute-1.amazonaws.com SSH

Tags:

Buka Putty, copy dan paste host name ke Putty

Cluster: My cluster Waiting Cluster ready after last step completed.

SSH

Connect to the Master Node Using SSH

You can connect to the Amazon EMR master node using SSH to run interactive queries, examine local logs, or access the Hadoop file system. [more ↗](#)

Windows Mac / Linux

- Download PuTTY.exe to your computer from: <http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html> ↗
- Start PuTTY.
- In the Category list, click Session.
- In the Host Name field, type **hadoop@ec2-3-95-180-214.compute-1.amazonaws.com**
- In the Category list, expand Connection > SSH, and then click Auth.
- For Private key file for authentication, click Browse and select the private key file (**golobok**).
- Click Open.
- Click Yes to dismiss the security alert.

Subnet ID: subnet-0434d549 ↗

Master: Bootstrapping 1 m3.xlarge Spot (may on-demand)

EC2 instance profile: EMR_EC2

EMR role: EMR_DefaultRole

PuTTY Configuration

Category: Session

Host Name (or IP address) hadoop@ec2-3-95-180-214.compute-1.a Port 22

Connection type: SSH

Load, save or delete a stored session

Saved Sessions

Default Settings

Close window on exit: Only on clean exit

Sqoop di EMR

- Kemudian koneksi putty seperti biasa hingga muncul tampilan seperti dibawah ini :

TERBUKA

```
hadoop@ip-172-31-19-46:~
```

Amazon Linux AMI

<https://aws.amazon.com/amazon-linux-ami/2018.03-release-notes/>
5 package(s) needed for security, out of 13 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEE MMMMMMMMM MBBBBBBBBB RRRRRRRRRRRRRRR
E:::::::::::E M:::::M M:::::M R:::::R R:::::R
EE:::::EEEEEEEEE::::E M:::::M M:::::M R:::::RRRRRRR:::::R
E:::::E EEEEEEE M:::::M M:::::M R:::::R R:::::R
E:::::E M:::::M::::M M:::::M::::M R:::::R R:::::R
E:::::EEEEEEEEE M:::::M M:::::M M:::::M R:::::RRRRRRR:::::R
E:::::::::::E M:::::M M:::::M M:::::M R:::::::::::RR
E:::::EEEEEEEEE M:::::M M:::::M M:::::M R:::::RRRRRRR:::::R
E:::::E M:::::M M:::::M M:::::M R:::::R R:::::R
E:::::E EEEEEEE M:::::M MMM M:::::M R:::::R R:::::R
EE:::::EEEEEEEEE::::E M:::::M M:::::M R:::::R R:::::R
E:::::::::::E M:::::M M:::::M R:::::R R:::::R
EEEEEEEEEEEEEEEEEE MMMMMMM RRRRRRRR RRRRRRRR

```
[hadoop@ip-172-31-19-46 ~] $
```

Sqoop di EMR

■ Cek versi Hadoop

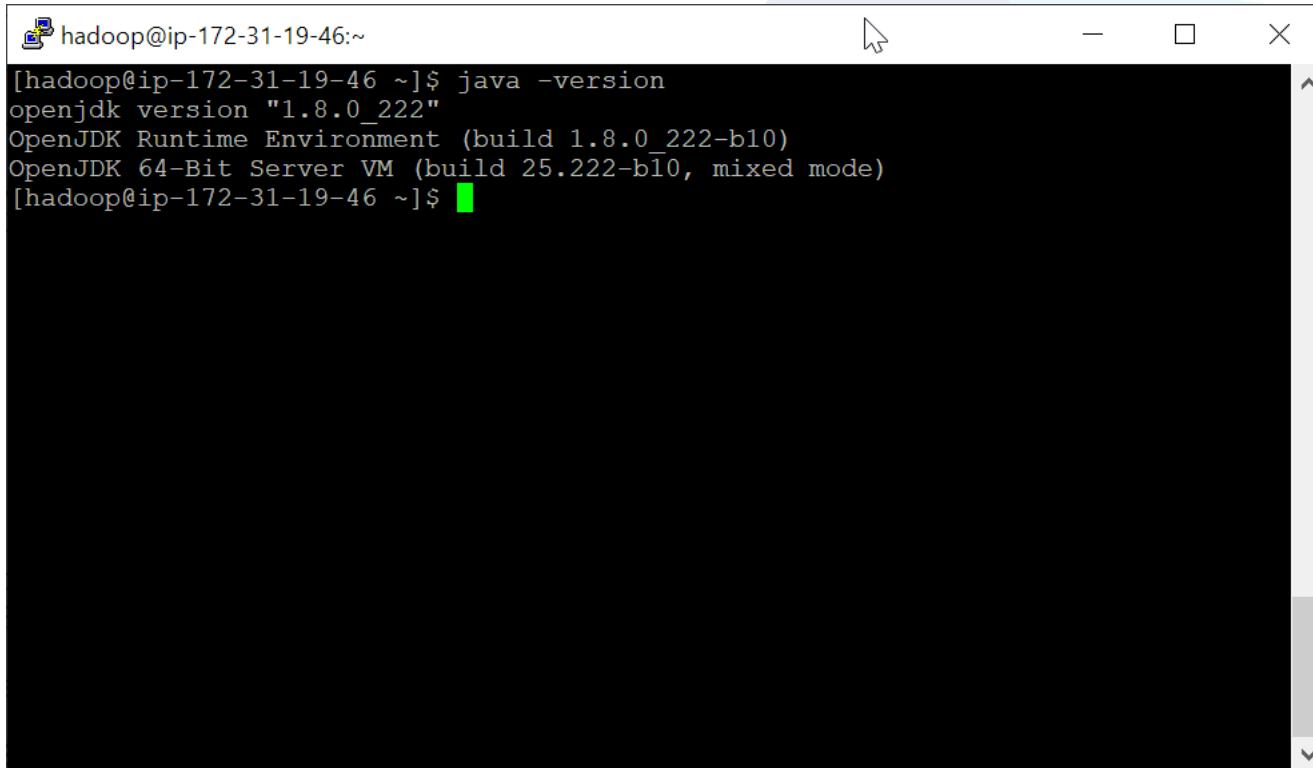
```
hadoop@ip-172-31-19-46:~]$ hadoop version
Hadoop 2.8.5-amzn-4
Subversion git@aws157git.com:/pkg/Aws157BigTop -r 082e98ac708024722b9a71933768ec
ad2e086acc
Compiled by ec2-user on 2019-09-10T22:35Z
Compiled with protoc 2.5.0
From source with checksum 9a7aa506e8de22caa8349aa16929
This command was run using /usr/lib/hadoop/hadoop-common-2.8.5-amzn-4.jar
[hadoop@ip-172-31-19-46 ~]$
```

■ Cek versi Sqoop

```
hadoop@ip-172-31-19-46:~]$ sqoop version
Warning: /usr/lib/sqoop/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
19/10/04 18:21:19 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
Sqoop 1.4.7
git commit id 082e98ac708024722b9a71933768ecad2e086acc
Compiled by ec2-user on Wed Sep 11 02:15:43 UTC 2019
[hadoop@ip-172-31-19-46 ~]$
```

Sqoop di EMR

■ Cek versi Java



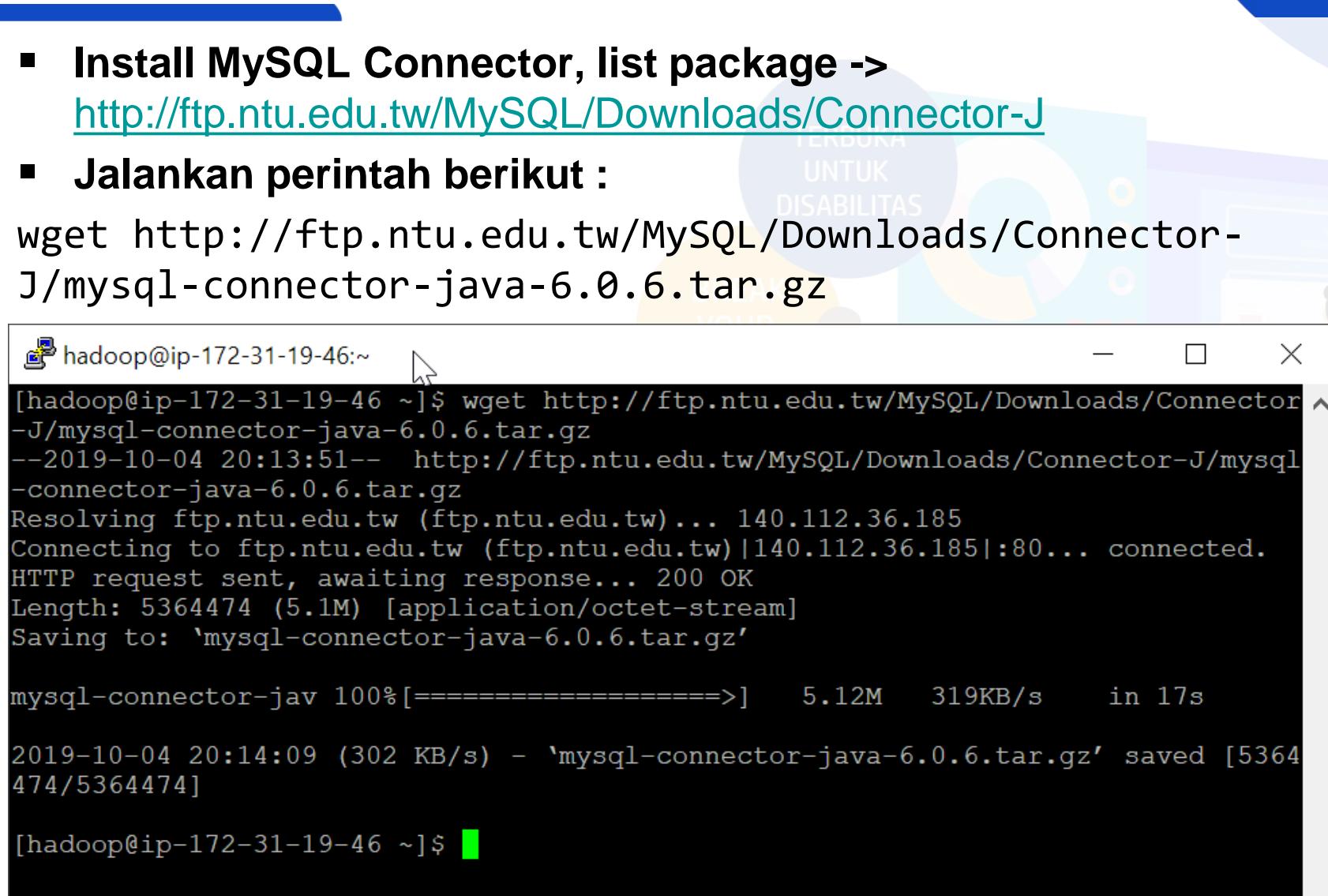
```
[hadoop@ip-172-31-19-46:~]$ java -version
openjdk version "1.8.0_222"
OpenJDK Runtime Environment (build 1.8.0_222-b10)
OpenJDK 64-Bit Server VM (build 25.222-b10, mixed mode)
[hadoop@ip-172-31-19-46 ~]$
```

Sqoop di EMR

- **Install MySQL Connector, list package ->**
<http://ftp.ntu.edu.tw/MySQL/Downloads/Connector-J>

- **Jalankan perintah berikut :**

```
wget http://ftp.ntu.edu.tw/MySQL/Downloads/Connector-J/mysql-connector-java-6.0.6.tar.gz
```



```
[hadoop@ip-172-31-19-46:~]$ wget http://ftp.ntu.edu.tw/MySQL/Downloads/Connector-J/mysql-connector-java-6.0.6.tar.gz
--2019-10-04 20:13:51--  http://ftp.ntu.edu.tw/MySQL/Downloads/Connector-J/mysql-connector-java-6.0.6.tar.gz
Resolving ftp.ntu.edu.tw (ftp.ntu.edu.tw) ... 140.112.36.185
Connecting to ftp.ntu.edu.tw (ftp.ntu.edu.tw)|140.112.36.185|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 5364474 (5.1M) [application/octet-stream]
Saving to: 'mysql-connector-java-6.0.6.tar.gz'

mysql-connector-jav 100%[=====] 5.12M 319KB/s in 17s

2019-10-04 20:14:09 (302 KB/s) - 'mysql-connector-java-6.0.6.tar.gz' saved [5364474/5364474]

[hadoop@ip-172-31-19-46:~]$
```

Sqoop di EMR

- tar -zxf mysql-connector-java-6.0.6.tar.gz
 - cd mysql-connector-java-6.0.6.tar.gz
 - sudo su
 - mv mysql-connector-java-6.0.6.tar.gz /usr/lib/sqoop/lib

```
[hadoop@ip-172-31-19-46:~/mysql-connector-java-6.0.6]$ tar -zxf mysql-connector-java-6.0.6.tar.gz
[hadoop@ip-172-31-19-46 ~]$ cd mysql-connector-java-6.0.6
[hadoop@ip-172-31-19-46 mysql-connector-java-6.0.6]$ sudo su
[REDACTED]
[hadoop@ip-172-31-19-46 mysql-connector-java-6.0.6]$ ls
[hadoop@ip-172-31-19-46 mysql-connector-java-6.0.6]$ sudo mv mysql-connector-java-6.0.6-bin.jar /usr/lib/sqoop/lib
```

Praktik Sqoop

- Sebelum praktik Sqoop, membuat database terlebih dahulu
- Pilih service **RDS** pada AWS, dan **Create database**

The screenshot shows the Amazon RDS Dashboard. On the left, a sidebar lists various services: Dashboard, Databases, Query Editor, Performance Insights, Snapshots, Automated backups, Reserved instances, Subnet groups, Parameter groups, Option groups, Events, Event subscriptions, and Recommendations (with 2 notifications). The main content area is titled "Amazon Aurora" and contains a brief description: "Amazon Aurora is a MySQL- and PostgreSQL-compatible enterprise-class database, starting at <\$1/day. Aurora supports up to 64TB of auto-scaling storage capacity, 6-way replication across three availability zones, and 15 low-latency read replicas." It features a prominent orange "Create database" button with a cursor hovering over it, and a smaller link "Or, Restore Aurora DB cluster from S3". Below this, there's a "Resources" section with a "Refresh" button. It displays resource counts: DB Instances (2/40), DB Clusters (0/40), Reserved instances (0/40), Snapshots (11), Recent events (10), Event subscriptions (0/20), Parameter groups (1), Option groups (1), Subnet groups (1/50), Supported platforms VPC, and Default network vpc-580c6122.

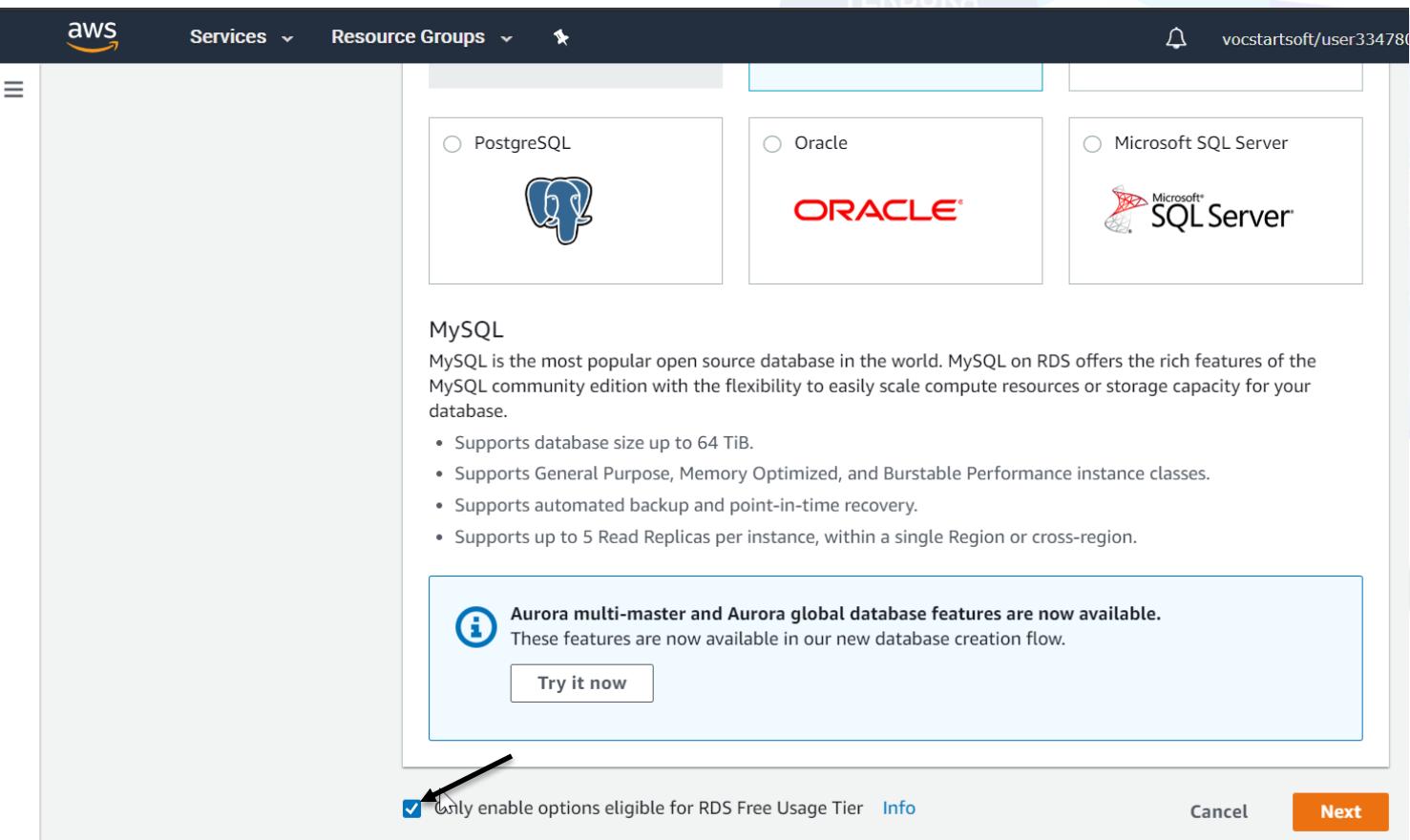
Praktik Sqoop

■ Pilih engine MySQL

The screenshot shows the AWS RDS 'Create database' wizard. The top navigation bar includes the AWS logo, 'Services' dropdown, 'Resource Groups' dropdown, and a user icon. A blue banner at the top says 'Switch to the new database creation flow.' Below it, the steps are listed: Step 1 'Select engine', Step 2 'Specify DB details', and Step 3 'Configure advanced settings'. The current step is 'Select engine'. The title 'Select engine' is centered above a grid of engine options. Each option has a radio button and an icon. The MySQL option is selected, indicated by a blue outline around the radio button and a larger blue outline around the entire box. Other options shown are Amazon Aurora (with a seal icon), MariaDB (with a seal icon), PostgreSQL (with a blue elephant icon), Oracle (with a red 'ORACLE' logo), and Microsoft SQL Server (with a red and white logo). At the bottom of the grid, the text 'MySQL' is followed by a small note about MySQL being the most popular open source database.

Praktik Sqoop

- Scroll kebawah lalu berikan centang seperti gambar dibawah, kemudian next



Praktik Soop

- Setelah next, scroll kebawah hingga menemui tampilan seperti dibawah. Isi setting seperti berikut kemudian **next**

Charges will apply when your database autoscales to the specified threshold
1000 GiB
(Minimum: 21 GiB, Maximum: 16384 GiB)

Settings

DB instance identifier [Info](#)
Specify a name that is unique for all DB instances owned by your AWS account in the current region.
mydatabase

Master username [Info](#)
Specify an alphanumeric string that defines the login ID for the master user.
root

Master password [Info](#) **Confirm password** [Info](#)
Master Password must be at least eight characters long, as in "mypassword". Can be any printable ASCII character except "/", "", or "@".
.....

Cancel Previous Next

Feedback English (US) © 2008 - 2019, Amazon Web Services, Inc. or its affiliates.

Praktik Soop

- Scroll kebawah lagi, hilangkan centang pada **Enable deletion protection**. Lalu **Create database**

The screenshot shows the AWS RDS 'Create database' wizard. The 'Maintenance' section is visible, with 'Enable auto minor version upgrade' selected. The 'Deletion protection' section shows a checkbox for 'Enable deletion protection' which is unchecked. At the bottom, there are 'Cancel', 'Previous', and 'Create database' buttons.

TERBUKA

aws Services Resource Groups Learn more

vocstartsoft/user334780=a

Maintenance

Auto minor version upgrade [Info](#)

Enable auto minor version upgrade
Enables automatic upgrades to new minor versions as they are released. The automatic upgrades occur during the maintenance window for the DB instance.

Disable auto minor version upgrade

Maintenance window [Info](#)

Select the period in which you want pending modifications or patches applied to the DB instance by Amazon RDS.

Select window

No preference

Deletion protection

Enable deletion protection
Protects the database from being deleted accidentally. While this option is enabled, you can't delete the database.

Cancel Previous **Create database**

Feedback English (US)

© 2008 - 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Praktik Sqoop

- Setelah next, scroll kebawah hingga menemui Database options dan isi seperti berikut

The screenshot shows the 'Database options' configuration page for an AWS RDS MySQL database. The 'Database name' is set to 'mydb'. The 'Port' is set to '3306'. The 'DB parameter group' is set to 'default.mysql5.7'. The 'Option group' is set to 'default:mysql-5-7'. Under 'IAM DB authentication', the 'Disable' option is selected. Below this, there is a section titled 'Encryption'.

aws Services Resource Groups TERRIKA vocstartsoft/user334780

Database options

Database name [Info](#)
mydb

Note: if no database name is specified then no initial MySQL database will be created on the DB Instance.

Port [Info](#)
TCP/IP port the DB instance will use for application connections.
3306

DB parameter group [Info](#)
default.mysql5.7

Option group [Info](#)
default:mysql-5-7

IAM DB authentication [Info](#)
 Enable IAM DB authentication
Manage your database user credentials through AWS IAM users and roles.
 Disable

Encryption

Feedback English (US) © 2008 - 2019, Amazon Web Services, Inc. or its affiliates.

Praktik Soop

- Klik pada **View DB Instance details**

The screenshot shows the AWS RDS 'Create database' interface. At the top, there's a blue banner with the text 'Switch to the new database creation flow.' Below it, the navigation bar includes 'Services' and 'Resource Groups'. The main content area displays a green box with a checkmark icon and the text 'Your DB instance is being created. Note: Your instance may take a few minutes to launch.' Below this, another box is titled 'Connecting to your DB instance' with the sub-instruction 'Once Amazon RDS finishes provisioning your DB instance, you can use a SQL client application or utility to connect to the instance.' A link 'Learn about connecting to your DB instance' is provided. At the bottom, there are two buttons: 'All DB instances' and a highlighted orange button labeled 'View DB instance details' with a hand cursor icon pointing at it. The footer contains links for 'Feedback', 'English (US)', and copyright information: '© 2008 - 2019, Amazon'.

Praktik Sqoop

- Tunggu status **Creating** berubah menjadi **Available**

Screenshot of the AWS RDS console showing the creation of a MySQL database named "mydatabase". The "Info" column shows the status as "Creating".

Summary			
DB identifier mydatabase	CPU -	Info Creating	Class db.t2.micro
Role Instance	Current activity	Engine MySQL Community	Region & AZ us-east-1f

Screenshot of the AWS RDS console showing the MySQL database "mydatabase" after it has been created. The "Info" column now shows the status as "Available".

Summary			
DB identifier mydatabase	CPU 2.83%	Info Available	Class db.t2.micro
Role Instance	Current activity 0 Connections	Engine MySQL Community	Region & AZ us-east-1f

Praktik Sqoop

- Untuk dapat di remote dari luar, edit security group dari database yang telah dibuat

The screenshot shows the AWS RDS MySQL instance configuration page for a database named 'mydatabase'. The top navigation bar includes 'Services', 'Resource Groups', and 'Support'. The instance details show it has 0 connections, uses MySQL Community engine, and is located in the us-east-1f region. The 'Connectivity & security' tab is selected, displaying the following configuration:

Endpoint & port	Networking	Security
Endpoint mydatabase.cxd9nnvuutpr.us-east-1.rds.amazonaws.com	Availability zone us-east-1f	VPC security groups rds-launch-wizard (sg-053428eb32b9ce256) (active)
Port 3306	VPC vpc-580c6122	Public accessibility Yes
	Subnet group default-vpc-580c6122	Certificate authority rds-ca-2015

Praktik Sqoop

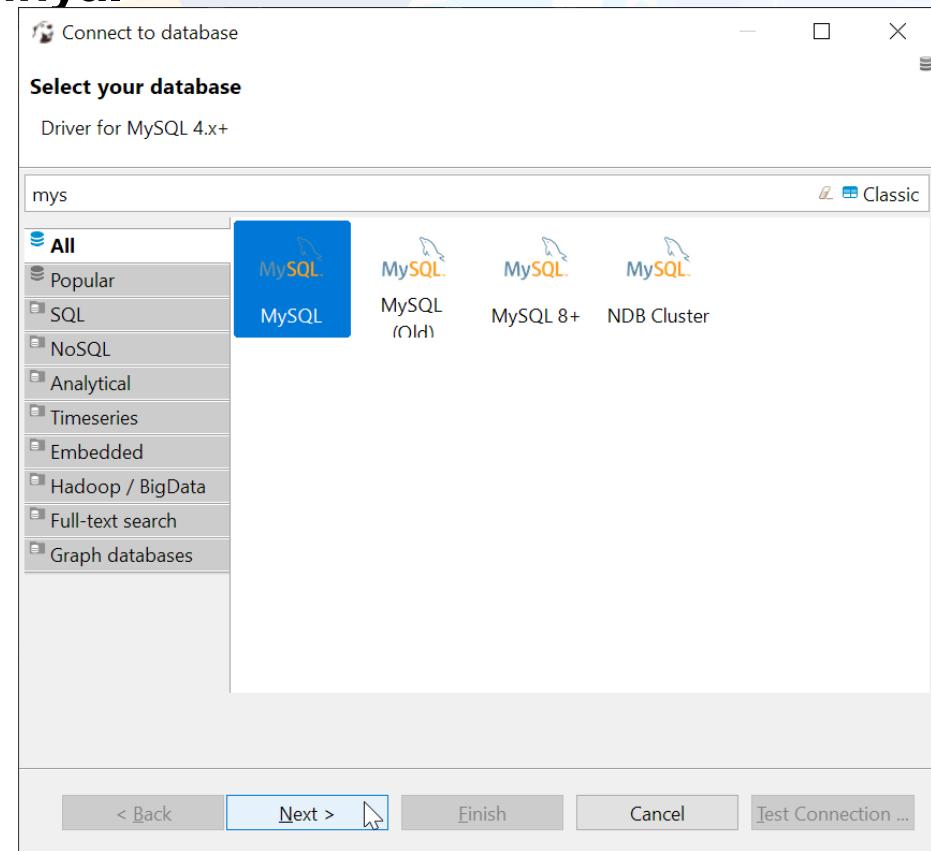
- Edit dan Add Rule seperti berikut dengan source Anywhere dan save

The screenshot shows the AWS EC2 Dashboard with the 'Resource Groups' tab selected. A search bar at the top right contains the identifier 'sg-053428eb32b9ce256'. The main content area displays the 'Edit inbound rules' dialog box. This dialog lists two existing rules and an 'Add Rule' button. The first rule is for MySQL/Aurora on port 3306 from a specific IP (36.85.75.110/32) with a custom description. The second rule is for MySQL/Aurora on port 3306 from 'Anywhere' (0.0.0.0/0, ::/0) with a similar description. A note at the bottom explains that edits on existing rules will result in the rule being deleted and a new one created, which may cause traffic disruption. At the bottom right of the dialog are 'Cancel' and 'Save' buttons.

Type	Protocol	Port Range	Source	Description
MySQL/Aurora	TCP	3306	Custom 36.85.75.110/32	e.g. SSH for Admin Desktop
MySQL/Aurora	TCP	3306	Anywhere 0.0.0.0/0, ::/0	e.g. SSH for Admin Desktop

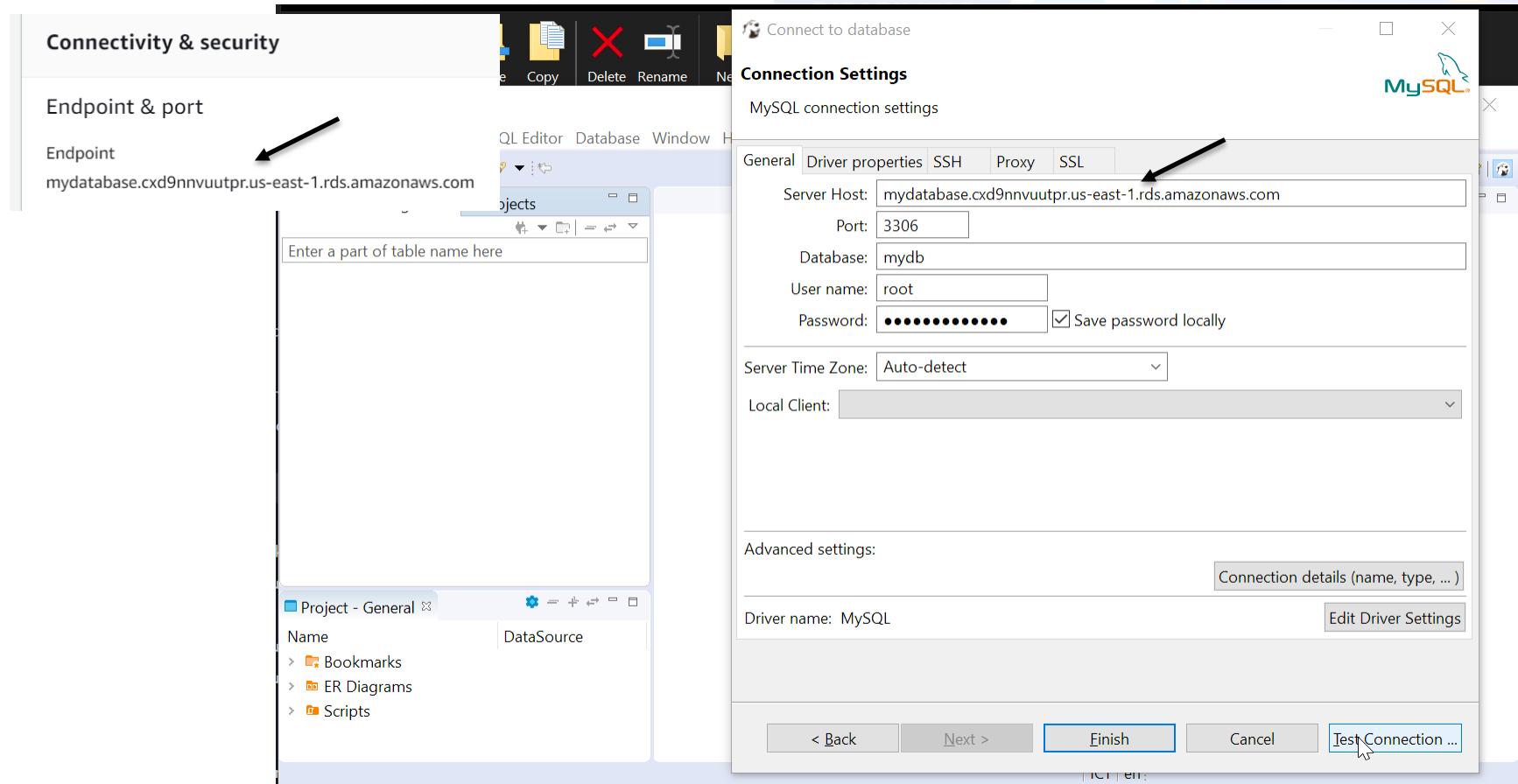
Praktik Sqoop

- Untuk dapat menggunakan MySQL yang ada pada AWS, perlu menggunakan aplikasi tambahan lagi yaitu DBeaver. Unduh pada link berikut -> <https://dbeaver.io/download/> . Kemudian install dan jalankan aplikasinya.
- Setelah aplikasi berhasil dijalankan, cari dan pilih MySQL kemudian klik Next



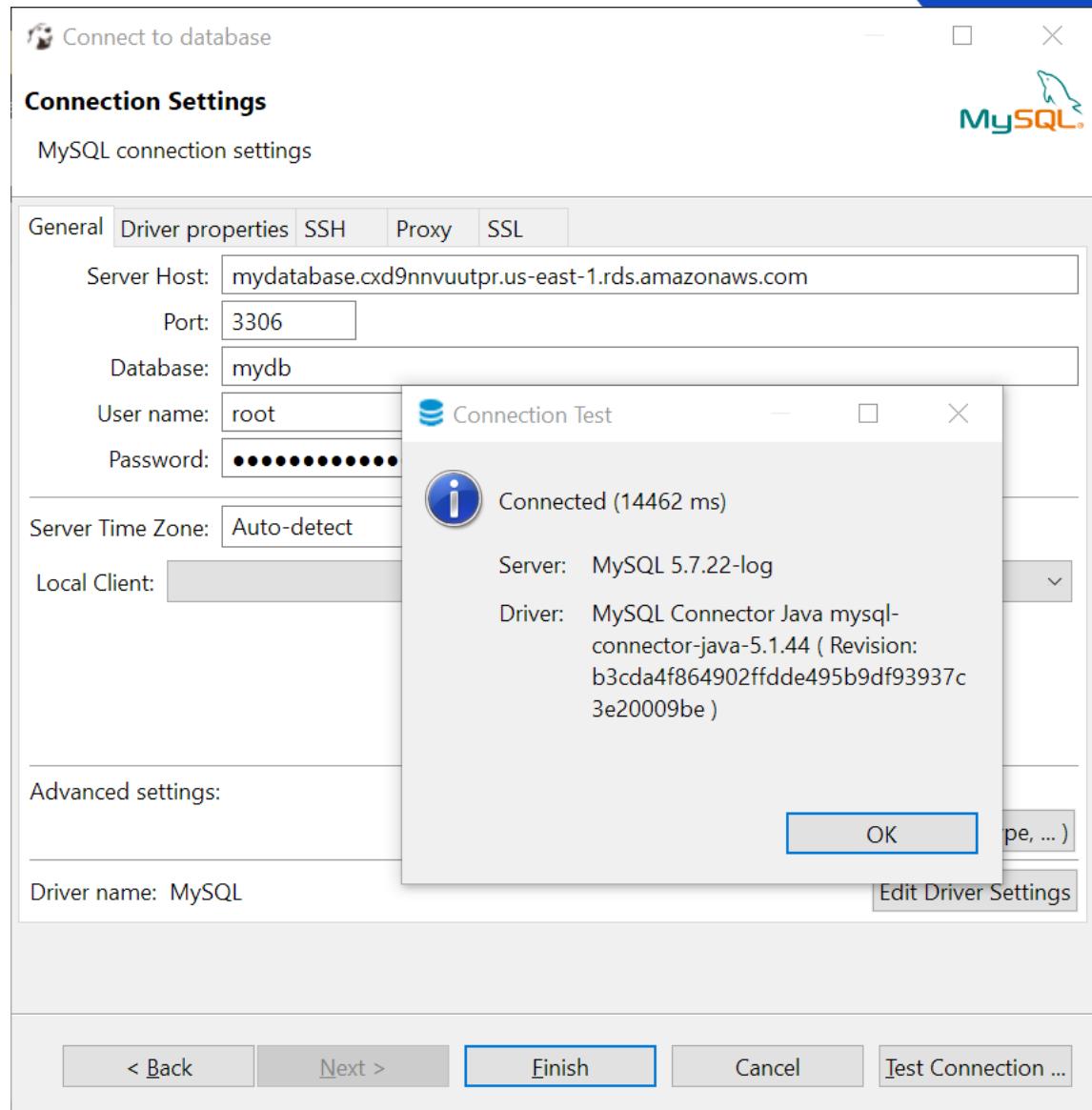
Praktik Sqoop

- Untuk konesnya, masukkan server host sesuai dengan end point RDS. Masukkan juga nama database, username dan password yang telah diatur sebelumnya.



Praktik Sqoop

- Kemudian klik Test Connection hingga berhasil seperti pada gambar disamping
- Jika sudah berhasil, klik OK dan Finish



Praktik Sqoop

- Buat tabel **db_mahasiswa** dengan kolom seperti dibawah

The screenshot shows the DBeaver 6.2.1 interface for MySQL database management. The left sidebar displays the Database Navigator with the 'mydb' database selected. The main workspace shows the 'Properties' tab for creating a new table named 'db_mahasiswa'. The table has two columns: 'nim' (int(11)) and 'nama' (varchar(1...)). The 'Columns' tab is also visible.

DBeaver 6.2.1 - db_mahasiswa

File Edit Navigate Search SQL Editor Database Window Help

MySQL - mydb mydb db_mahasiswa

Database Navigator Projects

Enter a part of table name here

MySQL - mydb

Databases innodb mydb Tables db_mahasiswa Columns Constraints Foreign Keys References Triggers Indexes Partitions db_nilai Views Indexes Procedures Triggers Events

Project - General

Name Data Source

Bookmarks ER Diagrams Scripts

Properties Data ER Diagram

Table Name: db_mahasiswa

Engine: InnoDB

Auto Increment: 1

Charset: latin1

Collation: latin1_swedish_ci

Description:

Columns

Column Name	#	Data Type	Not Null	Auto Increment	Key	Default	Extra	Expression	Comment
1 nim	1	int(11)	✓	✓	PRI				auto_incr...
2 nama	2	varchar(1...	✓	✗					

Constraints

Foreign Keys

References

Triggers

Indexes

Partitions

Statistics

DDL

Virtual

2 items

Sel: 0 | 0 Read-Only Smart Insert

filkom.ub.ac.id

Praktik Sqoop

Dan constraint seperti dibawah

The screenshot shows the DBeaver 6.2.1 interface with the following details:

- File Bar:** File, Edit, Navigate, Search, SQL Editor, Database, Window, Help.
- Toolbar:** Commit, Rollback, Auto.
- Database Navigator:** Shows MySQL - mydb selected.
- Table Properties:** Table Name: db_mahasiswa, Engine: InnoDB, Auto Increment: 1, Charset: latin1, Collation: latin1_swedish_ci.
- Constraints:** PRIMARY db_mahasiswa PRIMARY KEY.
- Columns:** A table with columns Name, Owner, Type.
- Project - General:** Shows Bookmarks, ER Diagrams, Scripts.
- Bottom Bar:** ICT en, FILKOM, DIGITAL TALENT, filkom.ub.ac.id, Save, Revert.

Praktik Sqoop

- Buat tabel satu lagi dengan nama **db_nilai** dengan kolom sbb :

DBeaver 6.2.1 - NewTable

File Edit Navigate Search SQL Editor Database Window Help

MySQL - mydb mydb *NewTable

Database Navigator Projects

Enter a part of table name here

MySQL - mydb

Databases innodb mydb

Tables db_mahasiswa NewTable

Columns nim (INT) tugas1 (INT) tugas2 (INT) tugas3 (INT) kuis1 (INT) kuis2 (INT) uts (INT) uas (INT)

Constraints Foreign Keys References Triggers Indexes Partitions Statistics DDL Virtual

Properties Data ER Diagram

Table Name: NewTable

Engine: InnoDB

Auto Increment: 0

Charset: latin1

Collation: latin1_swedish_ci

Description:

Column Name	#	Data Type	Not Null	Auto Increment	Key	Default	Extra	Expression	Comment
nim	1	INT	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>					
tugas1	2	INT	<input checked="" type="checkbox"/>	<input type="checkbox"/>					
tugas2	3	INT	<input checked="" type="checkbox"/>	<input type="checkbox"/>					
tugas3	4	INT	<input checked="" type="checkbox"/>	<input type="checkbox"/>					
kuis1	5	INT	<input checked="" type="checkbox"/>	<input type="checkbox"/>					
kuis2	6	INT	<input checked="" type="checkbox"/>	<input type="checkbox"/>					
uts	7	INT	<input checked="" type="checkbox"/>	<input type="checkbox"/>					
uas	8	INT	<input checked="" type="checkbox"/>	<input type="checkbox"/>					

8 items

ICT en Sel: 0 | 0 Read-Only Smart Insert ...

Save Revert

Quick Access

Praktik Sqoop

■ Tambahkan Foreign Keys

DBeaver 6.2.1 - db_nilai

File Edit Navigate Search SQL Editor Database Window Help

MySQL - mydb mydb db_nilai

Database Navigator Projects

Enter a part of table name here

MySQL - mydb

Databases innodb mydb

Tables db_mahasiswa db_nilai

Columns Constraints Foreign Keys References Triggers Indexes Partitions

Views

Indexes Procedures Triggers Events

Project - General

Name DataSources

Bookmarks ER Diagrams Scripts

Properties Data ER Diagram

Table Name: db_nilai

Engine: InnoDB

Auto Increment: 1

Charset: latin1

Collation: latin1_swedish_ci

Description:

Columns

Name	Owner	Ref Table	Type	Ref Object	On Delete	On Update
db_nilai_FK	db_nilai	db_mahasiswa	FOREIGN ...	PRIMARY	No Action	No Action

Constraints

Foreign Keys

References Triggers Indexes Partitions Statistics DDL Virtual

1 items

Sel: 0 | 0 Read-Only Smart Insert

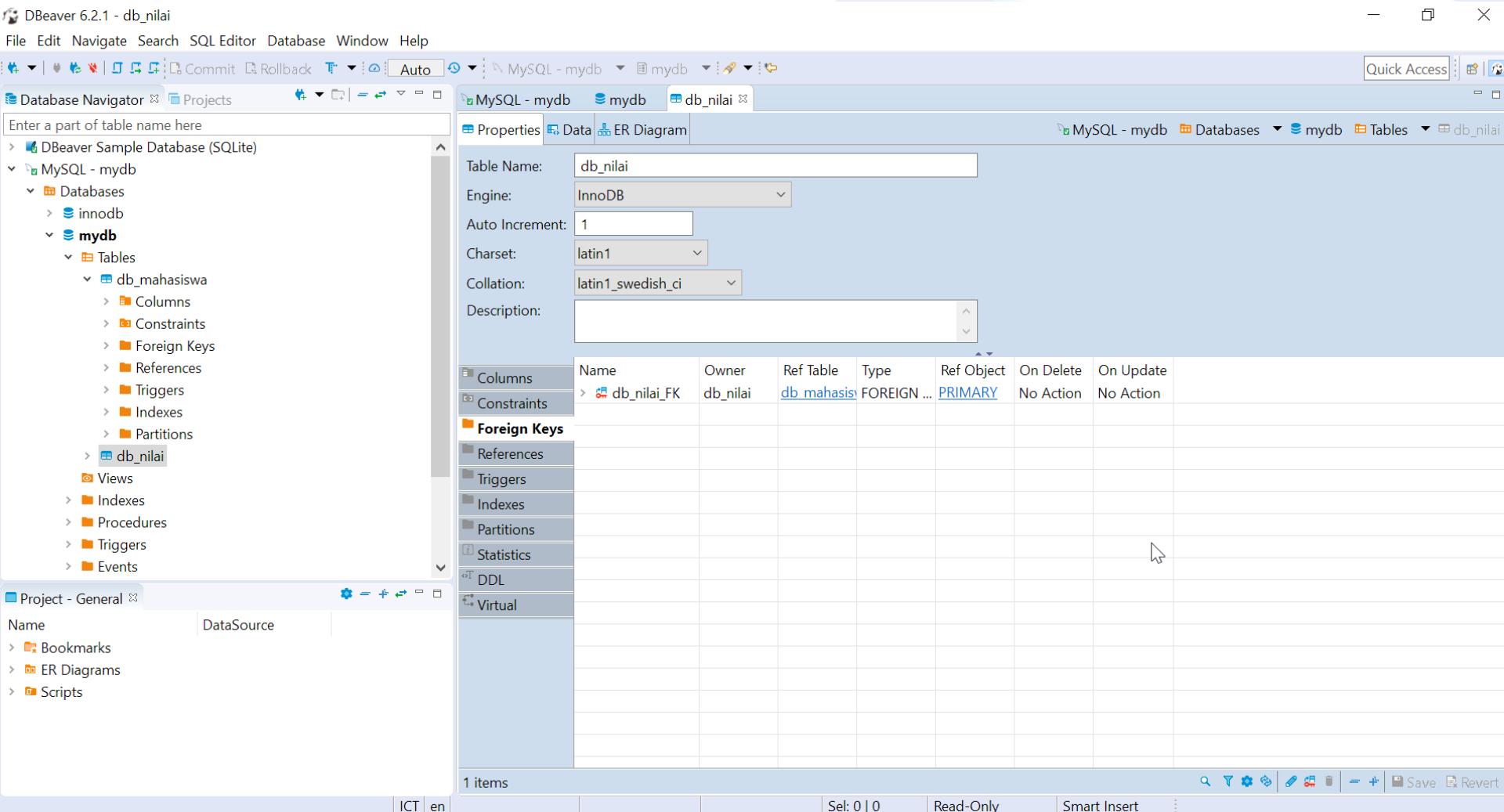
ICT en

FILKOM

DIGITAL TALENT KOMUNITAS

filkom.ub.ac.id

34



Praktik Sqoop

- Lakukan import data dari csv kedalam tabel db_mahasiswa

> DBeaver Sample Database (SQLite)

> MySQL - mydb

> Databases

> innodb

> **mydb**

> Tables

> **db_mahasiswa**

Create

Edit Table

Filter

View data

Read data in SQL console

Compare/Migrate

View Diagram

Export Data

Import Data

Generate SQL

Tools

Copy

Copy Advanced Info

Delete

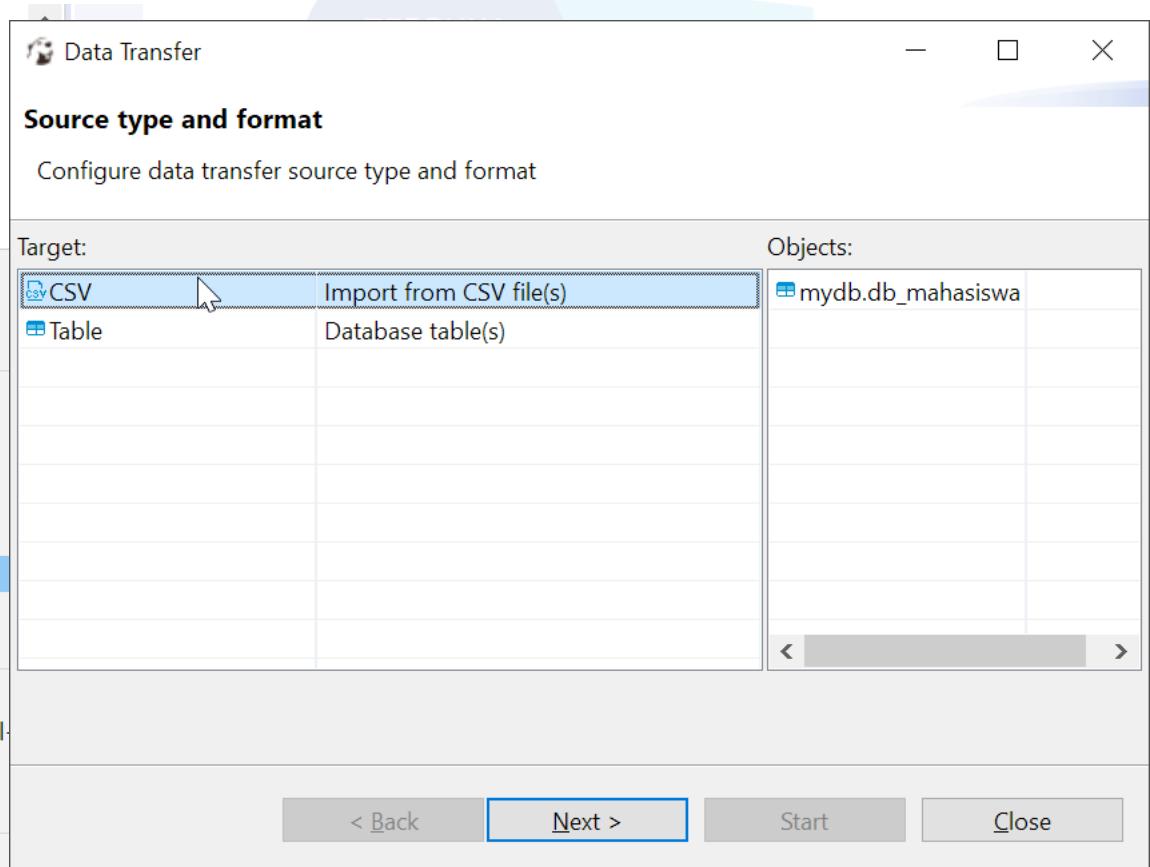
Rename

Refresh

Project - General

Name

Bookmarks



Praktik Sqoop

- Pilih file *.csv yang akan di import, next

The image displays two windows of a 'Data Transfer' application side-by-side, illustrating the process of selecting an input CSV file for import.

Left Window (Initial State):

- Input file(s)**: Configure input files or directories
- Input files**:
 - Source**: <none>
 - Target**: mydb.db_mahasiswa
- Importer settings**:

Name	Value
Extension	csv,tsv,txt
Encoding	utf-8
Column delimiter	,
Header position	top
- Buttons**: < Back, Next >

Right Window (After Selection):

- Input file(s)**: Configure input files or directories
- Input files**:
 - Source**: db_mahasiswa.csv
 - Target**: mydb.db_mahasiswa
- Importer settings**:

Name	Value
Extension	csv,tsv,txt
Encoding	utf-8
Column delimiter	,
Header position	top
- Buttons**: < Back, Next >, Start, Close

Praktik Sqoop

■ Klik next, next

The screenshot shows two windows of the Data Transfer application side-by-side.

Left Window (Preview data import):

- Entity:** mydb.db_mahasiswa
- Columns:**

Target	Source
123 nim	15001
abc nama	Mhs 1
- Preview data:**

nim	nama
15,002	Mhs 2
15,003	Mhs 3
15,004	Mhs 4
- Buttons:** < Back, Next (highlighted with a blue box)

Right Window (Settings (Files to Database, CSV)):

- Set export settings**
- Data load settings**
 - Transfer auto-generated columns
 - Truncate target table(s) before load
- Performance**
 - Open new connection(s)
 - Use transactions
- Commit after insert of :** 10000
- General**
 - Open table editor on finish
- Buttons:** < Back, Next (highlighted with a blue box), Start, Close

Praktik Sqoop

- Klik Start dan tunggu hingga proses import selesai

Data Transfer

Confirm

Check results

Objects

Source Container	Source	Target Container	Target
F:\Magister\FGA BIG...	db_mahasiswa.csv	MySQL - mydb	mydb.db_mahasiswa

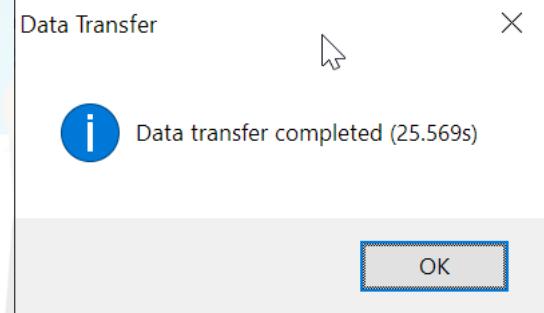
Source settings

Files settings:
CSV settings:
Extension: csv,tsv,txt
Encoding: utf-8
Column delimiter: .

Target settings

Database settings:
Open new connection(s): Yes
Use transactions: Yes
Commit after: 10000
Transfer auto-generated columns: Yes

< Back Next > Start Close



Praktik Sqoop

- Tampilan data saat import db_mahasiswa berhasil dilakukan

DBeaver 6.2.1 - db_mahasiswa

File Edit Navigate Search SQL Editor Database Window Help

MySQL - mydb db_mahasiswa

Properties Data ER Diagram

Enter a part of table name here

MySQL - mydb Databases mydb Tables db_mahasiswa

Quick Access

Database Navigator Projects

Enter a SQL expression to filter results (use Ctrl+Space)

Grid

	nim	nama
1	15,002	Mhs 2
2	15,003	Mhs 3
3	15,004	Mhs 4
4	15,005	Mhs 5
5	15,006	Mhs 6
6	15,007	Mhs 7
7	15,008	Mhs 8
8	15,009	Mhs 9
9	15,010	Mhs 10
10	15,011	Mhs 11
11	15,012	Mhs 12
12	15,013	Mhs 13
13	15,014	Mhs 14
14	15,015	Mhs 15
15	15,016	Mhs 16
16	15,017	Mhs 17
17	15,018	Mhs 18
18	15,019	Mhs 19
19	15,020	Mhs 20
20	15,021	Mhs 21
21	15,022	Mhs 22
22	15,023	Mhs 23
23	15,024	Mhs 24
24	15,025	Mhs 25
25	15,026	Mhs 26

Value

15002

Record

Save Cancel Script | 200 | 39 | 39 row(s) fetched - 275ms (+)

ICT | en | Sel: 0 | 0 | Read-Only | Smart Insert |

Praktik Sqoop

- Lakukan import pada db_nilai hingga berhasil

DBeaver 6.2.1 - db_nilai

File Edit Navigate Search SQL Editor Database Window Help

MySQL - mydb mydb db_nilai

Database Navigator Projects

Enter a part of table name here

MySQL - mydb

Databases innodb mydb

Tables db_mahasiswa db_nilai

Columns Constraints Foreign Keys References Triggers Indexes Partitions Views Indexes Procedures Triggers Events

Project - General

Name DataSources

Bookmarks ER Diagrams Scripts

Properties Data ER Diagram

MySQL - mydb Databases mydb Tables NewTable

Grid

	nim	tugas1	tugas2	tugas3	kuis1	kuis2	uts	uas
1	15,002	0	95	45	90	60	55	63
2	15,003	80	40	0	95	60	73	45
3	15,004	80	80	0	0	80	70	75
4	15,005	80	95	100	95	60	80	55
5	15,006	80	75	100	100	0	95	85
6	15,007	80	85	95	95	80	75	55
7	15,008	80	80	100	95	60	70	70
8	15,009	80	30	100	0	0	25	20
9	15,010	80	85	95	80	80	90	90
10	15,011	80	65	90	85	80	75	70
11	15,012	80	70	85	85	80	55	50
12	15,013	80	65	90	85	80	70	60
13	15,014	80	100	100	80	80	90	95
14	15,015	80	40	100	100	0	45	65
15	15,016	80	80	100	95	80	100	80
16	15,017	0	55	100	95	80	60	40
17	15,018	80	60	100	95	80	90	50
18	15,019	80	40	100	95	75	25	45
19	15,020	80	95	100	80	80	90	55
20	15,021	80	65	100	80	80	60	60
21	15,022	80	60	85	0	80	45	45
22	15,023	80	60	70	0	80	60	65
23	15,024	80	60	95	0	80	70	65
24	15,025	80	90	95	90	80	60	65
25	15,026	80	65	100	85	80	70	68

Save Cancel Script | 200 39 39 row(s) fetched - 276ms (+)

ICT en Sel: 0 | 0 Read-Only Smart Insert

Value 15002

Dictionary (db_mahasiswa): (Define Description)

Value	Description
15002	Mhs 2
15003	Mhs 3
15004	Mhs 4
15005	Mhs 5
15006	Mhs 6
15007	Mhs 7
15008	Mhs 8
15009	Mhs 9
15010	Mhs 10
15011	Mhs 11
15012	Mhs 12
15013	Mhs 13
15014	Mhs 14
15015	Mhs 15
15016	Mhs 16
15017	Mhs 17
15018	Mhs 18
15019	Mhs 19
15020	Mhs 20
15021	Mhs 21

Type part of dictionary value to search

Praktik Sqoop

- Kembali ke Putty dan lakukan cek database pada MySQL db server

```
sqoop list-databases --connect  
jdbc:mysql://endpointdbserveraws/ --username root --  
password isidenganpasswordanda
```



```
hadoop@ip-172-31-19-46:~/mysql-connector-java-6.0.6  
performance_schema  
sys  
[hadoop@ip-172-31-19-46 mysql-connector-java-6.0.6]$ sqoop list-databases --conn  
ect jdbc:mysql://mydatabase.cxd9nnvuutpr.us-east-1.rds.amazonaws.com/ --username  
root --password ██████████  
Warning: /usr/lib/sqoop/../.hcatalog does not exist! HCatalog jobs will fail.  
Please set $HCAT_HOME to the root of your HCatalog installation.  
Warning: /usr/lib/sqoop/../.accumulo does not exist! Accumulo imports will fail.  
Please set $ACCUMULO_HOME to the root of your Accumulo installation.  
19/10/04 20:44:52 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7  
19/10/04 20:44:52 WARN tool.BaseSqoopTool: Setting your password on the command-  
line is insecure. Consider using -P instead.  
19/10/04 20:44:52 INFO manager.MySQLManager: Preparing to use a MySQL streaming  
resultset.  
Loading class `com.mysql.jdbc.Driver'. This is deprecated. The new driver class  
is `com.mysql.cj.jdbc.Driver'. The driver is automatically registered via the SP  
I and manual loading of the driver class is generally unnecessary.  
information_schema  
innodb  
mydb  
mysql  
performance_schema  
sys  
[hadoop@ip-172-31-19-46 mysql-connector-java-6.0.6]$
```

Praktik Sqoop

■ cek tabel pada mydb

```
sqoop list-tables --connect  
jdbc:mysql://endpointdbserveraws/ --username root --  
password isidenganpasswordanda
```



```
[hadoop@ip-172-31-19-46:~/mysql-connector-java-6.0.6]  
[hadoop@ip-172-31-19-46 mysql-connector-java-6.0.6]$ sqoop list-tables --connect  
jdbc:mysql://mydatabase.cxd9nnvuutpr.us-east-1.rds.amazonaws.com/mydb --username root --password [REDACTED]  
Warning: /usr/lib/sqoop/../hcatalog does not exist! HCatalog jobs will fail.  
Please set $HCAT_HOME to the root of your HCatalog installation.  
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.  
Please set $ACCUMULO_HOME to the root of your Accumulo installation.  
19/10/04 20:49:25 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7  
19/10/04 20:49:25 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.  
19/10/04 20:49:25 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.  
Loading class `com.mysql.jdbc.Driver'. This is deprecated. The new driver class is `com.mysql.cj.jdbc.Driver'. The driver is automatically registered via the SPI and manual loading of the driver class is generally unnecessary.  
db_mahasiswa  
db_nilai  
[hadoop@ip-172-31-19-46 mysql-connector-java-6.0.6]$
```

Praktik Sqoop

- Import semua tabel pada **mydb** kedalam HDFS

sqoop import-all-tables --connect jdbc:mysql://
endpointdbserveraws/mydb --username root --password
isidenganpasswordanda

- Jika mendapat eror seperti berikut, lihat slide berikutnya

```
hadoop@ip-172-31-19-46:~/mysql-connector-java-6.0.6
Physical memory (bytes) snapshot=1085091840
Virtual memory (bytes) snapshot=12815433728
Total committed heap usage (bytes)=996147200
File Input Format Counters
    Bytes Read=0
File Output Format Counters
    Bytes Written=499
19/10/04 20:54:46 INFO mapreduce.ImportJobBase: Transferred 499 bytes in 25.4
seconds (19.5996 bytes/sec)
19/10/04 20:54:46 INFO mapreduce.ImportJobBase: Retrieved 39 records.
19/10/04 20:54:46 INFO tool.CodeGenTool: Beginning code generation
19/10/04 20:54:46 INFO manager.SqlManager: Executing SQL statement: SELECT t.
ROM `db_nilai` AS t LIMIT 1
19/10/04 20:54:46 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib
doop-mapreduce
Note: /tmp/sqoop-hadoop/compile/f07494ca0d53b616ff642c3ee6b761c6/db_nilai.jav
ses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
19/10/04 20:54:47 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-h
op/compile/f07494ca0d53b616ff642c3ee6b761c6/db_nilai.jar
19/10/04 20:54:47 ERROR tool.ImportAllTablesTool: Error during import: No pri
y key could be found for table db_nilai. Please specify one with --split-by o
r perform a sequential import with '-m 1'.
[hadoop@ip-172-31-19-46 mysql-connector-java-6.0.6]$ $HADOOP_HOME/bin/hadoop fs
```

Praktik Sqoop

- Jika terjadi eror lakukan perintah berikut

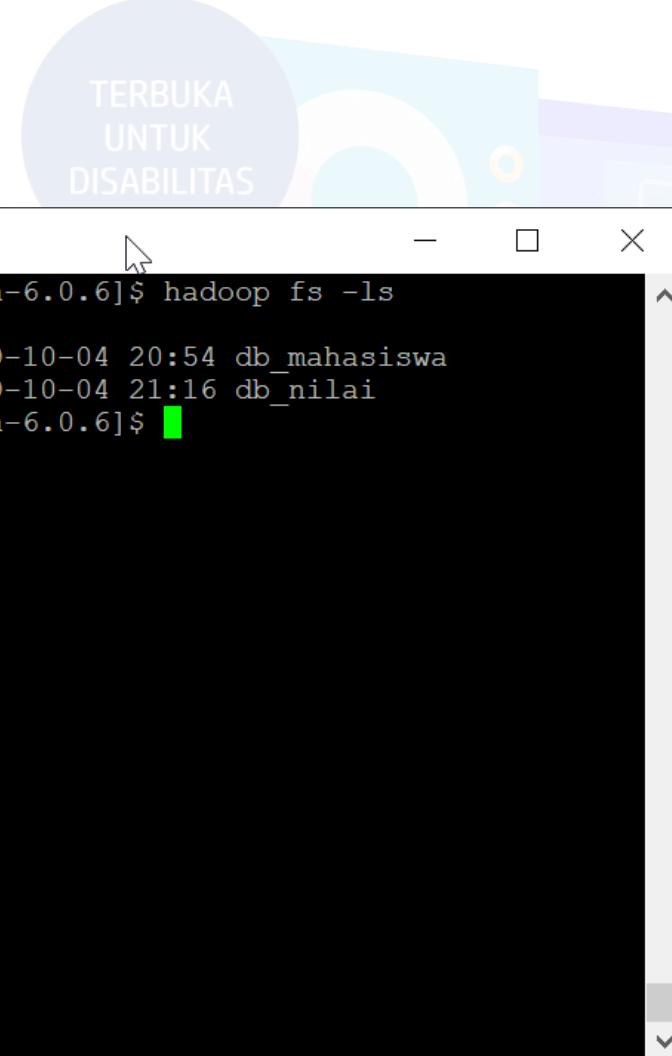
```
sqoop import --connect jdbc:mysql://endpointdbserveraws  
/mydb --username root --password isidenganpasswordanda  
--table db_nilai --m 1
```



Praktik Sqoop

■ Cek isi dari HDFS

Hdfs fs -ls



```
[hadoop@ip-172-31-19-46:mysql-connector-java-6.0.6]$ hadoop fs -ls
[hadoop@ip-172-31-19-46:mysql-connector-java-6.0.6]$ Found 2 items
drwxr-xr-x  - hadoop hadoop          0 2019-10-04 20:54 db_mahasiswa
drwxr-xr-x  - hadoop hadoop          0 2019-10-04 21:16 db_nilai
[hadoop@ip-172-31-19-46:mysql-connector-java-6.0.6]$
```

TERBUKA
UNTUK
DISABILITAS

BREAK
YOUR
LIMITS!

Lab Oozie



Praktik Oozie

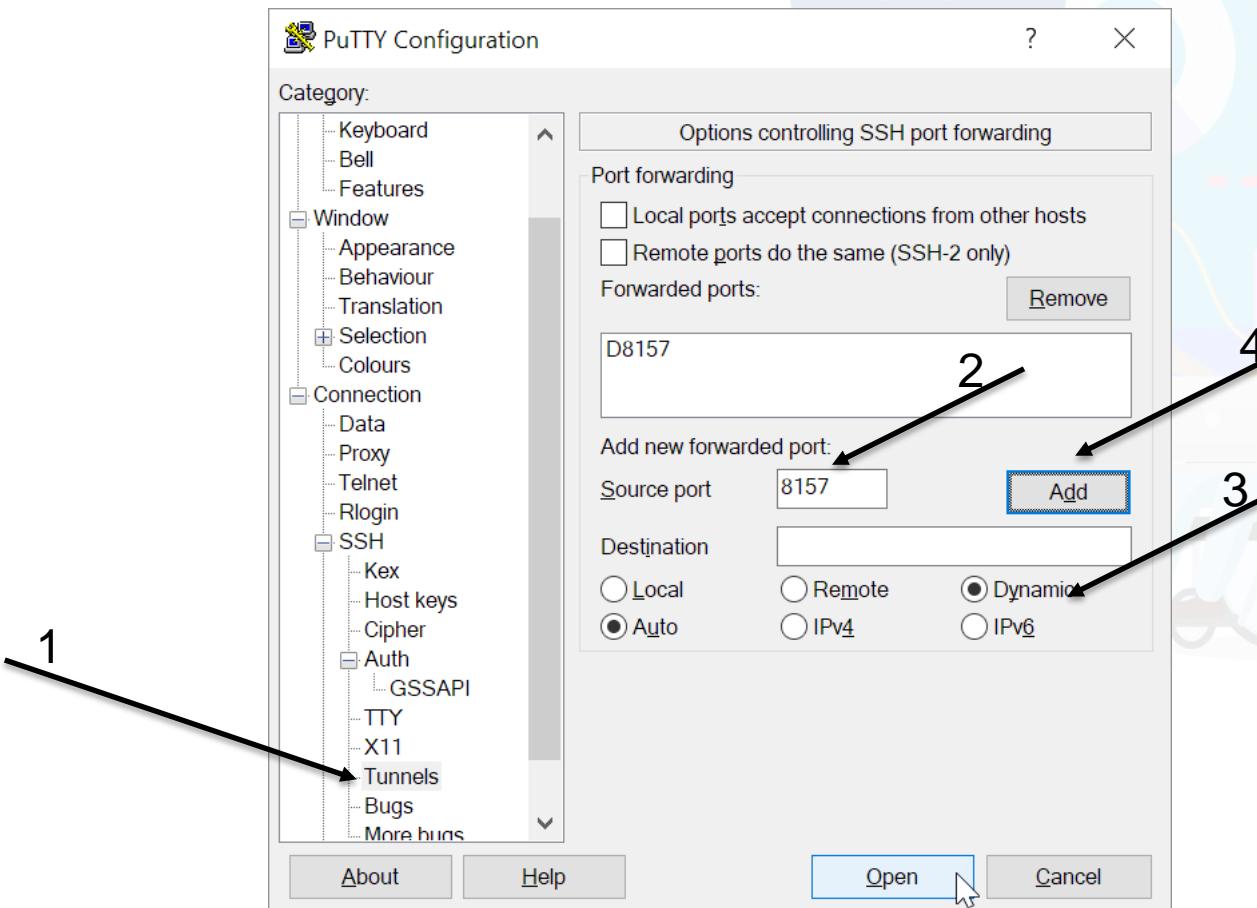
- Buat EMR Cluster terlebih dahulu yang didalamnya telah ada Oozie
- Kemudian aktifkan **Enable Web Connection** dan ikuti langkah-langkahnya

The screenshot shows the AWS EMR service console for a cluster named 'j-2UP80WFK03UAO'. The 'Connections' section is highlighted with a red arrow pointing to the 'Enable Web Connection' link. Other visible details include the Master public DNS (ec2-3-95-180-214.compute-1.amazonaws.com), SSH access, and various configuration parameters like Release label, Hadoop distribution, Applications, Log URI, and security settings.

Summary		Configuration details	
ID: j-2UP80WFK03UAO		Release label: emr-5.27.0	
Creation date: 2019-10-05 01:05 (UTC+7)		Hadoop distribution: Amazon 2.8.5	
Elapsed time: 7 minutes		Applications: Pig 0.17.0, Sqoop 1.4.7, Oozie 5.1.0	
Auto-terminate: No		Log URI: s3://aws-logs-056823078307-us-east-1/elasticmapreduce/	
Termination protection: On Change		EMRFS consistent view: Disabled	
		Custom AMI ID: --	
Network and hardware		Security and access	
Availability zone: us-east-1a		Key name: golobok	
Subnet ID: subnet-0434d549		EC2 instance profile: EMR_EC2_DefaultRole	
Master: Bootstrapping 1 m3.xlarge		EMR role: EMR_DefaultRole	

Praktik Oozie

- Buka Putty dan masukkan konfigurasi seperti biasa, Tambahkan pengaturan dibawah ini



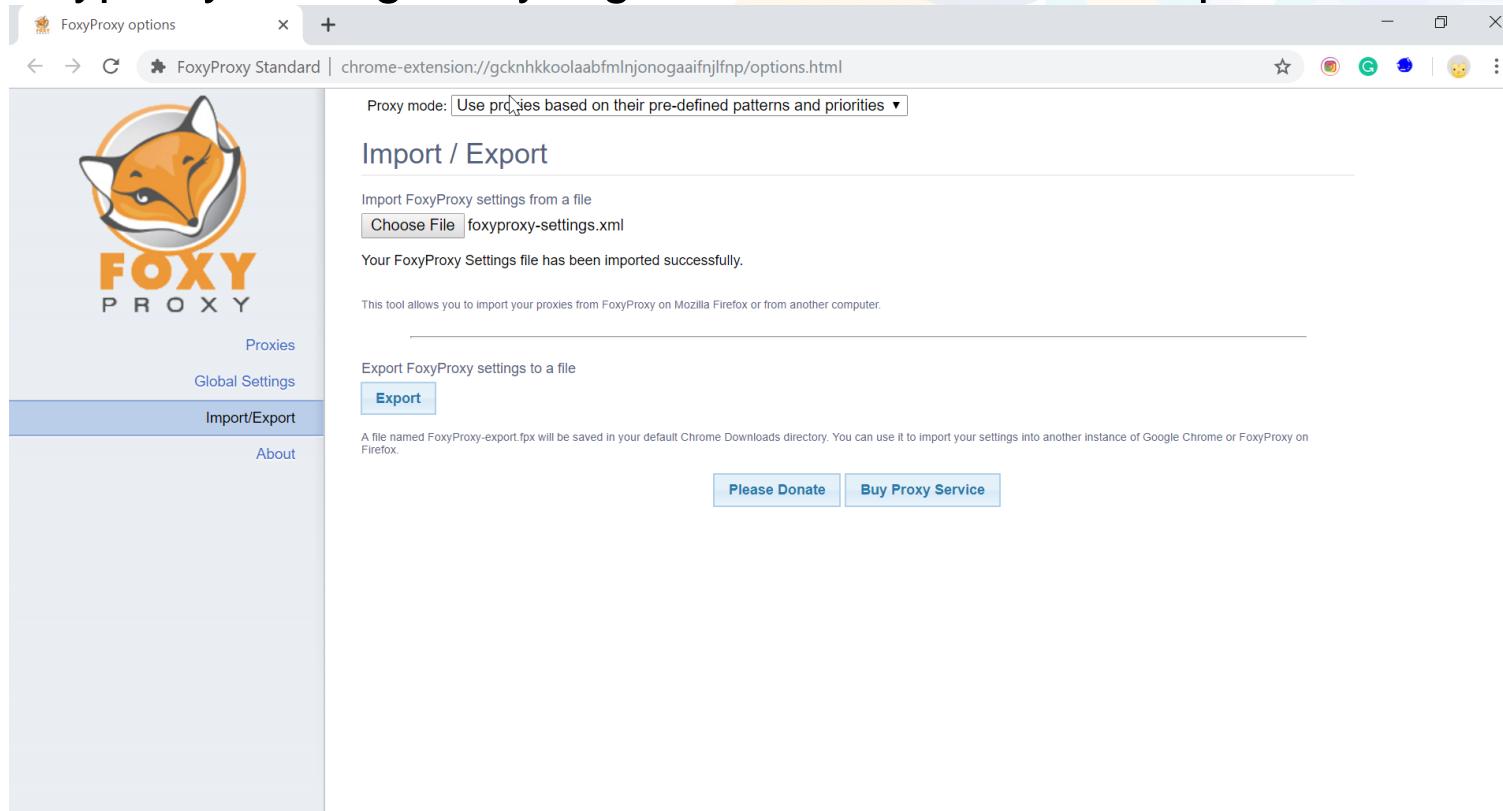
Praktik Oozie

- Buat file baru dengan nama “foxyproxy-settings.xml”,
Buka dan paster script berikut dan simpan

```
<?xml version="1.0" encoding="UTF-8"?>
<foxyproxy>
    <proxies>
        <proxy name="emr-socks-proxy" id="2322596116" notes="" fromSubscription="false" enabled="true"
mode="manual" selectedTabIndex="2" lastresort="false" animatedIcons="true" includeInCycle="true" color="#0055E5"
proxyDNS="true" noInternalIPs="false" autoconfMode="pac" clearCacheBeforeUse="false" disableCache="false"
clearCookiesBeforeUse="false" rejectCookies="false">
            <matches>
                <match enabled="true" name="*ec2*.amazonaws.com*" pattern="*ec2*.amazonaws.com*" isRegEx="false"
isBlackList="false" isMultiLine="false" caseSensitive="false" fromSubscription="false" />
                <match enabled="true" name="*ec2*.compute*" pattern="*ec2*.compute*" isRegEx="false"
isBlackList="false" isMultiLine="false" caseSensitive="false" fromSubscription="false" />
                <match enabled="true" name="10.*" pattern="http://10.*" isRegEx="false" isBlackList="false"
isMultiLine="false" caseSensitive="false" fromSubscription="false" />
                <match enabled="true" name="*10*.amazonaws.com*" pattern="*10*.amazonaws.com*" isRegEx="false"
isBlackList="false" isMultiLine="false" caseSensitive="false" fromSubscription="false" />
                <match enabled="true" name="*10*.compute*" pattern="*10*.compute*" isRegEx="false"
isBlackList="false" isMultiLine="false" caseSensitive="false" fromSubscription="false" />
                <match enabled="true" name="*.compute.internal*" pattern="*.compute.internal*" isRegEx="false"
isBlackList="false" isMultiLine="false" caseSensitive="false" fromSubscription="false" />
                <match enabled="true" name="*.ec2.internal*" pattern="*.ec2.internal*" isRegEx="false"
isBlackList="false" isMultiLine="false" caseSensitive="false" fromSubscription="false" />
            </matches>
            <manualconf host="localhost" port="8157" socksversion="5" isSocks="true" username="" password=""
domain="" />
        </proxy>
    </proxies>
</foxyproxy>
```

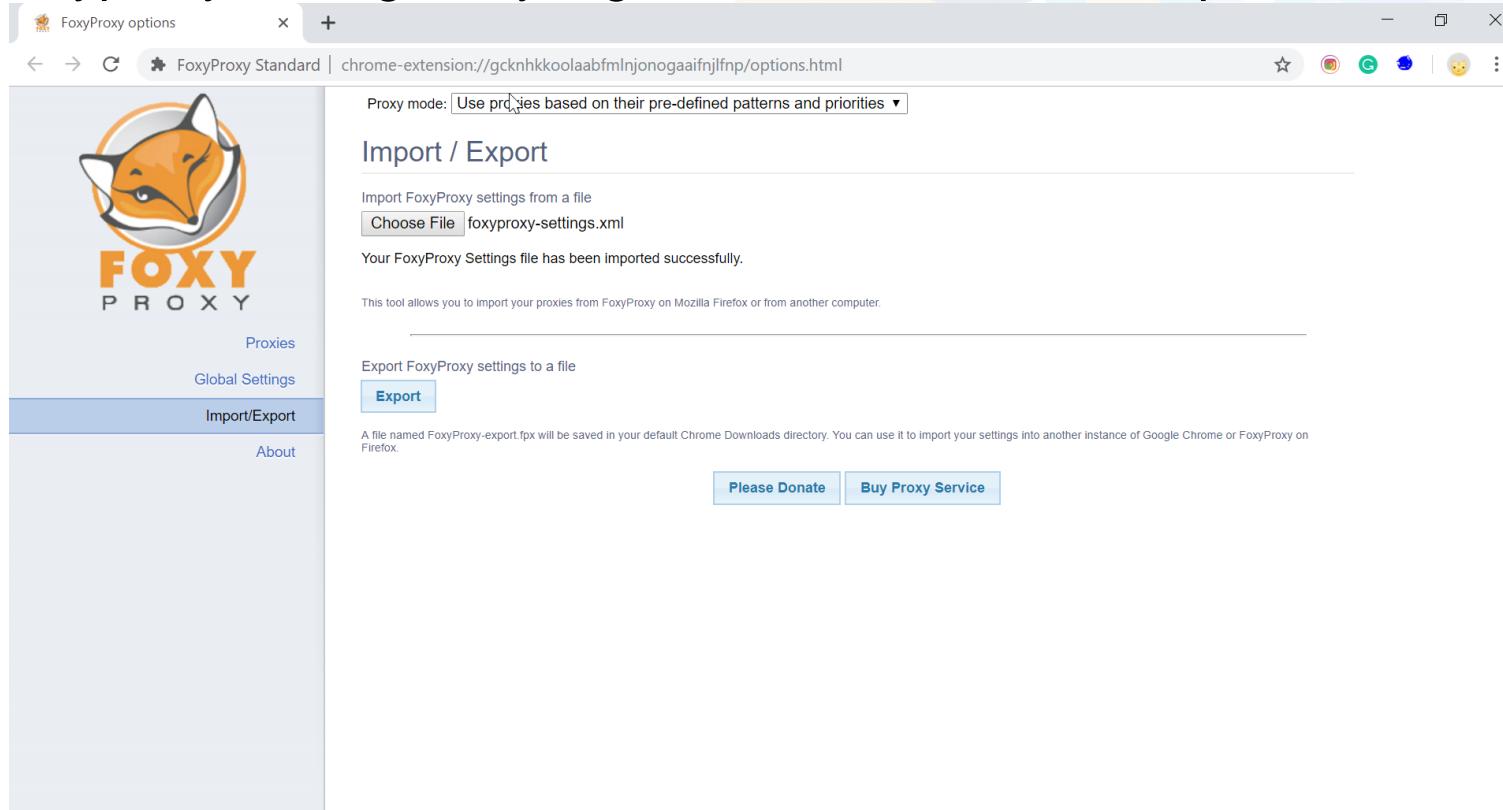
Praktik Oozie

- Buka Chrome, tambahkan ekstensi berikut :
<https://chrome.google.com/webstore/search/foxy%20proxy>
- Klik kanan ikon foxyproxy – options – Import/Export – Choose File – foxyproxy-setting.xml yang telah dibuat, kemudian pilih add



Praktik Oozie

- Buka Chrome, tambahkan ekstensi berikut :
<https://chrome.google.com/webstore/search/foxy%20proxy>
- Klik kanan ikon foxyproxy – options – Import/Export – Choose File – foxyproxy-setting.xml yang telah dibuat, kemudian pilih add



Praktik Oozie

- Masukkan alamat EMR Cluster yang sudah dibuat menjadi seperti berikut : alamatEMRClusterAWS.com:11000/oozie/

Oozie Web Console

Workflow Jobs Coordinator Jobs Bundle Jobs System Info Metrics Settings

All Jobs Active Jobs Done Jobs Custom Filter Server version [5.1.0]

Job Id	Name	Status	R...	User	Group	Created	Started	Last Modified	Ended
--------	------	--------	------	------	-------	---------	---------	---------------	-------

- Ini untuk memastikan bahwa Oozie telah siap untuk digunakan

Praktik Oozie

- Kembali ke Putty dan ketikkan perintah berikut dan tunggu hingga instalasi selesai berjalan:

install-oozie-example

Praktik Oozie

- Masuk ke direktori oozie dengan perintah berikut :

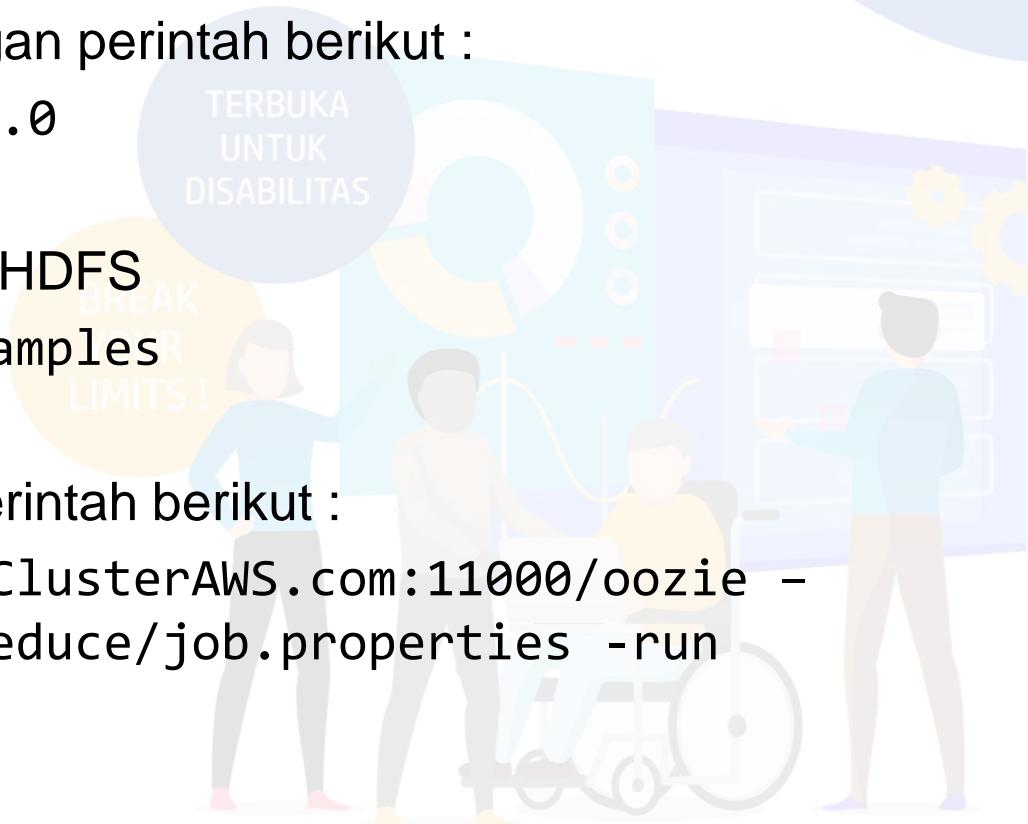
```
cd /usr/share/doc/oozie-5.1.0
```

- Setelah itu taruh file kedalam HDFS

```
hadoop fs -put examples examples
```

- Jalankan job oozie dengan perintah berikut :

```
Oozie job -oozie alamatEMRClusterAWS.com:11000/oozie -  
config examples/apps/map-reduce/job.properties -run
```



Praktik Oozie

- Berikut adalah hasil dari job oozie yang dijalankan
- Pada teks yang diblok adalah id job yang dijalankan oleh oozie

```
[hadoop@ip-172-31-19-46 oozie-5.1.0]$ hadoop fs -put examples examples
[hadoop@ip-172-31-19-46 oozie-5.1.0]$ oozie job -oozie http://ec2-3-95-180-214.compute-1.amazonaws.com:11000/oozie -config examples/apps/map-reduce/job.properties -run
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/oozie/lib/slf4j-log4j12-1.6.6.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/oozie/lib/slf4j-simple-1.6.6.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
job: 0000000-191004232348680-oozie-oozi-W
[hadoop@ip-172-31-19-46 oozie-5.1.0]$ clear
[hadoop@ip-172-31-19-46 oozie-5.1.0]$
```

Praktik Oozie

- Buka browser lalu masuk ke oozie anda dengan alamat -> alamatEMRClusterAWS.com:11000/oozie/
- Akan terlihat id job yang sama dengan yang ada pada putty

The screenshot shows the Oozie Web Console interface. At the top, there's a navigation bar with links for Documentation, Workflow Jobs, Coordinator Jobs, Bundle Jobs, System Info, Metrics, and Settings. Below that is a toolbar with icons for All Jobs, Active Jobs (which is selected), Done Jobs, and Custom Filter. The main area is titled "Workflow Jobs" and displays a table of jobs. The table has columns for Job Id, Name, Status, R..., User, Group, Created, Started, Last Modified, and Ended. There is one entry in the table:

	Job Id	Name	Status	R...	User	Group	Created	Started	Last Modified	Ended
1	0000000-191004232348680-oozie-oozi-W	map-reduce-wf	RUNNING	0	hadoop		Sat, 05 Oct 2019 07:19:06 GMT	Sat, 05 Oct 2019 07:19:06 GMT	Sat, 05 Oct 2019 07:19:07 GMT	

On the right side of the table, it says "Server version [5.1.0]".

Tugas Individu

1. Buatlah rangkuman materi dengan cara berikut:

- Lalukan ulang seperti yang ada di All slide
- Cek plagiasi diturnitin (tiap minggu) dari hasil rangkuman tersebut
 - > Register ke turnitin
 - > Masukkan **id class**: 21563495 & **enrool key**: filkomub9302



DIGITAL
TALENT
SCHOLARSHIP



DIGITAL TALENT SCHOLARSHIP 2019

Big Data Analytics



Terimakasih

Oleh: Imam Cholissodin | imamcs@ub.ac.id, Putra Pandu Adikara, Sufia Adha Putri

Asisten: Guedho, Sukma, Anshori, Aang dan Gusti

Fakultas Ilmu Komputer (Filkom) Universitas Brawijaya (UB)