


## Map Reduce: Hurma Mahmood-14855

### 1. Task 1:

Previous steps:

 Job Designer


Job Design (streaming type)

Name

paymentJob1

Description

Payment.Type.Frequency

 advanced

You can parameterize the values, using `$myVar` or `${myVar}`. When the design is submitted, you will be prompted for the actual value of `myVar`.

Mapper	<code>/usr/bin/python mapper_freq.py</code>
Reducer	<code>/usr/bin/python reducer_freq.py</code>

Hadoop job properties	Property name	Value
-----------------------	---------------	-------

Hadoop job properties

Property name	Value	
<input type="text" value="mapred.input.dir"/>	<input type="text" value="/user/cloudera/sim.data/PS_20174392719_14912"/>	<div><div>..</div><div>Delete</div></div>
<input type="text" value="mapred.output.dir"/>	<input type="text" value="/user/cloudera/sim.data/output"/>	<div><div>..</div><div>Delete</div></div>
<div>Add property</div>		

Files

Add file

Archives

..

Delete

..

Delete

Add archive

Save

Cancel

**HUE** Query Editors Data Browsers Workflows Search Security

**Oozie Dashboard** Workflows Coordinators Bundles SLA Oozie

**WORKFLOW** Workflow paymentJob1

paymentJob1 Graph Actions Details Configuration Log Definition

**SUBMITTER** cloudera

**STATUS** **SUCCEEDED**

**PROGRESS** 100%

**ID** 0000000-201206194654538-oozie-oozi-W [Back](#)

**VARIABLES** oozie.wf.applicatio...

**MANAGE**

Output:

Editors Data Browsers Workflows Search Security

Home Page 1 of 1

/ user / cloudera / sim.data / output / part-00000

```
CASH_IN 1399283
CASH_OUT 2237500
DEBIT 41432
PAYMENT 2151493
TRANSFER 532909
```

## 2. TASK 2:

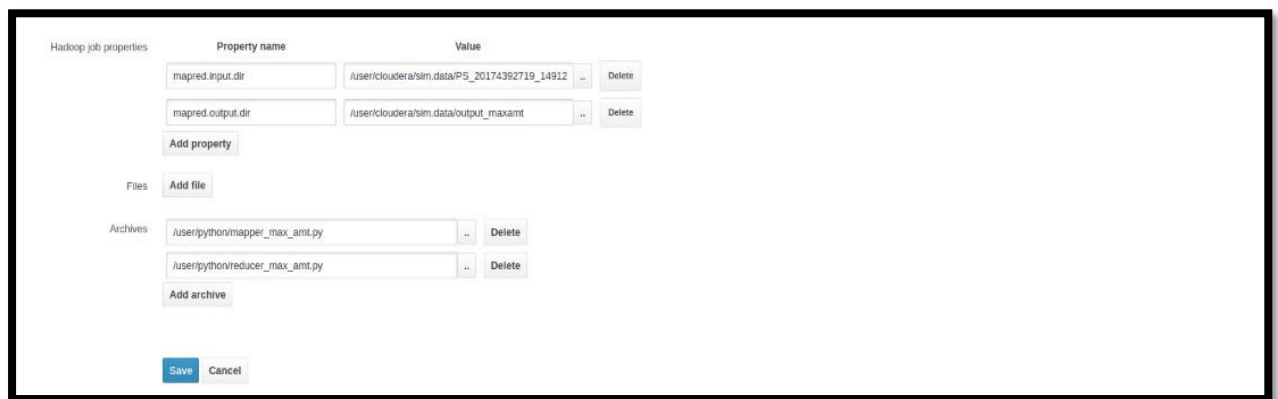
### Previous Steps:



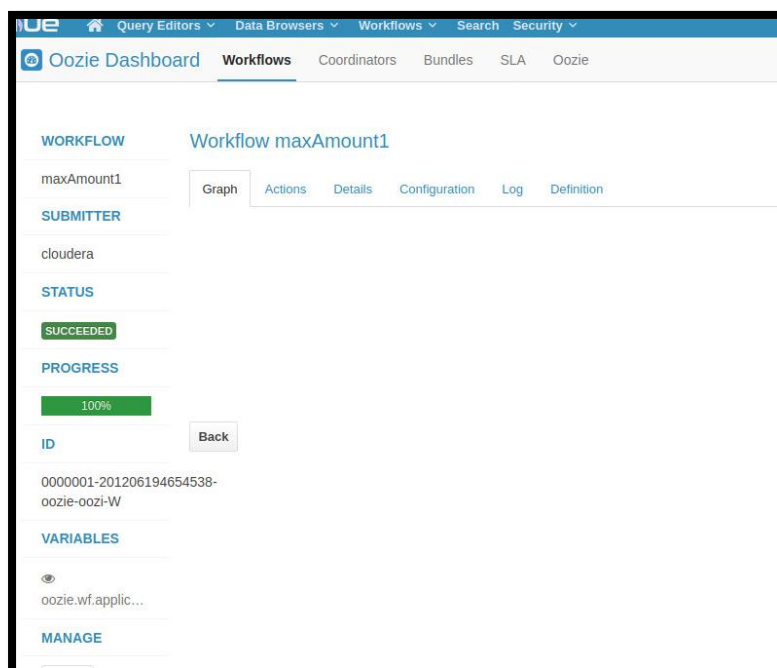
The screenshot shows the HUE Job Designer interface. The top navigation bar includes 'Query Editors', 'Data Browsers', 'Workflows', 'Search', and 'Security'. The main title is 'Job Designer'. Below it, there's a search bar and tabs for 'Overview', 'Edit', 'Log', 'Workflow', and 'Job'. The job is titled 'Job Design (streaming type)'. The 'Name' field contains 'maxAmount1' and the 'Description' field contains 'finding maximum amount'. There is an 'advanced' link.



This screenshot shows the configuration for the Mapper and Reducer. The 'Mapper' field is set to '/usr/bin/python mapper\_max\_amt.p' and the 'Reducer' field is set to '/usr/bin/python reducer\_max\_amt.p'.

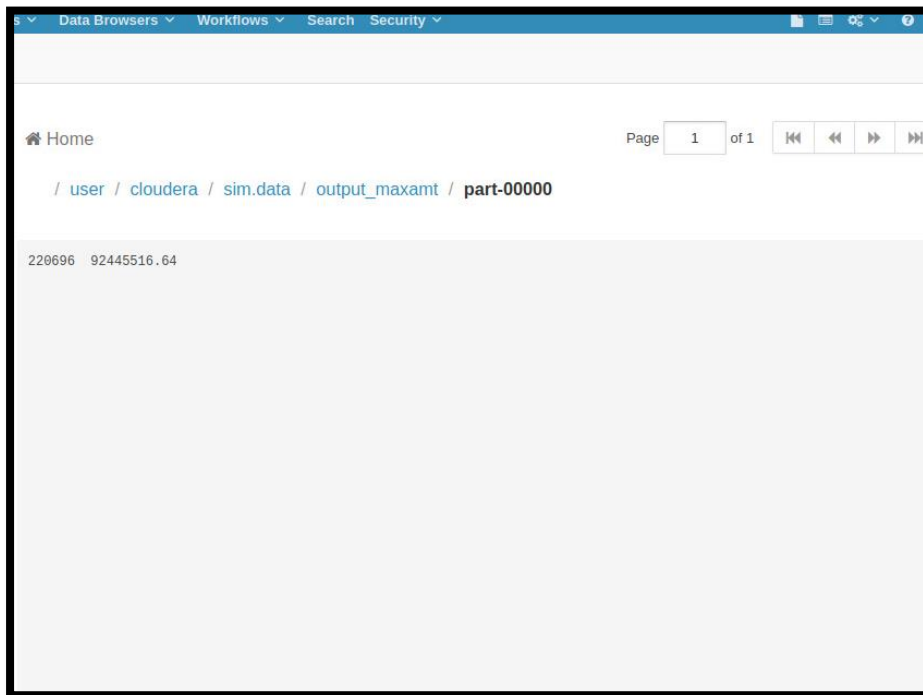


The screenshot displays the 'Hadoop job properties' configuration screen. It has a table with 'Property name' and 'Value' columns. The properties are: 'mapred.input.dir' with value '/user/cloudera/sim.data/PS\_20174392719\_14912' and 'mapred.output.dir' with value '/user/cloudera/sim.data/output\_maxamt'. There are 'Add property', 'Add file', and 'Add archive' buttons. The 'Archives' section shows two files: '/user/python/mapper\_max\_amt.py' and '/user/python/reducer\_max\_amt.py'. At the bottom, there are 'Save' and 'Cancel' buttons.



The screenshot shows the Oozie Dashboard. The top navigation bar includes 'Query Editors', 'Data Browsers', 'Workflows', 'Search', and 'Security'. The main title is 'Oozie Dashboard'. The 'Workflows' tab is selected. The workflow 'maxAmount1' is shown with a status of 'SUCCEEDED'. The progress bar is at 100%. The ID is '0000001-201206194654538-oozie-oozi-W'. The variables section shows 'oozie.wf.applic...'. There is a 'MANAGE' button.

Output:



3a. Why is shuffling needed in MR? Give an example to justify.

Shuffling is necessary in MapReduce because it is the process of transferring data from mappers to reducers through which the system performs the sort and transfers the map output to the reducer as input so it is necessary for reducers otherwise they would not have any input.

3b. Why is sorting needed in MR? Give an example to justify.

Sorting is necessary in MapReduce because sorting in Hadoop helps reducer to easily distinguish when a new reduce task should start. It basically saves time by stating a new reduce task when the next key in the sorted input data is different than previous.

Here in the given example both shuffling and sorting can be seen as explained above.

