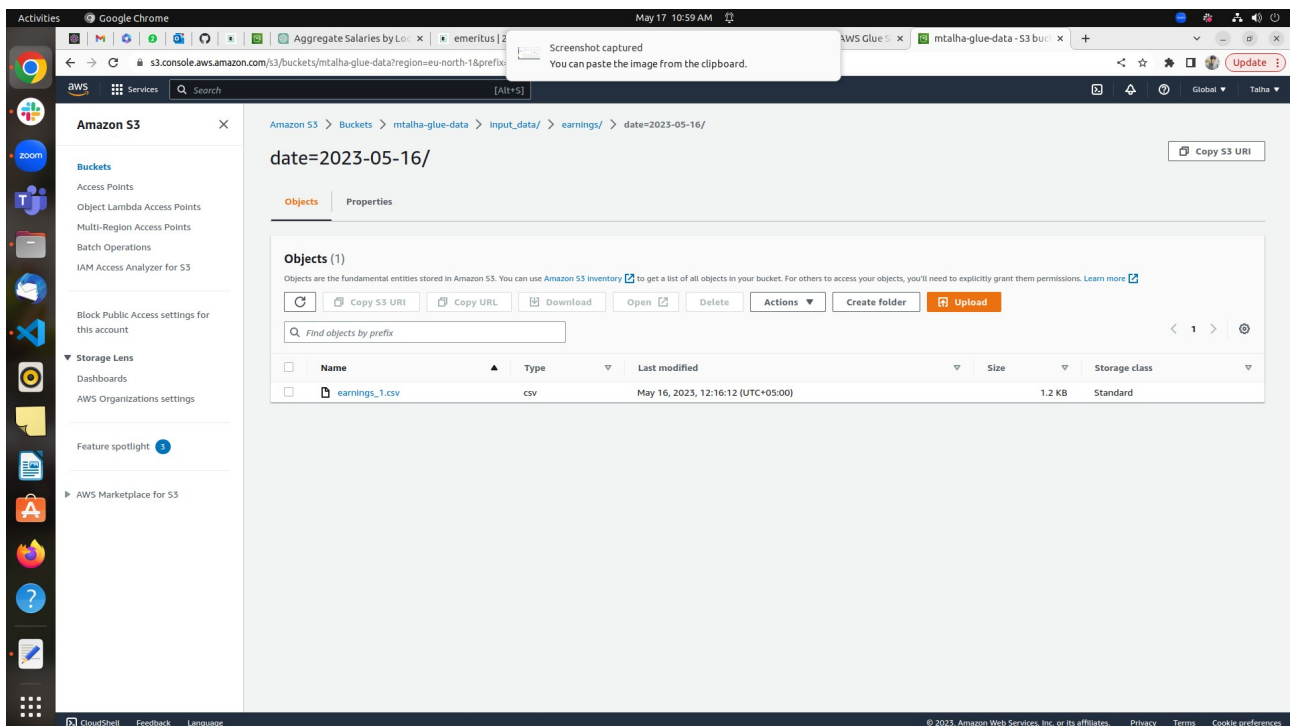
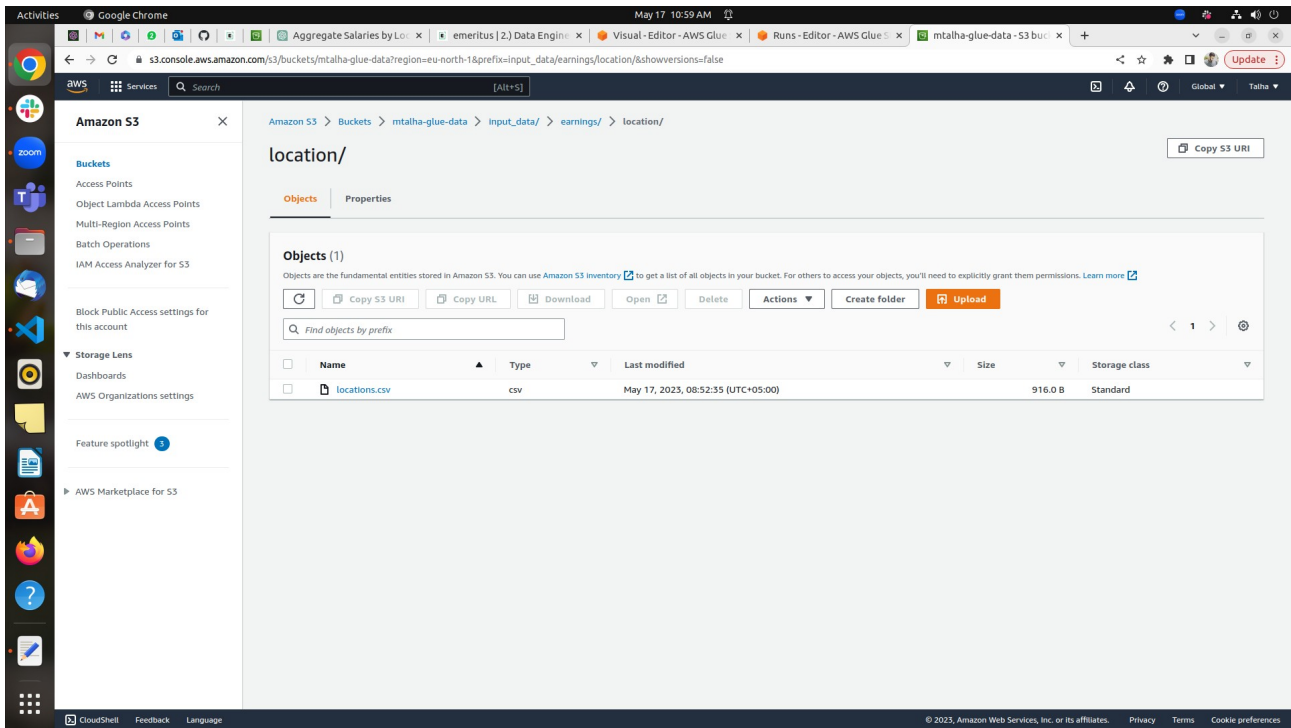
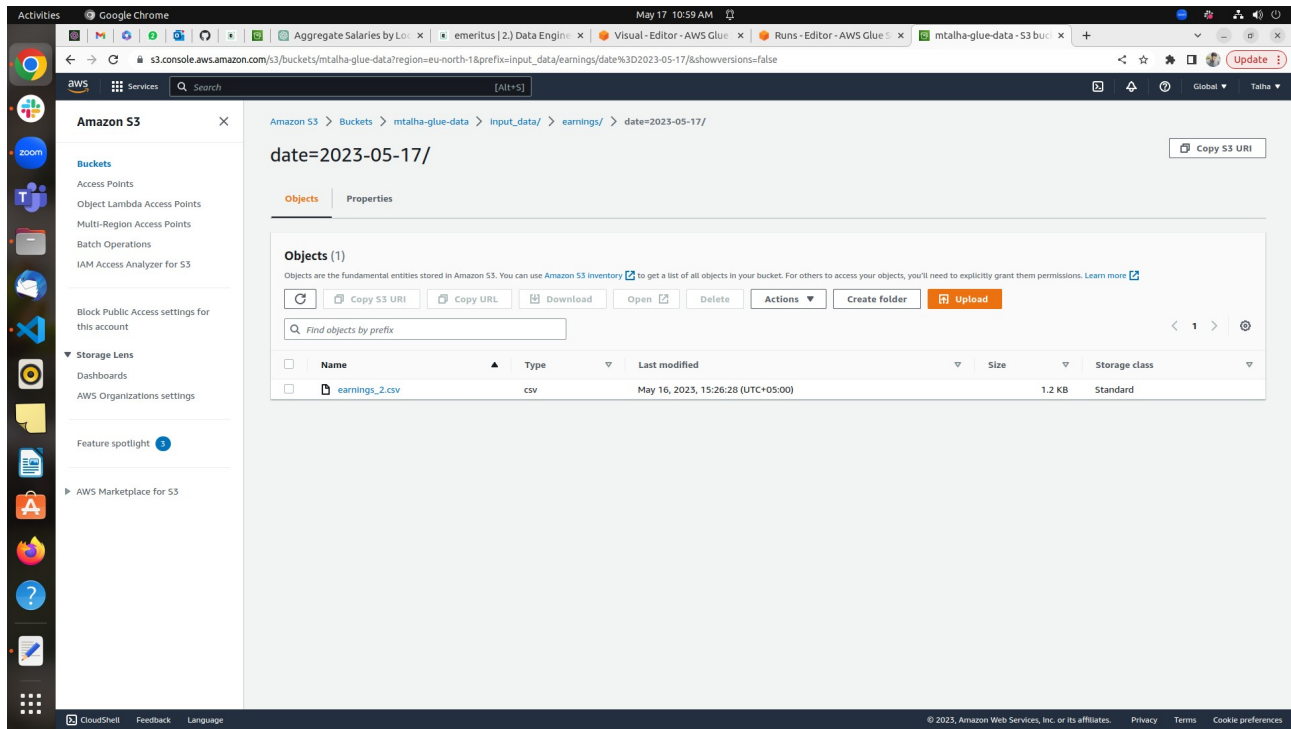


Using the earnings CSV as a base, prepare a new data file with employees' office locations. Make sure there are 5-6 distinct locations that are shared between employees.

Create a Glue job that aggregates the data based on the office location to calculate average salaries and raise percentages for these locations.

Uploading data into S3 bucket





Created Crawler

Activities Google Chrome May 17 10:59 AM

eu-north-1.console.aws.amazon.com/glue/home?region=eu-north-1#/v2/data-catalog/crawlers

AWS Glue

Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)

▼ Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings

► Data Integration and ETL
► Legacy pages

What's New
Documentation
AWS Marketplace

Enable compact mode
Enable new navigation

CloudShell Feedback Language

© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

One crawler successfully deleted
The following crawler is now deleted: "mtalha_rds_crawler"

Crawlers

A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog.

Crawlers (2) [info](#)
View and manage all available crawlers.

Filter crawlers

<input type="checkbox"/>	Name	State	Schedule	Last run	Last run timestamp	Log	Table changes...
<input type="checkbox"/>	mtalha_s3_earnings_crawler	Ready		Succeeded	May 16, 2023 at 10:33:09	View log	1 updated
<input type="checkbox"/>	mtalha_s3_location	Ready		Succeeded	May 17, 2023 at 04:05:31	View log	-

Created Glue Job

Activities Slack May 17 11:03 AM

eu-north-1.console.aws.amazon.com/gluestudio/home?region=eu-north-1#/editor/job/mtalha_employee_earnings_job/graph

AWS Glue

Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)

▼ Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings

► Data Integration and ETL
► Legacy pages

What's New
Documentation
AWS Marketplace

Enable compact mode
Enable new navigation

CloudShell Feedback Language

© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

mtalha_employee_earnings_job

Last modified on 5/17/2023, 11:01:27 AM Try new UI End session Actions

Successfully updated job
Successfully updated job mtalha_employee_earnings_job. To run the job choose the Run Job button.

Visual Script Job details Runs Schedules Version Control

Source Action Target Undo Redo Remove

Data source - S3 bucket Amazon S3

Data source - S3 bucket Amazon S3

Transform - Join Join

Transform - SQL Query SQL Query

Data target - S3 bucket Amazon S3

Transform Output schema Data preview

Data preview (5) [info](#)
Filter sample dataset

location	avg_earning
B	6286.75
C	5576.95
A	5926.05
D	5889.7
E	5599.2

Running the job

The screenshot shows the AWS Glue console interface for a job named `mtalha_employee_earnings_job`. The interface includes a left-hand navigation menu with options like 'Getting started', 'Data Catalog', and 'Legacy pages'. The main content area displays the 'Runs' tab for the job, showing a table of job runs. The most recent run is 'Failed'.

Job runs (1/26)

Run status	Retry	Start time	End time	Duration	Capacity	Worker type	Glue version
Failed	0	05/17/2023 11:01:28	05/17/2023 11:02:40	1 m 5 s	3 DPU's	G.1X	3.0
Failed	0	05/17/2023 10:56:39	05/17/2023 10:57:40	54 s	3 DPU's	G.1X	3.0
Succeeded	0	05/17/2023 10:49:16	05/17/2023 10:50:17	54 s	3 DPU's	G.1X	3.0
Succeeded	0	05/17/2023 10:45:02	05/17/2023 10:46:06	57 s	3 DPU's	G.1X	3.0
Failed	0	05/17/2023 10:43:23	05/17/2023 10:44:18	47 s	3 DPU's	G.1X	3.0

05/17/2023 10:58:36

AnalysisException: Expressions referencing the outer query are not supported outside of WHERE/HAVING clauses:

Job name	Id	Run status	Glue version
mtalha_employee_earnings_job	j_4ecc3ab6bda4cf06513a0a271e08e04951ffc61f8 9737e20bd16bd595acca	Failed	3.0

Retry attempt number	Start time	End time	Start-up time
Initial run	May 17, 2023 10:58:36 AM	May 17, 2023 10:59:28 AM	7 seconds

Execution time	Last modified on	Trigger name	Security configuration
45 seconds	May 17, 2023 10:59:28 AM	-	-

Sql Query & query output showing the result Avg salary based on the location

The screenshot displays the AWS Glue console interface for a job named `mtalha_employee_earnings_job`. The job has been successfully started. The workflow is visualized in the 'Visual' tab, showing a sequence of nodes: two 'Data source - S3 bucket Amazon S3' nodes feeding into a 'Transform - Join Join' node, which then feeds into a 'Transform - SQL Query SQL Query' node, and finally into a 'Data target - S3 bucket Amazon S3' node.

The 'Transform - SQL Query SQL Query' node is selected, showing its configuration in the right-hand pane. The 'Name' field is set to 'SQL Query'. The 'Node parents' section indicates that the two 'Data source - S3 bucket Amazon S3' nodes are the inputs. The 'Input sources' section shows the aliases 'Join' and 'myDataSource'. The 'SQL query' field contains the following SQL statement:

```
1 WITH averages AS (  
2   SELECT  
3     location,  
4     avg(earnings) as avg_earning  
5   FROM  
6     myDataSource  
7   GROUP BY  
8     location  
9 )  
10 SELECT  
11   location,  
12   (avg_earning - (SELECT MIN(avg_earning) FROM myDataSource)) / (SELECT MIN(avg_earning) FROM myDataSource) as avg_salary  
13 FROM  
14   averages
```

Saving the data

