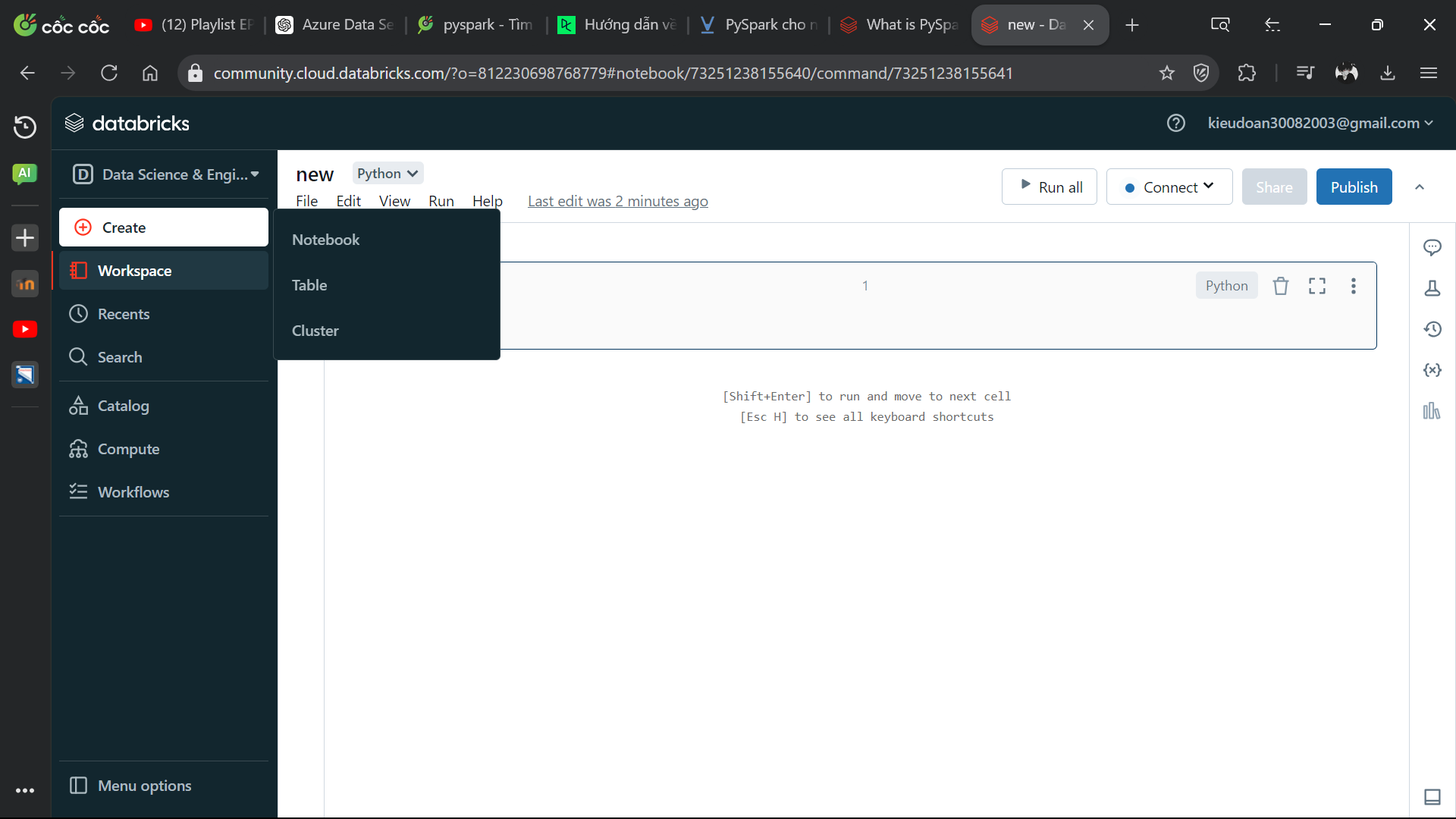
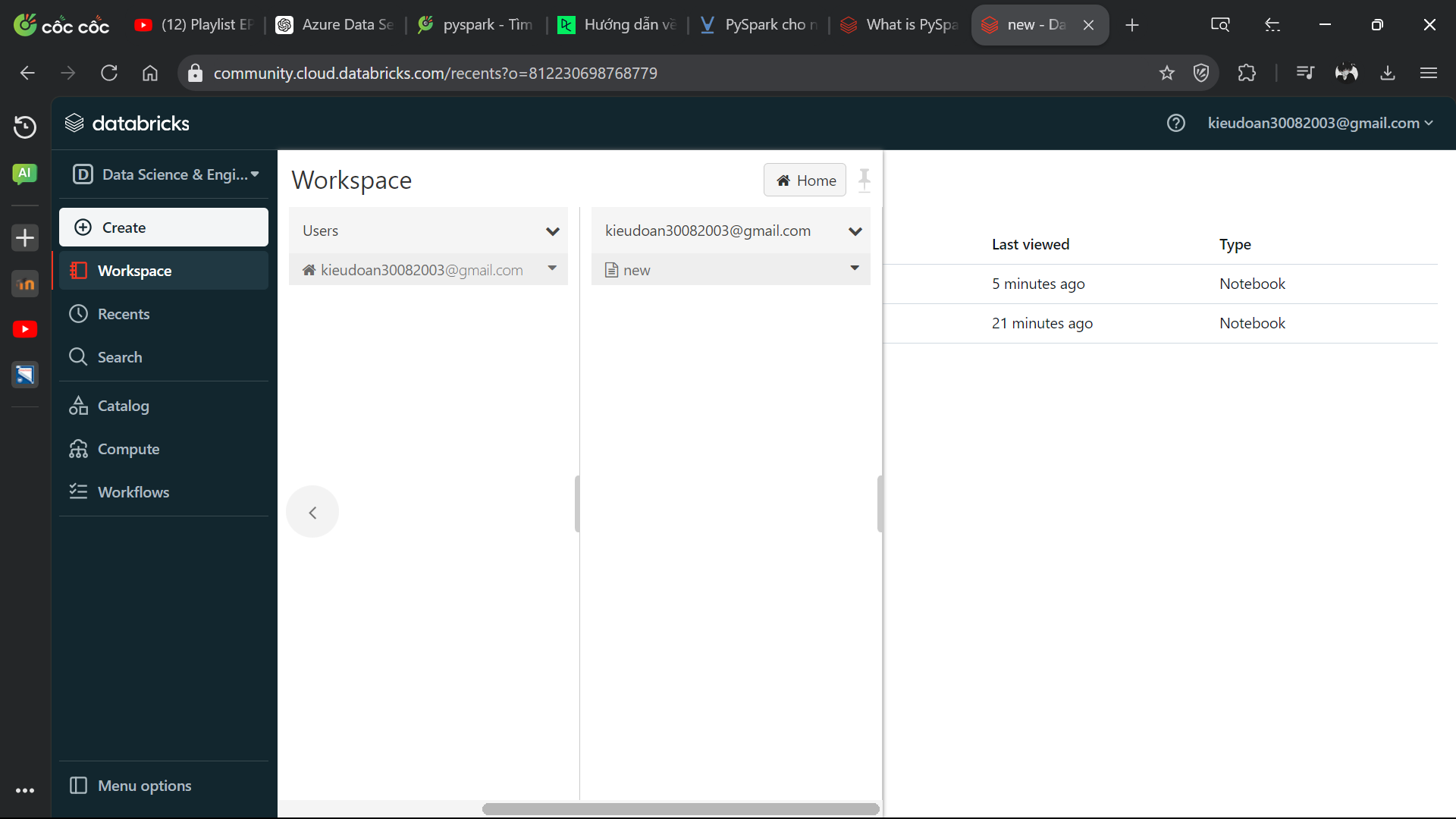
**REPORT**

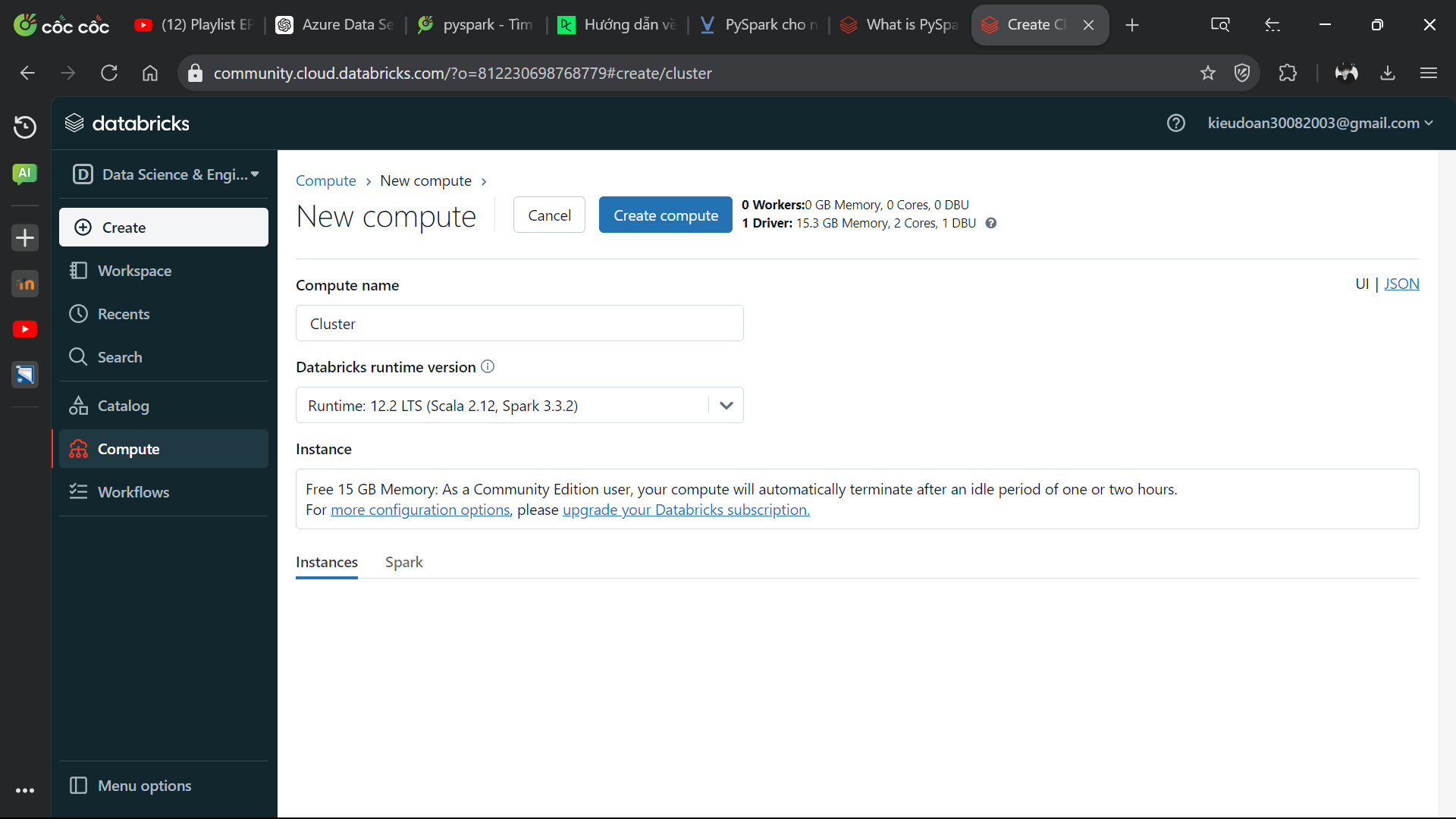
1. **Databricks**
2. Create file

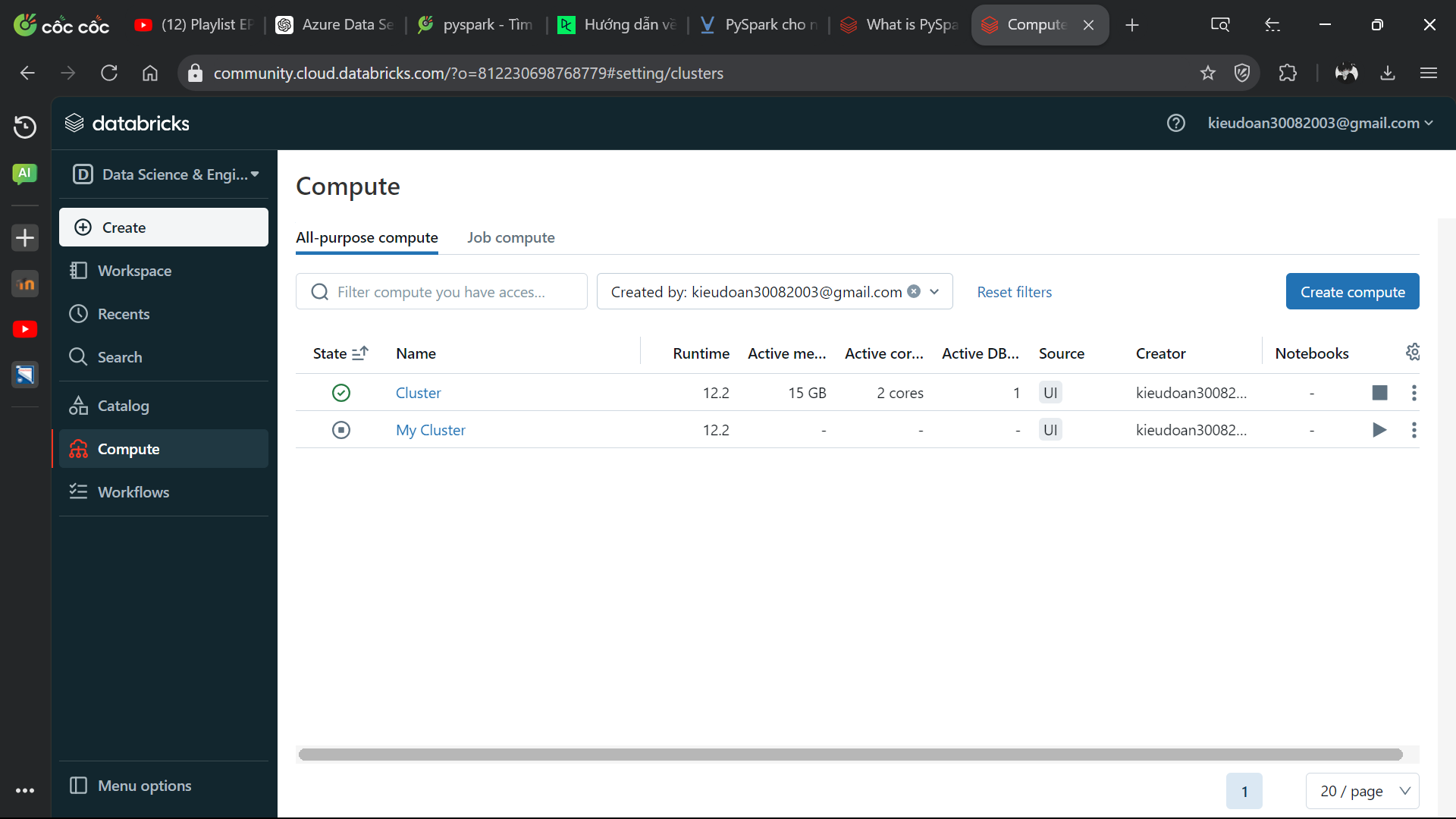
* Notebooks can be created from the create button in the sidebar on the main screen and the files created will go directly into the user section of the current workspace.



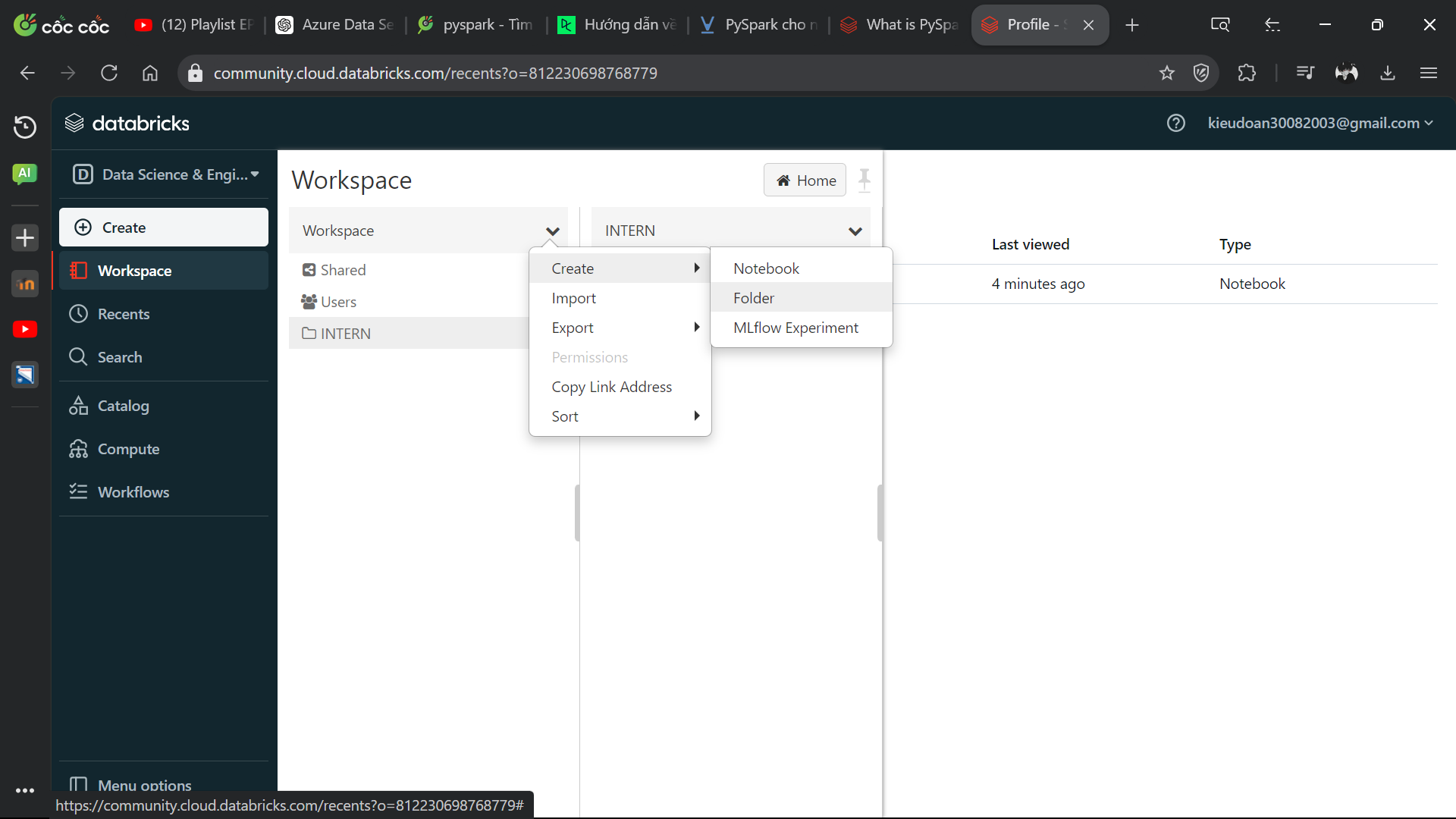


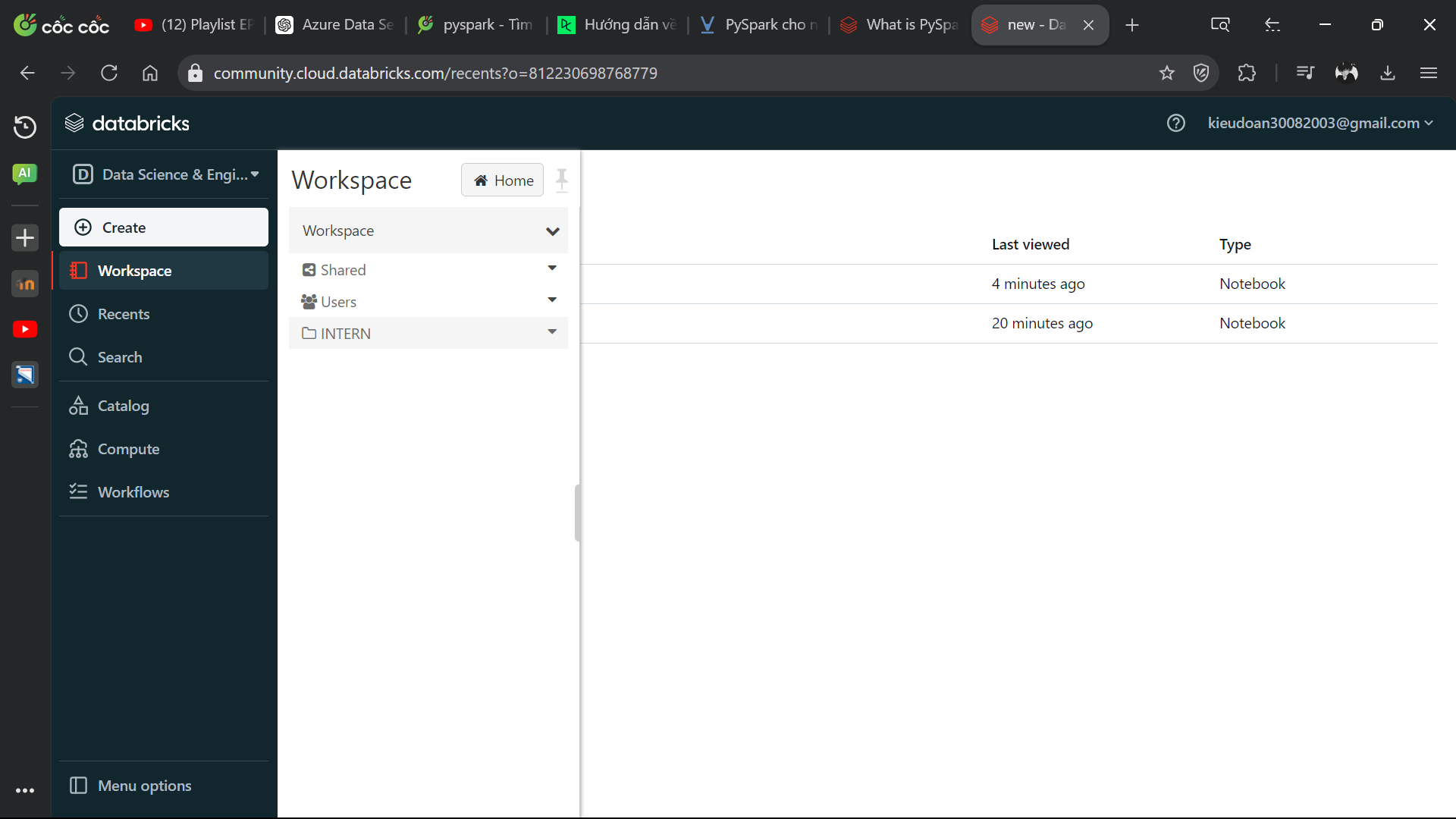
* Create cluster



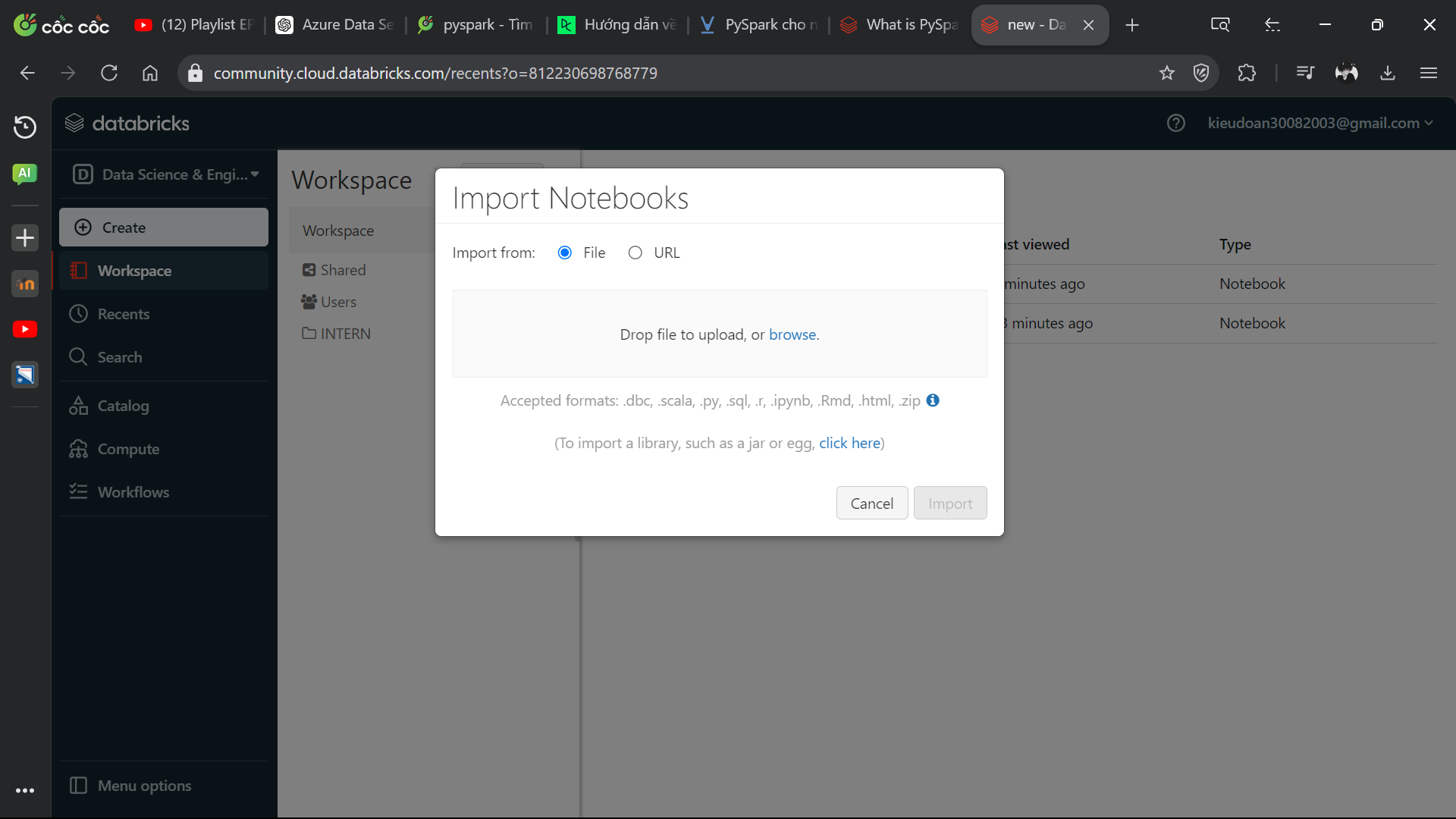


* You can create folders, notebooks, Mlflow Experiment in the workspace and the created files will go to the main section of the workspace

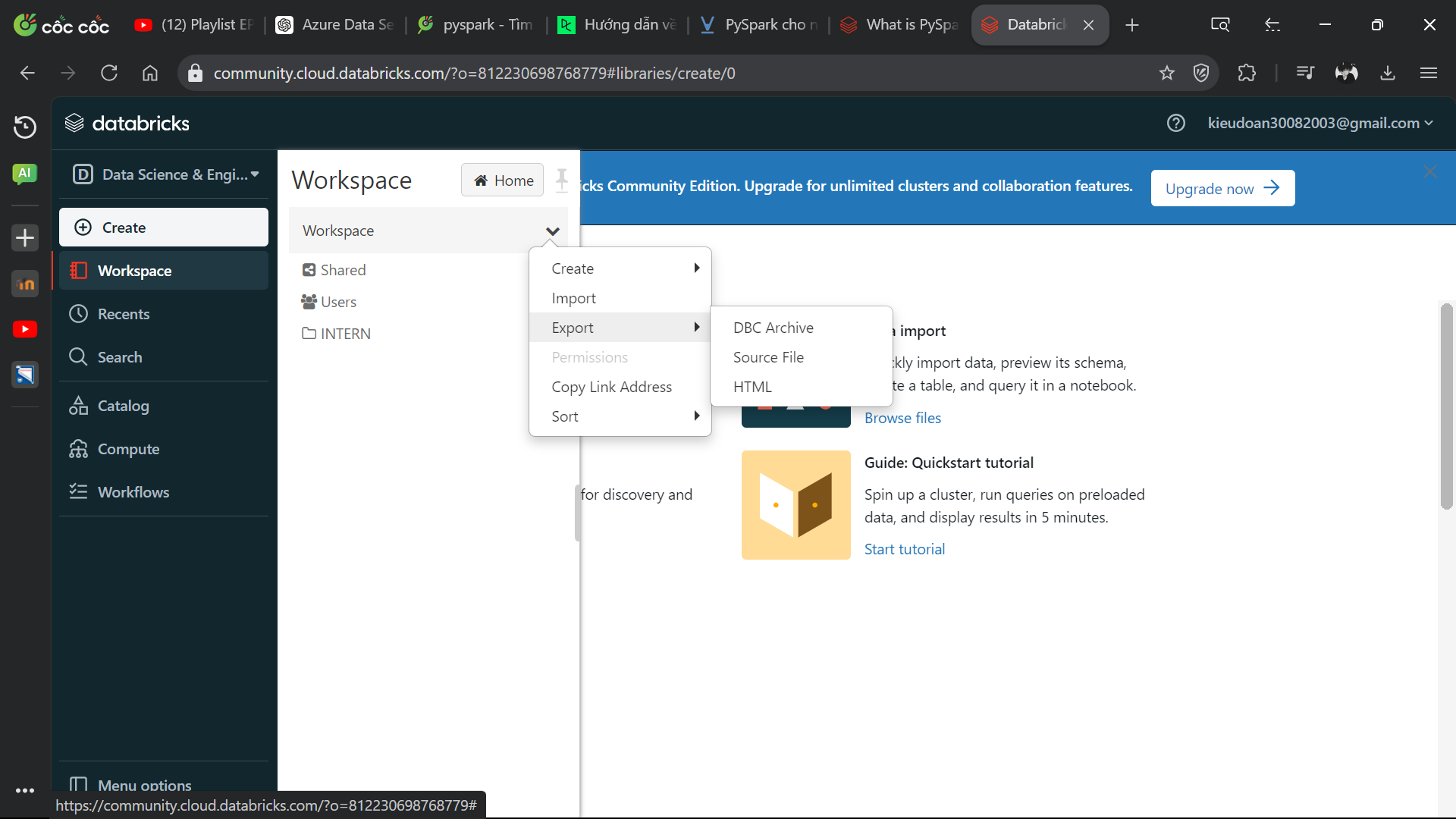




1. Import notebooks in your PC



1. Export file



1. **Pysparks**
2. Display data:

show(): Displays the first few rows in the DataFrame.

1. Select columns:

select(): Select one or more columns from the DataFrame.

selectExpr(): Allows the use of SQL expressions to select and transform columns.

1. Filter data:

filter(): Filter rows in DataFrame based on given conditions.

1. Add or change columns:

withColumn(): Add a new column or change an existing column.

withColumnRenamed(): Change the name of a column.

1. Grouping and aggregation:

groupBy(): Group data by one or more columns.

agg(): Perform aggregate functions like sum, avg, min, max on groups.

1. Sort data:

orderBy() or sort(): Sort data based on one or more columns.

1. Remove duplicate rows:

dropDuplicates(): Remove duplicate rows from DataFrame.

1. Combine DataFrame:

join(): Combine two DataFrames based on a given condition.

1. Handling missing data:

fillna(): Fill in values for missing data cells.

dropna(): Delete rows containing missing data.