PUC Minas

Preparando Ambiente - Spark

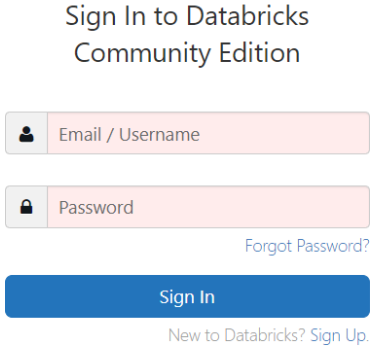
Diego Roberto Gonçalves de Pontes

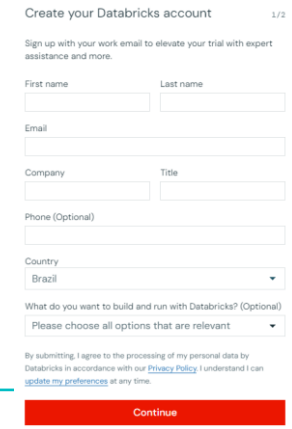
Apache Spark

- O Apache Spark roda localmente para testes, mas para os nossos experimentos vamos utilizar um computador em nuvem para todos poderem utilizar.
- O recurso em nuvem que vai ser utilizado tem um período de testes de 14 dias sem a necessidade de cadastrar com cartão de crédito.

databricks

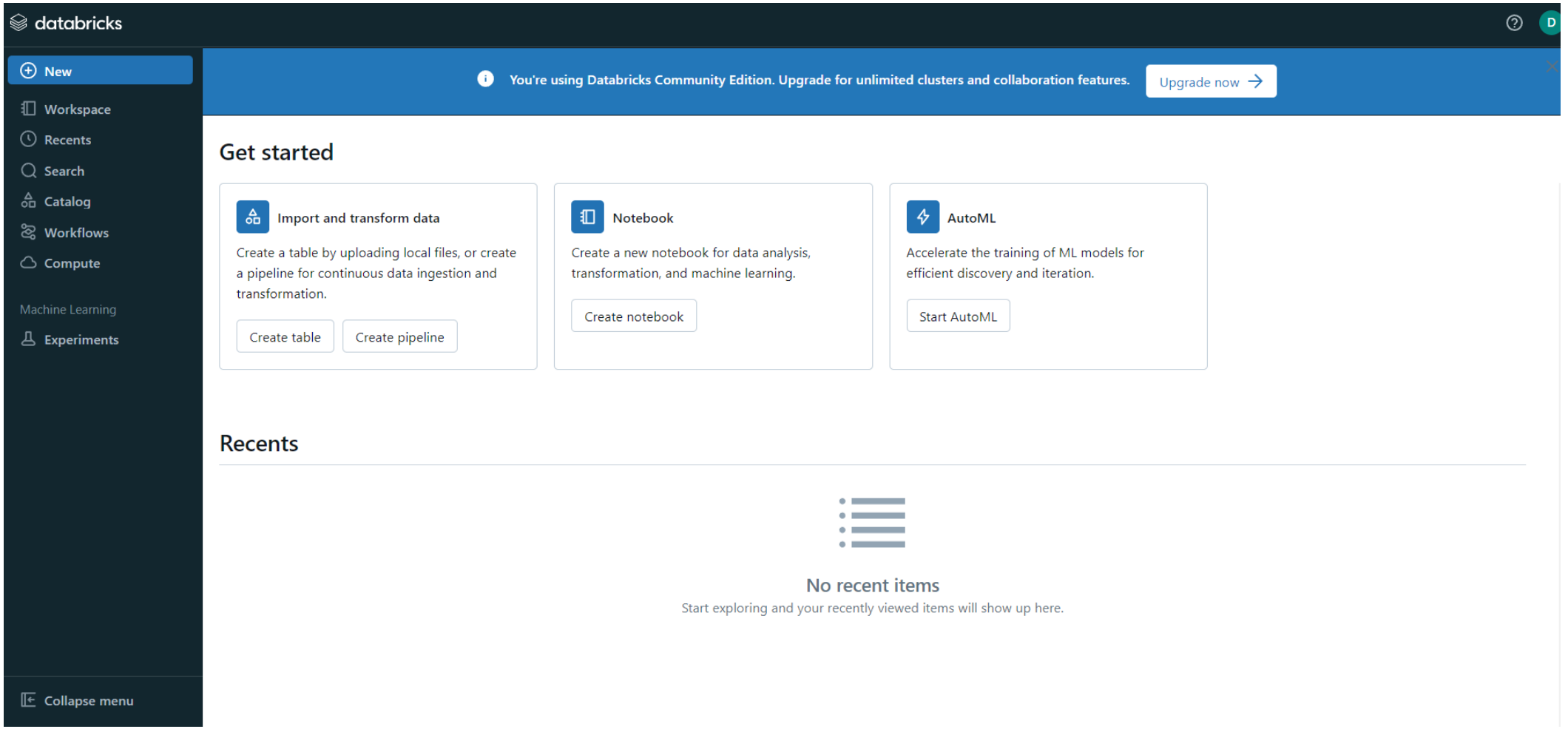
- A databricks já tem em seus clusters o Spark instalado e configurado.
- Acessar o link para fazer cadastro:
  - https://www.databricks.com/try-databricks#account
  - Basta realizar o cadastro e depois acessar o login
  - https://community.cloud.databricks.com/login.html?tuuid=e4292078-e760-4b82-af92-7ae3aa90aeea





PUC Minas

databricks



PUC Minas

Passos

- Criar um cluster;
- Realizar upload de arquivos;
- Criar notebook;
- Executar códigos.

Notebooks

- Notebooks Python são ferramentas interativas que combinam código, texto, e visualizações em um único documento. Eles são amplamente usados em data science, machine learning, educação e pesquisa para experimentar e documentar análises de dados de forma clara e colaborativa. Um dos ambientes mais populares para trabalhar com notebooks Python é o Jupyter Notebook.

PUC Minas