**I. Set up Dataproc để tạo cluster:**

**1) Enable Dataproc Metastore: à Click enable the API**

**Ảnh có chứa văn bản, Phông chữ, ảnh chụp màn hình, đại số

Mô tả được tạo tự động**

**2) Create Dataproc:**

**Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, số

Mô tả được tạo tự độngẢnh có chứa văn bản, số, phần mềm, ảnh chụp màn hình

Mô tả được tạo tự động**

**Ảnh có chứa văn bản, số, phần mềm, Phông chữ

Mô tả được tạo tự động**

**Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động**

**II. Tương tác với PySpark:**

1. **Chọn vào cluster vừa tạo:**

**Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động**

1. **Chọn Web Interfaces à Jupyter hoặc JupyterLab**

**Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, số

Mô tả được tạo tự động**

1. **Upload file:**

**Ảnh có chứa văn bản, phần mềm, Biểu tượng máy tính, Trang web

Mô tả được tạo tự động**

**4) Khởi tạo Spark và kiểm tra phiên bản:**

**Ảnh có chứa ảnh chụp màn hình, văn bản, phần mềm, Biểu tượng máy tính

Mô tả được tạo tự động**

**Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, số

Mô tả được tạo tự động**

**5) Đọc file từ GCS:**

**Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động**

**Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động**

**Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, phần mềm

Mô tả được tạo tự động**

**6) Tương tác với Cột (Columns) trong df:**

**Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, số

Mô tả được tạo tự động**

**Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, số

Mô tả được tạo tự động**

**7) Pyspark và SQL:**

**a) SELECT:**

**Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, số

Mô tả được tạo tự động**

**Ảnh có chứa văn bản, số, phần mềm, ảnh chụp màn hình

Mô tả được tạo tự độngẢnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Biểu tượng máy tính

Mô tả được tạo tự động**

**b) WHERE:**

Ảnh có chứa ảnh chụp màn hình, văn bản, phần mềm

Mô tả được tạo tự độngẢnh có chứa văn bản, ảnh chụp màn hình, Phông chữ

Mô tả được tạo tự động

**c) ORDER BY:**

Ảnh có chứa văn bản, phần mềm, ảnh chụp màn hình, số

Mô tả được tạo tự độngẢnh có chứa văn bản, ảnh chụp màn hình, phần mềm, số

Mô tả được tạo tự độngẢnh có chứa văn bản, ảnh chụp màn hình, phần mềm, số

Mô tả được tạo tự độngẢnh có chứa văn bản, ảnh chụp màn hình, phần mềm, số

Mô tả được tạo tự động

**d) LIMIT:**

**Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm

Mô tả được tạo tự động**

**e) GROUP BY:**

Ảnh có chứa văn bản, ảnh chụp màn hình, hàng

Mô tả được tạo tự động

**8) User Define Function (UDF):**

**Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Trang web

Mô tả được tạo tự động**

**9) Spark SQL:**

**Ảnh có chứa văn bản, ảnh chụp màn hình, số, phần mềm

Mô tả được tạo tự độngẢnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động**

**Ảnh có chứa ảnh chụp màn hình, văn bản, Phông chữ, hàng

Mô tả được tạo tự động**

**10) Ghi kết quả ra file Parquet:**

**Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động**

**11) Ví dụ thực tế:**

**Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, hàng

Mô tả được tạo tự động**

**Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự độngẢnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động**

**12) Resilient Distributed Datasets (RDD): à dùng để xử lý song song**

**Ảnh có chứa văn bản, ảnh chụp màn hình, số, phần mềm

Mô tả được tạo tự độngẢnh có chứa văn bản, ảnh chụp màn hình, phần mềm, số

Mô tả được tạo tự động**

**13) Map (“key”:“value”):**

**Ảnh có chứa văn bản, ảnh chụp màn hình, số, phần mềm

Mô tả được tạo tự độngẢnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Biểu tượng máy tính

Mô tả được tạo tự độngẢnh có chứa văn bản, số, phần mềm, Phông chữ

Mô tả được tạo tự độngẢnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động**

**14) Reduce:**

**Ảnh có chứa văn bản, phần mềm, ảnh chụp màn hình, số

Mô tả được tạo tự động**

**15) Word Count bằng RRD:**

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Biểu tượng máy tính

Mô tả được tạo tự độngẢnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Trang web

Mô tả được tạo tự độngẢnh có chứa văn bản, ảnh chụp màn hình

Mô tả được tạo tự động