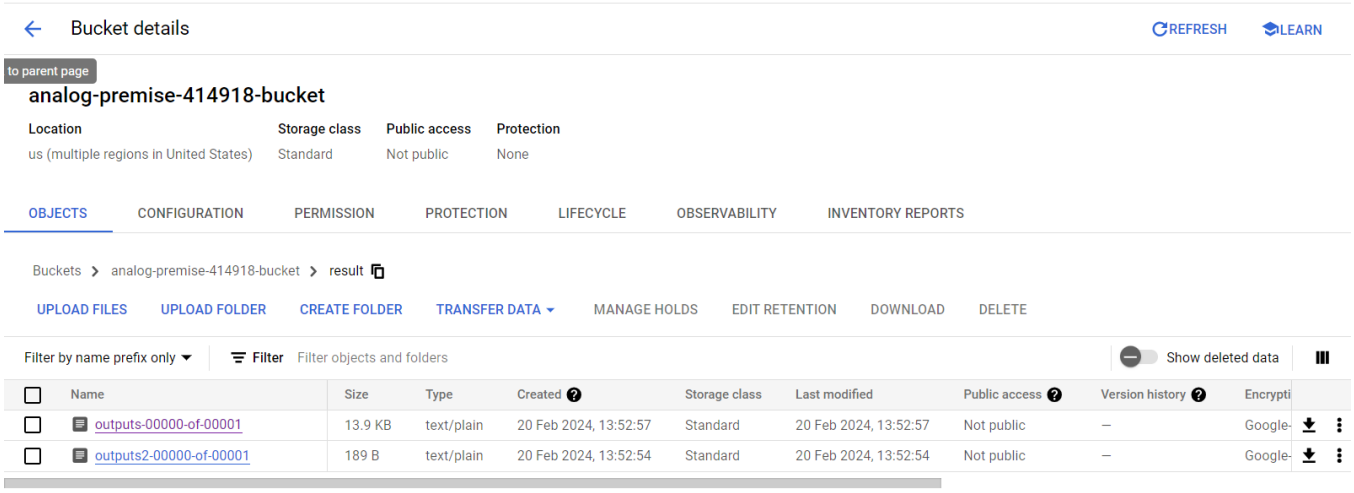
Cloud Computing – Lab

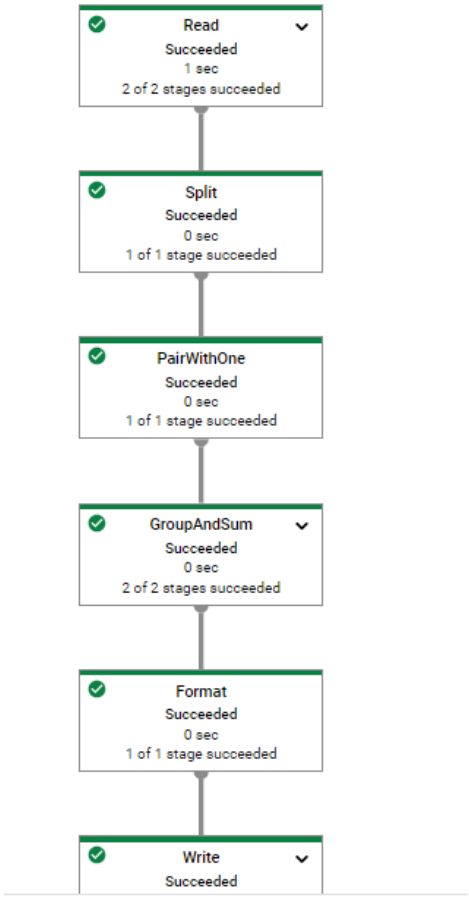
Description of wordcount2.py

What wordcount2.py does is take text as input and parses it into exact words. In essence, it performs word counting on a file. The imports are necessary modules for building data processing pipelines. The 'WordExtractingDoFn' class is used to process lines of text in the input file and parses it into words. The 'def run' function will run the main class using the Apache pipelines. The argument parser parses arguments to specify input and 2 output file locations. Then, the pipeline is constructed and configured. With pipeline running, the program reads from input file and saves it. Then, it will begin extracting and counting. It will split each line into separate words using the 'WordExtractingDoFn' class and converts all words to lowercase. Then it splits into two outputs/pipelines, one pipeline filters words starting from a-f. It PairsWithOne and GroupAndSum. The second pipeline extracts first letter and also does PairWithOne2 and GroupAndSum. Then, both pipelines will format the result. Then will write the file to text. As a result, there will be two different outputs.

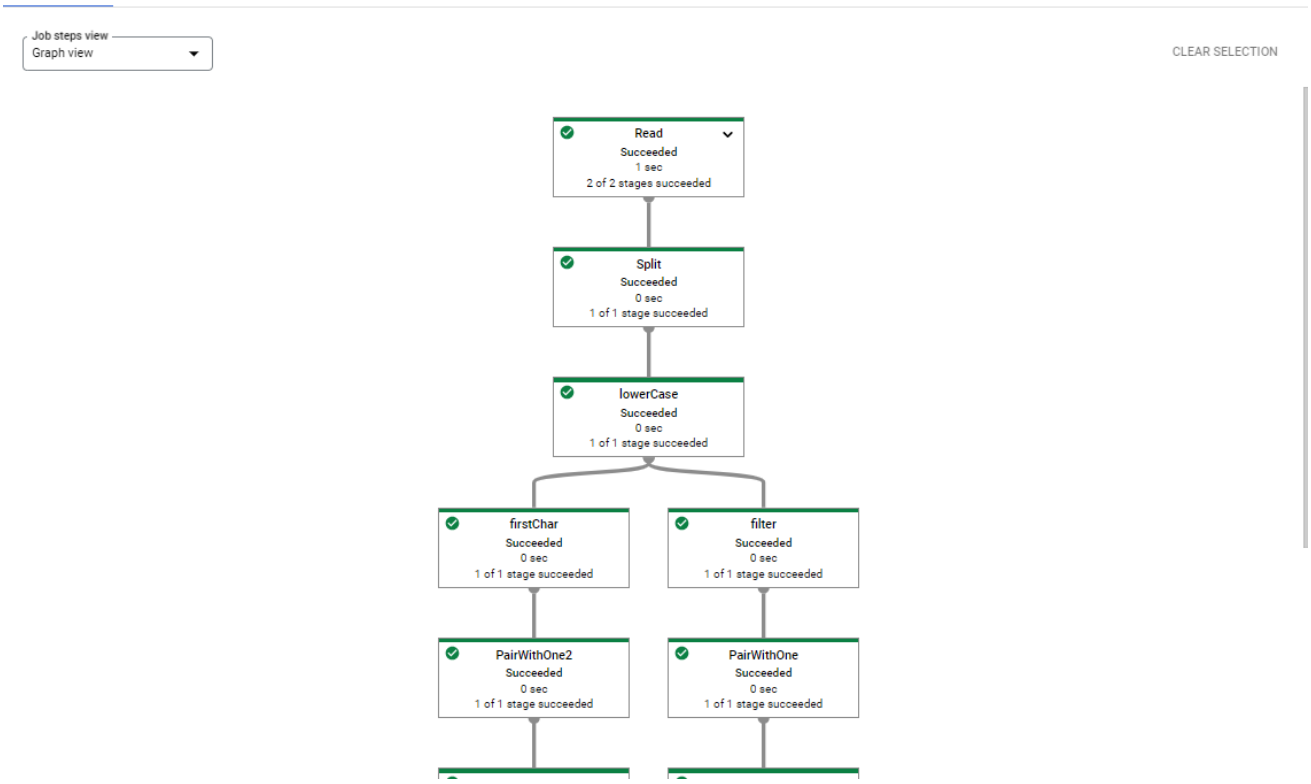
WordCount Examples

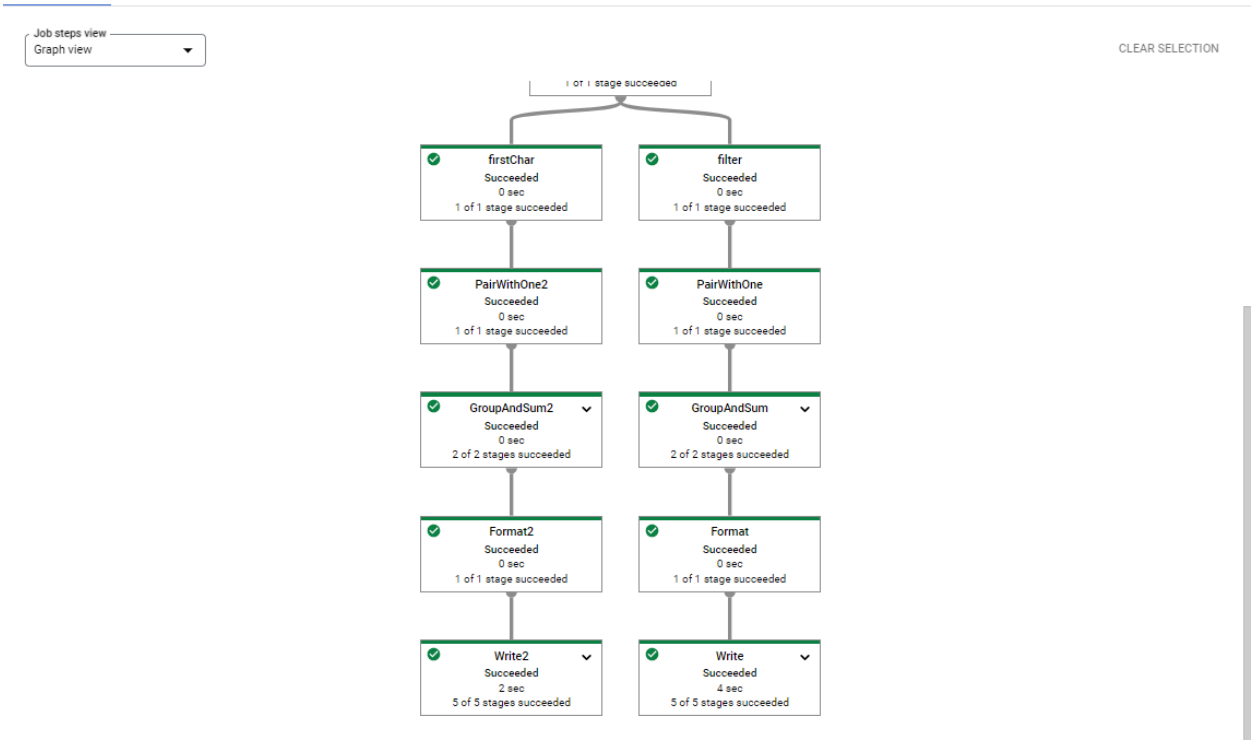


Bucket File Outputs



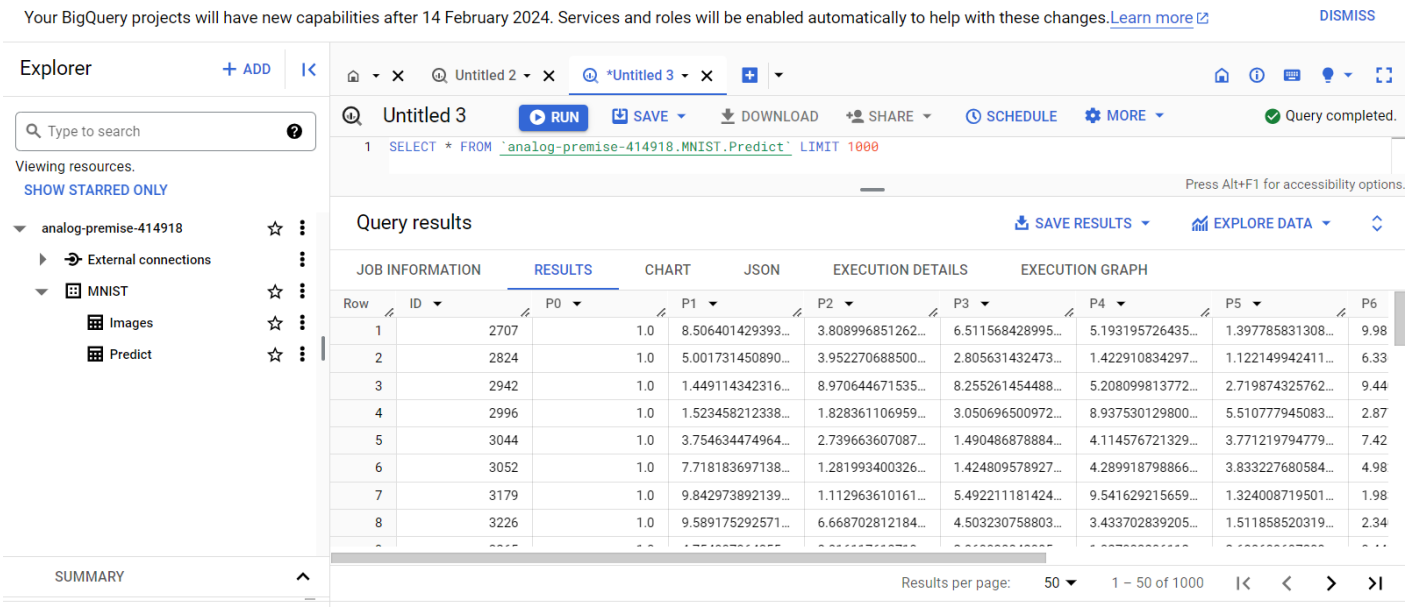
Data Flow Diagram for wordcount.py



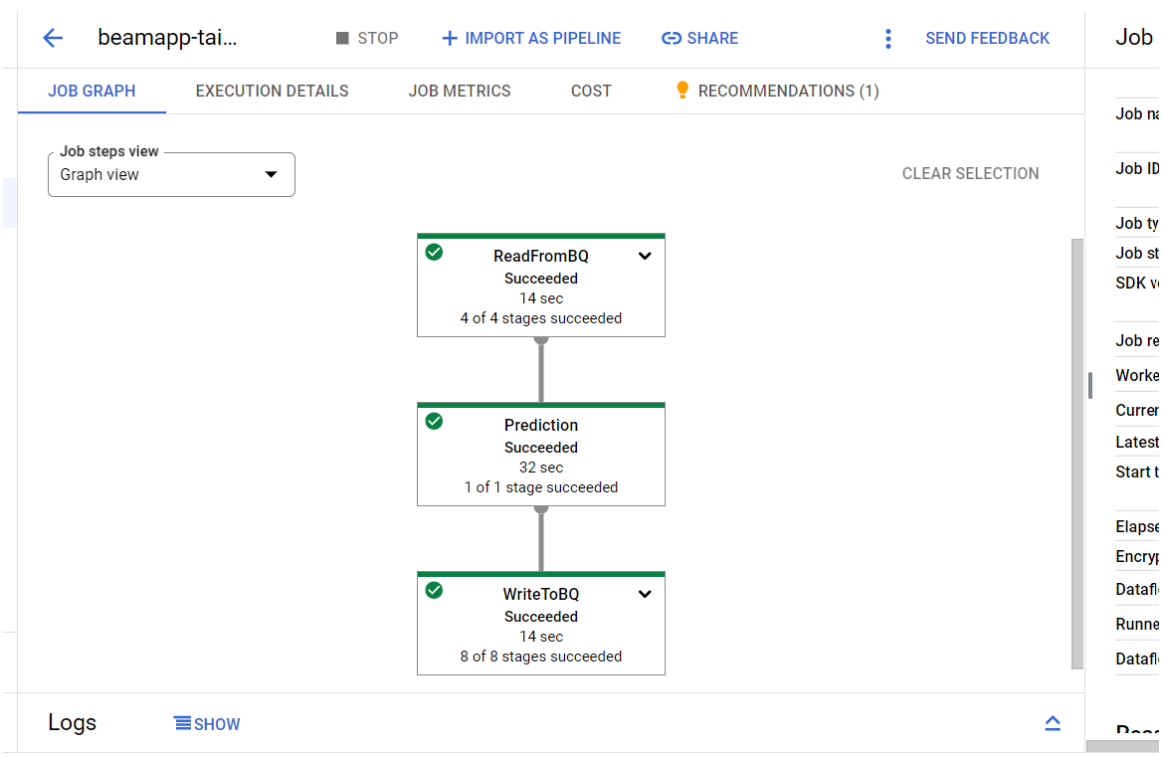


Dataflow Diagram for wordcount2.py

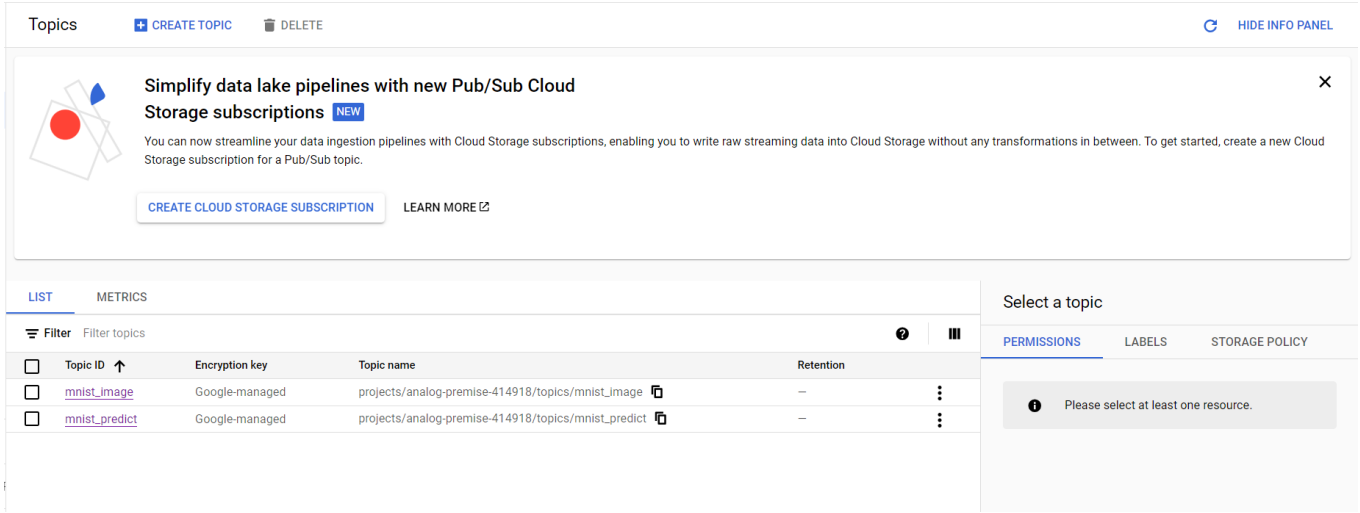
MNIST Examples



Row	ID	P0	P1	P2	P3	P4	P5	P6
1	2707	1.0	8.506401429393...	3.808996851262...	6.511568428995...	5.193195726435...	1.397785831308...	9.98
2	2824	1.0	5.001731450890...	3.952270688500...	2.805631432473...	1.422910834297...	1.122149942411...	6.33
3	2942	1.0	1.449114342316...	8.970644671535...	8.255261454488...	5.208099813772...	2.719874325762...	9.44
4	2996	1.0	1.523458212338...	1.828361106959...	3.050696500972...	8.937530129800...	5.510777945083...	2.87
5	3044	1.0	3.754634474964...	2.739663607087...	1.490486878884...	4.114576721329...	3.771219794779...	7.42
6	3052	1.0	7.718183697138...	1.281993400326...	1.424809578927...	4.289918798866...	3.833227680584...	4.98
7	3179	1.0	9.842973892139...	1.112963610161...	5.492211181424...	9.541629215659...	1.324008719501...	1.98
8	3226	1.0	9.589175292571...	6.668702812184...	4.503230758803...	3.433702839205...	1.511858520319...	2.34



MNIST Example 1 Dataflow Diagram



Topics Creation

MNIST Example 2 Dataflow Diagram

Videos

WordCount:

https://drive.google.com/file/d/1CqXCHB9jXUYFiFxhIFzWEvE4UHx_cIJS/view?usp=sharing

MNIST:

https://drive.google.com/file/d/1N6w7N3MVVMhOA2_6rq_mpDZUUYQHTWJC/view?usp=sharing

Design

GitHub Design: https://github.com/TaimourArshad1/cloud-lab2