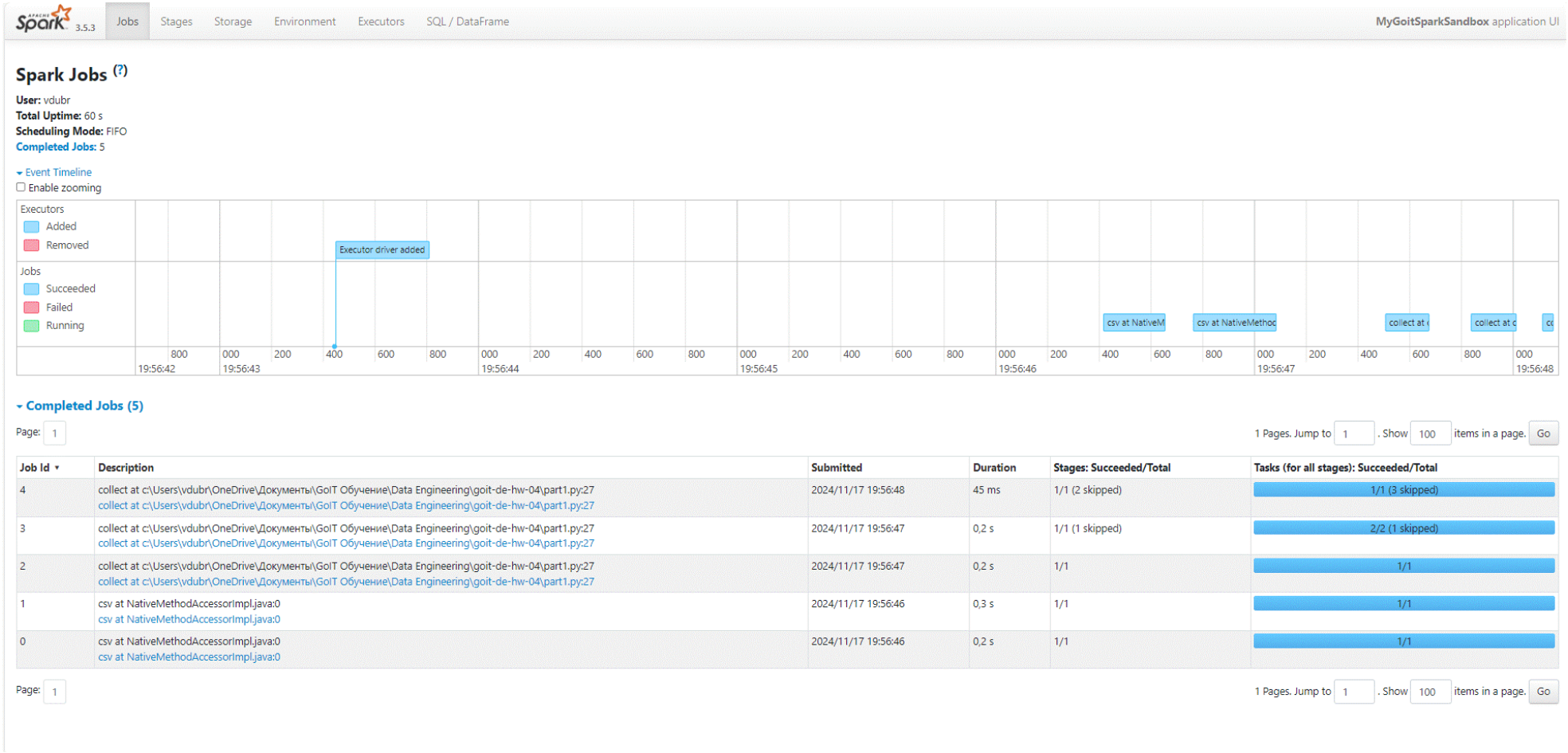
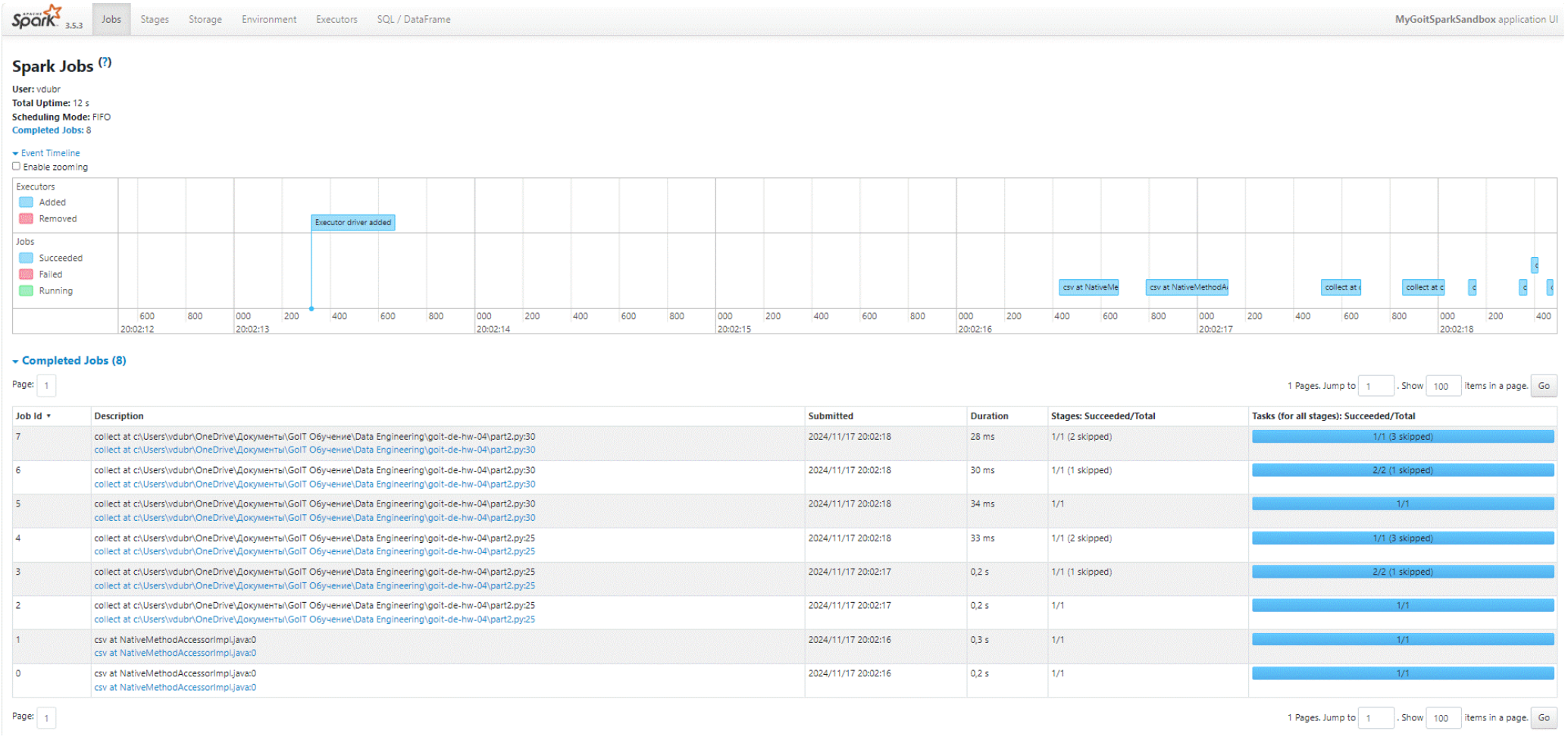
Частина 1:



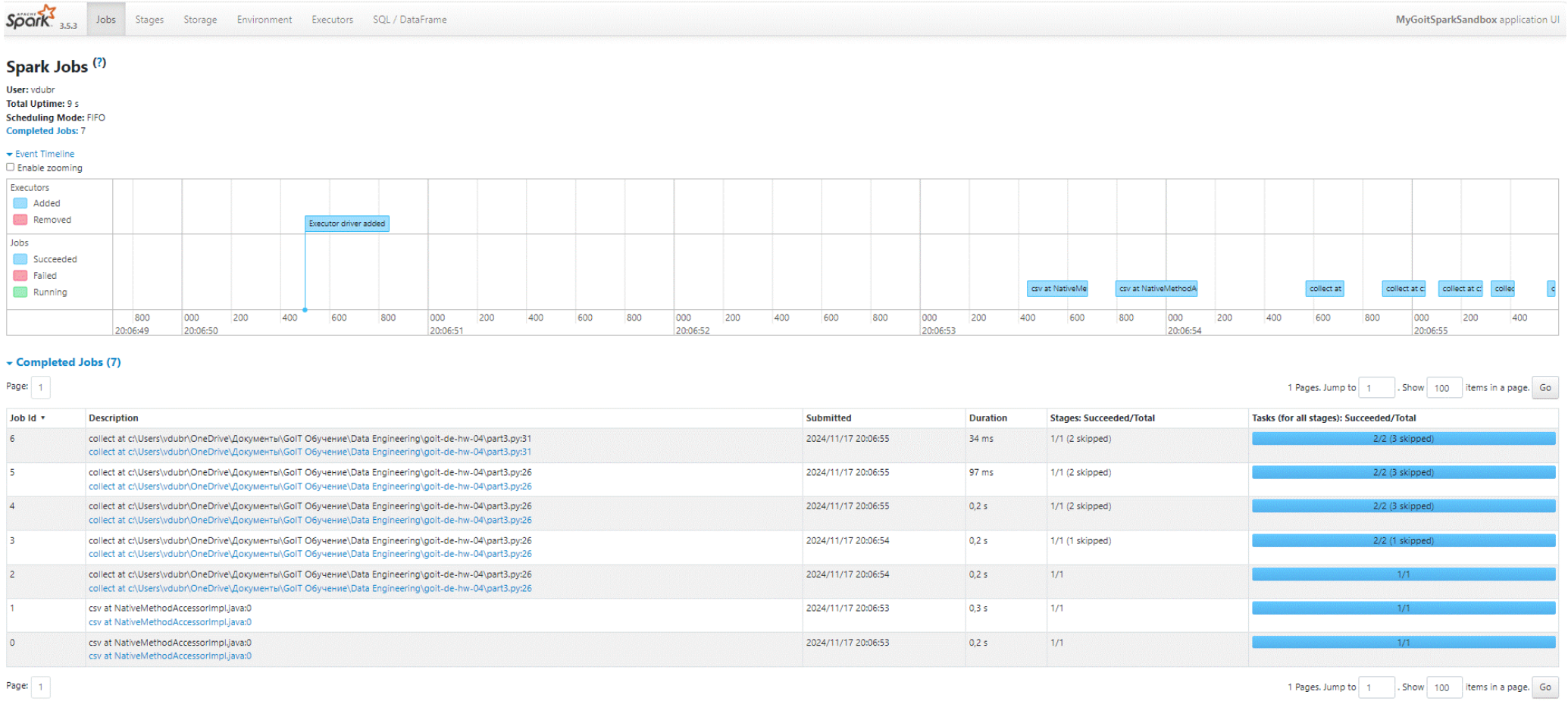
Job Id	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
4	collect at c:\Users\vdubr\OneDrive\Документы\GoIT Обучение\Data Engineering\goit-de-hw-04\part1.py:27 collect at c:\Users\vdubr\OneDrive\Документы\GoIT Обучение\Data Engineering\goit-de-hw-04\part1.py:27	2024/11/17 19:56:48	45 ms	1/1 (2 skipped)	1/1 (3 skipped)
3	collect at c:\Users\vdubr\OneDrive\Документы\GoIT Обучение\Data Engineering\goit-de-hw-04\part1.py:27 collect at c:\Users\vdubr\OneDrive\Документы\GoIT Обучение\Data Engineering\goit-de-hw-04\part1.py:27	2024/11/17 19:56:47	0,2 s	1/1 (1 skipped)	2/2 (1 skipped)
2	collect at c:\Users\vdubr\OneDrive\Документы\GoIT Обучение\Data Engineering\goit-de-hw-04\part1.py:27 collect at c:\Users\vdubr\OneDrive\Документы\GoIT Обучение\Data Engineering\goit-de-hw-04\part1.py:27	2024/11/17 19:56:47	0,2 s	1/1	1/1
1	csv at NativeMethodAccessorImpl.java:0 csv at NativeMethodAccessorImpl.java:0	2024/11/17 19:56:46	0,3 s	1/1	1/1
0	csv at NativeMethodAccessorImpl.java:0 csv at NativeMethodAccessorImpl.java:0	2024/11/17 19:56:46	0,2 s	1/1	1/1

Частина 2



Додавання nuek_processed.collect() на 25му рядку запускає обчислення попередніх дій у Spark (3 джоба). А другий виклик nuek_processed.collect() на 30 рядку заново обчислює всі дані (5 джобів, як у частині 1). 3 + 5 = 8.

Частина 3



До першого nuek_processed_cached.collect() на 26 рядку ми виконали 3 джоба. Плюс 1 джоб на кешування.

Другий виклик nuek_processed.collect() виконує 1 джоб на отримання даних з кещу, виконує джобом код: nuek_processed = nuek_processed_cached.where("count>2"), та збирає дані на драйвері ще 1 джобом.

Тобто: (3 + 1) + (1 + 1 + 1) = 7 джобів.