

24/11/18 15:29:13 WARN Utils: Your hostname, DESKTOP-NQFJ69T resolves to a loopback address: 127.0.1.1; using 172.19.170.19 instead (on interface eth0)

24/11/18 15:29:13 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address

Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties

Setting default log level to "WARN".

To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).

24/11/18 15:29:20 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Spark context Web UI available at <http://172.19.170.19:4040>

Spark context available as 'sc' (master = local[*], app id = local-1731907761502)

Spark session available as 'spark'.

WARNING: one deprecation (since 2.0.0); for details, enable ':setting -deprecation' or ':replay -deprecation'

ERROR StatusLogger Log4j2 could not find a logging implementation. Please add log4j-core to the classpath. Using SimpleLogger to log to the console...

Код предмета	Предмет	Учитель	Код студента	Фамилия студента	Имя студент
П01	Проектирование БД	Моисеев	С01	Рогов	Васили
П01	Проектирование БД	Моисеев	С02	Бахмутов	Паве
П01	Проектирование БД	Моисеев	С03	Васильев	Ле
П02	Машинное обучение	Щербань	С02	Бахмутов	Паве
П02	Машинное обучение	Щербань	С03	Васильев	Ле

24/11/18 15:29:31 WARN ExcelHeaderChecker: Number of column in Excel header is not equal to number of fields in the schema:
Header length: 6, schema size: 3
Excel file: file:///home/andrey/Sem6.xlsx

24/11/18 15:29:32 WARN ExcelHeaderChecker: Number of column in Excel header is not equal to number of fields in the schema:
Header length: 6, schema size: 4
Excel file: file:///home/andrey/Sem6.xlsx

task 1

00:00:07

root@DESKTOP-NQFJ69T:/home# cd andrey

root@DESKTOP-NQFJ69T:/home/andrey# ls

Sem6.xlsx airflow-venv s4.xlsx s4_2.xlsx s6.py s6s1.scala

root@DESKTOP-NQFJ69T:/home/andrey# python3 s6.py

Traceback (most recent call last):

File "s6.py", line 11, in <module>

DAGs

All 2 Active 1 Paused 1 Running 0 Failed 0

Filter DAGs by tag Search DAGs Auto-refresh

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	Actions	Links
AGanshin001	AGanshin	1	0 6 * * *	2024-11-18, 15:52:53	2024-11-18, 13:00:00	1	[Play] [Stop]	...
AGanshin002	AGanshin	0	0 6 * * *		2024-11-18, 07:00:00	0	[Play] [Stop]	...

Showing 1-2 of 2 DAGs

AGanshin001 - Grid - Airflow

localhost:8080/dags/AGanshin001/grid?dag_run_id=manual__2024-11-18T05%3A52%3A53.671707%2B00%3A00&task_id=pyspark&tab=logs

18.11.2024 17:22:03 25 All Run Types All Run States Clear Filters Auto-refresh

Press **shift** + **f** for Shortcuts

AGanshin001 / 2024-11-18, 13:00:00 +10 pyspark

Details Graph Gantt Code Logs

(by attempts)

1

All Levels All File Sources Wrap Download See More

```
DESKTOP-MQF369T.
*** Found local files:
*** = /home/airflow/logs/dag_id=AGanshin001/run_id=manual__2024-11-18T05:52:53.671707+00:00/task_id=pyspark/attempt=1.log
[2024-11-18, 15:52:54 +10] [taskinstance.py:1159] INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: AGanshin001.pyspark manual__2024-11-18T05:52:53.671707+00:00 [queued]
[2024-11-18, 15:52:54 +10] [taskinstance.py:1159] INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: AGanshin001.pyspark manual__2024-11-18T05:52:53.671707+00:00 [queued]
[2024-11-18, 15:52:54 +10] [taskinstance.py:1361] INFO - Starting attempt 1 of 1
[2024-11-18, 15:52:54 +10] [taskinstance.py:1382] INFO - Executing <Task(BashOperator): pyspark> on 2024-11-18 05:52:53.671707+00:00
[2024-11-18, 15:52:54 +10] [standard_task_runner.py:57] INFO - Starved process 14225 to run task
[2024-11-18, 15:52:54 +10] [standard_task_runner.py:84] INFO - Running: ['airflow', 'tasks', 'run', 'AGanshin001', 'pyspark', 'manual__2024-11-18T05:52:53.671707+00:00', '--job-id', '5',
[2024-11-18, 15:52:54 +10] [standard_task_runner.py:85] INFO - Job 5: Subtask pyspark
[2024-11-18, 15:52:54 +10] [task_command.py:416] INFO - Running <TaskInstance: AGanshin001.pyspark manual__2024-11-18T05:52:53.671707+00:00 [running]> on host DESKTOP-MQF369T.
[2024-11-18, 15:52:54 +10] [taskinstance.py:1662] INFO - Exporting env vars: AIRFLOW_CTX_DAG_EMAIL="alex@alex.ru" AIRFLOW_CTX_DAG_OWNER="AGanshin" AIRFLOW_CTX_DAG_ID="AGanshin001" AIRFLOW_CTX_TASK_ID="pyspark"
[2024-11-18, 15:52:54 +10] [subprocess.py:75] INFO - Running command: ['/usr/bin/bash', '-c', 'python3 /home/andrey/s6_2.py']
[2024-11-18, 15:52:54 +10] [subprocess.py:86] INFO - Output:
[2024-11-18, 15:52:54 +10] [subprocess.py:93] INFO - Hello
[2024-11-18, 15:52:54 +10] [subprocess.py:97] INFO - Command exited with return code 0
[2024-11-18, 15:52:54 +10] [taskinstance.py:1400] INFO - Marking task as SUCCESS. dag_id=AGanshin001, task_id=pyspark, execution_date=2024-11-18T05:52:53, start_date=2024-11-18T05:52:54, end_date=2024-11-18T05:52:54
[2024-11-18, 15:52:54 +10] [local_task_job_runner.py:228] INFO - Task exited with return code 0
[2024-11-18, 15:52:54 +10] [taskinstance.py:2778] INFO - 0 downstream tasks scheduled from follow-on schedule check
```

Airflow\spark\tasket1b\ - HeidiSQL 12.7.0.6850

айл Редактировать Поиск Запрос Инструменты Переход Помощь

Фильтр баз д Фильтр табл

Airflow База данных: spark Таблица: tasket1b Данные

Airflow

- Airflow
- information_s...
- mysql
- performance_...
- ✓ spark 48,0 KiB
 - tasket1a 16,0 KiB
 - tasket1b 16,0 KiB
 - tasket1b 16,0 KiB
 - tasket1b 16,0 KiB
 - sys

spark.tasket1b: 5 строк (точно)

#	Код предмета	Код студента	Фамилия студента	Имя студента
1	P01	C01	Рогов	Василий
2	P01	C02	Бахмутов	Павел
3	P01	C03	Васильев	Лев
4	P02	C02	Бахмутов	Павел
5	P02	C03	Васильев	Лев

Фильтр баз д... Фильтр табли...

Airflow База данных: spark Таблица: tasket11a Данные Запрос*

✓ **Airflow**

- > Airflow
- > information_s...
- > mysql
- > performance_...
- ✓ **spark** 48,0 KiB
 - tasket11a** 16,0 KiB
 - tasket11b 16,0 KiB
 - tasket14b 16,0 KiB
- > sys

spark.tasket11a: 2 строк (точно)

#	Код предмета	Предмет	Учитель
1	П01	Проектирование БД	Моисеев
2	П02	Машинное обучение	Щербань