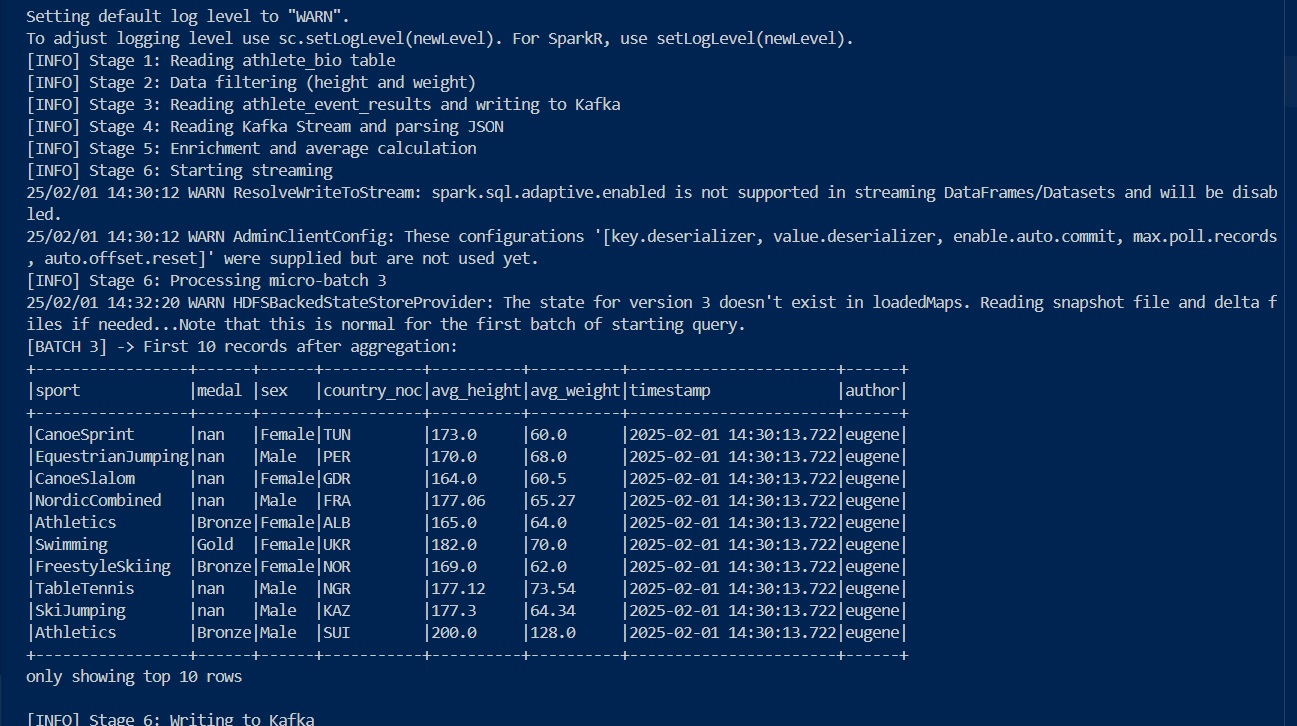
**Фінальний проєкт**

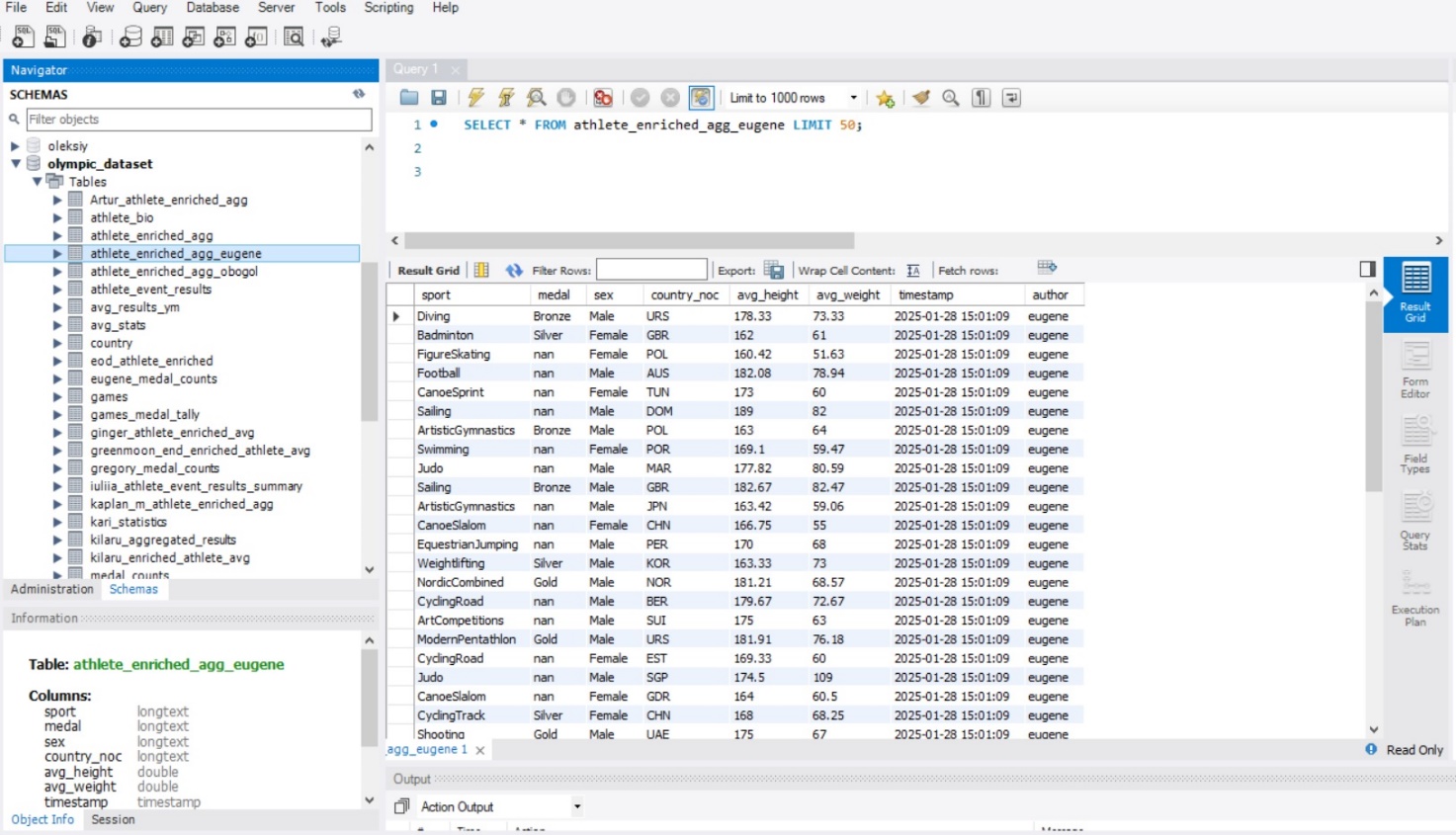
**Частина 1. Building an End-to-End Streaming Pipeline**



***Скриншот 1\_1.*** Фінальний DataFrame перед записом у MySQL та Kafka і виведення перших 10 рядків з агрегованими значеннями перед збереженням

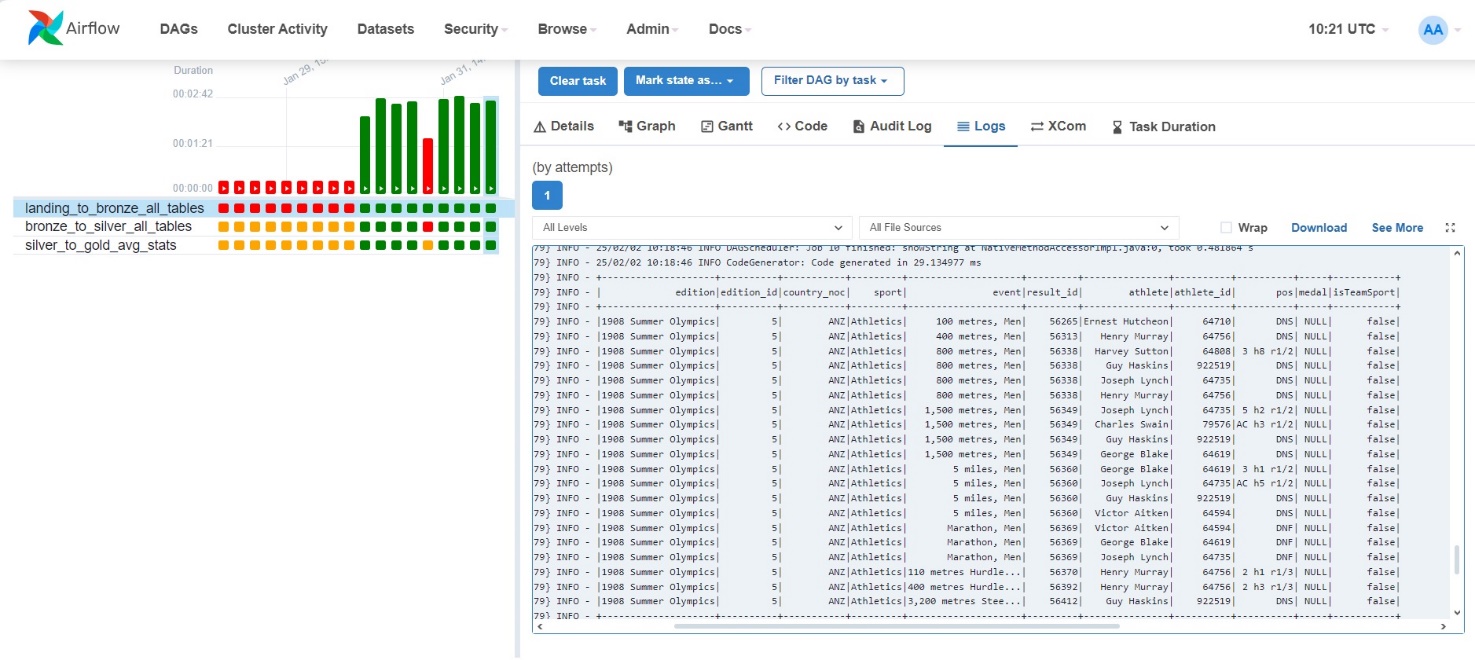


***Скриншот 1\_2.*** Лаги спарк підтверджують, що дані записалися у MySQL

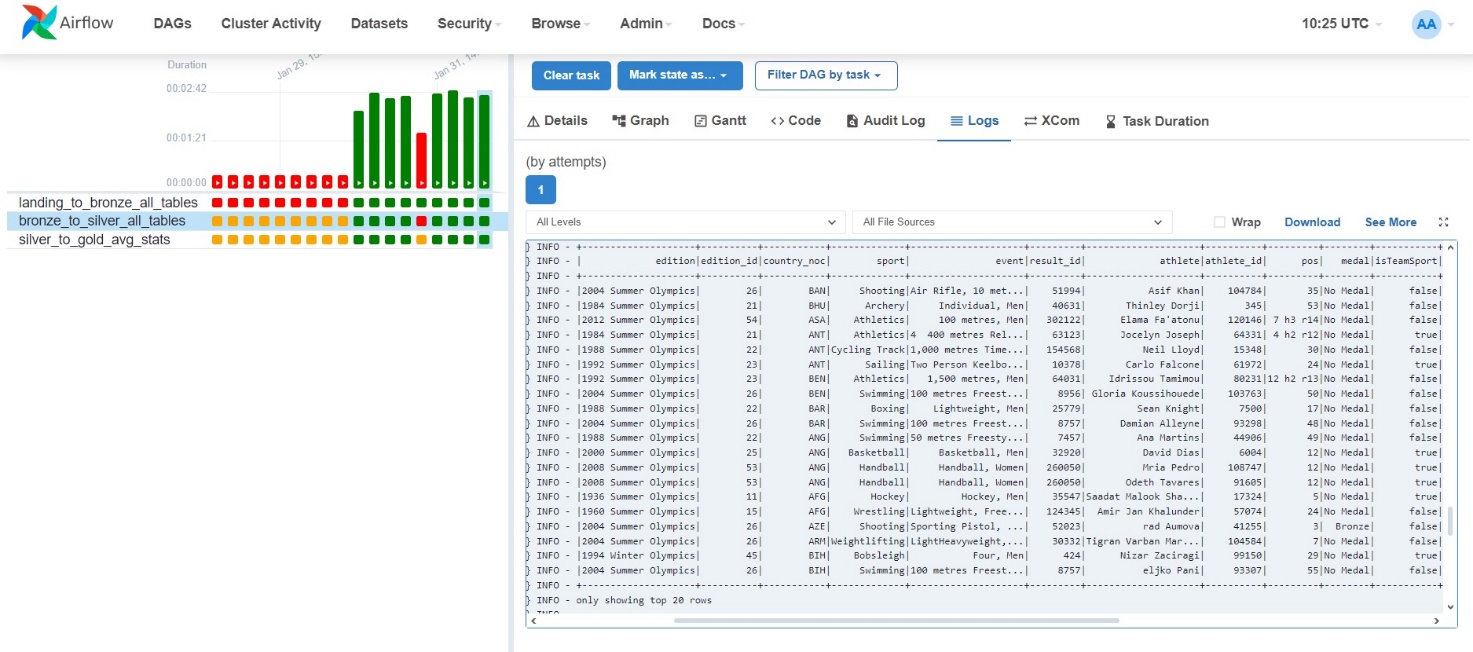


***Скриншот 1\_3.*** Результат збереження даних у у MySQL (таблиця athlete\_enriched\_agg\_eugene)

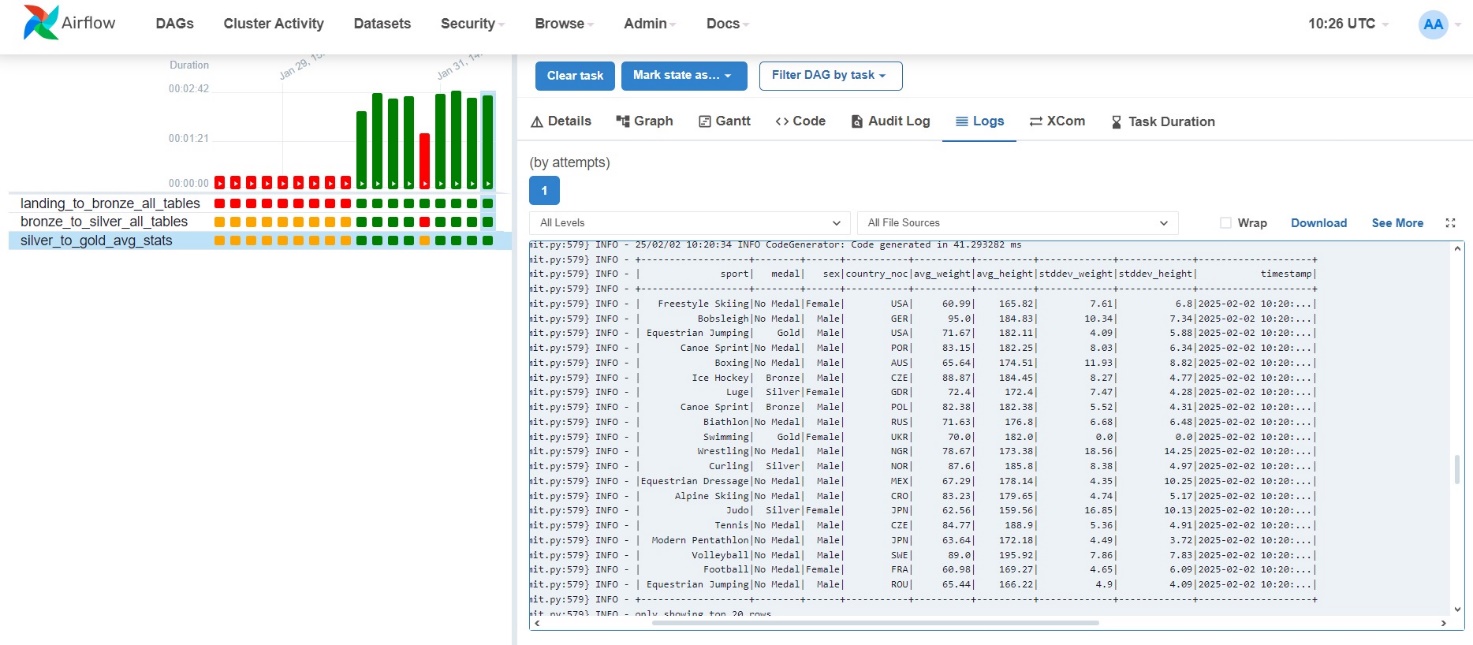
**Частина 2. Building an End-to-End Batch Data Lake**



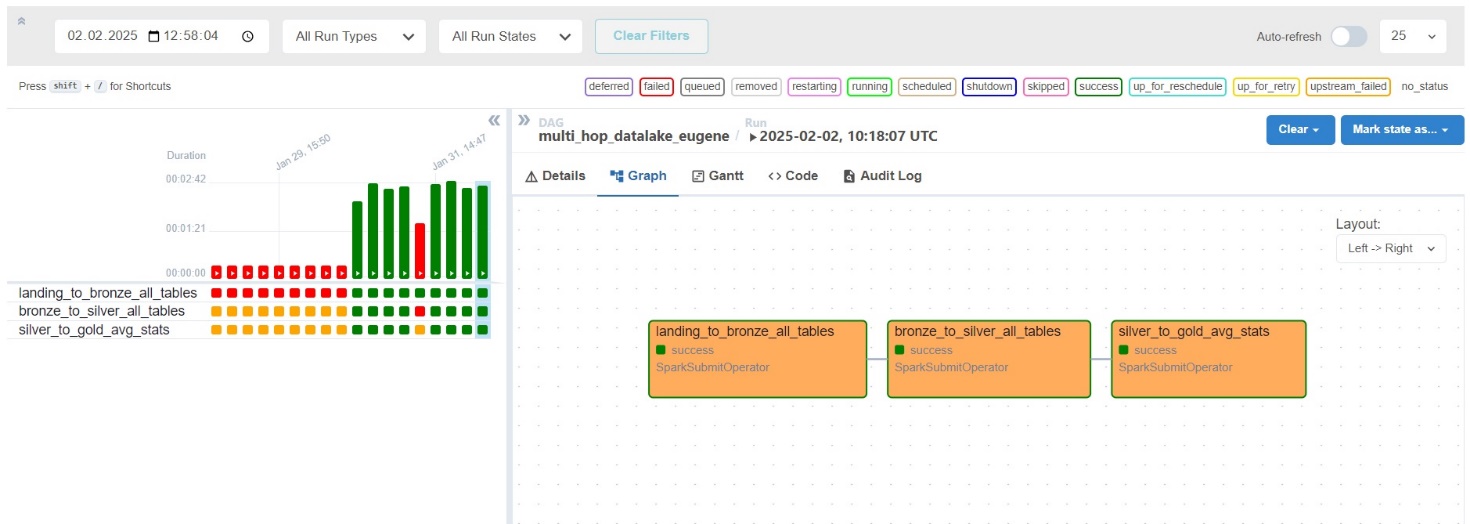
***Скриншот 2\_1.*** Логи першого етапу (landing\_to\_bronze\_all\_tables), де виконувалося завантаження даних з FTP-сервера та їх перетворення у формат parquet для зони bronze



***Скриншот 2\_2.*** Процес очищення даних (bronze\_to\_silver\_all\_tables), включаючи логування та роботу зі зчитаними parquet-файлами, очищення тексту та видалення дублікатів перед збереженням у silver



***Скриншот 2\_3.*** Третій етап (silver\_to\_gold\_avg\_stats), де дані з обох таблиць silver агрегуються для створення фінальної gold-зони зі статистикою атлетів



***Скриншот 2\_4.*** Граф виконання DAG-у, де всі три основні завдання успішно виконалися. Він ілюструє послідовність обробки даних: від завантаження даних у bronze-зону до фінальної агрегації в gold-зоні.