Technical Report: Store

To successfully store genomic data from CSV files into a MySQL database hosted on Clever Cloud, a detailed process was followed. This included configuring a Python development environment with necessary libraries such as Pandas for data manipulation, SQLAlchemy for database interaction, and PyMySQL as a database connector.

A MySQL database instance was set up on Clever Cloud. The database credentials, including host, port, user, and password, were securely configured.

A Python script was written to automate the process of reading data from the CSV files and storing it into the MySQL database. Using SQLAlchemy, a connection engine was created to interface with the MySQL database. The script utilized Pandas’ to\_sql method to seamlessly transfer data from DataFrame objects directly into the MySQL database, handling the conversion of data types and insertion process.

During the data upload, several challenges were encountered, particularly with database connections and permission settings. Detailed error logs provided by SQLAlchemy and PyMySQL were crucial in diagnosing issues such as access denials and configuration errors. Adjustments were made to the database user permissions, and connection strings were corrected, which resolved the issues and ensured a secure and stable data transfer.

After the data upload, verification processes were undertaken to ensure that the data was accurately uploaded. Queries were executed using tools like phpMyAdmin provided by Clever Cloud to visualize and check the data.