



#### Вывод :

DuckDB выигрывает по скорости выполнения запросов, благодаря хранилищу на основе столбцов и векторизованному выполнению. В целом pandas превосходит Postgres по времени выполнения запросов. Выбор столбцов в pandas эффективен с затратой  $O(1)$  времени, поскольку фрейм данных уже сохранен в памяти. Pandas является мощным инструментом для изученных задач анализа данных. Однако pandas имеет свои ограничения, и по-прежнему существует потребность в SQL. В pandas данные хранятся в памяти, и будет сложно загрузить CSV-файл, занимающий более половины памяти системы. Postgres и другие языки, основанные на SQL, были созданы для управления базами данных и предоставления пользователям удобного способа доступа к данным и их извлечения, особенно из нескольких таблиц. Pandas был создан для манипулирования данными, и его сильная сторона заключается в сложных операциях по анализу данных. В свою очередь SQLite продемонстрировала стабильную самую низкую производительность среди других 3 библиотек. Возможно, это связано с тем, что SQLite передаёт данные операционной системе, а операционная система обрабатывает эти данные. В это время SQLite ожидает и создает свою резервную копию данных в журнале памяти SQLite в оперативной памяти, и только после того, как данные будут отражены в базе на жёстком диске, продолжает выполнение операций.

PostgreSQL является масштабируемой системой управления базами данных с поддержкой формата JSON, обычно используется для обработки больших объемов информации. Однако, его производительность может снижаться в процессе выполнения пакетных операций и чтения данных.

SQLite представляет собой встроенную библиотеку, которую приложения используют непосредственно в своей основе. Это самодостаточная файловая БД, которая предоставляет инструменты для более простой обработки любых типов данных в сравнении с серверными БД. Когда приложение использует SQLite, оно взаимодействует с данными, содержащимися в файлах, напрямую, без интерфейсов, что повышает скорость и эффективность операций.

Однако при выполнении сложного анализа SQLite может работать не так быстро. Когда приложение использует SQLite, их связь производится с помощью функциональных и прямых вызовов файлов, содержащих данные (например, баз данных SQLite), а не какого-либо интерфейса, что повышает скорость и производительность операций. Так как база данных хранится в одном файле, это облегчает её перемещение. Несмотря на то, что SQLite является встроенной базой данных, при выполнении исчерпывающего анализа работает слишком медленно.

DuckDB - безсерверная система управления аналитической базой данных, которая обеспечивает быструю обработку запросов благодаря векторизованному выполнению запросов и ориентации на столбцы. Это делает DuckDB быстрой и простой в использовании. Также DuckDB не имеет внешних зависимостей и серверного программного обеспечения, что обеспечивает высокоскоростную передачу данных в базу данных и из нее. Однако она может иметь ограничения в поддержке SQL-функционала.

Pandas - одна из самых популярных библиотек Python, которая обеспечивает продвинутые инструменты для работы с данными. Однако она работает медленно на больших наборах данных из-за ограничений в поддержке многопоточности. Pandas использует только одно ядро процессора для выполнения задач по манипулированию данными, что также влияет на скорость ее работы.