Выбрал данную модель, потому что она схожа с моделью данных Инмана. В целом, так как процесс идет прозрачно, и он легок в понимании я выбрал его. Саму систему изобразил в файле “система.png”

В моменте, я забираю данные с источника и через питоновский скрипт трансформирую, преобразовываю и выгружаю в двх. В процессе, я забираю инкрементально данные с бд (4 таблицы) и уже после преобразовываю и также проверяю на дубли и очищаю/ обновляю данные с помощью библиотек Python. После выгружаю их в DWH, где в качестве примера используется Postgresql. Данные которые появились в DWH, в дальнейшем используются в витринах, для анализа и использования другими подразделениями.

В виде плюсов, могу сказать что в данном случае будет единая двх на которую можно будет ссылаться как основа и источник. Это будет плюсом так как в данном случае, все будет едино и не будет разветвлении по источникам данных. А дальнейшие анализы/исследования будут использоваться в отдельно создаваемых витринах.

Также хотел отметить что для регламентной выгрузки данных используется библиотека и система оркестрации данных Prefect, который очень прост в понимании. Выгрузка данных будет осуществляться раз в день и ее выполнение можно будет также посмотреть через сервер.

Помимо этого, изобразил взаимосвязь между таблицами в файле “physical\_final.png”. Там можно увидеть связь таблиц друг с другом и также типы данных, которые используются у меня в таблице (так как в процессе, некоторые из них я поменял для эффективности).

В случае вопросов можете написать, с радостью отвечу на них)