Take Home Interview Questions

Deliverables

Since these are candidate dependent systems, feel free to spin up any developer/free edition environment as connectors for development, there is no need to provide the credential.

- Solution design documentation, including considerations and limitations
- Solution metadata and codebase, preferable links to a github repository
- Any assumptions can be included, as long as it's called out

Question

Design a solution to replicate data from System A to System B. (ie. System A could be Salesforce.com and System B could be Snowflake). Data volume is considered to be high data volume, daily traffic is assumed to be over 100k records in average.
  - Replicate an object/table called Case
  - Case object has a column/field called Country
  - Case.Country has 3 values - USA, CAN and AUS

Requirements

- A scheduled job daily, in batches
- Pulls everything from the last scheduled run from System A to System B that replicates the object by case.country, each country belongs to different tables in system B.

    - To illustrate the flows:
    - A.case.country.USA -> B.case.USA
    - A.case.country.CAN -> B.case.CAN
    - A.case.country.AUS -> B.case.AUS

- (**Bonus**) Data replication in near real-time