TransferRoom

TransferRoom – Data Engineer Assignment

Version 1.0

Goal

1. Verify the candidate's ability to take high-level requirements and deliver a coherent architectural proposal
2. Verify the candidate's ability to articulate a clear rationale for architectural decisions
3. Verify the candidate's ability to propose appropriate tools for solving complex data problems

Deliverables

1. Candidates should deliver the solution to the assignment
    a. Detailed documentation, including a detailed architecture diagram
2. Public GitHub repository with any associated diagrams, comments, documentation and reference material related to the assignment for all parts of the assignment attempted.

Notes

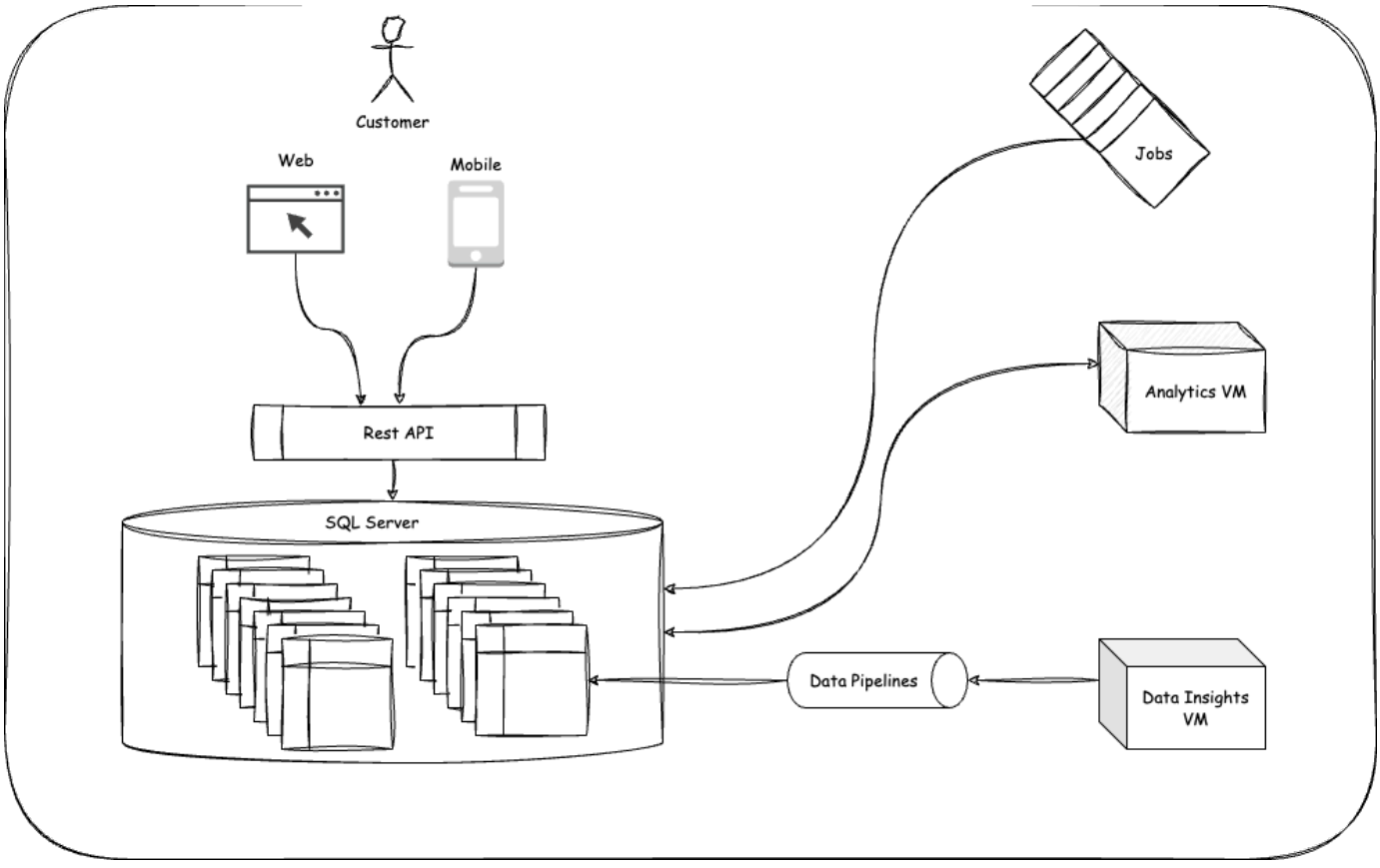
1. Provide as much detail as you feel necessary, including documentation and diagrams as you will be asked to present this back to us as part of the final stage of the process a. The assignment is open-book.
2. Please complete the assignment within 72 hours of receipt.

Objective

The objective of this exercise is to demonstrate the ability to review a high-level architectural problem and propose an improved architecture which satisfies additional new requirements with a detailed breakdown of the rationale for all elements of that new architecture proposal.

Assignment

1. Consider the following high-level data architecture diagram.



2. Customers leverage the platform via conventional native web and mobile experiences.
3. In addition, large and expensive OLAP and OLTP processes run on a continuous basis as well as a large number of scheduled processes and jobs.
4. All these processes connect to the primary production SQL Server database.

Problem Statement:

ShellCorp is a B2B business whose existing platform and architecture suffer from poor operational performance during periods of high traffic including SQL deadlocks and poor read and write performance.

There is an ever-expanding requirement to analyse, ingest and consume more data from analysts, data scientists and customers.

Questions and Tasks:

1. Propose a future-proof data architecture with tooling/vendor choices that deliver a scalable solution for internal and external customers.
2. Please provide additional detail regarding the data ingestion process, the tooling proposals and a sample implementation of how that data ingestion might work.
3. Please provide as much detail as possible regarding your rationale for any architectural technology choices you make.