

Case Study - Data of Fire v1.1

Thank you for your interest in *TrueNorth's Data Engineer* opportunity. We are excited to present you with a case study that will not only entertain you but also put your skills to the test. In this case study, we will dive into the fascinating world of data engineering, where you will have the opportunity to showcase your knowledge and problem-solving abilities.

At *TrueNorth*, we believe in providing challenging opportunities that push the boundaries of what is possible in the field of data engineering. This case study has been carefully crafted to simulate real-world scenarios and assess your capabilities in designing robust and scalable data solutions.

We encourage you to approach this case study with enthusiasm and creativity. As you tackle the challenges presented, demonstrate your expertise in building data pipelines, optimizing data storage and retrieval, and leveraging cutting-edge technologies to extract meaningful insights from large datasets.

So, without further ado, let's delve into the case study. We hope you find it both entertaining and challenging. Best of luck!

/ TrueNorth's Engineering Team

/ Description

A customer needs to analyze a [dataset](#) of fire incidents in the city of San Francisco. To do so, it requests you to make the data available in a data warehouse and create a model to run dynamic queries efficiently.

Requirements

The copy of the dataset in the data warehouse should reflect exactly the current state of the data at the source.

For the sake of this exercise, assume that the dataset is updated daily at the source.

The business intelligence team needs to run queries that aggregate these incidents along the following dimensions: time period, district, and battalion.

Deliverable

Present your work in a Git repository. You may use whatever platform you prefer (e.g. GitHub, Bitbucket, GitLab). If you use a public repository, don't mention TrueNorth anywhere.

Include at least one report that shows how to use your model.

Write a brief description of the provided solution, including your assumptions, thought process, and usage instructions. You may include this in a markdown file in the repository (e.g. [README.md](#)).

Use any technology you see fit to solve the problem.

Suggestions

If you are familiarized with Docker you could use it to spin up a database server locally since we're not providing a real data warehouse.

If you have worked with [dbt](#), you can use it to orchestrate the SQL scripts.