

# AlphaSense FinData Technical Interview Challenge

## Introduction

Before your technical challenge begins, we will send a PDF outlining the details of what is expected. You will have 90 minutes to write code, in any language of your choosing that can answer the questions that will be provided to you; you are not expected to get through all of them. You will need to fetch a handful of zipped CSV files from a public AWS S3 bucket that will be specified to you. The files are to be downloaded, parsed, have the data transformed and have output generated in some way. The content of the files will loosely resemble financial data in a very simplified format. We're concerned with logic and data structures more than a user interface of any kind, so use your discretion. Use of appropriate third-party libraries is welcome.

It is recommended to set up a personal development environment where you can comfortably write and run code in advance of receiving the questions. At the end of the 90 minutes, you will be expected to send your entire code solution over email or a public git repository of your choosing. In the live discussion portion of the interview, you will be expected to share your screen and provide a walkthrough of your code and demonstrate that it can execute and produce an output.

In the live discussion, interviewers may ask you hypothetical questions about designing, scaling, and integrating components like your solution into a broader system. Use of any whiteboarding tool that you may be familiar with while screen-sharing is encouraged.

## S3 Directory Structure

The data can be accessed from the bucket [as-findata-tech-challenge](#). This is in the [us-west-2](#) region. There are five zipped CSV files contained within the [company-data/](#) folder. An example of the full S3 key for one of the files is [company-data/CT4OAR0154.zip](#).

You are expected to write code to download the 5 files in [company-data/](#) and perform operations on the data specified in the questions below. [If you are having trouble connecting to AWS and downloading the files, you can request the interviewers to send the files to you directly \(over email\) – use this at your discretion if you feel blocked.](#) One sample file is provided to you at the start of the challenge ([sample-MNZIRS0108.csv](#)).

# The Data

The CSV files contain the data you will be parsing in order to complete the questions. Each row represents a specific ID's value over time. Each date cell in the header row represents a date range, where the value indicates the start date of a 3-month period. Each row's "scale" value indicates the factor that was applied to get the scaled value that is stored in each subsequent cell in the row; for example, a scale of 10 and a stored value of 10 indicates an actual value of 100.

## Questions

1. Return the value of the entry in row "MO\_BS\_INV" and with a start date of 2014-10-01 in file "MNZIRS0108.csv".
2. Return the mean of row "MO\_BS\_AP" in file "Y1HZ7B0146.csv".
3. Return the value of the entry in row "MO\_BS\_Intangibles" where the column's date range includes the date 2015-09-30 in file "U07N2S0124.csv".
4. Return a single number representing the mean of row "MO\_BS\_AR" over all input files.
5. Return a single number representing the mean of row "MO\_BS\_NCI" over all input files.
6. Return the coordinates and values of all entries for row "MO\_BS\_Goodwill" in any file where the value is greater than 20,000,000,000.
7. Return the IDs of rows where any of the values have increased by at least 50% from the first value over all time in file "Y8S4N80139.csv".
8. Return the IDs of rows where there is at least one value that differs by 20% or more from the mean of the row in file "CT4OAR0154.csv".