

Tech - Onsite Interview - Data Engineer/Data Scientist (fresh)

Purpose:

- To allow us to know how fast you can learn new concepts, technologies and how resourceful you are

Objectives:

Given 2 files, [blog.csv](#) and [comments.csv](#), they contain lists of blog posts and comments.

Your task is to:

1. Figure out the corresponding comments for each blog.
2. WRITE A PROGRAM in Java/Python to insert all the blog posts and their corresponding comments into MongoDB under collection `blogs` of database `interview`.

Each document must have the following fields:

```
id
title
author
content
comments : an array of respective comment documents
```

Each comment must have the following fields:

```
id
content
poster
```

3. Dump out the data in JSON format based on the results from #2, and submit both the code and data dump to us.

(see overleaf for sample dump)

Sample of a blog document:

```
{
  id: '1',
  title: "Hello",
  author: "Ella Chen",
  content: "Please comment me!",
  comments: [
    {
      id: '4',
      poster: "Peter",
      content: "Hi there"
    },
    {
      id: '5',
      poster: "Peter",
      content: "Hi, it's me again"
    }
  ]
}
```