

Technical Coding Challenge


SEEK is passionate about technical excellence, developing secure and scalable cloud-native solutions with clean, testable code while following the latest best practices. With a strong "DevOps and QA as a culture", each Engineer is responsible for code delivery, testing, documentation, package version management, etc.

With this in mind, consider how you would approach the coding challenge as a SEEKer. We are equally interested in seeing how you implement the solution; this is your opportunity to demonstrate your skills, experience, passion, and answers to the specific questions.

We ask that you complete the exercise using Apache Spark (written in vanilla pySpark or Python). Ensuring your submission can be compiled and run without any prerequisites and making no assumptions about the environment other than having access to docker.

While we recognise your time is precious, try to complete as much of the coding challenge as possible. Remember, we are equally interested in your approach to the problem domain emphasising clean, testable and documented code.

Please provide submissions through GitHub, with a clear commit history showing progress. If you need more clarification, please contact seekanalYTics@seek.com.au

 The test data required to complete the coding challenge can be found at <https://coding-challenge-public.s3.ap-southeast-2.amazonaws.com/test-data.zip>. To extract the contents, you will require the password `By9FNTZXP4j4izuufAs=`

Spark/SQL Questions

1. Please load the dataset into a Spark dataframe. You may want to look at the data using `jq` or a similar tool to get an idea of how the data is structured.
2. Print the schema
3. How many records are there in the dataset?
4. What is the average salary for each profile? Display the first 10 results, ordered by **lastName** in descending order.
5. What is the average salary across the whole dataset?
6. On average, what are the top 5 paying jobs? Bottom 5 paying jobs? If there is a tie, please order by **title**, **location**.
7. Who is currently making the most money? If there is a tie, please order in **lastName** descending, **fromDate** descending.
8. What was the most popular job **title** that started in 2019?
9. How many people are currently working?
10. For each person, list only their latest job. Display the first 10 results, ordered by **lastName** descending, **firstName** ascending order.
11. For each person, list their highest paying job along with their first name, last name, salary and the year they made this salary. Store the results in a dataframe, and then print out 10 results
12. Write out the last result (question 11) in `parquet` format, compressed, partitioned by the year of their highest paying job.

Reference Data:

```
{
  "id": "6bc63596-1ceb-485c-8277-42631b0c1aff",
  "profile": {
    "firstName": "Mirella",
    "lastName": "Breen",
    "jobHistory": [
      {
        "title": "sales representative",
        "location": "Perth",
        "salary": 67000,
        "fromDate": "2017-05-22",
        "toDate": "2019-02-22"
      },
      {
        "title": "corporate consultant",
        "location": "Perth",
        "salary": 58000,
        "fromDate": "2012-02-23",
```

```
    "toDate": "2017-04-23"
  },
  {
    "title": "financial counsellor",
    "location": "Perth",
    "salary": 53000,
    "fromDate": "2008-06-15",
    "toDate": "2012-02-15"
  },
  {
    "title": "project manager",
    "location": "Perth",
    "salary": 53000,
    "fromDate": "2005-02-12",
    "toDate": "2008-06-12"
  },
  {
    "title": "corporate consultant",
    "location": "Perth",
    "salary": 52000,
    "fromDate": "2000-07-25",
    "toDate": "2005-01-25"
  },
  {
    "title": "Sheetmetal Worker",
    "location": "Perth",
    "salary": 44000,
    "fromDate": "1994-10-21",
    "toDate": "2000-07-21"
  },
  {
    "title": "customer service officer",
    "location": "Perth",
    "salary": 42000,
    "fromDate": "1990-08-11",
    "toDate": "1994-10-11"
  },
  {
    "title": "financial counsellor",
    "location": "Perth",
    "salary": 38000,
    "fromDate": "1989-12-08",
    "toDate": "1990-08-08"
  }
]
}
```