**Data Source**

We have a table in our “person” schema called “job” that describes where 700 million people have worked across 20 million companies in over 1 billion roles. Each row represents one role a person held at a company.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Person ID** | **Person Name** | **Job Title** | **Company ID** | **Start Date** | **End Date** |
| Asdf5kjasdkfa2sd | Eric Aleman | Direct of Product | peopledatalabs | 2021-01-18 | NULL |
| d908u23nf9239g | Ben Eisenberg | Direct of Product | peopledatalabs | 2017-06-06 | NULL |
| 3jkbgd983jskksd | Varun Villait | VP of Product | peopledatalabs | 2020-09-15 | 2021-08-01 |
| 3jkbgd983jskksd | Varun Villait | Chief Product Officer | peopledatalabs | 2021-08-01 | NULL |
| kj2n3r90vur2oj34 | Victoria Boyd | Chief of Staff | peopledatalabs | 2022-01-18 | NULL |
| … | … | … | … | … | … |

You can make the following assumptions:

* People can work in multiple roles at one company (e.g. if someone gets a promotion), and this is represented by multiple rows in the table, and there can also be overlap in date ranges for a single person in two roles at the company
* Start date is never null
* End date is null if and only if the person is still in that role
* Every person has a unique ID in our dataset (the “person ID”), which is reused across any job record they have in this table

**Query**

Write a SQL query that shows the current headcount of each company every month since the beginning of 2000 assuming you have the table above. If possible, use PostgreSQL or Redshift syntax, but any SQL syntax is fine.

|  |  |  |
| --- | --- | --- |
| **Company ID** | **Month** | **Headcount** |
| google | 2000-01-01 | 359 |
| … | … | … |
| peopledatalabs | 2020-01-01 | 32 |
| peopledatalabs | 2020-02-01 | 35 |
| … | … | … |

**Query Problems**

* Suppose your query is timing out. What might be causing this?
* What are edge cases that might cause slight inaccuracies in your query?