

US Bureau of Labor Statistics

I was hired as a data engineer by the US Bureau of Labor Statistics to design a database to store information about the US workforce. This database should help them analyze and monitor labor market dynamics to provide useful information to policy makers.

Using a database to store such information is tricky because not only does it have to store data for the entire United States, but it must be updated weekly, and the data must be preserved for decades for historical analysis. Very quickly the problem of memory capacity can arise, which can create server congestion and slow down the execution of requests. switching to Big Data could be an effective way to solve the memory problem and help write optimal queries. In this case, the volume of data is the main reason to use Big Data.

The schema diagram shows how relationship between table.

