Data & Discovery - Database Engineer intern

Groupon’s mission is to become the daily habit in local commerce and fulfill our purpose of building strong communities through thriving small businesses by connecting people to a vibrant, global marketplace for local services, experiences and goods. In the process, we’re positively impacting the lives of millions of customers and merchants globally. Even with thousands of employees spread across multiple continents, we still maintain a culture that inspires innovation, rewards risk-taking and celebrates success. If you want to take more ownership of your career, then you're ready to be part of Groupon.

We might look like an ordinary ecommerce app, but under the surface we’re a cutting-edge technology company building products that positively impact the lives of millions of people and merchants globally. We have big goals to get to 2020 and beyond, and need you to help us get there.

Responsibilities:

- Takes care of the day-to-day running of Hadoop clusters
- Work closely with the database team, network team, BI team and application teams to make sure that all the big data applications are highly available and performing as expected.
- Responsible for capacity planning and estimating the requirements for lowering or increasing the capacity of the Hadoop cluster.
- Ensure that the Hadoop cluster is up and running all the time
- Monitoring cluster connectivity and performance
- Work collaboratively with the database operations team and the vendor to resolve incidents and mitigate risks
- Working on the strategy for the cloud migration

Requirements:

- Experience with shell scripting / UNIX / Linux, Terraform
- Cloud experience (GCP is a plus)
- Basic scripting experience in languages such as python
- Experience in building and managing Big data onprem and Cloud Clusters for Hadoop, EMR, Airflow
- SQL, RDBMS ( MySQL, Postgres etc.)

You’ll spend time on the following:

- Deploying a Hadoop cluster, maintaining a Hadoop cluster, adding and removing nodes using cluster monitoring tools like Ganglia Nagios or Cloudera Manager, configuring the NameNode high availability and keeping track of all the running hadoop jobs.
- Implementing, managing and administering the overall Hadoop infrastructure
- Automation on monitoring tasks, day to day to operational tasks
• Actively working on the cloud migration tasks