# Data Engineer - JD

**Job Description :**

**This is what you'll do :**
- The position grows our analytics capabilities with faster, more reliable tools, handling petabytes of data every day.
- Brainstorm and create new platforms that can help in our quest to make available to cluster users in all shapes and forms, with low latency and horizontal scalability.
- Make changes to our diagnosing any problems across the entire technical stack.
- Design and develop a real-time events pipeline for Data ingestion for real-time dash- boarding.
- Develop complex and efficient functions to transform raw data sources into powerful, reliable components of our data lake.
- Design & implement new components and various emerging technologies in Hadoop Eco- System, and successful execution of various projects.

**Skills that will help you succeed in this role :**
- Strong hands-on experience of 4+years with Spark, preferably PySpark etc.
- Excellent programming/debugging skills in Python.
- Experience with any scripting language such as Python, Bash etc.
- Good experience in Databases such as SQL, MongoDB etc
- Good to have experience with AWS and cloud technologies such as S3