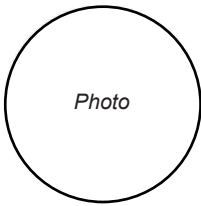# Name

Big Data Engineer

Photo

📱 +91 1212121212

✉️ xyz@gmail.com

📍 Hyderabad, IN

SUMMARY

4+ years of total experience in Software Development. 1+ years of experience as a Big Data Engineer. Experienced in developing, constructing, testing and maintaining big data projects majorly using Hadoop ecosystem and Apache Spark. Good Expertise in understanding Data Modeling, ETL development.

TECHNICAL SKILLS

- **Big Data Ecosystem:** HDFS, Hive, MapReduce, Spark, SQOOP, Zookeeper, Impala, Kafka (Intermediate)
- **Languages:** Java, C#, Scala(Intermediate)
- **Database:** MySQL, Oracle, SQLite, MariaDB, HBase, MongoDB
- **OS:** Windows, Unix
- **Scripting Language:** Linux
- **Tools:** Eclipse, Visual Studio, Tortoise SVN, Team Foundation Server (TFS), BitBucket, Git, Jenikins, Jira, Maven, SQL Server, Oracle SQL Developer.

KEY SKILLS

Product Design & Development
Root-cause Analysis
Research, Reporting & Documentation
Leadership & Team Management

PROFESSIONAL EXPERIENCE

Big Data Engineer

📅 Oct '19 - Present

DBS Asia Hub 2 Pvt Ltd.
Hyderabad, IN

DBS is a leading financial services group in Asia that has been at the forefront of creating and delivering best-in-class solutions for their customers. DBS Asia Hub 2 Pvt Ltd is the bank's first technology development centre outside Singapore.

Summary
- **Involved in developing a data pipeline** for processing batch data comprising terabytes of data using **Hadoop and Spark framework.**
- **Designed CI/CD pipeline** for our application by practicing DevOps methodologies.
- **Involved in writing automation scripts** that mitigate issues in Production.

Responsibilities
- Good working knowledge with different modules of the **Hadoop ecosystem (HDFS, Hive, Impala, Spark, SQOOP)**
- Expertise in **Mainframe technologies.**
- **Created Jobs/Schedules for Batch data processing** using IBM Workload Scheduler (TWS).
- **End-to-End ownership** of the project and looked into **development, sustenance, maintenance** of the application.
- Communicate with Business users and **translate the business requirements into working code**.
- Worked on huge **data processing running on hundreds of nodes** in the cluster.
- Responsible for the **CICD process** using Jenkins, Jira, Bitbucket, SonarQube, Maven.

Performance Optimization
- Worked on code migration from **Spark 1.6v to 2.3v**.
- Managed Hive tables from using **Avro to Parquet format**.
- Worked on **tuning improvement in Spark code & its configuration** which reduced our daily batch by half an hour.
- Involved in developing Dashboard using ELK stack, to ease L1 support batch monitoring task.

Key Achievements
- **Addressed various automation tasks in UAT and Production** environments which mitigates cascading effect to upstream and downstream projects.
- **Point-of-Contact for our upstream and downstream** projects.
- Received the **Star Performer** Award for extraordinary contribution to the project.

Application Engineer

📅 Jul '16 - Sep '19

Oracle India Pvt. Ltd.
Hyderabad, IN

In Oracle, I worked for Primavera P6 Application Development.

Role & Responsibilities:
- Involved in enhancing Primavera P6 Application performance.
- Contributed to major enhancements for the product.
- Involved in Writing and Reviewing code fixes and Unit test cases.
- Held Responsibility for Build Management and monthly patch set releases.
- Experienced in designing database schema and hands-on PL/SQL.
- Point-of-Contact for QA Team and acted as a Mentor for New Hires.

CERTIFICATIONS

- Databricks Training Certifications
- Data Structures & Algorithms certification course conducted by NPTEL IIT-Madras.
- Cognizant Certified Student.

Profiles

Linkedin
https://xxx//
.

GitHub
https://xxx//

HackerRank
https://xxx//
.

EDUCATION

B.E. - Computer Science

📅 Oct '12 - Jun '16

Vasavi College of Engineering (Affiliated to Osmania University)

Hyderabad, IN

- Percentage: **87.33**

PGP in Big Data Engineering

📅 Feb '20 - Present

NIT Rourkela

Hyderabad, IN

Currently pursuing PG in Big Data Engineering from NIT Rourkela, which includes the following modules.

Distributed and Parallel Computing for Big Data
Big Data Models and Management
Data Warehousing and Big Data Pipeline
Real-time Big Data Processing
Data Governance in Big Data
Big Data Analytics and Intelligence