Sonali Pednekar

Software Engineer

✉ pednekar123sonali@gmail.com

📞 8108823018

📍 Pune, India

in linkedin.com/in/sonali-pednekar-21978b17b/

SKILLS

- Hadoop, Map Reduce
- HDFS, YARN
- Spark, Spark SQL
- SQL, Hive, HBase, Impala
- Linux, Shell Scripting
- Scala, Python
- Git, SVN
- Machine Learning

AWARDS

Shining Star Award
Recipient of the Shining Star Award three times for exceptional project performance and dedication.

LANGUAGES

- Marathi
- English
- Hindi

CERTIFICATES

- Presented research paper entitled "Analysis of Impact of Big Data on our Daily Lives" in International Conference
- Data Science Training - Internshala
- Certificate of Data Science Bootcamp

PROFILE

Software Engineer with 2.6+ years of hands-on experience specializing in ETL design, development, and implementation within the Hadoop ecosystem. Having good knowledge of Hadoop, hdfs, Spark, Scala, Hive, HBase, Impala. Proficient in data processing, modelling, analytics, and visualization to deliver proven results.

EDUCATION

Post Graduate Diploma - Big Data Analytics (PG-DBDA), Centre for Development of Advanced Computing
August 2019 – February 2020 | Pune, India

Post Graduate - Master of Computer Application (MCA), Mumbai University
August 2016 – July 2019 | Mumbai, India

PROFESSIONAL EXPERIENCE

Saama Technologies Pvt Ltd, Software Engineer
March 2021 – present | Pune, India

Project Name - Building Next Generation Analytics
Description - The objective of the engagement is to implement Batch and Real time application for Blue Dart which will provide actionable insights about the shipment on various levels such as first mile, last mile, delivery monitor.
Tools - Apache Spark, Apache Kafka, Hadoop, HDFS, Apache Hive, HBase, Impala, Elastic Search
Environment - Eclipse, Hue, Linux, WinSCP, Putty, Yarn, mysql workbench, git, SVN
Roles & Responsibilities -

- Spearheaded the design and development of data pipelines using Apache Spark, Scala, Hive, and HBase for Big Data ETL.
- Managed end-to-end pipeline for real-time and batch processing, achieving optimal results and Elastic Search indexing.
- Created and managed Hive tables, loading data and crafting queries for storage and analysis.
- Developed automated shell scripts to streamline Spark Scala code execution.
- Experience in designing, creating, testing, evaluating, and maintaining data management & processing systems in order to achieve optimization.
- Proficient in data analysis using HiveQL and Impala.
- Conducted performance tuning and optimization of Spark jobs, exploring structure streaming for efficiency.
- Ensured seamless workflow integration with crontab, performed monitoring, debugging, and issue resolution.
- Collaborated with troubleshooting teams to resolve data discrepancies, enhancing code base and peer reviewing changes.
- Documented comprehensive pipeline architecture, data flows, transformations, and technical aspects.

Project Name - Celgene
Description - Celgene is all about Drug Discovery and Development, and clinical trial data.
Tools - Apache Spark, HDFS, Apache Hive, Elastic Search

Environment - Eclipse, Hue, Linux, WinSCP, Putty, Yarn, mysql workbench
Roles & Responsibilities -
- Monitoring data load
- Responsible for identifying issues, and debugging/fixing those issues.

Codium TechLabs Pvt Ltd, Intern

December 2018 – June 2019 | Navi Mumbai, India

Project Name - Loan Covenant System
Description - Loan Covenant System is built to maintain the loan work between Aditya Birla Finance Ltd (ABFL) and their Employees. The purpose of this System is to capture high-level business requirements for providing a solution for the development and maintenance of post disbursement tracking of loans using software hosted on the cloud or private server
Tools - PHP, HTML, CSS
Roles & Responsibilities -
- Designed user interface and data flow for loan work tracking between ABFL and employees.
- Developed and test interoperability for the Loan Covenant System.

ACADEMIC PROJECTS

Global Terrorist Attack Analysis
- The objective of the project is to build various classifiers and predict various terrorism activities using a machine learning algorithm.
- Used python for data extraction and manipulation, Hadoop for storage, Hive for analysis and querying, and Tableau for data visualization.
- Achieved most accurate predictions by performing machine learning algorithms such as KNN classifier, decision tree, random forest, XG boost
- Tools : Python, Hadoop, Hive, Machine Learning, Tableau

PERSONAL INFORMATION

Date of Birth : 22nd April 1996

Marital Status : Single

Address : Block-B, 1402 Sara Metro Ville, Tathawade Rd, Punawale, Pimpri-Chinchwad, Maharashtra 411033