SHWETHA P

Contact Info:
Email :
shwethap.dscience@gmail.com
Ph no : 9844988641

I myself passionate ,software professional with 3 years of experience on AWS Bigdata Engineer. A versatile and innovative in business awareness by being involved in product life cycle from development to deployment in an Agile Environment with well- versed Researched manner with feasible solutions. I am currently looking for BIG Data Developer job to hone my skills in the field of designing, development and analysis.

- Worked as Data Engineer to design, implement and Support different big data applications by using Apache Spark, Hadoop and AWS.
- Having 3+ years of experience in IT, experience in Apache Spark to implement Big Data Applications
- Involved in developing Big Data applications by using different frameworks like Hadoop, Hive, Sqoop, Spark, Flink, Phoenix, Cassandra, Kafka and Redshift
- Having extensive experience on SparkSQL, Spark streaming include tune-up of the Spark applications.
- Strong experience on AWS-EMR, Spark Installation, HDFS & Map-reduce Architecture. Along with that having a good knowledge on Spark, Scala and Hadoop distributions like Apache Hadoop, Cloudera.
- Strong experience in All Hadoop and Spark ecosystems include Hive, Pig, Shell Script, Sqoop, Nifi, Kafka, Oozie, Cassandra, Spark SQL, Spark Streaming and Flink.
- Having Primary knowledge & eager to implement Artificial Intelligence & IOT applications.

Educational Qualifications:

- Pursuing Phd Datascience in Christ Deemedto be University.

Degree	Board / University	College	Year of Passing
MCA	Visvesvaraya Technological University	Acharya Institute of Technology, Bangalore	**2017**
B.Sc (PMCs)	Bangalore University	GFGC, Chickaballapur	**2013**
PUC	Department of Pre-University Education	GJC, Gudibande	**2009**

Proffesional Experience :

Organization Name	Duration	Designation
Snipe IT Solution	2016-Sep – 2017 Jan	Intern
Snipe IT Solution	2017-Feb – 2020 Jan	Bigdata Developer
Empowersys (part time)	2017-Sep- 2018 Oct	Trainer for Bigdata Ecosystem

Technicial Skills :

Primary Skills	Secondary Skills	Languages
Spark: SparkSQL, Streaming	Kafka, Alluxio, Flink, Nifi, IOT & AI	Programming languages: Scala.
Hadoop: Hdfs, Yarn,Hive, Sqoop	Pig, Oozie ,	Database languages: Oracle, Mysql, Redshift
AWS: RDS, EC2, Redshift	S3, IAM	No SQL Databases: Cassandra, Hbase Phoenix
Analytics, Git Hub	R Studio	R Language, Git Repositeries.

Project Name	Digital Health Care	Team Size		7
Start Date	Jul 2017	End Date		Dec 2017
Project Description	Digital Health Care is a health care domain, Get the RDBMS data (Oracle/MySQL) and SFTP data, store in AWS-S3. Import and Export the data from OraOop (through Sqoop) and store in S3. Apply Transformation rules on the top of different datasets finally store in the desired output (CSV to JSON). Schedule the tasks in AWS-pipes, Scalup automatically based on data. Finally store the desired data in Dynamodb, Red Shift andS3 in desired format. Create the dashboards, reports using Splunk.			
Role & Contribution	- For injection data from Oracle, Mysql to S3 which can be queried using hive and spark SQL tables. - Worked on Sqoop jobs for ingesting data from MySql to AmazonS3 - Created hive external tables for querying the data - Use Spark Dataframe APIs to inject Oracle data to S3 and stored inRedshift. - Write a script to get RDBMS data to Redshift. - Process the datasets and apply different transformation rules on the topof different datasets. - Process the complex/nested Json and Csv data using Dataframe API. - Automatically scale-up the EMR Instances based on the data. - Apply Transformation rules on the top of DataFrames. - Run and schedule the Spark script in EMR Pipes. - Process Hive, csv, json, oracle data at a time (POC). - Validate and debug the script between source and destination. - Validate the source and final output data. - Debug & test the process is reaching Client's expectations ornot. - Query execution is trigger. Improve the process timing. - Based on new spark versions, applying different optimization transformation rules. - Debug the script to minimize the shuffling data.			
Technology &Tools	SparkSQL, Streaming, Sqoop, AWS (EMR,Elastic Search, S3, EMR, DynamoDB, Pipes, Redshift)			
Key Achievements	Successfully implemented and delivered to client.			

Project Name	HDFC Bank Project	Team Size	14
Start Date	Jan 2018	**End Date**	Jan 2020
Project Description	To process billions of the records, Abinitio takes a lot of time. That's why store Abinitio data in Hive using spark without any data change. Abinitio data type and all operations, data types must be same. Optimize the performance, schedule, automate this project using shell script and oozie.		
Role & Contribution	- Understanding the requirements for the project. - Based on Mapping document, import data from Abinitio, and store data in Hive based on business logic. - Schedule spark Job using Oozie and shell script. - Query performance optimization, Update data using spark. - Performance Analysis using spark web UI. - Write alternative code to optimize Spark performance.		
Technology&Tools	Spark 2.3, Cloudera, Oozie, Hive, Scala, Schell Script		
Key Achievements	Schedule Oozie job based on app generate data & Performance optimization.		

MAJOR RESEARCH PROJECT :

S.I no	Project Name	Technologies and tools	Description	Duration
1.	**Prediction of Stock Market Index Using Soft Computing & Bio-Inspired Computing**	**IDE** : MATLab **Algorithms** : ANN(Artificial Neural Network) and Bio Inspired computing algorithms.	The past ten years data of the companies listed under NSE are used for building, training and testing the prediction model using soft computing and biologically inspired computing. The goal of this project is to determine how accurate the predicting model is predicting data.	9 Months

Intrenship Projects :

S.I no	Project Name	Technologies and tools	Description	Duration , Roles and Responsibilities
1	**Social Networking Analysis using Hadoop and Apache Flume**	**Platform :** Ubuntu **Technology :** Big Data **Framework: :** Hadoop (Hadoop Ecosystem) **Tools:** Apache Flume, Hive, Pig, Scoop, R tool, R Studio, Putty.	Sentimental Analysis and opinion mining on big data using twitter data and lexicographic dictionaries.	6 Months
2	**Digital Health Care**	**Platform :** Windows8x **Technology:** DB and Java Frameworks. **Tools :** Postgres SQL, MySQL workbench, JustIn mind Prototyper, SQL server, GitHub	It is an application which is like **Practo** helps to the patients for booking the appointments with doctor. We also included Homecare, Ayurveda etc., and this app is only for the health Care purpose.	6 months **Roles and Responsibility :** Freezing the requirements in agile methodology by Scrum meetings and also integrated with Hibernate using tools like Postgres SQL and MySQLWorkbench.
3.	**Billing System**	**Platform:** Windows 8x **Technology:** DB and Java Frameworks. **Tools :** Postgres SQL, MySQL workbench, JustIn mind Prototyper, SQL server, GitHub	It is an application for hotel management for billing purpose and also push the data to the cloud to overcome the income tax.	2 Months **Roles and Responsibility :** Freezing the requirements in agile methodology by Scrum meetings and also integrated with Hibernate using tools like Postgres SQL and MySQL Workbench.

ACADEMIC PROJECTS :

S.I no	Project Name	Technologies and tools	Description	Duration
1.	**Online Publisher System**	**Languages**: php, Html, css **IDE** : Dream weaver **DB tool** : Xamp server with MySQL queries	Repository can be maintained in the website, the researchers can use it and also they can publish their research articles.	3 Months
2.	**Swacha Bharath Abhiyaan with Complaint Registration and Feedback System**	**Languages** : C#.Net **DB tool** : Microsoft SQL Server	The goal is to reach both village and city people keep the India clean. The various departments of Karnataka Govt. are included for citizen complaints and feedback system.	3 Months

Paper Publications :
- Presented and published a paper on "Predictive Analysis of the Twitter Data Using the Big data Technology with Hadoop and Apache Flume" in NATIONALCONFERENCE on Reaching the Unreached through Science and Technology organized by Surana College Bengaluru.
- Presented a Paper entitled "Case Study on Global Positioning System (GPS) Antenna" in (National Conference on VLSI, Signal Processing, Communication and Soft Computing (NCVCS-14)", organized by Dept. of ECE, SKIT, Bangalore.
- Presented a Paper entitled "Citizen Complaint Register" in National Conference on Advances in Modern Computing and Application Trends (TECHMAN-14), organized by Dept. of MCA, AIT, Bangalore.

- Published a Scopus Indexed journal in "**INTERNATIONAL JOURNAL OF ADVANCED SCIENCE AND TECHNOLOGY**" paper entitled "An efficient and holistic approach to reduce output and dependent parameters for multi–output Learning".

Awards and Achievements

- Won "BEST RESEARCH PAPER AWARD" in National Conference entitled "Predictive Analysis of the Twitter Data Using the Big data Technology with Hadoop and Apache Flume".
- "Social Networking Analysis using Hadoop and Apache Flume" Selected as Second Academic best project in the project exhibition -2016 at AIT.
- Awarded as Active member of AIT Research Forum.
- Won second prize in Poster Presentation on Climatic Changes in National Science Day - 2011 organized by M.S.Ramaiah College of Arts, Science, and Commerce.
- Won second prize in IT Seminar on "ARE YOU A SHADOW OF IT" in inter College annual talent meet, VRISHINI-2009. Organized by SBMJC.

Certifications

- Certification on Enablement on bluemix.net organized by IBM

PERSONAL DETAILS

Name   : SHWETHA P

Father Name   : Prasad R K

Marital status   : Married

Nationality   : Indian

Languages Known   : Kannada, Telugu, English, Hindi.

Permanent Address   : #No 15, Nagashree, 2nd cross Vinayaka
                                      Layout,Uttarahalli,Bengaluru, Karnataka - 560061

DECLARATION

I hereby declare that the above information is true to the best of my knowledge and belief.

DATE:

PLACE: Bangalore   (SHWETHA P)