

PySpark Developer

Address:

23, DR Gents PG, MGR Street, Sholinganallur
Chennai-600119

Chhavi Sharan Bansal

Email : chhavi677@gmail.com
Mobile: +91 9340453437

Looking towards employer as a boost to my career as Hadoop Developer wherein I can apply my knowledge in **PySpark Development** for the mutual growth of the company and me. An IT professional who is having the **Bachelor of engineering degree** in **Computer science** in first class with **honors**.

Professional Summary

- ✚ Have total 4.7 years of experience in IT industry, previously employed with **HCL Technologies Ltd. from last 3.9 years and Currently employed with Capgemini India Ltd.**
- ✚ Have exposure of working in **Logistics and Banking Domains**.
- ✚ Experience in **ingestion, storage, querying, processing and analysis of Big Data** with experience in **Hadoop Ecosystem including HDFS, HIVE, PIG, SQOOP and PYSPARK**.
- ✚ Hands on **AWS Cloud Data Platform services** like: - Glue, Lambda, Dynamo DB, Athena, S3, Step Function, CloudWatch.
- ✚ Have exposure on software development methodologies like **Agile, SDLC**.
- ✚ Have exposure to **HBase** shell commands.
- ✚ Good understanding/knowledge of Hadoop Architecture and its components such as **HDFS, Node Manager, Resource Manager, Name Node, Data Node, Secondary Name node, and MapReduce concepts**.
- ✚ Working in **various format (Json, Avro, CSV, Parquet, ORC) by Data Frame**.
- ✚ **Processing and Analyzing the** Historic data By Spark.
- ✚ Load and save data from different data source in **Spark data frames**.
- ✚ **Store the data in hive table through Spark**.
- ✚ Experience in importing data using **Sqoop from Relational Database Systems to HDFS**.
- ✚ **Good Programming and Communication Skills with excellent work ethics**, easy adaptability to demanding time schedules coupled with positive user interaction and excellent team spirit.
- ✚ Exceptional ability to quickly master new concepts.

Key Skills

Big Data Technologies	: HDFS, Pig, Hive, Sqoop ,Flume, PySpark.
AWS Technologies	: S3, Glue, Athena, CloudWatch, Step Function, Lambda.
Programing Language	: Python, Shell Script, object oriented concepts
SQL Databases	: Oracle 11g
NO SQL Databases	: HBase
Scripting language	: UNIX and Shell scripting
Operating Systems	: AIX, Sun Solaris, Windows 10X
Tools	: Cloudera 5, AWS, WINSCP, GIT Hub, PyCharm, Jupiter NoteBook

Work Experience

- My Previous Organization is **HCL Technologies Ltd. from May 2016 to Feb 2020** and currently working as **Consultant in Capgemini India Ltd, Chennai from Feb 2020 to till date.**

Project #1

- Project : **GE Healthcare**
- Client : GE Healthcare
- Environment : S3, PySpark, Glue, Lambda, Python, Step Function, Cloud Watch
- Role : AWS Developer
- Duration : Feb 2020 to till now

Description:

- The idea behind this project is to create the workflow/Pipeline to execute the job automatically with the help of Step function as well as scheduling will be done by CloudWatch rule and create the lambda function to generate the report of State machine as well as create the glue job with the help of lambda.

Responsibilities:

- **Developed the Python script to generate the Report of State Machine.**
- **Developed the Python script to generate the report of Glue Job.**
- **Development of Step Functions** in AWS Console as per Track teams requirement.
- **Creation of workflow** with the help of Step Function.
- **Creating the Master Step function** for executing workflow smoothly.
- **Creating Activities, Lambda, CloudWatch rules in Dev.**
- Code Migration Process like **Code Push/Commit to GIT and Raising PR for Migration.**
- Review code from Orchestration side (atleast one approval needed from our team).
- **Create Dynamo DB entries in case of new Source added to Orchestration team with the help of Lambda.**
- Dry run in Dev(1st run after creating SFN).
- **Monitoring & informing the respective track team in case of any errors.**
- Taking up Adhoc requests once we receive official Email communication.
- Will run only Child SF's for Adhoc runs.(NOT SUPPOSED TO TRIGGER MASTER STEP FUNCTIONS TWICE IN A DAY, SINCE THAT WILL TRIGGER THE NEXT SF IN FLOW. NEVER SUBMIT MASTER FOR JOB COMPACTION & COPY JOB.
- In Case if you are supposed to trigger the successful Master for that day again as part of adhoc run **MAKE SURE YOU DISABLED THE RESPECTIVE CLOUDWATCH RULE.**
- Prod migrations should be taken through CICD Pipeline(Don't entertain direct changes in Prod)
- Having an experience on GIT Versioning Tool.
- **Having an experience on Cloud Watch for scheduling the job.**
- Follow the Agile Methodology i.e. **Rally.**

Project #2

- Project : **FED-EX**
- Client : Fed-ex
- Environment : HDFS, Sqoop, Hive, Pig ,Unix Shell Scripting ,Java ,Spark SQL ,Ambari ,Scala
- Role : Hadoop Developer
- Duration : Aug 18 to Feb 20

Description:

- The idea behind this project is to minimize the cost involved in the items that are caged at customs across international borders and to improve user satisfaction for FEDEX.

Responsibilities:

- Developed end to end Data Pipeline for various data sources like from **RDBMS to HDFS**.
- Developed data ingestion script to ingest data from Oracle to HDFS/HIVE.
- Written API using Spark-SQL for various **data stages to analysis and maintain the data** in well formed.
- Used multiple **Spark-SQL functions** to clean the data & stored it into **optimized format**. Responsible for writing **Hive Queries** for analyzing data in Hive warehouse **using Spark- SQL**.
- Develop **unit test cases and test scripts** based on documented source.
- Performed **data analysis activity on huge size of shipment data for finding caging** reason of shipment Area wise.
- **Responsible for writing the data validation query for two different data sets in Hive warehouse using Spark-SQL**.

Project #3

- Project : **Citi Bank - Japan**
- Client : Citi Bank - Japan
- Environment : HDFS, Sqoop, Hive, Pig, Flume
- Role : Hadoop Developer
- Duration : July 17 to Aug 18

Description:

Citi Bank is leading Banking and credit card services provider, to use Hadoop technologies to handle And analysis large amount of data currently the organization has the data in RDBMS but to use the hadoop Ecosystem for storage, archival and analysis of large amount of data also stored in HIVE tables on Hadoop Cluster which is intended to hold historical data from various sources (100s of Terabytes of data), which would provide a consolidated view of the customer's products across the bank.

Responsibilities:

- Experience in **import and export the data from RDBMS to HDFS and HDFS to RDBMS by Sqoop**.
- Experience in **incremental data load mechanism in Hive tables** using **Sqoop**.

- Experience in **handling different file formats like Text files, ORC files, Parquet data files** in Hive.
- Developed HIVE queries for the **analysis and processing of data**.
- Experience in **converting all the huge hive tables in ORC format table** for better **query performance**.
- **Experience in Hive Query Optimization.**
- Implemented **Partitioning, Dynamic Partitions, Buckets in HIVE**.
- Experience in **creating Hive Tables, loading with data and writing Hive queries** which will invoke and run Map Reduce jobs in the backend.
- Experienced in performing analytics on **Structured, Semi structure and Unstructured data using Apache Pig**.
Developed Pig Latin scripts to **extract the data from the web server output files to load into HDFS**.
- Experience in **Data cleansing using Apache Pig**.

Project #4

- Project : **Citi Bank - Japan**
- Client : Citi Bank - Japan
- Environment : UNIX, Oracle(SQL).
- Role : L2 Support Engineer.
- Duration : Feb 17 to July 17

Description:

my project name is Secore japan interface, it is interface which is used to transfer the security messages to JASDEC, JASDEC checks that company is following the JASDEC policy or not if it follows then JASDEC sends the acknowledgment to transfer the fund and after receive the acknowledgment then it does the Fund Transfer with in Citi bank and other Bank also which comes from SECORE to CITIFT, RTGS.

Responsibilities:

- My role is **L2 support engineer** with 24*7.
- Need to resolve them if there are any alert otherwise make turn over tool entry to aware on this issue next shift person.
- Receive the ticket from L1 Team, which are application related common issue are : don't have access to particular user, user is able to login, application is not able to access.
- Interact with onsite team, if there are any **application related issue, attending the meeting with onshore as well as off shore**, need to join the call and need to update the issue which we have face in that week.
- Every quarterly we have **migration activity**.
- Need to work on 24*7 rotation wise
- **Continuous Check the health status** of file system if it goes above to 90% then send the mail to SA Team.
- Interacting with onsite client and scheduling the jobs using by **AUTOSYS**.
- Having hands on working experience on various tools Like **CITRIX, TECTIA, SERVICENOW, GENEIO CONSOLE, ISQL Express**.

Project #5

- Project : **Citi Bank - Japan**
- Client : Citi Bank - Japan
- Environment : UNIX, Oracle(SQL).
- Role : L1 Support Engineer.
- Duration : Jan 17 to Feb 17

Responsibilities:

- My role is L1 support engineer with 24*7.
- Resolving the tickets that are raised on the severity.
- Interacting with L2 team as well as L3 Team if any critical issue occurs.
- Providing support on failures and alerts for production applications during business hours or as part of a 24x and informing to L2 Team to work on issue.
- Sending the **daily Status reports and mails on stipulated time.**
- Generating monthly Report and forwarding to L2 Team.
- Performing the **COB activity** on Weekend.
- Monitoring the **jobs and Disk space provide the status** to L2 support Team.
- Raising the incident and change request as per client requirement.
- Forwarding the incident to L2 team if fix is required.
- Providing support on all application of Citi Bank i.e. RTGS, SJI, JBS, Z Bank, Flex cube etc.
- Having hands on working experience on various tools Like **CITRIX, TECTIA, SERVICENOW, GENEIO CONSOLE, AUTOSYS, ISQL Express.**

Project # 6

- Project : **Land O' Lakes**
- Client : Land O' Lakes
- Environment : UNIX, Oracle(SQL).
- Role : L2 Support Engineer.
- Duration : May 16 to Jan 17

Description:

- Production support is using for transfer the bank files by CASH job from WEBSERIES to WELLS FARGO as well as transfer the file from CONCUR to LOL in COSI job and some jobs are schedule in TIDAL it will be move the File/cash in described path with in the time and without the manual effort as well as it calls some PLSQL script which makes the changes in database.
- Address/escalate any issue to the concerned parties and resolve as within the SLA.
- Enhancement and maintains of WEBSERIES, COSI, LOL, CURX, ASJD & WEBX application which involves Analysis and Development.

Responsibilities:

- My role is **L2 support engineer** with 24* 7.

- Resolving the tickets based on **SLA**.
- **Resolving the tickets** that are raised on the severity.
- **Interacting with onsite client**.
- **Handle exceptions and alerts for production applications** during business hours or as part of a **24x7 on-call rotation**.
- Work directly with customers to triage their technical issues and **provide resolution on support issues within the scope of support per the Service Level Agreement (SLA)**
- **Resolving the job failures** based on SLA
- Sending the **daily Status reports**.
- **Checking the log files**.
- Having good experience on **RFC's (Request for Changes)** as per client assigned.
- Having hands on working experience on various tools Like **CITRIX, FILEZILLA, PUTTY, ORACLE 11g, SERVICE NOW, FTP and WINSCP**.

Key Achievements: Received 1 Spot awards for delivered task with in SLA from group project Manager.

ACADEMIC STATS

Name Of Institution	University/Board	Year	Qualification	Percentage
Vikrant Institute of Technology Management Gwalior, Madhya Pradesh	RGPV, Bhopal	2015	B.E. In (CSE)	78%
Shree Krishna Higher secondary school, Kailaras	M.P. Board	2011	Higher Secondary	77%
S.R.S.S. High School, Nirara	M.P. Board	2008	High School	66%

DECLARATION

I hereby declare that all the details above are true to the best of my knowledge.

Date: 23/ Sep /2020

Place: Chennai

(ChhaviSharan Bansal)