Poonam Yadav

E-mail: poonamy08.90@gmail.com
Mobile: + 91 9620885222/7483199634
Current Employer : IBM India Pvt Ltd

Professional Summary

- 6+ years of IT experience in Development, Support, Implementation of software applications which includes 3+ years of experience in **Big Data using Hadoop, Hive, PIG, Sqoop, PySpark & Python Programing**.
- Worked on different services **of AWS like S3 , Lambda, Glue(Pyspark) ,Athena, CloudFormation etc.**
- I have developed a web application using **Python, Flask, Pandas, Numpy, HTML & CSS**
- Written **Hive** queries for data analysis to meet the business requirements.
- Involved in creating tables, partitioning, bucketing of table and loading data from DB2 to Hive tables using **Scoop**.
- Written **PySpark** programs using **RDD** & **Dataframes**
- Involved in writing data transformations, data cleansing using Spark.
- Working experience on the **Agile and Waterfall models** and have worked closely with clients.
- Proficient in IBM mainframes **MVS, COBOL, JCL, VSAM** and **DB2.**

Skill Set

Languages/Technologies: PySpark, Hive, Sqoop, Python, HDFS, COBOL
Database: IMS and DB2,SQL
Defect Tracking Tools : Rally Dev, TDP, Jira
Remote Connection Tools: Putty, WinSCP
IDE: Pycharm, IDLE, Jupyter Notebook, Intellij
Tools: Vagrant, Virtual box, Bazel
AWS Services: S3, Glue, Lambda , Athena & CloudFormation

Project Experience

- Working with Deloitte USI as Consultant since Apr, 2019.
- Worked with IBM India Pvt Ltd from Apr,2014 to Mar 2019 as Senior Software Developer.

Dec 2019 – Present	Data Engineer	Deloitte USI

Project 1# - Enterprise Data Backbone

Description: The scope of the project is to create an enterprise Data Backbone(EDB) which is a shared platform built on AWS cloud based on OMOP CDM model, that would help multiple client teams to work coherently and perform BI/Visualizations and analytics.

Technical Environment: Amazon Web Services(AWS) – S3 , Lambda, Glue ETL , Aurora DB(Postgres engine) & Athena.

Responsibilities:

- Understanding functional requirements and build technical architecture to meet them.
- Design and build Audit, Balance and Control framework to store metadata and transactional audit and error information in a central repository on RDS (AWS Aurora DB).
- Design and build OMOP standard tables in conformed layer and map them from source data.
- Develop reusable and parameterized data ingestion/transformation jobs using AWS Lambda/Glue ETL in Python and Pyspark.
- Produce Unit test case document and results for code sign-off to higher environments.
- Create Cloud formation templates to create roles and permissions to Glue, Lambda, S3 & requires services of AWS.
- Consolidate artifacts and deployment trackers for code migration to higher environments.

Apr 2019– Dec 2019	Data Engineer	Deloitte USI

Project 2# - Knowledgebase for Stork

Description: The Takeda initiative, the Strategic Translational Oncology Research Knowledgebase (STORK), has the goal to enable the **integration of disparate assay and patient data** to inform and influence **drug program benchmarking** and data-driven decision making.

Technical Environment: Amazon Web Services(AWS) – S3 , Lambda, Glue ETL , Aurora DB(Postgres engine) & Elastic Search.

Responsibilities:

- Developed Spark scripts by using PySpark as per the requirement
- Involved in design phase to decide on size of EMR cluster
- Used RDD, Dataframe & SparkQL to write Spark scripts
- Used AWS EMR cli & Jupyter notebook to execute code & S3 as the Raw Layer to ingest data.
- Developed code to load the data to Elastic search using PySpark.
- Worked on Postgre Db for Configuration & Audit tables.
- Load data back to Conformed layer for further analysis in R & UI.

Sep 2018 – Mar 2019	Data Engineer	HPE

Project 3# - Enterprise system

Description: The object of this project is to build a scalable and robust real-time data ingestion and processing platform through which client SAP system can send real-time financial data and using Kafka we ingest those data, save those data into HDFS along with Data Lake and for real-time data processing send those data into Spark and for historical data we use Hive along with Oozie jobs to load data.

Technical Environment: Cloudera , Apache Tez, Oozie , Scala

Responsibilities:

- Developing real-time ingestion platform using Kafka, Spark, Scala and Tez.
- Schedule jobs by creating Oozie workflows.
- Transform data based on different use cases and load back to Hive table.
- Work closely with the Qlik team to keep the UI working.

Dec 2017– Aug 2018	Data Engineer	HCSC

Project 3# - Migration of DB2 tables to Hive tables

Description: This project is to load the data from DB2 tables to Hive Tables using Scoop & do the reconciliation once data is loaded correctly.

Responsibilities:
- Import data into HDFS and Hive using Sqoop.
- Involve in create Hive tables, loading with data and writing Hive queries which will run internally in MapReduce way.
- Load data into Hive tables.
- Work closely with the business and analytics team in gathering the system requirements.
- Reconciliation of tables i.e verifying whether the data is loaded correctly or not

Jan 2017– Nov 2017	Application Programmer	IBM Internal Account

Project 1# - Create a Web application which allows to upload excel, process data & download it.

Description: The project dealt with the Invoicing and billing exclusions of the IBM's Third party vendors. When a Vendors fails to honor the Invoice within the stipulated time period, an exclusion criteria is associated with the default.

Responsibilities:
- Design a front end in Flask, where an IBM reporting/processing team could upload a region-wise report of IBM's Vendor invoices. Post successful upload of the report, they could select/deselect multiple exclusion criteria's, valid for that particular region.
- Designed customized pages for each country in HTML,CSS having the exclusion criteria that applies for that country.
- Write python modules for different countries using Python, Pandas & Numpy.
- Work closely with the IBM clients to understand the requirement

Apr 2014– Jan 2017	Application Programmer	HCSC/AT&T

Project 1# - Group Finance Risk Assessment
Project 2# - RX/PBM Stoploss Project
Project 3# - Delinquency Project
Project 4# - FPSC : Increasing the Span Limit from 400 to 1000 Spans
Project 5# - Multilogon Project

Responsibilities:
- Gathering and analyzing new requirements from the Clients
- Understanding the requirement and finalizing the scope.
- Written programs independently for different functionalities.
- Loading of records in Test and UAT region.
- Worked in sync with frontend developer for system testing.
- Prepared test cases and test results documents and discussed with client.
- Worked in odd hours for UAT testing with onsite team.

Academics

- Acquired B.Tech**.** in Information Technology with an aggregate of 73.84% from Uttar Pradesh Technical University,
  Ghaziabad
- Passed class 10th (CBSE) from Kendriya Vidyalaya,Ghaziabad with 80.6%.
- Passed class 12th (CBSE) from CSHP Public School, Ghaziabad with 72%.

Passed

Personal Details

Date of Birth	:	08th May 1990
Permanent Address	:	L-420A, Sec-12, Pratap Vihar, Ghaziabad,UP-201009, India.
Current Location	:	Bangalore, India
Gender	:	Female
Passport Status	:	Available
Languages Known	:	**English, Hindi**