|  |
| --- |
| SUMMARY |
| I have recently joined Publicis Sapient's Gurugram office as Senior Associate L2- Data Engineering. |
|  |
| Prior to joining Publicis Sapient, I have been working as Data Engineer for over nine years. I have got extensive hands-on on Spark, Scala along with AWS and good exposure on Big Data, Hadoop, SQL, Python, GitHub, Batch Processing with Parquet files. I have worked in different domains like Insurance, Retail, Banking and Finance, etc. Apart from being senior developer, I was leading a team of 5 data engineers and led the L&D plan, also contributed to recruitment. |

|  |
| --- |
| WORK EXPERIENCE: |
| **Publicis Sapient September 2023 To Present** |
| **Data Engineering Team, Senior Associate L2** |
| **Description:**NA |
| **Roles and Responsibilities:**NA |
| **Skills/Domain:**Big Data, Hadoop, Spark, Scala, SQL, Sqoop, Cloud, AWS, Azure, Data Engineering |
| **AXA Business Services, Pvt Ltd June 2019 To September 2023** |
| **Sherlock, Consultant Data Engineering** |
| **Description:**End to End Fraud Claims Analytics Project to detect frauds among various Insurance claims e.g. Motor, Housing, etc. By detecting fraud claims we were able to make huge recoveries and more profit. |
| **Roles and Responsibilities:**- Following Agile Strategy to plan and deliver development activities. |
| - Extracting and transforming data from source files using Avro and Scala. |
| - Building business-features using Scala Datasets to create an analytical layer. |
| - Using Spark GraphFrames to generate graph for Web Application, to visualize links between various information linked to a particular claim/policy/person/vehicle, etc. |
| - Using ScalaTest to create unit test cases for major functionalities in the code. |
| - Using Lambda to orchestrate, EMR to execute the pipeline, S3 to store and Athena to analyze the data. |
| - Built the foundation layer for X-Res tool, using Python, which helped Reserving teams to calculate and visualize various reports on the same platform. |
| - Responsible for team management activities for AXA UK like standups, code review, technical discussions for improvements. |
| **Skills/Domain:**Spark, Scala, AWS, SQL, Analytics |
| **Toll Technology Centre July 2018 To June 2019** |
| **Internal, Data Engineer** |
| **Description:**Multiple Internal projects to cater the data needs of Reporting and Data Science teams |
| **Roles and Responsibilities:**- Requirement gathering and building the ETL solutions on Spark and scheduling the jobs using Oozie. |
| - Imported data from AWS S3 into Spark using Scala, performed transformations/actions on Dataframes. |
| - Developed scripts to automate the manual processes of ingesting data from Excel files to HDFS using Pandas and various generic libraries of Python. |
| - POCs for various projects using Informatica BDM, Spark Scala, Hive, Oozie, Sqoop to showcase the comparison and build ETL pipelines, finally that data being used for reporting and data science projects. |
| - Designed and created Hive external tables using shared meta-store instead of derby with dynamic partitioning and bucketing. |
| **Skills/Domain:**Spark, Scala, AWS, SQL, Informatica, Python |
| **Datametica Solutions Pvt. Ltd. January 2017 To July 2018** |
| **AngiesList, Yes Bank, Big Data Developer** |
| **Description:**NA |
| **Roles and Responsibilities:**- Implemented Spark using Scala and utilizing Spark Dataframe, Spark SQL API for faster processing of data instead of Mapreduce in Java. |
| - Used HIVE to do transformations, event joins and aggregations before storing the data onto HDFS. |
| - Involved in developing Shell scripts to orchestrate execution of all other scripts (Sqoop, Hive and Pig/Spark) and move the data files within and outside of HDFS. |
| - Worked on AWS and GCP cloud platforms for storage and processing purposes. |
| **Skills/Domain:**Spark, Scala, Hive, AWS, GCP, SQL |
| **Cognizant Technology Solutions Pvt Ltd February 2014 To December 2016** |
| **CareFirst, PepsiCo, STC, Programmer Analyst** |
| **Description:**NA |
| **Roles and Responsibilities:**- Analyzing the business requirement and developing the logic to meet the requirements. |
| - Used Sqoop and Oozie to import the structured data from RDBMS to HDFS and schedule the task for incremental loads. |
| - Developed mappingsin Informatica BDM that perform Extraction, Transformation and load of source data into Derived Masters schema using various transformations like Source Qualifier, Aggregator, Filter, Router, Sequence Generator, Joiner, Expression, Normalizer and update strategy to meet business logic in the mappings. |
| - Designed and implemented partitioning (Static, Dynamic), Buckets used Parquet, Avro file formats in Hive. |
| - Coach new team members on technical tools such as Informatica BDM and Hadoop ecosystem |
| **Skills/Domain:**Informatica, Shell Scripting, Hive, SQL, Spark |

|  |
| --- |
| EDUCATION: |
| B.Tech. (Bachelor of Technology), Computer Science and Engineering: AIT, Guru Gobind Singh Indraprastha University, 2013 |
|  |

|  |  |
| --- | --- |
| **SKILLS:** | |
| **Additional Skill(s):** | Data Analytics,Data Engineering,Data Platforms,DE-Big Data,DE-Data Warehouse and ETL |
| **Other Skills:** |  |
|  |  |

|  |
| --- |
| ADDITIONAL DETAILS: |
| **Certification/Trainings** |
| Microsoft Azure DP 203 -- 2023 |
| Microsoft Azure DP 900 -- 2022 |
| Functional Programming in Scala (Coursera specialization) -- 2020 |
| Lightbend Scala Language Professional- Level2 -- 2020 |
| **Awards** |
| Employee of the Month -- 2020 |
| Spot Award- Datametica -- 2018 |
| **Links** |
| LinkedIn | linkedin.com/in/ankmit09 |