|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |  | | --- | | **Education**  B. Tech, Mechanical Engineering, The Vazir Sultan College of Engineering, IN, 2012 |  |  | | --- | | **Languages**  English - Very Good  Hindi - Fluent | |  | **Trivedi Rajashekar**  **India Contact: +91 8919230873**  **Profile**  Around 4.5 years of experience in the field of IT using systems involved in JAVA, PYTHON, SQL and development of Big-Data and Hadoop framework.  **Key skills**  Hadoop, HDFS, Map Reduce and Yarn, Pig, Hive, SNOWFLAKE, BIG QUERY, Apache Spark, Core Python, R, Core Java, SQL  **Key courses and training**  Undertaking AWS Data Engineer.  **Work experience**  **Consultant** HCL, India | Mar 2019 - Jul 2022  PROJECTS UNDERTAKEN  Project: MTA (Multi Touch Attribution):  Gathering the interests of the public by reading the multiple touches (Selections) made by the  users will be captured on the web server and collecting the data. Applying transformations over the data to estimate the public  interests and choices and offering the benefits WRT to their choices.  Title: MTA  Client: DROPBOX.  Role: consultant  Team: 2  Duration: MAR-2021 to till date  Environment: AWS, SNOWFLAKE, SPARK, PYTHON, HIVE, AIRFLOW, ATOM, GIT  Responsibilities:  • Creating the drone jobs to load the data from Hive to S3  • Creating Stage and Stream tables.  • Creating the Snowpipes to load the streaming into s3  • Creating procedure which is checking the data for table status whether all the tables have up to date  data (Day - 1). Then only tasks get executed and then only data will be loaded into tables.  • Procedures will be called by an automated task and data will be loaded into S3  • Testing and pushing the tables into the prod environment.  Project: EMS (Electronic Mailing System):  Preparing the personalized Messages and streaming the messages with the customer inputs.  Understanding and analyzing the customers native language and communicating with the customer in his own language to feel  customer comfortable.  Title: EMS  Client: DROPBOX.  Role: consultant  Team: 5  Duration: MAR-2021 to till date  Environment: AWS, SNOWFLAKE, SPARK, ITERABLE, SMARTLING, BHOOMI, HIVE, AIRFLOW, GIT  Responsibilities:  • Creating the drone jobs to load the data from Hive to S3 TO SNOWFLAKE  • Creating Stage and Stream tables INTO SNOWFLAKE and to S3 bucket.  • Creating the Snowpipes to load the streaming into s3  • Creating procedure which is checking the data for table status whether all the tables have up to date  data (Day - 1). Then only tasks get executed and then only data will be loaded into tables.  • Procedures will be called by an automated task and data will be loaded into S3  • Testing and pushing the tables into the prod environment.  Title: ORION DATA FACTORY  Client: PALO ALTO.  Role: consultant  Team: 7  Duration: OCT-2020 to FEB-2021.  Environment: GCP, HDFS, SPARK, PYTHON, HIVE, AIRFLOW, ATOM, GIT  Description  Palo Alto Networks, Inc. is an American multinational cybersecurity company with headquarters in Santa Clara, California. Its  core products are a platform that includes advanced firewalls and cloud-based offerings that extend those firewalls to cover other  aspects of security. Our project role is to maintain a data pipeline and make the changes as per client requirements and deploy the  changes in all respective environments.  • Develop, construct, test and maintain architecture.  • Aligning architecture with business requirements.  • DataBase Assessment about objects and workloads to migrate  • Created an architecture on GCP technology stack  • Created VPC cloud setup in GCP  • Script changes in the existing production python scripts as per client requirements.  • Communicating with the client and the new changes have to be done as per the client expectation.  • Testing the changes and validating with client requirements.  • Taking approvals from clients to push the code to further environments.  • PR and CR creation.  • Job monitoring and reporting.  • Responsible for Incidents and CRs.  PROJECT: SINGULARITY  Title: SINGULARITY  Client: DBS bank Singapore.  Role: consultant  Team: 15  Duration: APR-2019 to SEP-2020.  Environment: HDFS, SPARK, PYTHON, HIVE, AIRFLOW  Description  DBS Bank has huge historical data in its own data base called TERADATA(T1). The actual data placed in the TERADATA  which is associated with different source systems like mainframe open source, SAS based files need to get those files data as  RDD’S and maintains as input datasets. After that create the pipelines and pastes the developed existing Teradata code and  replace the spark supportive code. Than we have a special tool called job server using this we can run Jenkins for creating  ADA(hive) table we can call it as T3 table. once we create the tables successfully trigger the jobs continuously we have a special  scheduler tool called AIRFLOW.  Responsibilities:  • Creating end to end Spark applications to perform various data cleansing, validation,  transformation and summarization activities on user behavioral data.  • Used Spark SQL and data frame API extensively to build spark applications.  • Imported data from different sources into Spark RDD for processing.  • Developed custom aggregate functions using Spark SQL and performed interactive querying.  • Developed Shell Scripts to import files from AWS and exported into various paths as per  requirement with small modifications.  • Developing and Managing Metadata using Collibra.  • Design the flow of data from source to destination (T1 to T3)  • Involved in loading data from Teradata to HDFS(alleuxio).  • Exporting the data from HDFS to HIVE by creating tables.  • Using the in-memory computing capabilities of spark using Scala, performed advanced procedures  like text analytics and processing.  • Used Spark-Streaming APIs to perform necessary transformations and actions on the fly which gets  the data from AIRFLOW in near real time and Persists into HDFS.  • Developed PYTHON scripts to generate Data frames/Data sets and RDD in Spark 2.3 for Data  Aggregation, queries and writing data back into HIVE.  **Software Engineer** Genpact With the payroll of IDC Technologies., India | Sep 2017 - Jun 2018  Project: UK Infrastructure Major  Team Size: 12  Environment: Hadoop, SSAS, SSIS, SharePoint, Power BI and Service Now  Description: Project involves managing the flow of data from different sources like Oracle, SQL Server(TAMI), TERADATA  and SalesForce to a Hadoop data lake that consists of multiple SQOOP Jobs that run on daily basis with stringent timelines.  Batches run on a OOZIE scheduler which involves running of multiple SQL and HQL scripts which will enable the flow of the  data from data lake to data Mart.  Once the batch completes the all the business logic and the calculation is being applied at SSAS level, this raised a requirement  of cube processing automation, to accomplish the SSIS packages are built to enable the processing of the cubes at scheduled time  Responsibilities  • Analyzing the batch performance on hourly basis and raise a ticket to top management in case of  any issues anticipated  • Estimating the batch completion time to estimate the report delivery within SLA  • Identifying the long running SQL and HQL scripts and optimizing them to minimize the batch  duration and maximize the performance  • Working on the dimension matrix to identify the duplicate entities that consume lot of space and  proposing for the removal of the dimensions  • Building and maintaining of shell scripts which will run through all the log files to identify the  failure of the batches  • Meeting the ad-hoc requirement of changes in the batch by changing the existing Python code |
|  |  |  |