**Rakesh Gottipati  
Email: rakesh.data098@gmail.com**

**PH: 801-696-6617**

**LinkedIn Profile: https://** **www.linkedin.com/in/rakesh-g-986576141/**

**About Me**

Big Data and Cloud engineering professional with 10+ years of experience, passionate to work in the intersection of cloud computing, Data Engineering, experienced in building distributed data solutions, data analytical applications, ETL and streaming pipelines leveraging big data ecosystem components, Databricks platform, AWS and Azure Cloud services, Hadoop Eco System.

Expert in Data Analysis, Data Profiling, Data Cleansing & Quality, Data Migration, Data Integration, Data Ingestion and Data Transformation

As a data engineer, I was responsible for Assess and document data requirements client-specific requirements to develop user-friendly BI solutions - reports, dashboards, and decision aids. Worked on Data warehousing, DatPython,ering, Feature engineering, big data, ETL/ELT, and Business Intelligence , specialize in Azure frameworks, Hadoop Ecosystem, Excel, Snowflake, relational databases, tools like Tableau, PowerBI Python and Data DevOps Frameworks/Azure Dev Ops Pipelines , specialize in AWS and Azure frameworks, Cloudera, Hadoop Ecosystem, Spark/Py Spark/Scala, Data bricks, Hive, Redshift, Snowflake, relational databases, tools like Tableau, Airflow, DBT, Presto/Athena, and Data DevOps Frameworks/Pipelines with strong Programming/Scripting skills in Python , Data Warehouse developer, Data modeler, Data Analyst ,Data Modeling, Implementation and Support and maintenance of various applications in both OLTP and OLAP systems.

**Professional Summary (8 + Years of Work Experience)**

**Hadoop Eco system**

* Worked on Hadoop ecosystem and different frameworks inside it – HDFS, YARN, MapReduce, Apache Pig, Hive, Flume, Sqoop, ZooKeeper, Oozie, Impala and Kafka Spark: Building the infrastructure requirement for optimal extraction, transformation, and loading of data from a wide variety of data sources using Apache Spark, SparkSQL, python technologies. Worked with Cloudera **Good** understanding of **Zookeeper** and **Kafka** for monitoring managing **Hadoop** jobs used **Cloudera** **CDH4**, **CDH5** for monitoring and managing Hadoop cluster.
* Good working experience on **Spark (Spark Core Component, Spark SQL, Spark Streaming, Spark MLlib, Spark GraphX, SparkR)** with **Scala** and **Kafka**. Experience in Analysis, Design, Development and **Big Data** in **Scala**, **Spark**, **Hadoop**, **Pig** and **HDFS** environment
* Expertise in developing applications using **Big Data ecosystem** – **Hadoop (HDFS, Yarn, MapReduce), Pig, Sqoop, Zookeeper, Spark, Flame, Hive.**
* Extensive experience in relational Data modeling, Dimensional data modeling, logical/Physical Design, ER Diagrams and OLTP and OLAP System Study and Analysis.
* Extensively used ERWIN and PowerDesigner to design Logical and Physical Data Models, to forward and reverse engineering data models and publishing data model to acrobat PDF files.
* Data Streaming from various sources cloud (AWS, Azure) on - premises by using the tools Spark Flume.
* Used Azure Data Lake, Azure Data Factory, Azure Machine Learning, Azure Databricks
* AWS cloud experience using EC2, S3, EMR, RDS, Redshift, AWS Sage maker, Glue.
* Implemented monitoring and established best practices around using **Elasticsearch** and used **AWS Lambda** to run code without managing servers.
* Built machine learning solutions using PySpark for large sets of data on Hadoop ecosystem.
* Adept in statistical programming languages like Python and R including Big-Data technologies like Hadoop, HDFS, Spark and Hive.
* Worked on technical stack like Snowflake, SSIS, SSAS, SSRS to design warehousing applications.
* Experience in data mining, including predictive behavior analysis, Optimization and Customer Segmentation analysis using SAS and SQL.
* Experience in Applied Statistics, Exploratory Data Analysis and Visualization using matplotlib, Tableau, Power BI, Google Analytics.

**AZURE**

* Experience in **AWS, Azure DevOps, Continuous Integration, Continuous Deployment and Cloud Implementations.**
* Experience on Migrating SQL database to **Azure data Lake, Azure data lake Analytics**, **Azure SQL Database, Data Bricks** and **Azure SQL Data warehouse, Azure HDInsight** and controlling and granting database accessandMigrating On premise databases to **Azure Data Lake store** using Azure Data factory
* Experience with **Azure** services like **HDInsight, Active Directory, Storage Explorer, Stream Analytics.**
* Extensive experience in **Azure Cloud Services (PaaS & IaaS), Storage, Data-Factory, Data Lake (ADLA** &ADLS), Active Directory, Synapse, Logic Apps, Azure Monitoring, Key Vault, and Azure SQL**.**

**AWS**

* Worked on Spark **Architecture** including **Spark SQL**, **Data Frames**, **Spark Streaming**, experience in analyzing Data with **Spark** while using **Scala**. Hands on experience in using other **Amazon Web Services** like **S3**, **VPC**, **EC2**, **Autoscaling**, **RedShift**, **DynamoDB**, **Route53**, **RDS**, **Glacier**, **EMR**.
* Expertise in **Amazon Web Services (AWS)** Cloud Platform which includes services like **VPC, DynamoDB, Route 53, Elastic Container Services (ECS), Security Groups, CloudWatch, EC2, S3, Security Groups, Kinesis, Red shift, IAM, CloudFormation, ELB, Cloud Front, Elastic Beanstalk (EBS).**
* Experience in using Python included **Boto3** to supplement automation provided by **Ansible** and **Terraform** for tasks such as encrypting Elastic Beanstalk volumes and scheduling Lambda functions for routine AWS tasks.
* Implemented monitoring and established best practices around using **Elasticsearch** and used **AWS Lambda** to run code without managing servers.

**Data Visualization**

* Worked on Text Analytics, generating Data Visualization using Python dashboards using Tableau
* Developed Consumer - based custom features and applications using **Python, Django, HTML, CSS**. Experienced with Software Development Life Cycle, Database designs, agile methodologies, coding, IDE's such as **Jupiter Notebook, PyCharm, Emacs, Spyder and Visual Studio**.
* Proficient in managing entire data science project life cycle and actively involved in all the phases of project life cycle including Data acquisition, Data cleaning, Feature scaling, Dimension reduction techniques, Feature engineering, Statistical modeling, and Ensemble learning.
* Worked on the principles of Data warehousing, Fact Tables, Dimension Tables, Star and Snowflake schema modeling.

**Technical Skills**

|  |  |
| --- | --- |
| **Hadoop Distributions** | Cloudera, AWS EMR and Azure Data Factory. |
| **Languages** | Scala, Python, SQL, Python, Hive QL, KSQL. |
| **IDE Tools** | Eclipse, IntelliJ, PyCharm. |
| **Analytical Techniques** | Hypothesis testing, Predictive analysis, Machine Learning, Regression Modelling, Logistic Modelling, Time Series Analysis, Decision Tree, Neural Networks, Support Vector Machines (SVM), Monte Carlo methods, Random Forest, Time series analysis. |
| **Analytical tool** | Rapid Data miner, Google analytics, IBM Watson, R Studio, SAS/STAT, Google Ads, Azure data lake analytics, SAS Enterprise miner, PyCharm, Jupyter notebook, NLP, MATLAB, GGPLOT, WEKA, Databricks, Jenkins. |
| **AWS Services** | VPC, IAM, S3, Elastic Beanstalk, CloudFront, Redshift, Lambda, Kinesis, DynamoDB, Direct Connect, Storage Gateway, EKS, DMS, SMS, SNS, and SWF |
| **Reporting and ETL Tools** | Tableau, Power BI, Talend, AWS GLUE. |
| **ETL Tools** | Microsoft SSIS, Talend Open Studio, Data Stage 11.x, Informatica PowerHouse 9.0, Informatica IDQ, Collibra, KAFKA, FLUME |
| **Databases** | Oracle, SQL Server, MySQL, MS Access, NoSQL Database (Hbase, Cassandra, Mongo DB), snowflake |
| **Big Data Technologies** | Hadoop, HDFS, Hive, Pig, Oozie, Sqoop, Spark, Machine Learning, Pandas, NumPy, Seaborn, Impala, Zookeeper, Flume, Airflow, Informatica, Snowflake, DataBricks, Kafka, Cloudera |
| **Machine Learning**  **And Statistics** | Regression, Random Forest, Clustering, Time-Series Forecasting, Hypothesis,  Explanatory Data Analysis |
| **Containerization** | Docker, Kubernetes |
| **CI/CD Tools** | Jenkins, Bamboo, GitLab CI, uDeploy, Travis CI, Octopus |
| **Operating Systems** | UNIX, LINUX, Ubuntu, CentOS. |
| **Other Software** | Control M, Eclipse, PyCharm, Jupyter, Apache, Jira, Putty, Advanced Excel, TOAD, Oracle SQL developer, MS Office, FTP, Control-M, SQL Assistant, Rally, JIRA, GitHub, JSON |
| **Frameworks** | Django, Flask, WebApp2 |

**Professional Experience / Work Experience**

**Role: Principal Data Engineer**

**Client: Nexus- Dallas, TX Feb 2020 to Present**

**Project Description:**  This project is an implementation of the Apache Hadoop technology stack on the Microsoft Azure cloud platform. I was involved in Data Analysis, Data Profiling, Data Integration, Migration, Data governance, metadata management, and Master Data Management and Configuration Management. The Project involved in developing the dashboards in Tableau, Power BI environments for Metrics and Compliance department. Pulling the data from various data sources into Hadoop platform and standardized all data through a series of master data management processes to reach client goal. As a Hadoop developer, involved in maintaining huge data and designing and developing predictive data models for business users according to the requirement.

Implementation is on hybrid cloud involving Azure, AWS

**Key Contributions**

* Using **Azure Data Factory(V2)**, created ingestion pipelines from different sources into **Azure Data Lake Storage.**
* Created and maintained Azure resources using a combination of **Windows PowerShell** and **Azure Resource Manager (ARM**) templates for unit testing during the DevOps process.
* Moving data from **Azure Blob storage** to **Azure Data Lake Storage** using **Azure Data Factory** pipelines.
* Configuring and developing Azure Databricks notebooks using **PySpark** and **Spark** **SQL** for **data transformation**, **aggregations**, and **extractions** from multiple file formats for analyzing the data.
* Involved into Application Design and Data Architecture using Cloud and Big Data solutions on **Azure**.
* Leading the effort for migration of Legacy-system to **Microsoft Azure** cloud-based solution and re-designing the Legacy Application solutions with minimal changes to run on cloud platform.
* Created several types of **data visualizations** using **Python** and **Tableau**. Performed **reverse engineering** using **Erwin** to redefine entities, attributes, and relationships existing database.
* Authored **Python (PySpark) Scripts** for custom UDF’s for Row/ Column manipulations, merges, aggregations, stacking, data labeling and for all Cleaning and conforming tasks.
* Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL.

**Responsibilities:**

* Transformed business problems into **Big** **Data** **solutions** and define **Big Data strategy** and **Roadmap**.
* Installed, configured, and maintained **Data Pipelines**
* Designed the business requirement collection approach based on the project scope and **SDLC** **methodology**.
* Created **Pipelines** in **ADF** using **Linked Services/Datasets/Pipeline/ to Extract, Transform,** and **load** data from different sources like **Azure SQL, Blob storage, Azure SQL Data warehouse,** write-back tool and backwards.
* Developed **Spark applications** using **Scala** and **Spark-SQL** for data extraction, transformation, and aggregation from multiple file formats.
* Used **kafka** and integrating with the **Spark Streaming**
* Used ORC, Parquet file formats on **HDInsight**, **Azure Blobs** and **Azure tables** to store for raw data.
* Involved in writing **T-SQL** working on **SSIS, SSAS, Data Cleansing, Data Scrubbing** and **Data Migration.**
* Worked on **Dimensional** and **Relational** **Data** **Modeling** using **Star** and **Snowflake** **Schemas**, **OLTP/OLAP** **system**, **Conceptual**, **Logical** and **Physical** **data** modeling using **Erwin.**
* Performed **PoC** for **Big** **data** **solution** using **Hadoop** for data loading and data querying
* Extract Transform and Load data from Sources Systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL and U-SQL Azure Data Lake Analytics. Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW ) and processing the data in Databricks.
* Designed and Developed ETL jobs to extract data from different sources and load it in data mart in **Snowflake** and managed **Snowflake** clusters such as launching the cluster by specifying the nodes and performing the data analysis queries.
* Experience as a big data engineer on SQL, HIVE, ETL(Informatica), Data modeling (Erwin),Snowflake, HBase, and HDFS.
* Implemented Spark using Python and Spark SQL for faster testing and processing of data. Loaded processed data into Snowflake for advanced data warehousing and analytics capabilities.
* Worked on real-time data processing using Spark Streaming and Kafka using Python. Integrated real-time data into Snowflake to enable real-time analytics and insights.
* Involved in the complete SSIS life cycle in creating SSIS packages, building, deploying, and executing the packages in both environments (Development and Production) with Snowflake as the data source and destination.
* Used **Sqoop** to channel data from different sources of **HDFS** and **RDBMS**.
* Involved in **Normalization** and **De-Normalization** of existing tables for faster query retrieval.
* Developed and maintained **data** **dictionary** to create metadata reports for technical and business purpose.

**Environment:** Spark, Scala, AWS, ETL, Kafka, Tableau, Hadoop, Python, Snowflake, HDFS, Hive, MapReduce, PySpark, Pig, Docker, Sqoop, Teradata, JSON, MongoDB, SQL, Agile and Windows, UNIX

**Client: Volkswagen – Auburn Hills,MI Dec 2018 – Jan 2020**

**Sr. Data Engineer**

**Project Description:**  Building stage, data warehouse and BIlayers, enriching the required matrices for the BI Team for building the dashboard and implementing the business logic, Performing ETL on AWS Glue, EMR and load the data into stage, Dw, BI tables, Executing the SQL scripts from S3 and automation of SQL scripts.

Here we are using different technologies to process the incoming data using Oozie, Spark, Python, Scala. We are processing the entire flow on Premises AS400 server and the final data we are loading to Teradata and loaded a few intermediate datasets into Hive tables.

**Data Migration:** Analyzing existing conversion processes and identified the problems impacting the quality of our data conversions and implemented with Python, SQL. Managed end-to-end complex data migration and ETL processes. Worked closely to the deadline, daily targets.

**Key Contributions:**

* Implement AWS Lambdas to drive real-time monitoring dashboard of Kinesis streams.
* Worked on Data Warehouse design, data integration transformation using Apache Spark and Python.
* Created/Setup EMR clusters for running data engineering work loads and data scientists.
* Utilized Spark SQL API in PySpark to extract and load data and perform SQL queries.
* Developing the data pipelines on Hadoop using SAS or Pyspark.
* Developed end to end **Spark** applications using **Scala** to perform various **data cleansing, validation**, transformation, and summarization activities according to the requirements.
* Designed and Developed **Spark** workflows using **Scala** for data pull from **AWS S3 bucket** and **Snowflake** applying transformations on it.
* Developed **Spark code** in **Python and SparkSQL** environment for faster testing **and** processing of data and loading the data into **Spark RDD** and doing In-memory computation to generate the output response with less memory usage.
* Developed Simple to complex **MapReduce** Jobs using **Hive and Pig.**
* Developed **PIG** scripts to transform the raw data into intelligent data as specified by business users.
* Utilized **Spark, Scala, Python** for querying, preparing from big data sources. Wrote pre-processing queries in **python** for internal **spark** jobs.

**Responsibilities:**

* Expertise in designing and deployment of **Hadoop cluster** and different Big Data analytic tools including **Pig**, Hive, HBase, Oozie, Sqoop, Flume, **Spark**, Impala.
* Ingested the data from Relational Databases to **HDFS** using **SQOOP**.
* Implemented advanced procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in python.
* Implemented **Spark** using python and Spark SQL for faster testing and processing of data.
* Involved in converting **Hive/SQL** queries into Spark transformations using Spark RDDs, Scala.
* Wrote queries in SQL and R to extract, transform and load (ETL) data from large datasets using Data Staging.
* Worked on data pattern recognition, data cleaning as well as data visualizations such as Scatter Plot, Box Plot and Histogram Plot to explore the data using packages Matplotlib, Seaborn in Python, ggplot in R and SAS.
* Chosen and produced information into csv records and put away them into **AWS S3** by utilizing AWS EC2 and afterward organized and put away in AWS Redshift.
* Developed **Glue** Jobs to read data in csv format in raw layer and write data to parquet format in publish layer.
* Developed a **Glue** job which does delete, update and incremental loads from source to target.
* Extract Real time feed using **Kafka** and **Spark** Streaming and convert it to RDD and process data in the form of Data Frame and save the data as Parquet format in HDFS.
* Experienced in transferring Streaming data, data from different data sources into **HDFS, No SQL** databases.
* Created **ETL Mapping** with Talend Integration Suite to pull data from Source, apply transformations, and load data into target database.
* Experience in connecting to SAS interface to Hadoop and querying results to SAS BI client tools.
* Creating Spark clusters and configuring high concurrency clusters using Azure Databricks to speed up the preparation of high-quality data.
* Used **PySpark and Pandas** to calculate the moving average and RSI score of the stocks and generated them into data warehouse.
* Chipped away at a python content to extricate information from Netezza data sets and move it to AWS S3.
* Developed multiple Kafka Producers Consumers from scratch to as per the SRS requirement specifications.
* Created on-demand tables on S3 files using **Lambda Functions and AWS Glue using Python and PySpark.**
* Optimized existing algorithms in Hadoop using **Spark Context, Spark-SQL, Data Frames and Pair RDD's**.
* Developed pipeline for POC to compare performance/efficiency while running pipeline using the **AWS EMR Spark cluster.** Extensive experience in building ETL jobs using Jupyter notebooks with Apache Spark.
* Running analytics on power plant data using **Pyspark API**with **Jupyter**notebooks in on premise cluster for certain transforming needs.

**Environment:** Hadoop, Hive, Flume, Map Reduce, Sqoop, Kafka, Spark, Yarn, Cassandra, Oozie, shell Scripting, Scala, Maven, MySQL AWS (Lambda, Glue, EMR), NoSQL, Python, HDFS, Amazon Elastic Compute Cloud, Amazon Simple Storage Service(S3), CloudWatch Triggers (SQS, Event Bridge, SNS), REST, ETL, DynamoDB, JSON, Tableau

**Client: Kellogg’s- Battle Creek,MI Apr 2017 – Nov 2018**

**Sr. Data Engineer / Big Data Developer / AWS Cloud Data Engineer**

**Project Description:**

As part of the checkout flow, Lowe’s offers various discounts on items we sell online. These discounts are available for customers to view/apply throughout the checkout flow. Discount- Engine is a key micro-service in the checkout flow that returns price adjustments and other Discounts-applicable information back to the up-stream clients to help present the discounted Information to our customers on the website (Online and In-store POS). Discounts are created in Discount-Admin UI and pushed to Cassandra (Google Cloud) using Google Pub/Sub and Redis (In-Store). Discounts data are validated before loading into Cassandra database using spring Validation framework. The discounts data available in Cassandra will be used for price adjustments during checkout (Online and POS). Active discounts will be cached using GUAVA cache to minimize time to process each checkout.

**Key Contributions:**

* Prepared ETL design document which consists of the database structure, change data capture, Error handling, restart, and refresh strategies.
* Worked on batch processing of data sources using Apache Spark, Elastic search.
* Enabling other teams to work with more complex scenarios and machine learning solutions.
* Involved in using Spark Data Frames to create Various Datasets and applied business transformations and data cleansing operations using Databricks Notebooks.
* Leveraged and integrated **Google Cloud Storage** and **Big Query applications**, which connected to Tableau for end user web-based dashboards and reports.

**Responsibilities:**

* Worked on code migration of quality monitoring tool from **Amazon EC2** to **AWS Lambda** and built logical datasets to administer quality monitoring on snowflake warehouses.
* Implemented and setup **AWS shield, AWS, config, Amazon Macie,** and **Amazon inspector** for security and protection of sensitive data.
* Automation of cloud infrastructure using **Terraform**, and application configuration and deployment.
* Creating and managing access to AWS services for IAM user accounts and for role-based users.
* Using **Tableau**, designing dashboard to show operational metrics.
* Hands-on experience integrating AWS services: **EC2, S3, Network Protocol, Transit VPC, VPC Peering, VPC Endpoints, VPC Private Link.**
* Developed multiple POC’s using **Spark, Scala** and deployed on the Yarn Cluster, compared the performance of Spark, with Hive and SQL.
* Used **Amazon Elastic Cloud Compute (EC2)** infrastructure for computational tasks and **Simple Storage Service (S3)** as storage mechanism.
* Evaluated **Snowflake** design considerations for any change in the application.
* Implemented Installation and configuration of the multi-node cluster on Cloud using **AWS.**
* Experienced with **Spark Streaming** and **AWS Kinesis** for real-time data processing.
* Configured the services **S3, AWS Glue, EC2** using python **Boto 3**.
* Implemented ETL pipelines in **AWS Glue** for various spark transformations.
* Involved in **PL/SQL** query optimization to reduce the overall run time of stored procedures.
* Worked on **Kibana** dashboards based on log stash data and integrated several source and target systems into **Elastic search** for near real-time log analysis of end-to-end transaction monitoring.
* Integrated **Apache Airflow** with AWS to monitor multi-stage machine learning processes with **Amazon Sage Maker** jobs.

**Environment:** Hadoop, Hive, Map Reduce, Kafka, Spark, HDFS, Python, Horton works Data platform, CDH4, AWS, Apache NiFi, Oozie, Hbase, JSON, CSV, XML, Hive, Sqoop, Pig, MySQL, Jira

**Data Analyst/Power BI Developer (Intern / Trainee) June 2014 to Dec 2015**

**Client: BlueJeans Network**

**Project Description:** Company gets data from different sources like Tonga/BPS/PAM/FASL/Intrader in the form of flat files and DB pulls. The primary aim of this project is to integrate the data sources from existing system and to provide analysis to make business related decisions.

**Key Contributions:** Developed Dashboard **Visualizations, cross tables, Bar charts, Waterfall, Tree map and complex reports** which involves custom Controls and Custom Expressions using bars, lines and pies, maps, scatter plots**, Gantts, bubbles, histograms, bullets, heat maps and highlight table**.

* Worked in Data Warehousing Informatica ETL using **Informatica Power Center, Teradata, Oracle**, **SQL Server** and Flat files.
* creating **mappings with various transformations, mapplets, SCD Type-1/2** and **reusable** objects using Informatica. · Created **Sessions, workflows** in Informatica.

**Responsibilities:**

* Utilized Spark SQL to perform advanced-level data extraction, data transformation, data management tasks providing on the go responses to some management questions by performing complex joins, queries.
* Extracted Transform and Load data from Sources Systems to **Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL, and U-SQL Azure Data Lake Analytics. Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW, Azure Synapse Analytics) and processing the data in In Azure Databricks.**
* Expert designer, coder and tester for data warehouse programming by using **Informatica Power Center, Informatica Developer Tool, Informatica Power Exchange (9.5.1 and above).**
* Involved in design, installation, administration and optimization of hybrid cloud components to ensure business continuity with **Azure AD**, **ADFS**, **SSO** & **VPN** Gateways.
* Led implementation of Office 365 and **Azure Active Directory** for single sign on, authentication, authorization and Azure Role-based Access Control (RBAC).
* Responsible for fully documenting, managing library of source code, algorithms for future use.
* Developed, tested hypotheses (t-test, F-test) using R to support research, product offerings and communicate findings to data reports/ visualization in a clear, precise, actionable manner using tools such as Looker, Power BI and Tableau.
* Extensive practical knowledge in importing data for use in reporting software (Looker/Tableau) graphs.
* Pulling ad-hoc reporting SSRS, sales force customize to client’s need by using Alteryx, advance Excel.
* Creating and maintaining various reports, tracking spreadsheets, and on-demand ad-hoc report requests, as needed, and within specific timeframes. Experience in developing Tableau and Looker dashboards.
* created new OLAP data mart for more concise data and more efficient report performance.
* **Created KPI, Partitions, Perspectives, and Calculated Measures, used DAX queries** in Power BI.

**Environment:** Power BI, Microsoft Dynamics CRM, Informatica Power Center 9.6.1/9.5, Power Mart, Teradata SQL assistant, PL/SQL, Tidal, Tableau, Hadoop, SAP BO, SSRS, Windows, IDQ, IDE, SDLC, UNIX shell scripting, AGILE,Erwin, TOAD, Oracle, Autosys, Business objects

**Master’s Course Work**

* **Analysis Tools: -** Advanced Excel (VBA, VLOOKUP, Macros), Power BI, Tableau, Adobe Analytics (Omniture), Google Analytics
* **Programming/Database:** - Python, R, Advanced SQL, MySQL, PostgreSQL, snowflake, C, C++, HTML,
* **Core Skills:** - Data Warehousing, Data Mining, Data Visualization, Requirement Gathering, Project
* **Software Skills:** - NetBeans, Android Studio, RStudio, Hadoop, STATA
* **Cloud: AWS**, Azure Python for Finance. Financial Analytics in Spreadsheets Machine Learning.
* Data Visualization. Cloud Computing.