**Manikanta Borra**



980(202)-2939  
[**Manikantaborra.workmail@gmail.com**](mailto:Manikantaborra.workmail@gmail.com)

**SUMMARY**

* Having around 10+ years of experience in Data Engineering, my experience encompasses a broad range of cloud platforms, including GCP, AWS, and Azure, with specialized expertise in the GCP Cloud Data Platform. I am also skilled in a variety of Big Data Ecosystem technologies, such as Hadoop, Map Reduce, Pig, Hive, and Spark, and have a strong background in data visualization, reporting, and data quality solutions.
* Extensive experience with different phases of the project (project initiation, project requirement, and specification gathering, designing system, coding, testing, and debugging existing client-server-based applications).
* Experience in AWS Cloud Computing services, such as EC2, S3, EBS, VPC, ELB, Route53, Cloud Watch, Security Groups, EKS, IAM, Cloud Front, RDS, and Glacier.
* Extensively involved through the Software Development Life Cycle (SDLC) from initial planning through implementation of the projects.
* Experienced with Google Cloud Platform (GCP) and all the big data products, including BigQuery, Cloud Dataproc, Google Cloud Storage, and Composer (Air Flow as a service).
* Strong proficiency in SQL concepts, Presto SQL, Hive SQL, Python (Pandas, NumPy, SciPy, Matplotlib), Scala and Spark to handle large volumes of data.
* Skilled in Shell/Bash scripting and building data pipelines on Unix/Linux systems.
* Proficient in various Hadoop ecosystem components such as HDFS, YARN, MapReduce, Spark, Pig, Sqoop, Hive, Impala, HBase, Kafka, and Crontab tools.
* Knowledgeable in various file formats in HDFS, such as Avro, ORC, and Parquet.
* Experience in using Stack driver service/Dataproc clusters in GCP for accessing logs for debugging.
* Proficient in using Hortonworks on-premises Hadoop distribution and have written Hive SQL and Sqoop scripts.
* Experience in GCP Dataproc, GCS, Cloud functions, BigQuery, Azure Data Factory, and DataBricks.
* Proficient use of Bash, Python, and Shell scripts for automating builds and deployment procedures, with a focus on CI/CD and AWS Cloud Architecture.
* Knowledge of performing case studies, functional analysis, and data flow and object modeling.
* Hands-on experience in designing end-to-end ETL Strategy utilizing SSIS as an ETL tool and authored a multitude of SSIS packages for data migration purposes.
* Experience in working on object-oriented analysis concepts including implementation, design, and programming technical skills.
* Expertise in ETL tools such as Redpoint Data Management,Redpoint Ineractions,IBM Data Stage and Neolane
* Data Engineering, Data Modeler, Data Science, Data Management role in
* Leadership in Data Management. Built complete Data Management, Big Data Migration, Data Governance team from ground up. Understand all file feed data issues and resolution around it.
* Implemented Big Data integration solutions with Talend, handling diverse data sources and formats.
* Hands-on experience on Google Cloud Platform (GCP) in all the big data products BigQuery, Cloud Data Proc, Google Cloud Storage, and Composer (Air Flow as a service).
* Hands-on experience with big data technologies such as Spark, Hive, Hadoop, HDFS, etc.
* Experience with building data pipelines in Python/Pyspark/HiveSQL/Presto/BigQuery and building python DAG in Apache Airflow.
* Experience working on Azure Services like Data Lake, Data Lake Analytics, SQL Database, Synapse, Databricks, Data factory, Logic Apps and SQL Data warehouse and GCP services Like Big Query, Dataproc, Pub sub etc.
* Worked with SAS Enterprise guide and have written numerous passes through Hive SQL queries.
* Diverse experience in all phases of the software development life cycle (SDLC) especially in the Analysis, Design, Development, Testing, and Deploying of applications.
* Utilized Spark, Scala, and Python for querying, and preparing from big data sources.
* Worked with different application teams on gathering relevant data and prepared data visualization and storylines for upper-level management.
* Expertise in tools such as Tableau, SQL, etc. for inspecting information, preparing dashboards, and making recommendations.
* Possess an excellent understanding of Quality Assurance processes and SDLC.
* Proficient in designing data pipelines that can capture data from streaming web data as well as RDBMS source data.
* Merit of building the ETL architecture and source to target mapping to load data into a data warehouse.
* Experience in manipulating data for data loads, extracts, statistical analysis, and modeling.

# TECHNICAL SKILLS:

|  |  |
| --- | --- |
| **Big-Data Technologies** | Hadoop, MapReduce, HDFS, Sqoop, PIG, Hive, HBase, Oozie, Flume, NiFi, Kafka, Zookeeper, Yarn, Apache Spark, Mahout, Sparklib |
| **Databases** | Snowflake, Oracle, MySQL, SQL Server, PostgreSQL, Teradata, Spark-Redis, DB2 |
| **NoSQL Databases** | HBase, Cassandra, MongoDB, DynamoDB, and Cosmos DB |
| **Hadoop Core Service** | HDFS, Map Reduce, Spark, YARN |
| **Hadoop Distribution** | Apache Hadoop 2.x/1.x, Cloudera CDP, Hortonworks HDP, Amazon EMR (EMR, EC2, EBS, RDS, S3, Athena, Glue, Elasticsearch, Lambda, DynamoDB, Redshift, ECS, Quicksight) Azure HDInsight (DataBricks, DataLake, Blob Storage, Data Factory ADF, SQL DB, SQL DWH, Cosmos DB, Azure AD) |
| **Programming** | C, Scala, Python, R, SQL, PL/SQL,Pyspark, Pig Latin, HiveQL, Unix, Shell Scripting. |
| **Cloud Technologies** | AWS, Microsoft Azure, GCP |
| **GCP Cloud Platform** | Big Query, Cloud Data Proc, GCS Bucket, G-Cloud Function, Apache Beam, Cloud Shell, GSUTIL, BQ Command Line, Cloud Data Flow |
| **Data Services** | Hive, Pig, Impala, Sqoop, Flume, Kafka |
| **Monitoring Tools** | Cloudera Manager |
| **Versioning tools** | SVN, Git, GitHub, GitBucket |
| **ETL/BI** | Tableau, ggplot2, matplotlib, SSRS, Power Bl, Informatica, SSIS, SSAS, QlikView, Erwin, and Arcadia |
| **Database Modelling** | Dimension Modeling, ER Modeling, Star Schema Modeling, Snowflake Modeling |
| **Machine-Learning Techniques** | Linear & Logistic Regression, Classification and Regression Trees, Random Forest, Associative rules, NLP and Clustering |
| **Operating Systems** | Windows 7/8/XP/2008/2012, Ubuntu Linux, MacOS, Unix |
| **Build Tools** | Jenkins, Maven, ANT, Apache Airflow |
| **Development Tools** | Eclipse, NetBeans, Microsoft SQL Studio, Toad |
| **Methodologies** | RAD, JAD, UML, System Development Life Cycle (SDLC), Jira, Agile, Confluence, and Waterfall Model |

**PROFESSIONAL EXPERIENCE:**

**CVS HEALTH, Irving, TX Sep 2024 to Present   
Role: Senior GCP Data Engineer**

**Responsibilities:**

* Developed and maintained full life cycle ETL solutions for high-volume data pipelines, integrating structured and unstructured healthcare datasets across multiple systems.
* Worked extensively with **delimited** and **EBCDIC multi-record file sources**, implementing **encryption/decryption** routines to handle sensitive data securely in compliance with **HIPAA** protocols.
* Migrate Teradata to BigQuery using optimized data extraction, transformation, and loading (ETL/ELT) strategies.
* Migrate data from Oracle, RDBMS, fixed-width files, and EBCDIC encrypted files to BigQuery.
* Convert Teradata BTEQ scripts to BigQuery SQL, ensuring compatibility and performance improvements.
* Optimize data ingestion pipelines for scalability, performance, and cost efficiency.
* Develop ETL/ELT workflows using Dataflow, Dataproc, and BigQuery.
* Ensure business logic and data quality are maintained across all transformations.
* Optimize BigQuery SQL queries for better performance, reducing execution time and cost.
* Design and manage end-to-end data pipelines using Cloud Composer (Apache Airflow).
* Develop and execute DAGs (Directed Acyclic Graphs) in Airflow to ensure seamless data movement and processing.
* Schedule and manage Tidal jobs to automate data ingestion and transformation tasks.
* Automate data processing pipelines using SQL, and GCP services to improve efficiency.
* Deploy changes to pre-prod and production environments every sprint following DevOps best practices.
* Implement CI/CD pipelines using Cloud Build and Terraform for automated deployments.
* Ensure version control and rollback strategies are in place to prevent disruptions.
* Work in a sprint-based Agile environment for iterative development and continuous improvement.
* Collaborate closely with BI teams, software engineers, and business stakeholders.
* Participate in sprint planning, stand-ups, retrospectives, and code reviews to ensure seamless project execution.
* Monitor data pipelines and job executions using Cloud Logging, Cloud Monitoring, and Stack driver.
* Proactively troubleshoot pipeline failures, performance bottlenecks, and query optimization issues.
* Implement BigQuery optimizations like partitioning, clustering, and materialized views to improve performance and reduce costs.
* Processed and transformed files from various sources including DT Share Channel files, NIX tables, denormalized tables, and drug reference datasets to support pharmacy and claims processing systems.
* Designed and implemented custom file parsers and loaders to process hierarchical, flat, and delimited file formats in alignment with healthcare data in ingestion standards.
* Developed reusable components in Python, C#, and Java for data validation, enrichment, and transformation, enhancing code efficiency and maintainability.
* Wrote complex SQL queries and stored procedures to extract, validate, and transform data from DBMS systems, supporting downstream reporting and analytics.
* Implemented audit and error-handling frameworks to monitor production workflows and reduce downtime by ensuring transparency and traceability.
* Collaborated with cross-functional teams including QA, analysts, and architects to gather requirements and deliver scalable, high-performance data solutions.

**Environment:** Red Point Data Management, Python, C#, Java, SQL, Teradata, Oracle, EBCDIC, Delimited & Hierarchical File Formats DBMS Systems, Informatica, HIPAA Compliance, DT Share Channel Files, NIX Tables, Denormalized Tables, Jenkins, Git, Linux/Unix, ServiceNow, Visual Studio, Healthcare Claims Systems, SQL Server, Stored Procedures, Audit/Error Logging Frameworks, Agile/Scrum.

**PAYPAL, SAN JOSE, CA Feb 2023 to Sep 2024**

**Role: LEAD GCP Data Engineer**

**Responsibilities:**

* Designed and implemented a highly scalable, real-time data ingestion pipeline using GCP Dataflow to ingest and process over 1 million events per second from IoT devices into a Databricks Delta Lake.
* Developed Databricks notebooks using pyspark and Spark Structured Streaming to process and transform real-time data streams, enriching data with machine learning models deployed on GCP AI Platform.
* Implemented monitoring and alerting mechanisms using Stack driver, enabling proactive issue identification and resolution in GCP data pipelines.
* Built data integration workflows using Red Point Data Management, streamlining data ingestion, cleansing, and transformation processes.
* Led the migration of large-scale data warehouses from data to Google Cloud Platform (GCP), utilizing BigQuery for seamless data transition and ensuring minimal downtime.
* Enhanced existing Big Data processing workflows by integrating Java-based applications with Apache Kafka for real-time data streaming, reducing latency and increasing throughput by 30%.
* Proficient in writing complex SQL queries for data manipulation and analysis. Experience with relational databases such as Oracle and PostgreSQL.
* Skilled in designing and managing NoSQL database applications with MongoDB. Familiar with performance tuning and database optimization.
* Designed and implemented robust database solutions using SQL and MongoDB, enhancing data retrieval efficiency by 30%.
* Designed and developed Java programs for efficient data ingestion, cleansing, and integration processes, ensuring high data quality and availability for analytics purposes in a healthcare analytics platform.
* Created customized Java applications tailored to specific business needs, such as predictive maintenance and customer behavior analytics, which helped increase customer retention rates by 15% in a retail environment.
* Utilized Java to implement sophisticated data processing algorithms, enabling complex event processing and real-time decision-making in a high stakes trading platform.
* Worked with Python, Azure ADF, IoT, Event Hub, Cosmos DB, SQL DB, Snowflake, and Bash scripts to develop custom data transformations and data quality rules, resulting in a 25% reduction in data processing errors.
* Expertise in designing and deploying Hadoop clusters and different Big Data analytic tools, including Pig, Hive, Sqoop, and Apache Spark with Cloudera Distribution.
* Used Redpoint Interaction to generate automatic emails to the various customers on daily and weekly basis.
* Most notable clients include Technology Crossover Ventures, Redpoint Ventures, VantagePoint Venture capital and Jafco Ventures.
* Developed and maintained CI/CD pipelines on GCP using Cloud Build and Cloud Run, enabling seamless code deployment and testing in a controlled environment.
* Implemented data versioning and lineage tracking using tools such as Data Catalog and Data Studio, enabling auditability and traceability of healthcare data in GCP.
* Conducted capacity planning and scaling of GCP data pipelines using Kubernetes and Cloud Autoscaling, ensuring optimal performance and cost-efficiency.
* Developed multi-cloud strategies in better using GCP (for its PAAS).
* Designed and developed Spark jobs with Scala to implement end-to-end data pipelines for batch processing.
* Developed data pipeline using Flume, Kafka, and Spark Stream to ingest data from their weblog server and apply the transformation.
* Developed data validation scripts in Hive and Spark and perform validation using Jupiter Notebook by spinning up the query cluster in EMR.
* Utilized Spark SQL API in Pyspark to extract and load data and perform Azure ADF, IoT, Event Hub, Cosmos DB, SQL DB and Snowflake.
* Developed Pyspark script to encrypt the raw data by using hashing algorithms concepts on client-specified columns.
* Developed Stored Procedures, Views, and Triggers, and was responsible for the design, development, and testing of the database.
* Developed Python-based API (RESTful Web Service) to track revenue and perform revenue analysis.

**Environment:** GCP, Gcs Bucket, G-Cloud Function, Databricks, Apache Beam, Pyspark, Cloud Dataflow, Cloud Shell, GSUTIL, Cloud SQL, Big Query, Cloud Data Proc, GCS, Cloud Composer, Talend for Big Data, Airflow, Hadoop, Hive, Teradata, SAS, Teradata, Spark, Python, SQL Server, Kubernetes, Docker, Java.

## KPMG, Atlanta , GA July 2021 to Jan 2023

## Role: GCP Data Engineer

**Responsibilities:**

* Architected a cloud-native, real-time data warehousing solution on GCP, ingesting data from multiple sources into a Databricks Delta Lake using Dataflow and querying it using Databricks SQL for ad-hoc analytics.
* Created Databricks notebooks using SQL, Python, and automated notebooks using Databricks Jobs to perform real-time data transformations and load data into BigQuery for low-latency reporting
* Enhanced query efficiency by implementing partitioning and clustering strategies on high-volume tables within **BigQuery**. Established integrations between applications, with a primary focus on Salesforce connectivity.
* Worked on Alert analysis, reduction, and tunning of **Datadog** alerts for Infrastructure efficiency.
* Diligently worked with Kafka Admin team to set up Kafka cluster and implemented **Kafka** producer and consumer applications on Kafka cluster setup with help of **Zookeeper**.
* Implemented **DevOps** and **CI/CD** practices using GIT, GitLab, DevOps, **Docker**, and **Kubernetes**, streamlining development workflows and enhancing collaboration.
* Leading the testing efforts in support of projects/programs across a large landscape of technologies ( Unix, Angular JS, , sause LABS, Cucumber JVM, Mongo DB, GIT Hub, Bit Bucket, SQL, NoSQL database, API, Java, Jenkins
* Developed under the scrum methodology and in a **CI/CD** environment using **Jenkins**
* Worked on implementing scalable infrastructure and platform for large amounts of data ingestion, aggregation, integration, and analytics in **Hadoop** using Spark and **Hive**.
* Built a system for analyzing the column names from all tables and identifying personal information columns of data across on-premises Databases (data migration) to GCP.
* Worked on writing **terraform** scripts from scratch for building **Dev, Staging, Prod, and DR** environments.
* Used **Kafka HDFS Connector** to export data from Kafka topic to HDFS files in a variety of formats and integrated with Apache Hive to make data immediately available for SQL querying.
* Utilized the **GCP Console** to monitor and manage resource usage, permissions, and access controls.
* Used **Cloud shell SDK in GCP** to configure the services Data Proc, Cloud Storage, and Big Query.
* Worked on GCP for the purpose of data **migration from Oracle database to GCP.**
* Uploaded and downloaded data to and from **Cloud Storage** using the command-line tools, and client libraries.
* Developed **pyspark** scripts to handle the migration of large volumes of data, ensuring minimal downtime and optimal performance and analyzed the SQL scripts, and designed solutions to implement using pyspark.
* Built **Tableau** dashboards with ODBC connections from various sources, including BigQuery and **Presto SQL engine**.
* Worked on querying data using **Spark SQL** on top of **pyspark** engine jobs to perform data cleansing, validation, and applied transformations and executed the program using Python API.
* Process and load bound and unbound Data from Google Pub/Subtopic to Big Query using cloud Dataflow with Python.
* Worked on partitions of **Pub/Sub** messages and setting up the replication factors.
* Developed **T-SQL** **(SQL)** queries, stored procedures, user-defined functions, and built-in functions.
* Migrated previously written cron jobs to airflow/composer in GCP.
* Used to write **Python DAGs** in **airflow** which orchestrate end-to-end data pipelines for multiple applications.
* Led incident response and postmortem analyses using **Cloud Logging**, identifying root causes and implementing preventive measures to enhance system reliability.
* Used **windowing functions** to order data and remove duplicates in source data before loading to DataMart for better performance.
* Worked on importing and exporting data from **Oracle** into **HDFS** and HIVE using Sqoop for analysis, visualization, and generating reports.
* Created **Hive** tables using HiveQL, then loaded the data into Hive tables and analyzed the data by developing Hive queries.
* Used **Oozie** Scheduler systems to automate the pipeline workflow and orchestrate the **map-reduce** jobs that extract and **Zookeeper** for providing coordinating services to the cluster.
* Used **SFTP** to generate detailed logs for file transfers, user activities, and file access.
* Worked on **NoSQL** Databases such as **HBase** and integrated with pyspark for processing and persisting real-time streaming.
* Used **Talend** to load transformed data into BigQuery and implemented data quality checks and data governance rules.
* Developed **Power Pivot/SSRS (SQL Server Reporting Services)** Reports and added logos, pie charts, and bar graphs for display purposes as per business needs.
* Regularly Worked on Data serialization formats for converting complex objects into sequence bits by using **Parquet, Avro, JSON, and CSV formats.**
* Also developed internal dashboards for the team using **Power** **BI** tools for tracking daily tasks.

**Environment:** GCP Console, Cloud Storage, Databricks , Big Query, CI/CD, PostgreSQL, Salesforce, Datadog, DevOps, Zookeeper, Terraform, Mongo DB,Jenkins, Tableau, Data Proc, Spark, Hadoop, Hive, Scala, Cloud SQL, Shell Scripting, SQL Server 2016/2012, T-SQL, SSIS, Visual Studio, Power BI, PowerShell, Oracle, Teradata, Airflow, GIT, Docker.

## M & T bank, Buffalo NY. Jan 2020 to June 2021

## Role: Data Engineer

**Responsibilities:**

* Worked on GCP's big data stack, including **Dataproc**, Dataflow, **Pub Sub**, GCS, Cloud Functions, Big Query, Stack driver, Cloud Logging, IAM, and Data Studio. Part of the **ETL** development team with a primary focus on supplying quality data to business users.
* Managed continuous integration and deployment using **DevOps**, improving efficiency and consistency.
* Experience in identifying and resolving production bugs using **Stack driver** logs in GCP.
* Designed data transformation and enrichment processes within **IBM** Streams to enhance raw data with contextual information before storing it in Google Cloud.
* Proficient in maintaining infrastructure across multiple projects in Google Cloud Platform using **Terraform**.
* Expertise in optimizing **Big Query** for reporting with **Tableau**, employing techniques like partitioning and scenario testing.
* Involved in Building **ETL** to **Kubernetes** with **Apache** **Airflow** and **Spark** in GCP.
* Optimized data warehousing architecture for enhanced performance, resulting in improved query response times and cost reduction.
* Experience with **Kubernetes** in GCP, designing reports in **Data Studio**, and developing monitoring techniques using Stack driver’s log router.
* Implemented Apache Airflow on GCP Cloud Composer to schedule and monitor real-time data pipelines, ensuring reliable and timely data delivery from Databricks Delta Lake to downstream systems like BigQuery.
* Leveraged Databricks Delta Live Tables to build and orchestrate real-time data pipelines, enabling continuous data processing and updates to target data marts and cubes.
* Developed under the scrum methodology and in a **CI/CD** environment using Jenkins.
* Proficiency in using Big Data related services in Google Cloud Platform like **Presto**, **Hive**, **Spark-SQL**, and Big Query, utilizing Python client libraries for efficient analytics.
* Developed a framework to generate daily **ad-hoc reports** and extracts from enterprise data in BigQuery, facilitating efficient data retrieval.
* Hands-on experience using **Apache Airflow** in GCP Composer environment to build data pipelines, working with various airflow operators.
* Experience building data pipelines with **Cloud Composer**, **Cloud Dataflow**, and Cloud **Data prep**.
* Set up proactive monitoring and alerting for **data integration** and **warehousing** pipelines using GCP's monitoring and logging tools.
* Contributed to the migration of microservices to Google Cloud Platform (**GCP**) and application migration to **microservices**.
* Utilized **Google Data Catalog** and other **GCP APIs** for monitoring, query, and billing-related analysis for BigQuery usage.
* Programmed to download SQL Dumps from equipment maintenance sites and loaded them into **GCS buckets**.
* Processed and loaded bound and unbound data from **Google Pub/Sub** to **BigQuery** using Cloud Dataflow with **Python**.
* Utilized **Confluence** and **Jira** for collaboration and **Git** for version control.
* Evaluated data profiling, cleansing, integration, and extraction tools (e.g., Informatica).
* Monitored daily, weekly, and monthly jobs and provided support in case of failures or issues.
* Used **Python** SDK with Apache Beam framework to stream **PubSub** messages into **BigQuery** tables.
* Worked on a self-service analytics platform using **Apache Druid/Superset,** enabling engineering and business teams to make timely product decisions.
* Worked on solutions for high-volume data stream ingestion, processing, and low-latency data provisioning using **Hadoop** Ecosystems.

**Environment:** GCP, GCS Bucket, G-Cloud Function,Databricks, CI/CD, DevOps Apache Beam, Jenkins, Cloud Dataflow, Big Query, Cloud Dataproc, GCS, GKE, Cloud Composer, Airflow, Hadoop, Hive,Tableau**,** Spark, Python, SQL, Kubernetes, Jira, Pub/Sub.

**Dollar General- Goodlettsville, TN June 2018 to Dec 2019**

**Role: Data Engineer**

**Responsibilities:**

* Extensively involved in Installation and configuration of Cloudera Hadoop Distribution.
* Implemented advanced procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in Scala.
* Developed spark applications for performing large scale transformations and denormalization of relational datasets.
* Integrated Databricks with GCP Stack driver for comprehensive monitoring and alerting of real-time data pipelines, enabling proactive issue detection and resolution.
* Developed custom monitoring and observability solutions using Databricks and GCP Cloud Functions to stream application logs and metrics to Pub/Sub and BigQuery for real-time analysis and debugging
* Created reports for the BI team using Sqoop to export data into HDFS and Hive.
* Implemented scheduled downtime for non-prod servers for optimizing AWS pricing.
* Optimized the Hive tables using optimization techniques like partitions and bucketing to provide better performance with HiveQL queries.
* Did Aggregations and analysis on large set of log data, collection of log data done using custom built Input Adapters.
* Evaluated data import-export capabilities, data analysis performance of Apache Hadoop framework.
* Involved in installation of HDP Hadoop, configuration of the cluster and the eco system components like Sqoop, Pig, Hive, HBase and Oozie.
* Created HBase tables to load large sets of structured, semi-structured and unstructured data coming from UNIX, NoSQL and a variety of portfolios.
* Tested raw data and executed performance scripts and Assisted with data capacity planning and node forecasting.
* Created reports for the BI team using Sqoop to export data into HDFS and Hive.
* Worked extensively with Sqoop for importing and exporting the data from HDFS to Relational Database system and vice-versa.
* Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports for the Data science team.
* Optimized Map/Reduce Jobs to use HDFS efficiently by using various compression mechanisms.

**Environment:** Hadoop, Hive, PIG, Sqoop, Kafka, AWS EMR, Databricks,AWS S3, AWS Redshift, Oozie, Flume, HBase, Hue, HDP, IBM Mainframes, HP Non Stop and RedHat 5.6,Kafka.

**Neo App Solutions, Hyderabad Sep 2013 to Dec 2017**

**Role: Data Analyst**

**Responsibilities:**

* Designed Business Intelligent dashboards using Tableau, Power BI and Excel to support the data quality and revenue initiatives.
* Developed complex Excel templates to improve the response time of the support requests from the business users.
* Experienced in maintaining and supporting operational reports, dashboards, and scorecards using Microsoft Power BI (Power Pivot, T-SQL, Excel pivot tables).
* Refactored the data model in Power BI reports.
* Created various metrics such as process metrics, risk metrics, and performance metrics. Identified metrics to be upgraded to KPIs and KRIs.
* Created enterprise-wide Power BI / Tableau dashboards and Excel reports that provided actionable insights which was used by the executives to make data driven strategic decisions.
* Developed database objects, including tables, views and materialized views using SQL.
* Designed Tableau visualization using various charts such as Cross tabs, Heat Maps, Pie charts, Bar charts, Geographic Maps.
* Analyze data and its trends from Salesforce (CRM) and build reports over the data.
* Experienced in developing and programming skills using SQL queries for Stored Procedures, Triggers, Functions, and Packages using SQL.
* Experience in Data transformation and Data mapping from source to target database schemas and data cleansing.
* Strong experience in Data Analysis, Data Cleansing, Data Import, and Data Export using Python.
* Used Python for the use cases of predictive analytics.
* Used packages like Pandas, NumPy and matplotlib to analyze the patterns for Data Exploration.
* Perform multiple research activities to troubleshoot/fix data inconsistencies; wrote multiple SQL queries to apply data fixes.

**Environment:** Tableau, Power BI, SQL, Python, and MS Suite.