**SUDHEESH NAKKALA**

**Data Engineer**

**Sudheeshnakkala4@gmail.com**

**(510) 945-5710**

**PROFESSIONAL SUMMARY:**

* Over 10+ years of experience in areas including **Data Analyst, Statistical** **Analysis, Machine Learning, Deep** **Learning with** large data sets of structured and unstructured data in travel services, strong functional knowledge, business processes, and latest market trends and manufactory industries.
* Developed predictive models using **Decision Tree, Random Forest, Naïve Bayes, Logistic Regression**, **and Cluster** **Analysis**.
* Involved in the entire data science project life cycle and actively involved in all the phases including data extraction, data cleaning, statistical modeling, and data visualization with large data sets of structured and unstructured data, created ER diagrams and schema.
* Experience with publishing reports to Power BI Service, setting up dashboards, and managing user access and permissions.
* Expert in the entire Data Science process life cycle including Data Acquisition, Data Preparation, Data Manipulation, Feature Engineering, **Machine Learning** **Algorithms**, **Validation, and Visualization**.
* Strong knowledge in Statistical methodologies such as **Hypothesis Testing**, Principal Component Analysis (**PCA**), Sampling Distributions and Time Series Analysis.
* Extensively worked on Data preparation, exploratory analysis, Feature engineering using supervised and unsupervised modeling.
* Integrating **Snowflake** with tools like Apache Airflow, dbt (Data Build Tool), and Talend for scheduling and orchestrating workflows.
* Skills in optimizing reports and dashboards for performance, including understanding query folding and managing large datasets.
* Experienced in **Machine Learning, Data mining** with large datasets of Structured and Unstructured **Data, Data Acquisition, Data Validation, and Predictive Modeling.**
* Proficient in Python and its libraries such as **NumPy, Pandas, Scikit-Learn, Matplotlib and Seaborn**.
* Expert in prepossessing data in **Pandas** using visualization, data cleaning and engineering methods such as looking for Correlations, Imputations, Scaling and Handling Categories
* Experience in building various machine learning models using algorithms such as Linear Regression, Gradient Descent, Support Vector Machines (**SVM**), Logistic Regression, **KNN**, Decision **Tree**, Ensembles such as Random Forrest, **AdaBoost**, Gradient Boosting Trees.
* Experienced the full software lifecycle in **SDLC**, Agile, and Scrum methodologies.
* Strong **SQL** programming skills, with experience in working with functions, packages, and triggers.
* Excellent understanding of machine learning techniques and algorithms, such as **k-NN, Naive Bayes, SVM**, Decision Forestretch.
* **Snowflake** integrates with many popular ETL tools like Fivetran, Matillion, Talend, and Informatica.
* Experienced in Data Integration Validation and Data Quality controls for **ETL** process and **Data Warehousing** using **MS Visual Studio SSIS, SSAS, and SSRS**.
* Expert in developing **Data Conversions**/Migration from Legacy System of various sources (flat files, **Oracle, Non-Oracle Database)** to **Oracle** system Using **SQL LOADER**, External table and Calling Appropriate Interface tables and **API's** Informatica.
* Drive core insights from available data to suggest **A/B tests** that drive improvements in site execution and experience
* Good Knowledge and experience in deep learning algorithms such as Artificial Neural network (**ANN**), Convolutional Neural Network (**CNN**) and Recurrent Neural Network (**RNN**), **LSTM** and **RNN** based speech recognition using Tensor Flow.
* Working knowledge on **Azure** **cloud** components (**HDInsight**, **DataBricks**, **DataLake**, **Blob Storage**, **Data Factory**, **Storage Explorer**, **SQL DB**, **SQL DWH**, **CosmosDB**).
* Experienced in developing Data Pipelines in **Azure Data Factory** and Datasets/pipelines during ETL process from **Azure SQL, Blob** **Storage, and Azure SQL Datawarehouse**.
* Possesses hands on experience **in Cloudera Hadoop, Hadoop**, various **ETL** tools, Cassandra, and various Confidential IaaS/PaaS services.
* Excellence in handling **Big Data** Ecosystems like **Apache Hadoop, MapReduce, Spark**, HDFS Architecture, Cassandra, HBase, Sqoop, Hive, Pig, MLlib, ELT.
* Expertise in writing **Spark RDD** transformations, actions, Data Frames, case classes for the required input data and performed the data transformations using **Spark-Core**, **Spark SQL**.
* Expertise in building **PySpark and Spark-Scala** applications for analysis, batch processing, and stream processing.
* Good understanding of **Big Data Hadoop** and **Yarn** architecture along with various **Hadoop** Demons such as Job Tracker, Task Tracker, Name Node, Data Node, Resource/Cluster Manager, and **Kafka.**
* Experienced in writing **Spark** scripts in both **Python** and **Scala** for development and data analysis.
* Experience in analyzing data from multiple sources and creating reports with Interactive Dashboards using **power BI**, **Tableau and Matplotlib.**
* Experience in Importing and exporting data from different databases like **MS SQL Server, Oracle, Cassandra, Teradata**, **PostgreSQL** Post into **HDFS** using **Sqoop**.
* Experience in Developing ETL workflows using **Informatica** **PowerCenter 9.X/8.X** and IDQ. Worked extensively with Informatica client tools- Designer, Repository Manager, Workflow Manager, and Workflow Monitor.
* Well versed with the design and development of presentation layer for web applications using technologies like **HTML, CSS, jQuery**, and **JavaScript**.

**EDUCATION:**

Bachelor of Computer Science

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
|  **Big Data Ecosystem** | HDFS, MapReduce, Yarn, Spark, Kafka, Airflow, Hive, Pig. |
|  **Hadoop Distributions** | Apache Hadoop 2.x/1.x, AWS (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, CosmosDB, Azure AD). |
|  **Machine Learning** |  Linear Regression, Logistic Regression, Naïve Bayes, Decision Trees, Random Forest, SVM, KNN. |
|  **Programming Languages** | Python, Scala, Shell Scripting, Pig Latin, HiveQL. |
|  **NoSQL Database** | MongoDB 3.x, Hadoop HBase 0.98. |
|  **Database** | Snowflake, AWS RDS, Teradata, Oracle 9i/10g, MySQL 5.5/5.6/8.0, Microsoft SQL, Postgres SQL. |
|  **ETL/BI** | Snowflake, Informatica, Tableau.  |
|  **Reporting & Visualization** | Tableau 9.x/10.x, Matplotlib, Power BI. |
|  **Web Development** | JavaScript, HTML, CSS, Postman, Flask. |
|  **Operating systems** | Linux (Ubuntu, RedHat), Windows (XP/7/8/10).  |
| **OLAP/Reporting Tools** | SSRS, SSAS, MDX, Tableau, PowerBI |
| **Relational Databases** | SQL Server database 2014/2012/2008 R2/2005, Oracle 11g, SQL Server Azure, MS Access, Postgresql |
| **Other Tools** | MS Office 2003/2007/2010 and 2013, Power pivot, Power Builder, GIT, CI-CD, Jupyter Note Book. |
| **Data Warehousing & BI** | Star Schema, Snowflake schema, Facts and Dimensions tables, SAS, SSIS, and Splunk |
| **Platforms (O/S** | Red-Hat LINUX, Ubuntu, Windows NT/2000/XP. |
| **Security Management:** | Hortonworks Ambari, Cloudera Manager, Apache Knox, XA Secure, Kerberos. |
| **Web-technologies** | DHTML, HTML, XHTML, XML, XSL (XSLT, XPATH), XSD, CSS, JavaScript, SOAP, RESTful, Agile, Design Patterns |
| **Data Modeling** | Rational Rose, Erwin 7.3/7.1/4.1/4.0 |
| **Code Editors** | Eclipse, Intellij |

**PROFESSIONAL EXPERIENCE:**

**Client: Anthem Nov 2022-Till date**

**Role: Data Engineer**

**Description:** Upon an unfinished world, humanity struggles to survive in a savage environment full of diverse threats. Equipped with incredible Javelin exosuits, a faction known as the Freelancers seeks to tip the balance in mankind’s favor.

**Responsibilities:**

* Perform data manipulation, data preparation, normalization, and predictive modelling. Improve efficiency and accuracy by evaluating model in **Python** and **R**.
* Used **Python** and **R** for programming for improvement of model. Upgrade the entire models for improvement of the product.
* **Snowflake** supports structured, semi-structured (e.g., JSON, Avro, Parquet), and unstructured data storage.
* Writing complex **SQL** queries for validating the data against different kinds of reports generated by Cognos.
* Filtered the discovered boundaries by implementing a non-max suppression algorithm to achieve an optimal bounding box per identified object.
* Evaluated models using Cross-Validation, Log loss function, **ROC** curves and used **AUC** for feature selection and elastic technologies like **Elasticsearch**, **Kibana, etc**
* Implemented statistical modeling with **XGBoostmachine** **learning** software package using **Python** to determine the predicted probabilities of each model.
* Created master data for modeling by combining various tables and derived fields from client data and students LORs, essays and various performance metrics.
* Used **NumPy, SciPy, pandas, NTLK (Natural Language Processing Toolkit**), **Matplotlib** to build the model.
* Worked in Hadoop Environment using **pig, Sqoop, Hive, HBase** and detailed understanding of **MapReduce** programs
* Involved in integration of various relational and non-relational sources such as **Oracle, XML and Flat Files**.
* Designed and implemented scalable **Cloud Data** and **Analytical** architecture solutions for various public and private cloud platforms using **Azure**.
* Created Data Pipeline using Processor Groups and multiple processors using **Apache Nifi for Flat File, RDBMS** as part of a **POC** using **AmazonEC2**.
* **Cloud Data Platform:** Understand how Snowflake is built for the cloud and how it leverages cloud providers like AWS, Azure, or Google Cloud.
* Formulated several graphs to show the performance of the students by demographics and their mean score in different **USMLE** exams.
* Knowledge of integrating Datadog with various data technologies (e.g., AWS, Azure, Kafka, databases) to gather comprehensive insights.
* Application of various Artificial Intelligence (**AI**)/machine learning algorithms and statistical modeling like decision **trees,text** analytics, natural language processing(**NLP**), supervised and unsupervised, regression models for Predictive Analysis of data.
* Performed Data Cleaning, features scaling, features engineering using pandas and **NumPy** packages in python and build models using deep learning frameworks.
* Created deep learning models using **TensorFlow** and **Keras** by combining all tests as a single normalized score and predict residency attainment of students.
* Using Datadog to collaborate with other teams (like DevOps and SRE) to resolve issues and improve overall system reliability.
* Keeping up with new features and best practices, as Power BI frequently updates with new functionalities.
* Used XGB classifier if the feature is a categorical variable and **XGB** repressor for continuous variables and combined it using **FeatureUnion** and **FunctionTransfomer** methods of **Natural Language** **Processing**.
* Created data layers as signals to Signal Hub to predict new unseen data with performance not less than the static model build using deep learning framework.
* Worked with the Data Governance group in creating a custom data dictionary template to be used across the various business lines.
* Familiarity with integrating Power BI with other Microsoft tools (like Azure, Excel) and collaborating with team members.
* Create statistical models using distributed and standalone models to build various diagnostics, predictive and prescriptive solution.
* Interface with other technology teams to load (**ETL**), extract and transform data from a wide variety of data sources
* Provides input and recommendations on technical issues to Business & Data Analysts, **BI Engineers, and Data** Scientists.
* Developed **SQL** procedures to synchronize the dynamic data generated from **GTID** systems with the **Azure SQL** Server.
* Expertise in building Azure native enterprise applications and migrating applications from on-premises to Azure environments
* Responsible for maintenance and monitoring of production/test/dev systems running on MS Azure

**Environment:** Python 2.x,3.x, Hive, AWS, Linux, Tableau Desktop, Microsoft Excel, NLP, Deep learning frameworks such as TensorFlow, Keras, Boosting algorithms,DB2, R, Python, Visio, HP ALM, Agile.

**Client: united health care**  **Aug 2021- Oct 2022**

**Role: Data Engineer**

**Description:** ​​​​​​​​​​​​​​​​​​​​​When it comes to your health, everything matters. That’s why UnitedHealth care is helping people live healthier lives and making the health system work better for everyone

**Responsibilities:**

* Extensively involved in all phases of data acquisition, data collection, data cleaning, model development, model validation and visualization to deliver data science solutions.
* Extracted data from **SQL Server Database** copied into **HDFS** File system and used Hadoop tools such as Hive and Pig Latin to retrieve the data required for building models.
* Performed data cleaning including transforming variables and dealing with missing value and ensured data quality, consistency, integrity using **Pandas, NumPy**.
* Developed predictive models on large scale datasets to address various business problems through leveraging advanced statistical modeling, machine learning.
* Hypothesis Testing, Statistical Modeling, Quantitative and Qualitative Analysis, Statistical Computing Methods, Regression Analysis Hadoop, Spark.
* Implemented public segmentation using unsupervised **machine learning** algorithms by implementing K-means algorithm by using **PySpark** using data munging. Analyzed and solved business problems and found patterns and insights within structured and unstructured data.
* Using **Snowflake's** error handling mechanisms in stored procedures and data pipelines.
* Implemented advanced computer vision techniques like distortion correction, thresholding techniques, and the sliding window method to identify the lane markings to highlight the entire lane.
* Tested the algorithm in a video to ensure that the lane boundaries are accurately identified.
* Utilized a diverse array of technologies and tools as needed, to deliver insights such as **R, SAS, MATLAB, Tableau** and more.
* Ability to connect Power BI to various data sources, including databases, APIs, and cloud services.
* Detected near-duplicated news by applying **NLP** methods and developing machine learning models like label spreading and clustering.
* **Data Modeling Techniques:** Knowledge of normalized and denormalized designs (e.g., star schema, **snowflake schema**).
* Employed the output of the semantic segmentation to perform drivable space estimation in 3D, lane estimation and to filter errors in the output of the 2D object detectors.
* Prototyping and experimenting with ML algorithms and integrating into a production system for different business needs.
* Knowledge of creating effective and insightful visualizations. Understanding best practices for data visualization to communicate findings clearly.
* Preparing reports, extracts and integration of existing data as needed to support analysis, visualizations, and statistical models.
* Building and productionizing predictive models on large datasets by utilizing advanced statistical modeling, machine learning, or other data mining techniques
* Implemented Porter Stemmer (**Natural Language Tool Kit**) and **NLP** bag of words model (Count Vectorizer) to prepare the data.
* Implemented number of customer clustering models and these clusters are plotted visually using **Tableau** legends for the higher management.
* Connecting Datadog with data technologies (e.g., databases, cloud services, Kafka) to gather insights across the data stack.
* Familiarity with Power Query for data extraction, transformation, and loading (ETL) tasks. Being able to connect to various data sources and perform necessary transformations is crucial.
* Developed SQL procedures to synchronize the dynamic data generated from **GTID** systems with the **Azure SQL** Server.
* **Data Transformation (ETL/ELT):** Designing transformation pipelines in SQL using **Snowflake Scripting** or **Snowpipe** (automated data ingestion).
* Process automation using **Python/R scripts** with Oracle database to generate and write the results in the production environment on a weekly basis.
* Used **Data Quality** validation techniques to validate Critical Data elements (**CDE**) and identified various anomalies.
* Performing **Data Validation / Data Reconciliation** between the disparate source and target systems for various projects.
* Utilizing Datadog's collaboration features to share insights with cross-functional teams and improve overall system reliability.
* **Data Sharing:** **Snowflake’s** ability to securely share data across different Snowflake accounts without needing to replicate the data.
* Writing complex SQL queries for validating the data against different kinds of reports generated by Cognos.
* Provides input and recommendations on technical issues to Business & Data Analysts, BI Engineers, and Data Scientists.

**Environment**: SAS, R, MLIB, Python, Data Governance, MDM, MATLAB, Tableau, Azure SQL Server.

**Client: Xfusion Technologies USA** Apr 2020-July 2021

**Role: Data Engineer**

**Description:** ​​​​​​​​​​​​​​​​​​​​​ xFusion Technologies, Inc is a California-based company specializing in exceptional business-technology solutions through the powerful fusion of innovation and human ingenuity. For over 15 years, we've been a trusted partner, crafting bespoke technology solutions

**Responsibilities**:

* Worked on **AWS Data pipeline** to configure data loads from **S3** to into **Redshift**.
* Using **AWS** Redshift, I Extracted, transformed and loaded data from various heterogeneous data sources and destinations
* Created Tables, Stored Procedures, and extracted data using **T**-**SQL** for business users whenever required.
* Performs data analysis and design, and creates and maintains large, complex logical and physical data models, and metadata repositories using **ERWIN** and **MB MDR**
* I have written **shell script** to trigger data Stage jobs.
* Assist service developers in finding relevant content in the existing reference models.
* Like Access, Excel, CSV, Oracle, flat files using connectors, tasks and transformations provided by **AWS Data Pipeline.**
* Utilized **Spark** SQL API in **PySpark** to extract and load data and perform SQL queries.
* Worked on developing **Pyspark** script to encrypting the raw data by using Hashing algorithms concepts on client specified columns.
* Responsible for Design, Development, and testing of the database and Developed Stored Procedures, Views, and Triggers
* Developed **Python**-based **API** (RESTful Web Service) to track revenue and perform revenue analysis.
* Compiling and validating data from all departments and Presenting to Director Operation.
* **KPI** calculator Sheet and maintain that sheet within SharePoint.
* Created **Tableau** reports with complex calculations and worked on Ad-hoc reporting using **PowerBI**.
* Creating data model that correlates all the metrics and gives a valuable output.
* Worked on the tuning of **SQL** Queries to bring down run time by working on Indexes and Execution Plan.
* Performing **ETL** testing activities like running the Jobs, Extracting the data using necessary queries from database transform, and upload into the **Data warehouse** servers.
* Pre-processing using **Hive** and **Pig**.
* Explore log collection, processing, and analysis. Understand how to create log-based metrics and visualize log data.
* 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 In **Azure Databricks.**
* Implemented Copy activity, Custom **Azure Data** Factory Pipeline Activities
* Primarily involved in Data Migration using **SQL, SQL Azure, Azure Storage,**and**Azure Data Factory, SSIS, PowerShell.**
* Architect & implement medium to large scale **BI** solutions on Azure using **Azure Data** Platform services (Azure Data Lake, Data Factory, Data Lake Analytics, Stream Analytics, Azure SQL DW, HDInsight/Databricks, NoSQL DB).
* Utilize Datadog's official documentation and training resources, including tutorials and webinars.
* Migration of on-premise data (Oracle/ SQL Server/ DB2/ MongoDB) to Azure Data Lake and Stored (ADLS) using Azure Data Factory (ADF V1/V2).
* Developed a detailed project plan and helped manage the data conversion migration from the legacy system to the target **snowflake** database.
* Design, develop, and test dimensional data models using Star and Snowflake schema methodologies under the Kimball method.

**Environment:**SQL Server, Hive, Hadoop Cluster, ETL, Tableau, Teradata, Machine Learning (Logistic regression/ Random Forests/ Collaborative filtering), Git Hub, MS Office suite, Agile.

**Client: Advanced Technology Group Inc USA Nov 2018 to Mar 2020**

**Role: Data Engineer/Data Analyst**

**Description:** ​​​​​​​​​​​​​​​​​​​​​ At ATG, we are dedicated to delivering winning solutions on every project. The quality and integrity of the work performed by ATG is embodied by our entire organization, from senior management through our skilled field personnel. Our culture of responsiveness and ownership strengthens our business. It creates innovation, cost savings, quality improvements, safe environments,

**Responsibilities:**

* Responsible for the execution of**big data analytics**, predictive analytics and machine learning initiatives.
* Implemented a proof of concept deploying this product in**AWS S3 bucket** and **Snowflake**.
* Utilize AWS services with focus on big data architect /analytics / enterprise Data warehouse and business intelligence solutions to ensure optimal architecture, scalability, flexibility, availability, performance, and to provide meaningful and valuable information for better decision-making.
* Developed **Scala**scripts, **UDF**'s using both **data frames/SQL and RDD** in **Spark**for data aggregation, queries and writing back into S3 bucket.
* Experience in**data cleansing and data mining**.
* Wrote, compiled, and executed programs as necessary using Apache Spark in Scala to**perform ETL jobs** with ingested data.
* Used **Spark Streaming** to divide streaming data into batches as an input to Spark engine for batch processing.
* Wrote Spark applications for data validation, cleansing, transformation, and custom aggregation and used Spark engine,**Spark SQL** for**data analysis**and provided to the data scientists for further analysis.
* Prepared scripts to automate the ingestion process using **Python**and **Scala**as needed through various sources such as**API, AWS S3, Teradata and snowflake.**
* Designed and Developed Spark workflows using Scala for data pull from AWS S3 bucket and Snowflake applying transformations on it.
* Implemented**Spark RDD transformations** toMap business analysis and apply actions on top of transformations.
* Automated resulting scripts and workflow using **Apache Airflow** and **shell scripting** to ensure daily execution in production.
* Created scripts to read**CSV, json and parquet files**from S3 buckets in **Python**and load into **AWS S3, DynamoDB and Snowflake.**
* Implemented**AWS Lambda** functions to run scripts in response to events in Amazon DynamoDB table or S3 bucket or to HTTP requests using Amazon API gateway
* Migrated data from AWS S3 bucket to Snowflake by writing custom read/write**snowflake utility** function using Scala.
* Worked o**n Snowflake Schemas and Data Warehousing and**processedbatch and streaming data load pipeline using **Snow Pipe**and Matillion from data lake Confidential AWS S3 bucket.
* Profile structured, unstructured, and semi-structured data across various sources to identify **patterns in data and Implement data quality metrics**using necessary query’s or **python** scripts based on source.
* Install and configure**Apache Airflow** for S3 bucket and Snowflake data warehouse and created **dags**to run the Airflow.
* Created DAG to use the**Email Operator, Bash Operator and spark Livy operator** to execute and in **EC2**instance.
* Deploy the code to**EMR** via**CI/CD using Jenkins**

**Environment:** AgileScrum,MapReduce,Snowflake,Pig,Spark,Scala,Hive,Kafka,Python,Airflow,JSON,Parquet,CSV,Codecloud, AWS

**Client: Fusion Global Solutions, NY USA Oct 2016 – Sep 2018**

**Role: Data Engineer/Data Analyst**

**Description:**

Fusion Global Business Solutions was founded in Feb 1998 as a subsidiary of [Horizon Technology Group plc](https://en.wikipedia.org/wiki/Avnet) In July 2002 the company was subject to a management buyout led by then managing director Mark Lyttle. And now operates as a private limited company.

**Responsibilities**:

* Experience in building and architecting multiple Data pipelines, end to end **ETL** and **ELT** process for Data ingestion and transformation in **GCP**
* Strong understanding of **AWS** components such as **EC2** and **S3**
* Implemented a Continuous Delivery pipeline with **Docker** and **Git Hub**
* Worked with **g-cloud** function with **Python** to load **Data** in to Bigquery for on arrival **csv** files in **GCS** bucket
* Process and load bound and unbound Data from **Google pub/sub** topic to Bigquery using cloud Dataflow with **Python**.
* Devised simple and complex **SQL** scripts to check and validate Dataflow in various applications.
* Performed Data Analysis, Data Migration, Data Cleansing, Transformation, Integration, Data Import, and Data Export through **Python**.
* Developed and deployed data pipeline in cloud such as **AWS** and **GCP**
* Performed data engineering functions: data extract, transformation, loading, and integration in support of enterprise data infrastructures - data warehouse, operational data stores and master data management
* Responsible for data services and data movement infrastructures good experience with **ETL** concepts, building **ETL** solutions and **Data modeling**
* Architected several **DAGs** (Directed Acyclic Graph) for automating **ETL** pipelines
* Hands on experience on architecting the **ETL** transformation layers and writing **spark** jobs to do the processing.
* Gather and process raw data at scale (including writing scripts, web scraping, calling APIs, write SQL queries, writing applications)
* Experience in fact dimensional modeling (Star schema, Snowflake schema), transactional modeling and SCD (Slowly changing dimension)
* Devised **PL**/**SQL** Stored Procedures, Functions, Triggers, Views and packages. Made use of Indexing, Aggregation and Materialized views to optimize query performance.
* Developed logistic regression models (**Python**) to predict subscription response rate based on customers variables like past transactions, response to prior mailings, promotions, demographics, interests, and hobbies, etc.
* Develop near real time data pipeline using **spark**
* Process and load bound and unbound Data from **Google pub/sub topic** to Bigquery using cloud Dataflow with **Python**
* Hands of experience in**GCP, Big Query, GCS bucket, G - cloud function, cloud dataflow, Pub/sub cloud shell, GSUTIL, BQ command line utilities, Data Proc, Stack driver**
* Implemented Apache Airflow for authoring, scheduling and monitoring Data Pipelines
* Proficient in **Machine Learning techniques** (Decision Trees, Linear/Logistic Repressors) and Statistical Modeling
* Worked on confluence and **Jira** skilled in data visualization like Matplotlib and Seaborn library

**Environment**: SQL Server 2008, Python 2.x (NumPy/Pandas/Scikit-Learn), GitHub, Scrum, JIRA.

**Client: KAS Services Hyderabad Sep 2014 - July 2015**

**Role: Data Analyst**

**Description:** Anko is the innovation and technology hub for Kmart Group Australia. We fuel customer experience for iconic Australian retail brands Kmart and Target, some of Australia’s largest and fastest growing retailers who provide great value, on trend products across apparel and general merchandise.

**Responsibilities:**

* Led the development of prescriptive data model to optimize operations and increase efficiency
* Analyzed requirements, developed and debugged applications using BI tools
* Worked with end users/customers to understand business requirements, recommended technical solutions and documented functional requirements.
* Designed, developed, tested, and maintained Tableau reports based on user requirements.
* Designing, developing and providing technical solutions through Tableau reports based on user requirements.
* Extensively worked on designing and creating tableau dashboards. Also, automated dashboards to help stakeholders in implementation of strategic planning in the organization.
* Extensively worked on advance visualizations techniques in Tableau to assist business users in represent data visually to solve complex problems.
* Worked with leadership management team for visualizing the key customer metrics, driving better solution in consumer POD and supply chain.
* Utilized various Tableau functions like **LOD, WINDOWS SUM, WINDOWS AVG**, conditional functions (**IF, Nested IF, CASE, etc.)**, **RANK**, Table calculations like Percentage difference, Percent of Total, Moving Average, **YTD, YOY, etc**. for various measures/calculated fields.
* Worked on view orientation, **Sizing, layout, Color, Fonts, Tooltips,** for building Interactive dashboards/reports.
* Extensively worked on Tableau dashboards optimization for better performance and monitoring. Specifically dealing with huge volume of data sets in tableau.
* Extract, transform, and load data from a variety of data sources into staging datasets in redshift environment and ultimately into tableau for reporting.
* Extensively worked in preparing strategical planning and implementation of successfully delivery projects within the **POD** and supply chain teams.
* Scheduled data refresh on Tableau Server for weekly and monthly increments upon business requirement and ensured views and dashboards display changes in the data accurately.
* Published dashboards to the tableau server and communicated with the end stakeholder for **PMO** Usage
* Integrated projects in the server and organized workshops on the projects for better usage of the capability built in the dashboards.
* Publishes reports/dashboards on Tableau server & embedded views in other web applications.
* Worked on administration tasks such as Setting permissions, managing ownerships and providing access to the users and adding them to the specific group.
* Worked on year end activity for Database audit and performed year end activities for Business Panorama.
* Performed database audits and created reports in tableau based on the audit logs.

**Environment:** PowerDesigner 15.0, Oracle 10g, Oracle SQL Developer, Rational ClearQuest, MS Excel and Access.