**Python Data Engineer**

***Ahteshamuddin Mohammad***

**Contact: 302-286-9010**

**PROFESSIONAL SUMMARY**

* Extensive experience in Analyzing, Developing, Managing and implementing various stand-alone, client-server enterprise applications using Python, Djangoand mapping the requirements to the systems.
* Well versed with Agile with SCRUM, Waterfall Model and Test-driven Development (TDD) methodologies.
* Experience in developing web applications by using Python, Django, C++, XML, CSS, HTML, JavaScript and jQuery.
* Experience in analyzing data using Python, R, SQL, Microsoft Excel,Hive, PySpark, Spark SQL for Data Mining, Data Cleansing, Data Munging and Machine Learning.
* Experience in developing machine learning models like Classification, Regression, Clustering, Decision Tree.
* Good experience in developing web applications implementing Model View Control (MVC) architecture using Django, Flask, Pyramid and Python web application frameworks.
* Experience in implementing with server-side technologies with restful API and MVC design patterns with node JS and Django framework.
* Experience in working with number of public and private cloud platforms like Amazon Web Services (AWS), Microsoft Azure, Rackspace Cloud and Openstack.
* Extensive experience in Amazon Web Services (Amazon EC2, Amazon S3, Amazon Simple DB, Amazon RDS, Amazon Elastic Load Balancing, Elastic Search, Amazon MQ, Amazon Lambdas, Amazon SQS, AWS Identity and access management, AWS Cloud Watch, Amazon EBS and Amazon CloudFormation).
* Proficient in SQLite, MySQL and SQL databases with Python.
* Experienced in working with various Python IDE’s using PyCharm, PyScripter, Spyder, PyStudio, PyDev, IDLE, NetBeans and Sublime Text
* Good experience in using WAMP (Windows, Apache, MYSQL, and Python/PHP) and LAMP (Linux, Apache, MySQL, and Python/PHP) Architectures.
* Experience with Requests, Report Lab, NumPy, SciPy, Pytables, Python-Twitter, Matplotlib, HTTPLib2, Urllib2, Beautiful Soup, Data Frame and Pandas python libraries during development lifecycle.
* Hands-on experience in handling database issues and connections with SQL and NoSQL databases like MongoDB, Cassandra, Redis, CouchDB, DynamoDB by installing and configuring various packages in python.
* Strong ability to conduct qualitative and quantitative analysis for effective data-driven decision making.
* Conducted ad-hoc data analysis on large datasets from multiple data sources to provide data insights and actionable advice to support business leaders according to self-service BI goals.
* Experience in data preprocessing, data analysis, machine learning to get insights into structured and unstructured data.
* Experienced in working on Application Servers like WebSphere, WebLogic, Tomcat and Web Servers like Apache server, NGINX.
* Good Knowledge in writing different kinds of tests like Unit test/Pytest and build them.
* Experienced with version control systems like Git, GitHub, CVS, and SVN to keep the versions and configurations of the code organized.
* Experienced with containerization and orchestration services like Docker, Kubernetes.
* Expertise in Build Automation and Continuous Integration tools such as Apache ANT, Maven, Jenkins.
* Strong experience in developing Web Services like SOAP, REST, Restful with Python programming language.
* Experienced in writing SQL Queries, Stored procedures, functions, packages, tables, views, triggers using relational database like Oracle, DB2, MySQL, Sybase, PostgreSQL and MS SQL server.
* Experience in using Docker and Ansible to fully automate the deployment and execution of the benchmark suite on a cluster of machines.
* Good Experience in Linux Bash scripting and following PEP-8 Guidelines in Python.
* Extensive Knowledge on developing Spark SQL jobs by developing Data Frames.
* Executed complex HiveQL queries for required data extraction from Hive tables and written Hive UDF’s
* Experience in building applications in different operating systems like Linux (Ubuntu, CentOS, Debian), Mac OS.
* Excellent Interpersonal and communication skills, efficient time management and organization skills, ability to handle multiple tasks and work well in a team environment.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| Operating Systems | Windows 98/2000/XP/7,8, Mac OS and Linux CentOS, Debian, Ubuntu |
| Programming Languages | Python, R, C, C++ |
| Web Technologies | HTML/HTML5, CSS/CSS3, XML, jQuery, JSON, Bootstrap, Angular JS |
| Python Libraries/Packages | NumPy, SciPy,Boto, Pickle, PySide, PyTables, Data Frames, Pandas, Matplotlib, SQLAlchemy, HTTPLib2, Urllib2, Beautiful Soup, Py Query |
| Statistical Analysis Skills | A/B Testing, Time Series Analysis, Marko |
| IDE | PyCharm, PyScripter, Spyder, PyStudio, PyDev, IDLE, NetBeans, Sublime Text, Visual Code |
| Machine Learning and Analytical Tools | Supervised Learning (Linear Regression, Logistic Regression, Decision Tree, Random Forest, SVM, Classification), Unsupervised Learning (Clustering, KNN, Factor Analysis, PCA), Natural Language Processing, Google Analytics Fiddler, Tableau. |
| Cloud Computing | AWS, Azure, Rackspace, OpenStack |
| AWS Services | Amazon EC2, Amazon S3, Amazon Simple DB, Amazon MQ, Amazon ECS, Amazon Lambdas, Amazon Sagemaker, Amazon RDS, Amazon Elastic Load Balancing, Elastic Search, Amazon SQS, AWS Identity and access management, AWS Cloud Watch, Amazon EBS and Amazon CloudFormation |
| Databases/Servers | MySQL, SQLite3, Cassandra, Redis, PostgreSQL, CouchDB, MongoDB,TerraData, Apache Web Server 2.0, NginX, Tomcat, JBoss, WebLogic |
| ETL | Informatica 9.6, Data Stage, SSIS. |
| Web Services/ Protocols | TCP/IP, UDP, FTP, HTTP/HTTPS, SOAP, Rest, Restful |
| Miscellaneous | Git, GitHub, SVN, CVS |
| Build and CI tools | Docker, Kubernetes, Maven, Gradle, Jenkins, Hudson, Bamboo |
| SDLC/Testing Methodologies | Agile, Waterfall, Scrum, TDD |

**Education & Certification**

* Data Science: Successfully completed training and certification on Python for Data Science.
* Master of Science,Information Quality, University of Arkansas Little Rock 2017
* Bachelor of Engineering, Information Technology, Osmania University 2012

**PROFESSIONAL WORK EXPERIENCE**

General Electric Healthcare, San Ramon, California August2018 – Present

**Role: Python Data Engineer**

* Developed Data pipelines using python for medical image pre-processing, Training and Testing.
* Developed Artificial Intelligence Platform which helps Data Scientist’s to Train, Test and develop A.I. models on Amazon Sagemaker.
* Used Pandas,Opencv,Numpy, Seaborn, Tensorflow, Keras, Matplotlib, Sci-kit-learn, NLTK in Python for developing data pipelines and various machine learning algorithms.
* Design and engineer REST APIs and/or packages that abstract feature extraction and complex prediction/forecasting algorithms on time series data.
* Implement AWS Lambdas to drive real-time monitoring dashboards from system logs.
* Cleansing the data for normal distribution by applying various techniques like missing value treatment, outlier treatment, and hypothesis testing.
* Perform Data Cleaning, features scaling, features engineering using pandas and numpy packages in python.
* Create several types of data visualizations using Python and Tableau.
* Collected data needs and requirements by Interacting with the other departments.
* Worked on different data formats such as JSON, XML.
* Performed preliminary data analysis using descriptive statistics and handled anomalies such as removing duplicates and imputing missing values.
* Developed various graph methods to visualize and understand the data like Scatter plot, Pi-plot, bar charts, box-plot, and histograms.
* Involved in development of Web Services using REST API’s for sending and getting data from the external interface in the JSON format.
* *Configured EC2 instances and configured IAM users and roles and created S3 data pipe using Boto API to load data from internal data sources.*
* *Developed rest API's using python with flask framework and done the integration of various data sources including Java, JDBC, RDBMS, Shell Scripting, Spreadsheets, and Text files.*
* Implemented Agile Methodology for building an internal application.
* Developed A.I machine learning algorithms like Classification, Regression, Deep Learning using python.
* Conducted statistical analysis on Healthcare data using python and various tools.
* Experience in cloud versioning technologies like Github.
* Worked closely with Data Scientists to know data requirements for the experiments.
* Deep experience in using DevOps technologies like Junkins,Docker, Kubernetes etc.

**Client: University of San Francisco July 2017 – August 2018**

**Role: Data Engineer**

**Responsibilities:**

* Develop a data platform from scratch and took part in requirement gathering and analysis phase of the project in documenting the business requirements.
* Worked in designing tables in Hive, MYSQL using SQOOP and processing data like importing and exporting of databases to the HDFS, involved in processing large datasets of different forms including structured, semi-structured and unstructured data.
* Developed rest API's using python with flask and Django framework and done the integration of various data sources including Java, JDBC, RDBMS, Shell Scripting, Spreadsheets, and Text files.
* Worked with Hadoop architecture and the daemons of Hadoop including Name-Node, Data Node, Job Tracker, Task Tracker, and Resource Manager.
* Used AWS data pipeline for Data Extraction, Transformation and Loading from homogeneous or heterogeneous data sources and built various graphs for business decision-making using Python matplot library
* Developed scripts to load data to hive from HDFS and involved in ingesting data into Data Warehouse using various data loading techniques.
* Scheduled Jobs using crontab, run deck and control-m.
* Build Cassandra queries for performing various CRUD operations like create, update, read and delete, also used Bootstrap as a mechanism to manage and organize the html page layout
* Developed entire frontend and backend modules using Python on Django Web Framework and created User Interface (UI) using JavaScript, bootstrap, Cassandra with MySQL and HTML5/CSS
* Importing and exporting data jobs, to perform operations like copying data from HDFS and to HDFS using Sqoop and developed Spark code and Spark-SQL/Streaming for faster testing and processing of data.
* Analyzed SQL scripts and designed the solutions to implement using PySpark.
* Used JSON and XML SerDe's for serialization and de-serialization to load JSON and XML data into Hive tables.
* Used SparkSQL to load JSON data and create Schema RDD and loaded it into Hive Tables and handled structured data using SparkSQL.
* Developing data processing tasks using PySpark such as reading data from external sources, merge data, perform data enrichment and load in to target data destinations.
* Added support for Amazon AWS S3 and RDS to host static/media files and the database into Amazon Cloud.
* Worked in development of applications especially in LINUX environment and familiar with all its commands and worked on Jenkins continuous integration tool for deployment of project and deployed the project into Jenkins using GIT version control system
* Managed the imported data from different data sources, performed transformation using Hive, Pig and Map- Reduce and loaded data in HDFS.
* Executed Oozie workflow engine to run multiple Hive and Pig jobs, which run independently with time and data availability and developed Oozie workflow to run job onto data availability of transactions.
* To achieve Continuous Delivery goal on high scalable environment, used Docker coupled with load-balancing tool Nginx.
* Used MongoDB to stored data in JSON format and developed and tested many features of dashboard using Python, Bootstrap, CSS, and JavaScript.

**Environment**: Hadoop, Hive, Sqoop, Pig, java, Django, Flask, XML, MySQL, MS SQL Server, Linux, Shell Scripting, Mongo dB, SQL, Python 3.3, Django, HTML5/CSS, Cassandra, JavaScript, PyCharm, GIT, Linux, Shell Scripting, RESTful, Docker, Jenkins, JIRA, jQuery, MySQL, Bootstrap, HTML5, CSS, AWS, EC2, S3.

**Client:University of Arkansas, Little rock, Arkansas March 2016 – May 2017**

**Role: Data science Analyst(Graduate Assistant)**

**Responsibilities:**

* Participate in requirement gathering and analysis phase of the project in documenting the business requirements by conducting workshops/meetings with various business users.
* Developed Python/Django application for Google Analytics aggregation and reporting and used Django configuration to manage URLs and application parameters.
* Worked on Python Open stack API's and used Python scripts to update content in the database and manipulate files.
* Generated Python Django Forms to record data of online users and detailed Understanding on existing build system, Tools related for information of various products and releases and test results information.
* Involved in using AWS for the Tableau server scaling and secured Tableau server on AWS to protect the Tableau environment using Amazon VPC, security group, AWS IAM and AWS Direct Connect.
* Configured EC2 instances and configured IAM users and roles and created S3 data pipe using Boto API to load data from internal data sources.
* Built a mechanism for automatically moving the existing proprietary binary format data files to HDFS using a service called Ingestion service.
* Involved and worked on Python Open stack API's and used several python libraries such as wxPython, NumPy and matplotlib
* Performed Data transformations in HIVE and used partitions, buckets for performance improvements.
* Ingestion of data into Hadoop using Sqoop and apply data transformations and using Pig and HIVE.
* Used Python and Django creating graphics, XML processing, data exchange and business logic implementation
* Used Git, GitHub, and Amazon EC2 and deployment using Heroku and Used extracted data for analysis and carried out various mathematical operations for calculation purpose using python library - NumPy, SciPy.
* Developed server-based web traffic using RESTful API's statistical analysis tool using Flask, Pandas.
* Used Pandas API to put the data as time series and tabular format for east timestamp data manipulation and retrieval.
* Added support for Amazon AWS S3 and RDS to host static/media files and the database into Amazon Cloud.
* Systems automation utilizing Control-M for scheduling and PowerShell/C# for script development and used Pandas library for statistical Analysis.
* Participate in the design, build and deployment of NoSQL implementations like MongoDB.
* Wrote and executed various MYSQL database queries from Python using Python-MySQL connector and MySQL dB package.
* Developed various Python scripts to find vulnerabilities with SQL Queries by doing SQL injection, permission checks and performance analysis and developed scripts to migrate data from proprietary database to PostgreSQL.
* Involved in development of Web Services using SOAP for sending and getting data from the external interface in the XML format.
* Performed troubleshooting, fixed and deployed many Python bug fixes of the two main applications that were a main source of data for both customers and internal customer service team.

**Environment**: Python, Hive, Oozie, Amazon AWS S3, MySQL, HTML, Python 2.7, Django, HTML5, CSS, XML, MySQL, MS SQL Server, GIT, Jenkins, JIRA, MySQL, Cassandra, Pig, Hadoop, AWS Cloud Watch, AWS Redshift, SQL, SOAP, Rest APIs, AWS EC2, XML, JavaScript, AWS, Linux, Shell Scripting, AJAX, Mongo dB

**WeberSoft Systems, India                            Sept 2012–July 2015**

**Role: Data Engineer(Python Developer)**

**Responsibilities:**

* Developed entire frontend and backend modules using Python on Django Web Framework.
* Used Django framework for application development.
* Designed and developed the UI of the website using HTML, AJAX, CSS and JavaScript
* Worked on CSS Bootstrap to develop web applications.
* Used update strategy to effectively migrate data from source to target.
* Moved the mappings from development environment to test environment.
* Designed ETL Process using Informatica to load data from Flat Files, and Excel Files to target Oracle Data Warehouse database.
* Interacted with the business community and database administrators to identify the Business requirements and data realties.
* Created various transformations according to the business logic like Source Qualifier, Normalizer, Lookup, Stored Procedure, Sequence Generator, Router, Filter, Aggregator, Joiner, Expression and Update Strategy.
* Created Informatica mappings using various Transformations like Joiner, Aggregate, Expression, Filter and Update Strategy.
* Improving workflow performance by shifting filters as close as possible to the source and selecting tables with fewer rows as the master during joins.
* Used connected and unconnected lookups whenever appropriate, along with the use of appropriate caches.
* Created tasks and workflows in the Workflow Manager and monitored the sessions in the Workflow Monitor.
* Perform Maintenance, including managing Space, Remove Bad Files, Remove Cache Files and monitoring services.
* Set up Permissions for Groups and Users in all Development Environments.
* Migration of developed objects across different environments.
* Designed and developed Web services using XML and jQuery.
* Improved performance by using more modularized approach and using more in built methods.
* Experienced in Agile Methodologies and SCRUM Process.
* Maintained program libraries, user’s manuals and technical documentation.
* Wrote unit test cases for testing tools.
* Involved in entire lifecycle of the projects including Design, Development, and Deployment, Testing and Implementation and support.
* Built various graphs for business decision making using Python matplotlib library.
* Worked in development of applications especially in UNIX environment and familiar with all its commands.
* Used NumPy for Numerical analysis for Insurance premium.
* Handling the day to day issues and fine tuning the applications for enhanced performance.
* Implement code in Python to retrieve and manipulate data.

**Environment:**Python, Django, MySQL, Linux, Informatica Power Centre 9.6.1, PL/SQL, HTML, XHTML, CSS, AJAX, JavaScript, Apache Web Server, NO SQL, jQuery.