**Utham Kumar Chinthanippu**

**Senior Python Developer**

**LinkedIn:** [**www.linkedin.com/in/utham-kumar-a0003317a**](http://www.linkedin.com/in/utham-kumar-a0003317a)

**Email:** **uthamkc@gmail.com** **|| Phone: +1 (213)-528-8804**

**Professional Summary**

* Highly accomplishing **Senior** **Python Developer/Data Engineer** with **over** **10 years** of experience. Recognized for expertise in statistical data analysis and a strong command of core technologies including **Django, Python, Flask, PyCharm, Pyramid** and **Restful APIs**. Possesses a proven track record of delivering high-level solutions and leveraging advanced tools to drive impactful data-driven insights.
* Proficient in leveraging a range of Python Integrated Development Environments (**IDEs**) to drive efficient and effective development processes. Skilled in utilizing **PyCharm, PySpark, Spyder, PyStudio** and **PyDev** to streamline coding workflows and enhance productivity.
* Experienced in analysis of project requirements, adept in crafting innovative solutions and skilled in strategic selection of optimal technologies and frameworks for implementation.
* Highly proficient in Python scripting with an extensive skill set in utilizing powerful libraries such as **NumPy, Matplotlib** and **Pandas** for advanced statistical analysis, data visualization and efficient data organization.
* Demonstrated expertise in optimizing and profiling **Python code** leveraging a wide range of libraries and frameworks to achieve superior performance enhancements.
* High-level proficiency in optimizing and profiling Python code, utilizing modern libraries like Beautiful S**oup, SciPy, Matplotlib, Pickle, PySide, Pandas Data frame, Bottle framework, urllib2, Pychart Highcharts.**
* Possessing extensive expertise in advanced development of highly concurrent andweb services, utilizing cutting-edge frameworks like **Cherrypy** and **BottlePy.**
* Demonstrating an exceptional level of mastery in the latest **Hadoop** ecosystem technologies, encompassing **HDFS, MapReduce, Hive, Pig, Mahout, Oozie, Flume, Sqoop, Zookeeper, HBase, Cassandra, Spark, Spark SQL, Spark Streaming, Kinesis, Airflow, Yarn,** and **Scala.**
* Proficiently adept in working with both **WAMP** and **LAMP** architectures, leveraging the power of **Python ORM libraries** such as **Django ORM** and **SQL Alchemy** for seamless data management.
* Demonstrating mastery in implementing the Model-View-Controller (**MVC**) architecture using Django, seamlessly integrating state-of-the-art concepts such as Multi-Threading, Exception Handling, and Collections to ensure optimal performance and scalability.
* Experience in seamless integration of databases, encompassing both **SQL** (e.g., SQL Server) and **NoSQL** (e.g., MongoDB, Cassandra) technologies.
* Proficiently leveraging cutting-edge big data technologies like **Hadoop** (HDFS, MapReduce, Hive, Pig) and **Spark** (RDD, Spark SQL, Spark Streaming) to enable scalable data transformations and advanced analytics.
* Experience in utilizing **Python OpenStack APIs** and adept at working with **JSON**-based **REST** **web** **services**, while also demonstrating proficiency in developing web services using Python and working with protocols like **SOAP** and **REST**
* Highly skilled in developing Spark applications, leveraging key components such **as RDD, Data Frame, Spark SQL, Spark MLlib, Spark GraphX,** and **Spark SQL API** to facilitate high-scalable data transformations.
* Demonstrating hands-on experience with a wide range of **AWS** services (EMR, EC2, RDS, EBS, S3, Lambda, Glue, Elasticsearch, Kinesis, SQS, DynamoDB, Redshift, Athena, CloudTrail, SNS, CloudWatch, ECS, and Step Functions), extensive expertise in IT data analytics projects, including migrating on-premises ETL processes to Google Cloud Platform (**GCP**) utilizing cloud-native tools (BigQuery, Airflow, Cloud Run, Cloud Build, Cloud DataProc, and Google Cloud Storage) and proficiency in **Azure services** (Azure Data Lake, Azure Storage, Azure SQL, Event Hub, Data Factory, Scala, Synapse Analytics, Delta Lake, Cosmos DB, and Azure Databricks for data processing).
* Demonstrating expertise in designing and implementing sophisticated data models, employing state-of-the-art **ETL** techniques and proficiently utilizing tools such as **Informatica Power Center** to facilitate efficient data management and processing.
* Highly proficient in crafting comprehensive unit tests and seamlessly integrating them into the build process, leveraging cutting-edge frameworks such as **Unit** **Test** and **PyTest**.
* Adept in maintaining exceptional code quality through optimization, debugging, and utilizing advanced version control systems like **Git** and **SVN** to ensure seamless collaboration and version management.
* Demonstrating expertise in proficiently diagnosing and resolving intricate **Python code issues** encompassing both development and deployment environments.
* Extensive experience in leveraging cutting-edge **DevOps** tools such as **Jenkins**, **Docker,** and **CI/CD pipelines** along with adept utilization of advanced version control systems.
* Well-acquainted with a range of visualization and reporting tools including **QuickSight**, **Business** **Objects**, **Tableau** and **Power BI** enabling seamless data visualization and insightful reporting capabilities.

**Technical Skills:**

|  |  |
| --- | --- |
| Programming Languages | Python, R, SQL, Java, .Net, Pyspark, HTML, CSS, JavaScript, Scala, Spark. |
| Python Libraries | Requests, Report Lab, NumPy, SciPy, Pytables, cv2, imageio, Python-Twitter, Matplotlib, HTTPLib2, Urllib2, Beautiful Soup, PySpark, Pytest, Pymongo, cxOracle, PyExcel, Boto3. |
| Frameworks | Django, Flask, Pyramid, PyCharm, Sublime Text. |
| Architectures | MVW, MVC, WAMP, LAMP. |
| DBMS | Oracle, PostgreSQL, Teradata, IBM DB2, MySQL, PL/SQL, MongoDB, Cassandra, DynamoDB, HBase. |
| Web Services | REST, SOAP, Microservices. |
| Big Data Ecosystem Tools | Cloudera distribution, Hortonworks Ambari, HDFS, Map Reduce, YARN, Pig, Sqoop, HBase, Hive, Flume, Cassandra, Apache Spark, Oozie, Zookeeper, Hadoop, Scala, Impala, Kafka, Airflow, DBT, NiFi. |
| Reporting Tools | Power BI, SSIS, SSAS, SSRS, Tableau. |
| Containerization / orchestration Tools | Kubernetes, Docker, Docker Registry, Docker Hub, Docker Swarm. |
| Cloud Technologies | Amazon EC2, Amazon S3, Amazon RDS, VPC, IAM, Amazon Elastic Load Balancing, Auto Scaling, CloudWatch, SNS, SES, SQS, Lambda, Step Functions, Cloud transformations, EMR, Big Query, GCS Bucket, G-Cloud function, Cloud Dataflow, Pub/Sub, Cloud Shell, GSUTIL, BQ command line utilities, Data Proc, Azure web application, App services, Azure storage, Azure SQL Database, Virtual machines, Azure search, Notification Hub. |
| Data Modelling Techniques | Relational data modeling, ER/Studio, Erwin, Sybase Power Designer, Star Join Schema, Snowflake modeling, FACT and Dimensions tables. |
| Streaming Frameworks  | Kinesis, Kafka, Flume. |
| Version Control and CI/CD Tools | Concurrent Versions System (CVS), Subversion (SVN), GIT, GitHub, Mercurial, Bit Bucket, Docker, Kubernetes. |

**Certifications:**

**Microsoft Certified Azure Data Engineer Associate**

**Professional Experience:**

**Professional Experience**

**Client: Wells Fargo, Charlotte, NC Jan2021-Present**

**Role: Senior Python Developer**

**Responsibilities:**

* Implemented object-oriented programming principles and multithreading techniques in **Python** to develop efficient and scalable solutions for diverse projects.
* Utilized the **Django** web framework and MySQL database management in Python to design and develop robust web applications, ensuring seamless data processing and management.
* Employed robust exception handling mechanisms in Python to enhance the reliability and stability of applications, ensuring smooth execution in diverse scenarios.
* Leveraged efficient data structures from the Collections module in Python to optimize data processing and manipulation tasks in projects.
* Developed PySpark DataFrames within Azure Databricks using **Python** to ingest and process data files from Data Lake or Blob storage, optimizing data transformation.
* Employed a comprehensive suite of Python-specific tools including Django, Django Rest Framework, Flask, Pandas, NumPy, SciPy, Matplotlib, and more to develop robust solutions and conduct in-depth data analysis tasks.
* Designed and implemented comprehensive data integration solutions using Python, encompassing various Azure services like Azure Data Lake Analytics, Azure Data Lake Storage, Azure Data Factory, SQL databases, and Azure SQL Data Warehouse.
* Implemented robust data governance and security policies leveraging Azure **Data Lake Storage, Azure Active Directory** and **Azure Key Vault** to ensure the highest levels of data integrity and compliance within the organization.
* Oversaw and managed precise control of database access, while executing seamless migration of on-premises databases to **Azure Data Lake Store using Azure Data Factory** and **Cosmos DB**.
* Utilized the Copy Activity feature in **Azure Data Factory** to facilitate seamless and secure data transfers between on-premises and cloud-based data stores, ensuring efficient and reliable data movement.
* Architected a robust solution leveraging **Azure** **EventHub** and Stream Analytics to design an advanced analytics architecture that catered to the organization's sophisticated data analysis requirements.
* Strategically designed and meticulously maintained comprehensive test cases for web applications, employing industry-standard tools such as **Selenium WebDriver, Python** and **Pytest.**
* Employed cutting-edge technologies such as Kafka, Storm, Spark streaming, and Java to enable real-time data processing and storage in **HDFS** (Hadoop Distributed File System).
* Extracted data from **HDFS** utilizing Hive and Presto and conducted extensive data analysis using Spark with **Scala** and **PySpark**.
* Leveraged Flume to effectively collect, aggregate, and store web log data from diverse sources, including web servers, mobile devices, and network devices. data was efficiently pushed to **HDFS** (Hadoop Distributed File System), ensuring centralized storage, and enabling seamless data analysis and insights generation.
* Played a key role in the conversion of **Cassandra** **Hive** and **SQL** queries into Spark transformations using **SparkRDD** in both **Scala** and **Python**.
* Developed **SSIS** packages incorporating advanced functionalities such as **Pivot Transformation, Fuzzy Lookup, Derived Columns, Condition Split, Term Extraction, Aggregate, Execute SQL Task, Data Flow Task** and **Execute Package.**

**Environment:** Python, Pytest, Pandas, NumPy, PyCharm, Visual Studio Code, Atom, MySQL Workbench, phpMyAdmin, Navicat, Azure portal, Azure Data Lake Analytics SDK, Azure portal, Azure Storage Explorer, Azure portal, Azure Data Factory UI, Azure portal, Azure SQL Database Management Studio, Django ORM, Flask, SQL Alchemy, Beautiful Soup, SciPy, Matplotlib, python-twitter, urllib2, Selenium WebDriver, Oracle, SQL Server, DB2, flat files, HDFS.

**Client: Guidewire, San Mateo, CA Oct2018-Dec2020**

**Role: Senior Python Developer**

**Responsibilities:**

* Developed **Python**, **Spark**, and **PySpark** scripts to architect and implement robust ETL pipelines, automating data ingestion and seamlessly updating relevant databases and tables. Incorporated performance tuning and optimization techniques to enhance Python applications' efficiency.
* Leveraged the **Pandas** library to execute efficient and scalable operations for reading and writing large datasets from CSV and Excel files, ensuring powerful data manipulation capabilities in Python.
* Enhanced server-side rendered web applications using **Django**/**Python**, HTML/CSS, and integrated various **Python** libraries like Matplotlib, MySQL dB, python-twitter, PySide, Pickle, and Pandas DataFrame to add advanced functionality.
* Utilized Google Cloud Pub/Sub service for real-time notifications on file arrivals in **Google Cloud Storage.**
* Played a key role in migrating on-premises **Hadoop** systems to Google Cloud Platform (**GCP**), leveraging **GCP** services including Cloud Storage, Data Proc, Dataflow, and BigQuery. Led proof-of-concept (**POC**) projects to assess the capabilities of **GCP's** offerings and conducted a comparative analysis between self-hosted Hadoop infrastructure and **GCP's** Data Proc.
* Designed and implemented robust data pipelines using Airflow in Google Cloud Platform (**GCP**) to efficiently handle **ETL** (Extract, Transform, Load) tasks. Leveraged a range of Airflow operators to ensure seamless data flow and integration. Utilized the **Cloud Shell SDK** in GCP to configure and manage essential services such as Data Proc, Storage, and BigQuery, ensuring optimal performance and scalability.
* Orchestrated the successful migration of an entire **Oracle** database to **BigQuery** within Google Cloud Platform (**GCP**) using industry-leading techniques and tools, ensuring seamless and accurate data transfer. Leveraged **Power** **BI** for insightful and visually compelling reports and dashboards based on the migrated data in **BigQuery**.
* Collaborated extensively with a diverse range of services including **Data Lake, Data Lake Analytics, SQL Database, Synapse, Data Bricks, Data Factory, Logic Apps** and **SQL Data** Warehouse to successfully implement large-scale data warehousing programs and end-to-end data integration solutions.
* Architected and implemented a comprehensive data model in Neptune, a highly scalable graph database, to efficiently manage and query complex relationships within the data. Proficiently loaded and organized data into the **Neptune** **database**, leveraging optimal data structures.
* Developed and deployed **Grafana** dashboards to provide comprehensive monitoring and visualization of key metrics extracted from the Cassandra database.
* Utilized **Teradata** **SQL** Assistant to write efficient and optimized **SQL** queries for ad hoc data retrieval and analysis in the **Teradata** database. Developed custom User-Defined Functions (**UDFs**) for Pig and Hive, leveraging Python and Java programming languages to extend the capabilities of Pig Latin and HiveQL (**HQL**).

**Environment:** Python, Spark, PySpark, ETL, CSV, Django/Python, HTML/CSS, MySQL dB, python-twitter, PySide, Pickle, Pandas DataFrame, network, urllib2, GCP, POC, Oracle, SQL, Neptune, UDFs, HQL, Glassfish, Tomcat and CGI.

**Client: Toyota Motor North America, Inc, Plano, TX. (Off Shore ) July 2015 - Sept 2018**

**Role: Python Developer**

**Responsibilities:**

* Developed **Python** scripts for **AWS** Lambda functions and utilized IAM roles to schedule and automate tasks using CloudWatch Triggers, ensuring efficient and reliable task execution within the AWS environment.
* Conducted data cleansing, validation, and quality checks on data stored in S3 buckets using Python, leveraging AWS Athena for interactive analytics to derive valuable insights from the data.
* Developed data tables using **PyQt**, a **Python** binding for the Qt framework, to effectively manage and organize records and information, contributing to robust data management capabilities.
* Utilized **Python's** pandas, NumPy, scikit-learn, and matplotlib libraries to construct machine learning statistical models and perform advanced data manipulation for data analysis and modeling tasks.
* Developed web applications using the **Django** Framework, following the Model-View-Controller (**MVC**) architecture, ensuring scalable and robust application development.
* Implemented Python programs/scripts for data preparation, transformation, and harmonization, tailored to the specific requirements of modeling tasks, enhancing data processing efficiency.
* Streamlined job automation and data pipelines through the utilization of **AWS** Step Functions and **AWS Lambda**, while implementing comprehensive performance metrics monitoring using **AWS CloudWatch** to ensure efficient and reliable orchestration within a professional **AWS** environment.
* Leveraged **AWS** **EMR** (Elastic MapReduce) to seamlessly perform large-scale data transformation and movement between various **AWS** data stores and databases, including Amazon Simple Storage Service (**Amazon S3)** and Amazon **DynamoDB** enabling efficient data processing and management within a professional **AWS** environment.
* Engaged in **AWS** **Glue** jobs to efficiently extract data from **S3** and load it into the designated **Redshift bucket** aligning with specific requirements and ensuring seamless data integration within a professional **AWS** environment.
* Leveraged Kinesis Data Firehose to effectively filter and pre-process data, enabling streamlined data streaming into **S3** data lakes while maintaining data integrity and adhering to specific data requirements within a professional context.
* Employed the Python Boto3 library to seamlessly configure and manage **AWS** **services**, including **AWS Glue, EC2** and **S3** enabling efficient and optimized utilization of cloud resources within a professional AWS environment.
* Developed and maintained data pipelines using industry-standard tools such as **Apache NiFi, Apache Kafka,** and **Apache Storm** to ensure smooth data flow and integration across systems.
* Configured **ZooKeeper** for efficient coordination and support of distributed applications, leveraging its high throughput, availability, and low latency features.
* Leveraged Apache Spark in conjunction with **Python** to develop and execute Big Data Analytics and Machine Learning applications, harnessing the power of distributed computing.

**Environment:** Python, AWS, IAM, S3 buckets, DynamoDB, Redshift bucket, DAG, EC2, AWS Glue, EC2, S3, NumPy, scikit-learn, matplotlib, MySQL, MVC, TDD, Netezza, Oracle, Spark SQL, Mload, Tpump, Fastload, Fast Export, Oozie.

**Client: Cox Communications, Inc. Atlanta, GA (Off Shore ) May 2013- Jun 2015**

**Role: Python Developer**

**Responsibilities:**

* Developed web-based applications using Python, Django, PHP, C++, XML, CSS, HTML, DHTML, JavaScript and jQuery.
* Played a key role in enhancing web application development by leveraging the Django web framework and Python programming language, along with HTML and CSS to create robust and scalable server-side rendered applications.
* Created user-friendly website interfaces by developing views and templates using Python and Django's view controller and templating language.
* Utilized a range of Python Integrated Development Environments (IDEs) including PyCharm, PyScripter, Spyder, and PyDev for effective development and debugging.
* Demonstrated proficiency in utilizing a wide range of SAS procedures, including Import, Export, Sort, Freq, Means, Format, Append, Univariate, Datasets and Report.
* Used SQL queries for legacy data retrieval jobs and successfully migrated the Django database from MySQL to PostgreSQL.
* Worked with both WAMP (Windows, Apache, MySQL, and Python/PHP) and LAMP (Linux, Apache, MySQL, and Python/PHP) architectures.
* Conducted comprehensive Exploratory Data Analysis (EDA) to unveil patterns, trends, and clusters within the dataset, employing statistical techniques and visualization tools.
* Developed highly scalable and deployable machine learning models, utilizing state-of-the-art techniques, and leveraging the power of Python libraries such as scikit-learn, MLlib and MLXtend.
* Utilized Python’s renowned data science packages, including Pandas, NumPy, Matplotlib, Seaborn, SciPy, Scikit-learn and NLTK.
* Employed a diverse range of modeling techniques, showcasing expertise in building sophisticated models for various tasks, including regression analysis, tree-based ensemble methods, time series forecasting, K-nearest neighbors (KNN), clustering, and isolation forest methods.
* Expertly implemented advanced machine learning models, including logistic regression and XGBoost, utilizing Python's renowned Scikit-learn library.
* Employed advanced algorithm optimization techniques, including stochastic gradient descent and automated tuning with Bayesian optimization, to enhance the performance and efficiency of machine learning models.
* Implemented Python alongside various libraries such as Matplotlib for charts and graphs, MySQLdb for database connectivity, python-twitter, PySide, Pickle, Pandas Data Frame, network and urllib2.
* Demonstrated proficiency in efficiently performing large-scale data read and write operations involving CSV and Excel files using the powerful data manipulation library, pandas.

**Environment:** Python, Django, PHP, C++, XML, CSS, HTML, DHTML, JavaScript, jQuery, IDEs, PyCharm, PyScripter, Spyder, PyDev, SQL, WAMP, LAMP, EDA, scikit-learn, MLlib, MLXtend, Pandas, NumPy, Matplotlib, Seaborn, SciPy, Scikit-learn, NLTK, KNN, MySQLdb, CSV.

**Education:**

**Bachelors in Electronics and Communication, JNTUH, India- 2012**