

Python Backend/ETL Developer Python Backend/ETL Developer Python Backend/Bigdata Developer Dallas, TX ? 5 years of IT experience in all phases of SDLC, along with experience in Application Design and software development. ? Capable of processing large sets of structured, semi-structured and unstructured data and supporting systems application architecture. ? Experience in Python OpenStack API'S ? Worked on Datasets related to retail, telecommunication and financial industries. ? Familiar with the Object-Oriented Programming concepts. ? Able to assess business rules, collaborate with participants and perform source-to-target data mapping, design and review. ? Experience in writing Sub Queries, Stored Procedures, Triggers, Cursors, and Functions on SQL-Server Cassandra, HBase (Phoenix SQL), Hive and PostgreSQL databases. ? Familiar with the AWS cloud services like EC2, Elastic Container Service (ECS), Simple Storage Service (S3) and Elastic MapReduce (EMR). ? Experience on analyzing the large datasets with In-memory data-structures using Pandas and spark. ? Written scripts for Read/Write for Hive and HBase through Thrift service. ? Worked as developer in agile environment with Git as Version Control. ? Familiar with the development Test Driven Development and Unit & Integration Testing. ? Hands on experience in parallel, concurrent and reusable programming techniques. ? Familiar with data ingestion pipeline design, Hadoop architectures and data modeling. ? Developed Web services using spark and Flask and Django frameworks. ? Developed and optimized ETL workflows in both legacy and distributed environments. ? Capable of writing analytical queries efficiently that helps analysts to spot the trends. ? Experience in working with the IDE's like Zeppelin, Notebook, PyCharm etc. ? Experience in using files JSON, XML, Pickle, ORC, AVRO, PARQUET file formats ? Configured Flume to extract the data from the web server and then loaded into HDFS. ? Developed UDFs (python) for Pig and Hive to preprocess and filter data sets for analysis in distributed environments. ? Imported and exported structured, semi-structured, unstructured data from HDFS and SQL databases by batch and streaming applications. ? Developed data Streaming applications in Hadoop or Bigdata environments using Kafka. ? Written Spark applications using Pyspark for real-time data analysis by connecting to the multiple data warehouse like Hive and HBase. ? Worked with docker services and creating application specific docker images. ?

Experience in creating the user interfaces using HTML, CSS and JavaScript ? Expertise in getting the web data through API's and web scrapping techniques. ? Capable of writing the configuration and Deployment Scripts using Fabric and Jenkins. ? Developed dashboards using Tableau Desktop and Bokeh and D3Js. Sponsorship required to work in the US Work Experience Python Backend/ETL Developer Capital One Inc October 2017 to Present Project description: This is a real time application which represents the data of customer and their performance from time to time. Also, the feedback is captured from some of the surveys conducted by the Enterprise. Where this helps the managers to make the decisions most effectively and understand their customer End-to-End. Responsibilities: ? Involved in architecture, flow and the database model of the application. ? Developed the ETL jobs as per the requirements to update the data into the staging database (Postgres) from various data sources and REST API's. ? Developed analytical queries in Teradata, SQL-Server, and Oracle. ? Developed a Web service on the Postgres database using python Flask framework which was served as a backend for the real-time dashboard. ? Partially involved in the developing the front-end components in the Angular and also editing the HTML, CSS and JavaScript. ? Wrote Unit and Integration Tests for all the ETL services. ? Containerized and Deployed the ETL and REST services on AWS ECS through the CI/CD Jenkins pipe. ? Worked on optimizing and memory management of the ETL services ? Developed Splunk Queries and the dashboards for the debugging the logs generated by the ETL and the REST services. Environment: Python, Postgres, Dockers, Teradata, Flask, Gunicorn, AWS, ECS, Jenkins, SQL Server, S3, Kafka, Angular4, D3Js, CSS, HTML5, JavaScript. Python/ETL Tester & Developer Capital One Inc May 2017 to September 2017 Project description: The objective of project is developing and writing the tests for the ETL applications, where the applications are deployed through Jenkins pipe (CI/CD) to the cloud. These ETL applications stream and batch load/write the data to various data sources and also generate the reports for the business users as per the requirements. Responsibilities: ? Created Integrated test Environments for the ETL applications developed in GO-Lang using the Dockers and the python API's. ? Appended the Integrated testing environments into Jenkins pipe to make the testing automated before the continuous deployment process. ? Installed data sources

like SQL-Server, Cassandra and remote servers using the Docker containers as to provide the integrated testing environment for the ETL applications. ? Also wrote Unit tests for the developed scripts for the getting through the quality checks before pushing to the deployments. ? Worked on optimizing and memory management of the ETL applications developed in Go-Lang and python and also reusing the existing code blocks for better performance. Environment: GO, python, Cassandra, Dockers, SQL-Server, GO, AWS, EC2, Mesos, Jenkins, S3, Kafka Splunk.

Python/Hadoop Developer DataMorhphix - Big Data Plattform for Analytics May 2016 to March 2017
Duration: May, 2016 to Mar, 2017 Real Time Integration with Health Care and Social Media data

Python/Hadoop Developer Project description: The aim of this project is to retrieve Health Care (HL7) and Social Media data and performing the sentiment analysis as well as interests of the user's feedback. Also providing the Realtime dashboards as per the interests of the users and represent the analysis with some BI tools and try to improve the quality of the service accordingly.

Responsibilities: ? Evaluated business requirements and prepared detailed specifications that follow project guidelines required to develop written programs. ? Wrote scripts in Python for Extracting Data from JSON and XML files. ? Developed the back-end web services for the worker using Python Flask REST APIs. ? Designed and Developed the CRUD scripts to load the transactional data into Hive and HBase using the thrift and python scripting. ? Performed Map/Reduce operations on the raw files located in HDFS for staging and transforming the data using Pig and spark. ? Collecting the social media data from various REST services and also scrapping the raw web pages using the web scrapping API's like Scrapy. ? Wrote the python scripts that get Sentiments and the Insights of the text data collected using the Watson Analytics API. ? Develop the spark jobs that aggregate the large datasets from HBase and store the aggregated the report into the temporary tables for reporting. ? Implemented Oozie workflow engine on Hortonworks Hadoop cluster to run multiple ETL jobs developed in python, Pig and spark in orderly manner. ? Worked on front end frameworks like JavaScript and Bokeh API for responsive web pages. Environment: Python, AWS, Hortonworks, HDFS, Hive, Kafka, HBase, Dockers, spark, Tableau, Bokeh, Phoenix SQL, Scrapy, XML, HTML, pandas, Watson-Alchemy. Python/Hadoop

Developer DataMorphix - Big Data Platform for Analytics September 2015 to April 2016 Duration: Sept, 2015 to Apr 2016 Integration with Spend analysis Omni Channel Solution Provider Python/Hadoop Developer Project Description: The main aim of the project is to identify the Goals, Usages and Transactions of the customer by looking at their spending and investments through the bank and try to improve the service quality accordingly. Performing the advanced analytics to generate predictive reports that helps in making future decisions to every individual customer. Responsibilities: ? Evaluated business requirements and prepared detailed specifications that follow project guidelines required to develop written programs. ? Developed an ETL service that looks for the files in the server and update the file into the Kafka queue. ? Developed a data consumer which takes the data from the Kafka queue and load it to the Hive tables. ? Worked closely with the data scientists for migrating the prediction algorithms/models to Python sciKit-learn API from R-studio and also Involved in the feature selection for creating the prediction models. ? Involved in designing the in the Hive using the optimizing techniques like bucketing/partitioning to stop the data across the cluster. ? Created the views in Hive to provide the datasets that are required for building the prediction models. Environment: Python, Hadoop, sciKit-learn, HDFS, Hive, Hortonworks, Oozie, MapReduce, Spark, Kafka, Tableau. Python Developer IntelliAsia Software Services Ltd May 2013 to April 2014 Project Description: This tool is used for advertising the scheduled timings. As part of the platform the advertising details and their scheduled times are stored in the Oracle database. Responsibilities: ? Writing Python scripts to parse XML documents as well as JSON based REST Web services and load the data in database. ? Writing ORM's for generating the complex SQL queries and building reusable code and libraries in Python for future use. ? Working closely with software developers and debug software and system problems ? Profiling Python code for optimization and memory management and implementing multithreading functionality. ? Involved in creating stored procedures that gets the data and help analysts to spot the trends. Environment: Python, Oracle, JSON, XML Education Technology Jawaharlal Nehru Technological University - Anantapur, Andhra Pradesh Master's in computer engineering in computer engineering Fairleigh Dickinson University Skills HTML (1 year), PYTHON (5 years),

KAFKA (3 years), ETL (4 years), SQL (3 years), Backend (3 years), Back End, Javascript (1 year)

Additional Information Technical Skills: Language Python, SQL, C++, GO-Lang, HTML, CSS, JavaScript, Jinja2 Technologies JDBC, NOSQL, Docker, AWS, Git Frameworks Tkinter, Flask, Django IDE PyCharm, IDLE, Notebook, Zeppelin Build Tools PyBuilder, Pip, Npm, VirtualEnv, Coverage, Jenkins, Docker Tools Tableau, Cron, Matplotlib, Pandas, Flume, Splunk, Bubbles (ETL), PySpark, Bokeh, Kafka, Boto3(AWS) Operating Systems Windows, Linux, OSX Big Data Technologies Hortonworks Hadoop, HDFS, Spark, Oozie, Sqoop, HBase, Hive, Impala, Pig, Flume and Hue, Cassandra, MongoDB

Name: Kelly Buchanan

Email: iwilliams@example.com

Phone: 001-820-286-4436x97455