

Sr. Python Developer Sr. Python Developer Sr. Python Developer - Fitch Ratings ? 5 years of professional IT experience which includes Big data ecosystem related technologies like Hadoop HDFS, Map Reduce, Apache Pig, Hive, Sqoop, Hbase, Flume, Oozie, Spark. ? Proficient in Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Data Node, Name Node and Map-Reduce concepts. ? Worked with Various Distributions like Cloudera, Horton works, and Amazon AWS. Worked with large sets of structured, semi-structured and unstructured data. ? Experience in working with tools like Attunity and having knowledge on ETL tools like Talend. ? Extensively worked on Spark SQL, Dataframes, RDD's to improve the performance of the application. ? Experience in using Apache Flume for collecting, aggregating and moving large amounts of data from application servers. ? Created Sqoop Jobs with incremental Loads to populate Hive External Table. ? Designed and implemented Hive and Pig UDF's using Python for evaluation, filtering, loading and storing of data. ? Developed web applications and Restful web services and APIs using Python Flask, Pyramid and Django. ? Good experience in Object oriented programming concepts in Python, Django and Linux. ? Experience in JSON based REST web services and SOAP for sending and getting data for the JSON format. ? Experience in using HUE for scheduling and monitoring oozie workflow and coordinator. ? Extensively worked on analyzing data using HiveQL, Pig Latin, and custom Map Reduce programs. ? Worked on using different file formats like JSON, Sequence files, AVRO file, Parquet file formats. ? Used Spark SQL and Hive SQL to process structured and un structured data. ? Extensive knowledge in writing and analyzing complex SQL queries, stored procedures, database tuning, query optimization and resolving key performance issues. ? Hands on Experience in Writing Python Scripts for Data Extract and Data Transfer from various data sources. ? Utilized standard Python modules such as CSV, Iter functions and pickle for development. ? Worked with Panda data frames and MySQL, queried MYSQL database queries from python using Python-MYSQL connector and MYSQL db package to retrieve information. ? Worked with several Python libraries like NumPy, Pandas and Matplotlib. ? Worked with Tableau to connect with Hive Data Warehouse to represent the data in dashboards. ? Knowledge of working with Python Django Forms to record data of online users. ? Experience in

Scrum, Agile models. ? Highly motivated and versatile team player with the ability to work independently & adapt quickly to new emerging technologies. Authorized to work in the US for any employer

Work Experience

Sr. Python Developer Fitch Ratings - New York, NY January 2016 to Present

Responsibilities: ? Analysis of requirements and implement different functions, models according to design. ? Extracted the data that are required for the segmentation using Pyspark environment. ? Worked on reading multiple data formats on HDFS using PySpark. Worked on creating custom ETL scripts using Python for business related data. ? Created custom new columns depending up on the use case while ingesting the data into Hadoop Lake using Pyspark. ? Performed Data Cleansing using python and loaded into the target tables. Involved in transformations using various Spark Actions and Transformations by Creating RDD's from the required files in HDFS. ? Worked on with spark dataframe operations that are required to develop a data format file. ? Analysed the sql scripts and designed it by using PySpark SQL for faster performance. ? Used python sub-process module to call UNIX shell commands to check directories or files exists. ? Worked on storing the dataframe into hive as table using PySpark. Developed data format file that is required by the Model to perform analytics using Spark SQL and Hive query language. ? Used spark parameters like number of executors, executor memory to execute the pipeline to increase the performance. ? Developed other scripts that will be moving the data from HDFS to Amazon S3. The Data format file is provided as input to Model which performs the analytics and the output will be placed in HDFS location. ? Created External Table in the HDFS location where the analytics data will be updating regularly. ? Monitor the Spark jobs in Spark URL and Involved in Unit test and debugging after development. ? Used Pandas API to put the data as time series and tabular format for easy timestamp data manipulation and retrieval. ? Used Tableau to develop the data visualization and developed dashboards according to the requirement. Involved in creating metrics, attributes, filters, reports, and dashboards with Tableau. **Environment:**

Hortonworks2.4, Spark 1.6, Hive, HiveQL, Zookeeper, Attunity, Python, Tableau, Pandas

Sr. Python Developer Community Healthcare Network - New York, NY August 2014 to December 2015

Responsibilities: ? Evaluated business requirements and prepared detailed specifications that

follow project guidelines required to develop written programs. ? Implemented python code for retrieving the Social Media data. Analyzed large amounts of data sets to determine optimal way to aggregate and report on it. ? Designed the data model for storing the data. ? Creating Hive tables, loading with data and writing Hive queries which will run internally in Map Reduce way. ? Designed and developed the scripts to load the data into Hive. Worked on Hive joins to aggregate social media data according to the requirement. ? Created tables in Hive which are used as staging table for the data loaded from the inbound feeds. Worked on different file formats like Text files and ORC files. ? Created Hive queries that helped market analysts spot emerging trends by comparing fresh data with EDW reference tables and historical metrics. Tested raw data and executed performance scripts. ? Developed Spark Code using python for faster processing of data. Segmentation and analysis of data from hive tables are done using Spark SQL. ? Implemented Oozie workflow engine to run multiple Hive and Python jobs. Developed Kafka where it will be receiving social media data and storing it in HDFS. ? Used Tableau to connect to required data warehouse and perform visualization. Environment: Apache Hadoop, Map Reduce, Hive, Pig, Zookeeper, Python, Oozie, Tableau, Spark, Kafka, Hue, HDFS, Tez. Sr. Python Developer Community Healthcare Network - New York, NY January 2013 to July 2014 Responsibilities: ? Requirement gathering and analysis phase of the project in documenting the business requirements by conducting workshops/meetings with various business users. ? Used Django/ Python web-framework to develop application. Developed the Python scripts in order to make the interaction between Server and MySQL Database. ? Used MVC framework to build modular & maintainable applications. Designed and developed transactions and persistence layers to save/retrieve/modify data for application functionalities using Django and PostgreSQL. ? Automated data movements using python scripts. Involved in splitting, validating and processing of files. Developed object-oriented programming to enhance product management. ? Developed templates using HTML, CSS, BOOTSTRAP, and JAVA Script based on client's request. ? Created and/or modified SQL Queries whenever required for change requests/enhancements. ? Responsible for writing scripts for merging large datasets using Panda data frames and MySQL. ? Developed the application using Python and MySQL. ?

Involved in writing the Python script files for processing data and loading to HDFS, and writing CLI commands using HDFS shell commands ? Involved in building database Model and Views utilizing Python, in order to build an interactive web based solution. ? Responsible for debugging and troubleshooting the web application. Designed and developed data management system using MySQL and built application logic in Python. Environment: Python 2.6, Django Framework 1.3, CSS, SQL, MySQL, JavaScript, JQuery. Skills APACHE HADOOP HDFS (4 years), HADOOP DISTRIBUTED FILE SYSTEM (4 years), HDFS (4 years), PYTHON (4 years), SQL (4 years) Additional Information SKILLSET: Big Data: HDFS, Map Reduce, Hive, Pig, Spark, Kafka, Sqoop, Flume & Oozie Cloud: AWS, Cloudera CDH, Horton works Sandbox Programming Languages: Python, Java (core), Unix Shell Scripting IDE'S: Eclipse, MATLAB, Scala IDE, SQL Developer, MS Office Database: Oracle 11g, SQL, MySQL Operating Systems: Linux, UNIX, Windows Tools: Tableau, Attunity Frameworks: Django Web Technologies: HTML, CSS, XML, JavaScript

Name: Richard Freeman

Email: xnguyen@example.org

Phone: (210)634-9965