

Python/Spark Developer Python/Spark Developer Data Engineer/ Python developer - Verisk Analytics 10+ Years IT experience - 4+ Years of experience in Python Knowledge on AWS EC2, S3 and EMR. Worked with Apache Spark and python for transforming the data and loading to HDFS. Experience in working with Perl CGI and UNIX Shell Scripting 8 years of experience in Perl Scripting Knowledge on Hadoop ecosystem. Used HDFS for storage and querying with Hive 1 Year of experience in Core Java Experience working on various databases: MS SQL Server, Teradata, Oracle and MySQL Experience working in NoSQL(MongoDB) and Python Experience in reading and writing xml reports with Perl XML modules Experience working with repositories: SVN and CVS Experience in working with Templating (HTML:Template, Template Toolkit) Experience working with large databases with high volumes of data Experience working in all phases of SDLC (System Development Life Cycle), planning, analysis, design, coding, implementation, testing, maintenance and documentation Experience implementing SCD's (Slowly Changing Dimensions) types I, II, III Experience Enforcing Coding Standards, Best Practices and User Security Possess Physical, Logical and Dimensional Data Modeling skills Familiar with Data Modeling concepts, Star Schema, Snowflake Schema, Facts and Dimensions Knowledge of OLTP, OLAP and other data warehousing concepts Knowledge and experience in SQL programming Experience creating, Tables, Views, Indexes, Partitions, Procedures, Functions, Triggers etc Experience devising SQL queries for Unit Testing, Data Validation, Integration Testing Experience in Job Monitoring & Scheduling in TWS (Tivoli Workload Scheduler), Autosys Experience creating HLD's, LLD's, (High Level & Low Level Designs) and Detailed Design Documents Experience with Onsite-Offshore Model Experience performing Maintenance, Support and Administrator related activities Strong technical, analytical, logical and communication skills. Work Experience Python/Spark Developer Verisk Analytics - Jersey City, NJ July 2017 to Present Forms Analytics - First part: To slice insurance documents(Coverage forms and Endorsements) and get the similarity and differences in the documents. Second part: Perform ETL operations on vendor files which are published to S3 and load aggregated data back to S3. First Part: ? Used python for end to end processing of documents with Spark Streaming,Kafka and AWS ? Documents are stored in MongoDB gridfs and

vendor data stored in AWS S3 buckets in the source system ? By using python scripts read mongo gridfs data and queue the document data to kafka ? Kafka consumer reads the data from queue and fetches the corresponding document in MongoDB Gridfs ? Slice the received document with BeautifulSoup and load the data to MongoDB, MS SQL and S3 ? Used Spark streaming using Python and Spark for calculating similarity and diff of documents

Second Part: ? Second part of project is processing of vendor documents from S3. ? Created AWS Lambda when vendor document is published to S3 bucket lambda sends it to Kafka topic ? Created AWS EMR cluster with Spark and using this cluster process the data coming from Kafka ? Have experience in creating RDDs, DataFrames and manipulating them ? Have working knowledge on Spark SQL ? Used Apache Spark and python to perform aggregations and transformations on the loaded data and load back to S3 ? Used Pymongo to perform CRUD operations on MongoDB ? Used pandas library to manipulate the data ? Consuming API s and get the JSON and parse the required information and load to MongoDB and SQL Server ? Used python modules such as requests, urllib, urllib2 for web crawling in Python

Python developer Citadel LLC - New York, NY September 2016 to July 2017

Project: Data Production - CDO DataProduction - CDO team helps data analysts providing data scraped from websites and loading to MongoDB and vendor data to database. Used python module Scrapy to scrape websites and get the data required information Used python modules such as requests, urllib, urllib2 for web crawling in Python Used other packages such as BeautifulSoup for data parsing in Python Used pandas to load vendor data to MSSQL after doing transformations Worked with data architects by creating scripts for transforming and loading data to MongoDB and MSSQL for their analysis Worked on connecting to MongoDB with pymongo module to perform CRUD operations Writing linux shell scripts for scheduling jobs in tidal Code deployment using Git and bitbucket Using tidal job scheduler to schedule crawler and parser for loading data to mongo and mssql Creating subscription process for loading data from vendor ftp to mssql and providing data to analysts

Environment: Python, MSSQL, Unix shell scripting, Putty, UNIX, Linux, WinXP, Git, MongoDB, Bit bucket, Tidal job scheduler

Python/Perl Scripting Engineer CISCO - San Jose, CA October 2012 to September 2016

Project: CEC (Cisco Employee

Connection) Migration CEC (Cisco Connection) is an integrated intranet website. It is Cisco's intranet starting point, providing the primary gateway to all Cisco internal websites and applications. Cisco CEC (www.in) platform which currently hosts many applications in Cisco San Jose Data Centers. Used python modules such as requests, urllib, urllib2 for web crawling in Python Used other packages such as BeautifulSoup for data parsing in Python Worked on writing and as well as read data from csv and excel file formats with Perl scripts Wrote Perl scripts to automate manual activities. Wrote python routines to log into the websites and fetch data for selected options. Loading vendor data into HDFS for storage and used Hive to query that data. Worked on connecting to Oracle database and fetch the data with Python Worked on connecting to MongoDB with pymongo module to perform CRUD operations Wrote Python/Perl scripts to parse XML/JSON documents and load the data in database. Managing off-shore team and assigning the work and monitoring with Rally Publish code and the supporting files in Stage and Production using Web updater tool Code deployment using Kintana Testing the code on State and Production environment. Checking the code by unit testing Documentation of User Requirement Specifications and System Requirement Specification Sending and Receiving Handovers to and from Offshore Knowledge Transfer to the newly hired Environment: Python, Perl, Oracle, Unix shell scripting, WinSCP, Putty, UNIX, Linux, WinXP, SVN, CVS, Hive, Hadoop Perl Scripting Engineer Morgan Stanley - New York, NY July 2011 to September 2012 Project: EDW OPS (Enterprise Data Warehouse Operations Support) Team supports the applications that come under the MSSB (Morgan Stanley Smith Barney) group under GWMG (Global Wealth Management Group). It supports Production, BCP (Business Continuity Plan), Dev, and QA regions. It is an Onsite-Offshore model with onsite mostly supporting Prod and BCP. Support activities include but not limited to Batch Monitoring in TWS (Tivoli Workload Scheduler), Job Recovery, Code Deployment, and Ticket Implementation. Roles and Responsibilities Performed Administrator related tasks: Granting access to App Dev teams to Informatica Repositories Bringing Infa services and repositories 'Up' when they go Down Removing Object Locks in Infa Admin Console whenever needed Handled Directory Space Issues Analyzed logs in Infa Admin Console during

outages and documented findings in Outage Tracker      Created new, Relational & Mload Connections as per requirement      Stay on top of emails sent to our group      Served as the Owner of CM (Change Management)      Scheduled & conducted weekly CM meetings/calls with App Dev teams for all Prod & BCP tickets      Gathered, Listed, Verified, Assigned, and Implemented daily Change Request & Service Request Tickets      Ensured all open tickets were implemented and closed by EOD      Deployed Informatica Mappings & Workflows to their target folders in Prod, BCP and QA Repositories      Recovered Failed Jobs in Informatica      Used WinSCP to migrate shell scripts, .Ini, .Config, .SQL files to their target directories in UNIX      Monitored the Daily Batch (around 1500 jobs) that runs in TWS (Tivoli Workload Scheduler)      Sending status emails to group with Batch Completion %, every 2 hours during system outages      Running Jobs On Demand in TWS and Shell Scripts in UNIX as per App Dev teams requirement      Analyzed Logs & Recovered Failed Jobs in TWS promptly and ensured a continuous and smooth batch run      Finding and removing invalid job/time dependencies in TWS      Sending and Receiving Handovers to and from Offshore      Knowledge Transfer to the newly hired      Environment: Perl, Informatica PowerCenter 9.1, Teradata, TWS (Tivoli Workload Scheduler), Teradata SQL Assistant, WinSCP, Putty, UNIX, WinXP Perl Scripting Engineer WISE - Pune, Maharashtra September 2007 to June 2011 Project: WISE Migration      The FWP Migration QA team was formed to focus on building a scalable, repeatable testing model, SDLC process improvement, environment provisioning /management as well as promoting greater education & awareness of QA best practices in use across Morgan Stanley for the back office warehouse migration.      Automating daily activities with Perl and Teradata      Developing HTML reports with Perl CGI      Extracting data from Teradata server by creating perl modules      Creating modules for base-lining the project      Verify data files, transaction log, table spaces, configuration etc.      Verify whether tables, views, stored procedures, triggers and other database objects created successfully      Verify schema, row counts, data reconciliation, data integrity, data correctness and validity      Verify ETL Jobs, middleware, application functionality, Bin to MSX Utility testing, Migration Suite testing      Verify performance meet user expectations/match or exceed existing system performance      Verify logins & users created correctly, permissions assigned

properly, data security at various levels      Generating test data from test specifications using TESTIFY tool      Checking coding compliance using Assent tool. This tool helps in checking if code is as per Clients standard. It helps in streamlining the entire process.      Implementing Quality Standards as per IQMS standards. This system provides guidelines for the conduct of every project and the means for monitoring it. It integrates the various internal processes and intends to provide a process approach for project execution      Setting up jobs in Autosys      Resolved the defects raised by the testing team in Mercury Quality Center      Technical mentoring of the ETL team and delegation of work      Defect Analysis and Interaction with Business Users during UAT/SIT      Created HLD's, LLD's, (High Level & Low Level Designs) and Detailed Design Documents      Environment: Perl, UNIX shell scripting, Teradata, Autosys, Putty, Mercury Quality Center, Windows XP, UNIX

Education Master of Computer Applications in Computer Applications Pondicherry central university - Puducherry, Puducherry 2007

Name: Robert Bishop

Email: stephaniecrawford@example.com

Phone: 5735765256