

Data Engineer Data Engineer Data Engineer - Capital One Richmond, VA ? Data Engineer with 2+ years of IT Industry experience in designing, developing and maintaining large-scale systems/applications with significant expertise in Python 2.7, 3.2 and 3.6. ? Experienced with full software development life-cycle, architecting scalable platforms, object-oriented programming, database design and agile methodologies ? Expert knowledge of and experience in Object Oriented Design and Programming concepts. ? Experience in writing data processing frameworks for large scale applications ? Hands on experience in using Spark ecosystem components like S3, RDS, Snowflake, REST, Kafka ? Experience in closely working with data analysis and Data scientists for converting POCs into production grade software. ? Experience in consuming data from Kafka into spark micro batches. ? Worked on spark application tuning and resource allocations based on use case ? Experience in Shell Scripting, SQL Server, UNIX and Linux. ? Experience in building applications in AWS infrastructure; Cloud Formation Templates, Cloud Watch Alarms, S3, RDS, Security Groups, VPCs, EC2. ? Familiar with JSON based REST Web services and Amazon Web services. ? Automated the continuous integration and deployments using Jenkins and AWS Cloud Templates deployment services (Lambda). ? Experience with web scrapping using beautiful soup; Created own flight logistic application using information from scrapped data. ? Experience in project deployment using Jenkins and using web services like Amazon Web Services (AWS) EC2, Cloud Formation Templates, AWS S3 and Cloud watch. ? Exposure to ML/DL ecosystem, algorithmic approaches along with design and performance constraints ? Exposure to Tensor flow and PyTorch frameworks & hardware for AI. Work Experience Data Engineer Capital One - Richmond, VA April 2018 to Present Capital One maintains an ecosystem of internal data processing applications designed to prevent fraud of various types including transactional and application fraud with dynamic fraud patterns. Ecosystem of highly available multiple spark clusters running rule on kafka streams and batch data. ? Developed multiplatform PySpark framework used to create spark jobs and provide SQL like interface for data analysts. This application can be used for both creating spark jobs and productionize them. ? Responsible for migrating findings from Data Scientist and fraud investigators into existing fraud defenses. And over seeing Data

Science Approaches so that functionalities can be converted to production environment. ? Designed data flow for new business needs and participated in architectural and workflow discussions. ? Responsible for maintaining allocating and tuning resources for faster performance for heavy data loads. ? Resolved Big Data small files issue in the organization using spark Hadoop configurations and dependency configurations. ? As part of Capital One Fraud Prevention team, worked on maintaining old and developing new rule engines along with application resiliency strategies. Created new Fraud defenses using existing data processing patterns ? Responsible for developmental and production data security aspects along with upgrading the system with new software and infrastructural features ? Converted exiting traditional Teradata fraud defenses into cloud architecture and spark SQL ? Create QA data by performing data analysis using Databricks notebooks with maintained entropy. ? Configured Apache Arrow for columnar data processing at processors in pyspark 2.3 ? Migrate exiting fraud defenses from Teradata to PySpark environment. Tune exiting spark jobs for performance. ? Created modules for switching cluster stack (active and inactive) using Jenkins, AWS lambda to EC2 cluster. ? Converted monolithic application into a pip installable package and incorporated it to schedule Jupyter Notebooks using papermill ? Built CI/CD pipeline using Jenkins and AWS lambda. Build secret management system. Built redundancy catch up in fraud case creation. ? Upgraded Spark Hadoop Version and participated in building Custom Assume Role Credential Provider in jar for spark session assume role auto renewal. Environment: Linux, Rel7, PySpark, Data Lake, Hadoop, Teradata, PSQL, Python, JSON, AWS, CI/CD, RISK management, Apache Arrow, Credit Card Fraud Prevention, REST, GitHub, Notebooks. Python Developer / Data Analyst WAFTS Solutions January 2017 to March 2018 Provide insights into data for Small to Mid-range businesses and present regular findings from data. ? Develop SQL code to be used with automated processes to identify revenue opportunities and financial issues. ? Data mine large datasets in Relational Databases to find emerging issues and root-cause in provisioning, marketing and billing systems ? Drive timely resolution of marketing issues for a seamless customer experience. ? Proactively monitor daily processes and results to ensure consistent coverage. ? Create REST APIs for web services using python Flask, Django frameworks. Environment:

Ubuntu, Python, SQL, JSON, AWS, REST, Jupyter Notebooks, GitHub. Jr Web Developer Repulsor Technologies - IN May 2015 to July 2015 Developed static websites for local businesses and organizations. ? Designed cover books and posters for technology events in undergraduate collages. Environment: Windows, HTML, CSS, XML, Adobe Photoshop Education Master of Science in Internet and Web Design in Internet and Web Design Wilmington University - Wilmington, DE 2017 Additional Information Professional skills Programming Languages: Python, Java, JavaScript Web services: RESTful Data bases: Oracle 10/11g, MySQL, SQL Server IDE's and tools: Eclipse, Pycharm, NetBeans OS &Environment: XP, windows, Linux, Unix, Ubuntu Unix Shell Scripting: Unix Shell Scripting Version control: GitHub Development Methodologies: Agile, Scrum Hadoop: HDFS, MapReduce, Spark Machine Learning: KNN, Gradient Descent, Back Propagation.

Name: Kristin Young

Email: blairchristopher@example.org

Phone: 2188360219