RAJIV PORUMALLA
408-718-9550
rajporu@gmail.com

SUMMARY:

- Around 14 years of strong experience in the Data Analysis, Testing, Design, Development.
- Involved in the Software Development Life Cycle (SDLC) phases which include Analysis, Design, Implementation, Testing and Maintenance.
- Experience on major components in **Hadoop** ecosystem (HDFS, MapReduce, Yarn, Pig, Hive, HBase, Flume, Sqoop, ZooKeeper, Oozie, Avro) and **Spark**.
- Experience in data analysis using **HiveQL, Spark SQL**, **Pig Latin** and custom **Map Reduce** programs.
- Experience in **Python, Java, Scala, MatLab, VBA, R, JavaScript, NodeJs, C, C++, HTML**.
- Solid Mathematics, Probability and Statistics foundation.
- Good Experience with **AWS** (EC2, S3, IAM, RDS, CloudWatch, AWS VPC, Route 53).

- Experience with **Git** and **Jenkins**.

- Proficient experience in different Databases like **Oracle, MariaDB, MySql, Postgres, Sybase, MS Access and DB2**.
- Experience in **NoSQL** databases like **MongoDB, HBase, Cassandra**.
- Good understanding of **Machine Learning** algorithms like Linear Regression, Logistic Regression, Bayes Models, Decision Trees, Random Forest and UnSupervised Learning.
- Good understanding of Deep Learning/ Neural Networks and technologies like **Keras** and **TensorFlow**.
- Good Experience developing and testing **SOAP and RESTful Web Services**.
- Proficient in coding **UNIX Shell Scripts**.
- Participate in designing the overall logical & physical **Data warehouse/Data-mart** data model and data architectures to support business requirements.
- Good understanding of **Data Warehousing** concepts and **ETL** (Extraction, Transformation and Loading) tools like **Informatica**, **Ab Initio and Talend**.
- Experience in Data Visualization using **ggplot2, D3.js, matplotlib, Tableau.**
- Good team player with excellent communication, analytical, interpersonal and writing skills.

EDUCATION:

- Master of Science in **Computer Science**, **Clemson University**, Clemson, USA.
- Bachelor of Technology in Engineering**, IIT Madras**, Chennai, India.

CERTIFICATIONS:

- **Deep Learning Nanodegree – Udacity**
- **Big Data for Data Engineers Specialization – Coursera**
- **Machine Learning – Stanford University – Coursera (Andrew Ng) – License WYXX3YDKAT2T**
- **The Data Scientist's Toolbox – Johns Hopkins University – Coursera – License GJZE87D7WZSL**
- **R Programming – Johns Hopkins University - Coursera**
- **Neural Networks and Deep Learning – Coursera (Andrew Ng) – License J4JQNAUF3RUZ**

- **Improving Deep Neural Networks: Hyperparamter tuning, Regularization and Optimization – Coursera (Andrew Ng) – License 8W5VQC8GQZAM**
- **Structuring Machine Learning Projects – Coursera (Andrew Ng) – License UYE2LXE8CWXP**
- **AWS Essentials – Udemy**
- **Taming Big Data with Apache Spark and Python – Udemy**
- **Apache Kafka Series – Udemy**
- **The Ultimate Hadoop Hands-On Hadoop - Udemy**
- **SUN Certified Java Developer**
- **C++ Programming for Financial Engineering – Baruch College (MFE), NY**
- **CFA (Chartered Financial Analyst) Level I**

TECHNICAL SKILLS:

Big Data/ML	Hadoop, HDFS, MapReduce, Spark, Python, Pig, Hive, Sqoop, Flume, Kafka, Yarn, Oozie, Zookeper, Machine Learning, Deep Learning, Keras, TensorFlow.
Languages	Python, Java, Scala, C, C++, JavaScript, NodeJs, R, MatLab, Octave.
Databases	MongoDB, Cassandra, CouchDb, Postgres, MariaDB, Oracle, MS SQL Server, MySql, Sybase, DB2
Tools	PyCharm, Postman, WebStorm, IntelliJ, Selenium, JMeter, Cucumber, Appium, Sauce Labs, SilkTest, Load Runner, Quality Center, Eclipse, Git, AWS, Postman, Splunk, MS Visio, SVN, Tandem Trading Applications.
ETL Tools	Ab Initio, Talend, Informatica
OLAP Tools	Tableau, Business Objects
Operating Systems	Mac OSX, Windows 8/7, Unix, CentOS, Linux, Ubuntu, MS DOS

EXPERIENCE:

Google, Palo Alto, CA **Dec 2017 – Present**
Software Engineer

Working on a project enhancing functionality and troubleshooting of Nest Devices like Indoor/ Outdoor Camera, Thermostat, Video Doorbell, Door Alarm.

Responsibilities:

- Developed Test framework using **Python** and **Unittest** for Nest devices like Camera, Thermostat, Video Doorbell, Alarm, Thermostat.
- Used Python libraries like **Multiprocessing**, **Subprocess**, **pexpect**, Queues to implement common Switchboard architecture for all devices.
- Developed Command Line Utility to test the features of Nest Devices like Factory Reset, Upgrade, Restart, etc
- Configuring and troubleshooting functionality of Nest Devices.
- Writing Bash Shell Scripts to assist in build automation.
- Build Automation using Jenkins.
- Used Git as Source Control repository.
- Followed Agile methodology with all tasks maintained in Jira. Participated in Sprint meetings.

Environment: Python, AWS, Git, Jenkins, Jira, Unix Shell Scripting, PyCharm, Ubuntu, MacOSX.

Symantec, Mountain View, CA **Nov 2013 – Nov 2017**
SDET

Worked on multiple projects like Hydra (AWS cloud based Data Loss Prevention Project) and SAM is a SAML based single sign on application that allows user to securely login to different applications. Facilitated insightful daily analysis of 60 to 80 GB of email data collected. Spawning violations and violation actions that increased the data security.

Responsibilities:

- Developed REST API test automation framework using **PyTest** and **Python**.
- UI Automation using **NodeJs**/ **Protractor** framework and **Java**/ **Selenium** framework.
- Extensively coded using **JavaScript** in writing UI automation test cases.
- Extracted the data from MySQL into HDFS using Sqoop and analyzed data by performing Hive queries (HiveQL) and running Pig scripts.
- Stored the data in Apache Cassandra Cluster. Used Impala to query the Hadoop data stored in HDFS.
- Managed and reviewed Hadoop log files.
- Implemented helper classes that access Cassandra using **Java** and Java API.
- Devops with **Ansible**.
- Developed UNIX shell scripts to assist in Automation.
- Used VMware Workstation and VMware Fusion virtualization to test the application in **OSX/ Windows/ Linux** images.
- Used **Jira** to maintain user stories and for bug tracking.

Environment: Python, AWS, Hadoop, JavaScript, Spark SQL, Sqoop, MongoDb, Java, Postman, Tableau, Talend 6.1/5.6, JMeter, REST WebServices, Sauce Labs, JSON, Jenkins, TestNG, Splunk, Oracle, Ansible, Postgres, MariaDb, Oracle, Git, Jira, AD, LDAP, Maven, Windows, Mac OSX and UNIX.

Visa, Foster City, CA **Jan 2013 – Oct 2013**
Sr QA Analyst

V.me allows customer to enroll, add a credit card and do purchase transactions online and helps merchant ensures safe and valid transactions. Various REST web services and business rules are invoked during customer transactions. This application is similar to the way we use **PayPal** for doing online purchases and adding/ maintaining various credit cards.

Responsibilities:

- Analyzed the Business Requirements Document and prepared detailed test cases for new Functionality.
- **Agile methodology** is used for this project. Attended daily stand-up meetings.
- All the business stories and bugs for the sprint are maintained in **JIRA**.
- Extensively used **SQL** queries for testing.
- Create **Hive** Queries which help analysts spot emerging trends by comparing fresh data with historical data.
- Load and transform large sets of structured and unstructured data.
- Create reports for BI team using Sqoop to export data into HDFS and Hive.
- Setup testing environments and prepare test data for testing flows to validate and prove positive and negative cases.
- Test and validate data at all stages of the ETL process.
- Tested **REST** web services using **SOAPUI (Groovy)** and **SOATest (Jython)**.
- Used Splunk tool in order to analyze the logs in the applications.
- Worked closely with developers and Product team to understand business stories and test the functionality properly.
- All the test scripts are maintained in **Git**. Business documentation is maintained in wiki pages.

Environment: Python, SOAPUI, MongoDb, Hadoop, HiveQL, SQL, Sqoop, Oracle, MySql, Git, ETL, REST, Maven, Jenkins, Splunk, UNIX.

Western Digital, Mountain View, CA **Aug 2012 – Dec 2012**
QA Engineer

Sequoia is a REST WebService based project with various additional features and enhancements for **WD NAS** devices (**WD MyBookLive**) before they are released to the market. This project involves two versions of NAS devices – Sequoia Mule and Sequoia Wireless. The code bases for these two devices are different. Also these two devices get different build version in each release.

Responsibilities:

- Analyzed the Business Requirements Document and prepared detailed test cases for new Functionality.
- Using Black box, Smoke, Sanity, Integration, Regression testing techniques and practices to test the application.
- Tested **REST** WebService calls made by the user/ NAS and automated them using **Python**.
- Installed **Sequoia Mule** builds and ran a smoke testing suite (Selenium scripts) to ensure the validity of the build.
- Ran the WebService test cases in **LAN/WAN** environment to ensure the proper functionality.
- **Jira** is used for defect tracking.

Environment: Python, REST WebServices, WD MyBookLive Mule, Oracle, Git, OSX, Windows7 and UNIX.

ADP, San Dimas, CA **Feb 2012 – Aug 2012**
QA Analyst

CCP Application has various modules like Security Services Administration, Payroll Tax Input, Agency Notice Manager, and Tax Reporting Online. These modules help the clients in their businesses by doing Payroll and Tax processing. Also various reports are generated in each module to assist the client.

Responsibilities:

- Experienced in defining Testing Methodologies, Designing Test Plans and Test Cases, Verifying and Validating Web based ADP applications, documentation based on standards for Software Development and effective QA implementation in all phases of Software Development Life Cycle (SDLC)
- Analyzed the Business Requirements Document, created the test plan and prepared detailed test cases for new Functionality
- Used **Agile methodology** (Scrum) to ensure the new features meet the customer requirements and acceptable quality assurance standards.
- Developed **Vuser scripts** with **Load Runner** for performance Testing. The scripts were developed using **Web, Web services, RTE, SAP GUI and Database protocols.**
- Performance and Functional testing of reports developed using **Business Objects.**
- Identified and documented defects using **Quality Center** to ensure applications functionality.
- Created Test Plan, generating test data, designing test cases, preparing test environment

Environment: Load Runner, Java, Business Objects, Quality Center, Oracle, Windows 7, Web Sphere and UNIX.

Chicago Mercantile Exchange, Chicago, IL **Jan 2008 – Jan 2012**

QA Analyst

EAGLE (Electronic Arbitrage GLOBEX Liquidity Enhancer) application for **FUTURES** market is a trading engine that operates within the **GLOBEX** Electronic Trading environment. EAGLE offers increased liquidity in the Eurodollars market by allowing complex trading strategies to be executed electronically. The major features of EAGLE: Implied Spreading, Protection Points, In-Flight Fill Mitigation, Stop Price Logic. Now EAGLE is merged into single **Match Engine** to handle both FUTURES and OPTIONS market.

Responsibilities:

- Prepared Test Strategy Document, detailed Test Plans and Test Cases which specify testing overview and testing approach depending on the Business and Technical requirements.
- Developed **Python Scripts** to assist in testing trading functionality.
- UI Automation using **Java**/ **Selenium** framework.
- Good experience with Market Data messages in **FIX** and **NSC** format.
- Good understanding of various tags of **FIX Order Entry and Market Data messages** (based on **FIXML** format). Extensively tested the validity of the messages generated and the data residing in it.
- Analyzed and tested the functionality of various **Trading Allocation Algorithms** (F, C, K, A, N, O, T, Q, S) used for quantity allocation to different traders, buying/ selling in a dynamic global trading environment. The instruments in different Trading Engines (EG1, IMP, EDO, NYMEX, CBOTY, CBOTZ) following different algorithms.
- Tested the functionality of Complex Trading Strategies like Soy Crush, Euro Dollars and Treasuries.
- All documents were maintained in **Rational Clear Case/ SharePoint/ GIT**.
- Proficient in using **Tandem Trading Software Applications**.

Environment: Java, JUnit, JIRA, ANT, Selenium, SOAP, Rational Clearquest, Jira, Rational Clearcase, Git, Tandem Trading Applications, FIX/NSC Protocol, Perl, Shell Scripting, XML, HTML, MS Office, SQL, PL/SQL, WSDL, Business Objects, Web Services, Unix and Windows.

Fannie Mae, Herndon, VA **Jan 2004 – Dec 2007**
QA Engineer

The Securities Cost Basis Sub-Ledger (**SCBSL**) is created based on the need for Fannie Mae to have one central location to calculate the results of different accounting events that affect the cost basis of the securities that it purchases, holds and eventually sells. The purpose of the SCBSL **Data Mart** project is to develop an **ETL** process that is capable of extracting Securities data from source accounting systems and populate an Integrated Data Store (IDS) which is the source of a second ETL process that populates a Data Mart at an accounting period level.

Responsibilities:

- Analyzed BRD and technical documentation. Developed **Test plans** and Test cases based on software requirements.
- Carried out extensive testing with different test cases, which reflect the various real time business situations.
- Participate in designing the overall logical & physical **Data warehouse/Data-mart** data model and data architectures to support business requirements.
- Created and tested **Ab Initio ETL** graphs performing data transformation, conditioning and loading. Validated the graphs, Performed unit testing and Integration test.
- Involved in the preparation of documentation for **ETL** using **Ab Initio** standards, procedures and naming conventions.
- Generated Database Configuration files (.dbc), **DMLs** which specify the record format with delimiters. These are used in components for building graphs in Ab Initio.
- Defect tracking is done using **Rational Clear Quest/ Quality Center**.

- Developed **PL/ SQL, Stored Procedures, Cursors, Packages, Functions and triggers** extensively in **Oracle**.
- **Quality Center** is used to maintain requirements and test cases. Also the test cases were executed in Quality Center for each new release.
- Developed **Shell Scripts** to automate report execution. Executed **Autosys jil** scripts to execute reports and ETL graphs.
- Used **UNIX** environment variables in various .ksh files.
- The reports were generated using Business Objects using various functionalities such as Sub reports, Formulae, Drill Down, Date based triggers etc.,
- Developed Shell Scripts to validate various business reports.
- Used inquiry and error functions like is_valid, is_defined, is_error and string functions like string_substring, string_concat and other string_* functions in developing Ab Initio graphs to perform data validation and data cleansing.
- Involved in Testing of GUI and Multiple Reports with different roles for security.
- **Business Objects** Report Testing using SQL queries to verify accuracy of various reports like Control, Management and Financial Reports.
- Involved in End-to-End testing of processes that migrate data.
- Interacted with developers to resolve issues.
- Actively participated in regular project status meetings.

Environment: Java, TestNG, ANT, Eclipse, Maven, Tortoise SVN, Quality Center, PL/SQL, Unix Shell Scripting, Oracle 10g, Business Objects, SQL Loader, PeopleSoft Tools 8.44, Ab Initio GDE 1.13 Co-op 2.13, Rational Clear Quest, Rational Clear Case, Shell Scripting, MS-Office (Word, Excel, PowerPoint), Unix.

MACHINE LEARNING PROJECTS (SELF):

- Detect Car Lanes and Traffic Signs for the **Self Driving Car Project** at Udacity (Ongoing).
- Evaluated six different algorithms (**Logistic Regression, k-Nearest Neighbors, Linear Discriminant Analysis, Gaussian Naïve Bayes, Classification and Regression Trees, Support Vector Machines**) on iris data set using R and Python. Used various packages like **pandas, sklearn, matplotlib, scipy**. Data Visualization using Univariate and Multivariate plots.
- Investigated Boston House Price dataset using Linear Algorithms (Linear Regression, Lasso Regression, ElasticNet) and Nonlinear models (Classification and Regression Trees, Support Vector Regression, k-Nearest Neighbors).
- Sonar Mines vs Rocks dataset (predict metal or rock objects from sonar return data): Binary Classification Predictive modeling problem in R and Python. Used data transforms to improve model performance. Performed algorithm tuning to improve model performance. Used ensemble methods and tuning of ensemble methods to improve performance.
- Working on Yahoo Stock Market data to make S&P predictions using Logistic Regression, Naïve Bayes and Long Only.