**Big Data Platform Lead**

[jay@quantomtech.com](mailto:jay@quantomtech.com)

720-805-0982

Over 16 years of working experience as specialized in providing Data & Analytics expertise that result in real and measurable business outcomes. Evaluates and implements Big Data and advanced analytics solutions on-premise and in the Cloud. Passionate about data, problem solving and driving value to make a difference

**PRIMARY RESPONSIBILITIES**

* 4 + Years of experience as Big Data Platform Lead at Wells Fargo
* I have managing 12 Application teams and about 1500 business users.
* Experience in design and build Big Data solutions in Banking and Financial industries.
* Experience in end-to-end design and build process of Near-Real Time and Batch Data Pipelines
* Experience in design and build data management lifecycle covering Data Ingestion, Data integration, Data consumption, Data delivery, and integrating Reporting, Analytics and System-System integration.
* Proficiency with tools like Alteryx, Amazon S3 and Dremio to automate workflow and data pipelines.
* Validate and select the technologies that support business requirements.
* Analyze the production defects, identifying root cause of the issues and coordinating with application teams to fix the issue
* Hands-on role in crafting and executing technical proof-of-concept initiatives to determine proper configurations, design patterns and general tool fit for the platform tenant.
* Work closely with the other architects on Big Data, Databases, Analytics or Security, as well as the Business Development teams, to enable large-scale customer use cases and drive the adoption of Data solutions for their data processing platforms.
* Create field enablement materials for the broader Data Analytics Technology population, to help them understand how to use the Data Capabilities.

**TECHNICAL EXPERTISE**

10+ years of hands-on experience in Hadoop, Hive, Spark, Pig, HBase, MongoDB, Oozie, MapReduce, Talend, Dremio, Jupiter Notebook, Alteryx, Power BI, Tableau, with MapR and Cloudera Hortonworks distributions.

4 + Years had experience as Big Data Platform Lead.

* Extensive experience Data analytical tools Dremio and Alteryx.
* Good Experience generating reports with Power BI
* Experience and detailed knowledge of resource managers: YARN
* Extensive experience in performance optimizations on Spark Jobs.
* Experience in file formats Avro, Parquet / with arrow data structure, unstructured data and/or text analytics.
* Hands on programming experience in Scala, Python, Shell Scripting.
* Experience and proficiency with Linux and Windows operating system.
* Experience using Source Code and Version Control systems like SVN, Git, bitbucket.
* Knowledge on CICD Tools and containerization technologies: Kubernetes, Dockers.
* Experience in UML diagrams to visualize the system designs.
* Delivered solutions using Agile methodology and Waterfall methodology.
* Knowledge on Terraform and Ansible to provisioning and configuring the cloud resources.

**SKILLSET SUMMARY**

|  |  |
| --- | --- |
| **Big Data Technologies** | Hadoop, Spark, MapReduce, Hive, Pig, Sqoop, HBase, Talend, Denodo |
| **Streaming Technologies** | Spark Structure Streaming |
| **Resource Managers** | YARN |
| **Programming Languages** | Python, SQL, Shell Scripting, Panda’s data frames |
| **Package Managers** | PIP, Conda, Maven |
| **Messaging Systems** | Kafka |
| **Databases** | HBase, Oracle, SQL Server, DB2, PostgreSQL, MongoDB, |
| **Data Visualizations** | Tableau, PowerBI |
| **Analytical Tools** | Dremio and Alteryx |
| **Workflow Tools** | Cron, Oozie, Autosys |
| **Cloud** | GCP Suite Services |
| **DevOps** | Git, Jenkins, Nexus, Ansible |
| **Business Domains** | Banking (Fraud & Risk, Operational Risk, Compliance) |

**PROFESSIONAL EXPERIENCE**

**Wells Fargo - Phoenix, AZ** 12/2021 to Present

***Role:*****Big Data Platform Lead**

**Risk Data Management Platform:** Risk Data Platform is responsible for providing consist, authorized, managed data in right format essential for enabling bank’s target Risk Management Framework. RDM architecture goal is to ensure that common functions across domains are managed through standardized, strategic technical capabilities that leverage Enterprise data and technical strategies and accurately and effectively produce critical outputs.

* Designed and implemented batch data processing pipelines using, Cloudera Hortonworks distributions and Apache Spark.
* Had good experience on implementing Ranger polices and KTABs
* Hands on experience in **Spark** and creating **DF** & applying operations transformations and Actions.
* Had Good Experience on onboarding tools like Alteryx, Power Bi, Tableau and S3 to support analytical use cases.
* Developed scripts to generate cluster reports like PROD and BCP compare and PROD inventory reports.
* Created Dashboards using Power BI to show cluster stats like no of applications, size and memory utilizations.
* Creating alerts for capacity utilizations using Grafana and Splunk

.  Had good experience on GCP tools like GCS, and Big query.

. Involved in Architectural designs discussions and conducting POC on different tools

**Wells Fargo - Phoenix, AZ** 12/2019 to 11/2021

***Role***: Ops Lead

**Risk and Finance Data Services (RFDS):** It is centered on core technological capabilities for Risk data integration, calculation processing and reporting. RFDS Platform is responsible for providing consist, authorized, managed data in right format essential for enabling bank’s target Risk Management Framework. RFDS architecture goal is to ensure that common functions across domains are managed through standardized, strategic technical capabilities that leverage Enterprise data and technical strategies and accurately and effectively produce critical outputs.

* Extensively involved in the fixing the performance issues.
* Developed scripts to generate cluster reports like PROD and BCP compare and PROD inventory reports.
* Created Dashboards using Power BI to show cluster stats like no of applications, size and memory utilizations.
* Creating alerts for capacity utilizations using Grafana and Splunk
* Developed **Spark** applications using **Scala** for easy **Hadoop** transitions.
* Used **Spark** and **Spark-SQL** to read the **parquet** data and create the tables in hive using the **Scala API**.
* Involved in converting **Hive/SQL** queries into **Spark** transformations using **Spark DF**, **Scala** and Python.

**Impetus - Phoenix, AZ** *06/2018 to 11/2019*

*Role: Sr Big Data Engineer*

**MatrixKRI:** Key risk indicators (KRIs) are measures used to provide a proactive, early warning of potential operational risk events or exposures that may be trending towards exceeding acceptable thresholds. Business units, staff groups and bank and regulated entities must prepare and maintain a relevant and appropriate set of KRIs to proactively monitor and reasonably anticipate operational risk events. KRIs tracks the exposure of most important operational risks and they serve as an early warning system that predicts large operational losses before they occur. KRIs are alert management to issues that must be addressed when certain thresholds are breached.

* Closely worked with business and product owners to understand the business requirements to generate key risk measures.
* Developed Pig scripts to create structure data from unstructured data.
* Streamed data from data sources Teradata, PostgreSQL to ODL (HIVE Warehouse - organization Data Layer).
* Developed Spark python scripts to create structure data from unstructured data.
* Data Exploring using Pandas and Spark and Feeding Data to Time series model to predict Anomalies
* Extracting Data from upstream application as json file and ingested into topics using Apache Kafka
* Spark Submit arguments are integrated with Unix Scripts to build and launch Spark Applications.
* Deployed/Scheduled use case scripts in event engine. It is internal scheduling engine.

**Cognizant - Phoenix, AZ** *10/2015 to 05/2018*

*Role: Sr Big Data Engineer*

**CR2020 ODL:** The vision of the CR2020 is to eliminate Credit Risk by analyzing attributes associated to accounts. There are two pieces to this project namely Data Ingestion (DI) and Data Management (DM). Data Ingestion team gets the data from SOR and places files/Tables in the Hadoop Ecosystem Cornerstone. The DM team Develops variables to be built for each use case as defined by the business users from the data available in cornerstone. Cornerstone is centralized database with organized in consumable and friendly formats.

* Design and Built Hive Data warehouse: Avro (staging tables), ORC/parquet (Final tables), data encryption.
* Design and Built Customer 360 logic in hive, Stored Final data hive external tables HBase as storage handler.
* Design and Built OOZIE workflow and Coordinators to schedule and orchestrate the Batch pipelines.
* Design and Built REST API serving layer: Performs ROEC logic iteratively and get optimum credit limit by consuming
* the Realtime inputs from Customer and pre-processed HBase customer 360 data.
* Design and Built Report and visualizations on decision making data through Tableau.

**Goldman Sachs - Jersey City, NJ** *08/2013 to 10/2015*

*Role: Big Data Engineer*

**QUANTUMM:** The purpose of the project is to perform the analysis on the Effectiveness and validity of controls and to store terabytes of log information generated by the source providers as part of the analysis and extract meaningful information out of it.

* Involved in moving all log files generated from various sources to HDFS and Spark for further processing.
* Written the Apache PIG scripts to process the HDFS data.
* Created Hive tables to store the processed results in a tabular format.
* Monitoring Hadoop scripts which take the input from HDFS and load the data into Hive.
* Created external tables in Hive.
* Spark used for processing data from HDFS, AWS

**UBS - Hyderabad, India** *03/2011 to 07/2013*

*Role: Technology Analyst*

**Infy Dashboard (Desktop Transformation Program):** Infy Dashboard is a web application that consolidates cleared transaction information across multiple Line of business groups. Provide ad-hoc query functionality via dashboard on newly created data repository for further investigation. Tool allows multiple users to query with AD authentication. The purpose of the project is to provide a reconciliation platform between clearing houses, clearing brokers and our internal booking system for Rates and Credit.

* Build a consolidated dashboard to show the status of each daily reports.
* Build a report monitoring section to monitor the status of parsing each report and show various metrics in each report.
* Developed a reconciliation tool to reconcile on various data fields between reports and P/S data fields.
* Developed a report delivery section to generate ad-hoc reports and download them on the fly from portal
* Developed a security model to authenticate access to only authorized users
* Developed archiving and log generating section to save reports and logs.
* Developed email delivery of reports to email boxes which are authorized by middle office.

**CRICO/RMF - Chennai, India 08/2008 to 03/2011**

*Role: Software Engineer*

**SSO**: CRICO/RMF is an insurance company and provides insurance to the doctors of Harvard Medical School and many other hospitals. Based on the ‘Harvard-affiliated medical institutions' twenty plus years of medical malpractice claims management experience, the RMFS “Claims Management, Analysis and Processing System” (C-MAPS) provides a web-enabled, user-friendly, and comprehensive medical malpractice transaction processing and clinical coding system

* Involved in writing Technical Design Documents based on Functional Specification document.
* Involved in writing Coding using ASP.NET and C#.NET, WCF, EL
* Using ADO.NET for interaction with database
* Used Validation controls and JavaScript in validating user inputs
* Involved in writing Unit test scripts and Unit test cases
* Use Style cop, a Microsoft developed tool, to ensure readability and styling for the code.
* Fixing the issues
* Involved in write Unit test scripts and Unit test cases.

**EDUCATION**

Bachelor of Technology in Electronics SVU (India), 2005

**TRAININGS & CERTIFICATIONS**

* Certified MapR Spark Developer
* Certified Google Cloud Associate Engineer
* Certified Microsoft Azure Fundamentals
* Certified Google Cloud Platform Big Data and Machine learning Fundamentals from Coursera
* Certified Leveraging Unstructured data with Cloud Data Proc on GCP from Coursera

Certified Serverless data analysis with Google Big query and Dataflow from Coursera