## **Purushothama Manchu**

**SR Big Data/AWS Developer Email ID:** [**pmanchu.dataeng@gmail.com**](mailto:pmanchu05@gmail.com) **Contact No: +1 571-789-3140**



SUMMARY

* 15 years of experience in Enterprise/Distributed Data/Application Development in diverse industries includes hands-on experience in AWS/Big Data/Hadoop/PySpark/Java-J2EE/Scala related technologies.
* Experience in using HDFS, MapReduce, Hive, HBASE, Sqoop, TDCH, Oozie, Spark, Scala/Java/ /PySpark, Cassandra, Kafka, Kerberos, AVRO/Parquet/txt file formats and Cloudera (CDH).
* Experience in using AWS EMR, EC2, Glue, S3, Lambda , AWS API’s and Athena for developing pipeline/ELT processes.
* Experience in writing Spark/Scala/PySpark ELT jobs, writing wrapper/shell scripts to execute hive/spark queries/jobs using spark.
* Experience in importing and exporting data using Sqoop/TDCH/JSON/Text files from Relational/Teradata Database Systems to HDFS and vice-versa.
* Extending Hive and Spark core functionality by writing custom UDFs.
* Worked on job workflow scheduling and monitoring tools Oozie and Zookeeper.
* Experience in designing, developing and implementing connectivity products that allow efficient exchange of data between our core database engine and the Hadoop ecosystem.
* Experience in developing Web/intranet, client/server applications using Core Java, J2EE, Servlets, JSP, JSF, React JS, Struts, Spring, JDBC, Hibernate and SQL.
* Worked on SOAP and Restful Web Services.
* Expertise in code quality & coverage by using SONAR, Fortify and Cobertura/check style plugin integration with Eclipse.
* Techno-functional responsibilities include interfacing with users, identifying functional and technical gaps, estimates, designing custom solutions, development, leading developers, producing documentation, and production support.

Excellent interpersonal and communication skills, creative, research-minded, technically competent and result-oriented with problem solving and leadership skills. Certified SAFe Agilist.

tools

**Hadoop/Big Data Tools: Hadoop 2.x, HDFS, MapReduce, Hive, HBase, Spark, PySpark, Snowflake, Scala, Sqoop, Flume, Cassandra, Oozie, Zookeeper and Cloudera.**

**Java & J2EE Technologies: Core Java, Servlets, JSP, JDBC, JNDI, Java Beans, SOAP Web Services.**

**Cloud Tools: AWS EMR, S3, Lambda and Athena.**

**Java/J2EE Frameworks: Struts, Hibernate, Spring, Spring Boot.**

**Programming Languages: Java, Scala, Python, Linux shell scripts.**

**Databases: Teradata, Cassandra, Oracle 11g/10g/9i, DB2, PostgreSQL, My-SQL.**

**Web Servers: NGINX, Apache Tomcat, WebLogic, WebSphere and JBoss.**

**Web Technologies: HTML, XML, JavaScript, AJAX, jQuery, DOJO, Node JS, React.**

**Atlassian Tools: JIRA, Confluence, GIT and Jenkins/Nexus.**

**Testing: SOA Test, SOAPUI and Spring4 JUNIT.**

**Build Tools: Gradle, Maven and ANT.**

**CICD Tools: Terraform, Ansible, Jenkin’s, Bamboo, Hudson , IBM UDeploy.**

**Scheduling Tools: Apache Airflow, Control-M, Autosys and Oozie.**

**professional work experience:**

* Working as Senior Data Engineer/AWS Developer in TeckPros LLC from MAR 2021 to present.
* Worked as Senior Data Engineer in C-HIT LLC from JAN 2018 to MAR 2021.
* Worked as Senior Software Engineer in Fidelity Business Services India Pvt Ltd( aka Fidelity Investments) from JUN 2011 to APR 2017.
* Worked as Senior Software Engineer in Mphasis from OCT 2010 to JUNE 2011.
* Worked as Software Engineer in Zensar Technologies from JUN 2008 to OCT 2010.

***Big Data Projects Executed:***

Client: Experian Mar 2021 to Present

**Project: GVAP/Ascend**

**Role: Senior BigData/AWS Developer (Consultant)**

Global Value Added Products(GVAP) enables the clients to submit the Campaigns to generate the data/files/reports that will be used for Ascend Batch Products by running Ascend Models like Vantage, FICO, Premier, Equity, Insight, Bankruptcy, Financial Personalities, HealthCare Payments Risk etc to identify the customers data eligible for Account Reviews, Pre-Screen, Quest and Archives.

Environment: Java, Scala, Python, PySpark, AWS EMR, Glue, S3, Lambda’s , Athena, GIT, JIRA, JUnit, Jenkins, Terraform , Apache Airflow.

**Responsibilities:**

* Design and development of new functionalities in the GVAP with implementation of micro services oriented Batch products and services using Java/Spring(REST/MVC) and AWS RDS..
* Developed the code to import/export the data from/to S3 (using Glue/Spark ELTjobs) that will be input/output for the Ascend Batch Products.
* Developing the Ascend Models using Spark/PySpark with AWS Lambda functions and storing the data in S3/Athena databases.
* Running the Ascend Models with 1% population data, verifying the sanity checks of the results and then working on co-ordinating for promoting the Ascend Models to run on the 100% population data.
* Creating airflow dags for the ETL Pipelines to invoke the python/shell scripts.
* Developed the demon threads to apply/purge the data in Athena/S3 as per the Retention Policy applied in the Campaigns.
* Developing the config files/scripts for creating the tables in DynamoDB and populating the reference data in it like Registering the Data Assets with CDA metadata and PII information, Service tables that are needed for ETL Flows.
* Working with the Business analysts team to capture the Data Model and mapping requirements, documenting the details and developing and verifying the details in implementations through ITSM.
* Mentoring and guiding the team members and coordinating with Product Owners/Architects on the requirements and solutions implementation.
* Working with SRE Team to onboard the new AWS features/Tools/Roles required for the applications, getting access provisioned for the new clients S3 inbound/outbound/deliver buckets, moving applications running in public IP’s to Private IP’s as part of Cloud migration.
* Building Cloud Formation templates/Terraform code for the AWS Cloud Infrastructure / code components deployments.

**Project: Common Data Access - Sandbox**

**Role: Senior Big Data/AWS Developer (Consultant)**

Common Data Access(CDA) pulls the data from multiple source/external systems , transforms the data into BU Specific requirements and enables it to be used for the downstream applications like GVAP/Campaign systems..

Environment: Snowflake, Snowpipe, AWS S3, .

**Responsibilities:**

* Worked on modeling the data attributes required for Clarity/First American Property data.
* Worked on developing on automation of Clarity/First American Property data ingestion from vendor provided files/feeds into S3 using Snowflake Snowpipe.
* Worked on developing the Snowflake ETL Process by consuming S3 data , applying the transformation to derive new fields and building the data in required format for the consumption of the downstream application.
* Worked on design and development of Snowflake ELT jobs to run in AWS Services - EC2, S3.

Client: JP Morgan Chase, OH Mar 2020 to Feb 2021

**Project: WorkForce Data Analytics**

**Role: Senior Big Data/AWS Developer (Consultant)**

Workforce Data Analytics is built to evaluate industry leading products that will allow JPMC to ingest Human Capital Management, Taleo from various systems and structures, semi structured and unstructured formats into Cloudera Hadoop Big Data Lake Platform.

Environment: Java, Scala, Spring, Spring Boot,Hadoop, Spark, HDFS,, Hive, Impala, Sqoop, Autosys, Maven, Jenkins, GIT, JIRA, JUnit, Mockito,, HBase, AWS EMR/Lambda and Terraform..

Responsibilities:

* Developing the code to run the pyspark jobs using AWS EMR Step Functions.
* Developing the workflows/Lambda functions to get the data from Oracle Cloud Platform Learning data using REST Webservices, ingesting the data into S3 by running spark jobs, applying the business logic to transform the raw data and building the data in Conformed zone and then pushing the data to Tableau Manager Dashboard and to SQL Server for downstream applications.
* Developed the code/workflows to detect the small hdfs files for the WDA data and Athena/Hive repartitions using Hadoop Filesystem API using Spark/Scala.
* Developing the CICD pipeline using Jules. Jenkin’s with unit test cases develop/execution and deploying the code in dev environments.

Client: Credit Suisse, NY May 2019 to Feb 2020

**Project: HARMONICA/Finance One Reconciliation**

**Role: Senior Big Data Developer (Consultant)**

Harmonica is a single calculator for harmonizing the source data and performs Basel regularity calculations for Risk Assessment for Investment Banking, Private Banking and Counterparty Credit Risk.

In Finance One Reconciliation we source the data from Ledge/Subledger , builds the dataset queries , combine the the Reconciliation attributes data using common aggregation key and builds/generates the aggregation keys unising the F1 Web UI Tools and then run the reconciliation jobs in Hadoop Cluster. Generates the Exception Data and reports the data to downstream and to the Qlik Dashboards.

Environment: Apache Hadoop 2.6, Kotlin/Java , Spark , Impala, Hive, HDFS, Sqoop, Oracle, Cloudera, Kerberos, Bit Bucket, Control M, Spring Boot, Spring, REST API, React JS and NODE JS.

Responsibilities:

* Developing the workflows to sqoop the data from Oracle, harmonizing the data using Spark Datasets using Java, applying the business rules/formulas/calculations and storing the data into Hadoop cluster that are used for the reporting layer.
* Developing the React UI pages to upload the various Adjustment and UnAdjusted files like Multipliers, DeductionTypes, AssetClass, SubClass, Collateral Type etc., writing the REST Controller, validation of data and persisting the data into Hadoop.

**Project: Capital Calculation Engine (COMET)**

**Role: Senior Big Data Developer (Consultant)**

Capital Calculation Engine (COMET) includes regulatory requirement calculations according to Basel 2 , Basel 3 and it performs capital calculations for different Regulatory’ s like FINMA, PRA, and FRB(IHC).

Environment: Apache Hadoop 2.6, Java, Spark, Spring, Hibernate, Impala, Hive, HDFS, Sqoop, Oracle, Cloudera, Kerberos, Bit Bucket, Bamboo and Control M.

**Responsibilities:**

* Participate in Project/Sprint planning, Sprint Demo’s/Grooming sessions and participates in development activities planning and execution.
* Involve in pulling data from multiple source databases using Sqoop and storing the data in Hadoop Distributed File System.
* With the stored data in Hadoop Distributed File System we apply data massaging, enrichment and transforms as per business rules/mapping by creating data workflows using MapReduce/Spark to build finalized data into Hive tables.
* Build the application frameworks with the combination of UNIX, Java and complex SQL queries with best design patterns, best practices for high performance, scalable, extensible applications.
* Involved in performing Data Quality checks using automated scripts to make sure we are getting correct data into our Hadoop storage layer from Source Systems.
* Create Test Plans, Test Cases and perform unit and integration testing.

Client: Anthem, VA Jan 2018 to April 2019.

**Project: Product Analytics 2.0 (PA)**

**Role: Lead Software Developer (Consultant)**

Product Analytics 2.0 is an enhanced platform to bring the data from multiple source systems like ACES/SPIDER/NASCO/OBIS/WEST-ISG-WPD, process and filter the data as per business needs and provide enhanced reporting for the Anthem Health Products that include subject areas like Product, Membership, Claims, Group, Broker and Commission, Revenue. This project is aimed to develop enhanced reporting on the products usage and products performance for the Product Managers and the company.

Environment: Apache Hadoop 2.6, PySpark, Hive, HDFS, Sqoop, TDCH, Streamset, Oozie, Python, Teradata, Cloudera, Kerberos, Kafka, Bit Bucket, Bamboo, UDeploy and Control M.

**Responsibilities:**

* As a Lead Developer, leverage daily experience with Enterprise Data Solutions (EDS) technology initiatives to develop an integrated approach to planning and problem solving for the current and future reporting needs.
* Develop Pyspark code for creating DataFrames by reading data from JSON/hive tables, adding the joins, applying the transformations and then persisting the transformed data to hive tables.
* Involved in updating existing/new Teradata scripts for tagging new Product fields over Mbrshp, Clm, Cpttn, Grp, Rvnu.
* Consuming the data produced by Kafka, validating the schema and consuming the data into Hive flatten structure.
* Develop processes to improve functionality, productivity and reduce defects.
* Coordinate between the development Data Model/Mapping and testing teams to synchronize development and testing, data mine for unit testing, review and provide project feedback to management and deliver the key modules.
* SAFe agilest Work with dependent agile teams to coordinate the requirements, development and delivery.
* Providing expert input and transfer of specialized knowledge will be vital to eliminate project delays, provide synergy between teams and deliver value to the business requirements.

**Project: Consolidate / Integrate Specialty Data warehousing and reporting (CISR)**

**Role: Lead Software Developer (Consultant)**

Consolidate / Integrate Specialty Data warehousing and reporting(CISR) is an initiative to bring the data from multiple source systems, process and filter the data as per business needs and provide enhanced reporting for the Large Group Rating System that include subject areas like Claims data, Membership and Group data, Broker and Capitation data, and Billing Data.

Environment: Apache Hadoop 2.6, HDFS, Sqoop, TDCH, Oozie, Spark, Hive, Java, Scala, Cloudera, Kerberos, Bit Bucket, Bamboo , UDeploy and Control M.

**Responsibilities:**

* As a Lead Developer, leverage daily experience with Enterprise Data Solutions (EDS) technology initiatives to develop and integrated approach to planning and problem solving for the Consolidate / Integrate Specialty Data warehousing and reporting.
* Involved in building Membership Incremental/historical data feeds to MDM team and scheduling these jobs for the generating Master Consumer Identifiers (MCID) and integrate the MCID back to Claim tables.
* Coordinate between the development Data Model/Mapping and testing teams to synchronize development and testing, data mine for unit testing, review and provide project feedback to management and deliver the key modules.
* Work with dependent agile teams to coordinate the requirements, development and delivery.
* Provide real-time decisions during the execution of the project.
* Provide expert input and transfer of specialized knowledge will be vital to eliminate project delays, provide synergy between teams and deliver value to the business requirements.

Company: Fidelity Investments, NC May 2014 to April 2017

**Project: Participant/Plan/Client Asset Allocation Calculation - High Volume Batch**

**Role: Senior Big Data Developer**

This project is used to Calculate Participant, Plan and Client Asset Allocation in a given High Volume Batch (Quarterly/Monthly) that we will need for existing/upcoming efforts including Rules Based Guidance, Retirement Readiness, GPS/Scoring Utility in Portfolio Analytics. ß

Environment: Apache Hadoop 2.4, HDFS, Sqoop, Oozie, Spark, Cassandra, Java, Scala, Cloudera, Kerberos, Git , Maven, Jenkins, UDeploy and Control M.

**Responsibilities:**

* Responsible for building scalable distributed data solutions using Spark with Scala programming language.
* Worked on Apache Spark with Scala (DataStax Spark Cassandra connector) to load batch data to/from Cassandra using Spark SQL Dataset, creating RDDs by pulling data from Cassandra for data aggregation using Spark Transformations/actions and then storing the aggregated data into Cassandra for the data usage in downstream applications.
* Worked on creating Scala Traits for common reusable components to get the spark session, Load Log Key, Load DTM, UUID and utility methods, used Scala to join the data from multiple tables, applying filters on datasets etc.
* Involved in Spark job configuration in Hadoop YARN Cluster, building and deploying the application in Hadoop cluster.

**Project: Rep Notes**

**Role: Big Data Developer**

Interaction HUB (iHUB) is the central repository for interactions between an individual and Fidelity across Personal and Workplace Investments (PWI). As part of a larger transformation effort to improve customer service and better predict the next best actions of a customer, Fidelity is creating a Hub in in Hadoop Hive (HCAP) to collect all interactions with the customers across all the channels i.e. campaigns, web, phone, customer service.

**Project: PWI Clickstream Query Tool**

**Role: Big Data Developer**

PWI Clickstream Query Tool is a Self-Service Tool used by PWI users for Legal and Compliance purposes. This tool uses the clickstream data stored in Nextgen Web Stats and leverages the reports to Fidelity’s Click2Compute (C2C) Cloud servers.

**Project: NextGen Web Stats**

**Role: Big Data Developer**

WebStats data warehouse has lived on a Legacy traditional ETL platform. The incoming web logs are processed using Perl and Informatica and the data is persisted in an Oracle Database. Migrating the legacy Web Stats to Hadoop will open doors for Advanced Analytics and allow Fidelity Business users to use multiple open source tools to analyze the web data and derive Business Value by finding Customer behavioral patterns and identifying key web signals.

Environment: Apache Hadoop 2.4, HDFS, Hive, Java, My SQL, Cloudera, Git , Maven, Jenkins, UDeploy and Control M.

**Responsibilities:**

* Responsible for building scalable distributed data solutions using Hadoop with Core Java programming language.
* Handled importing of data from various data sources, performed transformations using Hive, MapReduce, loaded data into HDFS and extracted the data from RDBMS into HDFS using Sqoop.
* Integrated HIVE with existing applications, involved in preparing data mappings and exploring data to clients with Hue.
* Analyzed the data for business use cases by performing Hive script queries.
* Created custom UDFs using Hive.
* Worked on Oozie workflow engine to run multiple Hive scripts.
* Worked with Data Engineering, Analysts and Business teams during application life cycle.

***Java/J2EE Projects:***

Fidelity Investments, NC June 2011 to May 2014

**Product Line: Plan Sponsor Webstation**

**Role: Senior Software Developer**

Plan Sponsor Webstation (PSW®) is a robust, easy-to-use, web based solution that provides clients with a window into their benefits programs giving them control and access to their information and helping them manage their employee benefits intelligently and efficiently. Products supported: Defined Contributions (DC), Defined Benefits (DB), Stock Plan Services (SPS), Health Savings Account (HSA) and Health & Welfare (HW).

Environment: JDK 1.7, JSP, Struts, Spring, SOAP/Restful Web Services, Angular JS, Web Sphere, Tomcat, DOJO, Git/Jenkins.

Third Party Tools: DCLe, SDL Tridion, EMC Documentum, OBIEE, Splunk.

**Responsibilities:**

* Analysis of requirements for new projects for all products. This includes working with Analysts and doing Impact Analysis.
* Involved in technical document preparation that includes preparing Data Flow/Class Diagrams using Microsoft Visio.
* Development of code as per the business requirements using relevant design patterns.
* Developed web pages using DOJO components for alert, confirm, popup and detailed tabular data with sorting and pagination.
* Involved in changing the Electra framework to integrate new changes for next generation applications using latest technologies like Spring AOP for Authorization and logging.
* Involved in generating soap web services client code, preparing code for unit testing and integrating the web service calls in web application code. For any issues testing web service using SOA test and SOAP UI
* Performing Unit testing and system integration setup for enterprise releases.
* Involved in Code review, coordinating with Secure Code Review team and providing fixed for SCR items.

Company: Mphasis, Bangalore India Oct 2010 to Jun 2011

**Project: Global Docs (GDocs)**

**Role: Senior Software Engineer**

Global Docs enables legal documentation from all regions to be accessible from one location and provides users with the ability to both download individual and legal documents and have the system combine multiple documents into one package, which can be tracked throughout the negotiations. Global docs system was updated to ZIP all package files (RTF format) generated by the system for saving disk space on application server and reducing time needed to download large packages.

Environment: Java, Spring,Struts, IBatis, My SQL, JBOSS, ANT .

Company: Zensar Technologies, Pune India June 2008 to Oct 2010

**Project: Banner Ad Manager (BAM)**

**Role: Software Engineer**

Banner Ad Manager is used to request your Ad to be placed on any of the pages on Cisco.com website (includes US, UK, India, Germany, France, China, Canada France, Japan and Canada country pages). This is a versatile tool that will take you through pages in every level of Cisco.com web-site thus allowing you to verify relevancy of content on these pages, at the same time allowing you to mark a page as the target location for your Ad.

Environment: JDK 1.5, JSP, Struts, Hibernate, Oracle, IBM ClearCase and IBM Web Sphere.

Third Party Tools: Google’s DoubleClick Ad Manager

**Responsibilities:**

* Responsible and active in the analysis, design, implementation and deployment of full Software Development Lifecycle (SDLC) of the project.
* Designed and developed user interface using JSP, HTML and JavaScript/jQuery.
* Developed Struts action classes, action forms and performed action mapping using Struts framework and performed data validation in form beans and action classes.
* Extensively used Struts framework as the controller to handle subsequent client requests and invoke the model based upon user requests.
* Defined the search criteria and pulled out the record of the Banner Ads Reports from Google’s DoubleClick Ad Manager.
* Used DAO and JDBC for database access, developed stored procedures and triggers using PL/SQL in order to calculate and update the tables to implement business logic.

education

* Masters in Computer Applications from ICFAI University in 2007
* Bachelor of Science in Computer Applications from Sri Venkateswara University in 2005.