****

 **Aravind Kumar**

 **Big Data Technical Architect**

 **EMAIL ID:** aravindkumar1980@gmail.com

 **Phone No :(** **732-447-4415**

**PROFESSIONAL SUMMARY**

* **Over 18 years of professional IT experience with emphasis on big data Technologies, working with many large-scale applications in various domains including Finance, Banking, Insurance and Health Care**
* **Cloudera Certified Hadoop Developer for Apache Hadoop (CCDH 410 – Version: 5)**
* Experience in complete Software Development Life Cycle process of application development. (Requirements gathering, analysis, design, development, testing and implementation).
* Expertise in Hadoop Distributed File System **(HDFS), Map Reduce, PIG, HIVE, HBASE, SQOOP**.
* Extensive experience **working on working in Big Data Hadoop Ecosystem comprising Apache Spark, PySpark API, Docker, Map Reduce, Hive, Pig, Apache Oozie, Sqoop, Flume, HDFS, Apache Avro**.
* Expertise in working on **AWS using Lambda, EMR, Redshift, SNS, SES, Glue, Data Pipeline, S3, API Gateway, Athena API, Amazon Kinesis and DynamoDB No SQL DB**.
* Excellent understanding of **Hadoop architecture and Hadoop ecosystem** such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node and MapReduce programming paradigm.
* Extensive experience in importing/exporting data from/to RDBMS the Hadoop Ecosystem using Apache Sqoop.
* Good Experience in analyzing data using **HiveQL, Pig Latin, and custom Map Reduce** programs in Java.
* Extensive experience in creating **Complete Workflow chain from scratch** for multiple projects within client domain using **Apache Oozie**. Workflow Scheduling involves Map Reduce Jobs, Hive, PySpark and Shell Script, Email actions with output of one workflow fed as input to another.
* Good experience in Cloudera platform and Cloudera Manager.
* Migration of the revenue data from Oracle to Hadoop, Hive, and **Amazon Redshift**
* Very strong industry experience in Apache Hive for data transformation.
* Strong experience in both Development and Maintenance/Support projects.
* Good team player with excellent communication skills to work in a team and individual environment.
* Strong exposure to IT consulting, software project management, team leadership, design, development, implementation, maintenance/support, and Integration of Enterprise Software Applications.
* Extensive experience in conducting feasibility studies, Plan reviews, Implementation, and Post Implementation Surveys
* Demonstrated ability to work independently, showing a high degree of self-motivation and initiative.
* Excellent team member with problem-solving and trouble-shooting capabilities, quick learner, result oriented and an enthusiastic team player.
* Extensive Experience in designing and developing in Spark using Python.
* Excellent in Analytical /problem solving skills.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| ***Big Data Technologies*** | Apache Spark 2.3, Python API for Spark Hadoop, HDFS, Map Reduce, Hive, Pig, Sqoop, Flume, Zookeeper, Oozie, Impala, Apache Avro |
| ***Programming languages*** | Python, Java, Cobol |
| ***Databases*** | Oracle, DB2,HBase, MySQL, Redshift |
| ***AWS*** | Lambda, EMR, Redshift, CFT, ECS, SNS, SES, Glue, Data Pipeline, S3, API Gateway, Athena API, Amazon Kinesis and DynamoDB No SQL DB.  |
| ***NoSQL Database*** | DynamoDB |
| ***Operating Systems*** | Windows & Linux, UNIX |
| ***Scheduling Tools*** | Control-M and Oozie |
| ***Other* Tools/Utilities** | TSO/ISPF, QMF, SPUFI, SDF II and Change man, CVS, SVN, GIT |
| ***Defect Tracking Tools*** | HP Quality Center |

**PROFESSIONAL EXPERIENCE:**

**Client: Panasonic Avionics Corporation** **Jan 2023 to till date**

**Big Data Architect**

**Project: Datahub**

**Environment: AWS Lambda, Amazon Redshift, Amazon EC2, Python-Spark, Glue, Lambda Functions, Code Commit, Code Build, Code Pipeline, SQL, S3 and Step-Functions.**

Connectivity and Usage tracks inflight Internet Availability for Each operator for its customers during in-Flight and Panasonic commitment on meeting SLAs with each Operator.

**Responsibilities**:

* Lead Onshore/Offshore Teams for their deliverables.
* Conduct Daily Stand up with the team and resolve if they have issues with their assignments so deliveries can be met.
* Conduct Backlog grooming and pointing meetings to refine Stories for EDW.
* Manage Leadership expectation on the deliverables.
* Participate in project startup meeting to determine the high-level estimate of work which will be coming in for the project.
* Performance Tuning of Existing slow running Glue jobs which helps in reducing runtime which translated to AWS resources Billing savings to the Client.
* Gather information from business partners about program functionality and capabilities.
* Design and Develop data pipelines to extract, load, and transform data using SQL and Python.
* Work with Data Lake Team on Source Data ingestion.
* Investigate, recommend, and implement data ingestion and ETL performance improvements.
* Document data ingestion and ETL program designs, present findings, conduct peer code reviews.
* Develop and execute test plans to validate code.
* Create DDL scripts to add Create Table or add columns to existing Tables in EDW as and when source data structure changes.
* Monitor all production issues and inquiries and provide efficient resolution.
* Create Mapping Document with source data structure and Target Tables and evaluate the technical debts to meet the requirements to be implemented.

**Client: Intuit Inc** **Dec 2019 to Jan 2023**

**Big Data Architect**

**Project: DAC**

**Environment: PySpark, Redshift, EMR, EC2, S3, CFT, Athena API, Hive, Tidal.**

DAC teams are currently in the process of migrating data to AWS.  This is part of a larger effort across Intuit to have better, easier to access, cleaner data for analytics and application purposes.  Ideally, the consolidation data on AWS will allow for a variety of use cases and it will make it easier for anyone trying to work with the data in the future.

**Responsibilities**:

* Designed and developed end to end applications for Data ingestion, Organized Data layers and business use cases.
* Designed and developed Framework to migrate data from RDS/Hive/Redshift to Data Lake
* Designed and developed various templates (**Full Merge/Truncate load/Append** Only) using **PySpark** to securely transform datasets in S3 curated storage into consumption data views.
* Designed and developed the Test Suite to test the Framework.
* Developed python program to ingest glance data using glance API to **S3,** thenusing framework glance data is ingested into data lake.
* Created **System architecture/ Design** and Software development for DAC.
* Leveraged **Amazon Athena** for ad-hoc query analytics.
* Analyze the **Business Requirements** and come up with Design/ Architecture identifying the different components, flow diagrams and discuss with the team.
* Participate in the **end-to-end** life cycle of the project right from requirements, design, development, and testing.

**Client: Panasonic Avionics Corporation** **Apr 2018 to till Dec 2019**

**Big Data Architect**

**Project: Insights**

**Environment: AWS Lambda, Redshift, CFT, ECS, SNS, SES, Glue, Data Pipeline, PySpark, S3, API Gateway, Athena API, Amazon Kinesis and DynamoDB No SQL DB.**

Insights is a SaaS Application + Analytics Consulting service that is offered to our customers.  To provide deep industry standard data consulting services and value-add services, that complement the Insights Platform, directly impacting the improvement of its client’s passenger experience, help airlines maximize their IFE investment and provide airlines an in-depth view and better understanding of product success and risk factors.

**Responsibilities**:

* Designed and developed end to end applications for Data ingestion, Organized Data layers and business use cases.
* Developed **DynamoDB** components to store the Insights Data
* Developed AWS **Glue ETL** jobs using PySpark to securely transform datasets in S3 curated storage into consumption data views.
* Worked on Continuous Integration, Continuous Deployment, Build Automation and Test-Driven Development to enable the rapid delivery of end user capabilities using Amazon Web Services (AWS) Stack (**Code Commit, Code Deploy, Code pipeline, Code Build, IAM, CFT**)
* Designed and developed the insights applications using **AWS using Lambda, SNS, Glue API, S3, API Gateway, Athena API**.
* Developed python Job to process glance data using glance API & aggregates data in Parquet and push the output to **S3** & **DynamoDB**.
* Created **System architecture/ Design** and Software development for Framework.
* Worked on AWS **Cloud Formation** to provision AWS resources(**S3, SNS, RDS, EMR, Glue, Lambda, DynamoDB**)
* Developed **Spark Code** to implement **Data Quality Checks** – to check the processed data across the system, flight count etc.
* Leveraged **Amazon Athena** for ad-hoc query analytics.
* Analyze the **Business Requirements** and come up with Design/ Architecture identifying the different components, flow diagrams and discuss with the team.
* Participate in the **end-to-end** life cycle of the project right from requirements, design, development, and testing.

**Client: Autodesk Sep 2016 to Apr 2018**

**Big Data Technical Lead**

**Project: BIC Finance Reporting**

**Environment: AWS EMR, AWS S3, AWS Cloud Watch, RDS(MYSQL), HDFS, Hive, Redshift, Sqoop, Oozie Workflows, Shell Scripts, Spark**

Born In the Cloud (BiC) is Autodesk's cloud licensing and entitlements platform that enables flexible and scalable business model offerings for customers. It provides a trial-to-purchase user experience that enables direct in-product purchases and instant access to the purchases.

BIC Finance Reporting project deals with ingestion of data from two data sources Pelican, Respro and MDS systems into AWS system, processing the data in AWS using Hive scripts and pushing the data into Redshift. The processed data will be available in Redshift in the form of views.

**Responsibilities**:

* Analyzing the requirements and the existing environment to help come up with the right strategy to build the BIC system.
* Developed **Spark** and **Hive** scripts for Data processing.
* Developed **Oozie** Workflow and Coordinator for integrating other systems like Denodo, Hadoop ETL (Hive, Sqoop), Redshift & CloudWatch
* Enabled the **Oozie** SLA feature to alert the long running job.
* Built and Owned Data ingestion process from different sources to Hadoop cluster
* Developed programs Python Spark job to process raw data in **Parquet** and push the output to S3.
* Worked on ETL scripts to pull the data from denodo Data Base into HDFS.
* Developed **hive** tables to upload data from different sources.
* Involved in Database Schema design.
* Developed script to load the data into **Redshift** from Hive tables.
* Created different views in **Redshift** for different applications.
* Stored the job status in MYSQL RDS
* Proposed an automated system using Shell script to **Sqoop** the job.
* Worked in **Agile** development approach.
* Created the estimates and defined the sprint stages.
* Mainly worked on **Hive** queries to categorize data of different claims.
* Created cloud watch to monitor the application.
* Monitored System health and logs and responded accordingly to any warning or failure conditions.
* Involved in the design of Distribution styles for redshift tables.

**Client: Caterpillar Inc Jan 2015 to Sep 2016**

**Big Data Technical Lead**

**Project: DDSW**

**Environment: CDH 5 , HDFS, Hive, Impala, Sqoop, Tableau, Oozie Workflows, Shell Scripts, IntelliJ, Gradle, Python, AWS**

Dealer Data Staging Warehouse (DDSW)

Caterpillar’s business model originates from a guide, issued in the 1920s, that established territory relationships with a number of Dealer affiliates. These largely autonomous relationships allowed the Dealers to develop their own models for tracking important data, such as customers and inventory, that relate to local market conditions, including government regulation and customary business practices.

This model has also led to conditions that disrupt Caterpillar’s markets, such as uniform pricing for replacement parts and efficient logistics planning for warehousing inventory. Caterpillar also sees value in analyzing Dealer data for its own uses, such as gaining a better understanding of customer applications of its machinery, predictive failure analysis, supply chain optimization, and customer purchasing patterns.

The Dealer Data Staging Warehouse (DDSW) platform stages the data received from Caterpillar’s Dealers and prepares them for consumption for a wide variety of uses, such as customer portal services, analytics for equipment monitoring, parts pricing, and customer lead generation, and other emerging applications.

**Responsibilities**:

* Analyzing the requirements and the existing environment to help come up with the right strategy to build the DDSW system.
* Designed and Executed **Oozie** workflows using **Hive**, **Python** and Shell actions to extract, transform and Load data into Hive Tables.
* Worked extensively with **Avro** and **Parquet** file formats.
* Involved Low level design for **MR**, **Hive**, **Impala**, Shell scripts to process data.
* Worked on ETL scripts to pull the data from Oracle Data Base into **HDFS**.
* Developed hive tables to upload data from different sources.
* Involved in Database Schema design.
* Involved Sprint Planning and Sprint Retrospective meetings.
* Daily Scrum Status meeting.
* Proposed an automated system using Shell script to **Sqoop** the job.
* Worked in Agile development approach.
* Created the estimates and defined the sprint stages.
* Developed a strategy for Full load and incremental load using **Sqoop**.
* Mainly worked on Hive/Impala queries to categorize data of different claims.
* Implemented Partitioning, Dynamic **Partitions**, **Buckets** in **HIVE**.
* Generate final reporting data using Tableau for testing by connecting it to the corresponding Hive table’s using Hive ODBC connector.
* Written **python** scripts to generate alerts.
* Monitored System health and logs and responded accordingly to any warning or failure conditions.
* Implemented POC on AWS
* Worked on Kerberos Authentication for Hadoop.

**Client: Citi Group** **July 2010 to Jan 2015**

**Sr. Software Technical**

**Project: Chemistry – Secore**

**Environment:**

Citi SECORE Custody is Citibank’s Core Safekeeping and Asset Servicing Company in North America, EMEA, CEEMEA & ASPAC managing both Domestic and Global Custody operations. The SECORE application is handling global custody for the customers for all the trade markets including Trade creation, corporate actions, and settlement, AFX / FX and class actions.

**Responsibilities**:

* Resources planning to work across the project team to set the appropriate schedule and

owning the accountability of the software application delivery.

* Strategic planning with Client Managers.
* Involved in (design, development, maintainability, quality) and innovation in setting project.

direction, generating DDL, SQL static Modules, Copy Books and BIND control cards from FSA.

files structures (CITI specific). Meta Data Analysis as part of VSAM to DB2 migrations.

* Developed backup/recovery procedures for application DB2 tables.
* Lead the team in migrating VSAM to DB2 .
* Extracting business rules from legacy COBOL-CICS-DB2 programs and preparing functional.

specifications in JAVA.

* Worked with BI teams in generating reports and designing ETL workflows for BI teams.
* Prepared, implemented plan, tracked defects in SIT & UAT and provided the required.

implementation support.

* Actively participated in software development lifecycle (scope, design, implement, deploy, test), including design and code reviews, test development, test automation.
* Documented the systems processes and procedures for future references.

**Client: Fidelity Business Services** **Feb 2007 to Jul 2010**

**Sr. Software Engineer**

**Project: ICS (Integrated Customer System).**

**Environment: Java, J2EE (JSPs & Servlets), JUnit, HTML, CSS, JavaScript, Apache Tomcat, Oracle**

Integrated Customer System is FBC’s (Fidelity Brokerage Company) “System of Record” for customer and non-monetary account information. It houses over 20 million customer accounts. It is a single source for this information to all FBC subsystems and user front end and houses business rules governing the entry and maintenance of this data. It has been designed to provide 24 X 7 availability while achieving the highest standards of quality and the greatest processing efficiency possible. The major functions that are supported by ICS are New Account setup Customer and account Maintenance Features & options for accounts (e.g. Checking, Debit card, Credit card additional names etc.) Customer reporting Business parameters Also supports interface to multiple front ends.

**Responsibilities:**

* Involved in requirements analysis and prepared Requirements Specifications document.
* Designed implementation logic for core functionalities.
* Developed service layer logic for core modules using JSPs and Servlets and involved in integration with presentation layer.
* Involved in implementation of presentation layer logic using HTML, CSS, JavaScript, and XHTML
* Design of Oracle database to store customer's & account’s details
* Used JDBC connections to store and retrieve data from the database.
* Development of complex SQL queries and stored procedures to process and store the data.
* Developed test cases using JUnit.
* Involved in unit testing and bug fixing.
* Used CVS version control to maintain the Source Code.
* Prepared design documents for code developed and defect tracker maintenance.

**Client: SwissRe Sep 2005 to Jan 2007**

**Sr. Programmer Analyst**

**Project: COJAK**

**Environment : Core Java, Java Batch, Service Beans, EJB, RMI/IIOP, J2EE, COBOL390, CICS, DB2**

SwissRe is one of the largest reinsurance companies. The project “COJAK” was to convert Cobol based services into Java based batch processing. This will eventually replace all Cobol programs (backend processing, logical request processing and batch processing) with Java. In addition, client server based front end applications were developed using Java that can integrate with both Java batch and Cobol processes, all running on mainframes using z/OS Host.

**Responsibilities:**

* Responsible for Proof of Concept, Planning, Designing new proposed Architecture.
* Worked on Java, Swing, Web services, XML in addition to Mainframe Technology.
* Extracted the business rules from Legacy COBOL programs to code in Java.
* Used latest methodologies to convert the existing Mainframe programs to Java &Java batch.
* Able to migrate with the limited resources available in Mainframe.
* Fine tuning of application programs with the help of DBA.
* Utilized transaction wrapper technology (EJB, Batch, Service Bean on WebSphere cluster).
* Attended functional meetings and prepared the high-level Detail Design Document.
* Designed high- and low-level Design documents for the new functions to be implemented.
* Supported the re-structuring of DB2 tables by re-writing the existing programs.
* Debugging and troubleshooting any technical issues while implementing the applications
* Implemented Java client based OLTP process with WebSphere server running on Mainframe z/OS Host.

**Client: HCA – Hospital Corporation of America Apr 2004 to Sept 2005**

**Programmer Analyst**

**Project: HCA – Patient Accounting**

**Environment :COBOL390, MVS-JCL, CICS, DB2, VSAM, TSO/ISPF, QMF, SPUFI, SDF II and Change man**

HCA is one of the premier healthcare service organizations in the world and operates approximately 200 hospitals and over 80 surgery centers in the U.S., England, and Switzerland. HCA has outsourced four modules (also referred to as towers of HCA) to Syntel viz. Patient Accounting, Financial Reporting, HR / Payroll and SMART.

Patient Accounting (PA) could be considered as the entry point for the HCA healthcare systems. It deals with the process’s admission details, personal details, insurance details, etc. of the patient and continues through to his treatment details and ends finally with his billing and payment details.

**Responsibilities:**

* Procuring the project requirements from business analyst & users, breaking up the project delivery into phases and meeting the deadlines as per the estimates.
* Transforming the Business requirements into design.
* Preparation Analysis, estimation, and design.
* Single point of contact between customer and offshore team members.
* Prepared high-level and low-level design based on business requirement document.
* Preparation of Technical Specifications by using high-level design and business requirement document.
* Providing Module Inventory and Estimates by identifying the impacted components.
* Business and Technical knowledge sharing with other Team members.
* Coded complex programs, report program (batch & Online) in COBOL/VSAM/DB2/CICS
* Preparation of analysis documents, modification of Programs / JCLs and peer review
* Preparing the Unit Test Case document, Coding and Unit Test Results document
* Developing maps, online and batch programs and perform Review of Test cases and code.
* Solving defects at SIT/UAT phases and giving the Implementation support.
* After implementation, preparing Defect log and Defect Action Plan documents.
* Mentoring and motivating team members in enabling the team to work independently on Tasks.

**EDUCATION**

Master of Computer Applications

**CERTIFICATION**

**Cloudera Certified Hadoop Developer** for Apache Hadoop

**AWS Certified Developer** – Associate

**AWS Certified Solutions Architect** - Professional