**UpendraRao.B**

**Bigdata Engineer**

# Profile

Software professional with **15+ years** of experience in IT industry. Strong Technical Lead experience in handling various development, maintenance & enhancements, migration and production support projects.

In-Depth knowledge and experience in design, development of **Big Data applications in AWS, Cloudera, Hortonworks platform** that perform large scale Distributed Data Processing/Analytics using Python, Spark, **Scala ,Kafka, Hadoop, Hive, Oozie, NoSQL’s** (HBase and Cassandra) and Big Data ecosystem tools.

Extensively worked on data ingestion projects, building data lake, ingesting data from various sources like RDBMS data bases **(DB2, MySQL server, oracle and PostgreSQL)** andfiles from mainframes/SAS and third party systems (Binary, ASCII delimited and fixed).

Hands on experience working on automation of ingestion frameworks which are handling different kinds of data formats from various sources using **Python, Spark, Scala, Unix/Shell scripting**. Experience in handling various file formats like **Sequence files, RC File, parquet and AVRO**.

**AWS Certified Solution Architect** - Experience in Cloud services using AWS including **EC2, Auto-Scaling** in launching EC2 instances, Elastic Load Balancer, Elastic Beanstalk, **S3, Glacier, Dynamo DB, Cloud Front, RDS**, VPC, Route53, Cloud Watch, Cloud Formation, IAM, SNS/SQS.

Experience in NoSQL and RDBMS databases and very good knowledge of various kinds of NoSQL DB used in Big Data space **(HBase, AWS Dynamo DB and Cassandra).**

Detailed technical knowledge and hands-on experience of DevOps, Automation, Build Engineering and Configuration Management. Designed and implemented **Continuous Integration (CI)** and **Continuous Deployment (CD)** workflows with **Git, Maven, Jenkins, Rundeck and Ansible.**

Having sound Technical and Functional knowledge in key functional areas such as **Banking, Financial domain and Insurance domains.**

Detailed technical knowledge and hands-on experience of various Configuration Management and scheduling tools. Implemented various projects using agile methodology.

Extensively Involved in Business requirement gathering, Function and Technical specification creation, Impact assessment and provided Estimations.

Designed and implemented Automation of tasks to improve cluster efficiencies **(Housekeeping/Compaction)** and to utilize **cluster/control-M resources by jobs efficiently.**

Involved in providing production support on various analytics and operational clusters, continuously monitoring and reporting various types of ingestion, transformation jobs. Handling on-call off hours support, analyzing and fixing job failures within SLA.

Experience in handling and supporting various platform activities like cluster upgrades, Gateways migration and Control-M jobs migration.

# Technical Skills

|  |  |
| --- | --- |
| Languages/Toolkit | UNIX scripting, JCL,COBOL, Scala and Python |
| Big Data Technologies | Spark, Hadoop 2.6, Hive, Impala, HBase, HDFS, YARN, Sqoop, Flume, Kafka and Oozie |
| DevOps(CI/CD) | Git, Maven, Jenkins, Rundeck and Ansible. |
| Databases(RDBMS/No SQL) | DB2, Oracle, MySQL, PostgreSQL, HBASE and CASSANDRA |
| Tools/Methodologies | STP, Control-M, Service Now, ALM, Agile, Scrum, JIRA and Confluence. |
| Big Data Stack | Cloudera and Horton Works |
| System Environments | MVS OS/390,Unix/Linux/Windows |
| Cloud | AWS including EC2, Auto-Scaling in launching EC2 instances, Elastic Load Balancer, Elastic Beanstalk, S3, Glacier, Dynamo DB, Cloud Front, RDS, VPC, Route53, Cloud Watch, Cloud Formation, IAM, SNS/SQS and Terraform |

# Professional Education & Certifications

* AWS Certified Solutions Architect (<http://aws.amazon.com/verification)> - **991X7RDLCBV4QX5M**
* Masters of Computer Applications (MCA) (2001-2004)
* B.Sc. (Bachelor of Science) Computers (1998-2001)

# Work Experience:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Company / Organization | Location | Role / Designations | From Date | To Date | No of Years |
| ITC Infotech India Ltd | Milton Keynes, UK | Lead consultant | Oct 2016 | Till date | 4.8 yrs. |
| FIS Global Business Solutions | ChennaiWellington, NZ | Project Lead - Development | Sep 2012 | Sept 2016 | 4 yrs. |
| Polaris Software Lab Limited | Chennai | Senior Project Lead | Mar 2010 | Sept 2012 | 2.7 yrs. |
| Mphasis an EDS Company | Chennai | Senior Software Engineer | Apr 2007 | Feb 2010 | 3 yrs. |
| VIT Consultancy Pvt. Ltd | Chennai | Software Engineer | Apr 2006 | Mar 2007 | 1 yrs. |

**Bigdata/AWS Cloud Engineer – Santander UK (ITC Infotech India Ltd) Dec 2016 – Current**

**Analytics Data Lake:** A single Data lake for the whole of Santander UK group (Retail, Corporate, Cater Allen, SGCB, Isle of Man, Jersey, SCF), with data segregation and high data security. A data lake is a cost-eﬀective, highly-scalable architecture for collecting and processing virtually any data format from any source. It enables new business insights and data use cases that were previously unachievable.

* Designed and developed ingestion frameworks (Batch/NRT) to ingest data into Data Lake.
* New requirements review and provide Big Data estimates to various projects
* Designed and developed the data processing pipeline to fetch data from various source systems and the data is transformed and ingested in to HDFS/Hive/HBase at various levels.
* Implemented CI/CD pipeline using Git, Jenkins, Ansible and Maven
* Automation of tasks to increase cluster efficiencies by implementing compaction techniques, Data classification analysis and by performing regular housekeeping tasks.
* Implemented NFS mount solution to achieve HA/DR by making data and codes available across clusters DC1&DC2.
* Involved in generating regular and adhoc reporting on lake data for business. Transformed data to use by MI tools.
* Designed and developed complex Jrecord (binary) fixed file ingestion framework to overcome limitations of existing frameworks. Legacy frame works are not supporting ingestion of COBOL generated extracts using copybook.
* Extending HIVE functionality by using custom UDF’s. Hands on writing the Hive scripts to reduce the job execution time.
* Imported/Exported data from various databases to HDFS through SQOOP. Tested raw data and executed performance scripts.
* MIPS reduction initiatives on DB2 ingestions, various third party product evolutions and provide suggestion to Client (Alibaba cloud/AWS EMR/Control-M workbench etc.).
* Planning and Support for Cluster Upgrades/Gateway version upgrades/Control-M jobs migrations.

**FCTP AWS (EMR Migration):**

* Setup Jenkins pipeline to bake AMI’s and created image out of it to use it in Terraform templates.
* Developed Terraform templates and created stack’s
* Configured and managed VPC, Subnet, DMZ, Network ACL, Security groups, Auto Scaling (Launch Config/User Data), ELB, EBS, Route 53 DNS, S3, Dynamo DB.
* Integrated GitHub, AWS Code Pipeline, Jenkins and AWS EMR to create a software development pipeline.
* Launching AWS EMR cluster using Terraform, testing of Bigdata applications on AWS EMR cluster.

**Development Technologies and Toolsets:** Hive, Sqoop, Kafka, Spark SQL, Spark, Scala, Unix Scripting, Kabana dashboards and Control-M.

**Big data Engineer- FIS Global Business Solutions Sept 2012 to Sept 2016**

**Technical Log Lake (ANZ Bank):** The scope of this Use case involved implementing Hadoop clusters for monitoring the access (fraudulent access) and Web server log (to analyze the customer preferences on banking products). The access logs and click through were loaded using Kafka to be processed by spark jobs and to generate the statistics reports. This helped in considerable reduction in the duration taken on analytical report generation.

* Designed and developed the data processing pipeline to fetch data from various source systems and the data is transformed and ingested in to HDFS/Hive/impala at various levels.
* Involved in Data process design, design, implementation and iteration of new processes around the big data platform.
* Experience in real time streaming data using Kafka producer, consumer, spark streaming and flume
* Experience in using Cloudera, Horton works distributions
* Experience in Jira and Control-M scheduling tool.
* Involved in generating regular and adhoc reporting on lake data for business. Transformed data to use by Kibana dashboards

**DIH (Data Integration Hub) - Professional Services (DIH), Littlerock USA:**

This project is about to build a Centralized Hub for bank's data in Big data using Hadoop. Currently it has multiple individual systems which serves the required outgoing feeds for other systems. The current system is lacking to handle and process huge volume of data and there is a bottleneck in keeping the system with the expected performance. Hence, bank wanted to build a hub for their enterprise data where it can serve efficiently for all the outgoing feeds for the required systems.

* Requirement gathering, Coding and testing.
* Developed Hive UDFs for business requirement.
* Worked on creating java classes using Hadoop API
* Worked on Sqoop ,Hive databases
* Involved in loading data from UNIX file system to HDFS.
* Monitoring Hadoop clusters and sending email using monitoring scripts.

**Development Technologies and Toolsets:** UNIX Scripting, Python, Spark, Kafka, HDFS, Elastic search, Kibana, Flume, Hive and Sqoop.

**BSP (Batch Simplification Project) - CITI bank:**

In current system various IM and ST systems extracts are generated by legacy systems are sent to downstream applications EDR and AML. Batch simplification project, purpose of this project is to replace the legacy mainframe systems with latest technologies for easy maintenance using IBM Data stage. Now the extracts from the ADS are extracted, transformed and sent to downstream systems AML and EDR for their business processing.

As a Technical Lead, involved in the following activities:

* Involved in Development, Code reviews and document reviews.
* Prepared High level design and low level design documents.
* Involved in Data modelling activities.
* Prepared Technical Estimation and Status Reports.

**Development Technologies and Toolsets:** Systematics, COBOL, JCL, CICS, VSAM, DB2, File-AID, XPEDITER, OPC, CA-VIEW and ENDEVOR.

**Technical Lead - (Polaris Software Lab Limited) – Mar 2010 to Sept 2012**

**EDC - Migration, Royal bank of Scotland, UK:** RBS Group is a universal bank with operations in retail, wholesale and investment banking, as well as wealth management, mortgage lending and credit cards.

The EDC Payments application interface with many other systems including Customer Information Application (CIA), International and New Authorization Process (NAP) in order to retrieve the information regarding the Customer accounts (both domestic and currency) and the individual available funds respectively. The International system will provide a daily extract of all Currency Accounts and this will be loaded into an EDC DB2 table to be used in the EDC online application. As a Technical Lead, involved in the following activities:

* Analysing the data requirements – BOM and BRD’s.
* Creation of Application Designs, Traceability Matrix, CD, UTP, UTC and UTR documents.
* Handled End to End activities for one Subject area.
* Prepared Functional Specification documents.
* Supported in Various Testing Phases.
* Involved in code and job reviews, Peer Reviews before delivery.
* Prepared all required documents and KT documents.
* Creating the CMR’s and implementation plans to implement the system and providing the support.

**Development Technologies and Toolsets:** COBOL, CICS Web services, JCL Jobs, SOAPUI, CA-View and DB2.

**Senior Software Engineer - Mphasis an EDS Company – Apr 2007 to Feb 2010**

CSAA – California State Automobiles Association, a part of AAA – American Automobiles Association is an Insurance company based at San Francisco, CA. CSAA was founded on sound principles, attracted good people, and offered important benefits which have grown to include Emergency Road Service, Insurance Services, Travel Services, Financial Services, Automotive Services, Education, and more. CSAA Account is mainly divided into three projects called Membership, Auto and Homeowner.

This is a Maintenance and support project and involves some enhancements made to the system. As a Primary Team Member, involved in the following activities:

* Collected the software requirements and specifications.
* Developing new modules, which enhance and improve the functionality of the system.
* Requirement Analysis, coding and Unit testing.
* Writing unit test cases and integration test cases.
* Investigate & Fix ABENDs and Remedy tickets
* Involved in sharing knowledge gathered with the team members
* Developed SQL statements to improve back-end communications.

**Development Technologies and Toolsets:** HUON, COBOL, JCL, CICS, VSAM, DB2, File-AID, XPEDITER, INTERTEST, OPC, CA-VIEW and ENDEVOR.

**Software Engineer - VIT Consultancy Pvt Ltd – Apr 2006 to Mar 2007**

**Citibank (NAS)** (New Applications System) is a robust credit decision system now functional in over 30 countries. NAS supports both the Cards as well as the Banking products like loans and overdrafts.

User-defined parameters allow the approval or decline of an application, with necessary alerts to exceptional or grey areas. These parameters can be set by product, credit scores, demographics, channel, or a host of other criteria.

As a Team Member, involved in the following activities

* Developing new modules, which enhance and improve the functionality of the system.
* Requirement Analysis.
* Coding and Unit Testing
* Involved in bug fixing of the existing Screens.
* Writing unit test cases and integration test cases.
* Preparing project related documents as requested by Team lead.

**Development Technologies and Toolsets:** Card Pac, Vision Plus and DB2