**Ankoji Uppala**

PROFESSIONAL SUMMARY

* Overall IT experience of around **12+** yearsout of which **7** years as **Big Data Engineer** experience and **5** years of Siebel development experience.
* Cloudera Certified Hadoop Developer and AWS Certified Solutions Architect Associate.
* Expertise in the various Hadoop Distributions such as Hortonworks, Cloudera.
* Professional experience in the design, development, support and maintenance of distributed systems
* Expertise in developing Spark SQL code with Data frame transformations to implement complex business logic.
* Expertise in implementing Ad-hoc queries using Hive QL. Implemented Data Quality checks using Hive queries
* Experience in migrating existing bigdata production applications from on-prem Hortonworks environment to AWS cloud by using AWS S3, EC2, EMR, Athena, lambda, glue services and shell scripting.
* Experience in enhancing the Pyspark based EMR framework to automate the data pipelines.
* Expertise in Spark core, Spark SQLandScala & PySpark, Azure Databricks**.**
* Expertise in Amazon S3, EMR and Athena.
* Implemented connectors for Salesforce and SAP using python code.
* Expertise in Shell scripting in implementing orchestration scripts and deployment scripts.
* Experience in using Apache Sqoop to import and export data to and from HDFS and external RDBMS databases.
* Expertise in production support model and implemented monitoring scripts and prepared SOPs.
* Competent in executing all phases of software life cycle in on-site and offshore based projects. Worked on agile methodology as well.
* Providing architectural and technical guidance to product development team and ensuring that sound software engineering design process and quality practices are followed.
* Domains exposed – Health Care, Digital Media, Telecom, Automotive.
* Expertise in Hadoop Eco-System elements using HDFS, MapReduce, Hive, Impala, Sqoop & Yarn
* Skill identification and training the existing team to be prepared for future releases in the project.

KEY WORK EXPERIENCE

**Project: Rapid Response HST Implementation**

**Client: Baxter, USA**

**Description**:

The high-level scope of this project is to provide a set of technical services that deliver structures, master, demand and supply data to support planning capabilities for Baxter HST.

Source Data comes from SAP, JDE and PIM systems. There is a PySpark framework (BIS) runs from AWS EMR cluster which extracts data from the source systems, does the ETL for the incoming data and generates tab delimited output files. These output files will be FTPed to Rapid Response FTP server.

**Project Duration**: 1 year

**Technologies Stack**: AWS S3, EMR, RDS, PySpark, Snowflake,DataStage, Postgres, Oracle, shell scripting.

**Role**: Bigdata Architect/TechLead

**Responsibilities**:

* + Worked on the enhancement of existing BIS pyspark framework and added new functionality to fetch data from JDE (Oracle) sources which will be a value added for client to reduce the usage of IBM DataStage.
  + Added connectors to BIS framework to fetch data from Salesforce and SAP using Python code.
  + Enhanced BIS framework with CDC functionality.
  + Enhanced the security of BIS framework by enabling SSL certificates for RDS Postgres authentication. Worked with AWS team to setup IAM DB Authentication setup.
  + Did a POC on AWS Secrets manager to store and retrieve sensitive informations like passwords and implemented these changes in BIS framework.
  + Worked on EMR cluster setup and configuring various security settings like encryption, auditing etc. Enhanced the existing bootstrap scripts
  + Worked on setting up AWS S3 buckets and folder structures required for the project. Worked on setting up Postgres databases, tables, functions, triggers, sequences etc
  + Worked on changes for BIS SAST scan report non-compliance comments.
  + Worked on the data pipeline developments using BIS framework and implemented several pipelines. Worked on performance tuning of the pipelines.
  + Performed code reviews and made sure that team followed coding standards. Also provided technical guidence to the team.
  + Worked on the CICD scripts to promote code to next higher environments
  + Implemented a shell script to download output files from AWS S3 and transfer the files to Rapid Response FTP server.
  + Worked on Athena database and table setup
  + Created detailed documents for environment setup, RDS IAM DB setup, handling julian dates, effective usage of EMR cluster etc
  + Trained freshers on BIS framework and made them as developers in the project
  + Worked on the creating techinal mappings for the given Functional specs
  + Worked on Data validation and bug fixes

**Project: DataStage to Azure Databricks migration**

**Client: OSHKOSH, USA**

**Description**:

Oshkosh Corporation, formerly Oshkosh Truck, is an American industrial company that designs and builds specialty trucks, military vehicles, truck bodies, airport fire apparatus, and access equipment.

One of the main objectives of the Netezza Migration Project is to migrate the existing ETLs running in IBM DataStage system to Azure Databricks platform.

**Project Duration**: 5 months

**Technologies Stack**: Azure Databricks, Delta Lake, PySpark, IBM DataStage.

**Role**: Bigdata TechLead/Data Engineer

**Responsibilities**:

* + Converted the Datastage jobs into Pyspark ETL pipelines
  + Validated the pyspark notebooks given by Dataswitch team and corrected the code as per the ETL logic written in DataStage
  + Handled data upserts and CDC using Azure Databricks and delta lake
  + Helped team technically whenever they stuck, performed code reviews
  + Performed data validations between DataStage and Databricks
  + Created orchestration notebooks for end to end execution of ETLs

**Project: Strategic BERT**

**Client: Credit Suisse Bank, USA**

**Description**:

Credit Suisse Bank is a leading financial services company in USA, advising clients in all aspects of finance, across the globe and around the clock.

This project is to re-platform various Basel calculators from a SQL based stack onto a big data platform consisting of Cloudera Hadoop, Spark, Core Java, Hive, Sqoop, Impala.

**Project Duration**: 1.5 years

**Technologies Stack**: Core Java, Apache Spark, Hive, Impala,Sqoop,Hadoop,Cloudera and Shell Scripting.

**Role**: Bigdata Architect/Senior Developer

**Responsibilities**:

* + Involved into detailed discussions and interaction with business stake holders /vendors in daily scrum calls.
  + Analysing existing SQL stored procedures & queries and development of equivalent Spark code.
  + Creation of logical and physical data models.
  + Implementation of data ingestion modules.
  + Implementation of data quality checks.
  + Implementation of data retention policy scripts.
  + Implementation of archestration scripts for scheduling the jobs.
  + Implementation of deployment scripts and validation scripts for deployment process.
  + Taking care of release management activities and preparing deployment documents, checklists, run documents.
  + Productionizing the basel3 calculators.
  + Taking care of production support activities for productionized applications.

**Project: CII (Customer Information Insights)**

**Client: Anthem**

**Description**:

Anthem is one of the world's leading medical insurance provider and one of the top player in Health Care Domain in US and provides health care services.

Anthem source data layer is in EDWARD ( Main Frame DB2 database) and monthly the raw data from EDWARD is loaded in to CDL (Common Data Layer) in Hadoop. From CDL layer, data will be loaded into ADL (Analytical Data Layer) in Hadoop using the Spark Framework by applying business transformations and the summary data is loaded to reporting layer which is Teradata using the Spark.

We have so many applicationslike Providers, Cost and Utilization, Cost and Value etc which process data for different subject areas. For all the applications, we are using Spark framework for data processing.

**Project Duration**: 6 months

**Technologies Stack**: Scala, Apache Spark, Hive, Teradata Bteqs and Shell Scripting.

**Role**: Tech Lead

**Responsibilities**:

* + Implementing Spark codes with data frame transformations as per the business uses cases.
  + Implemented a Spark framework which reads YAML file and runs the application specific DAO codes and loads tables mentioned in the application config file.
  + Worked on the Membership, Claims and Providers modules.
  + Preparation of Contrl-M setup for production jobs and implementing wrapper scripts to be triggered from Contrl-M jobs.
  + Taking care of release management activities like implementing automated build scripts, preparing deployment documents, checklists, run documents.
  + Ensure build phase of projects are delivered on time and with high quality.
  + Preparation of weekly status reports and metrics for each scrum deliverables.
  + Mentor the diversified team on Big Data Components and drive consistently.
  + Involved into detailed discussions and interaction with business stake holders /vendors in daily scrum calls. Reducing the vendor resources in the project.
  + Taking care of production support activities for productionized applications.
  + Preparation of automation scripts for data validations for all production loads.
  + Training Freshers in Big data and mentoring them.

**Project: Bigdata Application Support**

**Client: NBCUniversal**

**Description**:

NBCUniversal is one of the world's leading media and entertainment companies in the development, production, and marketing of entertainment, news, and information to a global audience.

NBCUniversal Big Data Architecture includes 30+ Vendor applications which feeds data on hourly, daily, weekly and Monthly into the Data Lake Service using the Data Management Tool Bedrock. The data feed has PreIngestion, Ingestion and Post Ingestion stages. During PreIngestion NBCU Big Data Systems performs data quality checks and takes care of data cleansing and upon post Ingestion data will be feed to down stream systems through Hive External Tables.

**Project Duration**: 2 years 8 months

**Technologies Stack**: Java, Scala, MapReduce, Apache Spark, Bedrock, AWS S3, EC2, EMR, Athena, lambda, glue, My SQL ,Sqoop, Hive Bedrock, HBase, Phoenix and Shell Scripting.

**Role**: Developer

**Responsibilities**:

* + Ensure build phase of projects are delivered on time and with high quality.
  + Involved into detailed discussions and interaction with business stake holders /vendors.
  + Work closely with multiple cross functional teams to effectively co-ordinate and manage business user expectations.
  + Ensure quality and time management processes are followed by team (e.g., metrics, RFC)
  + Help to document best practices in developing and deploying big data solutions.
  + Hands-on data analysis, trouble shooting of issues.
  + Design and Implement Tool upgrades (Bedrock 4.0.1 to Bedrock 4.2.1 etc ).
  + Develop/Configure Bedrock Workflows to meet the expected functionality.
  + Develop Spark-Scala Programs and Unit test the functionality. Converted Hive queries to Spark SQL for performance improvement.
  + Design, Develop and Implement AWS S3 backup across the applications.
  + Experience in migrating existing bigdata production applications from on-prem Hortonworks environment to AWS cloud by using AWS S3, EC2, EMR, Athena services and shell scripting.
  + Reviewing of code changes.
  + Implemented ITIL , which has enhanced Support model and changed deployment strategy.
  + Implemented upsert logic using HBase and Phoenix.
  + Implemented a generic script for column data type validations for hive tables for all production applications.
  + Implemented a generic script which pushes processed data from source cluster to destination cluster and to create hive partitions in hive tables in destination cluster.
  + Implemented generic script for Data Quality during the transfer of processed data from source cluster to destination cluster based on the file sizes.
  + Implemented code for retension policy to maintain data for specified no.of days for specific application in source cluster. Also active monitoring of cluster resources.
  + Implemented bedrock monitoring script which monitors bedrock services like WF executors, BDCAs, micro services etc and sends an alert email to support team when any of these bedrock services go down in production. This script is very useful for support team to handle any outages pro-actively.
  + Implemented a code change in production to improve the performance of files down load from a FTP server by changing SFTP to LFTP.
  + Prepared very detailed SOP document for Olympics which has manual steps to process the Olympics critical data with in SLA if workflow fails at any step. Also provided all DR options. Created individual workflows which can handle step level failures.
  + Worked on so many production deployments. Also provided support for production maintenances. Did impact analysis for each maintenance. Prepared detailed document with the jobs to hold during maintenance and post maintenance activities.
  + Implemented the generic script for S3 back-up of all applications data which can be called for all the NBCU applications. This is for DR purpose when HDFS goes down.
  + Converted existing hive queries to Spark SQL to improve the performance for critical applications (which plays a major role in Olympics).
  + Supported a new test cluster set-up and did regression testing and made sure that all the utilities are working fine in new cluster.
  + Set-up a disaster recovery cluster for Olympics and regression testing for Olympics critical applications.
  + Resolving of production issues for all production applications on time and preparation of detailed SOP documents for re-occuring issues which helps the other team members.
  + Conducting training sessions on Spark and scala to client team with lab to practice them.
  + Implementation of enhancements for existing 30 production applications using shell scripting, Hive, Spark, Scala, Bedrock etc.

**Project: PWM Quantum**

**Client: Goldman Sachs**

**Description**:

Quantum is a Java based framework built on top of Hadoop distributed file system and uses Hadoop echo components like Hive to process Goldman Sachs (GS) data. This framework is built for less duplication, one way to access data, better predictability and consistency of data load and data distribution, data lineage and entitlements.

**Project Duration**: 6 months

**Technologies Stack**: Hadoop 2.0, Java, Hive, Sqoop and Mongo DB

**Role**: Developer

**Responsibilities**:

* + Monitoring daily Quantum production loads.
  + Developed a java code to generate a producer manifest which is a JSON file and metadata to Quantum system.
  + Developed Hive UDF for converting Unix timestamp from milliseconds to seconds.
  + Developed a hive Generic UDF which accepts two parameters, date in long format and format string and returns the time stamp in given format.
  + Developed hive transformations for few data sets.
  + Developed a java program which will connect to hive, execute a quantum query generate an output file and compares it with AMR output file which is generated daily for each business date.
  + Imported from SyBase to HDFS using Sqoop
  + Created bucketing and partitioning as and when required
  + Implementing new requirements
  + Resubmitting the jobs if jobs fail
  + Cleaning old input data to avoid space issue in HDFS.

**Project: PMIX (Product Mix)**

**Client: Mc Donald**

**Description**:

From all the Mc Donald stores bills are generated in the form of XMLs. The main objective of this POC is to process the bills generated in all the Mc Donald stores across the globe and generate the daily and hourly sales using the Hadoop MapReduce framework.

**Project Duration**: 3 months

**Technologies Stack**: Hadoop 2.0, MapReduce

**Role**: Developer

**Responsibilities**:

* + Participated in daily call with client for gathering requirements
  + Prepared detailed design documents
  + Maintain files in HDFS
  + Developed MapReduce programs which reads XML files, parse them using Java DOM parser and generate daily sales.
  + Validated mapreduce output with the source data.

**Project: Vodafone Greece Pegasus Siebel 7.8 CRM**

**Client: Vodafone Greece**

**Description**:

Vodafone Greece Pegasus Siebel 7.8 CRM has been the replacement of the current billing system, Jupiter, the porting of the current functionalities of Siebel 6.3 to Siebel 7.8 and the enhancement of the integration architecture between the CRM, Provisioning and Billing systems.

**Project Duration**: 2.7 years

**Technologies Stack**: Siebel 7.8 Tools, Siebel 7.8 eCommunications, Windows 2003, Oracle 10g

**Role**: Developer

**Responsibilities**:

* + Worked on the bulk request functionality.
  + Worked on implementation of one change request on BIP for a New Connection order.
  + Worked on a requirement to call a stored procedure from Siebel using EAI ODBC Service
  + Worked on implementation of discount on MAF for barred MSISDNs
  + Worked on a requirement to restrict SR creation for given list of Divisions
  + Worked on provisioning of RED tariff line items.
  + Worked on some enhancements for the outbound web services Start voice provisioning, Send RFS, Restart failed provisioning and Get order status properties inbound web service.
  + Implemented two inbound web services for VOP customer and VOP Asset.
  + Worked on order management functionalities.
  + Implemented a solution to show bundle information from ALU in Siebel using VBC.
  + Implemented two new order types On Site Support and Change Offer.
  + Worked on order validations and submission.

**Project: Orion**

**Client: Navistar USA.**

**Description**:

Navistar is an International Truck and Engine Corporation, a manufacturer and marketer of medium and heavy trucks and mid-range diesel engines. They have separate divisions to handle Trucks, Engines, Parts, etc.

‘**Customer For Life**’ is the name for the project to implement Siebel SFA (Sales Force Automation) module for Truck division. The goal of this project is to provide Opportunity and Activity Management to business users. It allows business to set goals for different customers, manage the sales lifecycle from initial contact to sales proposal to order. A set of reports are provided to forecast the sales pipeline and compare it with the actual sales to identify the gaps.

**Project Duration**: 2 years

**Technologies Stack**: Siebel 8.0 Tools, Siebel 8.0 Automotive, Siebel 8.0 eDealer, Siebel 8.0 Sales Wireless, Windows 2003, Oracle 10g

**Role**: Developer

**Responsibilities**:

* + Involved in preparing the Detail Functional / Technical design documents for Customer, Contacts and addresses additions in Siebel.
  + Taken the complete ownership and customized vanilla Siebel Sales Wireless application as per navistar business rules and worked on .swt & .xsl files to display company logo on all the wireless applets. Configured link to Google Map from a wireless applet.
  + Taken the complete ownership and developed one more wireless application for Navistar dealer users. Prepared TD and learning documents.
  + Implemented Task UI for creation of Activities and Opportunities.
  + Implemented a workflow to update activity owner position Id for the activities created through ‘All Activities’ view.
  + Worked on Personalization and state models.
  + Worked on ‘New’ flag indicator.
  + Participated regular interaction with Clients and Users in resolving of the requests
  + Involved in the deployment of builds
  + Worked on performance tuning for account and contact entities

**EDUCATION**

* Post Graduation (M.C.A.) in Computer Science from Sri Venkateshwara University, Tirupathi in April – 2006 with aggregate of 68%.
* Bachelor Degree (B.Sc) in Computer Science from Sri Venkateshwara University, Tirupathi in April – 2003 with aggregate of 85%.
* Intermediate (M.P.C) from Board of Intermediate Education, A.P in April – 2000 with aggregate of 82%.
* S.S.C from Board of Secondary Education, A.P in April – 1998 with aggregate of 79%.