### **Surya Kumer**

**Email ID:12345@gmail.com PH: 987-675-1234**

**SYNOPSIS**

* Around 7+ years of experience in IT industry, including ETL Technologies, Abi Initio, Hadoop Ecosystem environment, Data Warehouse, RDBMS. Designing, developing and maintenance of various applications in verticals such as Financial, Retail and Banking.
* Highly experienced in the **Analysis, Design**and **Development of**Data warehousing solutions and in developing strategies for **Extraction, Transformation and Loading (ETL)**mechanism using **Ab Initio**.
* Good experience with Data Migration, Data Extract Transformation and Loading using **Ab Initio** into **Oracle, Hadoop/Hive, Cassandra, Teradata, DB2 Databases and Files**.
* Database experience on Teradata V13/V12/V2R6, Oracle 11g/10g/9i and DB2
* Analyze Hadoop clusters and developed ETL frameworks using big data tools using **Hive with scheduling, error handling and logging.**
* An avid self-motivated learner with interest in learning new technologies and concepts.
* Strong knowledge of Data Warehousing concepts and Dimensional modelling like Star Schema and Snowflake Schema.
* Good in **Hive** queries to load and processing data in Hadoop File system (HFS).
* Experience in integration of various data sources with Multiple Relational Databases(RDBMS) like **Oracle, Teradata and DB2**.
* Good understanding of NoSQL Database and hands on work experience in writing applications on NoSQL database like **Cassandra**
* Experienced in developing **UNIX shell** wrapper scripts to run **Ab Initio** graphs and scheduling jobs.
* Involved in full lifecycle of various projects, including requirement gathering, system designing, application development, enhancement, deployment, maintenance and support.
* Involved in Metadata mapping from legacy source system to target database fields and creating Ab Initio **Data Manipulation** Language DMLs.
* Good understanding of various Abi component groups such as Compress, Database, Dataset, Partition, De-partition, FTP, Sort and Transform.
* Experience in Query optimization, Performance Tuning.
* Experience in analyzing business requirements and make the changes on a fly in sync with the business team's dynamic needs.
* Experience with all stages of the SDLC and **Agile Development** model right from the requirement gathering to Deployment and production support.
* Hands on **Agile (Scrum),** **Waterfall** model along with automation and enterprise tools like JIRA, Confluence to develop projects and version control, Git.

**CORE COMPITENCIE**



|  |  |
| --- | --- |
| **Category** | **Skill** |
| **ETL Tools** | Ab Initio GDE Co-op, InformaticaPowerCenter |
| **Hadoop Eco System** | Apache Hive, Spark, Kafka, Scala, Kerberos, Sqoop, Yarn |
| **DBMS** | Teradata, Oracle, Hive, Cassandra, DB2, MS-SQL Server |
| **Reporting Tool** | Tableau |
| **Scheduling Tools** | Control M, Autosys, Tivoli Maestro, Redwood |
| **Other Tools** | IntelliJ IDEA, Eclipse, SVN, PostMan, Bitbucket, Github, Jenkins, Toad, Putty, Rational Clearcase, Swagger, Express>IT, Confluence, EME |
| **Programming Languages** | Unix shell scripting, Scala, HiveQL |
| **Data Modelling/Methodologies** | Data Warehousing, Logical/Physical/Dimensional,Star/Snowflake, ETL, OLAP, Complete Software Development Cycle. |
| **Operating Systems** | Linux, UNIX, Windows, Windows NT |
| **Development Frameworks** | Agile |

**PROJECTS EXECUTED**

**Client: Charter, NC** Jan 2017-till date

**Role: ETL/Data Engineer**

**Description:** POS ML – Point of Sale Machine Learning Model This project is to automate the development of Credit Card Fraud authorization detection score rule using Machine Learning method and the development of them into the production environment. The Machine Learning and automated score rule implementation will substantially improve the fraud authorization performance through rapid and comprehensive reaction to new fraud trends. Fraud strategy consists of two key components fraud models and fraud strategy.

**Responsibilities**

* Developing hive queries using historical data based on the fraud models designed by modelers. Working closely with Modelers if any requirement changes on the Machine Learning strategy and models.
* Extracting data from the source hive tables and landing the data in Hadoop Sematic Zone Layer and creating hive external tables on top of the data.
* Converting the hadoop files to local serial files using hadoop/hdfs commands and using the SFTP protocols to transmit the files to downstream application.
* Created shell scripts to automate the whole process starting from extracting the source data, creating the hadoop files, converting to serial files to transmitting the files.
* Automated the application creating scheduling using Control M to send files daily to downstream application.
* Consumed **avro/json** formats of data and parsed it and loaded to hive tables in orc format for data consumption.
* Used **Cassandra** as a service to load the transaction data into Cassandra.
* Responsible for building scalable distributed data solutions using **Hadoop**.
* Involved in creating Hive tables, and loading and analyzing data using hive queries
* Configured continuous delivery pipelines for deploying micro services using **CI/CD Jenkins**.
* Developed Hive queries to process the data and implemented Partitioning, Dynamic Partitions, Buckets in **HIVE**.
* Classifying the data for protection group and ranger policies.
* Performed functional testing using Junit and written cucumber test cases.
* Involved in high level design of processing SOR files from edge node to conformed and semantic zone.
* Created **Managed, External** and **ORC** HIVE tables on top of Hadoop file system. Experienced in handling large datasets using Partitions, Spark in Memory capabilities, Effective & efficient Joins, Transformations and other during ingestion process itself.
* History migration from Teradata to Hadoop using Sqoop.
* Performance tuning the hive queries and monitoring the CPU cycles of the process on the hadoop platform side to make sure the hive query processing is not impacting the other applications.
* Performed **Metadata mapping** from legacy source system to target database fields and involved in creating Ab Initio DML’s
* Involved in creating detail data flows with source and target mappings and convert data requirements into low level design templates.
* Responsible for setting up Repository projects using Ab Initio EME for creating a common development environment that can be used by the team for source code control.
* Implemented various levels of parameter definition like project parameters and graph parameters instead of start and end scripts.
* Developed graphs based on data requirements using various AB Initio Components such as **Rollup, Reformat, Join, Scan, Normalize, Gather, Broadcast, Merge** etc., making use of statements/variables in the components for creating complex data transformations.
* Performed **data cleansing** using AB Initio functions such as is\_valid, is\_error, is\_defind
* Extensively used string\_\* functions, date functions and error functions for source to target data transformations.
* Resolved various issues related to platform, performance, hive query optimization. Regular monitoring of jobs and making sure the file SLAs are met daily.
* Created **Tableau** reports from hive tables for showcase to clients.

**Client: EZ Worldwide**

Aug 2014- Dec 2016

**Role: ETL Application Developer**

**Description:** EADS – Enterprise Authentication Data Store maintains customer authentication information and focuses on constructing the supporting data model populating it with both operational and analytical data like IVR, Call Center/Phone, Correspondence, Authentication Events, Channel Contacts, etc. and leveraging it to enhance existing authentication layers and out sort work flows. This will in turn aid in defining a customer risk profile with score, behavior and customer value data as main input for the authentication purpose. Our project is focused on constructing the supporting data model, populating it with both operation and analytical data by leveraging the existing architecture which will aid in defining Customer risk profiles.

**Responsibilities**

* Performed various key roles in Business Area Risk and Fraud for Enterprise Authentication Data Store team in Chase Card Services.
* Involves setting up of Real-time and **Batch ETL** servers, OADS Database servers, IBM MQ bus servers for real-time data processing, creating the repositories for real time and batch servers.
* **Data Ingestion** into **Hadoop**(HDFS): Ingested data into Hadoop from various data sources like Oracle, Teradata using **Hive**. Created Hive jobs with incremental load to populate **Hive External tables**. Involved in importing the real-time data to Hadoop using Kafka.
* Strong experience on **Real time** application designing of Reader, Parser and Load process and developing associated ETL Abinitio graphs to pick up messages from MQ servers through loading to OADS database.
* Experience in creating **Hadoop entities** in **Confirmed, Semantic and Work** zones of Hadoop ecosystem and ingesting data to **HDFS**.
* Performed ad-hoc queries on structured data using **Hive QL** and used **Partitioning, bucketing** techniques and joins with Hive for faster data access.
* Created schema in Hive with performance optimization using bucketing & partitioning.
* Written Hive queries to transform data for further downstream processing.
* Worked on creating a daily LRF snapshot of OADS data which is loaded on to HADOOP servers using ST5 File Transfer Mechanism and Batch process.
* Exposed to **Hive Structured Streaming** and worked as a part of Agile spike to check the functionality that suits the business requirement.
* Understanding the specifications for **Operational Data Store** for mortgage banking and card services **ETL Processes** and interacting with the business analysts and end users to understand system requirements.
* Automated the ETL processes using **Unix Shell script** and **Scheduling** tools such as **Control M.**
* Designed error handling techniques and Data file level validations duplicate and TDQ checks, **Header/Trailer validation** of **JSON** files.
* Performed Data Analysis and created **High Level and Technical Design** documents based on system requirements.
* Developed graphs that will generate Load Ready Files used to load DB2 tables.
* Built a **semantic layer** on top of the DB2 target tables with specific roles to control access for loading, monitoring and querying the data.
* Performed reconciliation, technical data quality checks, error file check, bad file check using the shell scripts. Developed DB2 load strategy for the target tables. Created DML’s for all the target tables of ASCII, UTF8, Extended ASCII, EBCIDIC formats.
* Worked on improving the performance of Ab Initio graphs by using Various Ab Initio performance technique’s like using looks instead of Join’s etc.
* Extensively used the Ab Initio components like Reformat, Join, Partition by Key, Partition by Expression, Merge, Gather, Sort, De-dup Sort, Rollup, Scan, FTP, Lookup, Normalize, De-normalize, Input, Output and Join with DB. Used AbI features like MFS (8 way), check point, phases etc. Implemented Lookup’s, In-Memory Joins and rollup’s to speed up various Ab Initio Graphs.

**Client: Fabex Tech Solutions Pvt Ltd**

Feb 2013 – Mar 2014

**Role: ETL Application Systems Engineer**

**Description:** To improve Customer and Banker experiences by taking advantage of the full functionality of the Enterprise Multi-Channel Leads Solutions (MCL). Including: **Events data**: adding existing consumer event types for small business to ECP Event store and MCL prioritization logic/scoring engine to create priority leads for business ECNs which include the current product URL pages on wellsfargo.com and the related products. **Best Contact Information data**: add small business best contact data (also known as ECP customer demographic information) to MCL such that the “best contact” information can be systematically applied to Small Business leads which allows the small business LOB data files to stop sending the business contact data.

**Responsibilities**

* Involved in understanding the requirements of the End Users/Business Analysts and Developed Strategies for ETL processes.
* Developed and Implemented extraction, transformation and loading the data from different legacy systems using Ab Initio
* Developed Generic graphs for data cleansing, data validation and data transformation.
* Tune Teradata Development and Production systems for optimal performance.
* Used SQL Assistant (Query man) front-end tools to issue SQL commands matching the business requirements to Teradata RDBMS.
* Developed Ab Initio graphs for cleaning, validating and transforming the data received from various SME’s.
* Responsible for the automation of Ab Initio graphs using unix shell scripts and documented processes procedures as well as flow of the process.
* Extensively used ETL to load data using Ab Initio DB components from heterogeneous source systems from DB2, Oracle etc. to target Data Warehouse Teradata Database and other file systems.
* Participated in data model (Logical/Physical) discussions with Data Modelers and created both logical and physical data models.
* Responsible for trouble shooting, identifying and resolving data problems, worked with analysts to determine data requirements and identify data sources, provide estimates for task duration.
* Performed tuning and optimization of database configuration and application SQL.
* Worked on Auto-sys for scheduling and monitoring the jobs in all environments.
* Worked on various ab initio graphs using different Teradata utilities to load/update huge tables in Teradata. Worked with reporting team to generate reports for the specific requirements from business team.