**Saravanakumar Nanu Mobile:** +1 8472178229

**Data Engineer, USA Email:** [**nskumarin@gmail.com**](mailto:nskumarin@gmail.com)

# PROFILE SUMMARY

* More than 15+ years of experience in analytical and technical expertise in SDLC Automation, CI/CD pipeline, infrastructure design, implementation, service delivery of IAAS/PAAS/SAAS cloud solutions and solving complex with creative solutions.
* Experience with Snowflake and the current project makes use of Snowflake as a warehouse.
* Extensive experience in using IBM DataStage and use the same for implementing various data warehouse concepts.
* Experience in UNIX shell scripting and involved in implementation of the file manipulation, invoking of Oracle, Teradata and Netezza commands, SQL commands, etc.
* Experience in Legacy Architecture and Modern Architecture using Snowflake in cloud environments.
* Good at python scripting and handling large volumes of data using the same.
* Working experience with big data technologies like Sqoop, Spark, Hive and HDFS
* Strong problem-solving skills and analytical skills and have handled many productions Sev-1 tickets/lower environment issues.
* Experience in scheduling tools like Autosys, Tibco, Control-M and oozie workflow
* Experience in Maintain the code and various version control tools like GIT, SVN and ACCUREV
* Experience in retiring the legacy systems and replacing them with Modern architecture to save cost for organization.
* Quick learner of new technology and processes and great training skills.
* Working experience in Agile working model, DevOps, and Software Development Cycle Lifecycle (SDLC).
* Learning and exploring AWS related technologies and product offerings of the cloud platform to get certified on the same.

# TECHNICAL SKILLS:

|  |  |
| --- | --- |
| **ETL Tools** | DBT, IBM Infosphere DataStage 11.X, Talend open studio |
| **Languages known** | UNIX Shell Scripting, Python, Pyspark, SQL, PL/SQL, NZSQL, SOQL, NoSQL |
| **Database** | Snowflake, Netezza, Oracle, PostgreSQL, Hive |
| **Version / CI CD** | Tortoise SVN, ALM Bitbucket, GitHub, Jenkins. Maven |
| **Tools** | Snow SQL and Snowflake Web UI, Atlassian tools, Aginity Workbench, PL/ SQL Developer, Query Analyzer, Putty, Sqoop, Hue and Tableau reports |
| **Domain Knowledge** | Retail market research, Automobile, Banking |
| **Scheduling tools** | Control M, TIBCO, Oozie and Airflow |
| **Operating System/Cloud environment** | Linux, Windows, AWS and GCP |

**WORK EXPERIENCE:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Custom  er | Employer | Location | Period | Role |
| USAA | KFORCE INC | SanAntonio, TX | Sep 19 – Till Now | Advanced Software Developer |
| FMCC | Systems Technology Group Inc | Dearborn, MI | Oct 17 –  Sep 19 | Software Developer |
| Nielsen | Tata Consultancy Services | Schaumburg,  IL/Chennai | Feb 10 –  Sep 17 | Assistant Consultant |
| Citi Bank | Polaris Software Lab Ltd | Chennai, India | Apr 08 –  Feb 10 | System Engineer |
| India  bulls | ESoft Factory Pte Ltd | Chennai, India | Oct 05 –  Mar 08 | Software Engineer |

# Data Engineer Sep 2019 to till date.

USAA team creates data lakes, data marts and warehouses for the reports involving metrics of each subunit. The BA team creates stories based on requirements given by the business. Various kinds of dashboards and Hadoop loads are dealt with.

Project 1: Onboarding new Investment Manager sends the Intraday trade files to USAA after receiving the files, Staging the data and transforming & creating order files to accept the orders in the CRIMS System

Project 2: SLA is a project to retire legacy and replace it with a custom-built system to have a single vendor platform to avoid the complex integrations between subledger accounting system and General Ledger system (People Soft)

Project 3: ARM Project - Converting SLA files to CJT format to ingest data in Cloud. Roles and Responsibilities:

* Create ELT pipelines with the help of dbt and snowflake.
* Build custom materialization for replacing the existing reusable jobs for Type1, Type 2 load, s3 to snowflake file ingestion process.
* Involved in the planning, design, and development for the Oracle and Netezza to Snowflake data migration.
* Convert all the ETL projects to ELT model in order to facilitate easy movement from Netezza to Snowflake
* Involve in Data Modelling activities and create data models to be published in a centralized documentation repository.
* Create Python frameworks to automate the existing processes with multiple manual touch points.
* Prepare source to target mappings as per the business requirement.
* Develop python/shell scripts to parse complex input data from various input systems.
* Create DataStage jobs using the custom framework for both development and enhancement.
* Create data flow diagrams for the business processes at column level for offshore teams.
* Load huge chunks of data from Netezza to Hadoop for analysis by business teams.
* Review the ETL/ELT code with the operations team and other teams to ensure that the coding standards are incorporated.
* converted RDD into data frame and store the data into a hive table using Spark.
* Handled different files systems such as Parquet, ORC, Sequence Files, AVRO, Fixed Width, Delimited, JSON, XML.
* Create data models for the new requirements coming from the business side by following the project specific naming standards.
* Highly experienced in Snowflake across various features such as loading data with Snowflake and AWS, External Stages, Snow pipe, Streams & Tasks, Time Travel, zero copy cloning, handled various file formats including Parquet, Json etc.
* Determine the type of the dimension table (type 1/SCD) based on the business requirements.
* Data ingestion activities involved in AWS S3 vendor files to load data into Snowflake using lambda and convert JSON file using AWS Streamsets
* Created data pipelines for the cloud and on demand infrastructure.
* Analyze SAS models and convert them into IT supported DataStage projects.
* SQL query tuning for processes involving multiple tables and databases and performance improvement suggestions for long running queries
* Create Subject Area Marts and data marts out of the warehouse.
* Deploy Control M cloud jobs in Kubernetes clusters to support DBT data pipelines.

# Senior Backend Developer

# FMCC # Duration # **Oct 2017 –Sep 2019**

Project 4: Anacredit is a project to retire legacy system to replace it with a custom-built system DW in Teradata and data lake in Hadoop to provide better insights data for organization using modern architecture.

# Nielsen # Duration # **Feb 2010 –Sep 2017**

Project 5: AOD (Answers on Demand), a dynamic reporting platform of the Nielsen Company had huge growth which led to capacity issues due to multi-terabytes of data volume surge. This needed several initiatives to reduce the run time of existing modules and to improve the scalability of the application in general. Big data initiatives were tried out using Sqoop data load followed by processing using Hue tool with Hive, Spark, and Impala

Project 6: Answers On-Demand (AOD) is an end-to-end Nielsen data management and reporting process which source the Dimension and Fact data from legacy repository systems and transforms to drive the end-to-end data processing and reporting systems associated with the Netezza output database**.**

# Citibank # Duration # **Apr 2008 – Feb 2010**

Project 7: This project targets to generate Current portfolio value & purchased value of investment to understand the profit and loss of customer investment and send the updates to the customer through Mobile or Email or Both in Batch process.

# India bulls # Duration # **Oct 2005 –Mar 2008**

Project 8: IBOSS is a comprehensive solutions suite for stockbrokers and traders, giving absolute control over the entire trading process from order to settlement. It stores the complete client information and all the shared details with the appropriate document and all types of transaction details are also stored in the database. It is a centralized system. It integrates all your operations spanning front-office, management, and back-office.

Roles and Responsibilities:

* Involved in developing DataStage Designer Server and Parallel jobs for Extracting from Operational data store (ODS), Cleansing, Transforming, and Integrating/Loading Data into Data Warehouse (DW) for reporting**.**
* Analyze the existing code base and identify the areas that had scope for improvement and create the requirements for the same.
* Used NZSQL, NZLOAD, NZ\_MIGRATE scripts for day-to-day loading/Unloading and migration activities.
* Created PL/SQL stored procedures, functions, and packages to refresh the data in Metadata (MDM) Schema and publish the data to reporting schema.
* Created database objects such as tables, views, synonyms, indexes, sequences,
* SQL query tuning to reduce the full table scan scenarios and improve performance.
* Reduce the number of database connections by writing multiple SQLs into a file and execute at one shot.
* Identify and eliminate process dependencies to enable prod support to run numerous processes ahead of time.
* Chunking and parallelism of database operations by spawning multiple threads wherever possible.
* Enhance processes to track multiple configurable threads and introduce auditing wherever possible.
* Extensively used DataStage Manager and IBM infosphere Connector Migrations tool to Export/import/Migrate DataStage components.
* Provide performance improvement proposals and implement them.
* Led Data migration from Netezza to Hadoop using Sqoop commands.
* Led DataStage migration from DataStage 8.1 to DataStage 11.5 for multiple modules.
* Led Data migration from Oracle to Postgres database.
* DataStage to Talend migration POC.
* Created complex TERADATA SQL to perform Validation using jobs in each layer (ODS, Integration, Data Warehouse and Data Mart)
* Developed Job Sequences for automating and scheduling the DataStage Jobs.
* Involved in creating Hive tables, running the Ingestion jobs, and monitoring the jobs using Hadoop Web interface, HUE.
* DataStage performance enhancements like clearing up scratch disk space utilization regularly, Stage level changes, cleanup of intermediate datasets used in the project to save disk space and improve performance.
* Reduce manual touch points in the application and integrate them into TIBCO orchestration.
* Led the effort to retire legacy code and onboard all clients to core platform.
* Work on POCs to migrate to freeware and cost-effective tools and Hadoop open-source platforms.
* Move most database operations to UNIX file level operations to reduce connection times.

# Education (Period: Mar 2001 - Apr 2005)

Bachelor of Engineering in Computer Science Engineering from Anna University, India

