**SRI SMRUTHI VARATHARAJAN**

**949-771-6550**

**smruthivaratharajan@gmail.com**

**|** [**https://www.linkedin.com/in/sri-varatharajan/**](https://www.linkedin.com/in/sri-varatharajan/)

*Over 7+ years of experience in Product and Data Platforms, Data Integration, BusinessIntelligence and demonstrated experience in Data Mining of large datasets including structured and unstructured data. I have the expertise in Data Validation, Data acquisition, Data Visualization and Reporting and used Machine Learning tools in Predictive Modeling.*

SUMMARY

* Experience to managing technical support for product launches including building new data models, modifying existing data models, building Minimum Viable Product (MVP) dashboards with required KPIs.
* Extensively worked on data extraction, transaction and loading data from/to various DB like Redshift, Big Query, Netezza, Oracle, SQL Server, DB2 and Flat Files using DBT, DataStage, Python and Airflow.
* Proficient in creating Dashboard Reports for technical and non-technical teams using Tableau and Power BI.
* Hands-on experience with Airflow and Data built tools to implement ETL/ELT pipelines and orchestrations.
* Performed investigations into data and business questions and presenting findings into an effective story to be actionable for business users.
* Expertise in analysis, design, implementation, testing activities and production support of Enterprise Data warehouse and Business Intelligence applications.
* Proficient in Agile delivery model, worked with Delivery team and managed product Scope, Vision to meet deliverables, coordinated with Product Owner for Product Backlog, Release Backlog and Sprint Backlog.
* Experienced in using data visualization tools like Tableau and Power BI for advanced Analytics, create Dashboard reportswith charts and graphs, integrate data with complex data sources like Salesforce and related ERP platforms.
* Worked closely with internal teams and stakeholders in drafting and gathering requirements, executed data model development, and built dashboards to show the insights from the data gathered.
* Hands-on experience in data analysis, enterprise Data Quality, Data mapping, Data profiling and ETL workflows.
* Experience in partnering with Data Management Leaders, Interface Engineers, Database Engineers, BI Engineers, Product Owners, Business Analysts, Project Managers, Business Partners, and others across teams to achieve project and business goals.
* Defined and created application data dictionaries by working withbusiness and technical partners. Identify the quality and completeness of data by workingclosely with business, technical, and project teams.
* Developed SOPs for new processes implemented and documented re-usable tools on SharePoint, created a central repository of required documents for process control.
* Possess advanced programming skills in Python and used statistical parametric analysis methods such as Correlation, Regression, Analysis of variance and general linear models in data modeling and data mining.
* Good Knowledge of Machine Learning concepts and implementation like Feature extraction, Feature Selection, Dimensionality reduction and Classification.
* Experienced in using Erwin Data Modeler to create Data Models by Forward and Reverse Engineering.
* Extensive knowledge of time-series, trend and statistical analyses to python and MATLAB.
* Good understanding of Strategic Planning, Product Development, Project Management Agile/ Scrum Methodology timely delivery of product with high-quality rollouts to on-time and on-budget conclusion.
* Knowledge of Web and Mobile Application development using Salesforce, Power Appsand UI modeling.

EDUCATION

* MS in Bioengineering & Biomedical Engineering (Specialization: Data analytics) | University of California, Irvine
* BE in Biomedical Engineering | Anna University

TECHNICAL SKILLS

|  |  |
| --- | --- |
| ***Programming Languages*** | SQL, Python, R, HTML, JavaScript |
| ***VisualizationTools*** | Tableau, Power BI, SAP, Google data studio, Advanced Excel, Power Charts |
| ***Databases*** | PostgreSQL, MySQL, Microsoft SQL Server, MongoDB, IBM DB2 |
| ***Cloud*** | AWS, Salesforce CRM, GCP |
| ***ETL and modeling tools*** | Airflow, DBT, DataStage, Erwin data modeler |
| ***Software*** | HEAP, GitHub, Confluence, SharePoint, MS Visio,Lucidchart, MATLAB, SPSS |
| ***Miscellaneous*** | JIRA, ServiceNow, Remedy, StarTeam, Salesforce CRM, Data Loader |

PROFESSIONAL EXPERIENCE

Data Analyst, Product & Strategy | **Helix, *San Mateo, California*****Mar 2021– CURRENT**

**Key Projects:**

* New Product launch: SSO/Non-SSO workflow (SSO – Single Sign On)
* New Partner Launch: EHR partner launch
* Airflow:Data model launch
* HEAP integration and product analytics using Salesforce Object model and Data for ERP workflows
* Product, Operations and Business dashboards launch and maintenance
* MyHealth Workflow – Data pipeline
* Confirmatory Workflow – Data pipeline
* Blood sample Launch – data model updates and MVP dashboard launch

**Responsibilities:**

* Built ELT pipeline to ingest and transform data within redshift warehouse using Airflow and data build tool (dbt)
* Designed ELT pipeline for the new Airflow data model orchestration setup and using a data source outside of Tableau extracts, used SQL to conduct data validation and analysis of ETL completed data
* Built data engineering pipelines using MWAA (AWS Managed Airflow) and Python, reduced the data prep and monitoring effort by 70%
* Built data and reporting infrastructure from the ground up COVID and population genomics (PopGen) business lines using Tableau and SQL to provide insights into Partner Success and Business KPIs
* Worked on converting the existing full load ELT data load into incremental loads with downstream and upstream data extraction and loading changes to optimize the ELT pipeline
* Built operational reporting in Tableau to find areas of improvement, improved the next day Turn Around Time (TAT) to 97% and made sure no kits/samples are left behind in the workflow
* Managed HEAP data integration to redshift, built product analytics reports using Tableau and HEAP to understand user-behavior and track product KPIs like conversion, retention, user interaction
* Created events on HEAP using the user interface to track user behavior and measure product analytics metrics like retention, conversion, monitor choices users make on the final step of the registration/results-viewing
* Developed ingestion flow for partner data via SFTP, created dashboard to monitor Enrollments, Demand forecasts vs actuals, reduced manual workload of 90 hours
* Created a HAMUI ingestion worklist by creating a redshift data pipeline with daily data checks and generated automated report for Lab operations
* Co-ordinated end-to-end product and partner launches, updated data models and workflow changes using dbt, implemented MVP metrics tracking via tableau
* Created new data models, created, and updated Redshift databases for the confirmatory workflow launch, created data validation and check points to optimize the new pipeline
* Created specimen type fields and related data model changes to support the new workflow launch using Salesforce CRM and Salesforce App integration data.
* Created Tableau Dashboard reports using Salesforce integration data for health outcomes, optimized health services and managed Patient 360 Dashboard using advanced recovery systems and treatment data.
* Performed backfills of Redshift databases after an interruption of airflow DAGs for data flow updates or troubleshooting existing data updates in the pipeline
* Worked extensively in troubleshooting and removing data duplicates which proliferated in the downstream production data sources used for business intelligence
* Setup interactive Tableau dashboard to monitor the health of the exome workflow, identified potential for turnaround time optimization in the product workflow and communicated to the team for implementation
* Created automated reports with KPIs to be sent to business partners via SFTP on a daily and weekly schedule using Airflow DAG
* Created dashboard and reports using Tableau to satisfy ad-hoc requests and business questions on a product or business linefrom marketing and digital healthcare experience with health systems interoperability.
* Created monthly reconciliation reports using Tableau for Finance, billing, and settlement questions for the CFO team and partner invoice support
* Created post partner launch MVP dashboard in Tableau, gathered key metrics and requirements from dedicated advisors and product owners to setup the dashboard and monitor the performance
* Gathered business requirements, setup the data model for new partner launch for the population genomics business line, updated required databases in redshift and prepared for partner launch
* Troubleshoot existing ETL/ELT pipeline to ensure interrupted data flow in the pipeline, make sure upstream dependencies for data ingestion are taken care by the respective teams
* Generated SOP for new product launches, data model and process updates on Confluence for document keeping and internal sharing within the team
* Organized the tableau folder file structure at the company and business line level to enable controlled and limited access aligning to the privacy and security policies in the company

**Tools/Environment:** Airflow, DBT, Python, SQL, PostgreSQL, Tableau, HEAP, AWS (Redshift, S3), Google Sheets, Confluence, JIRA, Slack, Lucidchart, SSIS.

Lead Data Analyst| ***Department of Health, D.C, District of Columbia*** **May 2020 – Feb 2021**

DC Department of Health promotes health, wellness and equity across the district and protects the safety of residents. Open Data platform connects Health and Human Services and presents data to Dashboard screens from surveillance data, Vaccine and Reopening data sets. I was involved in AWS migration efforts of DC health data sets to AWS S3 buckets for use of ArcGIS Dashboards and MicroStrategy teams.

**Key Projects:**

* Data Surveillance – COVID Dashboards
* Application Development – Physician Portal
* IT Legacy to AWS migration for Department of Health

**Responsibilities:**

* Created workflows to build data mart for the analytics team for predicting future utilization measures, reduced the data prep effort by 70% for outbreak reporting requirements.
* Created SSIS platform requirements and deployed Stored Procedures for workflow tasks like data recovery.
* Created functional requirements for data visualizations using Power Query in Power BI for business KPIs that reduced manual efforts and improved efficiency in data acquisition workloads for Covid Dashboard display.
* Used Erwin data modeler to capture Reverse Engineering models, thus captured production schemas in ETL.
* Automated backups from on-premises legacy system to AWS using VEEAM, eliminating manual process, reduced burnout rates by 70%.
* Created migration tool, centralized repository on SharePoint to assist in resource/requirements gathering, reduced downtime, improved efficiency by 80%.
* Developed SOW, Business Requirements, cost analysis to manage the necessary infrastructure & AWS migration.
* Worked closely with the project manager and technical analyst to create architecture requirements and hosting plans for the legacy to AWS migration.
* Served as point of contact for escalations and outages, managed plan of action, developed project scope and executed rollouts.
* Managed projects scope, sprint planning and business requirements through Agile planning via Remedy.
* Performed statistical analysis (correlation, regression) on user data, built reporting infrastructure using SQL and Power BI to provide real-time insights into population health.
* Created project plans, converted business requirements to technical requirements for developers and created flow charts, UI requirements using Visio.
* Created Salesforce Custom objects to be used in reports to track data pipeline and report KPIs to management.
* Performed end-to-end testing of the contact tracing application, including the UI and the backend data pipeline, preparing for the launch of the mobile and web application
* Managed UI layouts, workflows, document templates and record types for the patient-physician portal application using MS excel, Visio
* Developed hypotheses for A/B testing for product optimization, analyzed and reported key insights using salesforce dashboards on design improvements to management
* Managed development of cutting-edge SaaS solution for real time transcriptions, reducing average physician workload processing time from an average 7 hours to less than 20 minutes
* Led development of a suite of web & mobile apps providing customers with the ability to complete fast, secure transactions as well as teleconsultation via mobile devices
* Led cross-functional core teams and development activities to meet objectives including scope, budget, and schedule for completion of healthcare applications and advanced healthcare units (AHU)
* Implemented strategic planning to provision the necessary resources for timely delivery of product with high-quality rollouts to on-time and on-budget conclusion and 30% reduction in burnout rates
* Built resource requirements, maintained inventory, managed budget $2.7M for completion of Phase I&II of product development
* Managed implementation of Health portal application, coordinated off-shore Salesforce development team, followed Agile/scrum methodology
* Defined key business metrics, created reports using MS Excel and worked with executive leadership to identify business objectives, set goals & milestones for long term and short term
* Developed prototype designs and development for advance healthcare units for COVID-19 isolation using visio and CAD
* Lead R&D team and coordinated with contract builders for completing the prototype complaint with CDC guidelines
* Authored various documents: Project Charters, Project Plans, Validation Plan, Validation Summary Reports, Training Materials, Release to Production Memos, Training Materials, and User Guides.
* Managed direct relationship with third party vendors and offshore development teams, created communication channel, and project’s review for management
* Lead of team of interns for validation and testing of developed application, defined acceptance criteria for the final product

**Tools/Environment:** SQL server, MySQL,Salesforce, AWS (Redshift, S3, EC2), MS Excel, Visio, Confluence, JIRA, Microsoft application suite, VEEAM, VMWare, PowerBI, SQL.

Algorithm Engineer | ***Cercacor Laboratories, Inc, Irvine, California*****Oct 2019 – May 2020**

Cercacor Laboratories, Inc is specialized in research and development of medical equipment, medical devices and including patient monitors, respiratory sensors, cardio-pulmonary recording devices and other equipment used in surgical procedures. I worked on projects relating to Cloud migration using Google Cloud resources and managed data processing from data acquisition to GCP platform for discovery.

**Key Projects:**

* Migration of existing DataStage dependencies to GCP
* Launch of Clinical Trials
* Predictive model testing and optimization

**Responsibilities:**

* Developed ETL pipelines to integrate customer health and wellness information from multiple data sources like Oracle, SQL server to google cloud, increased data quality by 40%
* Performed regression analysis using MATLAB on users’ vital parameters and behavioral patterns, effectively improved the prediction efficiency of the model by 30%
* Performed statistical analysis on physiological parameters such as pulse rate and its correlation to user patterns of sleep, diet, and activity
* Developed, executed test plans and test cases and continued evaluation of test cases for automation framework
* Built operational dashboards using SAP to improve scheduling the clinical trial cycle, and reduced wait times by 10 mins
* Generated output tables by developing ETL pipelines to extract values from tables in turn increasing data quality
* Built ETL pipeline to ingest and clean data from data sources like Oracle, SQL server to big query using dataflow.
* Automated the data transfer process to cloud storage using gsutil and python, reduced the transfer time by 50%
* Developed data cleansing script in python to remove the special characters, data type issues during data import, reduced the data load and failures by 70%.
* Converted Netezza stored procedures to big query routines and done process optimization, improved the processing time by 13%.
* Migrated DataStage jobs to GCP and automated using google composer improved the processing time by 22 %, troubleshoot issues and resolve tickets in development and testing environments.
* Led discussions with data architects, engineers, SMEs to understand, and translate business requirements into technical designs using Visio
* Improved efficiency by 40% in the process and automated manual reports using Tableau and other ETL tools
* Built ETL models to enhance business processes and automate data extraction which lowered manual works of 200hours, coordinated with DataStage development team to understand and execute ETL workflows.
* Responsible for preparation of design documents, Unit testing, performance review.
* Organized data collection methodology, pre-processed clinical trial data, and loaded into data marts for data scientist operations.
* Developed thorough product requirement documents including scope, user personas, user stories, wireframes, and mockups for new digital products .
* Understood user requirements, prepared product roadmap, pitch, estimated efforts, prioritized features & coordinated development sprints .

**Tools/Environment:** DataStage, GCP, Netezza, SQL Server, Oracle, MATLAB, SAP

Data Research Analyst | **University of California, Irvine** **Sep 2018 – Sep 2019**

I assisted the Clinical Research Coordinator and Data Management team in the research program with the Department of Family Medicine and other innovative programs that used ETL data pipelines in the data processing, data cleansing and data ingestion process.

**Key Projects:**

* Quality Management Systems – Electronic Data Capture (EDC)
* Tools Innovation for Data Engineering and Integration (EDC Tools)
* Clinical Research and Data Management

**Responsibilities:**

* Demonstrated ability in communication and worked with both technical and non-technical personnel at various levels in the design phase of the project for data acquisition and data engineering.
* Designed ETL pipeline for real-time data collection, cleaning and loading for statistical analysis.
* Designed a process for data collection, quantitative analysis to determine global statistics of network activity and connectivity using power and phase-based measures to estimate functional connectivity.
* Introduced a new service review and data gathering system, leading to changes that decreased the data load and errors in the process and improved efficiency.
* Converted data into actionable insights by modelling and statistical analysis, converted raw data from excel into useful dashboards and reports for the team to understand the workflow and process needs.
* Developed ETL pipelines to combine data from multiple data sources like flat files, CSV, MS Excel along with the real time data reading microelectrode system and increased data quality by 50%.
* Designed and built statistical analysis models on large data that helped to prove the hypotheses in the control testing panel.
* Analyzed the existing application and prepared the source file format that needs to be created for processing.
* Involved in understanding the scope of application, present schema, data model and defining relationship within and between groups of data.
* Extracted, cleaned, transformed, and validated the model data with the purpose of understanding or making conclusions from the data for decision-making purposes and hypotheses testing.
* Created visually impactful dashboards in Excel for data reporting by using pivot tables and VLOOKUP. Extracted, interpreted, and analyzed data to identify key metrics and transform raw data into meaningful and actionable information.
* Designed and executed algorithms to pre-process and analyzed data using python to prove desired hypotheses.
* Developed algorithm to employ power and phase-based measures to estimate functional connectivity .
* Performed cross-correlation based analysis to determine global statistics of network activity and connectivity
* Analyzed samples with MEA technology and statistical analysis using Python, maintained these instruments including cleaning and running controls and blanks to ensure accuracy.
* Used information value, principal components analysis, and Chi square feature selection techniques to identify
* Worked closely with Data Architect to review all the conceptual, logical and ETL design models with respect to functions, definition, maintenance review and support Data analysis, Data Quality and ETL design that feeds the logical data models.
* Interacted closely with Subject Matter Experts (SME) for requirements gathering and execution.
* Conducted Exploratory Data Analysis using Python Matplotlib and Seaborn to identify underlyingpatterns and correlation between features.
* Responsible for critically reviewing data analysis requests and identifying opportunities to support or impact lab operations decisions or management purposes.

**Tools/Environment:** MS Excel, MATLAB, Python: NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, MEA

Product Analyst Intern| **LIVMOR, Inc, Irvine, California** **Jan 2018 – Sep 2018**

LIVMOR, Inc is a digital health solutions company that builds medical device systems with data collection, remote data acquisition, continuous monitoring and data management solutions deployed as part of Halo project integration.

I was involved in the device monitoring systems enabled on Mobile platform that assisted physicians to improve patient outcomes and reduce healthcare cost with better quality data for diagnosis and treatments.

**Key Projects:**

* Testing of web and mobile applications
* Developed new BI solution using BI tools for process improvement
* Data gathering, cleaning protocols layout
* Product design and development with statistical analysis and data modelling
* FDA approval of LIVMOR Halo+ (PPG signal monitoring)

**Responsibilities:**

* Designed data model and data mapping in the new Digital integration and Data enablement platform.
* Automated SQL procedures to transfer 10,000 patient data from external applications to LIVMOR applications, eliminating manual errors in data acquisition.
* Performed statistical analysis using R programming, compared the LIVMOR smartwatch to an approved FDA standard, achieved 94.7% sensitivity and 94.3% specificity.
* Standardized ETL processes and tools to automate data extraction and manual reporting, reducing manual workload of 400hours across the entire operation.
* Enhanced the process of viewing reports in an effective manner by creation of real time dashboard in Tableau to analyze the total.
* Medicaid budget for different provider services resulting in better decision management for the business
* Developed KPI dashboards with key recommendations in Tableau, analyzing parameters such as user enrollment, region, and user complaints to support leadership in decision making
* Developed, executed test plans and test cases and continued evaluation of test cases for automation framework.
* Assisted, prepared regulatory compliance documents including 510Ksubmissions and simulation protocols.
* Performed risk analysis, prepared SOP, generated QA related documents-risk factor analysis
* Build detailed excel reports for Ad-hoc analysis and accreditation reporting
* Develop data strategy with the leadership to identify data gaps, gather & analyze data like demographics, enrollment, registration, and user behavior analytics.
* Extensively used SQL to extract, QC data for data migration and data processing to data warehouse.
* Present Data insights & BI solutions to leadership & non-technical teams
* Understand the source data and enterprise data models and create source-to-target mapping documentation.
* Collaborate with Enterprise Architects for identification and ingestion of Strategic Data Sources.
* Collaborate with business and technical stakeholders while defining solutions.
* Collaborating with the enterprise Architect, Program Lead for communication and program documentation.
* Strong experience in data profiling and data analysis using complex sql queries.
* Documented technical specifications, enhancements, created SOP for technical process change and control.
* Collaborated with development teams for design discussions and creation of data models.
* Collaborate with testing teams in creation of test cases, defined and reviewed acceptance criteria for test pass and failure, communicated results to the team to reduce the failure rate
* Collaborate with the data scientists to organize huge sets of data for data modeling including masking data to protect privacy of the users, cleaning thedata and reduced the data error

**Tools/Environment:** Tableau, SQL, T-SQL, SSIS, SSRS, SSMS, FDA, 510K, QMS, Documentum.

Health Data Analyst| **Max Healthcare, Chennai, India** **May 2016 – Aug 2017**

Max Healthcare is a hospital chain based in India and operates National Capital Region of Delhi and North India. I worked on the design of Supply Chain and Automated Order Management System from the armed services community and patient population to support insured beneficiaries with Government of India Health plans.

**Key Projects:**

* Developed new process for data automation and reporting
* Design and delivered new Supply Chain automation launch
* Design and development of new equipment maintenance launch
* Maintenance of existing pipelines and data models, statistical analysis to understand consumer behavior

**Responsibilities:**

* Collaborated with database engineers to implement ETL process, wrote and optimized SQL queries to perform data extraction and merging from SQL server database.
* Gathered, analyzed, and translated business requirements, communicated with other departments, created data models as needed for the new program launch
* Performed data visualization and Designed dashboards with Tableau, and generated complex reports, including charts, summaries, and graphs to interpret the findings to the team and stakeholders.
* Worked with external teams to implement automated SQL procedures to transfer 100,000 patient data from Sapphire to external applications, eliminating manual errors
* Resolved high priority tickets on JIRA periodically to reduce incident counts and breaches thus improving customer satisfaction
* Developed KPI dashboards with key recommendations in Tableau, analyzing churn parameters such as mode of payment, region, and user complaints to support top management in decision making
* Generated output tables by developing ETL pipelines to extract values from multiple tables in turn increasing data quality by 30%
* Developed reporting database and client interface, conducted sensitivity analysis of communication protocols to enable the team to perform ad hoc research on storage and server infrastructure environment
* Identified potential impacts on infrastructure environment and planned maintenance activity by utilizing two-tier architecture thereby increasing efficiency and reducing manual work by 85%
* Created Data stage jobs to extract, transform and load data from various sources like relational databases, flat files, etc. into the target tables.
* Participated with project teams through the planning, development, and test stages to ensure full understanding of business requirements.
* Gathered information from existing systems, analyzed data and time requirements, built effective dashboards for reporting key KPIs to the team and management
* Automated report for supply chain and inventory maintenance, reduced manual workload by 80hours, improved the supply chain process and QC process for the biomed equipment maintenance
* Prepared detailed specification, SOPs for new process development and code deployment and changes.
* Simplified and optimized complex queries involving multiple tables while optimizing procedures and triggers to be used in production

**Tools/Environment:** SQL Server, Tableau, Dbeaver, SQL, T-SQL, SSIS, SSMS, Stored Procedures, JIRA.

---