Vijayakumar Kannan

, Tampa, FL, 33637 \* (848) 702-2241 \* technologist.vijay@gmail.com

**Career Objective**

Driven and ambitious developer with a proven history of exceeding goals and achieving high client satisfaction ratings. I am seeking a senior python or snowflake developer position with a fast-growing technology organization where I can marry my enthusiasm for relationship-building and my passion for innovation.

**Experience**

Juul LABS, **Tampa, FL**

*Senior Data Engineer, Sep 2022 – Present*

* Leverage Python to design robust, reusable and scalable data solutions and data pipeline frameworks to automate the ingestion, processing and delivery of both structured and unstructured data  
  • Drive development of large-scale data engineering projects  
  • Create data pipelines in airflow, DBT and the general suite of Google Cloud Platform  
  • Build, manage, and support data models and ensure data quality with data tests in Monte Carlo and Data fold and Comfortable working in a scrum agile environment using Jira  
  • Partner with Data Scientists, Data Engineers and Business Analysts to build configurable, scalable, and robust data processing infrastructure  
  • Work closely with our sales, operations, research, and finance teams on data storage, retrieval, and analysis  
  • Develop new systems and tools to enable stakeholders to consume and understand data more intuitively  
  • Create and establish design standards and assurance processes to ensure compatibility and operability of data connections, flows, and storage requirements  
  • Validate model transformations for data integrity (source/target tables values and counts are expected, ensure proper data cleansing)  
  Keep Juul on the cutting edge of data technology

*Skillset: Google Cloud Platform (BigQuery,Pub Sub,Cloud Build. GCS, Data Studio, Data Fusion & Data Prep), Python, PySpark,Azure,DBT,Kubernetes,Docker,Databricks, Looker, Tableau*

IBM - DirecTV - Contractor**, Tampa, FL**

*Senior Data Engineer (Python, PySpark, ETL), Aug 2021 – Sep 2022*

* Design, development, and implementation of performant ETL pipelines using python API (pySpark) of Apache Spark on AWS and Azure.
* Good understanding of Spark’s RDD API Spark’s Dataframe and API, monitoring spark jobs using Spark UI
* Integration of data storage solutions in spark especially with AWS S3 object storage and Azure container.
* Performance tuning of pySpark scripts.
* Developed highly maintainable Hadoop code and followed all best practices regarding coding.  
  Inspected and analyzed existing Hadoop environments for proposed product launches, producing cost/benefit analyses for use of included legacy assets.
* Responsible for development of high performance, distributed computing tasks using Big Data technologies such as Hadoop, NoSQL, text mining and other distributed environment technologies.
* Developed the post-processor script in PERL for data ingestion process to move the unstructured files to the landing zone to targeted systems.
* Worked on building automated reports and dashboards with the help of Power BI and other reporting tools.
* Experienced in tools and systems on MS SQL Server BI Stack, including SSRS and TSQL, Power Query, MDX, Power BI, and DAX.
* Implemented the Big Data programming languages and technology, writes code, completes programming and documentation, and performs testing and debugging of applications.
* Analyzed and designed programs, debugs software enhancements and/or new products used in distributed, large-scale analytics and visualization solutions.  
  Interacted with data scientists and industry experts to understand how data needs to be converted, loaded and presented. Worked in highly agile environment.
* Developed and implemented the new flow of Component trades which includes loading the feed file in Perl/Unix shell, applying the business rules through stored procedures, and generating the report in Perl for the end users.
* Proficient in developing,publishing, and scheduling Power BI reports as per the business requirements.

*Skillset: Azure , Databricks,Apache NiFi,Data Router,AWS (S3,Lamba & Redshift), Snowflake, Unix Scripting,Python, PySpark ,Kubernetes,Docker, Tableau*

GATHI ANALYTICS - USAA - CONTRACTOR, Tampa, FL

*Data Engineer (Python & DBT Snowflake), Jan 2021 - Jul 2021*

* Adhere to IT Control Policies throughout design, development and testing and incorporate Corporate Architectural Standards into application design specifications.
* Participate in code reviews and ensure that all solutions are aligned to pre -defined architectural specifications; identify/troubleshoot application code-related issues; and review and provide feedback to the final user documentation
* Designed snowpipe and mappings for optimal performance in a large-scale Unix/RDBMS environment and knowledge of optimal mapping techniques
* Development in a data warehousing environment, with expertise in: CDC, type-2 processing, dimensional surrogate key handling, exception data processing, etc
* Design and implement performant data ingestion dataflows from multiple sources using Azure Databricks. Prepare deployment groups for migration of code to QA and Production environments
* Development using Teradata utilities (fast load, multiload, tpump to load data, BTEQ scripts, DML and DDL, Fast export) and Informatica Teradata optimizations (PDO)
* Developed ETL pipelines in and out of data warehouse using combination of Python and Snowflakes. Also, worked on the features such as SnowSQL , External Function, External Tables, Storage Integration, Stage and SnowPipes.
* Experienced in creating the Application Interface Design (AID) and High-Level Design (HLD) documents for the applications.

Worked on developing Modern Data Warehouse solutions using Azure Stack (Azure Data Lake, Azure Databricks) and PySpark Develop batch processing and integration solutions and process Structured and Non-Structured Data Demonstrated in depth skills with Azure Databricks and Spark and SQL Collaborate and engage with BI & analytics and business team.

* Established end-to-end CI/CD with dbt Cloud and Gitlab pipelines.

*Skillset: Datastage, Snowflake,DBT, Oracle, Python,Control-M, UNIX Scripting*

State Of Michigan DHHS - Contractor**, Lansing, MI**

*Data Engineer, Jul 2020 - Dec 2020*

* Design, development, and implementation of performant ETL pipelines using python API (pySpark) of Apache Spark on AWS.
* Designed and created ETL code installations, aiding in transitions from one data warehouse to another.
* Designed integration tools to combine data from multiple, varied data sources such as RDBMS, SQL and big data installations.
* Developed and acquired test data and runs tests to determine accuracy of program logic to produce desired results; debugs and revises programs based on test results.
* Used program specifications to prepare logic sequence of program source statements in development, modification, documentation, and maintenance of computer programs.
* Developed and prepared documents for general design of solutions that meet system and security requirements, including selection of alternative approaches.
* Develops applications to be used in client server environments.
* Developed and maintains automated data processing systems and computer application programs.
* Developed specifications to detail what new systems or enhancements to existing systems can accomplish.
* Worked on data masking using Data Discovery & Classification and Dynamic Data Masking techniques in on-premises database.
* Directed system enhancements and projects through SDLC life cycle, Agile methodologies, defining project objectives, aligning project resources to achieve goals, and ensuring established timelines are met.

*Skillset: Datastage, Informatica, Oracle, Python, XML,JSON,Azure,Databricks,PySpark,Azure, Looker, Power BI*

CitiusTech - Mayo Clinic - Contractor**, Rochester**

*Data engineer Feb 2017 - Jun 2020*

* Collaborated with quality assurance team to verify correct implementation of software development changes.
* Involved in Data mapping specifications to create and execute detailed system test plans
* Data mapping specifies what data will be extracted from internal data warehouse, transformed, and sent to external entity.
* Executed data cleansing through standardizing customer data, verifying international addresses, identifying duplicate records, and applying data rules by integrating stages from IBM Quality Stage.
* Involved in understanding the scope of application, presenting schema, data model, and defining relationships within and between groups of data.
* Possessed extensive experience in Data Warehousing concepts with emphasis on ETL and life cycle development includes requirement analysis, design, development, testing, and implementation.
* Responsible for building SQL queries for Oracle to fetch the data from the Database as per the business requirement and use the SQL queries/function/packages in the Data Stage for further processing.
* Developing visual reports, dashboards and KPI scorecards using Power BI desktop.

*Skillset: Google Cloud Platform (BigQuery,Pub Sub,Cloud Build. GCS, Data Studio, Data Fusion & Data Prep), Python, PySpark,Azure,DBT,Kubernetes,Docker,Databricks, Datastage, ETL tools, Denodo,SQL Server, EPIC (Caboodle & Clarity) , Hyperspace, Power BI, Tableau*

IBM - AT &T**, Middletown, NJ**

*Senior ETL Developer, Oct 2014 - Jan 2017*

* Designed and created ETL code installations, aiding in transitions from one data warehouse to another.
* Designed integration tools to combine data from multiple, varied data sources such as RDBMS, SQL and big data installations.
* Responsible for interfacing with the AT&T technical team, application system engineers to gather requirements, conduct impact analysis, documenting the requirements and present the conclusions to the technical and business teams
* Well versed in components of Hadoop ecosystem – HDFS, MapReduce, HIVE, PIG, Sqoop & HBase by attending several key trainings in Big data & unstructured data handling
* Functioned as a single point of contact for both AT&T and IBM technical teams and the senior management.
* Responsible for helping in Building and data extraction and Transformation solutions in accordance with EBDP (Enterprise Bulk Data Processing) methodology adopted by AT&T - IBM.Responsible for Build or enhance applications/data services in support of the enterprise target architecture for Enterprise Bulk Data Processing by getting ordering, provisioning and Billing data from various platforms to a single view.Responsible for Client Support: Interfacing with the customers for day to day operational issues and resolution of the same.
* Integrated Power BI reports into other applications using embedded analytics like Power BI service (SaaS), or by API automation
* Responsible for mentoring, coaching the client's support teams and Provides assistance working with them to resolve user issues.
* Responsible for tracking the project with the processes and quality standards for both IBM (QMS- Quality Management System) and client AT&T(PRISM).

*Skillset: Python, Azure, Hadoop - Hortenworks, Hive, Pig, Flume, Sqoop, SQL Server,Power BI,Datastage, Informatica & Talend*

IBM, Global Business Services**, Chennai, India**

*ETL Developer, Apr 2011 - Oct 2014*

* Designed and created ETL code installations, aiding in transitions from one data warehouse to another.
* Expertise in application maintenance and support for this incentive and commission payable application.
* Interacted well with the global clients and importantly handled all the critical production issue during installation & monthly production cycle and the impact of it earning their trust to a greater extent.
* Responsible for leading the Off-shore Production team to trouble shoot the issues faced in day-to-day activities.
* Developed many innovative tools in Mainframes to reduce the manual efforts and repetitive works.
* End to end release planning for firm wide regulatory releases for fiscal year 2015 to reengineer all product reference data re design.
* Exceptional analytical and problem solving skills
* Team player with the ability to communicate effectively at all levels of the development process.
* Extensive experience in IBM InfoSphere 11.3, IBM InfoSphere 9.1, IBM InfoSphere 8.1 (DataStage, QualityStage, Information Analyzer, Metadata Workbench), IBM WebSphere 8.0 (DataStage, Information Analyzer, QualityStage).
* Served a vital role as an offshore lead and my versatility on Datastage & Mainframe giving flexibility for the project to go beyond the expectations of the customers, providing quality deliverables before the deadlines.
* Worked with File Transfer Protocol (FTP) and Secure File Transfer Protocol (SFTP) to pull or send the files from one server to another server.
* Designed adhoc job in datastage which helps to reconcile the achievement data of the eligible employees in a proper reporting format which runs each week to update the client team.

*Skillset: Informatica, Datastage, Unix Scripting, Python,DB2,SQL Server*

**Education**

**Western Governors University**

**Salt Lake City, UT**

Master of Science Data Analytics (Jan 2021)

**Anna University**

**Chennai, India**

Bachelor of Science Information Technology (Jan 2009)

**Additional Skills**

* Python & PySpark
* Scripting (Shell & Perl)
* Snowflake - Cloud Data Warehouse
* ETL Integration - Datastage & Informatica
* SQL (DB2, SQL Server & Oracle)
* Azure
* AWS
* Hadoop (Cloudera)
* IBM Mainframes
* Power BI
* Tableau
* Cognos

**Certifications**

* SnowPro Certification – Snowflake
* SAS