**Sai Jnanesh**

[**maxyblue1@gmail.com**](mailto:maxyblue1@gmail.com)

**+1 (913) 703-6119**

**Senior Data Engineer**

**Summary**

* Senior Data Engineer with over 9 years of IT experience in Data Application Developments
* Fluent programming experience with PowerShell, Scala, Java, Python and SQL.
* Hands-on experience in developing and deploying enterprise-based applications using major Hadoop ecosystem components like MapReduce, YARN, Hive, HBase, Flume, Sqoop, Spark SQL, Kafka.
* Experience on developing data pipelines and flow on Cloud services like Azure.
* Strong development skills with Azure Data Lake, Azure Data Factory, SQL Data Warehouse Azure Blob, Azure Storage Explorer.
* Architect and implement ETL and data movement solutions using Azure Data Factory (ADF), SSIS
* Worked on Azure SQL and SQL administering on different management standards.
* Extensive working experience on Saas and Iaas platforms.
* Expert in providing ETL solutions for any type of business model.
* Proficient with Spark Core, Spark SQL and Spark Streaming for processing and transforming complex data using in-memory computing capabilities written in Scala.
* Worked with Spark to improve efficiency of existing algorithms using Spark Context, Spark SQL, Data Frame, Pair RDD's and Spark YARN.
* Extensive experience in web application development using Python with Django Framework and web technologies using Object Oriented Programming concepts to create scalable and robust components, which can be used across the whole framework.
* Hands on experience in writing advance scripting like Python, PowerShell, Bash for automating process.
* Experience working on Databricks platform for extensive use of Spark framework.
* Develop Power BI reports & effective dashboards after gathering and translating end-user requirements.
* Experience in troubleshooting and resolving architecture problems including database and storage, network, security and applications.
* Experience managing Big Data platform deployed in Azure Cloud.
* Wrote Python scripts to parse XML documents and load the data in database.
* Datacenter Migration, Azure Data Services have a strong virtualization experience.
* Rich knowledge of generating drill through and drill down reports with drop down menu option, sorting the data, and defining subtotals in Power BI.
* Experience developing on - premise and Real Time processes.
* Expertise in design, analysis, and development of business intelligence applications using Microsoft SQL server, BI Stack and Power BI.
* Expert in generating on-demand and scheduled reports for Business Analysis or management decision using SQL Servers Reporting Services (SSRS) and Power BI.
* Establish, design, and support an enterprise Cloud SQL architecture that provides the required capacity, control and supports the long-term strategic goals for an enterprise database solution.
* Ensure compliance with server-specific architectural standards and implementation practices.
* Advice on cost efficiency for future usage and cost optimization for current infrastructure.

**Technical Skills**

**Tools:** SSMS, Power BI, MS Office (Excel, SharePoint, Visio), Adobe Analytics, Azure Data Factories, Azure Data Lake, Azure SQL Databases, Azure Management tools

**Version Control:**GIT

**Programming Languages:**.NET, C#, JSON, U-SQL

**Version Control:**SVN, CVS, TFS, GIT, Bit Bucket

**Platforms:**Windows, Linux

**Databases:**SQL Server, MySQL, Oracle

**Cloud:** Azure SQL, Azure PowerShell, DataLake

**Languages:**C, C++, Python, Scala, UNIX Shell Script, COBOL, SQL and PL/SQL

**CI/CD**: GitLab

**Tools:** Teradata SQL Assistant, PyCharm, Autosys

**Operating Systems:**Linux, Unix, ZOS and Windows

**Databases:**Teradata, Oracle, DB2, SQL Server, MySQL

**ETL Tools:**IBM Infosphere Information Server

**Reporting:**Tableau

**Professional Experience**

**Client: Mayo Clinic Rochester MN May 2021 to Present**

**Role: Sr. Data Engineer**

**Responsibilities:**

* Performs data analysis and design, and creates and maintains large, complex logical and physical data models, and metadata repositories using ERWIN and MB MDR
* I have written shell script to trigger data Stage jobs.
* Assist service developers in finding relevant content in the existing reference models.
* Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL.
* Created Tables, Stored Procedures, and extracted data using T-SQL for business users whenever required.
* Responsible for Design, Development, and testing of the database and Developed Stored Procedures, Views, and Triggers
* Developed Python-based API (RESTful Web Service) to track revenue and perform revenue analysis.
* Migrate data from traditional database systems to Azure databases.
* Design and implement migration strategies for traditional systems on Azure (Lift and shift/Azure Migrate, other third-party tools.
* Utilized Spark SQL API in Pyspark to extract and load data and perform SQL queries.
* Worked on developing Pyspark script to encrypting the raw data by using hashing algorithms concepts on client specified columns.
* Interacts with Business Analysts, Users, and SMEs on elaborating requirements.
* Design and implement end-to-end data solutions (storage, integration, processing, visualization) in Azure.
* Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL.
* Propose architectures considering cost/spend in Azure and develop recommendations to right-size data infrastructure.
* Writing modules in Python to connect to PostgreSQL and doing CRUD operations with PostgreSQL.
* Worked on selenium testing framework also JIRA was used to build an environment for development and unit Test Python library was used for testing many programs on python and other codes.
* Develop programs to automate the testing of controller in CI/CD environment using Python, Bash script, Git and Linux command line.
* Experience in DWH/BI project implementation using Azure Data Factory.
* Architect, design and validate Azure infrastructure-as-a-Service (IaaS) environment
* Extract Transform and Load data from Sources Systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL, and U-SQL Azure Data Lake Analytics.
* Developed complex reports with multiple datasets and sources in Power BI.
* Refactored and extended large existing Django/Python code base by maintaining PEP-8 code standards and fix the bugs.
* Developed python scripts in order to migrate information from Oracle to MongoDB.
* Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in In Azure Databricks.
* Recreating existing application logic and functionality in the Azure Data Lake, Data Factory, SQL Database and SQL Datawarehouse environment. experience in DWH/BI project implementation using Azure DF and databricks.
* Deployed the Power BI reports to Power BI services and pinned visuals to make dashboards out of it.
* Implemented security and shared reports among different users with features such as content packs and workspace in Power BI service.
* Incorporated hierarchies, measures, calculated columns, parameters, functions feature in Power BI Desktop.
* Utilized Power Query in Power BI to Pivot & Un-pivot the data model for data cleansing and data massaging.

**Environment:** MS Azure, SQL Server Management Studio, Visual Studio, ETL, SSIS, SSAS, Advance Excel, Databricks, Spark, Python, Power BI, Hive/Hadoop, Snowflake

**Client: Fifth Third Bank, Evansville, IN March 2019 to April 2021**

**Role: Data Engineer**

**Responsibilities:**

* Designed and implemented configurable data delivery pipeline for scheduled updates to customer facing data stores built with Python
* Proficient in Machine Learning techniques (Decision Trees, Linear/Logistic Regressors) and Statistical Modeling
* Design & implement migration strategies for traditional systems on Azure (Lift and shift/Azure Migrate, other third-party tools).
* Performed Data Migration to GCP. Worked on confluence and Jira
* Implemented a Continuous Delivery pipeline with Docker and Git Hub
* Responsible for data services and data movement infrastructures
* Experienced in ETL concepts, building ETL solutions and Data modeling
* Experience in building and architecting multiple Data pipelines, end to end ETL and ELT process for Data ingestion and transformation in GCP
* Implemented security and shared reports among different users with features such as content packs and workspace in Power BI service.
* Configured Power BI gateway to keep dashboards and reports up to date with on premise data sources.
* Implement ad-hoc analysis solutions using Azure Data Lake Analytics/Store, HDInsight.
* Used various sources to pull data into Power BI such as SQL Server, Excel, Oracle, SQL Azure, etc.
* Collaborate with application architects on infrastructure as a service (IaaS) application to Platform as a Service (PaaS).
* Aggregated daily sales team updates to send report to executives and to organize jobs running on Spark clusters
* Used various sources to pull data into Power BI such as SQL Server, Excel, Oracle, SQL Azure, etc.
* Refined the data model utilizing Power BI model with help of DAX expression.
* Migrate data from traditional database systems to Azure databases.
* Build Complex distributed systems involving huge amounts of data handling, collecting metrics building data pipeline, and Analytics.
* Worked on architecting the ETL transformation layers and writing spark jobs to do the processing.
* Participated in the full software development lifecycle with requirements, solution design, development, QA implementation, and product support using Scrum and other Agile methodologies
* Identify potential problems and recommend alternative technical solutions.
* Participating in Technical Architecture Documents, Project Design, and Implementation discussions.
* Design and implement end-to-end data solutions (storage, integration, processing, visualization) in Azure.
* Built performant, scalable ETL processes to load, cleanse and validate data
* Implemented calculated columns and measures using DAX in Power BI.
* Consumed data in Power BI and created insights with help of Power BI features.
* Utilized sync slicers, conditional formatting and bookmarks to enhance reports in Power BI Desktop.

**Environment:** SSMS, Visual Studio, VSTS, Azure SQL, Azure Storage Explorer, Big query, Apache Beam, Cloud Dataflow, Cloud Shell, Cloud SQL, MySQL, Posgres, SQL Server, Python, Scala, Spark, Hive, Power BI

**Client: Early Warning's, Scottsdale, AZ January 2017 to February 2019**

**Role: Data Engineer**

**Responsibilities:**

* Monitor communications performance using visual, diagnostic equipment, status indicator checking methods, etc., to locate problems
* Manage the lifecycle for technical support documentation including Standard Operating Procedures and work instructions.
* Analyze and cleanse raw data using HiveQL
* Respond to Local Area Network (LAN) and Wide Area Network (WAN) user requests for system upgrades and changes.
* Perform software installations and upgrades to operating systems and layered software packages.
* Created Hive tables as per requirements, internal or external tables defined with appropriate static and dynamic partitions, intended for efficiency.
* Experience in fact dimensional modeling (Star schema, Snowflake schema), transactional modeling and SCD (Slowly changing dimension)
* Devised PL/SQL Stored Procedures, Functions, Triggers, Views and packages. Made use of Indexing, Aggregation and Materialized views to optimize query performance.
* Experience in data transformations using Map-Reduce, HIVE for different file formats.
* Involved in converting Hive/SQL queries into transformations using Python
* Performed complex joins on tables in hive with various optimization techniques
* Hands of experience in GCP, Big Query, GCS bucket, G - cloud function, cloud dataflow, Pub/suB cloud shell, GSUTIL, BQ command line utilities, Data Proc, Stack driver
* Implemented Apache Airflow for authoring, scheduling and monitoring Data Pipelines
* Proficient in Machine Learning techniques (Decision Trees, Linear/Logistic Regressors) and Statistical Modeling
* Schedule installations and upgrades and maintain them in accordance with established IT policies and procedures.
* Developed logistic regression models (Python) to predict subscription response rate based on customers variables like past transactions, response to prior mailings, promotions, demographics, interests, and hobbies, etc.
* Develop near real time data pipeline using spark
* Process and load bound and unbound Data from Google pub/sub topic to Bigquery using cloud Dataflow with Python
* Set up, configure, and troubleshoot AV collaboration systems and software used by staff.
* Ensure data/media recoverability by implementing a schedule of system backups and database archive operations. Implement and promote standard operating procedures.
* Use network management tools such as the Cisco Identity Services Engine (ISE) and Cisco Prime to identify, troubleshoot, and resolve wired and wireless networking issues on LAN equipment and/or PCs and other end-user devices.
* Track and report issues in our ticketing system.
* Participate in statewide projects to plan and implement system operation, optimization, enhancements, and upgrades.
* Write SOQL queries, create test data in salesforce for informatica cloud mappings unit testing.
* Administer data backup systems at remote offices to ensure data availability, security, and recoverability. Quickly restore data accidentally deleted by customers.
* Develop Informatica cloud real time processes (ICRT). Work with WSDL, SOAP UI for APIs
* Prepare TDDs, Test Case documents after each process has been developed.

**Environment:** SQL Server, Windows Server, Windows 2008 R2, /2000 Server(64bit), Windows 2000 Advanced Server, Windows XP, Unix, LINUX, Shell Script. Hadoop, HDFS, Hive

**Client: Avon Technologies Pvt Ltd Hyd India July 2015 October 2016**

**Role: Data Engineer**

**Responsibilities:**

* Performed ETL using Azure Data Bricks. Migrated on-premise Oracle ETL process to Azure Synapse Analytics.
* Worked on python scripting to automate generation of scripts. Data curation done using azure data bricks.
* Used Azure Data Lake as Source and pulled data using Azure Polybasic.
* Used stored procedures, lookup, execute pipeline, data flow, copy data, azure function features in ADF.
* Developed mapping document to map columns from source to target.
* Created azure data factory (ADF pipelines) using Azure polybase and Azure blob.
* Extensively used T-SQL in constructing user profiles, relational database models, data dictionaries, and data integrity.
* Worked with PowerShell and UNIX scripts for file transfer, emailing and other file related tasks.
* Experienced in azure Data Lake storage gen2 to store excel files, parquet files and retrieve using data using Blob API.
* Used azure data bricks to do ETL (workspace, hive, rdd, map, reduce so on).
* Tuned performance of Informatica mappings and sessions for improving the process and making it efficient after eliminating bottlenecks.
* Worked on complex SQL Queries, PL/SQL procedures and convert them to ETL tasks
* Worked as a developer in creating complex Stored Procedures, Triggers, Functions, Indexes, Tables, Views and other T-SQL code and SQL joins for applications.
* Prepared capacity and architecture plan to create the Azure Cloud environment to host migrated IaaS VMs and PaaS role instances for refactored applications and databases.
* Installing the Applications Insights tool on the web services and configuring an Application Insight workspace in Azure. Configured Application Insights to perform web tests and alerts.
* Handle deployment workflow including provisioning, load balancing, and health monitoring.
* Worked on creating star schema for drilling data. Created Pyspark procedures, functions, packages to load data.
* Conducted capacity planning and architecture of Storage Accounts.
* Deployed DSC from Azure Automation to On-premises or cloud environments.
* Configured Azure Traffic Manager to build routing for user traffic Infrastructure Migrations: Drive Operational efforts to migrate all legacy services to a fully Virtualized Infrastructure.
* Configure continuous integration from source control, setting up build definition within Visual Studio Team Services (VSTS) and configure continuous delivery to automate the deployment of ASP.NET MVC applications to Azure web apps
* Implemented HA deployment models with Azure Classic and Azure Resource Manager
* Configured Azure Active Directory and managed users and groups.

**Environment:** SQL Server, Azure, Pyspark, ETL, PowerShell, UNIX, ASP.NET MVC, Visual Studio, T-SQL

**Client: Careator Technologies Pvt Ltd Hyderabad, India May 2014 to June 2015**

**Role: Data Engineer**

**Responsibilities:**

* Performed Data analysis for building Reporting Data Mart.
* Worked with Reporting developers to oversee the implementation of report/universe designs.
* Worked with deployments from Dev to UAT, and then to Prod.
* Built reporting data warehouse from ERP system using Order Management, Invoice & Service contracts modules.
* Acted as SME for Data Warehouse related processes.
* Automated the application to run on Alteryx and QlikView servers.
* Worked with many up the stream and downstream systems and many customers around the world in interfacing various systems for data extractions, ETL, Analytics and Reporting needs.
* Applied the statistical data cleansing model to the production jobs using Alteryx.
* Identified various data sources, formulated data dictionaries, design and develop data models (physical/logical data models) based on the given specs and requirements.
* Applied statistical methods to clean up bad data /outliers and designed a model to alert on outliers
* Designed and developed multiple metrics using 3 sigma control charts, P-Charts and Box plots in QlikView with dynamic subgrouping to analyze the shift in the sensor data.
* Applied the statistical data cleansing model to the production jobs using Alteryx.
* Read semi-structured sensor data to analyze the quality of the feed
* Applied statistical methods to clean up bad data /outliers and designed a model to alert on outliers
* Designed and developed multiple metrics using 3 sigma control charts, P-Charts and Box plots in QlikView with dynamic subgrouping to analyze the shift in the sensor data.
* Worked on SQL Server Integration Services (SSIS) to integrate and analyze data from multiple heterogeneous information sources.
* Developed custom reports using SQL Server Reporting Services (SSRS) to review score cards, business trends based on the data from different Business or market locations.

**Environment:** SQL Server, SSIS, SSRS, Alteryx, QlikView, ERP, SQL Server