**Venkat Koushik Thota**

[**koushik5586@gmail.com**](mailto:koushik5586@gmail.com)

**+91-9295096950**

**Profile Summary:**

Accomplished **Data Engineer** with 2 Years of focused experience in Insurance domain, with a strong background in **Data warehousing** and **ETL** experience using Informatica Power Center and SQL. Well versed in shell scripting (UNIX), Python and **AWS** services (Solutions Architect-Associate certified) along with good knowledge in Bigdata technologies.

**Professional Summary:**

* Proficient knowledge and hands-on experience in building Data Warehouses, Data Marts, Data Integration using Amazon Redshift, Redshift spectrum and S3 as the data store and ETL processes using Informatica Power Center.
* Well versed in using Bash and python scripts along with AWS CLI to automate several functionalities in S3, Lambda and AWS Glue which help in cutting down time and effort in client deliverables.
* Knowledgeable in the entire AWS stack with hands on experience in core services like EC2, S3, Redshift, Lambda, CloudTrail, Athena, RDS, ASG, VPC, Glue.
* Good understanding of working in Redshift Keys, design components, copy activity, spectrum etc.
* Good Knowledge on Data Warehousing concepts like Star Schema, Snowflake Schema, and built data warehouse essential Dimensions and Fact tables using Redshift from different source databases like Oracle and PostgreSQL
* Possess good knowledge in Bigdata technologies – HDFS, Hadoop, Splunk, spark, mongo DB and its AWS counter parts Kinesis and EMR.
* Possess good domain knowledge in insurance domain and helped translating the business requirements into informatica mappings
* Involved in Phases like Requirement gathering, feasibility analysis, code build, code commit on shared repositories (SVN), Unit Test Case preparation and debugging defects with the accurate root cause analysis.
* Possess good communication skills with day-to-day involvement in Client meetings to discuss business requirements.
* Well versed with entire flow of SDLC (Software Development Life Cycle) following Agile Methodology.

**Experience Summary:**

Working as Data Engineer (Associate Consultant) for **Capgemini**, Hyderabad from May 2018 to till date.

**Technical Skills:**

|  |  |
| --- | --- |
| **Cloud Environment** | AWS- EC2, S3, Redshift, Spectrum, Cloud trail, Athena, Lambda ,VPC,ASG |
| **Programming Languages** | Bash, C, Java, Python |
| **Web Development Languages** | HTML,CSS5,XML |
|  |  |
| **Developer Tools** | Informatica Power Center, Teradata, Cognos, Jira, Confluence, Jenkins |
|  |  |
| **Domain Knowledge** | Insurance |
|  |  |
| **Databases** | Oracle, PostgreSQL |
|  |  |
| **Operating Systems** | Windows 7, Windows 8 Windows 10, UNIX, Mac OS |
|  |  |

**Certifications:**

* AWS Certified Solutions Architect - Associate (Issued December 2020)
* Oracle Database SQL Certified Associate (Issued August 2018)

**Summary of Works Involved:**

**Direct Line Group Data (DLG)**

|  |  |  |
| --- | --- | --- |
| Project | : Best4Customer | |
| Client | : Direct Line Group Insurance ,UK | |
| Domain | : | Insurance |
| Location | : | Hyderabad. |
| Role | : | ETL developer/Data Engineer |

**Description:**

Direct Line group is an insurance company based in the United Kingdom formed in 2012. The Company owns a number of subsidiaries, providing various insurance products, including Direct line and Churchill, as well as the roadside and vehicle recovery provider green flag. The Best4Customer program is one of the prestigious account for Capgemini aimed at migrating of various Mainframes systems into single Data Lake entity on AWS cloud using different technologies like, Guidewire, Mulesoft, and ETL etc. And is also about generating reports of various policy holders, their quotes and every aspect related to that policy depending on what type of an insurance policy it is like motor, rescue, home, travel, etc. by extracting data from guidewire source, process for cleansing, reformat, validating data and loading it into tables/files for reports to be generated according to the business requirement.

**Responsibilities:**

* Developed mappings that perform Extraction, Transformation and load of source data into Derived Masters schema using various power center transformations like Source Qualifier, Sequence generator, look up, Rank, Joiner, Expression, Stored Procedure, SQL, Normalizer and update strategy to meet business logic in the mappings in Informatica
* Developed slowly changing dimensions to maintain historical as well as incremental data using Type I, Type II and Type III strategies.
* Responsible for Code build, Code merges, CICD, Code migration across different environments along with Performance tuning in SQL queries and application performance in NFT environments and handling release management to production for 3 consecutive releases successfully
* Used AWS CLI commands and SDK’s to transfer S3 Files from Informatica server to S3 bucket.
* Automated s3 file deletion across multiple buckets and prefixes using AWS CLI and lifecycle policies in S3.
* Created automated shell and python scripts to Load and unload data from S3 to Redshift.
* Writing various data loading scripts, SFTP scripts, Backup/S3 transfer scripts in lambda and Glue.
* Defining data flows from Informatica, S3 and Redshift.
* Worked on Oracle Triggers to Sync data between to maintain historical data copy of the source database.
* Worked on table designing with Data distribution and sorting using DISTKEY’s, DISTSTYLE’s, SORTKEY’s.
* Created Redshift spectrum external tables based on the requirement.
* Own the responsibility of preparing vivid unit test cases covering various business scenarios, projecting the efficiency of code quality delivering the given business requirements.
* Adhere to Agile Methodology of Development that involves my presence in daily standup/scrum meetings, sprint discussions and sprint retrospection.
* Possessed good debugging and analytical skills by giving accurate pointers from performed root cause analysis.
* Proficient in handling defects that come up from different environment simultaneously with efficient tracking and solutions.
* Developed a unique automation that would ease the day-to-day activity of unit test case execution that considerably reduced manual efforts.
* Involved in different phases of detailed requirement gathering from Clients progressing from analysis to implementation.

**Projects:**

1. **Data Lake Solution:**

Involved Building a data lake solution to park all the insurance quotes, policies, claims and Billing details of users from Guidewire oracle tables to a robust and durable storage Amazon S3 by performing required transformations as per requirement using Informatica Power Center. Used Athena and Redshift spectrum to perform in the place analytics wherever required.

1. **Third Party Feeds:**

Built various feeds which involve processing ingress data from third party vendors who provide valuable information to the insurance providers regarding various metrics which decide the insurance premium for a policy holder. This involved using Informatica Power Center, AWS lambda to handle in S3 side wherever applicable to process the ingress files/table data and send egress files back to the vendor via SFTP with various bash and python scripts or else store in S3 or Redshift for analytics purposes.

1. **Datawarehouse Solution:**

Built Petabyte scale Data warehouse containing multiple CDC tables abiding to snowflake schema having a collection of dimension and fact tables with SCD type 1 and type 2 configurations in Amazon Redshift and PostgreSQL using Informatica PowerCenter, bash scripts and AWS Glue for Processing S3 files and developed stored procedures in redshift for data retention in redshift tables. All these tables are used for analytics by downstream teams.

1. **Data Analytics Lab (POC):**

Built robust analytics system to process real time data flowing from clickstreams via Apache Kafka and processing them using AWS Lambda (Python) and storing them in DynamoDB for future analytics.

**Academic Qualification:**

* **Bachelor of Technology (Electronics and Communication Engineering)**

Velagapudi Ramakrishna Siddhartha Engineering College – Kanuru, Vijayawada

Aggregate: 8.71 CGPA

* **XII- Board of Secondary Education (MPC)**

Sri Chaitanya Junior College – Gayatri Nagar, Vijayawada

Aggregate: 95%

* **Class X(CBSE)**

V.P.Siddhartha Public School – Moghalrajpuram, Vijayawada

Aggregate: 10 CGPA

**Achievements:**

Had been awarded with “**Wall of Fame**” twice (award given for exceptional performance for the month of January 2020 and July 2020)

**Personal Details**

Father’s Name : Prakash Thota

D.O.B : 22nd Jan, 1997

Marital Status : Single

Languages Known : English, Telugu, and Hindi

Residing City : Hyderabad, Telangana.