Shubham Khillari

Email : shubhamkhillari2000@gmail.com | **Mobile**: (+91)9975177011 | **City**: Pune (**MH**)

Linkedin: https://www.linkedin.com/in/shubham-khillari-440945270

Summary

I am an experienced IT professional with 3+ years of expertise in ETL processes, including data transformation, sourcing, mapping, conversion, and loading. I have extensive experience in writing and executing SQL queries and working with databases like SQL Server and BigQuery. My knowledge extends to various ETL transformations and Informatica mappings, along with hands-on experience in creating and managing ADF/Synapse Pipelines. I am   skilled in executing and monitoring Control-M jobs and have a fundamental understanding of UNIX commands and shell scripting. Additionally, I have experience in data quality frameworks to ensure accuracy, completeness, and consistency of data, as well as in data analysis, validation, cleansing, verification, and identifying mismatches. With   a strong focus on delivering high-quality code with minimal defects, I also possess excellent communication and documentation skills.

Work History

Trainee Programmer | Bitwise Solutions Pvt. Ltd.| Pune| *2022-05-01 To 2023-05-01*

- Requirement Analysis: Reviewed and analyzed user stories and UI specification documents to understand business and data requirements. Ensured all functional and non-functional requirements were accounted for before development.
- ETL Development: Designed and developed ETL workflows, including writing configuration entries for data extraction and loading processes for each table. Ensured efficient data movement between source and target systems.
- SQL Query Development: Created and optimized SQL queries for data extraction, transformation, and validation to maintain data integrity and improve performance.
- Data Extraction & Loading: Leveraged generic extraction scripts to retrieve data from the source system and used generic load scripts to load data into the target system, ensuring consistency and efficiency.
- Data Validation & Unit Testing: Conducted thorough data validation and unit testing to verify data accuracy and completeness. Ensured extracted and loaded data matched business expectations.
- QA Deployment & Testing: Deployed ETL workflows in the QA environment, monitored execution, and ensured no defects or issues were present before moving to production.
- Production Readiness & Deployment: After successful QA testing, prepared production readiness documentation and coordinated production deployment. Ensured a smooth transition with minimal downtime.
- Control-M Job Scheduling: Created and scheduled Control-M jobs for automated execution and monitoring of ETL workflows in production.
- Successful Production Deployment: Ensured successful deployment in the production environment with no defects, providing a stable and reliable data pipeline for business operations.

Programmer| Bitwise Solutions Pvt. Ltd.| Pune| *2023-05-05 To 2024-03-01*

- Requirement Analysis & Planning
  Analyzed user stories and user interface specification documents to understand business requirements and data processing needs.
  Collaborated with stakeholders to clarify expectations and gather necessary inputs for ETL development.
- Environment Setup & Configuration

Set up and configured the required environment, including shell configurations and system dependencies, ensuring a seamless execution environment.

Established necessary configurations for ETL workflows, including writing environment-specific configuration entries.

ETL Development & Data Processing

Designed and developed ETL workflows for data extraction, transformation, and loading while ensuring efficiency and scalability.

Created Data Quality (DQ) configuration entries to enforce validation rules and ensure data integrity.

Implemented the generic script to generate two target files as per project requirements.

Ensured all configurations and processing logic adhered to project standards and business rules.

- Data Quality Validation & Testing

    Conducted thorough validation of the generated DQ entries to ensure completeness and correctness.

    Performed rigorous data validation and unit testing to verify the accuracy of transformed data.

    Documented unit testing results, detailing scenarios, expected outcomes, and actual results for review and traceability.

- File Transfer & Automation

    Developed and configured SFTP commands to facilitate the secure transfer of two target files to the designated locations.

    Ensured proper logging and monitoring mechanisms for tracking file transfers and error handling.

- QA Deployment & Quality Assurance

    Successfully deployed the ETL solution in the QA environment.

    Conducted end-to-end testing, ensuring data accuracy, workflow integrity, and system performance.

    Verified that no defects or issues were found in the QA environment before proceeding to production readiness.

- Production Readiness & Deployment

    Created a comprehensive Production Readiness document outlining deployment steps, rollback plans, and operational guidelines.

    Received QA sign-off after successful testing and validation.

Programmer | Bitwise Solutions Pvt. Ltd. | Pune | *2024-03-10 To Current*

- Analysis of Informatica Mappings & Conversion to PySpark

    Conducted an in-depth analysis of existing Informatica mappings, transformation logic, and data flow to understand the business rules.

    Translated Informatica workflows into PySpark scripts, ensuring that all transformation logic and dependencies are accurately replicated.

    Optimized PySpark code for performance, scalability, and compatibility with Azure Synapse.

- Parameterization & Configuration Setup

    Designed and implemented parameterized ETL scripts to enable flexible configurations for different environments (DEV, QA, and PROD).

    Ensured dynamic control over source-target connections, file paths, and transformation parameters to enhance reusability   and maintainability.

- Data Extraction, Transformation & Loading (ETL)

    Extracted data from various sources, applying necessary transformations using PySpark before loading it into Azure Synapse.

    Designed scalable and efficient ingestion pipelines to handle structured and semi-structured data.

    Ensured data was loaded as per business logic, maintaining data integrity and consistency.

- Data Validation & Verification

    Implemented rigorous validation mechanisms to verify that data is transformed and loaded as expected.

    Ensured compliance with client requirements by validating data accuracy, completeness, and consistency.

    Compared source and target data to identify and resolve any discrepancies before proceeding to production.

- Ingestion Process into Azure Synapse

    Leveraged Azure Data Factory (ADF) and PySpark to orchestrate the ingestion of transformed data into Synapse.

    Ensured data partitioning and indexing strategies were optimized for query performance.

Implemented logging and monitoring mechanisms for tracking ingestion progress and identifying errors.
- Data Quality & Unit Testing
Conducted thorough unit testing to validate each step of the ETL pipeline, ensuring correctness and efficiency.
Verified business rules, transformations, and aggregations to match client expectations.
Created test cases to handle various data scenarios, including edge cases and error handling.

Skills & abilities

- **GCP-** Big Query, Dataflow, Composer, Cloud Functions, Cloud SQL, Bigtable, Shell Scripting

- **Azure-** Azure Databricks, Azure Synapse, SQL, Dataflow, Pipelines

- **Languages:** Python, Pyspark, SQL, Unix.

- **Tools:** Apache Airflow, Eclipse, VSTS, Jenkins, Putty Terminal, ADF/Synapse

Education

Course	SPECIALIZATION	Institute	Board	Score	Year
UG	B.SC (Computer Science)	ATSS CBSCA Collage Pimpri Chinchwad Pune	Savitribai Phule Pune University	74.21%	2021
Class XII	Science	Dnyan Mandir Junior Collage Ale, Junnar, Pune.	Maharashtra State Board	60.61%	2018
Class X	General	Shri Sadaguru Sitaram Maharaj Vidyalaya Pimpri Pendhar.	Maharashtra State Board	83.00%	2016

Awards & Recognition

Client Appreciation (2022-2023)

Client appreciation for defect free delivery, passion, collaboration, ownership of modules.

Client Appreciation (2022-2023)

Client appreciation for ownership, problem-solving ability, dedication, flexibility, and speedy responses when dealing with changing requirements, and handling multiple production Issues.

Client Appreciation (2023-2024)

Received Client Appreciation for Solving Medium and Complex Mapping with In Particular Time.