

SHEKHAR SUMAN
+91-9970589500
[**Simplyshekhar29@gmail.com**](mailto:Simplyshekhar29@gmail.com)

OBJECTIVE

A technically strong professional and a team player having strong technical skills looking for a more responsible and challenging position in the field of data analytics.

PROFESSIONAL SUMMARY

- **11 years** of IT experience in Data Analytics and Business Intelligence.
- **5 years** of experience in Big Data, AWS, Redshift, Python, Hadoop and Various Hadoop ecosystems like-Pig, Hive, Sqoop and Spark.
- **Over 9 years** of experience in ETL tools like Informatica power centre, Matillion, Streamsets, oracle application development using ORACLE, RDBMS, SQL, PL/SQL and Reporting tool such as SAP BO.
- Expertise in exploratory data analysis, data preprocessing and data modeling.
- Expertise in Big Data Analytics with Spark and Python.
- Experienced in formulating and refining business objectives, data selection, data preparation and model evaluation.
- Experienced in data mining on massive datasets by writing ad-hoc Python scripts, Hive and Pig queries using Hadoop framework.
- Experienced in big data analysis, move data between relational databases and Hadoop using SQOOP, manage data in HDFS, use Pig and Hive to run distributed queries on data.
- Expertise in Hadoop Administration and Big Data tools setup in distributed settings on AWS.
- Sound understanding of Hadoop Map Reduce and YARN framework and interaction among Big Data tools for data sharing and data mining.
- Proficient in Star and snow flake Schemas, Extract, Transform and Load of Data, Requirement Analysis, Design, Development, Testing, Documentation, Implementation of Business Applications and repository management and implementation.
- Expertise in Data Warehouse Concepts. Experience in Informatica Performance tuning of Target, Sources, Transformations and Sessions.
- Experience on Informatica 8.x/9.x. Developed good number of reusable Queries, Maplets and Sessions which are used across our Practice, involved in full life cycle development of Data Warehousing.
- Experience on Informatica migration projects.
- Proficient in Agile methodology.
- Conducted trainings on Hadoop and Big Data framework for different IDU's through training centre in the organization.
- Conducted trainings on Informatica for different IDU's through training centre in the organization.
- Good knowledge and experience in Databases ORACLE ,SQL and PL/SQL
- Exposure to UNIX shell script.
- Ability to communicate with non-technical staff and customers, translates their needs into appropriate systems, processes, and reports.
- Implemented Incremental Loads/Change Data Capture and Slowly Changing Dimensions.
- Possess excellent communication and interpersonal skills and ability to work in a team as well as individually.

EMPLOYMENT RECORD

Position held : Manager
Employer : Axtria India Pvt. Ltd.
Duration : November 2017– Present
Position held : Assistant Manager
Employer : Volkswagen IT Services India
Duration : January 2015 – November 2017
Position Held : ETL DATA Specialist-Informatica Developer
Employer : IBM Pvt.Ltd, Pune, India
Duration : Aug 2013-January 2015
Position held : Business Analyst
Employer : Tech Mahindra Ltd, Pune, India
Duration : June 2010 – Jul 2013

ACADEMIC QUALIFICATIONS

- **PGDM** (e-Business), 2016, Welingkar-Mumbai.
- **B.E.** (ELECTRONICS & COMMUNICATION), 2005-09, MIET-Gondia.
- **H.S.C.** C.B.S.E ,2004
- **S.S.C.** C.B.S.E,2002

TECHNICAL SKILLS

Data Analytics	-Big Data, Hadoop, Pig, Hive, Sqoop, Spark and Python
ETL Tools	- Informatica 8.x/9.x,Matillion, Streamsets
Reporting Tools	-SAP BO
Data Base	- Teradata, Redshift, RDBMS, Oracle (SQL, PL/SQL) - 10g/12c
Tools	- Control M, OPC, Autosys, Toad and PL/SQL Developer
Operating Systems	- All Windows OS and LINUX
Packages	- MS Office
Scripting language	- UNIX, Python

EXPERIENCE IN THE AREAS OF APPLICATION DEVELOPMENT

Project #1:

Duration : November 2017 – Present.
Title : Merck Data Migration
Client : Merck
Role : Manager
Software : AWS, Matillion, Streamsets, Redshift, Sqoop, Spark, Python, Informatica Power Centre 9.6.1, EMR,EC2,Athena, Teradata

Description:

Merck is world's leading pharmaceutical company which has wide variety of products in Health and Life-Sciences domain across the globe.

Existing data analytics system was built over on premise Teradata database and Informatica as ETL tool which was taking around 2-3 days for the weekly batch job to get completed. Objective of this data migration project was to build a cost-effective, scalable, easy to manage and efficient solution that can perform above operations in less than 8 hours by migrating existing DWH to AWS cloud.

Next generation system after migration uses AWS cloud components like EMR, Spark, Redshift, glue. Entire Informatica ETL logic is implemented in Matillion and PySpark. Athena is used for interactive query service that makes it easy to analyze data in Amazon S3 using standard SQL.

Following activities were carried out:

- Detailed analysis of system mappings to understand the data flows.
- Understanding the gaps in data across multiple systems – thorough interaction with Accounts Team for resolution of queries.
- Standardizing and de-duplicating the data in a single customer instance in the repository in the Test environment and getting approval from Accounts for the same.

Responsibilities:

- Understanding the requirements of client.
- Involved in requirement gathering and initial analysis.
- Involved in preparation of the Statement of work (SOW).
- Worked on architecture design and technology selection.
- Preparation of TDDs (Technical Design Documents).
- Configuration of AWS services, including Big-Data cluster (EMR).
- Leading the development of ETL processes in PySpark for the EDW.
- Performed PySpark job tuning using spark UI and logs, also tuned the EMR cluster.
- Preparing Unit Test Case document.
- Mine massive datasets in HDFS using Spark and Python.
- Develop code to clean, transform data (ETL) using Spark and Python and use Athena for querying data
- Create ETL process using AWS Matillion.
- Partner with cross site Operations leaders to identify key problem and provide solution through Analytics
- Contribute to data mining, reporting, and data analysis methodologies.
- Work with business domain experts and application developers to identify data relevant for analysis.
- Analyzed the sources, targets, transformed the data, mapped the data and loaded the data into the data warehouse.
- Performed the Tasks of Validating, Executing and scheduling the tasks.
- Involved in preparation of Low Level Design Documents (LLD).

Project #2:

Duration	: March 2016 – November 2017.
Title	: LOAD
Client	: Volkswagen
Role	: Big Data Developer/Informatica Lead
Software	: Informatica Power Centre 9.1, Workflow Manager and Workflow Monitor, Oracle 11g SQL and PL/SQL

Description:

LOAD is a Data Warehouse for logistic evaluations.

With provision of a uniform logistic data- and information-basis, using the standard software Business Objects, the user will be able to carry out logistic evaluations even without having programming knowledge. This project is successfully implemented in 49 countries.

Following activities were carried out:

- Detailed analysis of system mappings to understand the data flows.
- Understanding the gaps in data across multiple systems – thorough interaction with Accounts Team for resolution of queries.
- Standardizing and de-duplicating the data in a single customer instance in the repository in the Test environment and getting approval from Accounts for the same.

Responsibilities:

- Understanding the requirements of client.
- Analyzed the sources, targets, transformed the data, mapped the data and loaded the data into the data warehouse.
- Understanding end user requirements.
- Extensively used Informatica Power center for loading the data from various sources involving flat files and oracle tables to Data warehouse.
- Have good experience in understanding & creating source and target definitions, Mappings, Operators.
- Worked on Informatica tool -Source Analyzer, Data warehousing designer, Mapping Designer & Maplets, and Transformations.
- Created Informatica mappings to build data extract logic to load data. Most of the transformations were used like the Source qualifier, Expressions, joiners, Aggregators, lookups, Filters & Sequence Generators.
- Created sessions/workflows to run the mappings and set the session parameters to improve the load performance.
- Performed the Tasks of Validating, Executing and scheduling the tasks.
- ETL exercise in the development and Testing Environments.
- Responsible for unit testing activities.
- Involved in preparation of Low Level Design Documents (LLD).

Project #2:

Duration : Jan 2015 – Feb 2016.

Title : **KKP (Konzern Kennzahlen Production)**

Client : **Volkswagen**

Role : Informatica Developer

Software : Informatica Power Centre 9.1, Workflow Manager and Workflow Monitor, Oracle 11g SQL and PL/SQL

Description:

KKP (Konzern Kennzahlen Production) is a performance measurement & reporting system of production in all plants of the Group for strategic management team. Reports gets generated for logistics, finance, quality assurance, HR and spent hours in the form of traffic lights and trend arrows. KKP provided the overview of all departments in the form of key figures. The key figures are classified into "Top Key Figures" and "Supporting Key Figures" from different business areas.

Following activities were carried out:

- Travelled to Wolfsburg, Germany for project transition and requirement gathering.
- Detailed analysis of system mappings to understand the data flows.
- Monitoring the daily, weekly and monthly workflows for YPOD and resolving issues if any.
- Understanding the gaps in data across multiple systems – thorough interaction with Accounts Team for resolution of queries.
- Standardizing and de-duplicating the data in a single customer instance in the repository in the Test environment and getting approval from Accounts for the same.

Responsibilities:

- Understanding the requirements of client.
- Analyzed the sources, targets, transformed the data, mapped the data and loaded the data into the data warehouse.
- Understanding end user requirements.
- Extensively used Informatica Power center for loading the data from various sources involving flat files and oracle tables to Data warehouse.
- Have good experience in understanding & creating source and target definitions, Mappings, Operators.
- Worked on Informatica tool -Source Analyzer, Data warehousing designer, Mapping Designer & Maplets, and Transformations.
- Created Informatica mappings to build data extract logic to load data. Most of the transformations were used like the Source qualifier, Expressions, joiners, Aggregators, lookups, Filters & Sequence Generators.
- Created sessions/workflows to run the mappings and set the session parameters to improve the load performance.
- Performed the Tasks of Validating, Executing and scheduling the tasks.
- ETL exercise in the development and Testing Environments.
- Responsible for unit testing activities.
- Involved in preparation of Low Level Design Documents (LLD).

PERSONAL PROFILE

Date of Birth	:	24 th September 1986
Languages Known	:	English, Hindi and Bengali
Passport Number	:	T5255555
Current Location	:	Gurgoan
Alternate Number	:	+91-9595554615

Shekhar Suman.