****

**Poonam Malik**

|  |  |
| --- | --- |
|  | **PROFESSIONAL SUMMARY** |

* 9 years of IT experience in various ETL tools (DataStage, OWB, SSIS), Reporting Tools, Big Data Technologies (Hadoop).
* Extensive experience in Designing ETL workflows, Development, Enhancement, Maintenance, Production support.
* Extensive knowledge in Software Development Life Cycle (SDLC) involving Requirements Gathering, Design Analysis, Development, Maintenance, Implementation and Testing.
* Experienced in creating Databases, Tables, Stored Procedure, DDL/DML, Triggers, Views, Cursors and Indexes, Derived Tables using SQL programming to facilitate efficient data manipulation and data consistency.
* Good Experience in working with UNIX, SQL and various scheduling tools like Autosys and Tivoli.
* Good knowledge in Data Warehouse concepts and ETL process.
* Experience in working with Fast Track, DataStage, and Information Analyzer and Quality stage.
* Practical understanding of the Data modeling concepts like Star-Schema Modeling, Snowflake Schema Modeling, Fact and Dimension tables.
* Design and Developing UNIX Shell scripting for loading, running the mappings and validating the data in the data warehousing system
* Experience in data mining and data modelling for collecting, cleansing and analyzing structured and unstructured data for major project initiatives.
* Good Experience in analysis and understanding business requirements.
* Various Appreciation notes and Rewards (best team) from the Customer for the Quality Work done.
* Analyzing legacy System data to migrate data from old system to new system.
* Good understanding of Hadoop, Hive, Spark, Scala, Sqoop, Hadoopandof SAS, R statistical languages.
* Good knowledge of Data Analytics with Advanced Excel (Data analysis, solver, pivot tables, v-lookups, graphs/charts)

|  |  |
| --- | --- |
|  | **TECHNICAL SKILLS/CERTIFICATIONS** |

|  |  |
| --- | --- |
| **ETL Tools** | DataStage8.1/8.5/9.1/11.3/11.5, SSIS, OWB |
| **Big Data/ Hadoop** | Cloudera 4.x/5.x Manager/Hue, Apache Hadoop, HortonWorks HDP 2.3, Hive, Pig, Sqoop, Kafka |
| **BI Analytics / Visualization** | Business Objects, Cognos, Tableau Desktop, Power BI, D3, QlikView |
| **Database/File System** | Mongo Db, Teradata 14, Oracle 10g/9i, Netezza, SQL Server 2008, DB2, Hadoop HDFS |
| **Job Scheduling Tools** | TWS -Tivoli Scheduler, AutoSys |
| **Virtualization Technologies** | VMWare, Power VM, Virtual box, Citrix Xen |
| **Cloud Environments** | Cloudera, Azure, AWS |
| **Data Mining Tools** | SAS Enterprise miner, IBM Information Analyzer |
| **IDE Technologies** | Intellij, Eclipse |
| **Programming Languages** | SQL, PL/SQL, Python, R, SAS, Java, Shell Script |
| **Operating system** | Ubuntu, Windows |
| **Version Control** | CVS, GitHub, ClearCase |
| **Certifications** | OCA Certification, IBM DataStage 8.1 and IBM DataStage 8.5 Certification |

|  |  |
| --- | --- |
|  | **PROFESSIONAL EXPERIENCE** |

**Client: IBM Danbury, CT**

**Project: IDT Compliance June 2018 – Till Date**

IDT compliance project calculates indirect tax on IBM entities. It is a data migration project in which data from various countries and source is cleaned and migrated from mainframe DB2 to new system.

**Role and Responsibilities (ETL DataStage Developer):**

* Followed agile methodology during the development process of the data designing.
* Migration of data available in legacy mainframe DB2 using ETL process.
* Developed several ETL jobs in DataStage to transform the data from various legacy sources based on business requirements.
* Unit testing of the code before promoting to UAT/PROD.
* Encryption/Decryption of source data using UNIX shell scripting.
* Job scheduling using job sequence in DataStage.

**Environment:**IBM IIS DataStage 11.5, Putty, WinScp, DB2

**Client: AZTech TechnologiesWilmington, DE**

**Project: Renaissance May 2018 – June 2018**

Project Renaissance establishes a multi-tenant environment to ingest data from various sources into Premier Systems and transform and stage and store by using Spark/Scala in Big data platforms.

**Roleand Responsibilities (Big Data Engineer):**

* Followed agile methodology during the development process of the data designing.
* Migrated the existing DataStage Jobs into equivalent Spark-Scala with reusable artifacts.
* Ingested Netezza data into Hadoop and vice-versa by using the Sqoop scripts.
* Developed new complex Spark-Scala applications to transform the source input files to target output files.
* Designed Data parsing, data recovery, fault tolerance, exception module using Spark/Scala

**Environment:**Cloudera 5.12, HDFS, Hive, UNIX, Spark, Scala 2.11, Intellij, GitHub Netezza, IBM IISDataStage 11.5

**Client: Citi Bank (Wipro)Pune, India**

**Project: Rules HarmonyNov 2015 – Aug 2016**

Rules Harmony is Citi In-House ETL Automation tool used in creating business rules for data cleansing, data enrichment, data validations and transformations. DQP is an important feature of this tool which allows business to do data quality check on incoming and outgoing records. Rules and DQPs written in this tool are converted to DataStage/Spark and then integrated in the respective technologies.

**Roleand Responsibilities (BI Consultant):**

* BI Consultant for Analytical tool (Rules Harmony) development/implementation to generate reports for reconciliation.
* Provided recommendations from functionality perspective required for performing various data manipulation and cleansing tasks.
* Engaged with different stakeholders for understanding business requirements.
* Created Detail Level Documents for developers to build/code rules.
* Handled team of 10 developers and coached them in coding rules (using SQL logic) using Citi analytical tool.
* Validation of rules, data and reports generated before delivering it to client.
* Data from different data sourcesis extracted and analyzed using Tableau.
* Created dashboards for non-technical analysts and business users to convert data into understandable, interactive graphics.

**Environment:**Tableau Desktop 9.3, Microsoft Office 2010, SQL Server, SQL, Python, Rules Harmony, IBM IIS DataStage 11.3

**Client: Citi Bank (Wipro) Noida, India**

**Project: Genesis Sep 2013 – Oct2015**

To develop Data Warehouse for the CITI bank which involves integration of data from various countries source systems and cleansing of data according to the Customer Requirements and loading into the Warehouse having Teradata as the target database.

**Roleand Responsibilities (Senior DataStage Developer):**

* Developed several complex ETL jobs in DataStage to transform the data from various sources based on business requirements.
* Followed separate phases (SIT/ UAT/PAT) of testing before pushing code to PROD environment.
* Managed new/ongoing business requirements with performance tuning of complex DataStage jobs.
* CreatedTivoli schedules for scheduling jobs in PAT and PROD environment.
* Built various BTEQ scripts in Teradata required for loading data from holding to target database.
* Developed best practices for standard naming conventions and coding practices to ensure consistency of data models.
* Worked on a token system for automating the loading of data from holding to target database to prevent failure due to multi-load error.

**Environment:**Microsoft Office 2010, IBM IIS(DataStage 9.1/11.3), Oracle 10g, Oracle Toad, Teradata, Tivoli scheduler, Putty, Unix Shell Script, HP Quality Center, WinSCP, TextPad.

**Client: State Street(Wipro)Noida, India**

**Project: EDW Oct 2011 – Aug 2013**

State Street is one of the major client of Wipro and world’s leading provider of financial services to institutional investors with integrated range of services spans the entire investment spectrum, including research, investment management, trading services and investment servicing. Enterprise data warehouse (EDW) maintained by state street for its various IMS and GTS clients including Investec UK and SA, Martin curie, Lazard, AXA etc. for all their business and analytical needs.

**Roleand Responsibilities (DataStage Developer/Lead):**

* Understanding BI setup in production environment, createdapplication-based documents (SMTD/SOPs).
* Maintenance and Enhancement of Enterprise Data warehouse.
* Daily monitoring/scheduling of jobs via AutoSys scheduler to ensure meeting SLAs.
* Coached team of 6-8 developers in resolving ETL job failures and performed RCA on failures if any.
* Coordinated with development team to fix ETL (DataStage) jobs for any issue from design/performance perspective or if needed then coordinated with source team in case of any issue identified source data.
* Proactively participate in performance tuning of jobs which have shown degradation in performance or have started to impact SLAs.
* Automated several user requests to provide data that required manual and time taking process.
* Organized daily technical discussions with the Onsite team and including the individual offshore work stream leads and set expectations for offshore delivery.
* Daily status reports which show various metrics on support hours spent on application, number of issues, SLA miss, etc.

**Environment:**Microsoft Office 2007, IBM IIS (DataStage 8.5), DB2,Clear Case, AutoSys, Putty,UNIX Shell Script

**Client: Astellas(TCS) Noida, India/Leiden, Netherlands**

**Project: EDW Jan 2011 – Sep 2011**

EDW was a data integration project, collecting data related to drug discovery from various sources. Fetches the data from Argus and Polaris. Reports are sent to other bodies for checking whether the trials are being conducted properly or not. The data warehouse reads data from sources in, so called, pre-defined “feeds”. Each “feed” represents data from a source.

**Role and Responsibilities (DataStage Developer):**

* Figured out pain areas present in ETL, Reporting and DWH design where performance tuning is required.
* Suggested/implemented new design and created a new ETL design, which was more simplified and robust.
* Coached offshore team to create/implement ETL jobs in DataStage designer.
* Code review of ETL jobs from performance perspective and data from quality perspective before delivering.
* Implemented designsandworked on performance of long running jobs required to improve overall application.
* Conducted Knowledge Transfer sessions to train the resources and prepared various documents for new joiners.

**Environment:**Microsoft Office 2007, IBM IIS(DataStage 8.5), Oracle9i,UNIX Shell Script, FileZilla

**Client: Qantas (TCS) Noida, India**

**Project: Marlin Jan 2009 – Dec 2010**

Marlin is a legacy modernization project in QE (Qantas Engineering), implementing the new generation unified MRO (maintenance, repair and overhaul) system to standardize and optimizes its business process and manages the airworthiness compliance. Data migration is one-time data movement from legacy source system to the target system Maintenix. Since the legacy applications have vast amount of accumulated data over 35 years it needs cleansing both at the source & during data transformation. The age-old data needs enrichment, and transformation based on the business rules and target application data format specification technical rules. The data from Mainframe, DB.2 9.0, MS-Access, and PDF are migrated to target application Maintenix.

**Roleand Responsibilities (DataStage Developer):**

* Developed jobs performing migration of data from legacy system and transformations in DataStage Designer.
* Validated the data for quality and consistency loaded into tables to ensure quality.
* Created High Level, Low Level Design, Code review and testing documents.
* Worked on changes in requirements, enhancements and performance tuning of DataStage jobs.
* Standardize data to improve data quality through a sequence of data cleansing mechanisms using Quality Stage.
* Generating reports that helps to make critical business decisions from Information Analyzer that evaluates the content and structure of data for consistency and quality.
* Providing Knowledge Transition (Functional and Technical) to new team members and daily status reposting to onsite counterpart.

**Environment:**Microsoft Office 2007, IBM IIS 8.1(DataStage, Quality Stage, Information Analyzer, Fast Track), Windows 2007 server, Oracle 9i, Oracle PL/SQL developer, Ultra Edit.

**Client: Tata Capital Limited (TCS) Mumbai, India**

**Project:CDI**

**Apr 2007 –Dec 2008**

A TCL Corporate Portal is primarily aimed at enhancing customer experience and enabling the TCL user by providing availability of right information at the right time.

**Roleand Responsibilities (OWB Developer):**

* Module lead for the project, managing all the resources allocated to the module.
* Developed several mappings performing ETL in OWB and validated the data loaded into tables.
* Created various PL/SQL functions and procedures for validating the source data.
* Created High Level, Low Level Design, Code review and testing documents.
* Work on changes in business requirement, and on improvement of processes that are in place to get a task completed on time.

**Environment:**Microsoft Office 2005, Windows server, Oracle OWB, Oracle 9i, Oracle SQL developer, PL/SQL, Ultra-edit

|  |  |
| --- | --- |
|  | **ACADEMIC PROJECTS** |

* **Predicting the House Prices** [Scikit Learn - Logistic Regression, SAS Enterprise Miner]

Built analytical model using Linear Regression to predict house prices. Identifying patterns that influences housing prices.

* **Sentiment analysis using R**

Shiny app on sentiment analysis of the food reviews given by customer and compared to the rating given by customer.

* **Scheduling Analytics using Python**

Optimization of an Airline Schedule to maximize the number of flights, utilize aircrafts evenly, and utilize gates at airports as evenly as possible, and distribute flights among all markets.

* **Real-time Twitter Analytics using Big Data** [Hadoop, Hive, Sqoop, Flume, HDFS]

Built a near real time discovery platform to analyze sentiment from tweets using big data analytics framework Cloudera. It can continuously measure and compare sentiments of given input terms

* **Machine Learning with Big Data and Spark** [PySpark, Scala, ml]

Build a model that with cluster clients of a wholesale distributor based of the sales of some product categories.

|  |  |
| --- | --- |
|  | **EDUCATION** |

Master of Science in Business Analytics, University of TEXAS

Bachelor of Engineering, Maharshi Dayanand University