**Baskara**

[**Sateesh.g@paradigminfotech.com**](mailto:Sateesh.g@paradigminfotech.com)

Contact: 443-982 2951

**Summary:**

* A service oriented, customer focused Senior Data Architect, Data Engineer and IT professional with extensive hands on experience in support and delivery of Data Warehouse, Operational Data Store, BI solutions, Data Quality, Master Data Management, Big Data technologies and Cloud SaaS solutions
* Advanced data modelling skills including of Conceptual Data Model (CDM), Logical Data Model (LDM) and Physical Data Models (PDM) for OLTP’s and Dimensional Data Modelling (Star/Snowflake), normalization, de-normalization and implementation of Kimball and Inman Methodology
* Led DW teams and implemented multiple ETL, Business Intelligence and Data Analytics projects with resources from cross functional and technical teams. Scoped, planned, and prioritized multiple projects simultaneously.
* Over 15 + Years of experience in Banking, Financial, Trading and Healthcare Vertical.
* Led and participated in the development of short and long-term strategies to drive innovation, decrease speed to market, improve quality, and lower cost.
* Hands on Experience in data modeling tool using Erwin,Hackolade and Azure Purview ,Informatica AXON for data governance
* Hands on experience in BI Reporting Tools Power BI,R,Azure Synapse Analytics
* Experience in CI/CD pipe line tools for build, deployment automation using Terraform Scripts, Jenkins, GitHub,GitLab,Kubernetes in Developement,Testing,Production Environments
* Hands on Experience in Azure Databricks, Spark SQL, Spark Core,PySpark,DataFrame, Cluster Management, Delta ,Parquet tables and Data Lakehouse
* Hands on Experience in Webservices, Azure Data Factory, Azure Synapse, Azure SQL DB, Azure Data Lake and Azure App Service, Rest API using Azure API Management and Azure Functions,

Rest API,AWS S3,AWS Glue,Lambda,EC2,EMR,Redshift, HDFS,Pig,Hive,Hadoop,HBase,Spark,PySpark

* Developed Interactive Power BI Dashboards for variuos business units improving the data visualizations, analysis and decison making.
* Implemented Data Analysis Expressions (DAX) calculations, Power Query Transformations, models resulting perfomance for large data sets in Dashboards.
* Optimized Reports, Dashboards visualization’s, reducing dashboard loading time and enhancing the end user experience.
* Expertise in Informatica Power Center, Informatica Data Quality (IDQ), Informatica Analyst, Informatica Master Data Management (MDM), Informatica Meta Data Manager, Webservices
* Optimized Workbooks, Tableau Dashboards visualization’s, reducing dashboard loading time and enhancing the end user experience.

Hands on experience in BI Reporting Tools using Power BI, Tableau, R, SAS, RStudio, Tidyverse Packages, ggplot2

* Experience in AI/ML PyTorch,NumPy, SciPy, Matplotlib,Pandas,Scikit-Learn,NLP,RNN, Evaluated Decision Forests, Naïve Bayes Classification &k-Nearest Neighbors models
* Extensive experience in Database Development in Oracle,DB2,MySQL,PostgreSQL, Cassandra,MongoDB,NoSQL,Presto,Teradata
* Perform Business Analysis, User Requirement Gathering, User Requirement Analysis, complex data analysis in support of ad-hoc and standing customer requests, develop programs, methodologies for analyzing and presenting data
* Strong background in Autosys, AIX, UNIX, Linux, Shell Scripting (Korn Shell/Ksh) and Perl Scripting.
* Involved in full life cycle development and support of Data Warehouse projects. Involved in application support with SEI-CMM LEVEL 5 Software Quality, Process development and Test Methodologies.
* Experienced in establishing QA Procedures, QA Standards, proficient in writing and implementing test plans, documenting the testing procedures and writing in-depth test cases to meet the testing cycle.
* Extensive experience in conducting requirements gathering sessions to define system requirements via interviews, questionnaires, brainstorming sessions, workshops and prototyping.
* Hands on experience in Project planning, execution, monitoring and resource balancing skills with ability to support simultaneous projects in an organizational structure in both Agile/Scrum/Kanban and Waterfall Methods

**Education:**

Post Graduate Diploma in Computer Application, Loyola College, India.

BSc (Maths & Computer Science), Madurai Kamaraj University, India

**Certifications/ Courses:**

* Certified Informatica Designer.
* Informatica Data Quality and Informatica MDM Hub Certification.
* Salesforce, Visualforce, Apex
* Amazon Web Services (AWS) DW/Redshift Certified
* Databricks Certified
* AI/ML using TensorFlow

**Skills:**

**DATA**

**WAREHOUSING:** Informatica PowerCenter 10.5.4/10.2, Informatica Developer (IDQ) 10.5.4,

Informatica Analyst, Informatica Metadata Repository, Informatica Metadata

Manager,AXON,Informatica MDM HUB 10.2.x,/10.x/Siperian MDM Hub XU, SIF,

Informatica Power Exchange, Data Stage 7.x, Microsoft SSIS, DTS

**BI TOOLS:** Power BI,Tableau 2023.2/2021.4/2019.2/10.2/9.1, R, Business ObjectsX1i /R4 4.2,Crystal Reports 2011,

MicroStrategy 9. x/10.x Microsoft SSRS/SSAS, Hyperion, SAS

**DATABASES:** Oracle 19C/12C, Sybase ASE/IQ, DB2 UDB 9.5,Microsoft SQL Server 2014, MySQL5.x,PostgreSQL,Redshift

Teradata Vantage Cloud,Teradata 16.x/7.x

**LANGUAGES:** VBA,Python, R, PL/SQL,UNIX/LINUX Shell Scripting, C++, Java, Perl

**WEB TECHNOLOGIES:** J2EE, Java Script, ASP, ASP.NET, Web Services, SharePoint/MOSS 2007, AWS

**DATA MODEL:** Hackolade, Erwin 12.5,r9.6/R7, Visio 2007, ER Studio 9.5.1/8.0

**TOOLS:** Toad, Clear Case, Clear Quest, ALM, HP Quality Canter 10, QTP, DOORS 9.6, JIRA, GitHub, GitLab

Terraform, Kubernetes,Eclipse, MS Project 2007, SVN, Jenkins, Rally,GitHub,GitLab,XML Spy, RStudio

**ERP:** SAP R/3.SAP BW, SAP S/4 HANA

**CRM:** Salesforce, VisualForce, Apex

**SCHEDULER:** AUTOSYS, UC4, CONTROL M.

**MAINFRAME:** IBM 3090, MVS, JCL, CICS, DB2, COBOL.

**BIG DATA:** HDFS, Pig, Hive, Hadoop, HBase, NoSQL, Map R, Apache Spark, Kafka, Cassandra, MongoDB

**AWS/Azure:** Azure Databricks,Azure Rest API ,Azure API Management, Azure Functions,Microsoft Purview, Data Factory,Azure Synapse, Apache Airflow

**AI/ML** : TensorFlow 2.x,NLP,RNN,KNN,Naïve Bayes Classification, SciPy, Scikit-Learn

**Core Competencies:**

* Data Warehouse Application Design ,
* Data Quality Management
* Client management
* Data Analysis
* AI/ML
* Master Data Management
* Data Architecture
* Database Performance and Tuning
* Big Data
* Technical Project Management

**Professional Experience:**

**USPS, NC 11/20 – Till Date**

**Technical Lead/Data Architect**

**Application : ETL/Data Quality /MDM**

The Project is to Implement ETL, Data Governance, Data Quality and Master Data Management(MDM) for USPS Enterprise Data Governance team for

HR, Customer Relationships, Asset, Facility, Finance and Location Domains.

**Responsibilities:**

* Responsible for Data Architect, Data Modeling, scoping, mapping, logical and physical model efforts, planning and supporting the migration plan
* Involved in Data Analyst,Data Model for ETL,Data Quality and MDM Staging Schema
* Involved in Development of ETL, Rest API, REST Web Service,REST API,Data Service, complex Informatica Mappings,Mapplets,Workflows and Informatica Applications ,Workflow Job Scheduler
* Developed ETL data pipeline to ingest data into Delta Lake and Data Lakehouse tables using Databricks
* Used Spark SQL,PySpark,Spark Core,DataFrame,Cluster Management, Delta and Parquet tables in Databricks
* Create Databricks workspace,clusters,notebooks,jobs,AWS S3 buckets using Terraform for Developement,Testing,Production Environments
* Developed Rest API for MDM Golden Records and Vehicle Tracking Reports using Mulesoft
* Developed standard, ad-hoc and key performance indicator (KPI) reports and dashboards
* Create advance analysis actions, DAX calaculations,parameters,filetrs,trendlines,statistics,maps in the reports.Facilitating Power users in Power BI Dashboard Development
* Create groups, hierarchies, datasets to develop summary ,detail reports and Analytical Reports. Publishing the Reports and Dashboards into Power BI Server.
* Responsible for Design, Development, Implementation Data Science Applications, Dashboard Reports and Publishing the Reports and Dashboards into Azure
* Create various Charts like Stacked bar chart,Pie Chart,Donut Chart,Funnel Chart,Line Chart, Treemap Chart,Heatmap Chart, Histogram Chart with drilled down details
* Responsible for identifying Key Data Elements and data sets for Reports.
* Developed standard, ad-hoc and key performance indicator (KPI) reports and dashboards
* Create groups, hierarchies, datasets to develop summary ,detail reports and Analytical Reports.
* Create advance analysis actions, calaculations,parameters,trendlines,statistics,maps in the reports.Facilitating Power users in Dashboard Development.
* Developed SQL Scripts for ETL, Unix Shell Scripts (ksh/sh) for AFT process and DQ Application for Informatica Scheduler
* Responsible for identifying Key Data Elements for Informatica Analyst
* Created Data Catalog (EDC), Data Lineage using Azure Purview and Informatica AXON for data governance
* Worked with Data stewards to identify key data elements for Profiling,Data Masking, Standards,Data Cataloging,Data Quality rules
* Performed data quality checks, identify the issues, Identify areas for data quality improvements improvements, implement fixes and improvements.
* Involved in Performance tuning for Informatica Applications , Rest API,Database and SQL queries and Optimizations.
* Create groups, hierarchies, datasets to develop summary ,detail reports and Analytical Reports. Publishing the Reports and Dashboards
* Involved in migration of No SQL scripts into Relational SQL
* Create Complex PL/SQL,SQL scripts for ETL Load and QA Testing.
* Define and Describe Business terms, Concepts, DQ Rules ,Profiles and Scorecards in glossary for Informatica Analyst
* Developed Python scripts for standardize and cleanse the JSON,XML,Excel files for data anomaly

**Environment:** Azure Databricks, PySpark,Spark SQL, Oracle 19.x, Oracle PL/SQL, Erwin, Informatica Developer 10.5.4,Informatica PowerCenter 10.5.4, Informatica Analyst 10.5.4, Informatica AXON, Python 3.9, Linux, Unix, Hive, Informatica MDM Hub,ServiceNow,GitHub,GitLab,Anoconda 2.4.2,ALM, MongoDB, Cassandra, SAP S/4 HANA,Neo4j,R,RStudio, ggplot2, Apache Airflow, Presto, Mulesoft,Power BI, Power BI Desktop, Azure Purview, Jira,Jenkins,Terraform,Kubernetes

**CMS,MD 06/19 – 11/20**

**Technical Lead/ Data Architect**

Application : ETL/ Data Warehouse/Big Data

The Project is to Implement ETL, Big Data solutions on AWS for (Department of Human Services) and CMS Projects (Centers for Medicare and Medicaid Service) for Midas Market Place Model

**Responsibilities:**

* Responsible for Data Architect, Data Modeling, scoping, mapping, logical and physical model efforts, database planning, Data Lake and supporting the migration plan
* Implemented and support big data solutions, Design and development of cloud-based solutions, technical design and implementation of the architecture
* Responsible for Data Architect, Data Modeling, logical and physical model for Midas Marketplace Model
* Design, Development, Implementation of ETL pipeline using Spark, Hive,S3,Data Lake and Redshift solutions on AWS
* Create Databricks workspaces, AWS S3 bucket using Terraform for Developement,Testing,Production Environments
* Used Terraform scripts to create AWS S3 buckets,Databricks workspace,clusters,notebooks,jobs,AWS Glue services and Redshift Resources for Developement,Testing,Production Environments
* Create Lambda step functions to trigger events and to process data files.Scheduled Lambda functions
* Used AWS SNS,SQS for event notifications and messaging
* Developed parsing process to create DML, DDL for tables using PySpark for Oracle to PostgreSQL migration and conversion.
* Developed ETL Pipline using Spark SQL,PySpark in Databricks
* Design and Developed dashboards, create and maintain data visualizations reports using Power BI
* Responsible for data analysis, identifying the attributes ,Dimensions and Facts for the Reports
* Responsible for Design, Development, Implementation Data Science Applications, Tableau Dashboard Reports and Publishing the Reports and Dashboards into Tableau Server
* Create various Tableau Charts like Bar Chart, Pie Chart, Trend Line,Gnatt Chart,Scatter Plot,Heat Map,Tree Map, Cross Tab,Histograms with drilled down details,Views and Filters
* Identify data points for Machine Learning process for Predictive Visualization, Forecasting for Improper Payments, Tobacco Risk Ratings on Insurance Payments, and Fraud Detection for national producers related license information for Application and Policy Assistor
* Research modern technologies and products for their applicability to business processes, compare various solutions and determine the one which best fits the need.
* Used CloudWatch to monitor AWS Services and Cloud Watch logs to monitor applications
* Design and Architect Kafka Streams using Stream Sets for AWS Elastic Search.

**Environment:** PySpark,AWS,Lambda,EC2,EMR,Apache Spark, PostgreSQL,Oracle 12.2, Toad, Informatica PowerCenter BDM 10.2, Power Exchange, Python 3.7, Linux, Cloudera Hadoop HDFS, Hive, Impala, Kafka, GitHub, Rally, Tableau 2019.x, R, Erwin,PGAdmin,Databricks,JIRA,Jenkins, MongoDB, Tableau Prep, Hackolade,Redshift, Cloud Watch, Apache Airflow, SQS/SNS,EventBridge,Presto, ,Azure,Power BI, Power BI Desktop, Databricks, PySpark,Spark SQL, Jira,Jenkins,Terraform,Kubernetes

**Common Securitization Solutions,MD 05/17 – 05/19**

**Technical Lead/Data Architect**

Application : ETL/ Data Warehouse

The Project is to Implement ETL and Data Quality solutions on AWS for various modules like Disclosure, Bonds, Tax and Issuance, to automate the data validation process and to perform Reporting and Monitoring for Reconciliation Process.

**Responsibilities:**

* Responsible for Data Architect, Data Modeling, scoping, mapping, logical and physical model efforts, database planning, Schemas, Data Lake and supporting the migration plan
* Design, development of Informatica Mappings, Mapplets, Workflows, Worklets using Informatica PowerCenter and Implemented performance tuning logic on Informatica Targets, Sources, Mappings and Sessions to provide maximum efficiency and performance.
* Responsible for data analysis, identifying critical data elements, KPI, Measures and Metrics
* Develop Unix Shell Scripts to parse flat files, Python Scripts to parse JSON files and to download files from AWS S3 buckets.
* AWS Glue, Rest API was used for ETL and data ingestion was from AWS S3(Data Lake)
* Loaded the aggregate data into a relational database for reporting, dash boarding and ad-hoc analysis.
* Developed ETL Pipeline for Redshift Data Warehouse
* Create Queries, data transformations using Data frame API
* Create Lambda step functions for REST API
* Used Terraform to create AWS S3 buckets,AWS Glue services,Redshift Resources for Developement,Testing,Production Environments

**Environment:** AWS,Tableau 2019.1, Tableau Prep ,Oracle 12.2, Toad, Oracle PL/SQL, Informatica PowerCenter 9.6.1, Python 2.6, Linux, AWS, AWS Glue, Redshift, S3, Athena, Lambda, EC2,

Apache Spark, Tableau 9.1, Tableau Prep, Bash Scripting, R, Snowflake, GitHub, GitLab, Rally, Presto, Mulesoft, Teradata 16.x, Terraform, Kubernetes

**Fannie Mae,VA 04/11 –04/17**

**Technical Lead/Data Architect**

**Application :** Data Quality, Data Governance and EDI Applications

The Project is to Implement ETL, MDM for Security Master and Multifamily Business-focused Data Quality Rules to automate the data validation process and to perform Data Analysis, Profiling, Cleansing, Standardization, Matching, Consolidation, Reporting and Monitoring. Maintaining centralized control and management of data quality standards. Provide management level reports that measures the Key Performance Indicators (KPI), Subject Area, key data quality dimensions such as Reasonableness, Completeness, Consistency, etc. Reconciliation between various sources and target system.

* Conducted JAD sessions with Management, Business Analysts, SME, Users in capturing and understanding/analyzing business requirements, defining project scope and communicate with other stakeholders for open and pending issues.
* Developed Tableau Dashboards, Reports for multidimensional analysis and Enhancements and Maintenance of reports
* Generated interactive dashboards in Tableau with Quick filters, Parameters. Created adhoc reports to users by connecting various data sources. Published Workbooks in Tableau server
* Design the ETL processes to load data / real time data into the target Oracle database from multiple sources such as XML, flat files and databases.
* Involved in the development of Informatica Mappings, Mapplets and Workflows using Informatica PowerCenter and Informatica Developer for complex data quality rules. Created various tasks like Session, Command, Timer and Event wait.
* Implemented performance tuning logic on Informatica Targets, Sources, Mappings and Sessions to provide maximum efficiency and performance.
* Modified several of the existing mappings based on the user requirements and maintained existing mappings, sessions and workflows.
* Participate in establishing and evolving architecture standards, reference models, guidelines and principles
* Lead Role in Design, Architect and Implementation of Data Quality Projects.
* Responsible for implementing MDM standards, MDM best practices, procedures, enhancements and support for Security Master ETL/MDM application for EDI.
* Developed and implemented SOA using XML/XSD and Web Services for data integration.
* Developed Oracle PL/SQL Scripts ,SQL Scripts and UNIX Shell Scripts.
* Involved in development, implementation, and management of data strategies that optimize data quality and improve standardization and business information value derived from enterprise data.
* Perform detailed analysis to support the planning and impact assessment for Data Quality initiatives.
* Involved in Data Quality Management, Data Resource Management, Data Stewardship and Data Administration, Maintaining Data Standards, Data Governance.
* Responsible for Data Architect, Data Modeling, scoping, mapping, logical and physical model efforts, database planning, Schemas and supporting the migration plan
* Developed BOXI Crystal Reports, Tableau Dashboards, Reports for multidimensional analysis using R and Python Scripts.
* Generated interactive dashboards in Tableau with Quick filters, Parameters. Created adhoc reports to users by connecting various data sources. Published Workbooks
* Responsible for creating data profiling methodology, identifying known data quality issues and communicate with team for handling the data.
* Perform validation for Golden Records, Business Glossary, Catalog and Lineage analysis for IMR (Informatica Meta Data repository).
* Involved in data quality development and support initiatives by designing new data models, helping to track down sources of data anomalies, etc.
* Implemented DQ rules in Informatica Analyst (IDQ) for Adhoc Profiling and analysis for rules exceptions
* Work with the Agile team on complex reporting / dashboard solutions, developed self-reporting BI platform to the end user community

Involved in Master Data Management, production support, QA/Testing, developing test scripts for Data Quality projects and reports, executing test cases in Quality Center and ALM

**Environment:** Tableau 9.1, Tableau Desktop 9.1,Oracle 11g/12C,Toad, Informatica PowerCenter 9.6.1,Informatica MDM Hub 9.7.x, IDQ Developer 9.6.1, IDQ Analyst, 9.6.1, Informatica Metadata Manager, Power Exchange, Salesforce, TOAD, Linux, Shell Script(Ksh), Autosys, Tableau 9.1,R, RStudio, ggplot2,Business Objects R4 (BOXIR4) /Crystal Reports 2011,ER Studio 9.5.1,TIBCO, Eclipse IDE, Java, SVN, Jenkins, AWS, Netezza 6.1.x,Python 2.x

**Pfizer Inc ,NJ 09/10 – 04/11**

**Technical Lead/Data Architect**

**Application:** Heath Care Provider Information Systems (HCP)

Data feeds from Pfizer applications from different countries extracted, standardized, cleansed before loading into Siperian MDM for Customer and Address Intelligent Matches for Heath Care Provider Information Systems (HCP).

**Responsibilities:**

* Developed ETL procedures using Informatica PowerCenter and custom PL/SQL.
* Created Informatica mappings for populating the data into the dimension, facts, and history tables from different sources systems.
* Developed Workflows using task developer, worklet designer, and workflow designer in Workflow manager.
* Tuned the performance of mappings by following Informatica best practices and also applied several methods to get best performance by decreasing the run time of workflows.
* Developed procedures for monitoring of daily/weekly/monthly ETL processes and backup procedures for possible ETL load failures ,developed scheduler to run Informatica workflows
* Responsible for development of Informatica Mappings, Mapplets, and Workflows using Informatica PowerCenter and Informatica Developer for complex data quality rules
* Understand requirements to create, validate and implement Logical and Physical Data Models, DDL and Scripts.
* Created target state data model, identified attributes for Landing, Staging Area to be maintained in the Siperian MDM Hub.
* Develop process for data profiling, standardization, cleansing the data before loading into Siperian MDM Hub for match and merge for Customer and Address with HCP systems.
* Define, advice and write rules to enhance MDM efficiencies in support of client data program
* Analyze, validate, and interpret the content and quality of data within the Medical applications, Transactions, data stores and warehouses using Informatica Data Quality.
* Perform detailed data analysis and present findings in a cohesive, high impact format to drive and inform business and technical decisions
* Develop and maintain ORS Model and MDM rules and guidance etc.
* Responsible to perform functional and performance testing.

**Environment:** Oracle 11g,TOAD,Unix, Siperian MDM Hub XU,SIF, Business Data Director (BDD),Informatica PowerCenter 9.1,Informatica Data Quality 9.1,Erwin7.3.x, Windows 2003 Server, Informatica Metadata Repository, Netezza 6.1.x,Business Objects X1i,JBoss Application Server, Visio, Teradata

**Walt Disney,CA 03/10 – 08/10**

**Technical Lead / Data Architect**

**Application :** Enterprise Data Warehouse

Data Extract from SAP BW and Loading to ODS/FDM. Various Segments OPA, DCP (Disney Consumer Products) and Studios loads these data into their respective segment Data Marts, Cognos Cubes, Planning and Forecasting is used to generate reports using this data

**Responsibilities:**

* Developed Informatica mappings for extracting, cleansing, transforming and loading into Data Warehouse
* Involved in performance tuning for Informatica Mappings and databases to identify and remove processing bottlenecks, implementing the best practice for development
* Maintained source and target mappings, transformation logic and processes to reflect the changing business environment over time.
* Used various transformations like Filter, Router, Expression, Lookup Aggregator, Sequence Generator,
* Update Strategy, Joiner, Normalizer, Sorter and Union to develop mappings in Informatica Designer. Extensively used workflow variables, mapping parameters and mapping variables.
* Created sessions, batches for incremental load into staging tables and scheduled them to run daily.
* Used shortcuts to reuse objects without creating multiple objects in the repository and inherit changes made to the source automatically.
* Involved in Creating Stored Procedures, Functions, Views, Index for loading the staging tables.
* Performed Data Mapping and Data Migration from heterogeneous sources into the Data Warehouse
* Involved in development of Perl and UNIX (Ksh) Shell Scripts to generate various file formats and reports.
* Responsible for the design and delivery of a new Data Marts
* Involved in Data Architect, Data Model, ETL Design and Development
* Preparing release notes, deployment of Workflows and Jobs to test and production environment
* Created attributes for the dimensions/Fact tables based on the traceability matrix & source systems.
* Serve as a liaison between business and technical teams, database administrators and develop detailed functional, data and technical requirements
* Defined Target Load Order Plan and Constraint based loading.
* Developed various Reports, Interactive Dashboards with drill-down capabilities, with various charts and views using filters. Created parameterized reports, Drill down and Drill through reports.
* Automated and scheduled UNIX shell scripts for Informatica sessions and batches
* Coordination with offshore in development activities and QA testing.

**Environment:** Oracle 11G,AIX, DB2 UDB 9.5,SQL Server 2005, Shell Script, Windows 2003 Server, Informatica PowerCenter 8.6, Erwin7.3, SAP R/3,SAP BW, MicroStrategy 9.x,Teradata

**Previous Clients:**

Freddie Mac, ITG(Investment Technology Group),Standard Chartered Bank, Harvard Medical School