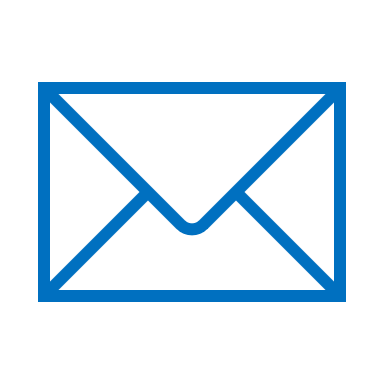
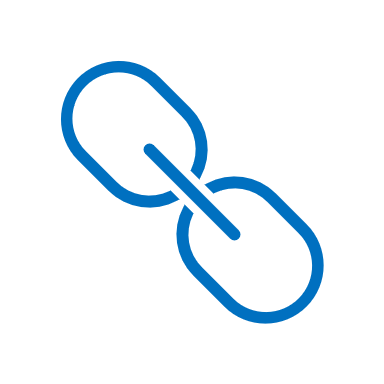
*ZAHIR SHEIKH*

+91 775-780-5314  [zahir.i.sheikh@gmail.com](mailto:zahir.i.sheikh@gmail.com)

www.linkedin.com/in/zahir-sheikh-4675358  Bangalore

**PROFESSIONAL SUMMARY**

* Around **15+ Years** of technical expertise in the IT industry having exposure to Financial, Communication, Insurance and Hospitality domain with prime focus to get answers from the data by building Advance Analytical solutions.
* Last **10 years** was in USA and worked as a Senior Technologist with strong business acumen and technical experience in areas like Data Integration/Quality, Data Management Consulting, AI/ML, Cloud Computing, Project Management and **Pre-Sales** for new proposal, effort estimation, cost estimates and resource estimates along with timeline and milestones to Customer.
* Hands-on experience in large scale Data Warehousing and well versed with Data consolidation/ ETL / ELT, Data Modeling, Data Quality, Data Enrichment, Data Security, Data Governance, Data Virtualization and Data Harmonization concepts, practices, and their implementation.
* Expertise on building **Ontology@Scale** platform for an organization to infer the Business Knowledge using **Data Fabric** solution with next generation technology by stitching a set of independent services to provide a single view of company’s data, irrespective of the repositories where data is generated, migrated to or consumed from either **on-premises** as well as **cloud**.
* Proven ability to analyze key business drivers and develop strategies for overall architecture governance, Solution Architecture, Implementation, and various application developments.
* Accountable for end-to-end project deliverables. Project delivery management through sprint planning – Scrum & Kanban; Software Development Life Cycle (SDLC) planning - Waterfall, Incremental, Spiral; intake management, release management and project planning - resources, tasks, requirements, milestones, and review points of Data Warehouse and Big data projects range from **1.5M to 10M USD**.
* Strong skills in Data Analysis, Data Requirement Analysis and Data Mapping for ETL/ELT processes.
* Experience in developing **PYSPARK** job using **Databricks** cloud notebook on **AWS** S3 cloud.
* Experience in implementing GROUP BY, ORDER BY, aggregated functions and join functionality between tables using **PRESTO** & **HIVE**.
* Experience in implementing batch query processing using **HIVE**.
* Experience in implementing **SPARK** data processing programs in Scala and Python to query and analysis data stored in **AWS S3 &** **Hadoop HDFS**.
* Extensively worked on data extraction, Transformation and loading data from various sources ranging from Oracle, Teradata, DB2, Netezza, MS SQL Server and Flat files.
* Responsible for all activities related to the development, implementation, and support of ETL processes for large scale data warehouses using Informatica Power Center **10.1.1** and Pentaho Data Integration Spoon **8.1** and database programming in **PL/SQL** (Stored Procedures, Triggers and Packages).
* Hands on experience in tuning mappings, identifying, and resolving performance bottlenecks in various levels like sources, targets, mappings, and sessions.
* Well versed in UNIX shell scripting and developing the complex SQL queries, unions and multiple tables joins and experience with Views.
* Excellent communication, documentation and presentation skills using tools like Visio and PowerPoint.
* Self-motivated, willing to assume responsibilities, a strong team player with great leadership qualities as well as comfortable individual with excellent communication and analytical skills, ability to adapt and learn new things quickly.

**ACCOMPLISHMENTS**

* **Apache SPARK** support subset of ANSI SQL 2003 and does not support recursive SQL like connect by clause to build the hierarchy and depth. Due to this limitation the project’s big data processing is at stack. Come up with innovative solution to build the hierarchy module. Used graph module to do network analysis in spark (i.e. java based **GraphX** library or python based **NetworkX** package). Identification of optimize existing numerical algorithms in graph module to study the structure and dynamics of hierarchy networks from employee to CEO. Analysis should be done at any level i.e., from top to bottom (CEO -> Employee) or from bottom to top (Employee -> CEO). Developed generic and highly optimize python script to build the hierarchical network for analysis purpose.
* Automated manual effort of environment setup activities take 32 hours for each delivery and team must prepare environment bimonthly. Identification of all the DB objects, backup of an existing ETL code, Scripts, Temp files and purging of old backup folders based on retention policy of the entire setup process. Built generic and highly optimize framework using shell & python script to take care of Sync Up activities and it is running on Linux Box. This framework implemented and had seen a drastic saving in an overall time from 32 hours to 4 minutes. The company’s Total Annual Saving of **$27,216.**

**INFORMATION TECHNOLOGY SKILLS**

* **Specialty** Solution Designing, Data Architecture, Data Modeling, Metadata Management, Master Data Management, Data Quality and Governance, System Migration, DWH/BI Architecture, Cloud Computing (Azure, AWS)
* ***Languages PL/SQL, SQL, BTEQ, Pyspark, Scala, Python, Panda, NumPy,***

***Unix Shell Script, Magic***

* **Cloud Services** AWS **Glue,** **EMR**, **S3, Redshift, Athena, Presto,** Databricks **Spark** Cloud**.**
* ***Big data*** Hortonworks **Hadoop 3.1, Hive, Sqoop, Spark 3.0.,** DBeaver **7.1**
* ***Data Modeling*** Oracle SQL Modeler, Erwin **9.2.**
* ***Database & Tool*** Oracle **11g/10g/9i**. Teradata **16.1**, MS SQL Server **2012**, DB2 **9.7.0.5**, Netezza **7.1.0.3**, Toad **9.6.1.1**, PLSQLDEV **7.0.0.1**, DbVisualizer **9.2.10**.
* **ETL Tool** Informatica **10.1.1,** Pentaho **(Spoon - Data Integration) 8.1**
* **Job Scheduling Tool** Autosys, Tivoli, Unix Cron Tab, Databricks Job Scheduler, Apache Airflow.
* **Tools**/**DevOps tools**  OmniStation 5.8, Azure DevOps, GitHub, Confluence Kanban Board.
* **Operating Systems** AIX 5.3/5.2, Linux 2.6.18, Windows

**EDUCATION & CREDENTIALS**

Bachelor of Engineering, Computer Engineering University of **Nagpur**, Maharashtra, INDIA.

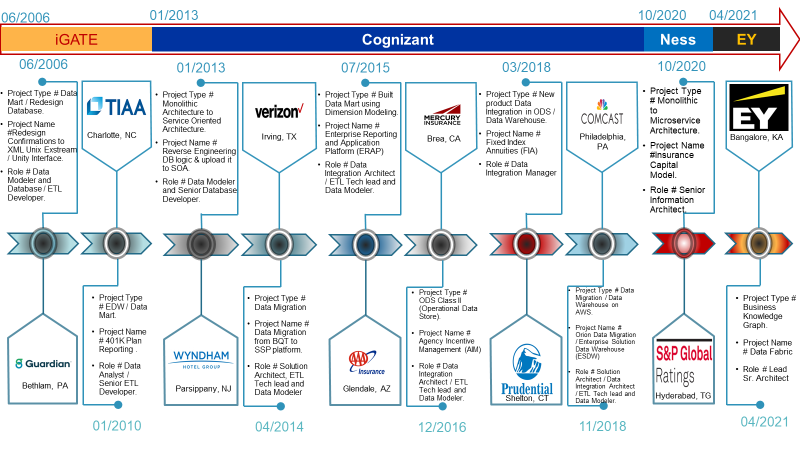
Microsoft Solution Architect Expert **Microsoft** Certified

Apache Spark – Structured Streaming **Databricks** Certified

Data Engineering with Databricks **Databricks** Certified

Delta Lake Rapid Start with Spark SQL **Databricks** Certified

**WORK EXPERIENCE**



* **Data Integration / Data Migration / Database Architect**
  + Worked as consultant such as Solution Architect, Information Architect, Data Architect, ETL Architect, Sr. Data Analyst, Sr. Developer, Data Modeler and lead the ETL/ELT development effort for Communication, Insurance (Life and Property & Casualty), Hospitality Enterprises’ business-critical projects by providing best solution and formulate strategies (i.e., Data -> information -> Knowledge) that leads to efficiency and effectiveness in business decision-making process.
  + Prepared Architecture blueprint for Data Readiness and Data as a service to internal / External stake holders by means of data is well catalogues and all translation and transformation are applied and available for consumption on **cloud** as well as **on-premises**.
  + Implemented data-based strategy for the corporate data (i.e., Structure, Semi-Structure & Unstructured) which needs to be aligned to the technology landscape. Understand and recommend the end-to-end process for identified data initiatives including requirements, feasibility, validation of data structure, compliance, and monetization.
  + Implemented data security strategy i.e., Advance Encryption Standard (AES -256) for data at rest, data in transit (SSL/TLS i.e., Socket Service Layer / Transport Security Layer with AES) and data in memory.
  + Executed User Access Management by defining data security policies on Cloud as well as On-premises data resources. It further leads to fine-grained access management of data object / entity at row level as column level access along with Role based Access Control (RBAC).
  + Define, configure, and optimize various Data Migration processes including landing, staging, hierarchies, foreign-key relationships, reference tables, cleanse functions using Referential Integrity and Erwin Technologies.
  + Execute against a data governance framework, with a focus on improvement of data quality and the protection of the corporate data assets through Enterprise / Local Metadata capture, modifications to organization behavior, policies and standards, principles, processes, governance metrics, related tools, and data architecture.
  + Provided various industry standard best solutions for Application decommission, OLTP to ODS, OLTP to EDW, Data Mart, Data lake and Data integration projects.
  + Designed Data warehouse to Data lake road map and implemented data ingest into **AWS cloud** through AWS Glue in batches and uses other AWS services like S3, EMR, Athena and Redshift.
  + **Databricks cloud** platform on AWS is used to process and transfer the data. Load curated data to consumption layer.
  + Improved performance of **ETL/ELT load** by redesigning the job or its SQL in ETL/ELT tools.
  + Improving applications stability through problem management. Crisis management to minimize the adverse impact of service interruptions and other incident on Client Business operations.
  + Transition management from the delivery to support organization during project rollouts.
  + Application rollouts activities such as preparation of User Training Manuals, Online Help Documentation etc. and conducting User Training.
  + Providing Services in a technically feasible and appropriately scalable manner which will accommodate bringing Applications in and taking Applications out of scope of the services.
* **Engineering Manager**
  + Lead the 8 members ODS, Big Data and EDW team with full accountability by sharing weekly and monthly project status report to stakeholders as well as facilitated resource onboarding.
  + Collaborate with business stakeholders to develop document policies and procedures surrounding Data Integration / Governance. Support key unit like KPMG audits etc.
  + Co-ordination between external team to understand the transformation logic of the new table structure design and its data necessary to produce KPI measures for BI reports.
  + Agile Transformation by setting up MS DevOps, Kanban, and Agile scrum model.
  + Executed over 10+ Projects with a mix of onsite-offshore engagement including Low-Cost Alternatives.
* **Pre-Sales**
  + Worked on multiple proposals for Data Integration / Data Migration projects.
  + Responding to client requests; Responses to clients include informal responses, pointers to publications/collateral, Request for Proposal (RFPs), Request for Information (RFI) and specific Statement of Work (SoW) or Work Orders.
  + Preparing Technical solution: Includes work estimation, designing architecture, preparing proof of concept (POC), project/staffing plan, resource estimates along with timeline and milestones to Customer.