**Deepa Hiremath**

Data Engineer Lead

**Email**- [deepashiremath@gmail.com](mailto:deepashiremath@gmail.com)

**Mobile** - 314 687 3938 [**SnowPro Core Certified**](https://www.credly.com/badges/221dbb43-6f91-4c58-8a0b-3597a3b80045/public_url)

**Professional Summary:**

* Over **13** years of IT experience in Data warehousing and Business intelligence with an emphasis on Project Planning & Management, Business Requirements Analysis, Application Design, Development, testing, implementation, and maintenance of client/server Data Warehouse.
* Experienced in optimizing data pipelines, improving performance, and ensuring data integrity.
* Experience with Snowflake cloud data warehouse and AWS S3 bucket integration using storage integration objects and pulling the various data formats from S3 into Snowflake tables.
* Experience in using Snowflake Clone and Time Travel.
* Experience in building Snow pipes for continuous data loads.
* Experience in migrating data from various sources to Snowflake Data Warehouse.
* Experience as a Big Data Developer involved in big data processing using Hadoop, Apache Spark, Hue, Hive, HDFS, and Spark job monitoring.
* Experienced in writing complex Spark SQL and Hive queries to meet business requirements.
* Experienced in optimizing Spark SQL and Hive queries by applying best practices and efficient querying techniques.
* Experienced in Business Intelligence ETL tools (Talend and Pentaho).
* Created Talend Mappings to populate the data into dimensions and fact tables.
* Implemented CDC/SCD methodology in Talend to load deltas to a Data Warehouse.
* Design, develop, test, implement, and support Data Warehousing ETL using Talend.
* I have a very good knowledge of RDBMS topics, and the ability to write complex SQL.
* Hands-on experience in coordinating with architects and building ELT and ETL complex  
  solutions, starting from identifying the required sources, source data preparation,  
  transforming data as per business logic, populating the BI LAYER for report extraction
* Expertise in migrating the code across various environments using – GitHub, Bitbucket
* Hands-on experience in scheduling and operationalizing applications using ControlM, Tidal, source availability, downstream readiness, platform resources, and transition applications to support teams with detailed documentation, and adequate automated notifications to avoid manual monitoring efforts.

**Technical Skills:**

|  |  |
| --- | --- |
| **Cloud Platform** | Snowflake, AWS (S3, RDS, IAM) |
| **Hadoop Ecosystem** | HIVE, HDFS, Hadoop, Spark, Impala |
| **Database** | MySQL, SQL Server, Teradata, Oracle, PostgreSQL |
| **Version Control Tools** | SVN, GIT, GIT Hub, Bit Bucket |
| **Bug Tracking Tools** | JIRA, ServiceNow, Bugzilla |
| **Languages & Scripts** | SQL, Python, JavaScript, XML, Java |
| **SDLC** | Agile, Scrum |
| **ETL/ELT Tools** | Talend, Pentaho. |

**Certifications:**

* SnowPro Core Certification

**Professional Experience:**

**Client: Hitachi Vantara,Dallas,TX.**

**Role: Data Engineer Lead** **01/2016 – Present**

**Responsibilities:**

* Experience with Snowflake cloud data warehouse and AWS Cloud S3 bucket integration using storage integration objects and pulling the various data formats from S3 into Snowflake tables.
* Evaluate Snowflake Design considerations for any change in the application.
* Bulk loading from the external stage (AWS Cloud S3) to the snowflake cloud using the COPY command.
* Used SNOW PIPE for continuous data ingestion from the S3 bucket.
* Data validation has been done through information schema and optimized workloads using clustering keys.
* Design architecture for End-to-End data flow from source to target including masking, and encryption which provides analytical data insights.
* Cloned Production Snowflake Tables to facilitate code modification and establish a testing environment, reducing testing setup time from 10 days to 1 day.
* Instrumental in automating manual processes, streamlining operations, and transforming them into repeatable, efficient capabilities. Enhances data management processes through insightful changes.
* Skillfully designs, builds, and manages AWS Cloud solutions that capture, manage, and utilize structured and unstructured data from diverse client data sources. Ensures optimal utilization of AWS resources.
* Experienced in optimizing Spark SQL and Hive queries. Tuned long-running jobs from 70-80 mins to 15 mins, achieving 80% performance gain.
* Experienced as a Big Data Developer involved in Big Data Processing to develop data pipelines using Hadoop, Apache Spark, Hue, Hive, HDFS, and Spark job monitoring.
* Proficient in Python for Data frames, RDDs, and Hive tables creation. Expertise in data extraction, transformations, and loading from HDFS, and Local FS to meet business needs, including HDFS and S3 data loading.
* Creating Spark scripts (Data frames, RDDs) in Spark 2.7 for complex business needs: transformations, joins, queries, analytics, and writing to Hive on HDFS.
* Creating data frames on Spark by importing data from HDFS and Oracle to develop data pipelines.
* Handling large datasets using partitions, Spark in Memory capabilities, Broadcasts in Spark, Effective & efficient Joins, Transformation, and others during the ingestion process itself.
* Creating Hive tables, preparing External tables, and loading them on HDFS from the local system
* Exporting of data from HDFS using Impala Shell with delimiter options and load into Teradata using BTEQ and TPT utilities
* Expertise in Data warehousing, Business Intelligence, Project Planning, Business Requirements, Technical Requirements, Design, Development, Testing, & Maintenance for client/server Data Warehouse.
* Defining data and tools standards, including tools for use in data acquisition and preparation, data modeling, metadata management, structured queries, and online analytical processing (OLAP) of both structured data and unstructured data.
* Analyze the Python code, which is used for machine learning algorithms, and Converted the Python code to Talend jobs to replicate the same functionality.
* Developed the NLP (Natural Language Processing) logic in Talend jobs. Validated the Results of the Python Code with that of the Talend jobs to ensure the accuracy of the functionality of the jobs that are converted from Python.
* Developed End-to-End data processing jobs in Talend which includes source data processing, data transformation, and data load to target. Also automated the entire process to load the data at regular intervals.
* Implementation of the IOT project which involved the Real-time processing of data from 20 different sensors using Pentaho ETL and loading the data to the database. We have set up the physical lab for real-time processing, and it’s been showcased globally, maintained, and running successfully for years.

**Technologies:** HDFS, Hive, Spark, Oracle, S3, Teradata (BTEQ and TPT), Bit Bucket, JIRA, Talend Big Data 7.2, Talend DI, Snowflake Cloud database, AWS Services (S3, EC2), MS SQL Server, Python

**Client: Samartha Info Solutions, Bangalore,India**

**Role:Senior Software Engineer 04/2013 - 12/2015**

**Responsibilities:**

* Developed Talend ETL jobs for transforming data from flat files to Staging databases.
* Conducted detailed analysis of flat files (Excel, Txt), improving data processing by 50%.
* Mapped data columns from files to a database, reducing data mapping errors by 30% through Talend ETL.
* Developing the required transformations to load the data from flat files to the tables.
* Developed the transformations to automate the data loading process reducing the manual effort of data loading from 7 days to 1 day.
* Played a pivotal role in Product Development, contributing to application design and application development.
* Experienced in the Deployment of data solutions, ensuring smooth integration and operation of different data sources.
* Adept at Operations management, ensuring efficient and reliable system performance.

**Technologies:** Talend DI, Oracle, MySQL, SVN, JIRA

**Client: Manipal Softwares, Bangalore,India**

**Role:Data Engineer Lead** **02/2010 - 03/2013**

**Responsibilities:**

* Created Pentaho ETL jobs, boosting data transformation efficiency by 40%.
* Integrated diverse bank apps (Trade Finance, FX, Private Banking, etc.), streamlining processes by 30%.
* Optimized Full Load strategy, reducing initial data setup time by 50%.
* Developed the Fraud Detection Engine using CEP Rules and Pentaho ETL Jobs.
* Designed, developed & tested daily Fraud Scenarios across applications, analyzing daily data patterns for precise fraud detection within hours.

**Technologies:** Pentaho DI, Java, HTML, CSS, SVN, JavaScript, Struts framework

**Client:IBM Softwares, Bangalore, India**

**Role:Technical Support Engineer** **04/2009 - 02/2010**

**Responsibilities:**

* Communicated with end-users to assess technical issues, determine root causes, and create actionable resolutions.
* Escalated unresolved issues to appropriate internal teams for resolution.
* Leveraged internal knowledge bases and resources to deliver accurate technical solutions.
* Diagnosed and troubleshot complex technical issues related to industry-specific environments.

**Education:**

* Bachelor of Engineering (Information Science) – 2008, Peoples Education Society University (PESU), Bangalore - India