**Chandra Bhaskar**

**Email:bhaskar4jha@gmail.com**

**Mob: 07667536182**

**09739430365**

**Lead Data Engineer**

Experienced Data Engineer with a focus on Cloud & big data. Having hands-on experience with Snowflake, Databricks, dbt, Azure, Python, Denodo, Talend, DataStage, Hadoop, Apache Spark, Hive, Sqoop, SQL  
Smart enough to get the high-level context, connect with all cross-functional partners like data scientists, engineers, and product owners and deliver the highly available & scalable data pipeline in a rapid manner. Dedicated and determined person who can adapt to any environment, also worked on delivering Multiple POC’s as well as internal projects with the changing technologies in Modern Data Stack. Building new practice related to cloud Data named as Cloud Data Factory taking multiple candidates interviews and mentoring & training juniors.

**Check out mine blogs published on company’s website:**

"Moved to the cloud, what next?" Click here - <https://hubs.ly/Q01m380x0>

"The Enterprise Strategy to Unlock Big Data Potential" Click here - <https://hubs.ly/Q01mvpP40>

"Modern Data Architecture: An Overview of Lambda and Kappa Architectures" Click here - <https://bit.ly/3LxKTKx>

**TOOLS & TECHNOLOGIES**

|  |  |
| --- | --- |
| **Cloud Data Warehouse/Lake House** | Snowflake, Databricks |
| **Big Data** | Sqoop, Hive, Hadoop, Spark, HDFS |
| **Programming Language** | Python, SQL |
| **Data Integration & ETL/ELT** | DataStage, Talend, dbt |
| **Data Replication** | IIDR (CDC) |
| **Database** | Oracle, MySQL, SQL Server |
| **Data Virtualization** | Denodo |
| **Scheduling** | Autosys, Oozie, Airflow, Cron |
| **VCS & Deployment** | GitHub, Azure Devops, Jenkins |
| **Container** | Docker |
| **Visualization** | Tableau, PowerBI |
| **Cloud** | Azure |
| **Operating System** | Linux, Windows |

**PROFESSIONAL EXPERIENCE**

**Lead Data Engineer– USEReady (April 2020 – Current)**

USE Ready is a Tableau Gold partner headquartered in NYC. USE Ready was recognized as “Partner of the year” by Tableau Software. USE Ready is also a Global Services alliance partner with Tableau and delivers several engagements on behalf of Tableau globally.

I’m currently working as Senior Data Engineer (ETL/Bigdata and Virtualization) and responsible for Consulting, Development of Data Integration and Datawarehouse projects for USE Ready’s clients.

**Key Project:**

**Client:** Aspire Lifestyles

**Title:** Campaign Management

**Role:** Lead Data Engineer

**Tools:** Talend, Snowflake, MoEngage, Bitbucket, Postman, Azure Blob

**Description:** The applications were developed to smooth the process of Campaign Management and Analytics for Booking Reminder alerts for Visa Customers under each category and welcome and Introduction Toolkit for Luxury Card Customers.

**Responsibilities:**

* Created data pipelines which includes Ingestion from Azure Blob, data loading to Snowflake, transformations using Snowflake Procedures as well as Talend, sending to MoEngage Bulk load Data API and design the job configurable so that any new clients can be included in no time.
* Understanding the KPIs and Requirements Documents.
* Creation solution Design document.
* Worked with business user to understand and create data model.
* Testing and validating the load in collaboration with Business Users.

**Key Project:**

**Client:** Aspire Lifestyles

**Title:** Design framework and Pipeline to Load EPC files till Datawarehouse and make data available to business users via access layer.

**Role:** Lead Data Engineer

**Tools:** Talend, Snowflake, Bitbucket, Azure Blob

**Description:** Design framework and Pipeline to Load EPC files till Datawarehouse and make data available to business users via access layer.

**Key Project:**

**Client:** Aspire Lifestyles

**Title:** SQL Server to Snowflake Migration

**Role:** Lead Data Engineer

**Tools:** Talend, Snowflake, Bitbucket, Azure Blob

**Description:** Migration of SQL Server Object to Snowflake.

**Client:** ABB

**Title:** Double Dip Invoice

**Role:** Senior Data Engineer

**Tools:** Denodo, Sqoop, Scala, Hadoop, Spark, Hive, Oozie, IntelliJ Idea

**Description:** The applications were developed to help the branch, complex and regional managers to detect Fraud for pricing analytics.

**Responsibilities:**

* Created data pipelines which includes Ingestion via Sqoop, data loading to Hadoop, transformations using spark with Scala, loading to hive, exposing tables to Denodo.
* Understanding the KPIs and Requirements Documents.
* Creation solution Design document.
* Worked with business user to understand and create data model.
* Testing and validating the ETL load in collaboration with Business Users.
* Installed Denodo components on Azure instance of windows server 2012.
* Created and implemented Security model on Denodo.
* Created and implemented caching model using impala.
* Establish and Enable SSL connection to and from Denodo to maintain higher security during data transfer.
* Integrating LDAP with Denodo for SSO.
* Enable https for web services component of Denodo like Data Catalog, Scheduler Administration tool as well as Diagnostic Monitoring tool.
* Installed Solution Manager component of Denodo and integrated all 3 environments (Dev/Qa/Prod).
* Created Deployment process for Denodo Objects using solution Manager for migrating to higher environments.
* Integrated Azure Devops with Denodo on VDP as well Solution Manger to maintain VCS.

**Prior Experience:**

**Ness Digital Engineering (Nov 2016 – March 2020)**

Ness Digital Engineering is an American [Israeli](https://en.wikipedia.org/wiki/State_of_Israel)-based provider of IT services. It provides information technology and business planning services in North America, Europe, the Middle East, and the Asia Pacific.

**Client: Canadian Tire Corporation (Mar2018-Mar2020)**

**Designation: Senior Data Engineer**

**Responsibilities:**

* Responsible for building scalable distributed data solutions using Hadoop.
* Used Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive.
* Experienced in performance tuning of Spark Applications for setting right Batch Interval time, correct level of Parallelism and memory tuning.
* Experienced in handling large datasets using Partitions, Spark in Memory capabilities, Broadcasts in Spark, Effective & efficient Joins, Transformations and other during ingestion process itself.
* Worked on Cluster of size 130 nodes, Analysed the SQL scripts and designed the solution to implement using PySpark
* Worked with various Hdfs file formats like Avro, parquet, Orc and sequence files.
* Developed Spark jobs to parse the JSON or XML data.
* Used Spark for interactive queries, processing of stream data and integration.
* Involved in converting Hive/SQL queries into Spark transformation using Spark API in Scala and python.
* Implemented task dependencies by writing python code for scheduling jobs using airflow.

**Client: Marks & FGL (Nov2016-Feb2018)**

**Designation: Bigdata Developer**

**Responsibilities:**

* Involved in Evaluation and analysis of Hadoop cluster and different big data analytic tools including pig, HBase, Sqoop, Flume, Spark
* Used Sqoop to extract the data back to relational database for business reporting.
* Worked with HiveQL on big data of logs to perform a trend analysis of user behavior on various online modules.
* Developed data pipeline using flume, Sqoop, Hive, Spark to ingest customer behavioral data and purchase history to Hdfs.
* Designed and created Hive external tables using shared meta-store instead of derby with partitioning, dynamic - partitioning and buckets.
* Used Impala for querying HDFS data to achieve better performance.
* Used Oozie workflow to co-ordinate Pig and hive scripts.
* Configured Sqoop Jobs to import data from RDBMS into HDFS using Oozie workflows.
* Involved in creating Hive Internal and External tables, loading data and writing hive queries, which will run internally in map, reduce way.
* Involved in all phases of the Big Data Implementation including requirement analysis, design, development, building, testing, and deployment in Hadoop cluster in fully distributed mode.
* Developed Hive jobs to transfer 8 years of bulk data from DB2, My SQL, Oracle to HDFS layer.
* Monitored workload, job performance and capacity planning using Cloudera Manager.

**Tesco Bank: (Nov2012-Oct2016)**

Tesco Bank is a British retail bank which was formed in July 1997, and which has been wholly owned by Tesco plc since 2008. The bank was formed as part of a 50:50 joint venture between The Royal Bank of Scotland and Tesco, the largest supermarket in the United Kingdom.

**Designation: Senior ETL and IIDR Developer**

**Responsibilities:**

* Understanding the functional requirements.
* Preparing the LLD based on Functional Requirements and Mapping Sheets.
* Develop ETL jobs for Interfaces using various stages like filter, aggregate, SCD, Change Capture, Lookup, Join, Sort etc.
* Implemented SCD Type 2 Using Change capture as well as SCD stage.
* Generated Surrogate key through transformer, Surrogate key generator as well as from Oracle Sequence.
* Worked with Multiple file stages like sequential stage, datasets, Complex Flat File stages, Xml stages etc.
* Unit testing and Integration testing of ETL jobs.
* Versioning and then Promotes the code to all higher environment (QA, Pre-Prod, Prod).
* Supporting defect fixes for QA, Pre-Prod and Prod Environments.

**Designation: DataStage and IIDR Admin**

**Responsibilities:**

* Installation and configuration of IBM Information Server and fix packs for both Data Stage and IIDR on Dev., Test, Pre-Prod and Prod servers.
* Created users, groups, credential mappings and assigned administrator and user roles, privileges in  
  Different environments.
* Configured Environment variables, security, and connectivity. Tuned the deployment for performance  
  And backed up the installation.
* Writing scripts to automate the day-to-day admin activities and scheduled it through cron.
* Added, deleted and setup Data Stage projects from Data stage administrator.
* Manage space allocation and usage on the Data Stage and CDC application Servers.
* Configured the server files uvobdc.confg and .odbc.ini files and setup TNS entries.
* Created Configuration files to increase parallelism of Data Stage jobs.
* Configuring password less ssh connection between the servers whenever needed in Project.
* Responsible for starting and stopping Data Stage and CDC services during maintenance window.
* Monitoring the complete Autosys Flow regularly.
* Handling the defects like Code Fixes, Performance tuning, Code Enhancements as well as adhoc change requests from business.
* Raising Change, Representing Change in CAB and get the approval from Change manager.
* Create Problem ticket, find out the root cause of problem.
* Resolving Incident and RF raised within SLA.

**Significant Achievements**

* Awarded with Star of Month in recognition of inspiring performance in development team at Tesco.
* Awarded with Star of Moment in recognition for extra ordinary performance improvement in travel batch 8hrs at Tesco.
* Awarded Tesco Value award for being a service superstar.
* Awarded Certificate of appreciation from SAP Labs.
* Awarded Spot Award at Ness from Canadian tire corporation towards the contribution in Data Integration Project.
* Awarded Spot award at Use Ready for contribution in ABB project.
* Received Quarterly Awards twice from CEO at USEReady for demonstrating excellence in delivery and deep customer centricity.

**CERTIFICATIONS**

|  |  |
| --- | --- |
| Hands on Snowflake – Web UI Essentials | Snowflake |
| Hands On Essentials - Data Warehouse | Snowflake |
| Hands On Essentials - Data Applications | Snowflake |
| Databricks Lakehouse Fundamentals | Databricks |
| dbt Fundamentals | Dbt Labs |
| Power BI Essential | LinkedIn |
| Microsoft Azure Fundamentals – AZ900 | Microsoft Azure |
| Microsoft Azure Data Fundamentals – DP900 | Microsoft Azure |
| Cloud Analytics Practitioner | Snowflake |
| Cloud Analytics Master | Snowflake |
| Cloud Analytics Architect | Snowflake |
| Cloud Analytics Academy – Executive Fast Track | Snowflake |
| Cloud Analytics Academy - Data Lake | Snowflake |
| Google Cloud Platform Fundamentals: Core Infrastructure | Google Cloud |
| Essential Cloud Infrastructure: Foundation | Google Cloud |
| Essential Cloud Infrastructure: Core Services | Google Cloud |
| Elastic Cloud Infrastructure: Scaling and Automation | Google Cloud |
| Elastic Cloud Infrastructure: Containers and Services | Google Cloud |
| Apache Spark2 with Scala | Udemy |
| SQL for Data Science | Data Camp |
| Python Programming Bootcamp | Udemy |
| Hands on Hadoop | Udemy |
| IBM Certified Solution Developer – Infosphere DataStage v8.5 | IBM |
| ITIL Foundation for Service Management Program Certified | QAI |

**EDUCATION**

|  |  |  |
| --- | --- | --- |
| **Degree** | **Institute** | **University** |
| B. E. in Computer Science and Engineering | BMS College of Engineering, Bangalore | VTU |