**Bharath Kumar Gattu**

**LOCATION: AUSTIN, TX-78613** mail:bharathg.dataengineer@gmail.com

phone:678-975-8207

**Professional Summary**

Overall, 10 Years of Software Development/Bug fix experience in Bigdata (Spark,Python, Pyspark,Hadoop, Hive and other Hadoop components),Azure, SQL Server,Snowflake and Teradata.

* Worked as a Developer in Information Management Delivery Team for IBM India Private Ltd.
* Experience in Hadoop Eco System and PySpark using Cloudera and Azure Data Bricks.
* Worked on most of the Azure components like Storage accounts, resource groups, Azure Data Factory (ADF), Azure Data Warehouse (ADW), Azure Synapse, Azure Databricks, Azure Datalake.
* Worked on different Data Ingestion tools like Azure Data Factory (ADF) and NIFI for ETL activities.
* Worked on different scheduling tools like Airflow, Jenkins and Control-M.
* Extensively worked on Spark (Pyspark) framework for Data Wrangling to implement different use cases based on the business requirement.
* Worked on Optimization techniques to save Azure cost.
* Confident and poised in interacting with individuals at all levels coupled with excellent coordination and liaison skills with the Management of the Client to accomplish the project
* Highly organized to work at various levels in a project.
* Having more than 6 years of experience with Banking Industry (Australia and New Zealand Bank client and IDFC First Bank)
* Excellent design, development and production support skills in Azure, Oracle, SQL Server, Teradata, and Unix applications
* Strong Data analysis and transformation skills
* Worked as a Module lead for Data Wrangling Project.
* Excellent time management and organizational skills with the ability to work on multiple projects and to meet business needs and timelines.
* Strong experience on Python and Cloud Technologies.

**Skills**

|  |  |
| --- | --- |
| * **Cloud Platform**: Microsoft Azure (ADF,Databricks,Synapse,Data Lake, Data Factory, SQL Data Warehouse),AWS,Snowflake. * **ETL Technologies**: IBM Infosphere DataStage, NIFI,Informatica,Azure Data Factory,Azure Synapse. * **IT Service Tools**: JIRA,GIT | * **Databases/DWH:** SQL Server, Oracle, Teradata,Snowflake * **Reporting Tools**: Power BI,Tableau,Cognos * **Programming Languages:** SQL, Python, Spark (Pyspark) * **Operating Systems:** Windows, Linux * **NoSQL:** Cassandra |

**Work History**

Senior Data Engineer: 02/2023-Till Date

AT&T-USA,Project Name: GIS Marketing ,Austin,Texas

* Working as a Senior Data Engineer Consultant at AT&T.
* Worked on End to End of the project.
* Working on a complete migration project from Teradata to Snowflake.
* Ingesting the batch data from different servers using Azure Data Factory to Azure Datalake.
* Implemented python framework function to ingest the data from Teradata to Databricks and from Databricks to Snowflake.
* Working in close partnership with cross-functional teams across the organization to define business rules and cleansing requirements for the data transformation process.
* Working on Data Enrichment from different source to Datalake for data analysis.
* Understand the Data Model and prepare the transformation logic to generate the data for business Insights.
* Extensively used on Azure Databricks,Data Lake, Azure Data Factory and Snowflake as a part of project requirements.

**Environment**: ADF , Pyspark, Python, Azure Stack, Databricks, Azure Data Lake,Delta Lake,Snowflake,Teradata and PowerBI.

Senior Data Engineer: 06/2022-01/2023

United Health Group(Optum), MN-USA,Project Name: DIGICOM & OnePaaS

* Worked as a Senior Data Engineer Consultant at UHG.
* Ingesting the batch data from different servers using Azure Data Factory to Azure Datalake.
* Extensively working on framework implementation using python to ingest batch/streaming data.
* Working on Data Enrichment from different source in Datalake for data analysis.
* Implemented De-Duplication process to avoid data skewness in the downstream.
* Worked on migration projects from Mapr clusters to Azure.
* Understand the Data Model and prepare the transformation logic to generate the data for business Insights.
* Extensively used on Azure Databricks,Data Lake and Azure Data Factory as a part of project requirements.
* Worked on Unity Catalog Migration.

**Environment**: ADF,Scala, Pyspark, Python, Azure Stack, Databricks, Azure Data Lake,Delta Lake,Terraform,Snowflake,Teradata and PowerBI.

Data Engineer: **11/2020-06/2022**

**IDFC FIRST BANK-Bangalore**

* Involved in end to end of the project architecture.
* Implemented different use cases as per business requirement from the data in the datalake.
* Involved in ingestion(ETL) process from different sources to Datalake using Nifi.
* Extensively worked with Pyspark on different usecases based on business requirement.
* Implemented tokenization on the sensitive bank data to push the data to AWS Cloud.
* Worked on detokenization from the AWS based on the business requirement.
* Involved in architecture discussions to optimize the solution.
* Worked on POC with comparing the performance on DBT queries and Spark Queries.

**Environment**: Nifi, Pyspark,Python,AWS, Airflow,S3, Nutanix-ObjectStore,DBT,Tableau

Senior Consultant-Analytics &Cognitive **10/2019 to 10/2020**

Deloitte India / Gulf Agency Company (Dubai) – Bangalore, India

* Involved an end to end of the project architecture.
* Implemented different use cases as per business requirement from the data in the data lake.
* Involved in implementing the data from on-premises database to Staging Layer, Incremental logic transformation in Azure
* Worked on different Ad hoc requirements of the project.
* Implemented the use cases with Azure Databricks.
* Optimized the code by using spark optimizations.
* Implemented the Optimize and Vacuum to save the costs from Azure.
* Implemented the metastore creation on the top of the Datalake files to avoid loading of data into DB.
* Used Azure Data Factory for ETL activities to load the data from on-premise to Data Lake.
* Used Azure Synapse for DWH implementation.

**Environment**: Azure Databricks, Pyspark, Azure Datalake,Azure Delta Lake,Azure CLI,Jenkins ,Synapse, other Azure components,Power BI, Grafana.

Module Lead-Digital**, 10/2018 to 10/2019**

Mindtree Limited/Procter & Gamble, USA and IMEA – Bangalore,India

* Understanding all the technical requirements in the project.
* Have used Big Data Platform (HDInsight cluster, Pyspark, Blob storage account, Data Factory v2.0)
* Source-To-Target Mapping (STTM) Owner.
* Played an important role in Data Wrangling.
* Preparing Technical Solution Document.
* Responsible for wrangling and send the data to Salesforce and PBI.
* Have used Pyspark for data wrangling in order to build the business logic as per the requirement.
* Have used Data Factory pipelines to transfer data from blob storage into SQL db of parquet format Using Pyspark scripts.

**Environment**: Azure HD Insights,Azure Databricks, Pyspark, ADF 2.0, Blob storage, Python.

Software Engineer: **12/2013 to 10/2018**

IBM India Pvt Ltd/ANZ Bank, Australia – Bangalore, India

* Ingest the data from different source system into HDFS
* Data from all sources was stored and processed in Spark.
* Designed a MDM application for Data validation, Data filtration.
* Developed the scripts for Data Transformation, Data Validation and Attribute Validation.
* Created Hive tables for the storage of processed data.
* Worked on different layers of the project for Data Transformation, Data Enrichment and Data Validation.
* Implemented the CDC logic using Pyspark and validated the data.
* Have used Data Frames to implement the logic for CDC in Pyspark.

**Environment**: Microsoft Azure,Hadoop HDP 2.4, Pyspark, HDFS, Hive,Python,Informatica,Teradata

**Education**

Bachelor of Technology: Electronics and Communication Engineering, 04/2011

SKD College of Engineering– India

Master of Engineering: Electronics and Communication Engineering, 05/2013

PSG College of Technology – India