Gayatri A blue and white sign

Description automatically generated A blue and white sign

Description automatically generated

Senior Data Engineer

gayatri.azurede@gmail.com

**Employer-details:[Globexus963@gmail.com](mailto:Globexus963@gmail.com)Mobile :** 732-427-8156

**PROFESSIONAL SUMMARY:**

* 14 years of development IT experience.
* Thorough understanding of Cloud Architecture, especially the Azure implementation
* Vast exposure in analyzing the application bottlenecks or issues and proposing resolutions to eliminate it.
* Hands-on experience in understanding and implementing Dataware house/Data Mart projects and end-to-end data integration pipelines.
* Strong in documenting the architecture design/use cases/solution recommendations.
* Worked in building application platforms in the cloud by leveraging **Azure, Databricks,** and open-source technologies and with **best engineering practices**.
* Created Pipelines in ADF using Datasets/Pipeline/ to Extract, Transform, and load data from different sources like Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool and backward.
* Hands-on experience in Installing and configuring Hadoop components.
* Install and maintain the native Hadoop Cluster and Cloudera Manager Cluster.
* Hands-on experience in writing Map Reduce jobs in Java, Pig, Hive.
* Hadoop integration with Business intelligence– Informatica Cloud, MSBI
* Created Oozie workflow for automating the process using Jenkins and Autosys.
* Imported and exported data into HDFS using Sqoop from various like Teradata, Oracle, Netezza.
* Involved in creating reusable components (for transformations and aggregations) like Universal key components, Data Quality, Data Integrity,using some frameworks like PIG, HIVE, MapReduce, and Spark.

**EDUCATION**

* B.Tech in Computer Science(OU)
* Master’s in computer science

**Certifications**

* [Microsoft Certified: Azure Data Engineer Associate](https://docs.microsoft.com/en-us/learn/certifications/azure-data-engineer)
* [Microsoft Certified: Azure Data Scientist](https://docs.microsoft.com/en-us/learn/certifications/azure-data-engineer)
* Oracle Certified Professional Java SE5 Programmer (OCPJP)
* ATG Certified Professional

**Technical Skills:**

**Programming Language:** PL/SQL, Python, Java,COBOL

**Data Processing:** Spark (SQL, Streaming, Spark API)

**Datawarehouse:** Hive, Snowflake, Synapse, Delta lake, DataMart

**Distributions/Cloud:** Spring boot,Apache Hadoop, Cloudera, AWS EMR, Databricks

**Container Orchestration:** Docker,

**ETL:** ADF, INFORMATICA, Sqoop, Data Stage

**Cloud:** AWS S3, AWS EMR, Azure Databricks, Azure Blob

**RDBMS/Storage/NoSQL/Search:** Oracle, MySQL, DB2, SQL Server

**Source Control:** Git/Bitbucket/Git lab

**Build/Other Tools:** Tomcat,Maven, Gradle, SBT, Jenkins Eclipse,VS Code,IntelliJ

**Operating System:** Windows, Ubuntu, CentOs, RedHat.

**Visualization Tools:** Power BI , Tableau

**PROFESSIONAL EXPERIENCE:**

**Client: McKee Risk Management Insurance Oct’2022 – Oct’2023**

**Lead Cloud Data Engineer**

**Project: Biz Dynamics Cloud Product**

**Responsibilities:**

* Writing Pyspark programs to load, clean and transform data using Databricks.
* Extensively worked on Spark execution tuning to improve the performance of the complete application.
* Involved in designing pipeline Migrating from on-prem to Azure.
* Build and Configurable Spark ETL components apply transformations and business rules for the consolidated dataset.
* Snowflake is used in staging, transforming, and loading data into BI tools and dashboard.
* Informatica to stream data using CDC form source tables in real times integrations.
* Tables on snowflake are updated by informatica in 3 phases Initial load, incremental and updates on snowflake data
* Initial load is performed only once when the job is triggered for the first time.
* Customized Snowflake Bulk Load and Snow pipe component Insert Dataset into Snowflake.
* Configured and launched the jobs from Control-M.

**Technologies:** PySpark, Spark SQL, Azure Data Factory, Azure Data Lake, Azure Databricks, Snowflake, Power BI

**Client: P&G Oct’2021–Oct’2022**

**Cloud Data Engineer**

**Project: Supply Chain Transformation**

**Responsibilities:**

* Designing and writing Pyspark code for creating end to end pipelines to Load, Extract and Transform Data.
* Performed duties as onsite lead and offshore coordination activities.
* Performed duties of Data Analyst / Business Analyst role by gathering requirements and conducting Data analysis by inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.
* Involved in Business Users discussions while gathering the requirements and analyzing the requirements to prepare the High-level Design Documents and backlog Stories.
* Create Stories for Backlog in Agile Model.
* Analyze current Systems used by client and involved in designing Data Model with data architect.
* Design EDW(Datawarehouse) Model using Star Schema.
* Involved in Developing reports using Power BI and Report Builder.

**Technologies:** PySpark, Power BI, SSIS, SSRS, Azure Blob, Snowflake, Azure Data Factory, Azure Data Lake, Azure Databricks,DataMart

**Client: Care Source Nov’2020 – Sep’2021**

**Cloud Data Engineer**

**Project: Data Modern Platform**

**Responsibilities:**

* Key responsibilities involved building ETL pipelines using ADF and Azure DataBricks.
* Migrating from On-prem to azure Cloud, involved and collaboratively working in the deep dive sessions with different teams to cloud enablement.
* Created Pipelines in **ADF** using **Linked Services**/**Datasets**/**Pipeline**/ to Extract, Transform and load data from different sources like **Azure SQL, Blob storage, Azure SQL Data warehouse**,DataMarts write-back tool and backwards.
* Build event messages real-time data pipelines, integrate with Azure data services
* Leveraging Azure Synapse to build analytics ML model data pipeline
* Azure Bot Service and Bot Framework provide tools to build, test, deploy, and manage intelligent bots, all in one place.

**Technologies:** PySpark, Spark SQL, Kafka and Hive, Mongo DB, Java, Azure Blob, Azure Synapse, Azure Cosmos DB, Azure Data Factory, Azure Data Lake, Azure Databricks, Azure Event Hubs Azure AI, Azure ML, Tableau

**Healthgrades, Austin, TX October 2019 – October 2020**

**Data Engineer**

**Project: CRM - AI/CPI Patient Analysis**

**Technical Environment:** Databricks 7.3, Spark/Scala/Python, PySpark, Spark Streamin, Kafka, Rest API, AWS EMR, AWS Lambda, AWS Step Function, AWS S3, Snowflake, Airflow

**Responsibilities:**

* Design and develop application and effectively contributing end to end Pipeline development.
* Business needs or requirement discussions with the Business users to come up with effective architectural design and timelines and managing the development team.
* Create development builds and deployable and helped in setup the environments like Production, Development, Integration, etc. Worked on creating and managing Pipelines in the Databricks Environment
* Extensively work on Spark execution tuning to improve the performance of complete application.
* Develop various security utilities and generic utilities to launch jobs for Impala, Spark, etc from API.
* Jobs Orchestration using Airflow

**Optum Platform, MN (for Optum Health) April 2017 - October 2019**

**Project: Optum Platform Development**

**Technical Environment:** MapR, Hadoop, HDFS, HBase, Apache Spark 1.6, Hive, Pig, Scala, Sqoop, SQL, Eclipse, Zookeeper, Java, JAXB, Scala 2.10, Talend 5.1, Mark Logic, Oozie, Splunk, AWS Redshift, AWS Snow ball, AWS Glue, AWS EMR

**Responsibilities:**

* Architectural discussions, design, implementation, etc. Explored or did POC’s to suggest on the right approach to Business team, Clients and development teams.
* Business needs or requirement discussions with the Business users to come up with effective architectural design and timelines. Interacted with management on the goals and its tracking. Discussion with Clients (ember clients) and their needs.
* Managed and guided the development team along with few implementations and helping them on POC’s like below Spark session cluster aware inside the docker container and providing impersonation for the users logged in based on their roles using TTL cache.
* Controllers, etc. Wrote many generic or reusable components along with logging wrapper generic exception handling, etc.
* Tested and validated various integrations to cloud (AWS) and big data distributions (MapR, Hortonworks).
* Developed a data pipeline using Kafka and Storm to store data into HDFS.
* Uploaded streaming data from Kafka to HDFS, HBase and Hive by integrating with storm.
* Integrated Spark with Mark Logic for MLCP jobs to push the documents. Used Splunk to visualization of Dashboards based on MLCP Logs. Used Talend to integrate with Hive/Map Reduce Jobs to Schedule Jobs.

**Client : NGM Insurance**

**Valumomentum Hyd Nov’2014 – Jan’2015**

**Big Data developer**

**Project: Claims Administration System** .

**Responsibilities**

* Working in an Agile - Scrum environment accomplish stories on time
* Understanding and participating in requirement gathering sessions with business and providing inputs to Business on new data sources systems.
* Prepare High Level Design document collaboratively work with groups
* Participating in design decisions, technology/tool decision, Sprint planning and product backlog grooming sessions.
* Working with different Hive file formats like Text file, Sequence file, ORC file, Parquet and Avro to analyze the data to build data model

**Environment:** HDP 2.6, HBase,Azure Services, IBM DB2 BigSQL, Hive, Pig, IBM TWS , IntelliJ, Java, Scala, Spring Boot,Spring MVC,AngularJS,JAX-RS,HTML 5.0,BootStrap

**Client : UHG July’ 2011 – Jun’2014**

**Senior Consultant**

**Hyderabad ,CSC**

**Responsibilities**

* Developed and Implement HBase capabilities for Big de-normalized data set and then apply a transformation on the de-normalized data set using hive.
* ELDM and ECMM Jobs developed using Spark RDD and Data Frame
* Developed a data pipeline using Kafka and Storm to store data into HDFS.
* Developed UDF using PIG and Hive depending on the complexity of input data.
* Using Splunk to visualization of Dashboards based on MLCP Logs
* Using Talend to integrate with Hive/Map Reduce Jobs to Schedule Jobs
* Developed UDF using PIG and Hive depending on the complexity of input data.
* Unit Testing with Hive and Pig Scripts Apple to Apple testing

**Environment:** MapR, Hadoop, HDFS, HBase, Apache Spark 1.6, Hive, Pig, Scala, Sqoop, SQL,

Eclipse, Zookeeper, Java, JAXB, Scala 2.10, Talend 5.1, Mark Logic, Oozie, Splunk

**Client : Erie Insurance Dec’2009 – Jul’2011**

**Sr. Sofrware Engineer**

**Hyderabad/Erie PA (Onsite) ,Valuemomentum**

**Project: Claims Administration System**

**Responsibilities**

* Working in an Agile - Scrum environment accomplish stories on time
* Understanding and participating in requirement gathering sessions with business and providing inputs to Business on new data sources systems.
* Prepare High Level Design document collaboratively work with groups
* Participating in design decisions, technology/tool decision, Sprint planning and product backlog grooming sessions.
* Development of Web application follows MVC pattern internal framework to implement the Controller layer and to assist with rendering the View.
* Developing and coding J2EE Components with JSP, Java Beans, business objects with EJBs and Servlets.
* Implement Hibernate as the ORM tool to communicate with the database.
* Involve in writing session beans, message driven beans and hibernate mapping files and hibernate configuration files.
* Involve in Bug fixing and functionality enhancements.

**Environment:**

Java, Dozer Mapping, Spring MVC, Spring Boot, JAX-WS, JAX-RS, JAXB, WAS 8.0, Jasper Reports, Apache Tomcat,IBM Portal, amCharts, AWS, Oracle 10g, SQL Server, MS BI, SSIS, AngularJS, HTML5

**Client : Kodak Jan 2008 – Dec 2009**

**Java/JEE Consultant**

**Hyderabad, Wipro**

**Project :** Kodak Gallery & Store

**Responsibilities**

* Involve in Requirements Gathering, Analysis, Development and Documentation.
* Development of Web application follows MVC pattern internal framework to implement the Controller layer and to assist with rendering the View.
* Developing and coding J2EE Components with JSP, Java Beans, business objects with EJBs and Servlets.
* Development of front-end validations using JavaScript.
* Implement Hibernate as the ORM tool to communicate with the database.
* Involve in writing session beans, message driven beans and hibernate mapping files and hibernate configuration files.
* Involve in Bug fixing and functionality enhancements.
* Develop Web pages using JSF, JSP and Ajax.

**Environment:**

Struts, JSP, Servlets, HTML, CSS, Ajax, EJB, XSLT, XML, XSD, JMS, Castor, Oracle 9i, Sun Solaris2.11, Weblogic 9, TIBCO, Eclipse