# TRILOK KUMAR TELLA

# Marker with solid fill Cincinnati, OH| +1(513)-568-3484 | [triloktella@gmail.com](mailto:triloktella@gmail.com)

# LinkedIn: <http://linkedin.com/in/tella-trilok-81aa47196>

**EDUCATION**

**University of Cincinnati**, Cincinnati, Ohio **GPA 3.9/4.0** Master’s in information technology **DECEMBER2018 VR, Siddhartha Engineering College,** Vijayawada  **GPA3.34/4.0** Bachelors in Electronics and communication **DECEMBER 2014**

**PROFILE SUMMARY:**

* **9+** years of IT professional experience as **Software developer** with core expertise in building **Bigdata** and **Hadoop** data pipelines.
* Over 4+ Years of **Big Data** experience in building highly scalable data analytics applications.
* Experience in working with Hadoop components like **HDFS**, **Map Reduce**, **Hive**, **HBase**, **Sqoop**, **Oozie**, **Spark**, **Kafka**.
* Strong understanding of **Distributed systems** design, **HDFS** architecture, internal working details of **MapReduce** and **Spark** processing frameworks.
* Solid experience developing **Spark** Applications for performing high scalable data transformations using RDD, **Dataframe** and **Spark-SQL**.
* Good hands-on experiencing working with various Hadoop distributions mainly Cloudera (**CDH**), Hortonworks (**HDP**) and Amazon **EMR**.
* Expertise in developing production ready Spark applications utilizing **Spark-Core**, **Dataframes**, **Spark-SQL**, **Spark-ML** and **Spark-Streaming** API's.
* Strong experience troubleshooting failures in **spark** applications and fine-tuning **spark** applications and **hive** queries for better performance.
* Good experience utilizing various optimization options in Spark like **broadcast joins**, **caching** (persisting), sizing executors appropriately, reducing **shuffle** stages etc.,
* Worked extensively on **Hive** for building complex data analytical applications.
* Strong experience writing complex map-reduce jobs including development of custom Input Formats and custom Record Readers.
* Good experience working with AWS Cloud services like **S3, EMR, Redshift, Athena, Glue,** etc.,
* Deep understanding of performance tuning, partitioning for optimizing spark applications.
* Worked on building real time data workflows using Kafka, Spark streaming and HBase.
* Extensive knowledge on NoSQL databases like HBase, Cassandra and Mongo DB.
* Solid experience in working with csv, text, avro, parquet, orc, json formats of data.
* Extensive experience in performing ETL on structured, semi-structured data using Pig Latin Scripts.
* Strong understanding of Data Modelling and experience with Data Cleansing, Data Profiling and Data analysis.
* Involved in database design, creating Tables, Views, Stored Procedures, Functions, Triggers, and Indexes.
* Good Knowledge about scalable, secure cloud architecture based on Amazon Web Services (leveraging AWS cloud services: EC2, Cloud Formation, VPC, S3, etc.
* Good knowledge in the core concepts of programming such as algorithms, data structures, collections.
* Good understanding of Service Oriented architecture (SOA) and web services like XML, XSD, XSDL, and SOAP.
* Analytical, organized, and enthusiastic to work in a fast paced and team-oriented environment.
* Expertise in interacting with business users and understanding the requirement and providing solutions to match their requirement.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| Hadoop/Big Data: | * **Spark, Hive, HDFS, MapReduce, Sqoop, Oozie, Kafka, Impala, Data Bricks, ADF, AzureSynapseAnalytics,AzureHDInsight,SQL,Python,ApacheSpark,DataModeling,ETLProcesses,Data Warehousing,Power BI** |
| Programming languages: | **Java, Scala, Python** |
| Cloud | **AWS-EC2, S3, EMR, RDS, Lambda, Redshift, Athena, Glue Metastore** |
| Database: | **NoSQL (HBase, Cassandra, MongoDB), Teradata, Oracle, SQL Server, MySQL, PostgresSQL** |
| IDE Tools: | **Eclipse, Intellij, PyCharm, VS code** |
| Development Approach: | **Agile, Waterfall** |
| Version Control: | **CVS, SVN, Git** |
| Reporting Tools: | **Tableau, QlikView** |

**PROFESSIONAL EXPERIENCE**: MARCH, 2022-CURRENT

**Client: Cigna, Bloomfield, CT**

**Data Engineer**

**Responsibilities:**

* Designed and implemented complete Azure data solutions from start to finish, including data ingestion, transformation, and storage utilizing Azure Data Factory, Azure Databricks, and Azure Data Lake Storage.
* To combine data from many sources and guarantee data quality and correctness, extensive ETL workflows, data pipelines, and data transformation algorithms were developed.
* Utilized Azure Event Hubs and Azure Stream Analytics to implement real-time streaming solutions for processing and analyzing massive data streams.
* Developed data models and optimized data architectures for effective data processing and analytics in collaboration with cross-functional teams.
* Power BI data visualizations were built and maintained to give stakeholders reports and insights they could act on.
* To ensure compliance with legal requirements, access control, encryption, and data masking were used as data security and governance methods.
* performed performance tuning and process optimization on data pipelines and systems to increase throughput and decrease latency.
* Utilizing Azure DevOps and Azure Monitor, I worked with DevOps teams to automate the deployment and monitoring of data engineering solutions.

**Client: Sysco, Houston, TX** July/2021-March/2022

**Data Engineer**

**Responsibilities:**

* Designed and developed data loading strategies, transformation for business to analyze the datasets.
* Processed **flat** **files** in various **file formats** and stored them as in various partition models in **S3**.
* Responsible for **building, develop, testing** reusable spark components that will be used across modules.
* Responsible in performing sort, **join, aggregations, filter**, and other transformations on the **datasets** using **Spark.**
* Involved in building a **linear** regression model using Spark ML for predicting purchase propensity.
* Responsible in implementing advanced procedures like **text analytics** and processing using the in-memory computing capabilities like **Spark**.
* Experience in extracting appropriate features from data sets to handle **bad, null, partial records** using **Spark SQL**.
* Collected data using Spark Streaming from **AWS S3** bucket in near-real-time and performs necessary Transformations and Aggregations to build the data model.
* Involved in converting **Hive/SQL** queries into Spark transformations using Spark Dataframes and Spark SQL.
* Expertized in implementing Spark using **Spark SQL** for faster testing and processing of data responsible to manage data from different sources.
* Using **spark - Cassandra** connector to load data to and from Cassandra.
* Experience in building Real-time **Data Pipelines** with **Kafka** Connect and Spark Streaming.
* Responsible in development of **Spark Cassandra** connector to load data from flat file to **Cassandra** for analysis.
* Responsible in creating **Hive tables**, loading with data, and writing **Hive** queries.
* Developed end to end data processing pipelines that begin with receiving data using distributed messaging systems **Kafka** for persisting data into **Cassandra**
* Responsible in creating **mappings** and **workflows** to extract and load data from relational **databases**, flat file sources and legacy systems using **Talend**.
* Developed and designed ETL Jobs using**Talend** Integration Suite in Talend 5.2.2
* Involved in creating **External Hive tables** and involved in data loading and writing **Hive UDFs**.
* Experience in using various **compression** techniques like **Snappy Codec, Lzo, and Gzip**to save data and optimize data transfer over network using **Avro**, **Parquet**.

**Client: Nordstrom, Seattle, WA** January, 2020 - June, 2021

**Hadoop Developer**

**Responsibilities:**

* Ingested user behavioral data from external servers such as FTP server and S3 buckets on daily basis using **custom Input Adapters**.
* Created **Sqoop** scripts to import/export user profile data from RDBMS to **S3** Data Lake.
* Developed various **spark** applications using Scala to perform various enrichments of user behavioral data (click stream data) merged with user profile data.
* Involved in data **cleansing**, event **enrichment**, data **aggregation**, **de**-**normalization** and data preparation needed for downstream model learning and reporting.
* Utilized Spark Scala API to implement batch processing of jobs
* **Troubleshooting** Spark applications for improved error tolerance.
* **Fine-tuning** spark applications/jobs to improve the efficiency and overall processing time for the pipelines
* Created **Kafka** producer API to send live-stream data into various Kafka topics.
* Developed **Spark-Streaming** applications to consume the data from Kafka topics and to insert the processed streams to **HBase**.
* Utilized Spark in Memory capabilities, to handle large datasets.
* Used broadcast variables in spark, effective & efficient Joins, transformations, and other capabilities for data processing.
* Experienced in working with **EMR** cluster and **S3** in AWS cloud.
* Creating Hive tables, loading and analyzing data using hive scripts. Implemented Partitioning, Dynamic Partitions, Buckets in **Hive**.
* Involved in continuous Integration of application using **Jenkins**.

***Environnent****:* AWS EMR, Spark, Hive, HDFS, Sqoop, Kafka, Oozie, HBase, Scala, MapReduce.

**Citi Group, India** January, 2013 – December,2017

**Hadoop Developer**

**Responsibilities:**

* Experience in Importing and exporting data into HDFS and **Hive** using **Sqoop.**
* Developed **Flume Agents**for loading and filtering the streaming data into HDFS.
* Moving Bulk amount data into **HBase** using Map Reduce Integration.
* Developed **Map-Reduce** programs to clean and aggregate the data
* Developed HBase data model on top of HDFS data to perform real time analytics using Java API.
* Developed different kind of custom filters and handled pre-defined filters on HBase data using API.
* Strong understanding of Hadoop eco system such as HDFS, MapReduce, HBase, Zookeeper, Pig, Hadoop streaming, Sqoop, Oozie and Hive
* Implement counters on HBase data to count total records on different tables.
* Experienced in handling **Avro**data files by passing schema into HDFS using Avro tools and Map Reduce.
* Worked on custom Pig Loaders and Storage classes to work with a variety of data formats such as JSON, Compressed CSV, etc.
* Implemented Secondary sorting to sort reducer output globally in map reduce.
* Created Hive Dynamic partitions to load time series data.
* Experienced in handling different types of joins in **Hive** like Map joins, bucker map joins, sorted bucket map joins.
* Created tables, partitions, buckets and perform analytics using **Hive**ad-hoc queries.
* Experienced import/export data into HDFS/Hive from relational data base and Tera data using **Sqoop**.
* Handling continuous streaming data comes from different sources using **flume**and set destination as HDFS.
* Experience with**CDH** distribution and Cloudera Manager to manage and monitor Hadoop clusters.

**Environment**: Hadoop, HDFS, Map Reduce, Hive, Pig, HBase, Sqoop, RDBMS/DB, Flat files, MySQL, CSV, Avro data files.