



# TARUN TEJA S.P - DATA ENGINEER

[TARUNSTEJA@GMAIL.COM](mailto:TARUNSTEJA@GMAIL.COM) +91-8639793435

## PROFILE SUMMARY

- Overall 3 years of experience in working with data applications using big data frameworks and data science.
- Experience of working on a project end-to-end: Problem scoping, data gathering, delivering advanced analytics solutions and insights.
- Collaborate with business and data owners to formulate problems whose solution are impactful to the business. Ability to set and meet project objectives & milestones.

## EDUCATION

### BACHELOR OF TECHNOLOGY

Tirumala Engineering College, Hyderabad (2011-15)

## WORK EXPERIENCE

**Data Engineer** – Accenture AI, Hyderabad

Oct 2017 – Till date

## TECHNICAL SKILLS

- Machine Learning – Regression, Classification
- Exploratory Data Analysis
- CAR creation
- Azure Databricks
- Certified GCP Data Engineer
- Big Data Stack: Hadoop, Spark, Hive, Impala, Sqoop
- Unified-Analytics Platform Databricks

## NON-TECHNICAL SKILLS

- Analytical Problem solving
- Curiosity Driven
- Automation
- Team Player
- Communication skills

## TOOLS

- Java
- Python
- Shell

## KEY PROJECTS

### Building Unified Data Layer

This is health care project where we are integrating silo data layers as unified data layer. The aim is to integrate different data sources to common layer there by increasing the velocity of analytics work loads and generate reports at faster phase.

<b>Customer</b>	Novartis
<b>Duration</b>	Ongoing
<b>Team Size</b>	3
<b>Location</b>	Hyderabad
<b>Technology/Tools</b>	pySpark, Python, AWS Databricks
<b>Activities</b>	<ul style="list-style-type: none"><li>• Perform data assessment for source and generate report.</li><li>• Able to do ingestion from source to Raw layer of s3 bucket.</li><li>• Part of Data modeling team.</li><li>• Was also assisting data Testing team in test validations</li><li>• Actively part of discussions with different business owners in building data model</li><li>• Perform API Validations using Postman and python</li><li>• Worked on aws lambda function</li></ul>

### Theft Identification

This is a utility project where we need to identify the energy theft cases using consumer data over 2 years. The aim is to identify patterns which indicate if a person is likely to theft or not, which may be used for taking actions such immediate disconnection of supply.

<b>Customer</b>	DTE (Detroit Energy USA)
<b>Duration</b>	4 months
<b>Team Size</b>	6
<b>Location</b>	Hyderabad
<b>Technology/Tools</b>	pySpark, Python, Impala, Azure Databricks, EDA
<b>Activities</b>	<ul style="list-style-type: none"><li>• Understand the concepts Theft identification.</li><li>• Data preparation by merging different source of data.</li><li>• Performed EDA on different types of data.</li><li>• Data Assessment in ADLS and Impala</li><li>• Developed SQL's in impala for various hypothesis</li><li>• Took the ownership of entire databricks notebooks and memory allocations.</li><li>• Ingesting large size csv files from impala shell and vice versa for analytics</li></ul>

## Price Recommendation

Deriving price recommendation engine by analyzing past 3 years data and provide best possible offers respective to seasons for cruise hotels

**Customer** Royal Caribbean Cruise Pvt Ltd, USA

**Duration** 8 months

**Team Size** 9

**Location** Hyderabad

**Technology/Tools** Pyspark, Python, Azure Databricks, Java

**Activities**

- Perform Data Analysis
- Data pipeline creation from source to destination using Azure databricks
- Data Factory
- pyspark
- Developed API using AWS Server less framework

## Date Lake

Dealing with data standardization and validation steps before pushing it to Data lake. Performed Data Migration from Teradata to Amazon S3 and Alluxio as cache layer. Applied Spark and shell Scripting to standardized data.

**Customer** Development Bank Of Singapore

**Duration** 4 months

**Team Size** 15

**Location** Hyderabad

**Technology/Tools** Collibra, Airflow

**Activities**

- Involved in data requirement analysis
- Applied transformations and loaded data from different file formats.
- Used Sqoop get portion of data
- Involved in data standardization and Governance.

**Date Migration**

Data migration from traditional databases to bigdata ecosystems to improve processing capabilities and speed up business movement and analysis.

Customer	Axis Bank
Duration	6 months
Team Size	9
Location	Mumbai (On site)
Technology/Tools	Oracle, Informatica, ETL, Hive, Python
Activities	<ul style="list-style-type: none"><li>• Developed Spark SQL</li><li>• Migrated PL/sql code to Pyspark code for performance optimization</li><li>• Implemented informatica mappings in BDM</li></ul>

