

## Job Description

|                       |  |
|-----------------------|--|
| <b>Company</b>        | C5i                                      |
| <b>Position Title</b> | Data Engineer – Hadoop                   |
| <b>Reporting to</b>   | Manager / Director, Platforms & Solution |
| <b>Schedule</b>       | Full time                                |
| <b>Work Location</b>  | Bengaluru / Mumbai                       |

## Overview

### **C5i**

C5i is a pure-play AI & Analytics provider that combines the power of human perspective with AI technology to deliver trustworthy intelligence. The company drives value through a comprehensive solution set, integrating multifunctional teams that have technical and business domain expertise with a robust suite of products, solutions, and accelerators tailored for various horizontal and industry-specific use cases. At the core, C5i's focus is to deliver business impact at speed and scale by driving adoption of AI-assisted decision-making.

C5i caters to some of the world's largest enterprises, including many Fortune 500 companies. The company's clients span Technology, Media, and Telecom (TMT), Pharma & Lifesciences, CPG, Retail, Banking, and other sectors. C5i has been recognized by leading industry analysts like Gartner and Forrester for its Analytics and AI capabilities and proprietary AI-based platforms.

## Global offices

[United States](#) | [Canada](#) | [United Kingdom](#) | [United Arab of Emirates](#) | [India](#)

### **Job Summary:**

This role will be part of a dynamic organization that is continuously exploring and innovating in the areas of Augmented Analytics, Generative AI, Intelligent Automation around fascinating domains of Competitive Intelligence, Digital Experience Optimization, Conversation Rate Optimization, Market Insights, Supply Chain etc. across multiple industry verticals viz; Technology, Media, Telecom, Retail, CPG, Manufacturing and Lifesciences.

### **Job Responsibilities:**

- Design, develop, and maintain scalable data pipelines and systems for data processing.
- Utilize Hadoop and related technologies to manage large-scale data processing.
- Perform data ingestion using Kafka & spark, Sqoop and various file formats and process data into Hive using Beeline/spark.
- Develop and maintain shell scripts for automation of data processing tasks.
- Implement full and incremental data loading strategies to ensure data consistency and availability.
- Orchestrate and monitor workflows using Apache Airflow.
- Ensure code quality and version control using Git.
- Troubleshoot and resolve data-related issues in a timely manner.
- Stay up-to-date with the latest industry trends and technologies to continuously improve our data infrastructure

### **Requirements & Qualifications:**

- Proven experience as a Data Engineer (ETL, data warehousing).
- Strong knowledge of Hadoop and its ecosystem (HDFS, YARN, MapReduce, Tez and spark).
- Proficiency in Kafka & spark, Sqoop and Hive.
- Experience with shell scripting for automation.
- Expertise in full and incremental data loading techniques.
- Excellent problem-solving skills and attention to detail.
- Ability to work collaboratively in a team environment and communicate effectively with stakeholders.

**C5i** is proud to be an equal opportunity employer. We are committed to equal employment opportunity regardless of race, color, religion, sex, sexual orientation, age, marital status, disability, gender identity, etc. If you have a disability or special need that requires accommodation, please keep us informed about the same at the hiring stages for us to factor necessary accommodations.