Welcome to *HackSprint 26!*

Join us for a day of innovation and collaboration!

Date: 5th February, 2026

Time: 9:30 AM - 5:30 PM

Venue: 1st Floor, C-DAC Hyderabad

Welcome & Agenda

Dive into a dynamic day filled with coding, creativity, and teamwork!

Task 1: Data Analytics

10:00 AM – 11:00 AM

Task 2: RAG Application

11:15 AM – 1:00 PM

Lunch Break

1:00 PM – 2:00 PM

Task 3: Crawling & Classification

2:00 PM – 5:00 PM

Hackathon Guidelines

Team Structure

8 teams in total

Work collaboratively within your assigned team

Rules & Expectations

- Submit via GitHub repository
- Document your approach clearly
- Code must be original work.
- Prefer local LLM's over proprietary

Evaluation Metrics:

Field Completeness

Whether or not all the relevant fields are scraped.

Robustness

How well the system handles errors. It should not crash and must always produce a valid JSON output as specified.

Accuracy

The quality and correctness of the LLM-generated summary, category, and sentiment.

System Prompt Analysis

The prompt given to LLM/model for Analysis.

Choice of Stack

Assessment of the technologies and frameworks used.

Efficiency

The speed and resourcefulness of your data processing pipeline from crawl to final JSON.

Innovation

Creative approaches to solving the problem, such as advanced error handling, caching strategies, or additional data extractions (e.g., named entity recognition).

Task 1: Data Analytics

Time: 9:30 AM – 11:30 AM

Problem statement: Objective is to perform a comprehensive data analysis of a given dataset to discover meaningful insights and present them in an interpretable manner which includes the following steps:

Objectives:

1. Data Ingestion: Load the dataset, understand the structure, schema and content

2. Extracting key information: Analyze the data to identify significant patterns, trends, anomalies or relationships that provide value or support decision-making.

3. Data Visualization: Use visualization tools to present the insights for interpretation by both technical and non-technical persons.

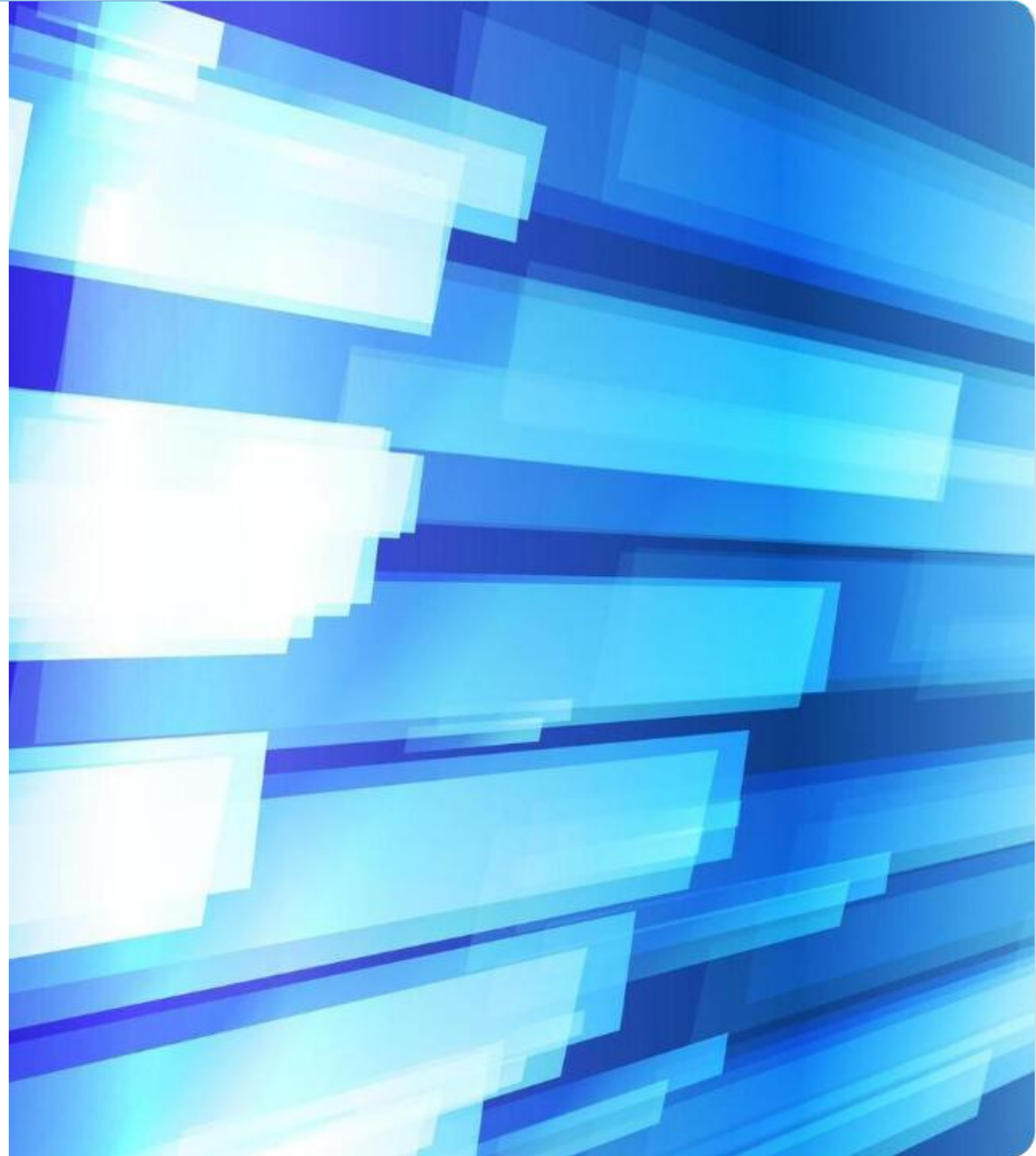
TASK 2 • 11:15 AM - 1:00 PM

RAG Application using multiple sources.

Build a context-aware RAG application that bridges the gap between generic LLMs and proprietary corporate data.

CORE OBJECTIVE

Create a unified pipeline to intelligently navigate **Structured Logs**, **Semi-Structured Events**, and **Unstructured Policies**.



Multi-Source Knowledge Base

The system must ingest and cross-reference these diverse data formats to provide grounded answers.

Structured Data

`CSV / Excel`

Employee Master: Role, Dept, Joining Date.
Leave Tracker: History & Balances.

Semi-Structured

`JSON Logs`

Attendance: Daily check-in/out timestamps.
Challenge: Identify "Incomplete" logs.

Unstructured

`PDF Manual`

HR Policy 2026: Rules for conduct, bonuses, and leave types.

Evaluation & Constraints

Your solution will be judged on its ability to handle complexity while maintaining strict accuracy.

- **Data Integrity:** Robustly handle varied date formats across CSV and Excel without errors.
- **Logic Reasoning:** Successfully cross-reference static rules (PDF) with dynamic data (Logs/CSV).
- **Hallucination Check:** Gracefully `REFUSE` to answer if information is not in the provided files.
- **Citations:** Every claim must cite a specific *"Section"* or *"Database"*.

TASK 03

// CHALLENGE STATEMENT

NewsStream AI

Real-Time News Intelligence Platform

RSS Ingestion | LLM Processing | RAG Retrieval | Live Analytics

System Architecture

01. INGEST

RSS Monitor

- Target: News RSS Feeds
- Extract: Full Text
- Mode: Continuous

02. PROCESS

LLM Core

- Summarize Articles
- Classify Category
- Analyze Sentiment

03. STORE

Knowledge Base

- Structured/Unstructured (MongoDB)
- Index for RAG

04. SERVE

User Interface

- Interactive Chatbot
- Real-time Dashboard
- JSON API Output

Core Objectives

Data Ingestion

Build a robust scraper to continuously monitor RSS feeds and handle extraction failures gracefully.

LLM Intelligence

Deploy a model to generate concise summaries and accurate sentiment tags for every article.

RAG Strategy

Implement search to allow the chatbot to retrieve context-aware answers.

Live Dashboard

Visualize data trends in real-time. Show sentiment breakdown and topic clustering.

Evaluation Criteria

Robustness & Validity

- Valid JSON output format
- Error handling (No crashes)
- Complete metadata extraction

System Quality

- RAG Retrieval Accuracy
- Summary conciseness
- Sentiment correctness

Engineering

- Low Latency Pipeline
- Chunking Strategy
- Prompt Optimization

Innovation

- Trend Analysis Features
- Dashboard UX/UI
- Topic Clustering