Excellent — splitting the work effectively is crucial to deliver the full project in 3 days **without burnout and with full marks for contribution**.

Here’s a balanced and efficient **task distribution for 4 team members**, aligning each person with:

* Their strengths (based on previous role assignments)
* A logical pipeline stage
* A measurable output for your final report/notebook

**✅**

**Task Distribution Plan (Team of 4)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Team Member** | **Assigned Module** | **Tasks** | **Deliverables** |
| **Yash Dubey** | **Data Ingestion & Preprocessing** | - Load all .plt files using Spark  - Parse files and create unified DataFrame  - Add user ID, traj ID, timestamp  - Assign zones to each point (grid-based)  - Optional: Resample trajectories | - Cleaned DataFrame  - Data loading code  - Documentation of grid/partition strategy |
| **Onkar Hemant Thete** | **Clustering Implementation** | - Implement DBSCAN in PySpark (or MLlib baseline)  - Implement simplified TRACLUS (partition + group)  - Tune parameters (epsilon, minPts)  - Run clustering per partition | - Clustered output  - Cluster visualizations  - Code and clustering insights |
| **Sanjeeva Srinivasa Rao** | **Similarity Measures & Partitioning** | - Implement Euclidean distance (baseline)  - Implement DTW using Pandas UDF  - Apply different partitioning strategies (region, user, time)  - Compare DTW vs Euclidean on runtime & accuracy | - Similarity functions  - Distance matrix or pairwise comparisons  - Partitioning performance table |
| **Jaswant Singh** | **Evaluation, Analysis & Reporting** | - Run runtime analysis (time.time() + Spark UI)  - Compute Silhouette Scores  - Collect results into tables/graphs  - Write evaluation + conclusion sections  - Format report (IEEE) & finalize notebook | - Final PDF report  - Results table  - Charts for poster/report |

**📋 Shared Tasks (Collaborate or Rotate)**

| **Task** | **Owner** |
| --- | --- |
| 📁 Folder setup, Git/Drive organization | Yash (initial setup) |
| 📊 Final result table/chart assembly | Jaswant (with inputs from all) |
| 🧪 Experiment coordination (what configs to run) | Sanjeeva + Onkar |
| 📄 Final Report Proofreading & Submission | Everyone reviews one section |

**🗓 Suggested 3-Day Schedule**

|  |  |  |
| --- | --- | --- |
| **Timeframe** | **Focus** | **Who Works On What** |
| **Day 1 Morning** | Data ingestion + parsing + DTW | Yash + Sanjeeva |
| **Day 1 Afternoon** | Clustering code (DBSCAN) + similarity eval | Onkar + Sanjeeva |
| **Day 2 Morning** | Run clustering experiments with partitioning | Everyone tests on different subsets |
| **Day 2 Afternoon** | Evaluation metrics, runtime, silhouette | Jaswant collects results |
| **Day 3 Full** | Report writing, notebook cleaning, backup visuals | Jaswant leads writeup; others supply figures/code |

**✅ Tips to Ensure Equal Contribution**

* Use a **shared Google Doc/Drive** for report editing
* Use **Git or Databricks Repos** to track who coded what
* Have everyone include a **contribution section** in the final report
* Take turns presenting findings or writing paragraphs

Would you like a ready-to-use **Google Sheet task tracker** or a **Gantt-style chart** to keep everyone aligned?