Nagar Yuwak Shikshan Sanstha's
Yeshwantrao Chavan College of Engineering
(An Autonomous Institution affiliated to Rashtrasant Tukadoji Maharaj Nagpur University)
Hingna Road, Wanadongri, Nagpur - 441 110
NAAC A++
Ph.: 07104-237919, 234623, 329249, 329250 Fax: 07104-232376, Website: www.ycce.edu

Department of Artificial Intelligence & Data Science

Vision of the Department

To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

Mission of the Department

To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.

Session 2025-2026

Vision: Dream of where you want.	**Mission:** Means to achieve Vision

Program Educational Objectives of the program (PEO): (broad statements that describe the professional and career accomplishments)

PEO1	**Preparation**	**P: Preparation**	**Pep-CL abbreviation pronounce as Pep-si-lL easy to recall**
PEO2	**Core Competence**	**E: Environment (Learning Environment)**	
PEO3	**Breadth**	**P: Professionalism**	
PEO4	**Professionalism**	**C: Core Competence**	
PEO5	**Learning Environment**	**L: Breadth (Learning in diverse areas)**	

Program Outcomes (PO):

1. Understand and Apply Parallel Programming Concepts
2. Analyse and Improve Program Performance.
3. Demonstrate Practical Skills in HPC Tools and Environments.

Keywords of POs:
Engineering knowledge, Problem analysis, Design/development of solutions, Conduct Investigations of Complex Problems, Engineering Tool Usage, The Engineer and The World, Ethics, Individual and Collaborative Team work, Communication, Project Management and Finance, Life-Long Learning

PSO Keywords: Cutting edge technologies, Research

"I am an engineer, and I know how to apply engineering knowledge to investigate, analyse and design solutions to complex problems using tools for entire world following all ethics in a collaborative way with proper management skills throughout my life." *to contribute to the development of cutting-edge technologies and Research*.

Integrity: I will adhere to the Laboratory Code of Conduct and ethics in its entirety.

Name and Signature of Student and Date
Sakshi Gokhale
04/11/25

Nagar Yuwak Shikshan Sanstha's
Yeshwantrao Chavan College of Engineering
(An Autonomous Institution affiliated to Rashtrasant Tukadoji Maharaj Nagpur University)
Hingna Road, Wanadongri, Nagpur - 441 110
NAAC A++
Ph.: 07104-237919, 234623, 329249, 329250 Fax: 07104-232376, Website: www.ycce.edu

Department of Artificial Intelligence & Data Science

Vision of the Department

To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

Mission of the Department

To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.

Session	2025-26 (ODD)	**Course Name**	HPC Lab
Semester	7 AIDS	**Course Code**	22ADS706
Roll No	16	**Name of Student**	Sakshi Gokhale

Practical Number	9 (Mini Project)
Course Outcome	1. Understand and Apply Parallel Programming Concepts 2. Analyse and Improve Program Performance
Aim	Performance Comparison of Data Serialization Formats
Problem Definition	Performance Comparison of Data Serialization Formats
Theory (100 words)	In High Performance Computing (HPC), huge volumes of structured numerical data are generated from simulations, scientific experiments, sensors, and analytical workloads. Efficient storage and fast data movement are essential to reduce execution time. Data serialization formats define how data gets converted into byte stream for storage or transfer. Different formats provide different performance characteristics. Text-based formats like CSV and JSON are human-readable but very slow in I/O operations and consume more disk space. Pickle is fast but it is Python specific and not suitable for cross-platform HPC environments. Parquet is a modern columnar binary format which stores data in compressed, vectorizable columns. This improves caching, minimizes storage size and supports parallel read operations. Therefore selection of the right serialization format is important for HPC pipelines. Parquet achieves high throughput and low memory usage, making it suitable for large scale data analysis and distributed computational systems.
Procedure and Execution (100 Words)	Code: import pandas as pd, numpy as np, time df = pd.DataFrame({ "A": np.random.rand(5_000_000), "B": np.random.rand(5_000_000), "C": np.random.rand(5_000_000)

Vision of the Department

To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

Mission of the Department

To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.

```
})

# write timings
for fmt, fn, save in [
   ("CSV","test.csv",lambda:df.to_csv("test.csv")),

("JSON","test.json",lambda:df.to_json("test.json",orient='records')),
   ("Pickle","test.pkl",lambda:df.to_pickle("test.pkl")),

("Parquet","test.parquet",lambda:df.to_parquet("test.parquet",engine='pyarrow'))
]:
   t=time.time(); save(); print(fmt,"write:",time.time()-t)

print("------READING------")

# read timings
for fmt, load in [
   ("CSV",lambda:pd.read_csv("test.csv")),
   ("JSON",lambda:pd.read_json("test.json")),
   ("Pickle",lambda:pd.read_pickle("test.pkl")),
   ("Parquet",lambda:pd.read_parquet("test.parquet"))
]:
   t=time.time(); load(); print(fmt,"read:",time.time()-t)
```

Output:

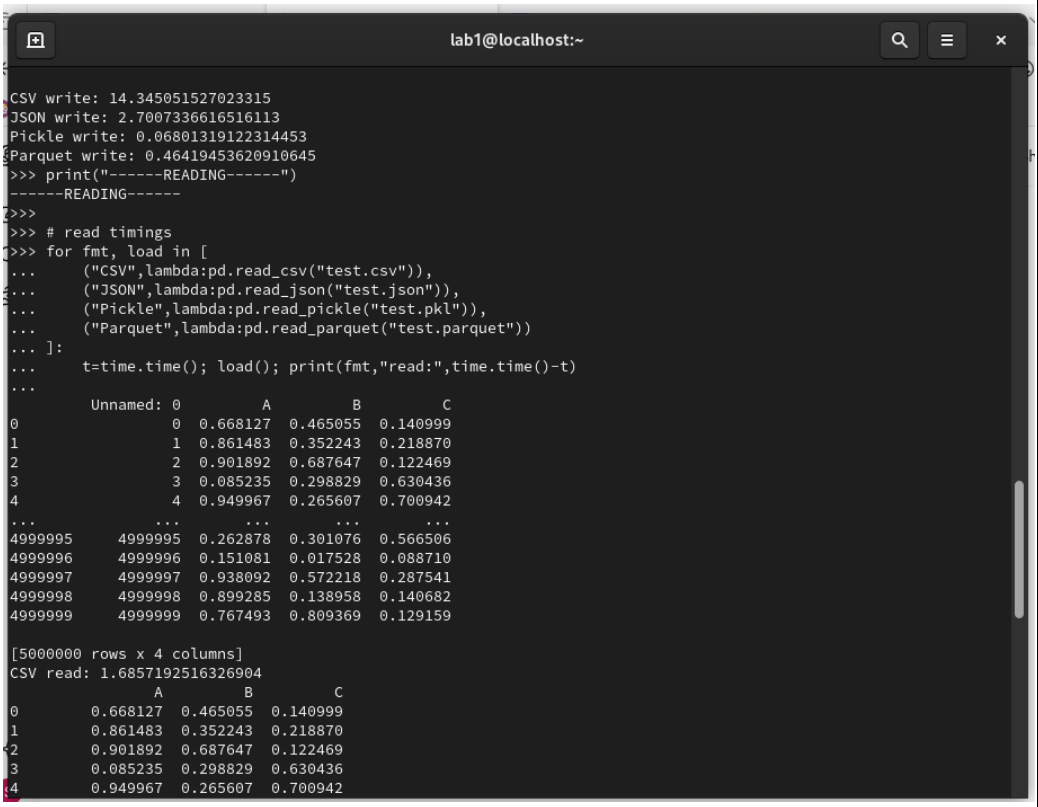
```
[lab1@localhost ~]$ python3 -m ensurepip --user
Looking in links: /tmp/tmpwjw5yd1k
Requirement already satisfied: setuptools in /usr/lib/python3.9/site-packages (53.0.0)
Processing /tmp/tmpwjw5yd1k/pip-21.3.1-py3-none-any.whl
Installing collected packages: pip
Successfully installed pip-21.3.1
[lab1@localhost ~]$ python3 -m pip install --user pandas pyarrow
Collecting pandas
  Downloading pandas-2.3.3-cp39-cp39-manylinux_2_24_x86_64.manylinux_2_28_x86_64.whl (12.8 MB)
     |████████████████████████████████| 12.8 MB 6.0 kB/s
Collecting pyarrow
  Downloading pyarrow-21.0.0-cp39-cp39-manylinux_2_28_x86_64.whl (42.7 MB)
     |████████████████████████████████| 42.7 MB 256 kB/s            ^[[24~
Collecting python-dateutil>=2.8.2
  Downloading python_dateutil-2.9.0.post0-py2.py3-none-any.whl (229 kB)
     |████████████████████████████████| 229 kB 271 kB/s
Collecting tzdata>=2022.7
  Downloading tzdata-2025.2-py2.py3-none-any.whl (347 kB)
     |████████████████████████████████| 347 kB 280 kB/s
Collecting pytz>=2020.1
  Downloading pytz-2025.2-py2.py3-none-any.whl (509 kB)
     |████████████████████████████████| 509 kB 250 kB/s
Collecting numpy>=1.22.4
  Downloading numpy-2.0.2-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (19.5 MB)
     |████████████████████████████████| 19.5 MB 14 kB/s
Requirement already satisfied: six>=1.5 in /usr/lib/python3.9/site-packages (from python-dateutil>=2.8.2->pandas) (1.15.0)
Installing collected packages: tzdata, pytz, python-dateutil, numpy, pyarrow, pandas
Successfully installed numpy-2.0.2 pandas-2.3.3 pyarrow-21.0.0 python-dateutil-2.9.0.post0 pytz-2025.2 tzdata-2025.2
```

Vision of the Department

To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

Mission of the Department

To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.

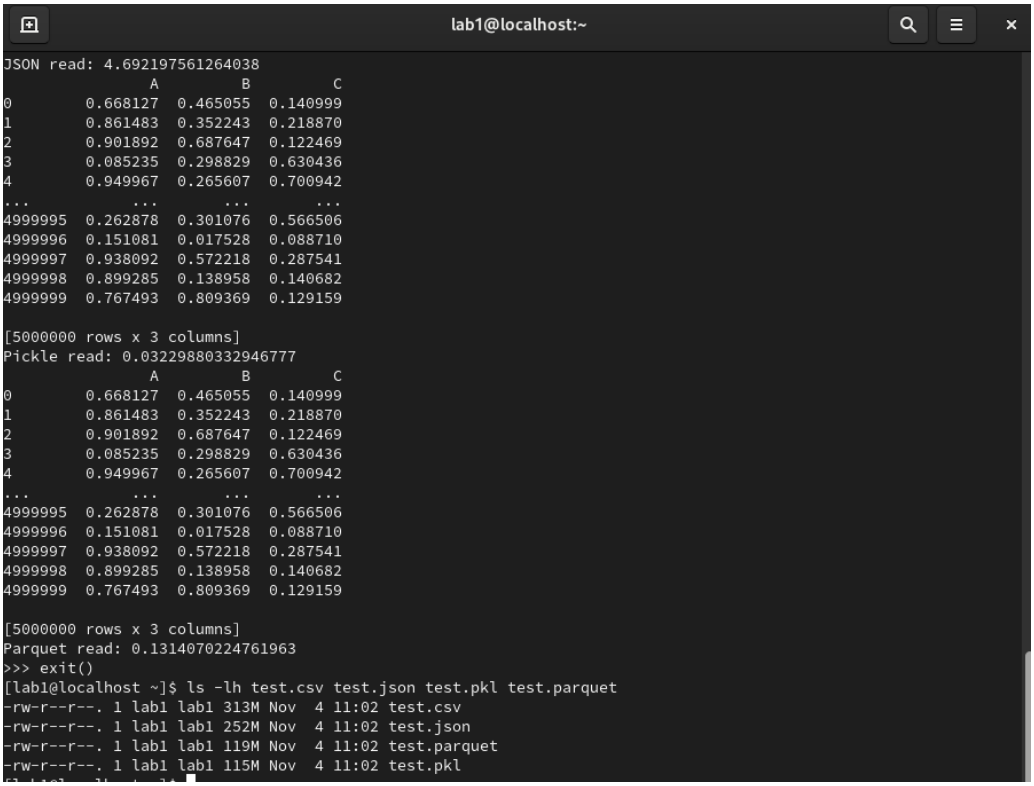


Vision of the Department

To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

Mission of the Department

To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.

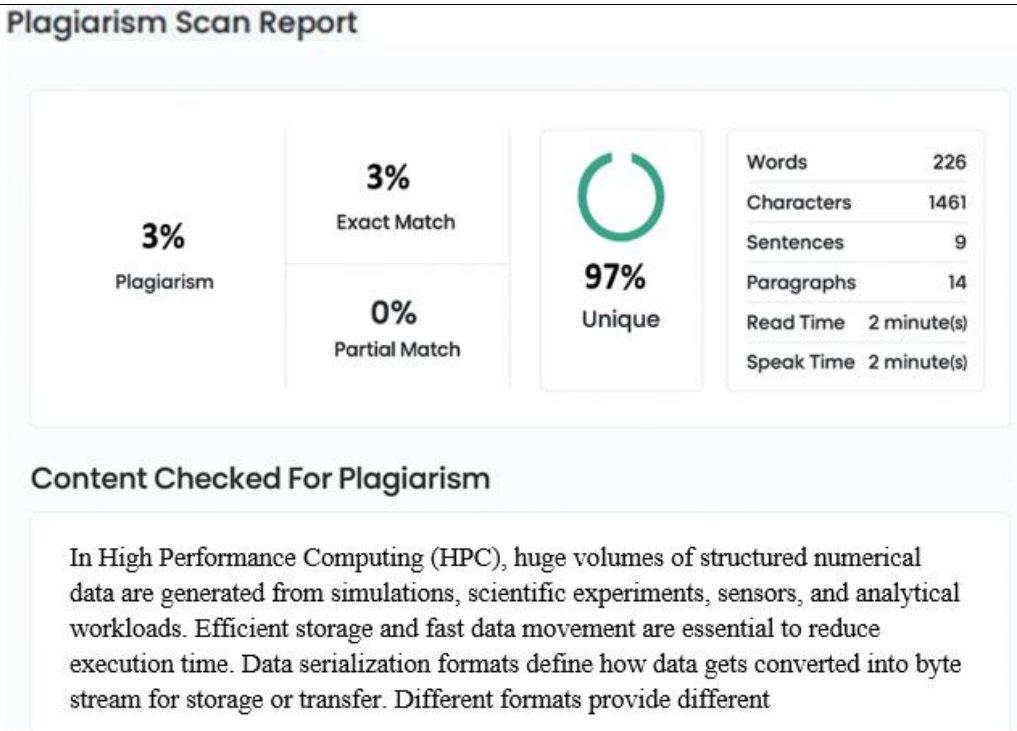
<table>
<tr><td></td><td> 
</td></tr>
<tr><td>Output Analysis</td><td>CSV and JSON formats produced the largest file sizes and slowest read/write times. Pickle was comparatively faster but it is language dependent. Parquet generated the smallest file size and showed the best performance in both write and read operations. Therefore, the experimental observation proves that columnar optimized formats perform significantly better for large HPC datasets.</td></tr>
<tr><td>Link of student Github profile where lab assignment</td><td>https://github.com/sakshi-gokhale/Lab-HPC</td></tr>
</table>

Vision of the Department

To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

Mission of the Department

To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skills through emerging technologies.

has been uploaded	
Conclusion	Parquet format is the most efficient serialization format for HPC because it is columnar, compressed and supports very fast reading operations. CSV and JSON are slower and not ideal for HPC scale data. Pickle is fast but not portable. Hence, Parquet is recommended for large scale scientific and HPC workflows.
Plag Report (Similarity index < 12%)	
Date	04/11/25