Data Processing Model to Perform Big Data Analytics in Hybrid Infrastructures

*

Team Members:
1.Anil Varikuppala
2.Vivek Reddy Suresh Puttireddy
3.Pavan Sai Kumar Jalluri
4.Om Sai Kumar Vaddi

***Abstract*—This project explores the implementation of a hybrid data processing model integrating cloud computing and volunteer computing infrastructures for efficient big data analytics. The approach aims to optimize resource allocation and performance while minimizing costs, based on the methodology proposed by Dos Anjos et al. (2020).**

***Index Terms*—Big Data, Cloud Computing, Volunteer Desktop Cloud, HR Alloc, Network Latency, Data Movement, Data Processing Model**

I. Research Statement and Conjecture:

Our approach will work by leveraging the benefits of both cloud and volunteer computing infrastructures to address the challenges of resource management and cost efficiency in big data analytics. By integrating these environments, we hypothesize that our model will achieve significant improvements in performance and cost savings compared to traditional cloud-based solutions.

This hybrid approach is aimed at optimizing the allocation of computing resources and reducing the costs associated with data processing tasks. We conjecture that by leveraging the combined strengths of both cloud and volunteer computing platforms, our model will significantly improve efficiency and cost-effectiveness over traditional cloud-only solutions. Specifically, we anticipate our model will demonstrate superior scalability and flexibility in managing diverse and voluminous data sets, leading to more effective big data analytics implementations. This approach is substantiated by Dos Anjos et al. (2020), who explored similar integrations, providing a foundational framework for our hypothesis

II. Preliminary Literature Survey and Methodology:

We will conduct a comprehensive literature review focusing on big data analytics, cloud computing, and volunteer computing, based on the paper by Dos Anjos et al. (2020). Our methodology will include implementing and evaluating the HR Alloc algorithm for data placement and resource allocation in a hybrid infrastructure setup.

Our methodology will involve the development and implementation of the HR Alloc algorithm, specifically designed for data placement and resource allocation in a hybrid computing infrastructure. This algorithm is expected to facilitate the dynamic distribution of resources across cloud and volunteer computing platforms, potentially leading to enhanced processing capabilities and reduced operational costs. We will evaluate the algorithm's effectiveness through a series of performance tests comparing it to traditional cloud-based data processing models. This empirical analysis will help us quantify the improvements in performance and cost-efficiency achieved by our hybrid model.

III. To Do List with Deadlines/Milestones:

- June 26: Complete the literature review and finalize the methodology.
- July 2: Develop a prototype for the HR Alloc algorithm.
- july 10: Conduct initial performance testing and analysis.
- July 17: Prepare progress report and abstract submission.
- July 24: Finalize project report and submit for evaluation.

IV. Project Management:

A. Roles:

- Project Leader: [Pavan Sai Kumar Jalluri].
- Research and Literature Review: [Anil Varikuppala]
- Algorithm Development: [Vivek Reddy Suresh Puttireddy]
- Performance Testing and Analysis: [Om Sai Kumar Vaddi]

B. Meetings:

Weekly meetings to review progress, address challenges, and adjust milestones as needed.

C. Communication:

Utilize Slack for daily updates and GitHub for version control and collaborative development.

REFERENCES

[1] 1. Julio C. S. Dos Anjos et al., "Data Processing Model to Perform Big Data Analytics in Hybrid Infrastructures," IEEE Access, DOI: 10.1109/ACCESS.2020.3023344, September 2020.

[2] 2. Marz, N., Warren, J. (2015). Big Data: Principles and best practices of scalable realtime data systems. Manning Publications

[3] 3. Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., Stoica, I. (2010). Spark: Cluster computing with working sets. HotCloud, 10(10-10), 95.