



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 2, February 2025

Advanced Data Analytics Framework for Insights into HTCondor Clusters Usage Patterns

**Prof.Dr .R.V Babar, Prof.A.Joshi, Pranali Kolhe, Abhijeet Chavan, Omkar Hajare,
Dnyaneshwari Jambukar**

Department of Artificial Intelligence and Data Science (AIDS), Suman Ramesh Tulsiani Technical Campus-Faculty of
Engineering, Kamshet, Pune, India

ABSTRACT: High-Throughput Computing (HTC) plays a crucial role in various scientific and industrial domains, enabling large-scale computations across distributed systems. HTCondor, a widely used workload management system, helps schedule and execute compute-intensive jobs. However, efficient job tracking, resource allocation, and real-time monitoring remain significant challenges. This research proposes an advanced data analytics framework that integrates Elasticsearch, Logstash, Kibana (ELK Stack), and Condor Adstash Daemon to analyze HTCondor cluster usage. The framework collects, processes, and visualizes system logs to provide real-time monitoring, anomaly detection, and resource optimization. The system employs Use Case Diagrams, Activity Diagrams, Data Flow Diagrams (DFDs), and Sequence Diagrams to structure the workflow and validate system performance. By offering a scalable, user-friendly interface, this framework enhances decision-making, fault detection, and workload balancing in HTCondor clusters. The ultimate goal is to improve computational efficiency, optimize resource utilization, and support future advancements in HTC environments.

I. INTRODUCTION

High-Throughput Computing (HTC) has revolutionized scientific simulations, big data analytics, artificial intelligence (AI), and cloud computing by allowing researchers to distribute jobs across multiple computing nodes. HTCondor, developed by the Center for High Throughput Computing at the University of Wisconsin-Madison, is a robust workload management system designed for handling large-scale job submissions and executions. However, existing monitoring solutions lack comprehensive real-time analytics and visualization capabilities. System administrators and researchers face difficulties in:

- Tracking job performance and failures
- Understanding resource consumption patterns
- Identifying workload inefficiencies
- Optimizing computational resource allocation

To address these challenges, this research integrates big data analytics tools (Elasticsearch, Logstash, Kibana) with HTCondor to enhance job tracking, performance monitoring, and decision-making capabilities.

II. LITERATURE SURVEY

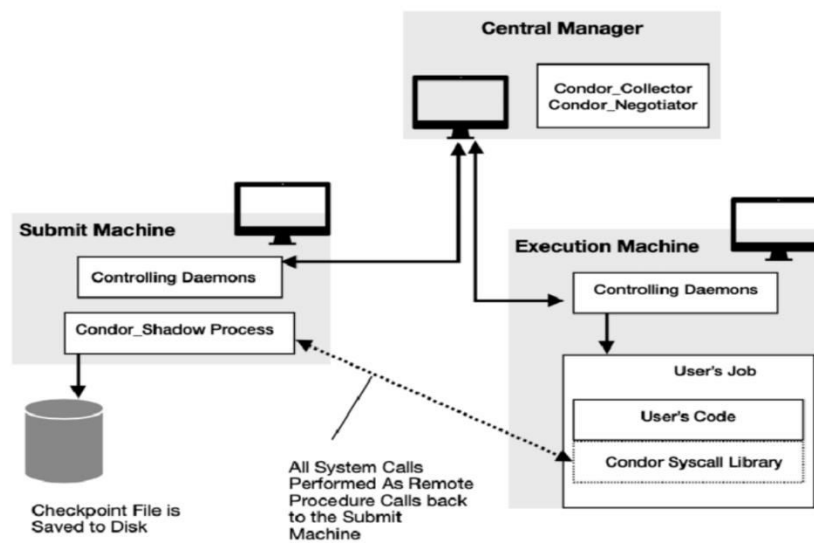
Several studies have explored efficient job scheduling, workload balancing, and performance optimization in High-Performance Computing (HPC) and HTC environments.

1. Traditional Monitoring Approaches
 - Most HTC environments rely on log-based tracking and static reports that provide limited insights into system performance.
 - Existing tools such as Ganglia, Nagios, and Prometheus offer monitoring but lack real-time interactive visualizations for deep analysis.
2. Big Data in Workload Management
 - Research highlights the integration of distributed analytics platforms like Elasticsearch and Hadoop for managing cluster-wide data.

- Advanced log processing tools like Logstash and Apache Kafka are increasingly used for real-time data ingestion.
- 3. Visualization Techniques in HTC
 - Studies demonstrate that interactive dashboards, anomaly detection models, and AI-driven forecasting can significantly improve system monitoring and fault prediction.

Our research builds on these findings by integrating real-time analytics, predictive modeling, and visual dashboards tailored specifically for HTCondor-based computing environments.

III. SYSTEM ARCHITECTURE



IV. METHODOLOGY

The proposed framework follows a structured approach for data collection, processing, visualization, and system optimization.

1. Data Collection

- The Condor Adstach Daemon continuously collects logs from the HTCondor system, including:
 - Job submission records
 - Resource usage metrics
 - Execution time and performance logs

2. Data Processing

- Logstash processes and transforms raw logs into structured data, which is indexed in Elasticsearch for advanced querying and storage.

3. Data Analysis and Storage

- Elasticsearch enables real-time search, aggregation, and filtering of job execution trends, resource allocation metrics, and failure reports.

4. Visualization

- Kibana provides interactive dashboards displaying:
 - Job success/failure rates
 - CPU, memory, and disk utilization
 - User-based job submission trends

5. Modeling and Validation

- The system architecture is validated using Use Case Diagrams, Activity Diagrams, Sequence Diagrams, and Data Flow Diagrams (DFDs) to illustrate workflow efficiency and data movement.

V. PROBLEM DEFINITION

HTCondor is widely used for high-throughput job execution, but it lacks advanced real-time analytics and predictive monitoring. Key challenges include:

- Difficulty in identifying inefficient job executions and failures
- Lack of real-time visualization tools for system administrators
- Limited automation for workload optimization and fault detection

The proposed solution addresses these issues by integrating big data analytics to provide:

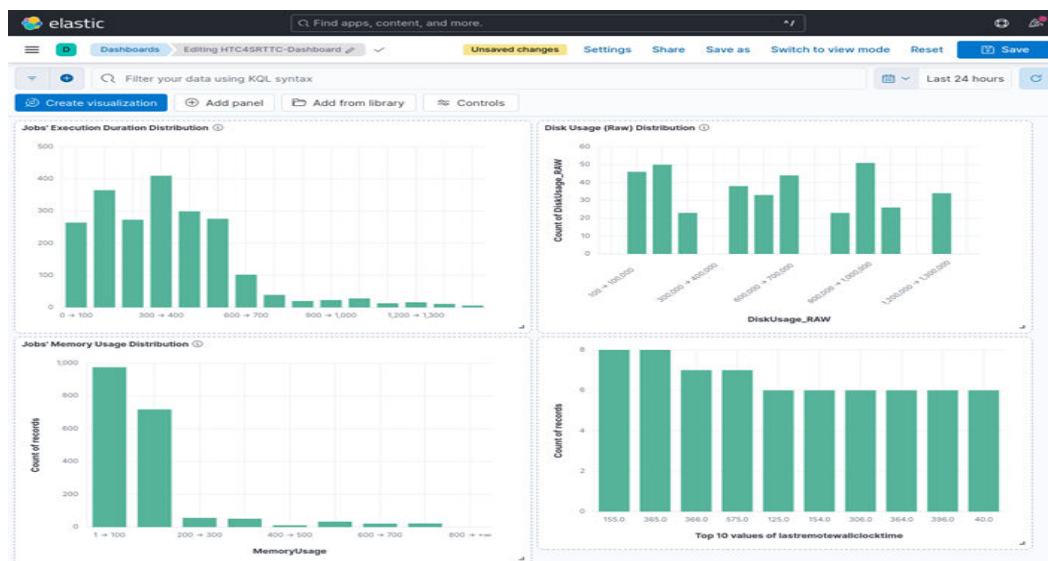
- Real-time system monitoring
- Efficient resource allocation strategies
- Predictive failure detection

VI. DISCUSSION AND RESULT

The framework introduces a modern approach to HTCondor monitoring and management by leveraging Elasticsearch, Logstash, and Kibana for enhanced data analysis and visualization.

Key Benefits of the Proposed System

- Real-time Monitoring
 - Provides live system statistics for administrators to track job statuses.
- Anomaly Detection
 - Detects unusual job failures or high resource consumption spikes.
- Scalability and Flexibility
 - Adapts to different cluster sizes and integrates with other data analytics platforms.
- Improved Decision-Making
 - Enables data-driven optimizations for better job scheduling and resource utilization.
- User-Friendly Dashboards
 - Visualizes complex data for easy interpretation by system administrators and researchers.



VII. CONCLUSION

This project presents a scalable, real-time analytics framework for HTCondor clusters using Elasticsearch, Logstash, Kibana, and Condor Adstash Daemon. The system provides real-time job tracking, resource utilization insights, and anomaly detection, enhancing decision-making and operational efficiency.

Future enhancements include:

- Integration of AI-driven predictive models for job failure forecasting.
- Automated workload balancing using historical data trends.
- Extended compatibility with cloud-based HTC environments.

By leveraging big data analytics, this framework aims to revolutionize HTCondor monitoring, ensuring optimized resource management and improved computational performance in high-throughput computing environments.

REFERENCES

- [1] HTCondor Documentation: <https://htcondor.org/manual/>
- [2] Elasticsearch-Documentation: <https://www.elastic.co/guide/en/elasticsearch/reference/current/index.html>
- [3] GitHub Repository : <https://github.com/dmwm/cms-htcondor-es>



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details