VOLUME 05 ISSUE 09 Pages: 113-123

OCLC - 1368736135













Journal Website: http://sciencebring.co m/index.php/ijasr

Copyright: Original content from this work may be used under the terms of the creative commons attributes 4.0 licence.



Resilient Fog-Cloud Architectures and Fault-Tolerant Test Infrastructure for Large-Scale GPU Manufacturing: Integrating Edge Intelligence, Energy Prediction, and **Verification Paradigms**

Submission Date: September 01, 2025, Accepted Date: September 15, 2025,

Published Date: September 30, 2025

Dr. Marcus R. Ellison

Global Institute of Systems Engineering, University of Edinburgh

ABSTRACT

This article synthesizes contemporary conceptual foundations and applied methodologies for designing resilient fog-cloud computing architectures and fault-tolerant test infrastructure targeted at large-scale GPU manufacturing. The abstract summarizes the research problem, methodological approach, principal findings, and implications. Background: modern GPU manufacturing and deployment increasingly rely on distributed computation across cloud, fog, and edge tiers for tasks including in-line inspection, predictive power management, real-time quality control, and cryptographically secure telemetry aggregation (Prakash, Suresh, & Dhinesh Kumar, 2019; Deepika & Prakash, 2020). Problem statement: existing test infrastructures for GPUs often fail to simultaneously satisfy the competing demands of scalability, low latency, energy efficiency, and rigorous formal verification (Designing Fault-Tolerant Test Infrastructure for Large-Scale GPU Manufacturing, 2025; Huang, Tsai, Paul, & Chen, 2005). Methods: this work develops an integrative framework combining fog computing paradigms, machine-learning based power prediction, hybrid cloud storage and serverless HPC patterns, and formal verification tools (Labelled Transition System Analyser; Lambers, Ehrig & Orejas, 2006; Lohmann, Sauer & Engels, 2005). Results: descriptive analysis demonstrates how layered redundancy, consensus-inspired decision rules, and blockchainaugmented telemetry can jointly improve fault tolerance, reduce energy volatility, and maintain throughput under realistic manufacturing perturbations (Iyer, 2020; Yehia & Aljaafreh, 2023). Discussion: the paper interprets findings in light of trade-offs between latency, trust, and computational cost, examines limitations of current verification approaches, and maps a research agenda for automated model checking and fog-native testing. Conclusion: by synthesizing theoretical and applied work across distributed

VOLUME 05 ISSUE 09 Pages: 113-123

OCLC - 1368736135











computing and verification communities, the proposed architecture offers a practical route to robust, costeffective, and verifiable GPU testbeds capable of meeting future manufacturing scale-up.

KEYWORDS

Fog computing, fault tolerance, GPU manufacturing, power prediction, formal verification, serverless HPC

Introduction

The contemporary manufacturing landscape for high-performance graphics processing units (GPUs) has evolved into a complex socio-technical system where test, verification, and continuous quality assurance must be executed at scale, with minimal latency and high confidence in correctness. Historically, verification and testing have been centralized processes performed postfabrication; however, this approach is increasingly unsuited to present needs because of demands for real-time defect detection, adaptive power management, and geographic distribution of manufacturing and testing resources. computing — the distribution of computation, storage, and services toward the network edge provides a compelling architectural model to decentralize testing, enable low-latency reaction to faults, and offload cloud resources for heavy analytics (Prakash, Suresh, & Dhinesh Kumar, 2019; Prakash et al., 2017). Simultaneously, advances in serverless high-performance computing and hybrid cloud solutions provide new models for elastic test execution and secure longterm data storage (Petrosyan & Astsatryan, 2022; Sandeep & Thangam, 2022).

This paper addresses an integrative problem at the intersection of distributed systems, verification, and manufacturing engineering: how to design a fault-tolerant test infrastructure for large-scale GPU manufacturing that capitalizes on fog computing's low-latency benefits, machine learning for energy and fault prediction, integrates secure telemetry aggregation, and applies rigorous model checking and executable contracts for correctness. The core research questions are: (1) What architectural patterns best reconcile low latency and rigorous fault tolerance in distributed test infrastructures for GPU manufacturing? (2) How can machine learningbased power and fault prediction be embedded into fog-cloud pipelines to optimize scheduling and reduce energy consumption? (3) What role do formal verification tools and techniques play in ensuring correctness of composite testing workflows that span fog, cloud, and serverless lavers?

The literature articulates several critical components relevant to these questions. Early work on composite web service testing and automated model checking highlights the importance of systematic verification of distributed workflows (Huang et al., 2005). The Labelled Transition System Analyser (LTSA) and similar tools demonstrate that modeling the control flow and concurrency properties of test orchestration is viable for detecting deadlocks, and other correctness violations livelocks. (Labelled Transition System Analyser, Version 2.2). Graph transformation conflict detection and executable visual contracts provide further techniques to express and check constraints in model space (Lambers, Ehrig & Orejas, 2006;

114

VOLUME 05 ISSUE 09 Pages: 113-123

OCLC - 1368736135











Lohmann, Sauer & Engels, 2005). On the systems side, empirical and conceptual treatments of fog computing illustrate its advantages and open challenges, including resource management, security, and integration with blockchain for trust (Prakash et al., 2017; Prakash, Suresh, & Dhinesh Kumar, 2019; Yehia & Aljaafreh, 2023). Finally, studies on power consumption prediction, hybrid cloud storage, and serverless HPC show the feasibility of predictive and elastic approaches to resource provisioning (Deepika & Prakash, 2020; Sandeep & Thangam, 2022; Petrosvan Astsatryan, 2022).

The contribution of this paper is threefold. First, it proposes a comprehensive fog-cloud-serverless architecture specifically tailored to the faulttolerant testing needs of GPU manufacturing, combining low-latency in-line checks with heavy cloud analyses. Second, it details how predictive models — particularly for power consumption can inform scheduling and reduce test energy footprints while maintaining throughput. Third, it maps a verification strategy that uses model checking, executable contracts, and conflict detection to ensure the correctness of test workflows across heterogeneous execution environments, offering concrete design patterns and resilience mechanisms.

Methodology

This work adopts a layered descriptive and normative methodology combining architectural synthesis, algorithmic pattern design, and formal method integration. The approach is text-based and conceptual, drawing from the theoretical and empirical findings in the provided literature to assemble an end-to-end design and to reason about its expected behavior and failure modes. The methodology has four primary components: architectural patterning, predictive modeling integration, verification and testing semantics, and deployment/operational considerations.

Architectural Patterning. The architecture is structured into three primary tiers: edge/fog, cloud, and serverless orchestration. The edge/fog tier hosts devices proximate to the GPU manufacturing lines, responsible for data ingestion (sensor arrays, vision systems), immediate anomaly detection, and local mitigation. This tier is crucial for low-latency interventions (Prakash, Suresh, & Dhinesh Kumar, 2019). The cloud tier provides long-term storage, federated model training, and heavy analytics that require larger computational resources (Sandeep & Thangam, 2022). The serverless orchestration layer enables ephemeral, elastic compute tasks for on-demand testing and batch verification jobs, reducing idle resource costs (Petrosyan & Astsatryan, 2022). Communication patterns between tiers include asynchronous message buses, prioritized telemetry channels for critical alerts, and secure ledgered channels when auditability is required (Yehia & Aljaafreh, 2023).

Predictive Modeling Integration. Predictive models particularly for power consumption and failure likelihood — are embedded within the fog tier to support proactive scheduling and test mitigation (Deepika & Prakash, 2020). The methodology describes how models are trained in the cloud on historical telemetry, validated using crossvalidation techniques, and then deployed to fog nodes in lightweight formats suitable for edge inference. The models provide per-unit predictions that inform both local decisions (e.g., throttling,

VOLUME 05 ISSUE 09 Pages: 113-123

OCLC - 1368736135











delayed execution) and cluster-level scheduling (e.g., which units to route to which test harness). A feedback loop updates models with labeled results from tests to continually refine prediction accuracy.

Verification and Testing Semantics. We adopt a formal semantics for test workflows, representing workflows as labeled transition systems amenable to LTSA-style analysis (Huang et al., 2005; Labelled Transition System Analyser, Version 2.2). The methodology prescribes a multi-stage verification pipeline: (a) static model checking of workflow specifications to detect concurrency errors and policy violations: (b) runtime contract monitoring using executable visual contracts that assert invariants during execution (Lohmann, Sauer & Engels, 2005); and (c) conflict detection using graph transformation techniques to identify incompatible operations or negative application conditions off-line (Lambers, Ehrig & Orejas, 2006). The verification layer is tightly coupled with the orchestration layer so verification failures can trigger compensating actions and reconfiguration.

Deployment and Operational Considerations. Realistic deployment addressing requires heterogeneity across manufacturing sites, limited privacy/regulatory network budgets, and constraints for manufacturing telemetry. The methodology therefore includes strategies for adaptive compression of telemetry, selective ledgering of critical events (to reduce blockchain cost while retaining audit trails), and hybrid cloud architectures for secure storage (Sandeep & Thangam, 2022; Yehia & Aljaafreh, 2023). Security hardening leverages cryptographic attestation for fog nodes and sandboxing of test harnesses to prevent lateral contamination.

Each component of the methodology is developed in narrative detail below, elaborating the design choices, expected failure modes, and mitigation strategies.

Architectural Design: Fog-Native, Cloud-**Augmented Testbeds**

The core architectural thesis is that a multi-tiered fog-cloud architecture yields significant resilience and operational advantages for GPU manufacturing test infrastructure because it balances the immediacy of edge responses with the scale and analytic power of cloud resources (Prakash et al., 2017; Prakash, Suresh, & Dhinesh Kumar, 2019). In practice, the fog tier contains sensor aggregation nodes collocated with test benches and vision systems. These nodes host lightweight inference engines for anomaly detection applied to video surveillance streams and telemetry channels to detect thermal excursions, early signs of defect, or communication anomalies (Prakash, Suresh, & Dhinesh Kumar, 2019). Fog nodes are designed to execute three classes of functionality:

- 1. Real-time screening: Low-latency filters classify incoming data into pass/hold/fail categories. This stage uses compact models and rule-based heuristics trained for quick decisions.
- 2. Local mitigation: When a fault is suspected, fog nodes can initiate immediate actions such as isolating a device, restarting test harnesses, or halting certain test sequences. These mitigations prevent propagation of faulty test states and reduce wasted downstream compute.
- 3. Telemetry pre-processing: Fog nodes compress, annotate, and prioritize telemetry before reliable delivery to the cloud, preserving bandwidth and reducing storage costs.

VOLUME 05 ISSUE 09 Pages: 113-123

OCLC - 1368736135











These responsibilities are congruent with the design practices proposed for smart city surveillance and other fog-driven applications, where fog nodes perform essential filtering and rapid control decisions while delegating complex analytics to remote clouds (Prakash, Suresh, & Dhinesh Kumar, 2019).

The cloud tier operates as the strategic analytics backbone. It houses federated model training pipelines for power and fault prediction, long-term storage of labeled test results, and global coordination services such as scheduling policies and configuration management (Deepika & Prakash, 2020; Sandeep & Thangam, 2022). The cloud uses hybrid storage approaches to balance security and efficiency; sensitive trace logs remain in controlled private storage while aggregate metrics and anonymized data feed public model improvements (Sandeep & Thangam, 2022).

Serverless HPC and on-demand test execution bridge the speed/scale gap. Serverless functions and short-lived containers are used for ephemeral test workloads that require burst compute — for example, re-running a suite of hardware diagnostics across a population of GPUs following a detected anomaly (Petrosyan & Astsatryan, 2022). Serverless orchestration reduces the capital costs associated with idle test rigs and facilitates scaling to meet peak demand.

Communication between layers emplovs prioritized channels. Critical alerts from fog nodes use low-latency, guaranteed delivery channels; bulk telemetry uses best-effort transfer with eventual consistency semantics. For auditing and tamper evidence, a hybrid blockchain pattern is where only essential cryptographic commitment records (hashes of test logs, key state transitions) are ledgered to reduce cost, while full logs remain off-chain (Yehia & Aljaafreh, 2023). This hybrid approach balances veracity and practicality for long-term traceability manufacturing contexts.

Predictive Models for Power and Fault **Management**

A central lever for operational efficiency in largescale GPU testbeds is the integration of predictive models that anticipate power consumption and failure likelihood. Accurate power prediction allows the system to schedule tests to avoid energy spikes, allocate cooling resources proactively, and reduce test failures due to thermal stress (Deepika & Prakash, 2020). The methodology envisions a layered modeling pipeline:

Model Development and Training. Historical telemetry and labeled failure data are aggregated in the cloud. Supervised learning methods are trained predict short-horizon to consumption for individual units and to estimate failure probability during a given test profile. Feature engineering focuses on key signals: temperature traces, current draw patterns, sensor fusion outputs from vision systems, and prior test outcomes. Cross-validation and out-of-sample evaluation are used to assess model robustness.

Edge Deployment. Trained models are distilled into lightweight formats for fog deployment — for example, via quantized neural networks or small gradient boosted ensembles — to provide local inference with millisecond latency, enabling onsite scheduling decisions (Deepika & Prakash, 2020). The distillation process preserves predictive fidelity while meeting memory and compute constraints on fog nodes.

VOLUME 05 ISSUE 09 Pages: 113-123

OCLC - 1368736135











Closed-Loop Adaptation. Models are continuously updated using labeled outcomes. When fog nodes execute tests and observe ground truth, they forward these to the cloud where models are retrained in batch or incrementally updated using federated learning techniques to respect privacy and bandwidth constraints.

Operational Policies Informed by Predictions. Power-aware scheduling policies use predictions to stagger tests such that aggregate power demands remain below thresholds. In practice, this takes the form of a constrained optimization problem where tests are scheduled to minimize makespan while satisfying power budgets. The policies can be expressed as heuristics suitable for local implementation: if an incoming test is predicted to exceed the local power budget in the next time window, the fog node either defers the test, routes it to another test bench with spare power headroom, or reduces the test intensity (e.g., using a reduced test pattern) until resources are freed.

Predictive models also inform risk scoring. Units with elevated failure probability might be fasttracked for more stringent checks, quarantined for extended burn-in cycles, or tagged for more dispatch. Embedding conservative intelligence at the fog level reduces unnecessary transport of faulty units, conserving throughput and lowering downstream rework costs.

Verification and Formal Methods Integration

Automation and scale in testing demand robust approaches to ensure that orchestration logic, compensation protocols, and monitoring contracts behave correctly under concurrency and partial failure. The verification approach synthesizes

automated model checking, executable visual contracts, and graph transformation conflict analysis to provide multi-layered assurance (Huang et al., 2005; Labelled Transition System Analyser, Version 2.2; Lambers, Ehrig & Orejas, 2006; Lohmann, Sauer & Engels, 2005).

Workflow Modeling. Test workflows are modeled as labeled transition systems (LTS) capturing states (test stages, device statuses) and transitions (events, commands). This formal model represents across concurrency multiple devices Model checkers orchestrators. verify can properties such as reachability (a certain pass/fail state is reachable only under certain conditions). absence of deadlocks (the system cannot reach a state where no progress is possible), and adherence to safety invariants (critical constraints like "if test harness is in isolation mode, network writes are blocked").

Static Verification. Automated model checking tools — conceptually similar to LTSA — are used to exhaustively explore the state space of the workflow specification. This analysis identifies latent concurrency issues and protocol violations that may not be evident in ad hoc testing. For composite workflows involving multiple orchestration microservices and patterns. automated checking has proven effective in detecting subtle composition errors (Huang et al., 2005).

Runtime Contracts and Monitoring. Tο complement static analysis, executable visual contracts are embedded into runtime components to assert invariants and pre/post conditions (Lohmann, Sauer & Engels, 2005). These contracts are evaluated at runtime and violations trigger compensatory actions such as rollbacks, retries, or

118

VOLUME of ISSUE on Pages: 113-123

OCLC - 1368736135











reconfiguration of test parameters. Because manufacturing environments experience nondeterministic perturbations, runtime monitoring is essential for resilience.

Conflict Detection via Graph Transformation. Graph transformation techniques — especially those handling negative application conditions are applied to detect conflicts in resource allocation and operation sequencing (Lambers, Ehrig & Orejas, 2006). For example, negative application conditions can express that a device should not be assigned to two conflicting test harnesses simultaneously; conflict detection algorithms can preemptively identify these allocation errors in the orchestration plan.

Integration Strategy. The verification pipeline is designed to be incremental. Workflow changes first pass through static model checking; then, during deployment, a combination of staged rollouts and contract monitoring observes real executions. Telemetry from runtime contracts and monitoring is fed back as labeled traces to refine models and update static specifications when real-world behaviors expose modeling gaps.

Fault Tolerance Mechanisms and Patterns

Fault tolerance in distributed test infrastructures must contend with both device-level failures (component defects, thermal runaway) and infrastructure-level faults (network partitions, fog node crashes). The proposed design uses several patterns to achieve robust behavior:

Redundant Observation and Majority Voting. Multiple, independent sensors and algorithmic detectors observe the same device. This redundancy supports majority voting to reduce false positives from noisy sensors. Majoritarian

decision rules are supplemented by confidence thresholds from predictive models, enabling probabilistic reconciliation when sensors disagree.

Local Containment and Quarantine. Upon anomaly detection, fog nodes place devices in quarantine, isolating them from the test pool and disabling operations that could exacerbate the failure. Quarantine mitigates cascading failures and preserves test harness availability.

Graceful Degradation and Fallback. When fog nodes cannot access cloud services due to network issues, they operate in degraded modes using cached models and local policies. These fallback mechanisms trade some accuracy for continuity of operations. Importantly, the system logs all degraded actions for later reconciliation.

Compensation and Rollback Protocols. complex transactions spanning multiple components (e.g., multi-stage tests requiring crossdevice coordination), compensation actions are defined to reverse or mitigate partial progress when failures occur. The formal verification laver ensures that compensation sequences preserve invariants.

Task Rebalancing and Elasticity. Serverless orchestration enables elastic reallocation of compute resources for heavy re-tests following failure spikes. The system automatically scales to process backlog while prioritizing critical quarantine re-tests based on risk scores.

Ledgered Audit Trails for Forensic Analysis. Critical events and state transitions cryptographically committed using hybrid ledgering to enable non-repudiable audit trails without incurring the full cost of storing bulk logs on the chain (Yehia & Aljaafreh, 2023). Ledgering

119

VOLUME 05 ISSUE 09 Pages: 113-123

OCLC - 1368736135











supports post-mortem analyses, root cause investigations, and regulatory audits.

Operational Policies and Scheduling Heuristics

Operationalizing the architecture requires practical scheduling heuristics that are implementable at fog nodes with limited lookahead. Derived from predictive power models and system state, the paper outlines several heuristics:

Power-Aware Time Slicing. Divide the schedule into time slices and allocate tests such that the predicted aggregate power in any slice stays within threshold T. If a new test would cause an overrun. it is deferred or routed elsewhere. This is a local, greedy heuristic that approximates global power management while remaining computationally inexpensive.

Risk-Prioritized Testing. Use failure probability predictions to assign priority. Units with higher predicted failure probabilities are routed to more comprehensive test sequences earlier to avoid false negatives reaching downstream production.

Progressive Intensification. Begin with lightweight, fast checks; escalate to full diagnostics if anomalies are detected. This reduces average test durations and conserves energy while preserving probability of detecting defects.

Distributed Load Balancing. Fog nodes coordinate via a gossip protocol to share load and headroom information, permitting lightweight load balancing among adjacent nodes when network conditions permit.

Audit-Aware Routing. Tests whose results must be ledgered for contractual or regulatory reasons are routed through nodes that support secure ledger interfaces, ensuring traceability where needed.

These heuristics are designed to be composable and adaptive: they can be combined using policy engines in fog nodes, and their parameters tuned based on site characteristics.

Results (Descriptive Analysis of Expected Behavior)

Because this work synthesizes established techniques rather than reporting new empirical field trials, the results are presented as a descriptive analysis grounded in the evidence from the referenced literature. The analysis focuses on how the combined architecture and policies are expected to influence three critical operational metrics: latency (time to detect and respond), energy consumption, and correctness (absence of critical workflow violations).

Latency. Fog-deployed anomaly detection reduces detection latency by orders of magnitude compared to cloud-only analysis because local inference operates near the data source, avoiding round-trip delays (Prakash, Suresh, & Dhinesh Kumar, 2019). The architectural pattern ensures that time-critical actions — such as halting a test harness on detection of thermal excursions — can be executed within the fog node's control loop. This localized control is consistent with findings in fog applications where immediate actions were necessary for safety-critical responses.

Energy Consumption. Integrating predictive power models into scheduling heuristics is expected to reduce peak power draws and smooth aggregate consumption. By deferring or rerouting highpower tests when predicted to coincide, the system reduces the need for emergency cooling and peak

VOLUME 05 ISSUE 09 Pages: 113-123

OCLC - 1368736135











power provisioning. Prior studies have demonstrated the practical gains of such schemes in cloud data centers; when adapted to fog-native manufacturing environments, these gains translate into lower operational energy costs and reduced failure rates caused by thermal stress (Deepika & Prakash, 2020).

Correctness and Robustness. The multi-stage orthogonal verification pipeline provides assurances. Static model checking prevents many classes of composition errors before deployment (Huang et al., 2005). Runtime contracts ensure contracts are upheld during execution and provide immediate triggers for mitigation when violated (Lohmann, Sauer & Engels, 2005). Graph transformation conflict detection further reduces resource allocation errors (Lambers, Ehrig & Orejas, 2006). Together, these measures reduce the probability of systemic failures and increase the ability to recover from partial faults.

Trade-offs and Resource Costs. The benefits come at the expense of increased architectural complexity and engineering effort. Fog nodes must be provisioned, secured, and kept in sync with cloud models. Ledgering and verification add compute overhead. infrastructure Nevertheless, hybridization strategies — such as hybrid ledgering and model distillation ameliorate many resource costs while preserving the key advantages of decentralization (Yehia & Aljaafreh, 2023; Petrosyan & Astsatryan, 2022).

Discussion

Interpretation and Theoretical Implications. The proposed architecture synthesizes distributed systems theory with verification and predictive analytics to create a resilient testing infrastructure adapted to the unique constraints of GPU manufacturing. Theoretically, the design embodies principles from distributed fault tolerance (redundancy, consensus, graceful degradation) and from formal methods (explicit modeling and run-time assertion). This dual orientation combining probabilistic machine learning predictions with deterministic formal verification — addresses a fundamental tension: machine learning excels at prediction under uncertainty but lacks formal guarantees; formal methods offer guarantees but often struggle with stochastic, highcompartmentalizing dimensional data. By responsibilities — using ML for probabilistic scheduling and verification tools for protocol correctness — the architecture harnesses the strengths of both paradigms.

Practical Trade-offs. **Operational** realities introduce several trade-offs that organizations must manage. First, model accuracy at the edge depends on the timeliness and representativeness of training data; data drift can erode trust in local predictions, requiring frequent revalidation (Deepika & Prakash, 2020). Second, ledgering and cryptographic commitments create persistent audit trails but can impose latency and storage costs; the hybrid approach recommended here mitigates these costs while preserving key auditability properties (Yehia & Aljaafreh, 2023). Third, the complexity of integrating verification tools into continuous deployment pipelines can be non-trivial; however, the benefits in detecting subtle concurrency bugs make the investment worthwhile, especially given the high cost of faulty GPUs entering the supply chain.

Limitations. The analysis has limitations. The framework is conceptual and synthesizes results

VOLUME 05 ISSUE 09 Pages: 113-123

OCLC - 1368736135











from the literature; empirical validation under diverse production conditions is necessary to quantify gains precisely. The scalability limits of some verification tools when applied to very large spaces remain a practical challenge; abstraction techniques and compositional verification must be explored to handle industrialscale workflows (Huang et al., 2005). Model robustness under adversarial conditions — for example, when sensors are tampered with requires further study and integration adversarial detection mechanisms.

Future Research Directions. Several promising avenues emerge from this synthesis. Federated learning approaches tailored for fog environments can improve model freshness while respecting data locality and privacy constraints. Research into compositional verification methods and automated abstraction can make model checking feasible for industrial-scale orchestration logic. Exploring specialized ledgering primitives that optimize for manufacturing telemetry — balancing auditability and throughput — is another direction. Finally, empirical pilot studies and controlled field trials will be necessary to quantify energy savings, defect detection improvement, and the effect throughput.

Conclusion

This article has presented an integrated architecture and methodological framework for fault-tolerant test infrastructure in large-scale GPU manufacturing. It argues that combining fog computing for low-latency decision making, cloud resources for large-scale analytics and model training, serverless patterns for elastic test execution, predictive models for power and failure

forecasting, and formal verification tools for correctness yields an infrastructure capable of meeting the stringent demands of modern GPU manufacturing. Key design elements include model distillation for edge deployment, hybrid ledgering for auditability, executable contracts and model checking for correctness, and adaptive scheduling heuristics informed by predictive analytics.

While thorough empirical evaluation remains a priority for future work, the descriptive analysis grounded in extant literature suggests the proposed architecture can materially improve latency, reduce energy costs, and enhance the reliability and traceability of testing operations. The synthesis provided here aims to guide engineers and researchers in constructing resilient, verifiable, and energy-aware testing ecosystems for high-volume GPU production.

References

- 1. Prakash, P., R. Suresh, P. N. Dhinesh Kumar. Smart City Video Surveillance Using Fog Computing. International Journal of Enterprise Network Management, Vol. 10, March 2019, pp. 389-399. DOI: 10.1504/IJENM(2019).
- 2. Prakash, P., K. G. Darshaun, P. Yaazhlene, M. V. Ganesh, V. Vasuda. Fog Computing: Issues, Challenges and Future Directions. International Journal of Electrical and Computer Engineering (IJECE), Vol. 7, December 2017, No 6, pp 3669-3673.
 - https://DOI:10.11591/ijece.v7i6.pp3669-3673
- 3. Singh, B. S., M. Pratap, D. K. Sangeeta. Hardware Setup for VLC Based Vehicle to Vehicle Communication under Fog Weather Condition. International Journal of Advanced Science and Technology, Vol. 29, 2020, No 3s.

VOLUME of ISSUE on Pages: 113-123

OCLC - 1368736135











- **4.** Designing Fault-Tolerant Test Infrastructure for Large-Scale GPU Manufacturing. (2025). International Journal of Signal Processing, Embedded Systems and VLSI Design, 5(01), 35-61. https://doi.org/10.55640/ijvsli-05-01-04
- 5. Deepika, T., P. Prakash. Power Consumption Prediction in Cloud Data Center Using Machine Learning. International Journal of Electrical and Computer Engineering, 2020, pp. 1524-1532. http://doi.org/10.11591/ijece.v10i2
- 6. Sandeep, H. R., S. Thangam. A Hybrid Cloud Approach for Efficient Data Storage and Security. In: Proc. of 6th International Conference on Communication and Electronics Systems (ICCES'22), 2022.
- 7. Iver, G. N. Evolutionary Games for Cloud, Fog and Edge Computing – A Comprehensive Study. in Intelligent Systems Advances Computing, Vol. 990, 2020, pp. 299-309. http://doi.org/:10.1007/978-981-13-8676-3 27
- 8. Yehia, I., A. A. Aljaafreh. Block Chain-Fog Computing Integration Applications.

- Cybernetics and Information Technologies, Vol. 23, 2023, No 1, pp. 3-37.
- 9. Petrosyan, D., H. Astsatryan. Serverless High-Performance Computing Cloud. over Cybernetics and Information Technologies, Vol. 22, 2022, No 3, pp. 82-92.
- 10. Huang, H., W.-T. Tsai, R. Paul, Y. Chen. Automated Model Checking and Testing for Composite Web Services. In Proc. of 8th IEEE Intern. Symp. on Object-Oriented Real-Time System Computing (ISORC'05), 2005, pp. 300-307.
- 11.Labelled Transition System Analyser (Version 2.2) http://www.dse.doc.ic.ac.uk/concurrency/ ltsa-v2/index.html
- 12. Lambers, L., H. Ehrig, F. Orejas. Conflict Detection for Graph Transformation with Negative Application Conditions. In Proc. of ICGT2006, pp. 61-76, 2006.
- 13. Lohmann, M., S. Sauer, G. Engels. Executable Visual Contracts. In Proc. IEEE Symposium on Languages Visual and Human Centric Computing (VL/HCC 05), pp. 63-70, 2005.