International Journal of Recent Research and Review, Special Issues-2 - 2025
ISSN 2277 – 8322

LEVERAGING AMAZON WEB SERVICES FOR NASA'S SCIENTIFIC MISSIONS: ARCHITECTURE, APPLICATIONS, AND OUTCOMES

Vimal Daga
CTO, LW India | Founder, #13 Informatics Pvt Ltd
LINUX WORLD PVT. LTD.

Preeti Daga
CSO, LW India | Founder, LWJazbaa Pvt Ltd
LINUX WORLD PVT. LTD.

Krish Sharma
Research Scholar
LINUX WORLD PVT. LTD.

Abstract- NASA's collaboration with Amazon Web Services (AWS) represents a transformative shift in how space agencies manage, process, and distribute vast volumes of scientific data. With missions like NISAR (NASA–ISRO Synthetic Aperture Radar) generating over 70 terabytes of satellite data daily, traditional on-premises infrastructure proved insufficient to meet the demands of scalability, agility, and cost-efficiency. AWS's cloud-native architecture enables NASA to dynamically scale compute resources using Amazon EC2 Spot Instances, which offer up to 90% cost savings compared to on-demand pricing. This allows NASA's Jet Propulsion Laboratory (JPL) to process raw radar signals into high-resolution imagery and derived data products such as displacement maps, all within mission timelines1. Central to this architecture is Amazon S3, which serves as the backbone for storing raw, intermediate, and final data products. The use of AWS Lambda, Amazon SQS, and Amazon SNS orchestrates real-time event-driven workflows, ensuring that data ingestion, transformation, and distribution occur with minimal latency and high fault tolerance. These services are tightly integrated with AWS Step Functions, which visually manage complex workflows and automate decision logic across distributed systems.

Keywords: NASA leverages Amazon Web Services (AWS) for scalable, cloud-native mission workflows. EC2 Spot Instances and S3 support HPC and centralized data lakes, while serverless tools like Lambda and event-driven orchestration (SNS, SQS, Step Functions) enable real-time analytics and automation.

I. INTRODUCTION

In the era of big data and cloud-native infrastructure, the collaboration between NASA and Amazon Web Services (AWS) reflects a pivotal shift in how scientific institutions manage, process, and distribute vast amounts of mission-critical information. As space-based Earth observation missions like NASA–ISRO Synthetic Aperture Radar (NISAR) produce over 70 terabytes of satellite data daily, traditional on-premises systems lack the elasticity and throughput needed for scalable scientific operations. AWS provides a robust and cost-optimized cloud ecosystem that supports high-performance computing (HPC), serverless orchestration, and open science dissemination through services such as EC2 Spot Instances, Amazon S3, AWS Lambda, and Step Functions.The transition to cloud-native architecture has enabled NASA's Jet Propulsion Laboratory (JPL) and Earth Science divisions to implement real-time event-driven pipelines for data ingestion, processing, and global delivery. By leveraging event orchestration tools like Amazon SNS and SQS, combined with centralized data lake architecture in S3, NASA ensures minimal latency, fault tolerance, and operational resilience across distributed workflows. Additionally, services like Amazon SageMaker and OpenSearch enable machine learning-driven anomaly detection and metadata indexing to enhance responsiveness and analytical capabilities.

NASA's Earthdata Cloud—a cloud-hosted evolution of the Earth Observing System Data and Information System (EOSDIS)—embodies the agency's commitment to the FAIR data principles and democratization of satellite-derived datasets. Researchers can interact directly with cloud-hosted analysis-ready data using Jupyter Notebooks and containerized applications, fostering reproducibility and collaborative science across domains such as climate modeling, disaster response, and planetary surveillance. Security and compliance are upheld through AWS GovCloud, IAM role-based policies, and RBAC protocols, ensuring adherence to federal data sovereignty standards.This paper explores NASA's cloud-enabled methodology for managing large-scale geospatial datasets, assessing both the technical architecture and the scientific implications of deploying Infrastructure-as-a-Service (IaaS) for Earth and space missions. Through case studies, performance metrics, and architectural analysis, we evaluate the efficiencies, challenges, and future potential of AWS-driven data ecosystems in enabling scalable, intelligent space exploration.

II. LITERATURE REVIEW

NASA's initial attempt to modernize its compute infrastructure came through Nebula, an early OpenStack platform. Although innovative for its time, Nebula was ultimately discontinued in favor of more scalable and cost-efficient commercial solutions, with AWS emerging as the primary provider. Several case studies demonstrate NASA's successful use of AWS services.A prominent example is the JPL NISAR mission, which processes approximately 70 terabytes of Earth observation data per day using EC2 Spot Instances. The dynamic nature of AWS's compute resources allows NASA to scale compute jobs flexibly based on data influx while minimizing costs. Moreover, NASA uses AWS for advanced machine learning pipelines, including real-time anomaly detection of solar superstorms—a critical capability for protecting satellites and infrastructure on Earth.NASA's Astrophysics Data System (ADS) has also undergone modernization using microservices and Kubernetes clusters deployed on AWS. This move significantly improved performance, scalability, and maintainability for one of the world's most-used astronomical metadata repositories. These developments underscore AWS's integral role across scientific disciplines within NASA.

III. METHODOLOGY (WITH THINKING POINTS)

To understand the practical implementation of AWS within NASA operations, it is essential to examine the architecture of specific workflows. In the case of NISAR and similar missions, raw satellite data is ingested into Amazon S3 buckets, where it is securely stored with versioning and lifecycle policies. This ingestion process triggers AWS Lambda functions or notifications via Amazon SNS and SQS to begin processing tasks.

Compute-intensive workloads, such as data analysis and image processing, are handled by EC2 Spot Instances—virtual machines that offer significant cost savings compared to on-demand pricing. These are orchestrated using AWS Step Functions to ensure reliability and fault tolerance. Post-processing data is again written to S3 for archiving or public dissemination. Additionally, metadata and analytics outputs are stored in Amazon RDS or indexed in Elasticsearch/OpenSearch for querying and dashboard visualization.

Quantitatively, this methodology enables NASA to process upwards of 70 TB of data daily with minimal latency. These services are further supported by CloudWatch for monitoring and IAM policies to ensure security and access control. Importantly, NASA's commitment to open science is reflected in the Earthdata Cloud platform, where processed datasets are made publicly available with no egress cost to users accessing data from within the AWS cloud.

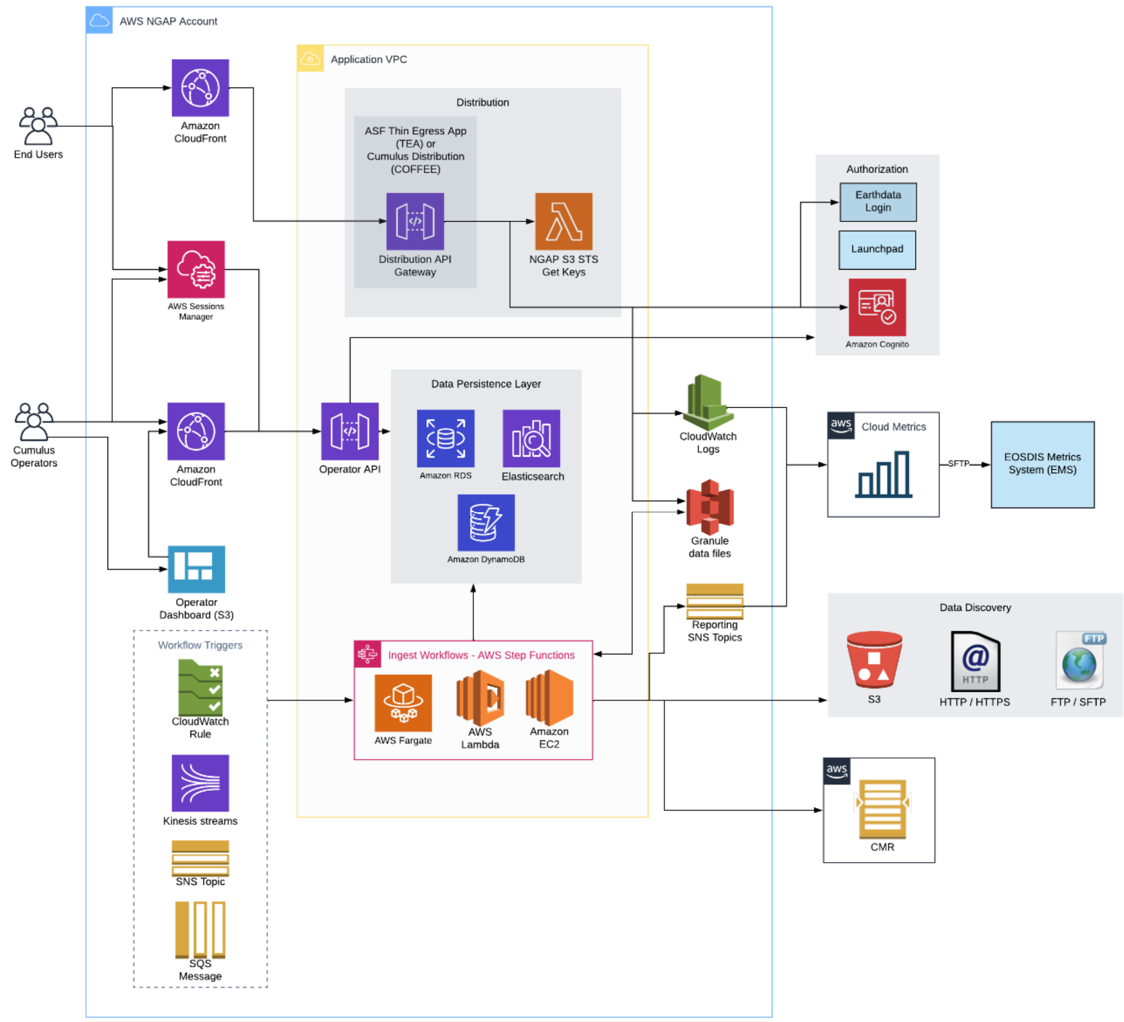

Figure 1: The image above (Cumulus architecture diagram) illustrates NASA's Cumulus system deployed on AWS

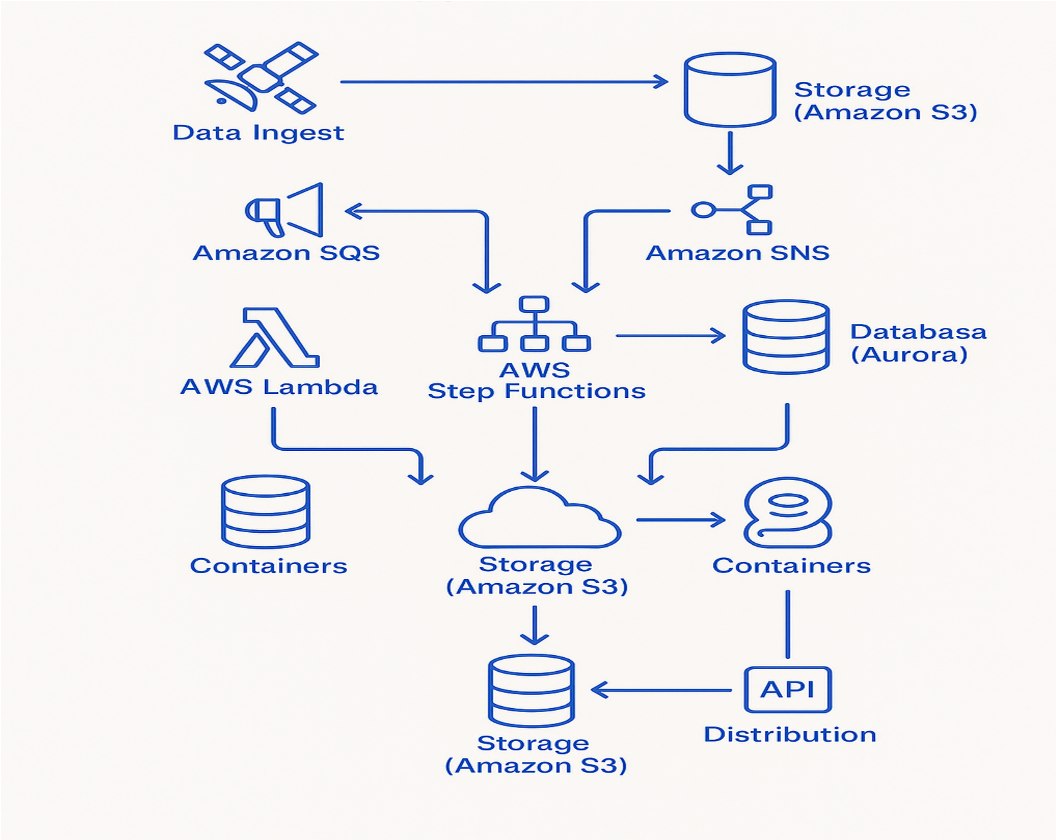

Figure 2:This architecture embodies a cloud-native approach to mission operations, highlighting NASA's use of serverless functions, messaging queues,

distributed storage, and managed orchestration to minimize infrastructure overhead while maximizing scalability and reproducibility.

IV. ADVANTAGES

NASA’s use of AWS brings several notable advantages. First, scalability is vastly improved. Missions like NISAR and the Solar Dynamics Observatory produce data on a scale that traditional infrastructure cannot efficiently manage. By leveraging AWS, NASA scales workloads elastically, deploying thousands of virtual cores as needed.

Second, significant cost savings are achieved through Spot Instances, which allow NASA to access computing power at up to 90% less cost than on-demand instances. This approach is ideal for interruption-tolerant workloads where jobs can be restarted without losing critical progress.

Third, AWS facilitates faster time-to-insight, particularly for anomaly detection and real-time analysis. By automating pipelines through serverless functions, NASA reduces manual overhead and latency in data processing.

Another advantage is global accessibility. Through Earthdata Cloud, NASA enables scientists worldwide to download and analyze data with minimal friction. The use of microservices and serverless patterns also enhances maintainability, allowing NASA teams to iterate faster and modularize complex systems.

V. DISADVANTAGES

espite its benefits, cloud adoption presents some challenges. Spot Instances, while cost-effective, can be interrupted without notice. This requires robust fault-tolerant design and checkpointing to avoid data loss or delays.

Vendor lock-in is another concern. NASA's growing reliance on AWS-specific services and APIs may limit future flexibility or migration to other platforms. Additionally, mission-critical and ITAR-compliant data requires strict access controls, often necessitating the use of AWS GovCloud and tailored IAM policies.

Costs must also be carefully managed. Although public users can access Earthdata freely, internal operational costs related to storage, compute, and data transfers within NASA still require close monitoring.

Lastly, the transition to a cloud-native paradigm involves a steep learning curve. Teams need expertise in Infrastructure as Code (IaC), serverless development, distributed systems, and cloud security best practices.

VI. RESULTS

This table summarizes the Amazon Web Services (AWS) components leveraged by NASA to support scalable, cost-efficient, and compliant mission operations. The services are grouped based on their functional roles across compute, storage, orchestration, monitoring, and security domains. Each entry details the service type, its specific use case within NASA workflows, and the corresponding benefits or implementation highlights.

VII. FUTURE WORK

The future work in this research will focus on enhancing the granularity and interoperability of cloud-hosted Earth observation data systems. This includes integrating real-time data streams with predictive analytics models for improved early warning systems in disaster scenarios. Expanding machine learning pipelines across additional NASA missions and optimizing algorithms like Random Cut Forest and deep learning frameworks for anomaly detection remain pivotal. A dedicated effort will also explore cost-aware cloud architectures by modeling data egress patterns and implementing intelligent compression and caching strategies. Additionally, emphasis will be placed on developing standardized metadata schemas that align with FAIR principles to support global collaboration. Research will aim to simulate hybrid-cloud deployments, combining on-premises compute with AWS services for latency-sensitive applications. Finally, future initiatives will investigate deploying containerized workflows using Kubernetes on AWS, enabling portable, reproducible science in a highly scalable cloud environment.

VIII. CONCLUSION

NASA's strategic partnership with Amazon Web Services (AWS) exemplifies a paradigm shift in space science and Earth observation, marked by a transition from legacy infrastructure to cloud-native, scalable systems. By embracing Infrastructure-as-a-Service (IaaS), NASA not only meets the growing computational demands of missions like NISAR, but also pioneers a model of data management that is resilient, cost-efficient, and globally accessible. The use of EC2 Spot Instances, Amazon S3, and serverless orchestration tools has enabled near real-time processing of massive satellite datasets, streamlining data pipelines while optimizing performance and cost.Through initiatives such as Earthdata Cloud, NASA advances the principles of open science by allowing researchers to conduct analysis directly in the cloud using modern tools like Jupyter Notebooks and Docker containers. This methodology not only promotes transparency and reproducibility, but also empowers the global scientific community with timely access to high-quality geospatial data. Security frameworks involving IAM, RBAC, and GovCloud reinforce compliance and governance standards for sensitive scientific operations. As future missions demand even greater agility and insight from data infrastructure, NASA's cloud-enabled workflows serve as a robust blueprint for innovation in both governmental and commercial aerospace sectors. The integration of machine learning, containerized environments, and automated orchestration continues to redefine what's possible in satellite analytics, planetary science, and climate intelligence—making AWS not just a provider, but a catalyst for scientific transformation.

REFERENCES

[1] Accomazzi, A. (2024). Decades of Transformation: Evolution of the NASA Astrophysics Data System's

Infrastructure. *arXiv preprint arXiv:2401.09685*.

[2] Elbehiery, K., & Elbehiery, H. (2020). Cloud technology: Conquest of commercial space business. In *Internet of Things—Applications and Future: Proceedings of ITAF 2019* (pp. 53-72). Singapore: Springer Singapore.

[3] Elbehiery, K., & Elbehiery, H. (2020). Cloud technology: Conquest of commercial space business. In *Internet of Things—Applications and Future: Proceedings of ITAF 2019* (pp. 53-72). Singapore: Springer Singapore.

[4] Pomerantz, M., Boettcher, A., Hans, M., Sandoval, M., & Wenzel, S. (2019, July). Multi-mission, interactive 3d visualization in a web browser for robotic system and space flight mission development and operations. In *2019 IEEE International Conference on Space Mission Challenges for Information Technology (SMC-IT)* (pp. 17-23). IEEE.

[5] Pomerantz, M., Boettcher, A., Hans, M., Sandoval, M., & Wenzel, S. (2019, July). Multi-mission, interactive 3d visualization in a web browser for robotic system and space flight mission development and operations. In *2019 IEEE International Conference on Space Mission Challenges for Information Technology (SMC-IT)* (pp. 17-23). IEEE.