AWS Whitepaper

Guavus 5G-IQ NWDAF on AWS

Guavus 5G-IQ NWDAF on AWS: AWS Whitepaper

Table of Contents

Guavus 5G-IQ NWDAF on AWS

Publication date: **May 5, 2022** (*Document history*)

Abstract

This whitepaper introduces the concept of Network Data Analytics Function (NWDAF), covers the benefits of deploying NWDAF on Amazon Web Services (AWS), and explains how the Guavus NWDAF solution is architected on AWS.

The AWS Well-Architected Framework helps you understand the pros and cons of the decisions you make when building systems in the cloud. The six pillars of the Framework allow you to learn architectural best practices for designing and operating reliable, secure, efficient, cost-effective, and sustainable systems. Using the AWS Well-Architected Tool, available at no charge in the AWS Management Console (sign-in required), you can review your workloads against these best practices by answering a set of questions for each pillar.

For more expert guidance and best practices for your cloud architecture—reference architecture deployments, diagrams, and whitepapers—refer to the AWS Architecture Center.

Are you Well-Architected?

The AWS Well-Architected Framework helps you understand the pros and cons of the decisions you make when building systems in the cloud. The six pillars of the Framework allow you to learn architectural best practices for designing and operating reliable, secure, efficient, cost-effective, and sustainable systems. Using the AWS Well-Architected Tool, available at no charge in the AWS Management Console, you can review your workloads against these best practices by answering a set of questions for each pillar.

For more expert guidance and best practices for your cloud architecture—reference architecture deployments, diagrams, and whitepapers—refer to the AWS Architecture Center.

Introduction

Mobile network operators (MNOs) around the world are building 5G networks to usher in a new era of gigabit-speed, low-latency, and ultra-reliable mobile services. MNOs envision 5G powering a broad range of applications and uses for consumers and businesses that are simply not possible in today's 4G LTE networks. 5G is designed to connect billions of devices through next-generation New Radio (NR) technology operating in new spectrum bands. 5G networks will also feature dense deployments of small cells, facilitating the delivery of high-bandwidth services in distance-limited, high-frequency bands.

All of this will be allowed by 5G cloud-native infrastructure that is disaggregated, virtualized, and software-defined. Using a cloud-native stack powers digital transformation, allowing 5G MNOs to benefit from the flexibility and speed afforded through adoption of modern DevOps techniques employed by the leading cloud service providers and cloud-based digital enterprises. The continuous integration and continuous delivery (CI/CD) approach allows MNOs to conceive, develop, test, deploy, and enhance 5G services at speed and scale, which holds the potential to unleash a wave of unprecedented service innovation in the telecom industry.

In a fully autonomous 5G network, where services are provisioned and enabled in a fully automated manner, without human intervention, all software components will be instrumented so that they generate the metrics, logs, and traces needed to monitor the state of the network and the behavior of connected devices. All of these data types will be standardized and passed between various software functions using standard application programming interfaces (APIs).

To make this happen, you first need to understand the Network Data Analytics Function (NWDAF).

Network Data Analytics Function

The NWDAF is the 3GPP (3rd Generation Partnership Project) standard network function that provides real-time operational intelligence in the 5G Core (5GC). NWDAF enables closed-loop network automation. MNOs require this to successfully deploy and operate 5G networks at scale, to continuously validate the network in real-time, identify problems, and take automated actions.

NWDAF data enhances the relationship between an MNO and its enterprise customers, providing an enterprise application with a network view of user equipment (UE) behavior, enabling the enterprise to confirm correct device operation and security.

NWDAF is a network function (NF) introduced as part of the 3GPP standards for the 5GC that:

- Is a centralized data collection and source for real-time operational intelligence about the 5GC network functions—unifying multiple tools into a single, logically centralized, distributed analytics solution. NWDAF brings a standard protocol and format for core network analytics.
- Collects data from various 5GC NFs and application functions as well as operations, administration, and management (OAM) systems, and operational support systems. Data types include 5GC service-based interface (SBI), control or user planes, logs, and data from UE.
- Acts as an analytics service for other NFs in the 5GC and for the OAM—deriving analytic insights from the data (telemetry, measures, and reference data). The output is designed specifically to be consumed by other 5GC network functions to allow *fast* control loop for automation.

  Analytics functions specified in the current release (Release 16) are:

  - Abnormal UE behavior
  - Expected UE behavior
  - Network Performance
  - NF load
  - Observed Service Experience
  - QoS Sustainability
  - Slice load level
  - UE Communication
  - UE Mobility
  - User Data Congestion

Why did the 3GPP define NWDAF as part of the 5G Service-Based Architecture?

The 3GPP recognized the need for machine intelligence in conceiving its 5G Service-Based Architecture (SBA).

This 5G SBA provides the overarching framework for the set of standards which govern the operation of the 5G radio access network (RAN) and 5GC. In contrast to 5G, the 3GPP standards for 4G LTE were conceived at a time when real-time streaming analytics, artificial intelligence (AI) and machine learning (ML) were not fully developed.

The 3GPP 5G SBA explicitly defines data analytics functions governing operation of 5G RAN and 5GC to deliver operational intelligence for network automation and service orchestration functions in the 5GC. These analytics functions inform the provisioning, performance, and fault management functions in the OAM layer.

What 3GPP has done in the 5G SBA is new, specifying standards for collecting and analyzing 5GC network and management data. On the collection side, this much-needed advancement circumvents the efforts related to data ingestion, data formatting, and data normalization that operators have in 3G and 4G networks, because they were forced to graft data analytics onto existing infrastructure.

On the analytics side, the standards specify the set of services a data analytics function must provide to other 5GC functions, and this is codified in a set of well-defined APIs. Think of NWDAF and management data analytics functionality as standard *form factors* that must be used to *plug* analytics processing into 5G infrastructure. This normalization accelerates the adoption of data analytics, which enables MNOs to automate the operation of 5G networks across multiple operation domains.

NWDAF use cases

3GPP Release 16, which was frozen in 2020, identifies the following NWDAF analytic IDs, which each address particular operational requirements:

- Current and predicted congestion for a specific location, device, or group of devices
- Load analytics information and prediction for specific NF instances or types
- Network performance computation and prediction
- Network slice instance load level computation and prediction
- Quality of service (QoS) sustainability—reporting and predicting QoS change
- Service experience computation and prediction for an application or network slice
- UE abnormal behavior or anomaly detection
- UE communication analytics and pattern prediction
- UE mobility analytics and expected behavior prediction

Although these use cases encompass a broad range of 5G operational requirements, this list is only a starting point. The 3GPP will define more use cases in future releases. But what considerations drove the 3GPP to focus on this initial set?

NWDAF operational intelligence

NWDAF in 3GPP Release 16 defines a set of use cases for network data analytics that encompass a range of 5G operational requirements, which can be logically grouped into three categories:

- Network conditions
- Device behavior
- Service experience

Release 16 NWDAF use cases provide real-time insights into these critical aspects of network and service operations. At a high level, the objectives are straightforward.

Operators need to know—in real time—how users and applications are impacting the 5G network, or how adverse network conditions or unusual UE behavior is impacting users and applications.

Regardless, operators need to constantly monitor service experience metrics for each type of 5G service.

The ultimate goal is network automation, but each Release 16 NWDAF use case is also applicable to operator-centric 5G network operations, in which humans evaluate the operational intelligence generated by NWDAF and make decisions to manually initiate service orchestration and network operations workflows.

Here is a quick run-down of the types of monitoring and analytics that NWDAF is concerned with in each category of operational intelligence.

Note that NWDAF calls for using statistical analytics to determine what is currently happening in the 5G network (or what has happened in the past) and predictive analytics to determine what is expected to happen in the future, based on current or historical trends.

Predictive analytics can play a key role in alerting operators to looming problems that could result in service degradation or even possible outages.

Network conditions

Real-time network monitoring is typically per slice and involves generating performance metrics such as throughput, latency, and connection setups, measuring network load and detecting congestion or other types of network performance anomalies.

- Network slice instance load level computation and prediction
- Load analytics information and prediction for a specific NF
- Network performance computation and prediction
- Congestion information—current and predicted for a specific location
- QoS sustainability—reporting and predicting QoS change

Device behavior

Human users as well as machines will be connected to the 5G network. 5G operations require monitoring device connectivity, mobility, and communications patterns.

Abnormal usage or unanticipated behavior by human users or smart machines could have a negative impact on network performance. The 3GPP specified analytic IDs that correspond to device behavior provide data about:

- Network slice instance load level computation and prediction

- UE mobility analytics and expected behavior prediction
- UE abnormal behavior detection and anomaly detection
- UE communication analytics and pattern prediction

Service experience

MNOs need to measure and track service experience to ensure that the 5G network is meeting user expectations as well as stringent performance and reliability service level agreements (SLAs). NWDAF-generated metrics will measure service experience by user, application type, device group or geographic location. The specification calls for tracking QoS sustainability to monitor the current quality of service and predict future changes.

- Service experience computation and prediction for an application or UE group
- Application service experience computation and prediction

5GC and NWDAF

The Network Data Analytic Function provides operational insights to other 5G Core NFs as well as OAM (orchestration and management) systems. These same 5G NFs, and OAM systems provide the input data on which NWDAF analytics are built. This creates some requirements for the NWDAF in the 5GC.

NWDAF inputs

The analytics IDs defined by the 3GPP in TS 23.288 use inputs from sources in the 5GC network including the Access and Mobility Management Function (AMF), Session Management Function (SMF), and Network Repository Function (NRF), as well as inputs from application functions (directly for trusted application functions or indirectly via NEF for untrusted application functions), the User Plane Function (UPF), and OAM functions.

Input data from the 5GC sources is collected through the SBI defined for all 5GC functions. While the input data from the application functions, UPF, and OAM is defined in 23.288, the mechanism to collect UPF input data is not defined in the Release 16 specification. The OAM services and reporting mechanisms accessed by the NWDAF are defined in TS 28.532, TS 28.550, and TS 28.545. Mechanisms such as managed object retrieval and file transfer are used by the NWDAF for OAM data collection.

These requirements drive the need for an NWDAF that has flexible ingestion processing, is able to parse a variety of data formats through a variety of interfaces. It is also valuable for the data to be collected close to the point of data generation.

This avoids the cost - in both time and network bandwidth - to transport all generated source data to a remote collection point. Data collected at the point of generation can be aggregated in a way that serves the analytics use case, thereby reducing the amount of data that must be transmitted and collected for analytics.

Taking this approach also reduces cost, latency, and network bandwidth requirements. For example, per-event data, such as per packet, can be aggregated into micro batches for each UE, while still offering the necessary granularity to provide analytics insights, while transmitting significantly fewer bytes across the network.

Data collection at the point of generation also enables analytics scoring to be performed locally. This scoring can be returned to analytics consumers for local decision-making

For example, mobile edge computing (MEC) management will be making optimizations and resource allocations locally, while centralized MEC management will be performing network-wide MEC resource optimization and management.

NWDAF outputs

The 3GPP has defined microservices that are offered by NWDAF to allow analytics consumers to obtain desired information through subscription or request-response interactions.

In Release 16, these are Nnwdaf_AnalyticsSubscription and Nnwdaf_AnalyticsInfo. Additional services are added in subsequent releases. These microservices can be used by applications outside of the 5GC, including orchestration systems and management data analytic service.

Some scenarios are not well served with a service-based interface. For example, a reporting system that makes the operational data available for medium-term and long-term analysis, such as customer churn analysis or network planning, benefits from receiving a copy of the data generated and stored by the NWDAF.

Outputs beyond the 3GPP SBI, such as data export or message bus interfaces, are valuable in serving the variety of applications that consume and process the operational intelligence information produced by NWDAF.

NWDAF deployment challenges and recommendations

There are many choices that must be made in the creation of a 5G NWDAF, and some capabilities were taken into consideration in the design of the Guavus 5G-IQ NWDAF. Guavus is a telecom analytics pioneer focused on enabling 5G MNOs to realize business value by using cloud-scale analytics and state-of-the-art AI/ML to increase operating efficiency, satisfy stringent 5G SLAs, and ensure QoS experience for users, machines, and Internet of Things (IoT) devices. Guavus has been a member of the AWS Partner Network (APN) since 2019.

Non-standard data inputs

The 3GPP specification has defined data to be produced by the 5GC network functions to enable automation. This data is collected by the NWDAF, enriched, aggregated, and provided to the consuming applications in the 5GC management domain and to outside application functions through the Network Exposure Function (NEF).

To provide timely data to allow decision making by the application and network functions, the NWDAF must produce operational intelligence in real time. The Guavus 5G-IQ NWDAF relies on stream processing to avoid the store-index-retrieve steps that are required in a typical relational database management system data lake.

Even with this specification, not everything has been standardized. For example, TS 23.288 §6.5.2 notes “How NWDAF collects information [for UPF load analytics] is not defined in this Release of the specification.”

The ability to ingest 3GPP SBI data and a variety of non-standard data from 3GPP functions such as OAM and UPF, as well as non-3GPP sources such as synthetic transaction generators and network probes is necessary as 5G networks are deployed.

Analytic algorithm selection

NWDAF use cases are designed to enable operation of 5G networks in a smarter way, by adopting a data-driven decision-making strategy. The efficiency and intelligence of the 5G network increases when real-time operational intelligence and predictive analytics can be used by network functions, orchestration and management functions in admission control, session management and provisioning decisions.

This moves from the deterministic calculation domain to a probabilistic ML domain. Thus, CSPs must also deal with uncertainties in ML-based predictions, since the output is now represented as the probability of some event to occur, unlike deterministic, statistical (past-based) analytics where there is no ambiguity.

ML is a data-driven exercise, where every model generated by an ML algorithm is tied to the data patterns observed during the model training phase. As data patterns change with time, or the relationship between the predicted variable and input data changes, performance of ML predictions will degrade, and the output will reflect a lower confidence score.

The Guavus 5G-IQ NWDAF MLOps framework automatically monitors model quality, enabling the entire system of model training, prediction, and evaluation to run in an automated self-correcting loop.

5G-IQ NWDAF MLOps provides for A/B testing before pushing a trained model to production to ensure that only the best performing model is in action on field. This framework is enhanced by Guavus’ proprietary drift detection libraries over streaming data to catch degradation signals early.

In addition to the predictive analytics defined by the 3GPP, Guavus has developed a load index that can be used to dynamically upscale and downscale network slices per the predicted load demand.

This goes beyond regular load prediction modules and analyzes changes in current load patterns. This system is more effective than a typical rule-based threshold system because it can differentiate between momentary load spikes versus a systematic trend toward a new load pattern.

The 5G-IQ analytic module that uses this index uses a patent pending technology developed in Guavus research labs to collect behaviors of multiple load key performance indicators (KPIs) for each NF per slice and collect them into one single index for the entire slice.

By extending the same capability to various services in the 5G network, multiple service KPIs can be collected into an ensemble to create an index that can alert degradations in service health before they manifest in the network.

For ML algorithms to work efficiently in an automated 5G network, there are some key considerations in choosing and designing them. So, all ML-based solutions for NWDAF are designed to adhere to the following conditions:

- ML implementation at high speeds
- ML implementation at high volumes
- ML implementation at low footprint

The Guavus solution applies all of the previously mentioned methodologies for analytics algorithm selection to ensure efficient execution of NWDAF use cases.

Deployment scenarios

It is important that NWDAF provide insights to analytic consumers quickly, to allow them to make operational decisions. This must be done while minimizing the burden of providing data on source functions.

Efficient and effective collection, enrichment, aggregation, and scoring by 5G-IQ NWDAF at the point of data generation minimizes resource consumption and allows localized decisioning and optimizations based on the local data. This decreases the network bandwidth, transmission time, and storage that would otherwise be required to transmit the data to a central server for processing.

To make 5G operations more effective, the Guavus ML architecture incorporates key design ideas to work on real-time streaming data in a distributed manner over the edge. To ensure data security for sensitive applications, the distributed architecture uses federated learning to enable distributed ML along with preserving data privacy at the edge. The data is not brought to the center, and partial model training happens at the edge. Only the learnings are consolidated at the central servers.

There are a wide variety of other use cases in 5G that can make operations more efficient if ML is applied at the UE level to do some inferencing at that level. Then the network can be improved if it is able to track UE level mobility at an individual or group level to predict velocity and direction for efficient signaling and reduced power consumption. Then, ML applied at the edge can be used to lower network traffic and optimize links.

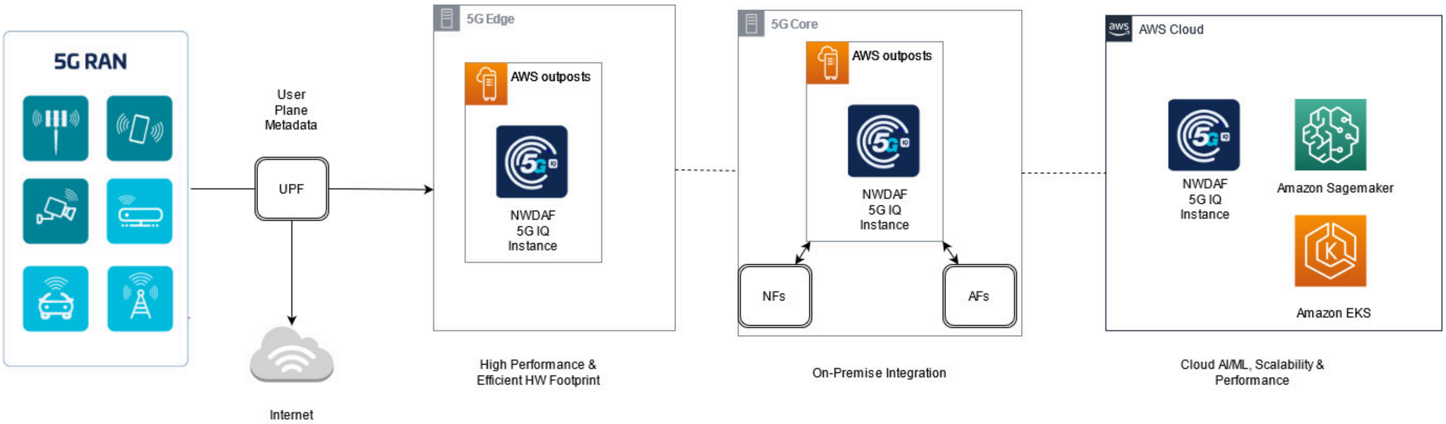
To enable this, some key ML design considerations for 5G use cases are:

- Distributed inference on the edge
- Lightweight inference over devices
- High-speed inference over real-time streams
- Distributed training over a compute cluster
- Low communication bandwidth over interconnected machines
- Federated learning

5G-IQ data collection manager

The Guavus 5G-IQ NWDAF data collection manager is built on Guavus SQLstream technology, Guavus SQLstream acquires data from multiple, disparate sources with automated discovery features customers can connect to data streams and start interacting with them with no coding necessary. The 5G-IQ NWDAF data collection manager can be distributed to the network edge, co-located with the data source. With this Guavus streaming analytics engine, the 5G-IQ data collection manager can collect data from 5G SBI interfaces as well as proprietary interfaces from network-specific or service-specific data sources.

The 5G-IQ data collection manager also can be co-located with the centralized 5G-IQ NWDAF components, supporting streaming data collection, enrichment, aggregation, and scoring as data reaches it.



NWDAF in the 5GC

Whether the 5G-IQ data collection manager is distributed to the network edge or co-located with the rest of the 5G-IQ NWDAF components, data is transmitted into the rest of the 5G-IQ NWDAF application through a Kafka bus. This Kafka bus serves as an intermediate storage solution, providing elasticity for any processing or data transmission speed mismatch among the components, as well as being an efficient way to connect multiple internal data consumers to the data source.

NWDAF benefits on the AWS Cloud

NWDAF workloads infrastructure consumption can be unpredictable, depending on the NF subscription, number of slices, number of devices, and other variables, building to maximum capacity that is an inefficient use of capital and creates poor return rates.

In addition, workloads that are unpredictable and may fluctuate from day to day, or even during a day, make it difficult to predict exact requirements. To overcome this uncertainty, a level of contingency is always added, allowing for expansion where needed, but at the expense of additional costs.

NWDAF requires running large queries and one-time analytics for use in creating reports, and exposes APIs for NF consumption. Rapid scaling and automatic scaling are not possible with in-house systems, and careful scheduling is needed to resolve competition for fixed resources.

AWS Regions runs efficiently and much more flexibly in scaling to meet demands than in-house on-premises data center solutions. AWS services such as Amazon Elastic Compute Cloud (Amazon EC2), Amazon Simple Storage Service (Amazon S3), or Amazon EMR, offer an ability to dynamically expand and shrink capacity depending on application needs and higher utilization rates help to reduce costs.

AWS analytics (Amazon Kinesis, Amazon EMR, Amazon Athena, Amazon Redshift, or Amazon SageMaker) deliver processing power efficiently, and mean that communications service providers (CSPs) do not have to invest in expensive data servers to support peak workloads but can grow capacity when required. AWS supports native implementation of DevOps processes, which ensure that operators can natively support greater operational efficiencies.

NWDAF promotes intelligent orchestration

Automation will play a vital role in 5G network operation, especially with the creation of network slices. AWS is promoting several strategic initiatives on 5G, including network orchestration, that helps CSPs to deploy and manage 5G network automation. AWS telco cloud solutions provide CSPs the ability to reduce the time to market for new services, assure service performance, increase revenue, and stay ahead of the competition.

AIML enabled automation ensures error-free deployment of new services, smarter self-healing capabilities, efficient and optimized use of network resources, and can free up human resources from manual and repetitive tasks associated with planning, design, and deployment.

Although specifications don't necessarily refer to the interdependency of NWDAF and orchestration, CSPs expect NWDAF will empower intelligent orchestration and activation of dynamic end-to-end (E2E) network slices across the RAN, X-haul, 5GC, and cloud; seamlessly scale bandwidth requirements of 5G deployments; optimize the network to enable better load spreading; and assure customer experiences are constantly optimized for their varied requirements.

Because of 5G standalone deployment with a combination of edge and cloud, NWDAF architecturally spans edge to cloud in support of NFs being deployed at the Edge or Region, and AWS infrastructure supporting edge services such as AWS Outposts, AWS Wavelength, and AWS core services in the AWS Regions.

NWDAF deployment at MEC

AWS offers AWS Local Zones, AWS Outposts, and AWS Wavelength to CSPs, depending on various use cases. It is foreseen that NWDAF, apart from listed use cases in the specification, will be become essential for all analytical needs, at the Edge or Region, for network optimization, network planning, or customer experience. Some of these additional use cases listed in the *About Guavus* section use NWDAF closer to edge (MEC) AWS Wavelength, Outposts, and Local Zones. Refer to NWDAF edge architecture.

Data location and speed is critical in operational systems requiring real-time decision making. The ability to capture, transform, and load data in real-time, when CSPs have many legacy systems in place; the timeliness in which operational intelligence outputs are required; operations in the face of missing or uncertain data; and the sheer volume of data emitted by the mobile network combine to create requirements that cannot be met by traditional analytic systems.

Support of real-time, complete streaming data pipelines (ingest, prepare, analyze, and act) for NWDAF is provided in AWS without CSPs needing to deploy new tools or infrastructure. Amazon Kinesis supports streaming data ingestion and analysis without the need for network or IT operations teams to manually scale infrastructure.

AWS solves data burst and staffing challenges

The scale and breadth of available data from NFs, and the amount of data that can be streamed in real time, is unique to NWDAF. Data volumes are much larger than existing analytics because NWDAF can be called as NFs interacting with other NFs and is part of the 5G call flow, which puts significant pressure on data infrastructure.

These problems can be addressed by using highly scalable AWS Cloud offerings and flexible consumption-based pricing models. NWDAF functionality is more than slice level information.

For example, as the self-optimized networks (SON) concept in 5G expands to the Core and RAN complete optimization, NWDAF can be used as an additional information source for SON—in particular, for collection of E2E service quality measurements and slice level performance indicators.

This would bring in additional demand for skilled staff to develop new use cases to further monetize NWDAF. Many CSPs outsource high-cost data science and data mining specialists to vendors to overcome staffing shortages.

NWDAF on AWS can overcome the staffing challenge for developing new use cases with the help of AI/ML stack and SageMaker Studio, which helps CSP data scientists and developers reduce data cleaning jobs and focus on building algorithms at a faster pace.

AWS Lake Formation supports the simplification of incumbent data and analytics solutions into a fully integrated and highly automated infrastructure, allowing current staff to focus on analytics and data issues, not systems administration activities, and therefore be more productive.

AWS overcomes data quality challenges

Because NWDAF is still at a nascent stage, there is a high possibility that NF vendors interacting with NWDAF are not 100% aligned to specifications. Data quality issues can be addressed by data consolidation to create a single version of the operational data with AWS Lake Formation, avoiding data synchronization and versioning issues and enabling the use of advanced data preparation using automated ML techniques to improve data quality.

NWDAF total cost of ownership

AWS services run at scale and afford efficiencies that are harder to achieve in a CSP's own data infrastructure or any other cloud infrastructure. AWS ensure that hardware is updated, maintenance schedules are performed, and integrations are provided to partnerships.

A pay-as-you-grow model ensures that costs are more aligned to revenue within the business. Although volume-based pricing can be more complex than traditional licensing and purchases, hybrid pricing approaches based on Reserved Instances and volume-based pricing enable the best of both worlds, and link usage and business needs better with the growth of a telecom operator or in support of the services that it is offering.

Reference architecture on AWS

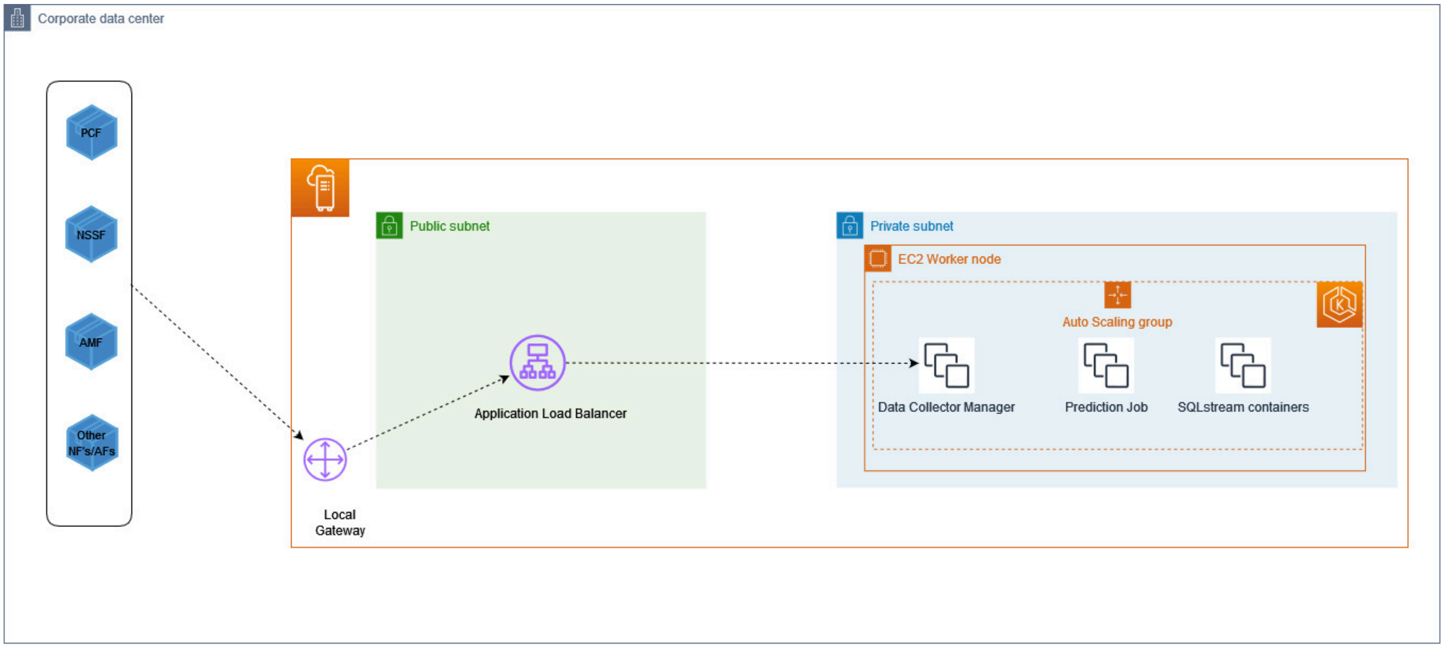
5G-IQ NWDAF uses AWS services to provide a flexible, containerized solution that can follow the distributed deployment of the 5G Core. NWDAF architecturally spans “edge to cloud” to provide a cost and time-effective solution, putting data collection, enrichment, aggregation and scoring where it is most effective for the CSP’s operational needs at that instant.

This is done with a Central NWDAF Module in communication with one or more Edge NWDAF Modules that can be co-located with the Central Module or be distributed to the edge of the 5G Core. The Edge and Central Modules provide storage, 5G Core registration, Machine Learning Operations (MLOps) as well as supporting the standardized NWDAF 5GC APIs.

Edge NWDAF module

As discussed previously, having an NWDAF sitting on the edge, close to data source and analytic consumer NFs, is key when addressing low latency use cases. This enables collection of data feeds from the NFs in real time, as well as hosting the prediction models near the analytic consumer, hence exposing the inferences to the analytic consumer locally, rather than having the need to queue a request for prediction back to a distant, centralized server.

AWS Outposts will host the NWDAF edge solution, located in the customer’s on-premises data center. AWS Outposts supports Elastic Load Balancing, providing a single and scalable point of entry for Rest API messages from the NFs to the NWDAF. In addition, AWS Outposts supports Amazon Elastic Kubernetes Service (Amazon EKS), hence the EC2 instances running on AWS Outposts will serve as the worker nodes for EKS, where the actual components will run.



Guavus 5G-IQ Edge NWDAF AWS architecture

Central NWDAF Module

All edge modules communicate with the central module for getting trained models, NRF registration data, and data collections requests.

The 5G-IQ NWDAF central module is architected across multiple Availability Zones to achieve high availability in case of an Availability Zone failover. All EC2 worker nodes run inside an Amazon Virtual Private Cloud (Amazon VPC) in private subnets. This allows isolation of the NWDAF central module from the public internet.

Specific security group and access control can be defined to the private subnet from AWS Outposts which hosts the NWDAF edge module.

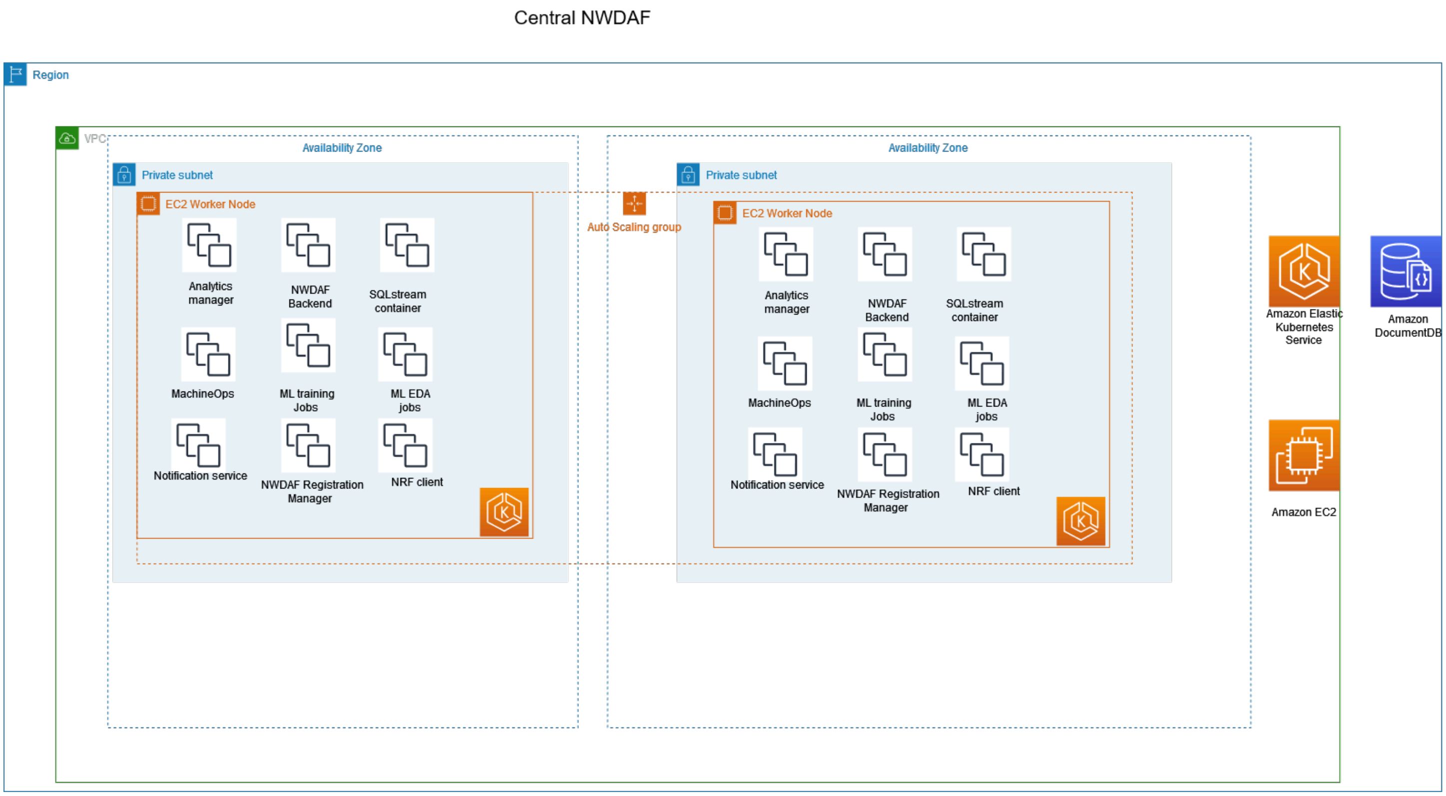
The Guavus NWDAF solution is packaged as Docker images and deployed through Helm charts to Amazon Elastic Kubernetes Service (Amazon EKS), an AWS managed Kubernetes cluster. Helm charts are available in the Guavus NWDAF installation guide, and are provided to customers during deployment.

Amazon EKS provides a scalable and highly available control plane that runs across multiple Availability Zones to eliminate the chance of a single point of failure. Kubernetes master nodes are managed by AWS. Kubernetes worker nodes are deployed on Amazon EC2 instances, which are scaled and distributed across Availability Zones independently. Docker images are pulled from an Amazon Elastic Container Registry (a fully managed Docker container registry).

EC2 worker nodes will run across multiple AWS Availability Zones. Running Kubernetes in a single Availability Zone can make the application highly available with EC2 instance failures, but it will not protect against outages that affect the entire Availability Zone. It is best practice to make each cluster highly available in multiple Availability Zones so that applications are still available in case of zone failures.

In AWS, the recommended way to run highly available Kubernetes clusters is by using Amazon EKS with worker nodes spread across three or more Availability Zones. For globally available applications, running separate clusters in different regions with multi-zone worker nodes is recommended.

AWS has published the blog post *[Creating Kubernetes Auto Scaling Groups for Multiple Availability Zones]()* with additional information on this topic.



Guavus 5G-IQ NWDAF AWS central architecture

Document-oriented database (Amazon DocumentDB)

Guavus NWDAF uses Amazon DocumentDB for its persistence infrastructure. Amazon DocumentDB (with MongoDB compatibility) is a fast, reliable, and fully managed database service. Amazon DocumentDB makes it easy to set up, operate, and scale MongoDB-compatible databases in the cloud. With Amazon DocumentDB, you can run the same application code and use the same drivers and tools that are used with MongoDB.

Amazon DocumentDB allows fast retrieval of data being written. KPIs and data can be accessed later by custom Lambda functions that can expose the data for various use cases.

It will store the following data:

- Live streaming data from NFs and AFs
- Predicted results for the notification subscribed services from NFs
- Models and model metadata
- Subscription data and unsubscription data
- Other application configurations

NWDAF registration manager

NWDAF registration manager handles all the NWDAF subscription registration requests. NFs will be calling the registration API exposed by NWDAF registration manager for NWDAF Events Subscription.

MLOps

MLOps manages the lifecycle of model training and fine-tuning model quality. It reads the historically predicted data and live NFs and AFs data from an Amazon DocumentDB database and computes the delta between predicted and actual values. The accuracy of prediction results will be calculated based on the difference between predicted values and live data. It will invoke the retraining of model in case the prediction discrepancy is beyond a configured or desired threshold.

Exploratory data analysis jobs

Exploratory data analyzes various data properties to assess data quality and statistics. It compares current data distributions with the data distribution observed while training the model as well as.

It also analyzes correlation and strength of features with respect to the predicted variable at a high level and finally, it ensures the data patterns remain suitable for the ML algorithm.

NWDAF backend APIs

The NWDAF backend APIs will read the event subscriptions from the subscription Kafka topic and will perform the subscription.

Conclusion

This whitepaper introduced the concept of NWDAF, covered the benefits of deploying NWDAF on AWS, and explained how the Guavus NWDAF solution is architected on AWS. 5G Standalone deployments are in their early days of adoption globally. NWDAF is also in its early days.

Because NWDAF is a key component in 5G standalone architecture, using AWS edge services (such as AWS Outposts) for the NWDAF edge capability — as well using the AWS data analytics and ML stack — it is key to enabling scalable NWDAF deployments that empower CSPs to act on real-time insights from their core networks.

About Guavus

Field-proven, carrier-grade telecom analytics experience

Guavus is a telecom analytics pioneer that is focused on enabling 5G MNOs to realize business value by using cloud-scale analytics and state-of-the-art AI/ML to increase operating efficiency, satisfy stringent 5G SLAs, and ensure QoS experience for users, machines, and Internet of Things (IoT) devices. The Guavus team combines in-depth telecom domain experience with data science and software engineering expertise to deliver high-performance, carrier-grade analytics solutions that the world's leading network operators are using to collect and analyze petabytes of data every day. For more than 15 years, Guavus has been at the forefront of applying big data, streaming analytics, and AI/ML to solutions that enable telecom network operators to overcome critical business challenges impacting customer experience, service quality, operating efficiency, revenue, and profitability.

Lean innovator inside a multi-billion euro technology leader

A Thales company, Guavus is a lean innovator inside a global technology leader with 80,000 employees in 68 countries, revenue of 17 billion euros (in 2020), and a self-funded research and development budget of 1 billion euros. Guavus works closely with non-telecom Thales business units to provide analytics solutions for applications in industrial IoT, aerospace, transportation, and digital security. Guavus also partners with leading network equipment vendors and cloud service providers to further extend its reach to customers around the world.

Contributors

Contributors to this document include:

- Andrew Colby, Head of 5G Product Management and Strategy, Guavus
- Miraj Godha, Principal Lead Architect, Guavus
- Guy Ben-Baruch, Senior Telco Partner solutions architect, 5G and Network Analytics, AWS
- Naresh Rao, Principal – Portfolio Lead, Telecom (Service Assurance and Network Analytics), AWS

Document revisions

Notices

AWS Glossary

For the latest AWS terminology, see the AWS glossary in the *AWS Glossary Reference*.