Planning for the ngVLA Cyberinfrastructure

Jennifer Schopf, Dan Stanzione, Niall Gaffney

Texas Advanced Computer Center, University of Texas at Austin

Rachel Rosen

National Radio Astronomy Observatory

May 7, 2026

Version 2.0

Corresponding Author: Jennifer M. Schopf, jms@tacc.utexas.edu

1. Introduction

The goal of this document is to detail a cost analysis set of tradeoffs for the conceptual design of the next generation Very Large Array (ngVLA) Data Processing Center (DPC). The DPC will host computing and storage resources, and as well as supporting the data transfer architecture to support data processing for ngVLA operations. This document focuses on the data path from the Correlator Back End (CBE), the on-site collection point for initial analysis of telescope data, to the Data Processing Center, the production of Data Cubes, and serving that data to the broader science community.

This document is a result of ongoing conversations and consultation between a team at the Texas Advanced Computing Center (TACC) and the ngVLA Computing and Software Subsystem (CSS) Integrated Product Team (IPT). It captures the understanding of the ngVLA system requirements as of August 2025, as made available, and will be revised as the design continues. The collaboration is ongoing, and the TACC team will be continuing to support the ngVLA project as it progresses.

We detail four possible approaches, first by presenting the tradeoffs involved with procuring **systems** that will provide the storage and processing needs, or by meeting the scientific needs of the facility through procuring **services** from a 3rd party provider. Secondly, each of these top-level choices triggers additional possible choices. For a systems approach, should an ngVLA-staff supported facility (assumed to be in Albuquerque) be constructed, or should a colocation provider be leveraged? What are the associated staffing costs beyond the hardware, and who will provide them? For a services approach, is the best source of services a commercial cloud service or academic service provider?

For each choice, we translate the scientific requirements of the ngVLA into technical requirements: describe the data transport, networking, processing and storage needs, then consider various ways to satisfy these needs through different solutions, complete with a look at lifecycle costs. For the purposes of the analysis in this document, we consider a scenario where project construction begins in 2030 with initial datacenter deployment, with live data sets to begin arriving (at a minimal rate) from the first antennas in 2032, reaching full production capacity from all antennas by 2039. To provide a full lifecycle analysis, we also consider the cost implications of 5 years of operations after the end of construction, through 2044, though we detail costs for the Construction and Operations periods separately. Any prices listed are based on estimates in August 2025, except where noted.

The document presents two potential hardware architectures we recommend carrying forward from Conceptual Design to Preliminary Design, and compares two ways to deploy and operate them, providing four choices. An additional four choices are presented to acquire the capabilities of these two architectures through commercial or non-commercial cloud services.

Table 1: Summary of final comparison for four alternative architectures.

Solution for System and Service Options	Cost through 2044
Option 1: Purchase Hardware deploy to CoLo	~2.3x Option 2
Option 2: Purchase Hardware deploy to LCCF	Least Expensive

Option 3: Purchase Service through Commercial Cloud	~12.4x Option 2
Option 4: Purchase Service through LCCF	~2x Option 2

The final comparison of the four alternatives is for a CPU-based architecture. The solutions include acquisition of systems or services, maintenance costs, housing, power, and cooling costs, and estimates for staffing needs are detailed in the text. We also outline the staffing needs for the project, both to support the basic system and if other support, such as software development and tuning, is required by the project.

Cost is not the only implication to consider. Different solutions provide different levels of support and investment in a successful outcome. Commercial cloud solutions often come with lock-in due to the cost of egress of data. Hosted services do provide a more flexible scale-up of the system, as you only pay for what you are using. We provide these as examples of the information needed to do a future cost-benefit evaluation of the different possibilities. We also discuss the projected cost-benefit of GPU-based computing and the need to monitor that closely before procuring any hardware.

TACC
TEXAS ADVANCED COMPUTING CENTER

ngvla

Table of Contents

1. Introduction 2
2.Translation of Science Requirements to Technical Requirements 6
    2.1 Terminology/Glossary 6
    2.2 High Level Workflow and Document Scope 7
    2.3 Data Center Ingest Rates 8
    2.4 Workflow Production Rates 9
    2.5 Compute Requirements 9
    2.6 Out of Scope Items 10
    2.7 Technical Requirements Summary 10
3. Hardware Solutions Analysis 12
    3.1 Technology Forecasting 12
    3.2 Network Planning 13
        3.2.1 Network from Socorro to a Metro Area 14
        3.2.2 Network from Metro area to the DPC 16
    3.3 Data Storage and Archive Capacity Planning 18
        3.3.1 Magnetic Tape 19
        3.3.3 Solid State Drives 21
        3.3.3 Hard Drives 22
        3.3.4 ngVLA Storage System Design Recommendation 23
    3.4 Compute Systems 24
        3.4.1 Pure CPU-Based System 25
        3.4.2 Pure GPU-Based System 25
        3.4.3 Interconnect 26
    3.5 Recommendation 26
    3.6 Additional Considerations 27
        3.6.1 Power and Cooling Loads 27
        3.6.2 Floor Space 28
        3.6.3 Staffing Requirements 28
4. Colocation/Hosting Model Options 29
    4.1 ngVLA Colocation Facility in ABQ (or elsewhere) 29
    4.2 TACC Hosted Facility 31
5. Service Model 32
    5.1 Commercial Service Forecasting 32
        5.1.1 Cloud CPU Compute Cost Calculation 33
        5.1.2 GPU Compute Cost Calculation 34

5.1.3 Cloud Storage 34
5.2 TACC Service Forecasting 35
5.3 Cloud Staffing Estimate 35
6. Other Cyber Requirements 37
7. Analysis and Summary 38
8. Conclusion 38

2.Translation of Science Requirements to Technical Requirements

2.1 Terminology/Glossary

Archive storage refers to data that must be kept for the long term, but where a long retrieval time (minutes to days) is acceptable. Archive or offline storage can be achieved through tape, “cold” disk (not currently powered on), or through cloud services. Archive storage capability is defined by capacity, performance, reliability (uptime and replicas) and cost.

CBE is the ngVLA Correlator Backend computing stage that takes raw correlator output, integrates and formats it, and produces raw, uncalibrated visibility data for post-processing.

CoLo, or **Colocation,** is the practice of renting space in a third party data center to house IT infrastructure such as compute and storage.

A **data cube** is a 3D structure composed by a stack of images, each one for a different frequency, that are produced as the outcome of the workflow at the Data Processing Center.

Data transfer rates, or **bandwidth**, are described in Gigabits per second (Gb/s) or Terabits per second (Tb/s). Note 1 byte = 8 bits. Therefore, moving 1TB of data in 10 minutes requires (10TB/600s)*8= 130Gb/s of bandwidth.

Data volume amounts are historically described in units of GigaBytes (GB - roughly 10^9 bytes, exactly 2^30 bytes), TeraBytes (TB, 10^12, 2^40), PetaBytes (PB, 10^15, 2^50), or ExaBytes (EB, 10^18, 2^60) of storage.

DP - Double Precision, sometimes also referred to as double-precision floating-point format, FP64, or float64, is a floating-point number format usually occupying 64 bits in computer memory.

FP64 - Floating Point 64, sometimes also referred to as Double Precision (DP), double-precision floating-point format, or float64, is a floating-point number format usually occupying 64 bits in computer memory.

An **image**, in this context, is a 2D structure of the quantity of interest, usually brightness intensity measured as Jy/beam.

LTO- Linear Tape Open standard [LTO].

NVME - Non Volatile Memory Express storage.

SSD - Solid State Drives, sometimes also called Flash drives.

Storage refers to the total data kept by the project. Storage is kept online (via disk, solid state, or a cloud service) and is the working data space for the project. Online storage is expected to

have relatively quick retrieval times for a wide variety of data sets. Online storage capability is defined by capacity, performance, reliability (uptime and redundancy), and cost.

TPC - Total Project Cost per the NSF Research Infrastructure Guide (https://nsf-gov-resources.nsf.gov/files/Research-Infrastructure-Guide-January-2025.pdf).
Visibilities - Also referred to as **uncalibrated visibilities**, **raw visibilities**, or **formatted visibilities**, these are the data products that are sent from the Correlator Back End (CBE) to the Data Processing Center (DPC).

2.2 High Level Workflow and Document Scope

The primary data product of the ngVLA array are uncalibrated ("raw") and formatted visibilities, referred in this document as "**visibilities**". Visibilities are sent from the Correlator Back End (CBE) to the Data Processing Center (DPC) at rates described in Table 2. A copy of the Raw Visibilities should be archived in long term data space (archive). The Raw Visibilities are also run through the compute workflow to produce **data cubes**, a set of stacked images. Each **image** is a 2D structure of the quantity of interest, usually brightness intensity measured as Jy/Beam. Each data cube is a series of stacked images, where each image is a different frequency. **Data Cubes**, which are made available to collaborators more broadly through online storage. In most cases, a one-year **proprietary period** on the data products is expected, so the products will only be made available to the PI and team during the first year.

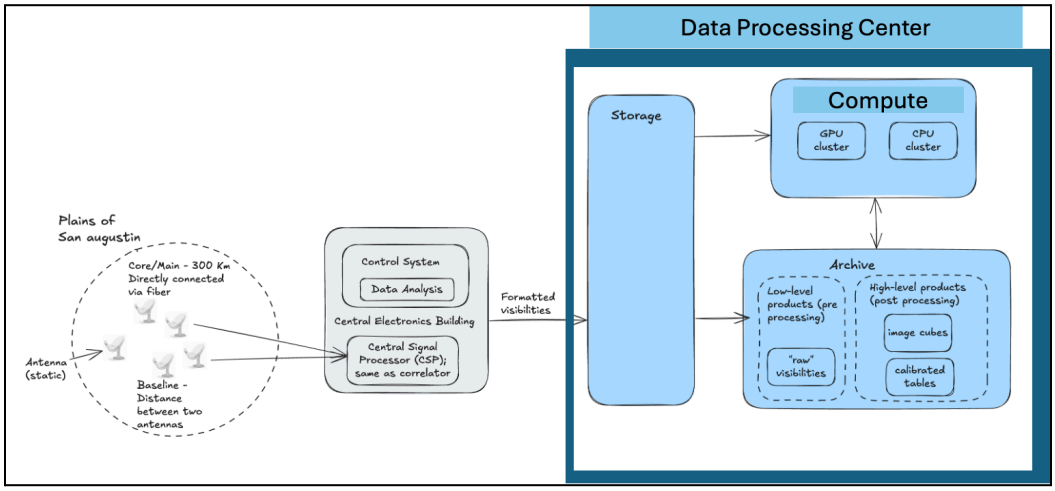


Figure 1: Rough schematic showing the flow of data from the Antennae in the array, located primarily in the Plains of Augustin, which then flows to the Central Electronics Building (CBE) for initial data processing and correlation. This produces formatted, uncalibrated ("raw") visibilities, which will be sent to a Data Processing Center, as depicted.

This document is focused on the workflow of the data once it leaves the CBE at the rates projected in Table 2, the data storage and longer-term archiving at the Data Processing Center, and the needed compute resources to process the data at mostly line rates.

2.3 Data Center Ingest Rates

One of the key factors in defining the networking, storage, and compute capabilities needed by the ngVLA project is the rate of data generation for the project. This value will vary with the number of antennae that are part of the array and for the application-mix that make up the observing pattern for the array, as different science experiments may create very different data set sizes. An estimate of the data rates during the construction phase of the project is given in Table 2. These estimates assume that the extreme high frequency resolution use cases are not part of Early Science; there is a linear progression of data rates from Early Science to Final Science; and commissioning activities assume 10% of Early Science.

Table 2: Summary of estimated data rates during project construction using EOP profile. Expected 28 days of peak data production each year.

Data Ingested by Workflow				Data Produced by Workflow				
Year	Avg Ingest (Gb/s)	Peak Ingest (Gb/s)	Total est. data ingest (PB/year)	# output Data Cubes/year	Avg size of Data Cube (TB) /year	Max size of Data Cube (TB) 28 days/year	Avg output data (PB/yr)	Total req. storage (PB/year)
2030	0.0227	1.57	0.085	219	1	2.4	0.2	0.285
2032	2.63	77.18	9.8	657	10.4	23.8	6.67	16.47
2036	80.89	1,513.86	304	1,533	10.4	203.7	15.6	319.6
2039	168.8	2,588.80	635	2,190	10.4	383.5	21.73	656.73

Our estimations assume that peak data rate observations are spread out across the year so as to not overwhelm the disk caches, and compute capacity, and that the system will be scaled to support the average data rate plus some overhead for node failures, networking issues, or storage system issues. We also note that the observations scheduling system may need to consider the network and compute infrastructure as part of its observing optimization algorithm.

More generally, ngVLA mentions plans to produce approximately 40PB of data per month when it reaches full production, which refers to the data streaming from the array for the Reference Observing Program (ROP). The ROP consists of a set of 27 observations that will be necessary to achieve the ngVLA key science goals. The fuller set of experiments, the Envelope Observing Program (EOP), consists of 63 observations, and defines what a typical year of observing is estimated to represent. It is expected that the EOP will produce 54.7 PB of data each month. One of the Key Science Goals included as part of the ROP is a spectral scan that uses high resolution and high bandwidth, and is a survey for looking for specific spectral lines in regions, such as looking for chemical signatures in star forming regions. Because this particular application produces so much data, the project will limit the number of proposals/observations of this type that are accepted, and they will be scheduled so that the buffers are not overwhelmed. Estimations in this document are based on the EOP [Hiriart24].

Because of the high variance for the data rates, and the fact that the peak applications will run for 28 days each year, the network will have to be designed to handle the peak rate (or close to

it). However, since the peak compute times will be scheduled next to less intensive observations, we will have some space in designing the compute side of the infrastructure.

2.4 Workflow Production Rates

After being run through the computational workflow, the data output rates are dependent on many factors that have not yet been fully determined. They depend on the image resolution,, the science experiment mix (high spectral resolution cases generate much larger data cubes than other experiments), and the number of antennae online and collecting data for any particular use case. Table 2 gives estimates of these values for the purpose of this document, but we know these will be adapted as additional instrument planning takes place.

It should also be noted that these datasets produced by the workflow may or may not be downloaded externally from the DPC. Outside the scope of this document, the project is determining the best ways to grant PIs access to the produced Data Cubes. This may be via something like a jupyter notebook, which allows the PI to run additional analysis and visualizations, or it may involve sending full Data Cubes to the PIs institutions.

2.5 Compute Requirements

Final compute requirements are unclear, but our estimate is based on the best 2025 modeling of the workload from existing VLA pipelines, extrapolated to data volumes of ngVLA, at a usable effective performance of 40PF at 64 bit double precision steady state capacity, with additional capacity for re-processing, estimated at 20% additional, and for external user data analysis.

Early experiences with the software stack imply this could improve somewhat in the Preliminary Design phase. We base this on:

- Early Benchmarking results that imply room for further optimization
- Runtime data that implies GPUs could be loaded more heavily with multiple copies.
- Ongoing research in mixed-precision and ML-based approaches to improve the computation.

Predicting the compute needs for ngVLA a decade before full operations is complicated by the fast shifting landscape of computing and algorithms.

Benchmarking with the 2025 gridding application (RoadRunner) on simulated ngVLA data on both an H100 GPU-based system and on a pure CPU-based system has shown that the GPU code executes approximately 2.5 times faster than the pure CPU code. This has implications for architecture selection. For instance, driven by AI demand, a quad-GPU server costs roughly 8 times more than the price of a top-of-the-market CPU server with 384 cores. Based on profiling and scaling from older V100 GPUs, early indications lead us to strongly suspect that this code is memory bandwidth limited, which means GPUs might provide more computation than their memory buses can effectively sustain for this application.

A complementary research project is utilizing machine learning to reduce the computational demands associated with processing interferometric radio observations from ALMA data. This is

taking place at the NSF-Simons AI Institute for Cosmic Origins by a team of researchers from the University of Texas, TACC, NRAO, and the University of Utah. Further, as we approach the end of Moore's law and with the market being primarily driven by AI, which focuses on lower precision computations, computing in a decade will certainly look as different from today as today seemed a decade ago. We describe how we see throughput computing being done in the future, combined with recommendations for where focus should be made on following and reaping benefits from the changes to come.

Based on this ongoing work, we are carrying multiple architectures forward at Conceptual Design, pending further refinement of the software and requirements.

Our 20PF baseline includes the following assumptions:

- The "RoadRunner" code captures the dominant compute code in terms of total execution time (more than 60% of wall clock time)
- Other pieces of code (other than RoadRunner) are as of August 2025 are CPU based, with no existing GPU implementations.
- Re-processing of data is an additional 20% of compute load.

In addition to the processing of the data, the data products are large enough that users would likely opt to do analysis at the ngVLA datacenter, rather than transfer vast amounts of data to their home institution for analysis. Based on discussions with the science team, we believe users will likely use interactive systems co-located with the data for their research. We are adding an additional 20% to the total estimate of computing capability to support this functionality.

2.6 Out of Scope Items

This document does not address:

- The location of the Science Operations Center, which may or may not be co-located with a Data Center, even if a non-Cloud solution is preferred.
- Support for the workflow needed for PIs to understand where in the process their proposal/data is.
- How the scheduling of additional processing will be handled. This document only addresses the initial data intake and calculation of the standard data outputs.
- What type of portal or user access will be needed for the Data Cubes, during and after the embargo period for data.
- How additional science outcomes will be computed beyond the workflow examined by the staff leading up to this report.

2.7 Technical Requirements Summary

Table 3: Data bandwidth requirement set to accommodate peak bandwidth given in Table 2. Data storage requirement adjusted from Table 2 to be cumulative estimates.

Year	Ingest Data Bandwidth Req't	Cumulative Data Archive Req't (Tape) (PB)	Cumulative Online Storage Req't (Disk/SSD) (PB)	Computing Req't (Sustained PF)

2030	23.15 Mb/s	0.3	0.2	5
2032	2.61 Gb/s	27	11	10
2036	80.86 Gb/s	517	58	30
2039	168.91 Gb/s	2,143	118	60

3. Hardware Solutions Analysis

This section looks at potential design options for hardware systems to meet the technical requirements of ngVLA operations. It details trade-offs to consider when designing the networks, storage, and compute, as well as projected staffing. This analysis, with the exception of certain networking sections, is *location independent*, i.e. the costs and capabilities would be the same if the data center were supported fully by ngVLA staff or leased through a co-location facility. Later sections will consider costs and trade-offs based on how this hardware is procured and housed, and compare to a purchased-services model. Note all forecasting was done in the August 2025 timeframe.

3.1 Technology Forecasting

It is critical to note that the details of the hardware solution, from technology choice to, crucially, costs, are dependent on forecasting the far future. Where possible, we detail the August 2025 costs and assumptions used to determine projected future costs, but, historically, a 2-3 year shift in timeline can create a 2x change in quantities and costs. This document takes a conservative and technically informed approach to cost forecasting, but it is critical to note this is a particularly unsettled time in forecasting technology trends.

After nearly 50 years of “Moore’s Law” improvements in transistor density, and the perceived accompanying benefits of 50% per year in cost, power, and capacity, the technology curves have changed. Clock rate improvements ended in 2008. While transistor density continues to improve in each generation of silicon, small feature size has meant that leakage current offsets the savings of shrinking transistors – meaning that new processor generations have twice the transistors, but at twice the power and twice the cost. The rise of solid-state storage displacing traditional spinning hard drives, has put storage on the same curve with processors. Even the traditional definition of Moore’s Law (transistors per area) is under stress – while the demise of Moore’s Law has been forecast for decades, the 2025 silicon processes build features just 10 silicon atoms across. Short of a fundamental transition in our understanding of semiconductor physics, these trends will slow and eventually end.

Further, the pandemic supply shock, followed by the AI demand bubble, have deeply distorted the technology marketplace over the last four years. In 2023 and 2024, the cost of digital storage (per bit) **increased** for the first time since the 1940s. The price of GPUs has increased by a factor of 8 over the last four generations (roughly 5-6 years, strongly influenced by a post ChatGPT environment). Memory and other components have seen similar trends, all the way out to the scale of data center transformers and air handlers. Cloud costs have seen a corresponding increase, with even basic CPU time increasing by more than 50%. While the 2025 incredible demand is very unlikely to continue through the planned construction of the ngVLA, the reality is that pricing is incredibly volatile, which complicates forecasting.

Geopolitical considerations also complicate forecasting – virtually none of the semiconductor supply chain is domestic. Though domestic sources can be found for some conventional processors, all server memory, most solid state storage, and all datacenter-class GPUs are

manufactured outside the United States, as of August 2025, and therefore subject to fluctuations in tariffs and higher costs due to potential domestic relocation of the supply chain.

The assumptions used here assume a baseline of moderate performance improvements to technology over time. The “AI bump” is figured into baseline prices, but our forecast in Agust 2025 is that the rapid rate of increase is expected to fall closer to historical trends as manufacturing capacity balances demands.

Forecasts are also based on extrapolation of current technology trends, within foreseeable physical limits. There are no assumptions that, for instance, Quantum Computing will disrupt the planning for data processing. While the project is in the timeframe where this is *possible*, it is not necessarily *likely* that general purpose quantum devices will be available. If they were to be available, there is a high uncertainty as to what their cost performance characteristics might be, or the related cost of re-engineering the software pipelines for the expected radically different programming models they will require. Nor are we assuming an AI solution to re-formulate the software for hypothetical quantum computers. While it is important to keep abreast of developments in these areas, it is not responsible to do cost and risk planning on the assumption of their availability.

3.2 Network Planning

The design of the ngVLA network sub-system is primarily dictated by the bandwidth requirements of the instrument and the expected peak rate of data production, as described above. There are multiple approaches available that can meet the project's needs, but each comes with differing levels of cost and staffing requirements.

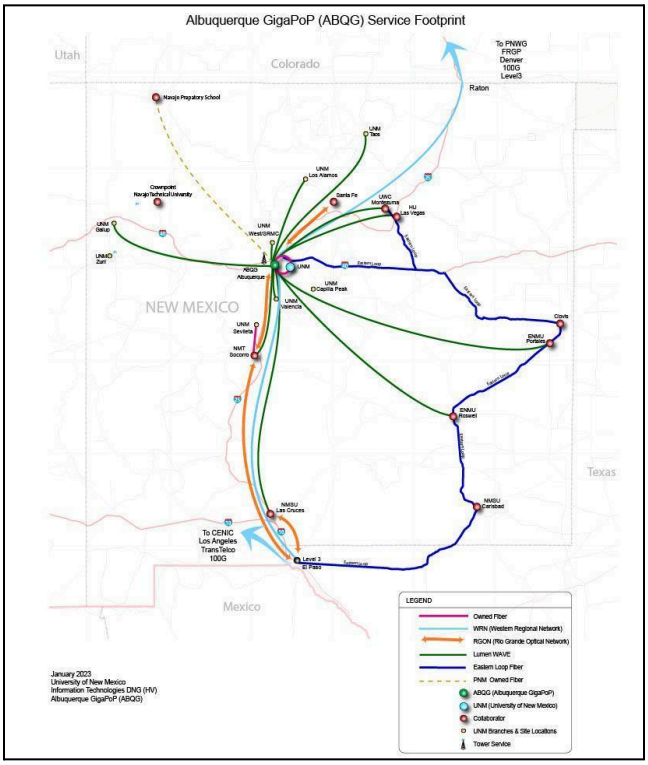


Figure 2: Albuquerque GigaPop Networking footprint as of August 2025, [ABQG].

In general, there will need to be reliable, redundant, high speed connections from the CBE to the Data Processing Center (DPC). This connection must be able to reliably **meet the expected peak data rate**, as it may occur up to 28 (non-consecutive) days per year depending on the science applications using the instrument. In addition, the DPC should be strongly connected to the collaborating sites that will be retrieving data, either data cubes or raw visibilities.

A networking solution needs to be divided into two components for the project. The most difficult part of designing a network will be the connectivity from Socorro to a metro area such as Albuquerque or El Paso (ideally both). See Figure 2 for a 2025 network map of regional connectivity. Once in a metro area, there are many commercially available options available, and the project can simply use peering and standard R&E networks to carry traffic to the DPC and secondary archive sites. In general, given the needed data rates stated in Tables 2 and 3, there will be several options and no special hardware or equipment will be needed for the majority of the network path.

3.2.1 Network from Socorro to a Metro Area

For the Socorro-Metro portion of the network, there are several options to be considered as a starting point in the network space. The first is a choice between **leased capacity (Layer 3)** or the purchase of **dark fiber or network IRUs (Layer 1)**. In general, leased capacity is negotiated with a commercial telco or regional research and education (R&E) network such as Internet2

[Internet2] or the Albuquerque GigaPoP [ABQG]. The advantage of this approach is that the project will need to acquire minimal additional equipment, since contracts generally include all equipment, maintenance needs, and service level agreements (SLA) for uptime and access.

Layer 3 leased capacity over a commercial vendor network may be less expensive or have a stronger SLA, but the network will not be very performant for the workload of the project. This type of network is generally tuned to achieve the best performance for very small flows, emails or interactive applications, such as traditional business applications. Capacity leased from a regional R&E network will be developed with research needs in mind and generally will perform better for large project flows. In addition, the associated contracting group is likely to be more responsive to R&E project needs and ensuring that the network settings needed for research flows are established and maintained

Layer 3 leased capacity is dedicated to the project, it is recommended that average network use stay below 30% to be able to adapt to overflow conditions [EB25, CW25] . In addition, without the use of dedicated links and bespoke protocols, even test data transfers rarely achieve 100% of the network use [FD24], so the project will need to plan accordingly. If the leased capacity is shared, strong SLAs and monitoring will ensure the project is receiving the needed bandwidth.

In August 2025, the project has sourced leased Layer 3 Services from Socorro to ABQ/El Paso, and have been quoted rates from Internet2 for 100G services. As of September 2025, the commercial companies listed in the area are limited (ATT, Cogent, TDS, maybe Sacred Wind). General pricing was quoted to be approximately $3,000-$4,000 monthly (plus one time set up costs) for 10G connectivity and $14,000-$17,000 monthly for 100G connectivity. In general, higher capacity bandwidth in rural areas is “quite variable based on existing contracts and purchases, saturation of the routes, and financial needs of the provider at that moment” [Quilt25].

The second general option for capacity is to purchase dark fiber or a longer-term lease of a wave on a fiber optic cable (Layer 1). This approach may offer additional avenues for upgrades as the project continues at a reduced cost, however the start-up can be challenging and expensive. In order to use Layer 1 services, additional investments must be made in terms of Layer 1 routers and equipment, which will be significantly more expensive than traditional Layer 3 and will need updating every 10 years. The main advantage of purchasing dark fiber specifically for a project is that there will be greater capacity for expanded use over time. With a dark fiber, only a fraction of the possible capacity would need to be lit when the project begins, and as the use and needs of the project grows, additional capacity can be made available for only incremental costs. How much capacity is achievable on a dark fiber will also be dependent on the age of the fiber.

However, with a Layer 1 approach, the project would be responsible for all monitoring and maintenance of a much larger equipment footprint, and at a very fine granularity. There are R&E teams that sell monitoring and maintenance as a service, for example, the IU Global NOC [GNOC] or the iNOC [iNOC], which should strongly be considered if the project decides to proceed with a Layer 1 approach. In addition, there are many vendor choices available for this equipment, so the team would be wise to take advantage of a consultant in the R&E space to

assist with the evaluation of choices, bidding, and final selection process if this approach is decided. Internet2 has offered services like this to other R&E projects in the past for a consulting fee.

In all cases, ideally, the end-to-end paths for the links would be fully redundant - meaning that the failure of any single router, switch, or cable would have a reduced impact on the overall data flow. History has shown that especially in isolated or rural environments, it can be challenging to get separate paths for network cables since, for example, multiple vendors may use the same network conduit alongside a highway.

Recommendation:

With the expected data rates, and the time frame of the project, **Layer 3 Services is likely the best option for the project if it is available.** We make the assumption that the local network connections continue to grow as they have over the recent few years. In addition, with a Layer 3 approach, we strongly recommend working with a regional R&E network whenever possible, as if the contract is with a commercial vendor the project will not be able to tune the network performance in such a way that will enable faster data transfers, for example, using jumbo frames or larger MTU buffers.

3.2.2 Network from Metro area to the DPC

Once the project has a network path to a data hub in either Albuquerque or El Paso (or both), the rest of the network path can be determined. Overall, the data demands outlined in Tables 2 and 3 are within reason for the R&E networks and backbones available to the project, so no special equipment or planning is likely needed. Most R1 institutions have 100 Gb/s networks to regional R&E networks, with state R&E backbones ranging from 400 Gb/s to 1.2 Tb/s. For example, TACC has 2x100G for external connectivity: one link through Dallas and another through Houston. At those points, TACC peers with a variety of network providers including Internet2, ESnet, LEARN, and commodity providers. The national Internet2 backbone has moved from 100Gb/s in 2012, to 400Gb/s in 2021, and has portions that are 800Gb/s, with the expectation of 1Tb/s in the next 2-3 years.

The project will need to establish peering with standard R&E networks from whichever exchange point is used in the metro area connection. The location of the DPC will somewhat influence both the cost and the reliability of the network capacity, simply in terms of the number of paths available. In general, since the flows will be very large and batch processed (not interactive), the latency of the connection will not play a strong role in the data transfer performance. However, there will be a need for redundant, high bandwidth connections.

Locating the DPC in a major metropolitan area, preferably an Internet2 hub or an area with a strong and active regional R&E network, would be advantageous. In all cases, the DPC will likely need to support some Layer 3 equipment, with planned refresh rates of 5-7 years to avoid end-of-life and end of support issues. Looking at the ngVLA data production rates, additional capacity will need to be acquired as the project grows. The community standard for shared resources is adding capacity when shared links reach 30% saturation.

Network costs will vary widely on the selection of the DPC, as some commercial sites include this as part of their quote and some don't. As a high level comparator, for an education institution to connect to Internet2 at 100G, the 2025 charge rate is roughly $225,000/yr and at 400Gb/s is roughly $400,000/year, just for the connections and peering.

As of August 2025, most R1 institutions support 100Gbps connectivity, and large commercial datacenters might have 8-12 of those connections. Many leading edge sites have or are upgrading to 400Gbps in 2025, which is more than sufficient to meet the project requirements through ~2034.
The shift from 100 Gbps to 400 Gbps as standard connectivity has taken roughly 10 years, not dissimilar from the move from 10-100Gbps in the previous decade. Switches and optics that support 800Gbps are in the marketplace, albeit at a premium, however prices are dropping every quarter as installations grow.

The full project operation in 2039 will require 2,500Gbps, however by that time frame a typical R1 will likely support at least 1,600Gbps. Any reasonable commercial or academic facility should have the bandwidth to handle what the project requires in the timeframe defined. At a worse case scenario, the project would need twice the standard connection.

To compare, the August 2025 TACC 100Gbps connection is ~$200k/year now (plus switching and optics at approximately $1M for 8 years). Networking costs will be trivial compared to the total cost of the project, even considering leasing bandwidth from Socorro or owning the fiber.

Please note that the costs considered here are only for the purchase/lease of capacity. The project will need to maintain some staffing for higher level monitoring and user support, for example, to ensure proper network routes are being used and that the flows are achieving the needed end-to-end performance, even when low level monitoring is generally the responsibility of the contracting party. This is the approach used by several other astronomy instruments, including SKA, VRO, and NoirLab [SA3CC25] in the 2025 timeframe.

Attention will still need to be paid to ensure good performance. For example, recent transfers between NRAO and TACC only achieved an average data transfer rate of 748Mb/s [NS1]. Of course, this may not be an apples-to-apples comparison due to the file sizes and such, but in any case show that data transfer speeds could likely be strongly improved with minimal additional oversight. In 2025, most sites can achieve 1.5 - 2 Gb/s transfer rates with minimal tuning and the use of basic software for data transfers, such as Globus [Globus] or hpn-ssh [Rapier25].

Data transfer performance would likely be able to be improved with regular check-ins on the paths being used, the buffer settings of the routes along the paths, and regular checks on data transfer protocols. This is the approach taken by many other astronomy projects in 2025 [SA3CC25]. For example, the Vera Rubin Observatory has contracted with several R&E organizations to provide capacity on existing leased network infrastructure. Primary among these is AMPATH [AMPATH], a project led by Julia Ibarra, Florida International University, that supports 100G networks between the US and South America [Ibarra25]. Early in their planning,

the VRO organized a separate networking support group, VRO-Net, and contracted not only with Ibarra but with the network support staff at the IU GlobalNOC for additional support, in coordination with R&E network staff in Chile and Brazil. As the data center at SLAC was identified as the main data receiver, additional team members were brought in from both the US Internet2 and ESnet, the US Department of Energy Network.

Similarly, some networks contract for additional networking and cybersecurity services for their networks. For example, NOIR lab is using the IU OmniSOC [OmniSOC] as a network Security Operations Center [RT25]. As with contracting for additional network support, the staffing needed to keep the network and data infrastructure secure is often out of scope for a traditional science team.

3.3 Data Storage and Archive Capacity Planning

The design of the ngVLA data storage systems is primarily dictated by the bandwidth requirements of the instrument, and secondarily by the needed storage capacity. Multiple storage technologies are available that can meet the needs of the project, but with widely varying costs.

Typical storage designs are hierarchical, which allows the use of a mixture of technologies to get desired performance and costs, as shown in Figure 3. For instance, while tape is the most affordable option for data storage, it is correspondingly slow. However, by introducing a tier of more expensive solid state storage, we can design a tape system that runs at the *average* data speed, with a tier of SSD drives that meets the *peak* speed, but is sized to only store data for the duration of the peak data burst.

Similarly, we can make tradeoffs about the reliability of data – permanently archived data needs to be kept very safe from data loss, but need not be as performant; a “working set” of data pulled from the safe archive can be faster yet less reliable, as data can always be restored from the archive.

Careful analysis of data rates, reliability requirements, and other assumptions are necessary for an optimum storage assumption. It is also important to consider the full usage model of data – a storage system that can keep up with the instrument to write the data, but leaves no bandwidth to read it for processing or further use, would not be a functional solution.

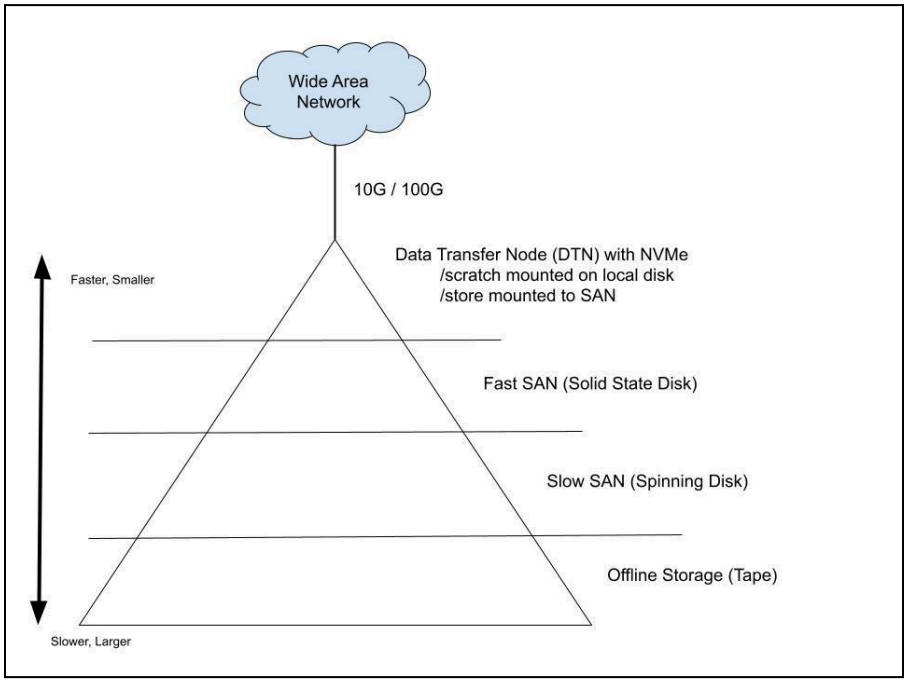


Figure 3: Hierarchical Storage Pyramid, compliments of Jason Zurawski, ESnet.

3.3.1 Magnetic Tape

Tape storage is too slow for any data that is in active use, but sensible for longer term and archival storage. While it might be easier to always have all data available on higher speed devices, the volume and expense is likely to be prohibitive for the project.

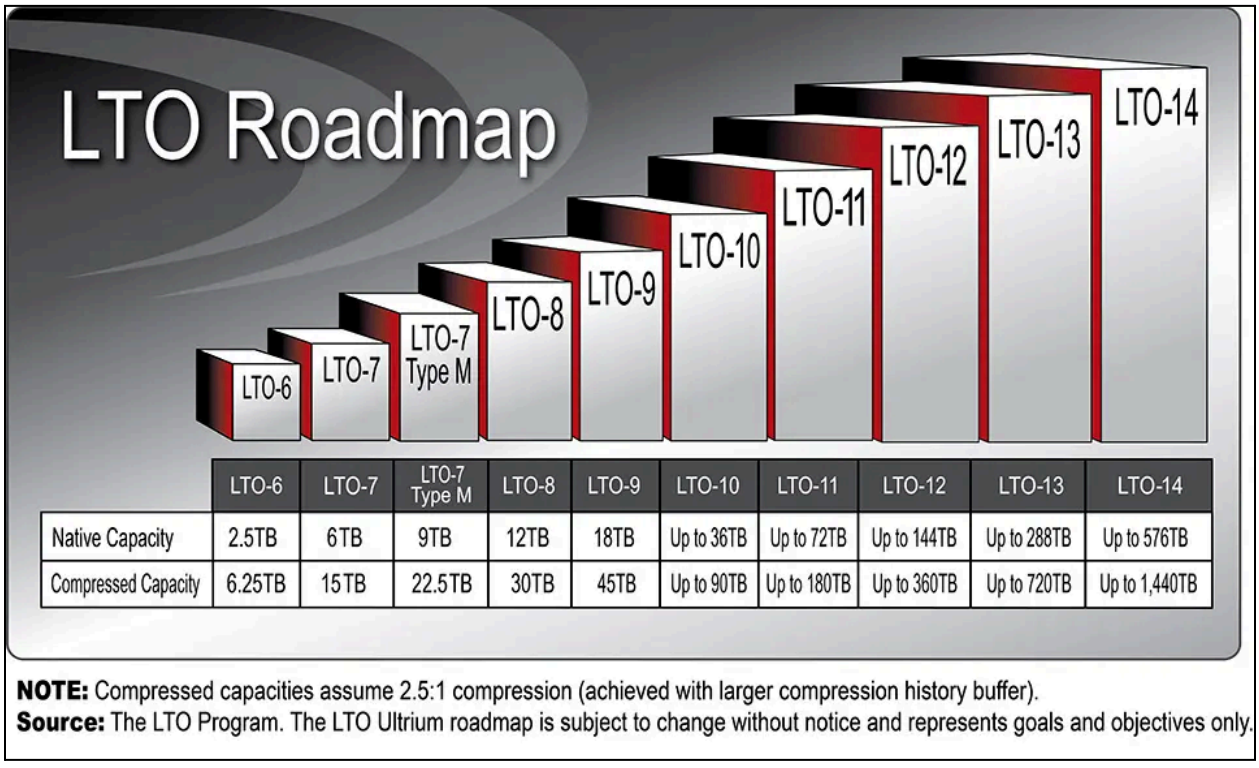
A typical tape archive system has several components; the tapes themselves (the physical media), the tape drives that read and write them, the tape library (a facility to store large numbers of tapes with robot arms to fetch the appropriate tape to the appropriate drive), and a front end disk cache system that users write directly on before the system sequences the files to tape. Typically, some tape management filesystem runs the cache and the libraries.

An advantage of tape is that once an archive is purchased, media (tapes) can be added on demand, allowing the cost to be spread over several years. Further, as with most technology, the price of tapes falls as technology gets older. New tapes in cutting edge technology typically cost $250, but when the next technology is released, historically they fall to ~$85. Tape drives in the 2025 market follow the Linear Tape Open (LTO) standard (Figure 4) [LTO], and capacity has improved twelvefold in the past 15 years, the data rate of tape has improved by a factor of 2.67 [IBM22].

The LTO standard assumes compression yields a factor of 2.5 over the raw media capacity. Our measurements using 2025 VLA data show compression in the range of 1.5. LTO-10 is on the verge of being released today, with 30TB tapes, meaning about 45TB of project data can be written per tape. We expect progress on the roadmap to continue, with new generations roughly

every 2.5 years, through at least the LTO-14 generation. There is clear technical evidence the roadmap will reach that far. This means we would expect to see 90TB per tape (compressed) by the start of construction, 180TB per tape by 2032 when the project begins to collect data, and approaching 1 Petabyte per tape by the time the full data rate is hit ~2039.

Performance improvements will be less aggressive since there is a limit to the rate tape can be pulled past the head without breaking the media. While capacity can improve with longer tapes, performance only improves with more density per linear length of tape. In August 2025, tape drives (LTO-9) achieve write rates (measured at TACC) of roughly 350MB/s. We expect to see only ~15% performance improvement per generation.



	LTO-6	LTO-7	LTO-7 Type M	LTO-8	LTO-9	LTO-10	LTO-11	LTO-12	LTO-13	LTO-14
Native Capacity	2.5TB	6TB	9TB	12TB	18TB	Up to 36TB	Up to 72TB	Up to 144TB	Up to 288TB	Up to 576TB
Compressed Capacity	6.25TB	15TB	22.5TB	30TB	45TB	Up to 90TB	Up to 180TB	Up to 360TB	Up to 720TB	Up to 1,440TB

NOTE: Compressed capacities assume 2.5:1 compression (achieved with larger compression history buffer).
Source: The LTO Program. The LTO Ultrium roadmap is subject to change without notice and represents goals and objectives only.

Figure 4: From https://spectralogic.com/spectra-stack-tape-library-capacity/. Note LTO-6 came out in December 2012 with a data rate of 160 MB/s.

It is worth noting that although tapes *may* last for longer periods, data can only be retrieved with high reliability from a tape for 7-8 years, so a replacement cycle – at least for the media – needs to be taken into account. Given the large volume of data that the project will put on tape, a ~1 year migration time is not an unreasonable requirement.

Another consideration is that data written to tape will also be needed to be retrieved. A single drive can only be writing *or* reading data, not both. Because of this, the additional needs of

retrieval, e.g. for reprocessing of deep storage data, needs to be incorporated into the system sizing.

Tape has several other advantages over other storage options besides raw cost per byte that should be factored into any decision:

- Tapes consume zero power when idle.
- Tape systems can grow in capacity simply by adding additional tapes or drives.
- Tapes can be ejected – it is easy to make copies (offline) and remove them to store elsewhere for extra data integrity.

In August 2025, a tape library capable of holding ~4,000 tapes with 20 state of the art tape drives cost roughly $1.5M, of which roughly $1M is for hardware, with the rest consisting of highly variable licensing costs, depending on software and vendor options.

Given the pricing for media above, we can extrapolate the cost of filling that library with tapes to be roughly $1M additional (that need not be acquired all at once), for a $2.5M total for the tape portion of the storage system, and rough capacity of 180PB assuming a 1.5 compression ratio for ngVLA data. This gives a rough hardware storage cost of $14k/Petabyte, with an expected halving (before accounting for inflation) every ~2.5 years for roughly the next decade.

3.3.3 Solid State Drives

The 2025 state of the art for digital storage is Solid State Drives (SSD), also referred to as non-volatile memory express (NVMe) storage, which utilizes NAND Flash technology. While the most expensive option, it has been the technology that is most quickly improving performance, capacity, and price, with the exception of a demand driven bump around AI. While also subject to the limitations of silicon technology, the use of 3D NAND has allowed vendors to stack dies to increase density, a trend that is expected to continue.

Traditionally, SSDs have been much smaller than traditional hard drives, but between 2023 and 2025, this trend has reversed. While hard drives are approaching 30TB per drive, SSD has moved from 15 to 30 to 60 to 120TB per drive, with 240TB drives expected in mid-2026. It is not unreasonable to expect individual SSDs to pass 1 Petabyte in the early 2030s, though the curve may slow substantially after that per industry roadmaps.

In mid-2025, at a near-wholesale cost on a recent quote, a server built with 60TB drives yields 1.35PB of raw storage, and costs around $170k. A server with twice as many 120TB drives costs roughly $380k, but will provide 5.4PB of capacity; 4 times the capacity for twice the price, or half the cost per byte stored.

While aggressive, this $70k/PB price is a reasonable starting point for estimations (roughly 5x the per byte cost of tape technology).

<u>However, it should be noted that at the end of 2025 into 2026, the cost of SSD drive costs increased by at least 500%,and the volatility of the AI market leaves all pricing quotes</u>

impossible to estimate with certainty even weeks out. For example, in March 2026 for another project we were given an estimate of over $100M for a 200PB SSD drive solution.

In addition to cost, it is important to note the relative performance advantages of SSD storage – individual servers can drive multiple GB/s of bandwidth, and can be aggregated in large arrays to provide TB/s of bandwidth. Unlike tape, all addresses on an SSD drive can be retrieved in equal time, giving them orders of magnitude advantages on small data reads and writes. Any address on a SSD drive can be retrieved on the order of milliseconds; a particular byte on the end of a tape in a library slot can take minutes to hours to retrieve.

SSD drives tend to be the top of any hierarchical storage system today; indeed, different SSD technologies may be integrated across multiple tiers to hide the latency of older storage technologies.

3.3.3 Hard Drives

Traditional hard drives, or spinning disk, has been the most common storage technology, but in August 2025, appeared to have been supplanted by SSD as the dominant datacenter technology. However, since the original version of this document was written, we have witnessed unprecedented volatility in the storage market as a result of the AI boom. As of April 2026, over the past 18 months, SSD storage costs have increased by at least 500%, and spinning disk by 60%, with year+ delays for delivery. Please note all cost and size estimates are based on August 2025 pricing predictions.

While spinning disk continues to improve, the pace of progress has become incremental. In the early 2000s, hard drives improved in density from one to two to four terabytes. In the last decade, the generations have moved from twenty to twenty-two to twenty-six TB. The price has remained roughly constant, meaning the improvements in cost, once halving every two years, now improve at closer to 10% every two years. Increasingly exotic technologies are being employed to keep them competitive, for example, 30TB hard drives use Heat-Assisted Magnetic Recording (HAMR) technology, using a laser to pre-heat the spot on the drive surface where the head will next write to improve performance, as traditional Perpendicular Magnetic Recording (PMR) has stopped delivering improvements.

Unlike SSD, spinning drives are also mechanical – they rely on spinning a platter and physically moving a drive head across the platter; mechanical processes will not keep pace with solid state ones. This also means the device is inherently sequential. To get to a new random address, the drive must spin to the right spot, and the head must move to it - collectively known as the *seek time*. As workloads, particularly the small random-access workloads typical of AI, prioritize the speed of individual operations over the total throughput, hard drives will fall further behind. The moving parts of a hard drive also make it much more prone to mechanical failure than solid state devices, and the sequential nature of drives make the time to rebuild a failed drive increase with each new generation – a full rebuild on a crashed hard drive takes more than 30 hours to complete in August 2025. Market trends already show volumes declining as this technology is squeezed between SSD at the high end and tape at the low end.

3.3.4 ngVLA Storage System Design Recommendation

A core component of the DPC will be a tiered, hierarchical storage component that will both archive the long-term data as well as store the data for use when computing the Data Cubes. In general, a Data Storage pyramid, such as the one depicted in Figure 3 is recommended. The project will need to take special care in evaluating the tradeoffs between the different storage mediums.

For the needs of ngVLA, as described in Section 2, we recommend an archive system with two geographically replicated tape libraries, to be minimally populated with drives and tapes in 2032, with a full replacement in 2039 to handle full rate data throughout operations. A cache should be put in place using SSD drives to handle the bursts of data and smooth traffic to tape.

A second, primary storage system should be put in place for the online data access, and scratch space for the processing cluster, completely in solid state technology. Based on the requirements analysis, we estimate this as being the size of the 2025 set of data cubes being shared, with additional space for ingest, initial processing, and re-processing equal to about 2 months data at the 2025 data rate.

The design of the storage system presented here assumes:

- Data rates and cumulative totals as shown in Section 2
- “Bursts” of the Max data rate do not exceed 24 hours.
- The LTO tape roadmap holds, with roughly 2.5 years between generations, through LTO-14, at roughly constant cost per tape, adjusted for inflation.
- SSDs continue to improve to at least 1.5PB per drive, under the current cost conditions, adjusted for inflation.
- No disruptive technology impacts the storage market.

The conceptual design proposes an initial archive system to be deployed in time for start of data acquisition (est. 2032) with ability for an upgrade in 2036 to handle the increasing data volumes and rates, and a replacement in 2039 to meet the full data volume of the completed instrument.

From a design point of view, and based in part of August 2025 pricing, what is recommended for the storage system includes:

2032 initial deployment:

- Online storage system of 25PB upgradeable to 60PB:
  - 20 servers of 24 SSD per server (480 total drives)
  - Each server should provide ~4GB/s bandwidth, aggregate 80GB/s
- Archive system capable of 50PB upgradeable to 500PB
  - Estimated archive BW: 8GB/s
  - Libraries + Drives
  - Front End cache with 30 GB/s bandwidth
  - Media + Licenses (500 tapes)

2036 Additional media deployment:

- Add 35PB of Online storage - Drives only, keep servers
- Add 450PB of Archive storage - Would need drive upgrades and tape Media (2500)

2039 Replacement/Upgrade
- Online storage system of 120PB
    - Online BW ~200GB/s
- Archive system capable of 2.5EB
    - Tape BW: 18GB/s
    - Libraries + Drives
    - Front End cache with 50GB/s bandwidth
    - Media + Licenses (2,000 tapes)

This design assumes a single site, single storage system. While the project requires a second archive, we recommend the consideration of having a **secondary archive at a partner site** as a backup site for both the data cubes and a second tape archive to mitigate loss of data due to a catastrophe at one site. Even an international site would be able to maintain data ingest at line rates for the project. Of note, this is how several other large scale astronomy projects are maintaining additional data storage, such as VRO and SKA [SA3CC25]. The second site need not also match the architecture of the primary, e.g. one could be cloud hosted.

3.4 Compute Systems

The requirements for the computing system are a delivered performance of 60PF sustained [Hiriart24] at the full data rate at the end of construction, with a near-linear ramp-up of compute capability from the start of construction ~2030.

The fundamental architectural decisions are the correct mix of CPU vs. GPU architectures, though in service models this can be tweaked dynamically. Given the uncertainty around code optimization at this stage, **we propose two design alternatives, fully costed, to set a rough budget for a more late-binding hardware decision**. The first option is a **pure CPU design**. The second option is a **pure GPU design**. Note that hybrid options mixing the node types remain a possibility – any hybrid will fit between the pricing bounds provided by option 1 or 2. We recommend carrying the higher cost of the two forward at Conceptual design for further refinement. Additional design considerations are the timing and number of hardware acquisitions - it rarely makes sense to acquire capability substantially ahead of the demand, given the comparatively short life cycle of hardware.

We propose a 3-stage hardware acquisition, with a 10PF Compute capability delivered by 2030, 30PF in 2035, and the full 60PF by 2040. This puts compute hardware in place aligned with demand, while taking advantage of likely performance improvements over time.

Note this analysis assumes a hardware lifespan of 5-6 years (less in the last upgrade). It is likely that useful value could be extracted from the hardware beyond this lifespan. However, service costs tend to increase greatly beyond 5 years, indeed, our projected pricing builds in a 5-year hardware warranty. This means that there would still be useful hardware, but it may not be

reliable enough for mission critical applications. However, the project could likely assume another 3 years of lifetime from the hardware without warranty, projecting a 5% failure rate each year, with corresponding reduction in capacity, perhaps retaining some contingency funds for emergency purchases.

This also means the hardware procured would cover the ramp-up period of production antennas, 2030-2039, have full warranty through 2044, and likely be able to handle 80-85% of production demands through 2047.

3.4.1 Pure CPU-Based System

For a CPU based system in 2030, if we assume:

- 80% of peak FLOPS advertised is “usable” FLOPS
- Assume a 2030 CPU node provides 9.46TF
  - Peak Flops in a 2025 AMD Milan 128 core: 7TF/node (max boost)
  - 7*0.8=5.6TF
  - Assume two generations of ~30% peak perf improvement per generation.
- Compute target of 10PF could be met with ~1100 nodes.

Extrapolating these trends to 2035, and increasing our target performance to 30PF:

- Node performance should rise (over 3 generations) to 20.8TF/node
- Power will increase to 4KW per node
- Node cost will increase ~20%.

This yields a roughly 1,450 node cluster to reach 30PF, using 5.6MW of power.

By 2039, when 60PF is needed, and another two generations of processors become available:

- Compute node performance will be 35PF
- Power will reach 6.7KW/node
- Cost will increase an additional 20%

3.4.2 Pure GPU-Based System

GPU applications tend to get a lower percentage of peak, so we will use 60% of peak Flops as our effective Flops for meeting the compute needs. A 2026 Blackwell GPU will deliver 40TF of double precision peak, for an effective 24TF per GPU. A Blackwell board contains 2 GPUs in 2025/2026, which will increase to 4 with Rubin in 2027 (with flat Double Precision (DP) performance).

We will assume:

- 96TF usable/node.
- Roughly 100 nodes will be needed.
- Estimate per node power at 6KW.

The 2035 upgrade requires 30PF. Here, we are disadvantaged by the fact that GPU roadmaps project no more 64-bit floating capability into the future, forcing us to use flat performance, but

continued inflation-adjusted pricing. We do not believe this is the *most likely* scenario, but is the most conservative for a long range forecast.

This scenario could improve in reality due to:

- Widespread incorporation of emulated 64 bit calculations using lower precision hardware (likely).
- Changes in roadmaps providing additional FP 64 performance (relatively unlikely).
- Improvements to the code exploiting mixed or lower precision algorithms for ngVLA software (somewhat likely).

Nevertheless, we will use our conservative approach, assuming GPU FP64 performance remains relatively fixed past the Rubin generation. We will also fix the power and adjust the price only for inflation – while it is likely “top end” datacenter GPUs will continue to increase in power, if there is no floating point improvement at 64 bit, there would be no point in investing in these top end GPUs any longer, and likely a lower end item would be available maintaining the max DP performance at the power of previous generations.

Given these constraints, the 2035 assumptions would be:

- 96TF usable/node
- Roughly 300 nodes will be needed.
- Estimate per node power at 6KW.

In this scenario, the GPU is more cost effective in the early users, with CPU-based systems more effective in later years. It is unlikely the market would sustain this dynamic, unless low-precision AI dominates all of computing (certainly that is the 2025 trend). **We do think the potential of emulation of FP64 would likely reduce the cost of this system by a factor of 3 in the later years**.

3.4.3 Interconnect

The cluster-level networking interconnect architecture is not specified at this time. The workload, as defined in August 2025, involves processing large batches of single node jobs, and the interconnect is only used for ingest and I/O, that is, there is no in-job message passing to stress the interconnect network. In our pricing models, we assumed a typical Infiniband interconnect cost was added in to the base node price, on the assumption that given the relatively modest cluster size, it would be appropriate to use an INfiniband or Ultra-Ethernet cluster with a fat-tree topology, which would not change the base price. By the latter stages of the project, commodity speeds will be well into the Terabits for internode connections, which should well exceed the projected needs, so no “exotic” network solutions should be required.

3.5 Recommendation

Given that the GPU roadmap is likely to have upside in FP64 not captured here, we recommend using the CPU cost (which is higher in the first purchase year, but lower in the out years) as a safe basis for estimate. If GPU FP64 performance does not improve, and the ngVLA

applications cannot exploit lower precision, then selecting GPUs would be a poor choice for the project, and CPU capability would be procured instead.

3.6 Additional Considerations

The costs above are complete for hardware *acquisition*, however, operations cost must also be considered. The design above gives additional requirements that will inform the hosting decision and operations costs.

3.6.1 Power and Cooling Loads

Most of the power and cooling load will come from the compute racks, so we will deal with them first. Per-node power has been skyrocketing in recent years, even for CPU systems, so while these systems will be physically compact, the power requirement is significant.

For the CPU systems, our estimate is that a single node will take roughly 1,450W. The server load will vastly outstrip network and switching power, so we will use this as the basis for our power computations. We are assuming steady growth in this over time, to 4KW/node in 2035, and 6.7KW per node in 2040. Due to the flattening of projected GPU FP64 computation, we are projecting a flatter curve for GPUs, but dramatically more per node. This is summarized in Table 4.

Table 4: Power requirement estimates.

Year	Compute CPU (Megawatts)	Compute GPU (Megawatts)	Primary Storage (Megawatts)	Archive Storage (Megawatts)
2030	1.6	0.6	0.04	0.07
2035	5.6	1.8	0.05	0.09
2039	11.4	3.6	0.1	0.2

Note that unlike costs, power estimates are not cumulative. The stated power would run all required hardware in a given year (unless, as earlier discussed, there is a desire to run the compute hardware beyond its normal lifetime). However, facilities calculations should consider growth to the max power range by the end of the next decade. Note: the GPU option in this scenario gives dramatically better power usage, though at a large additional capital cost.

The cooling loads may present more of a challenge. It is likely that all of the compute components will be liquid cooled throughout the project lifecycle. The actual configuration of cooling will be dependent on how dense the deployment is. However, in any scenario, sufficient chilling to remove the heat load generated by the max projected power load (~12MW) will be required. For reference, in today's direct-liquid cooling technology, removing a heat load of 100KW with warm supply water requires a flow rate of 1,200 Liters/Minute with a 1 bar pressure drop.

3.6.2 Floor Space

Floor space requirements can vary with the density of the solution – a data center not capable of high density racks will simply use more of them at lower density. Power capacity is likely to dominate any discussion of facilities. However, given a modern liquid cooled datacenter, 72 nodes per rack is a reasonable expectation for compute density.

Using that metric, the compute space would require 16 racks for the initial deployment, growing to 25 racks by 2039. The primary storage systems would require 2 racks, and the front end for the tape library would require 2 additional racks. Assume 2 additional racks for switching and management, and a maximum of 31 traditional datacenter cabinets would be required.

The tape libraries are physically much larger; a modern library is roughly 30'x8' in diameter, with sufficient aisle clearance required on all 4 sides for service. The design of the system would require two libraries. Note if a failover option is selected, twice the libraries would be required, but they would also need to be in two physically distinct locations.

3.6.3 Staffing Requirements

This analysis considers staffing requirements to keep the hardware and system software functional – **it does not take into account the need for software development, external science user support, or other functions that would be independent of the systems solution**. We also do not consider the need for depth of staff to cover vacations, on-call limits, or other similar issues that might vary based on how the solution is implemented (within a larger organization, or completely stand alone).

The general categories of staffing required are: Systems Administration (2 FTE), Storage Administration (1), Network Administration (0.5), Cybersecurity Operations (0.5), and 24x7 Support Staff (5). It is not uncommon for one person to take on more than one role in smaller IT organizations. However, given the scale of the system, our estimates to keep a system functioning across these categories are substantially larger. Note that the 24x7 support personnel tend to be substantially lower cost than the above categories, as the primary function is to simply monitor the systems for anything that would require locating the on-call person in any of the above categories.

The **total staffing required would be roughly 9 FTE,** assuming a single person can be found that can blend the cybersecurity and network skills, or other similar overlaps. Note in that scenario, three of the roles would not have backup when the primary person was on leave, except through cross-training of the administrators. In a dedicated shop, it would also not be uncommon to remove one shift from 24x7 operations to be covered during the 40 hours of week one can expect the remaining administration team to be on-site.

Also note that this staffing does not include personnel for software development, science user support, or other functions independent of the systems solution, only those staffing required to run the system.

4. Colocation/Hosting Model Options

If hardware is to be procured, it must also be housed and managed. There are again, multiple options in how this can be achieved:

- A datacenter can be built or purchased
- A commercial colocation option can be provided
- Hardware can be sited at a partner institution

A detailed examination of construction costs is beyond the scope of this document. However, as a quick “back of the envelope estimate”, and assuming that any site would need a new substation given the lack of datacenter power availability in most locales in 2025, a reasonable estimate of the construction costs of a 12MW facility is likely $100M. Power and operations costs would likely be similar to the analysis below. In the event an existing facility is refitted, utility-related costs and power and cooling are likely 80% of the new build cost, with the building shell providing the other 20%. Also to be considered, most equipment used in datacenters, from Uninterruptible Power Supplies to power distribution units, have an expected production lifespan of 10-12 years – a second refit may be required to meet the full lifecycle of ngVLA.

The hardware analysis shows a peak datacenter requirement of 12MW for power. This is direct IT power; depending upon the availability of water, climate, and options for cooling, it is safe to assume 10-20% additional total building power to run chilling equipment and remove excess heat. Most datacenters also charge a space fee, but given the rise in power density, energy and cooling costs tend to be the dominant factors.

Prior to the analysis, it is also worth noting unsettling trends in power costs. Wholesale costs have increased in many parts in the US by 15% in the last 3-4 months (as of the August 2025 drafting of this document). The push of AI-datacenter demand, coupled with recent cancellations of various renewable energy investments and an aging power grid, have led to many forecasts of dramatic increases in electrical costs going forward. Whether these will be “one-time” or ongoing increases will be subject to the fate of the AI boom, future public policy with regard to energy, etc. Power costs are an area of high uncertainty at this time. The ongoing datacenter construction underway today in the US will add 75GW of baseload to the grid, roughly the equivalent of the peak power demand of Texas. While it is possible a fossil fuel boom could offset some of this increase, it will take years of consistent policy for that to happen. Natural gas turbines in 2025 have a 4+ year construction backlog, new gas plants are not coming soon, and changes in policy of each presidential administration will likely prevent long term capital allocation in any direction.

4.1 ngVLA Colocation Facility in ABQ (or elsewhere)

One possibility is to rent space from a colocation facility. This could be in the nearest metro to the ngVLA instrument, or further afield.

Table 5 lists publicly available information for several of the data centers local to Albuquerque. Based on our initial assessment of CPU and storage needs, none has sufficient power or space

to support the project in the long term, though several could host the initial phases, and perhaps could be persuaded to upgrade over the next decade.

Table 5: Data centers local to Albuquerque.

Data Center Company	Square Feet	Power
Center Square ABQ1	13K	8.6 MW
H5 Data Centers	225K	n/a
Oso Grande	10K	750 KW
bigbyte.cc	16K	3 MW
Connect Albuquerque	90k (planned)	

None of these facilities have public pricing as of August 2025; however, we can pull pricing from other markets from past analysis done by the NSF Leadership Class Computing Facility.

Datacenter costs have a variety of components; power, add-on charges to support cooling, Monthly Recurring Costs (MRC), and upfront installation/modification costs, as well as network bandwidth costs.

Using the LCCF analysis of public markets for existing colocation facilities (last updated in 2020), the projected cost for this option would be an initial install cost at $650/KW, and an annual run cost (all-inclusive) of $8,120/KW. This assumed the pre-2020 (pre-ChatGPT) market for low density racks – likely a 50% large scale discount could be achieved vs. this pricing.

The rough hardware model scales datacenter needs for ngVLA from 2MW initially to 12MW by 2039, with that is continued steady state.

Emergencies involving power and hardware failure could be dealt with by colo-site staff, which may reduce the need for 24x7 monitoring, the staffing needs would be largely unchanged. The hosting would cost substantially more than the hardware being hosted over a decade.

An additional model to consider is the hybrid model adopted by the LCCF, where the datacenter is being built in a shell building by the datacenter operator, with the project footing the bill. In addition to getting a space built to spec for high density power/cooling, the operating costs are substantially lower as the operator is no longer retiring debt from the initial build.

4.2 TACC Hosted Facility

An additional colocation option would be to buy the hardware, but situate it at the facility being built for the NSF Leadership Class Computing Facility at the Texas Advanced Computing Center in Austin, Texas. In this model, TACC/LCCF would need to pass through direct power/cooling/MRC costs from the datacenter, but this would be “at cost”. Existing TACC staff could be leveraged for the operations under a staff “buyout” model (at Austin rates, but leveraging deeper teams and availability of fractional FTEs).

In this model, there is no initial outlay for datacenter costs, but the annual run rate is similar to option (2) above, as the pass through costs for multiple MW of datacenter space would be the same. **Working with TACC would reduce the cost by roughly half through the decade of construction, and if another 5 years of operations are factored in, reduces costs further by approximately 70%.**

5. Service Model

An alternative to buying hardware is to acquire the same capability through the cloud as a service. There are several fundamental differences to consider. If you don't own hardware, you also do not need to cover hardware maintenance, power, cooling, and housing, which has substantial costs beyond the hardware itself. You also have no residual value that comes from owning hardware – with a physical computer, you can simply choose to keep running it past the end of the projected lifespan; with a cloud, you reserve capacity by the month or year, and the exact time your billing period expires, all of your capability is gone.

While many things about cloud services can be done generically, many other services require customizing to meet a particular vendor's interfaces and capabilities, so switching vendors can be difficult. Some staffing costs disappear in the cloud model, but many remain – while you no longer fix a broken server, the cloud operator does not ensure that any of your services are orchestrated, up, running properly, and only billing for useful work.

5.1 Commercial Service Forecasting

A commercial cloud facility refers to cloud computing services offered by third-party commercial providers, most commonly Amazon Web Services (AWS), Microsoft Azure, or Google Cloud Platform. These providers charge a set of fees for compute, storage, and storage access, but allow a project to delegate running computing resources, storage, and network to a commercial entity, rather than paying for their own infrastructure. In general, cloud facilities are thought to enable better scalability, flexibility, and cost savings over long term projects. Unfortunately, experience shows that these approaches can limit flexibility due to vendor lock-in [OST16] and cost much more in practice than originally estimated.

Cloud storage can be thought of as virtually unlimited in size, although of course limited in access rate. When working with a commercial cloud provider, ingest of data for storage can often be offered for free or for only a nominal cost. The primary cost results from 3 factors: the monthly cost to maintain the stored data, additional charges based on the rate of access within the cloud, an egress costs, the amount to move data *out* via a network. The cost to store 1 petabyte (PB) of data in the cloud varies significantly based on several factors, including the cloud provider, storage tier, and data access patterns. At the time of writing, cloud storage costs for 1PB ranged from a few thousand dollars per month for archival storage to tens of thousands of dollars per month for frequently accessed storage. The ngVLA project will also have to negotiate set fees (to be paid by the project or its end users) for access to both the Raw Visibilities and the Data Cubes in order for access to be enabled.

Following the analysis of the previous sections, let us consider our compute, storage, and networking costs separately for commercial cloud costs.

5.1.1 Cloud CPU Compute Cost Calculation

Referring back to the hardware requirements established in earlier sections, it is relatively straightforward to convert the compute load to a cloud-friendly unit, whether with CPU or GPU nodes.

For the initial deployment, we forecast 1,100 state of the art CPU nodes. If we assume those nodes will be used for the same five year period, our compute load is:

(Node Count)*(Node rate)*(Total Hours used per node per year)*(Years of operation).

For the 2030-2034 window, this would simply be 1,100 nodes * 5 years * 8,000 hours/per year.

The price per node hour is less straightforward to compute. Fortunately, the exact hardware node used in the base performance analysis is also available in the cloud (an AMD Milan 128 core node with 256GB of RAM). The 2025 Amazon rate for this node is $11.05/hour (Microsoft and Google adjust rates to match).

A more complicated cost to determine is how these rates change over time. Once again, due to the AI boom, cloud costs have increased per node (or per core) over the last several years. While historical pricing data is not available from the vendors, we have performed this analysis several times over the last several years and use that data for an extrapolation. While there has been some variance, within a given vendor's processor, and though *per socket* performance has varied substantively, *per core* performance has remained, to first order, flat. This allows us to look at historical x86 processor performance to find equivalent performance to a 128 core node.

Considering 128 cores of AMD processor, with 2GB/RAM per core, the price from 2014-2023 was roughly constant, at $8.55-$8.70 per hour (for an on-demand) instance. Since 2023, the 2025 price is $11.05 per hour, a sharp increase after a decade of relative stability (and, notably, no cost decrease per core over the full decade). It seems safe to assume the "AI bump" is likely a one-time jump in price, and the price will remain roughly constant from 2025 – though the specter of higher electricity costs for the cloud providers makes it seem prudent to factor in some inflation (in this analysis, we assume cloud costs will increase 3%/year from the baseline).

Note also that the on-demand rate is a "premium" rate – discounts may be available if commitments are made in advance, although time is billed whether or not it is *actually* used, based on reserving an instance from 1 month to 3 years. However, these discounts tend to be less on the "latest" processors in their first year, as they are most likely to be sold on-demand and a multi-year lock in on older hardware makes good economic sense for cloud providers.

With a 1 year reserved instance, the $11.05 price can be reduced to $7.30/hour, and the 3 year reserved instance to $5.01/hour, though you must assume a 1 generation old device for these rates (we will ignore this for now).

For on-demand, our projection assumes 8k hours per year, which is roughly the yield of a server you own with 95% uptime. For the reserved instance, you must pay for 8,760 hours per year, with annual payments at the start of the year.

A last consideration before projecting price – our CPU analysis assumed we would buy the latest processor socket available in the timeframe of the purchase. For example, AMD CPUs have already increased from 128 to 256 cores. This cloud analysis is based on a fixed *per core* price. For the second and third compute prices, we assume substantial per-socket improvements that are not included in a constant cloud price. While per core performance is not improving, cores per socket is. Therefore, we use a fixed 2025 core pricing for a 128 core system for the initial system purchase, but add an (extremely cloud-friendly) 50% premium in 2035 and 2039 for additional cores.

Our recommendation would be to do an initial compute hardware purchase of 1100 nodes for 5 years. Then in 2035, to increase the node count to 1450 for the next four years. And in 2029, increase the node count again to 1700 nodes.

5.1.2 GPU Compute Cost Calculation

While we are using CPU costs for the hardware estimate, regardless of the final decision on deployment, it is worth repeating the analysis for the cloud to see if there is a substantial change in cost. Unfortunately, GPU pricing is even more sensitive to AI demand. A state of the art Grace-Blackwell rack with 72 GPUs in August 2025 costs $1.65M for *90 days of cloud usage*, which is the maximum length of reservation. Even the conventional x86-Blackwell instances are only available on-demand (no discounted reservation rates), and at $113/hour for an 8 GPU node.

Based on that rate, our initial 100 node 400GPU system, would cost 30% more than the cloud CPU system.

5.1.3 Cloud Storage

Cloud storage has a large variety of rates. The default is an object store, which requires moving data to instance-local storage in order to actually compute with it (in Amazon, to a Lustre or Elastic Block Storage, or EBS tier). The object store is further subdivided into roughly a dozen tiers based on access pattern and retrieval time.

Compute costs likely make the cloud path infeasible, so this analysis will be cursory, with the most favorable assumption for the cloud. The most cost effective tier of storage, S3 Glacier “Deep Archive” with a 12-hour retrieval time for a single file, costs a mere $0.00099 per GB per month – which translates to $12 per TB per year. There is an additional 10 cent charge for each 1,000 accesses to the data, which can add substantially to the cost, but we will ignore that factor in this analysis. This would be suitable for the vast majority of the archive storage, assuming latency of up to a day or two for datacube retrieval is tolerable. This tier would not, however, support the expected ingest rates, and is not suitable for directly computing data products. For this analysis, we will use the Elastic Block Storage tier for primary storage, S3 Standard storage for our archive “front end”, and Glacier Deep Archive to replace the tape tier.

The 2025 pricing for S3 standard is $0.023 per GB per month, or $282/TB per year. The 2025 pricing for EBS, at the minimum performance (125MB/s), is $0.08 GB/month, or $983TB/year (the likely target I/O load to process data would involve a tier that is 50% more, with a variable charge based on actual IOPS used, which could further double the cost).

We would recommend:

- 2032: 25PB primary storage with 2PB archival cache and 50PB of archive capacity
- 2036: 60PB primary storage with 4PB archival cache and 500PB of archive capacity
- 2038: 120PB primary storage with 8PB archival cache and 2500PB of archive capacity

Substantial optimization is possible by reducing the EBS and doing more aggressive swapping to S3 (at additional software cost), but given the very friendly assumptions, and the likely cloud cost exceeding the full ngVLA project budget, further analysis seems unnecessary.

5.2 TACC Service Forecasting

A second service-based alternative would be to purchase the needed capacity from the NSF LCCF at TACC. TACC also offers cloud-based services, where the needed hardware capacity could be provided through existing computing and storage systems (or at this scale, additions to them). Lower costs can be passed-through due to lower requirements for uptime for capacity computing, economies of scale, and zero profit margin.

The 2025 rates for cloud computing, based on the same node type (AMD 128 core, 256GB RAM) at TACC are $0.45/hr for compute with CPUs, $1.80/hr for GPU computing (based on Hopper GPUs), with storage costs at $59.50/TB-year for online storage, and $16/TB-year for archive storage.

Using an identical analysis to the commercial cloud above, the projected costs to meet the ngVLA needs can be computed. We need similar adjustments for likely future costs, but we assume costs will rise proportional to cloud costs. We use the same adjustment factors for CPU nodes, simply substituting the rate. For GPUs, the 2025 rate is based on Hopper GPUs, and the Blackwells are substantially more expensive; we add one additional cost doubling beyond the commercial cloud analysis. For storage, as with cloud, we simply use a constant-cost assumption, though real rates are likely to fall as drive and tape density improves.

The total 2030-2044 cost for cloud-based services at TACC show that GPU costs are roughly double CPU costs.

5.3 Cloud Staffing Estimate

A cloud model reduces the need for staff to repair hardware – but most roles remain the same; the total ngVLA infrastructure would require spinning up and down thousands of cloud instances and multiple storage volumes, monitoring the state of software on each instance and overall system services, and maintaining cybersecurity across the infrastructure.

The primary expected reduction is in the 24x7 monitoring, where automation could be used (assuming systems staff will respond to pages 24x365 from automated monitoring). Staffing in this scenario would include: Systems Administration (1.5 FTE), Storage Administration (1), Network Administration (0.5), Cybersecurity Operations (0.5), and 24x7 Support Staff (2). To maintain/monitor cloud operations, we estimate the team would consist of at least 5 staff members, for a net savings of 4 FTEs from the buy-build hardware scenario.

6. Other Cyber Requirements

In conversations with the ngVLA project team, the need for **additional coding assistance** was a factor that came up repeatedly as planning for this document took place. Traditionally, many research data centers have staff on hand that can help software engineers for specific science use cases tune their codes to take the most advantage of the hardware at that center. Some funding programs have even formalized this sort of assistance [CC], as it has been recognized that using the resources in a smarter way reduces wasted cycles and enables additional capacity for end users. Given the projected cyberinfrastructure costs forecast in this proposal, it can be said without hyperbole that even a 10-20% improvement in code efficiency could have hundreds of millions of dollars of impact on TPC estimates.

We have seen even larger payoffs, though they should not be expected. A recent multi-year project for the LCCF spanning 11 applications had one code (part of a common open source quantum materials package) yield a speedup of over 100x due to an algorithmic change (on the same hardware). Several other codes saw speedups in the 20% range, with others closer to 5%.

As part of the work defined by the ngVLA data center, we strongly recommend that whatever data center location is selected, that they consider the staffing on the coding help side in addition to other factors.

This type of expertise can be hired or can be made available through a number of centers and National Labs. Again, for reference, we price this using typical experienced HPC developer staff from LCCF/TACC. We estimate that no more than 2-3 FTE would be needed, with likely some “surge” in the first few years of construction, then some tapering in later years.

7. Analysis and Summary

In this document, we have looked at the costs of procuring hardware to meet the ngVLA scientific requirements, various options to site and host the hardware, and various options to purchase the same capability through a service model. A summary is given in Table 1, replicated here.

Table 1: Summary of final cost comparison for four alternative architectures.

Solution for System and Service Options	Cost through 2044
Option 1: Purchase Hardware deploy to CoLo	~2.3x Option 2
Option 2: Purchase Hardware deploy to LCCF	Least Expensive
Option 3: Purchase Service through Commercial Cloud	~12.4x Option 2
Option 4: Purchase Service through LCCF	~2x Option 2

Of course, there are always options beyond cost that must be considered. Different solutions provide different levels of support and investment in a successful outcome. Commercial cloud solutions often come with lock-in due to the cost of egress of data. Hosted services do provide a more flexible scale-up of the system, as you only pay for what you are using. We provide these as examples of the information needed to do a future cost-benefit evaluation of the different possibilities. We also discuss the projected cost-benefit of GPU-based computing and the need to monitor that closely before procuring any hardware.

8. Conclusion

Both the construction phase through 2039 and the operations phase from 2039 through 2044 currently exceed the projected programmatic envelope by a significant margin. These estimates carry substantial uncertainty - they depend on assumptions about hardware pricing trajectories, software efficiency, ingest rates, and the eventual mix of observing programs - and should be treated as a planning baseline rather than a fixed forecast. The proposed architecture is, however, fully scalable: once a programmatic envelope is established, the Data Processing Center can be sized to fit it. Several descope levers are available to bring the design into envelope, including (1) shortening the online retention window for visibility data, which is among the dominant drivers of storage cost; (2) excluding or capping a small subset of high-demand use cases - most notably the high-bandwidth spectral-scan program, whose peak data rates and compute requirements drive a disproportionate share of total cost; (3) continued investment in software and pipeline optimization, where even a 10–20% efficiency gain has been shown to translate into hundreds of millions of dollars of TPC impact; (4) selection of the most cost-effective hosting model (purchased hardware sited at the TACC LCCF) over commercial cloud or stand-alone colocation alternatives; and (5) emerging ML-based approaches to imaging and calibration that may further reduce sustained compute demand. Taken together, these levers provide meaningful flexibility to align the final design with whatever programmatic envelope is ultimately approved.

9. References

[ABQG] Albuquerque GigaPoP (ABQG), https://abqg.unm.edu

[AMPATH] AMPATH, https://ampath.net/

[CC] Campus Champions: Uniting Research Computing Facilitators, https://campuschampions.cyberinfrastructure.org/

[CW25] Chris Wilkinson, Internet 2, Senior Director of Network Infrastructure and Operations, Personal communication,8/7/25

[EB25] Ed Balas, ESNet Group Lead, Personal communication, 8/15/25

[FD24] FasterData, “Performance Expectations for a 100G Host”, November 5, 2024,https://fasterdata.es.net/performance-testing/performance-expectations-for-a-100g-host

[Globus] Globus, https://globus.org

[gNOC] GlobalNOC at Indiana University, https://globalnoc.iu.edu/

[Hiriart24] Rafael Hiriart, “ngVLA Data Rates and Computational Loads (Update)” ngVLA Computing Memo 11, August 30, 2024, https://library.nrao.edu/public/memos/ngvla/NGVLAC_11.pdf

[Ibarra25] Julio Ibarra, “AmLight: The Next Frontier Towards Discovery in the Americas and Africa”, NSF Award #2537489, 2025-2030, $9M, https://www.nsf.gov/awardsearch/showAward?AWD_ID=2537489

[IBM22] Performance specifications for LTO tape drives, 2022, https://www.ibm.com/docs/en/ts4500-tape-library?topic=performance-lto-specifications

[iNOC] iNOC, https://www.inoc.com/

[Internet2] Internet2, https://internet2.edu/

[LTO] Linear Tape Open, https://www.lto.org/what-is-lto/

[NS1] NetSage Dashboard, Sample data transfers between TACC and NRAO, https://tacc.netsage.io/grafana/d/-l3_u8nWk/what-do-individual-flows-by-organization-look-like?var-src=National%20Radio%20Astronomy%20Observatory%20%28NRAO%29&from=2025-02-01T06:00:00.000Z&to=2025-08-01T04:59:59.000Z&timezone=browser&orgId=1&var-src_resource=$__all&var-dest=Texas%20Advanced%20Computing%20Cent

er%20%28TACC%29&var-dest=tacc&var-dst_resource=$__all&var-subnet=&var-sensors=$__all&var-country_scope=$__all&var-is_net_test=yes

[OmniSOC] OmniSOC at Indiana University, https://researchsoc.iu.edu/

[OST16] rJ. Opara-Martins, R. Sahandi, and F. Tiag, "Critical Analysis of vendor lock-in and its impact on cloud computing migration: a business perspective", Journal of Cloud Computing: Advances, Systems, and applications, 5:4, 2016.

[Quilt25] Quilt25 Mail Discussion List, August 2025

[Rapier25] Chris Rapier, HPN-SSH: High performance SSH/SCP, https://www.psc.edu/hpn-ssh-home/

[RT25] Mauricio Rojas, Eduardo Toro, "NOIRLab - IT Ops", SA3CC Meeting 2025, https://www.amlight.net/wp-content/uploads/2025/05/NOIRLab-ITOps-SA3CC-2025.pdf

[SA3CC25] South American – African Astronomy Coordination Committee (SA3CC) Meeting 2025, May 5-9, 2025, https://www.amlight.net/?p=6016