DESI ARD Schema Justification Document: Architectural Defense for Edge-Constrained High-Dimensional Spectroscopy

Executive Summary

The following document serves as the comprehensive architectural defense and schema justification for the "Perfect" DESI Scientific Analysis Ready Dataset (ARD). This report moves beyond high-level conceptualization to provide a rigorous, engineering-grade validation of the data structures, storage formats, and computational pipelines required to democratize petascale spectroscopy.

The design philosophy is predicated on a specific, constrained hardware profile—a 4-node "Edge-Cloud" cluster typical of university research groups—comprising 144GB RAM per node, NVIDIA A4000 (16GB VRAM) accelerators, and Samsung PM893 SATA Enterprise SSDs. This constraint envelope is not treated as a limitation, but as the primary forcing function for architectural efficiency. The central thesis of this defense is the "One-Time Cost" philosophy: the strategic decision to front-load high-friction computational tasks (Bayesian inference, topological reconstruction, and neural embedding) into the ingestion pipeline, thereby converting processor-intensive operations into storage-efficient, queryable columns.

By synthesizing the latest research in astrophysical foundation models (e.g., Universal Spectrum Tokenizer, AstroCLIP), accelerated Bayesian sampling (Bagpipes/Nautilus), and computational topology (DisPerSE), this document demonstrates how the proposed schema reduces the "time-to-science" from months to seconds. The analysis justifies the selection of Apache Parquet and HDF5 over competing formats like Zarr, defends the use of specific floating-point precisions based on GPU tensor core capabilities, and provides a detailed throughput model proving the feasibility of this architecture on the specified hardware.

---

1. Architectural Philosophy and Hardware Constraints

The modern extragalactic survey landscape is defined by a fundamental friction: the divergence between data volume and analysis latency. The Dark Energy Spectroscopic Instrument (DESI), a Stage-IV dark energy experiment mounted on the Mayall 4-meter telescope, exemplifies this regime change.[1] With Data Release 1 (DR1) delivering approximately 18.7 million spectra of galaxies, quasars, and stars, the community now possesses a statistical sample of the universe that is orders of magnitude larger than previous surveys like SDSS or VIPERS.[3]

However, the raw spectra provided in standard data releases are merely the feedstock of scientific inquiry. Converting a flux array into a physical understanding of galaxy evolution—deriving stellar masses, star formation histories, or environmental metrics—requires complex post-processing pipelines. A single full Bayesian Spectral Energy Distribution (SED) fit can consume CPU-hours.[1] Multiplied by the 18 million objects in DR1, the computational cost to extract these standard metrics approaches millennia of single-core compute time.

This friction creates a "compute wall" that effectively locks the scientific potential of the archive behind a barrier of High-Performance Computing (HPC) allocation requirements. The ARD proposed herein is designed to dismantle this wall. By shifting the burden of computation from the end-user's query time to the curator's ingestion time, the ARD transforms the DESI archive from a static library into a dynamic engine for discovery.

1.1 The Target Infrastructure: The 4-Node Cluster

This schema justification is strictly bound by the physical realities of the target hardware. We assume a "24/7 On-Premise Cluster" with the following specifications, which represent a realistic mid-scale infrastructure accessible to departmental research groups:

- **Compute Nodes:** 4x interconnected nodes.
- **Memory (RAM):** 144 GB DDR4 ECC RAM per node.
- **Acceleration:** 1x NVIDIA RTX A4000 (16 GB GDDR6 VRAM) per node.
- **Storage:** Samsung PM893 1.92TB SATA 6Gb/s Enterprise SSDs.
- **Networking:** 10GbE LACP (Link Aggregation Control Protocol).

Every architectural decision—from the choice of embedding dimension to the file compression algorithm—is a direct response to the limitations and capabilities of these

components.

1.2 Constraint Analysis and Design Implications

To defend the schema, we must first rigorously analyze the bottlenecks imposed by this hardware profile.

1.2.1 The Storage Throughput Limit (SATA vs. NVMe)

The most critical bottleneck in this architecture is the storage interface. The Samsung PM893 is an Enterprise **SATA** SSD, not an NVMe PCIe 4.0 drive.[5] This distinction is paramount. While modern NVMe drives can achieve sequential read speeds exceeding 7,000 MB/s, the SATA 6Gb/s interface physically caps the PM893 at approximately **550 MB/s** for sequential reads and **520 MB/s** for sequential writes.[8]

Furthermore, the random read performance is rated at approximately **98,000 IOPS** (Input/Output Operations Per Second) for 4KB blocks.[5] While this is significantly faster than mechanical spinning rust (HDD), it is an order of magnitude slower than enterprise NVMe drives (e.g., the Samsung PM9A3 achieves 1,000,000 IOPS and 6,800 MB/s).[10]

Architectural Defense: This bandwidth constraint makes the use of **Apache Parquet** non-negotiable for the metadata catalog. A row-oriented format (like CSV or FITS tables) requires reading the entire row to access a single field. If a user wants to plot Stellar Mass vs. Redshift for 18 million galaxies, a row-oriented scan would necessitate reading hundreds of gigabytes of irrelevant data (fluxes, errors, flags), saturating the 550 MB/s SATA bus and resulting in query times measured in minutes or hours. Parquet's columnar storage structure allows the system to read *only* the specific column chunks required for the query. This reduces the effective I/O load by 95-99% for typical analytical queries, artificially inflating the "effective" throughput of the SATA drives to levels comparable to NVMe performance for specific workloads.[1]

1.2.2 The 16GB VRAM Limit (The AI Bottleneck)

The NVIDIA RTX A4000 is a powerful Ampere-architecture GPU, but it is constrained by **16 GB**

of VRAM.[1] This memory ceiling imposes hard limits on the "AI Layer" of the ARD. It categorically rules out the training of massive, multi-billion parameter Large Language Models (LLMs) or the inference of extremely large Vision Transformers (ViTs) with large batch sizes without complex strategies like model parallelism or aggressive quantization.[12]

Architectural Defense: The ARD design accommodates this by relying on **Pre-Trained Foundation Models** specifically selected for their parameter efficiency (e.g., the Universal Spectrum Tokenizer at ~300M parameters and AstroCLIP at ~43M parameters).[14] Furthermore, this constraint validates the "One-Time Cost" philosophy: users cannot be expected to run inference on these models dynamically for millions of objects. The embeddings must be pre-calculated and stored as static vectors. The 16GB limit also dictates the use of **Float16 (Half-Precision)** inference and storage for the embeddings, which halves the memory footprint and leverages the A4000's specialized Tensor Cores, which offer 76.7 TFLOPS of FP16 performance compared to only 19.2 TFLOPS for FP32.[11]

1.2.3 The 144GB System RAM Limit (The Topology Bottleneck)

While 144 GB of system RAM is substantial, it is the "middle ground" of computing. It is sufficient to hold the metadata catalog and search indices for 10-20 million objects in memory if optimized (approx. 20-40 GB). However, it is fundamentally insufficient for "Big Data" operations that require loading the entire spectral dataset (Terabytes) into RAM or calculating the Delaunay tessellation for the full survey volume in a single pass, which requires maintaining a massive complex of simplices and vertices.[16]

Architectural Defense: This limitation mandates a **Partitioned Schema**. The data must be physically sliced (sharded) by sky region (using the **HEALPix** scheme) so that users or pipeline processes can load and analyze a specific chunk of the universe without exhausting the node's memory. It also requires the use of **Domain Decomposition** for the topological layer, where the cosmic web is reconstructed in overlapping sub-volumes and stitched together, rather than processed as a monolithic block.[17]

1.3 The 4-Node Cluster Strategy: Distributed ingest

The presence of four distinct nodes allows for a parallelized ingestion strategy that linearly scales the "One-Time Cost" reduction. By distributing the HEALPix pixels across the four nodes, the total time required to generate the ARD—processing embeddings, running

Bayesian fits, and calculating topology—is divided by four. With 18 million spectra, a single GPU would struggle to complete the embedding generation in a reasonable timeframe. With 4x A4000s, the throughput (benchmarked at hundreds of spectra per second per GPU for transformer encoders) allows the full DR1 to be processed in days rather than months.[18] This parallel capability defends the feasibility of creating such a dense, pre-computed dataset on modest hardware.

---

2. The Data Ingestion Challenge: Taming the DESI Archive

To justify the ARD schema, we must first understand the scale and structure of the raw material: DESI Data Release 1 (DR1). The volume and complexity of DR1 necessitate a rigorous ingestion pipeline to transform raw FITS files into analysis-ready Parquet and HDF5 structures.

2.1 DESI DR1: Volume and Structure Analysis

DESI DR1 is a massive repository. It includes spectra for over **18.7 million unique targets** observed during the main survey (May 2021 – June 2022), plus a reprocessing of the Survey Validation (SV) data.[3] The breakdown of these objects is critical for sizing the database:

- **Galaxies:** ~13.1 million.
- **Quasars (QSOs):** ~1.6 million.
- **Stars:** ~4.0 million.[3]

The data is distributed in two primary formats: **Exposures** (single observations) and **Coadds** (combined spectra). The ARD focuses on Coadds for the primary catalog but must retain linkage to exposures for time-domain science.

2.1.1 File Sizes and Storage Footprint

The raw data volume for DR1 exceeds **300 TB**.[19] However, the spectroscopic data relevant for the ARD is more compact.

- **Coadd FITS Files:** These are grouped by HEALPix or Tile. A typical coadd file containing

spectra for multiple targets is approximately **213 MB**.[20]

- **Single Spectrum Size:** A single DESI spectrum covers the wavelength range **3600 Å to 9800 Å** across three arms (Blue, Red, NIR).[21] With a resolution $R \sim 2000-5000$, this results in approximately 4,000-6,000 pixel bins per arm.
- **Total Spectral Storage:** Storing the full resolution flux, inverse variance (ivar), and resolution matrices for 18 million objects consumes significant space. If we estimate ~100KB per compressed spectrum (flux+ivar only), the spectral payload alone is nearly **2 TB**. This fits comfortably within the aggregated **7.68 TB** storage of the 4-node cluster (4 x 1.92 TB), but it leaves little room for inefficiency. This tight margin reinforces the need for efficient compression (HDF5/LZF) and the discarding of redundant raw data in favor of the processed ARD products.

2.2 Ingestion Pipeline Design

The ingestion pipeline is the mechanism that populates the ARD schema. It must run continuously on the 4-node cluster, fetching raw FITS files, processing them through the three analysis layers, and committing the results to the optimized storage.

1. **Fetcher Daemon:** Monitors the DESI SAS (Science Archive Server) mirror or local raw dump. It identifies new or updated HEALPix pixels.
2. **Distributor:** Assigns HEALPix pixels to one of the 4 compute nodes based on hashing, ensuring load balancing.
3. **Processor (The "One-Time Cost"):**
    - **Load:** Reads the FITS coadd.
    - **AI Pass:** Batches spectra into the A4000 GPU for embedding generation (Shen'25, AstroCLIP).
    - **Physics Pass:** Runs Bagpipes (CPU) on the spectra to generate scalar properties.
    - **Topology Pass:** Updates the local density field for DisPerSE processing.
4. **Writer:** Commits metadata to Parquet (columnar) and spectra to HDF5 (chunked).

2.3 The "Exposures vs. Coadds" Decision

A critical schema decision is whether to index every single exposure or just the coadds.

- **Context:** DESI observes targets multiple times. The "Coadd" is the weighted sum of these exposures, providing the highest Signal-to-Noise Ratio (SNR).
- **Decision:** The ARD primary table indexes **Coadds**.
- **Justification:** For 99% of extragalactic science (redshifts, masses, clustering), the

Coadd is the optimal product. Indexing individual exposures would multiply the row count by a factor of 3-5, diluting the "Analysis Ready" nature of the dataset and bloating the catalog beyond the RAM limit. Exposure-level data is relegated to a secondary "Time Domain" table linked by TargetID, loaded only on demand.

---

3. The AI Layer Schema: Democratizing Latent Spaces

The first and most transformative layer of the ARD is the integration of Deep Learning (DL) embeddings. These vectors represent a compression of the high-dimensional, noisy spectral data into a dense, semantic latent space. This layer moves the schema from a "Data Archive" to a "Neural Atlas."

3.1 Model Selection: Balancing Physics and Morphology

The choice of Foundation Models is constrained by the A4000's 16GB VRAM and the scientific requirement for diverse utility. We defend the inclusion of two distinct models that offer orthogonal views of the data.

3.1.1 The Universal Spectrum Tokenizer (Shen et al. 2025)

- **Architecture:** This model utilizes a **Transformer** architecture designed specifically for sequential data. Crucially, it processes spectra on their **native wavelength grids**.[14]
- **Scientific Justification:** Traditional ML approaches require resampling spectra to a fixed grid (e.g., 10,000 bins from 3600-9800Å). This resampling introduces interpolation artifacts and correlates noise, degrading the signal in narrow emission lines. The Shen et al. model tokenizes the spectrum as a sequence of (flux, wavelength) pairs, preserving the raw instrumental resolution. It is trained via **Self-Supervised Masked Modeling**, where it learns to predict missing segments of the spectrum. This objective forces the model to learn the underlying physics of stellar atmospheres and galaxy SEDs (e.g., "if I see H-beta and OIII here, I expect H-alpha there").[1]
- **Hardware Defense:** The model size is moderate, typically **50M to 300M parameters**.[24] On the A4000, a 300M parameter model in FP16 precision requires roughly 600MB-1GB of VRAM for weights. This leaves ~15GB of VRAM for activation overhead and large batch sizes. Benchmarks for similar Transformer architectures (e.g., BERT-Large) on the A4000

suggest inference throughputs of hundreds of sequences per second.[25] This ensures the cluster can process the 18M spectra in DR1 within a reasonable timeframe (days).
- **Schema Entry:** emb_shen (Array<Float16>, 768 dimensions).

3.1.2 AstroCLIP (Parker et al. 2024)

- **Architecture:** AstroCLIP is a **Multimodal** Foundation Model. It consists of a spectral encoder (1D Transformer) and an image encoder (Vision Transformer, ViT-B/16, initialized with **DINOv2**).[26] These encoders are aligned using a **Contrastive Loss (InfoNCE)** function, which maximizes the cosine similarity between the spectrum of a galaxy and its corresponding image.[15]
- **Scientific Justification:** While the Universal Tokenizer captures *physical* properties (SED shape, lines), AstroCLIP captures *morphological* concepts. It learns to associate spectral features with visual shapes (e.g., associating strong emission lines with spiral arms or clumpy star formation). Including this embedding allows users to query the spectral database using morphological concepts (e.g., "Find spiral galaxies") without needing actual imaging data or morphological catalogs.[28]
- **Hardware Defense:** The spectral encoder component of AstroCLIP is lightweight, approximately **43M parameters**.[1] This is computationally inexpensive to run in inference mode. The image processing (DINOv2) is heavier but only needs to be run once per target if images are available; often, we only need the spectral encoder for the ARD if the goal is spectral retrieval.
- **Schema Entry:** emb_astroclip (Array<Float16>, 512 dimensions).

3.2 Operationalizing Embeddings on the Edge Cluster

Storing high-dimensional vectors for 18+ million objects creates a significant data management challenge that must be addressed by the schema.

- **Dimensionality and Storage:**
  - Shen et al.: 768 dimensions $\times$ 2 bytes (Float16) = 1,536 bytes/object.
  - AstroCLIP: 512 dimensions $\times$ 2 bytes (Float16) = 1,024 bytes/object.
  - Total per object: ~2.5 KB.
  - Total for 18M objects: ~45 GB.
- **Justification:** This 45 GB footprint fits entirely within the **144 GB System RAM** of a single node (and easily across the 576 GB aggregate RAM of the cluster). This validates the decision to store embeddings as simple columns in the Parquet files.

- **Indexing Strategy:** Because the entire embedding dataset fits in RAM, the ARD does not strictly require complex disk-based vector indices (like IVF-ADC on disk). We can utilize **In-Memory FAISS** (Facebook AI Similarity Search) indices constructed dynamically or cached on the NVMe/SATA drives. The schema allows for a "brute-force" scan or exact k-NN search in memory, which provides the highest possible recall accuracy.
- **Float16 Precision:** The A4000 GPUs include **Tensor Cores** specifically optimized for FP16 math (76.7 TFLOPS).[11] Storing embeddings in Float16 matches the native precision of the inference engine and reduces the I/O load on the Samsung PM893 SATA SSDs by 50% compared to Float32, without any meaningful loss in retrieval quality for scientific similarity tasks.

---

4. The Physics Layer Schema: Bayesian Posterior Compression

While embeddings allow for powerful qualitative analysis (clustering, anomaly detection), quantitative astrophysics requires robust physical scalars: Stellar Mass (M_*), Star Formation Rate (SFR), Metallicity (Z), and Dust Attenuation (A_V). The derivation of these parameters is the classic "High-Friction" bottleneck.

4.1 The Computational Wall and the Accelerated Solution

The "Gold Standard" for deriving physical properties is Bayesian Full Spectral Fitting (SED fitting). Codes like **Prospector** or **Bagpipes** generate synthetic spectra from stellar population models, redden them with dust laws, and compare them to observations.[1]

- **The Problem:** A standard MCMC (Markov Chain Monte Carlo) fit can take hours to converge for a single galaxy. Scaling this to 18 million objects is computationally intractable for a user on a standard workstation.
- **The ARD Solution:** The ARD pipeline utilizes **Bagpipes** coupled with the **nautilus** nested sampling algorithm.[4] Nautilus utilizes neural networks to approximate the posterior distribution boundaries during sampling, accelerating convergence by orders of magnitude compared to traditional samplers like MultiNest or dynesty. This reduces the runtime from hours to minutes per galaxy.
- **Throughput:** With 4 nodes $\times$ roughly 32-64 cores per node (standard for 144GB nodes), the cluster offers ~128-256 vCPUs. At 5 minutes per fit, the cluster can process ~30,000-70,000 galaxies per day. Over a year, the full BGS (Bright Galaxy Survey) and

key LRG (Luminous Red Galaxy) samples can be fully characterized.

4.2 Schema Structure: Posterior Percentiles

Standard catalogs often provide only the "Best Fit" value (Maximum A Posteriori), discarding the critical uncertainty information inherent in Bayesian analysis. Conversely, storing the full posterior chains (which can be Gigabytes per object) is impossible on the 1.92TB drives.

Schema Definition: The ARD employs a compression strategy that stores the **16th, 50th, and 84th percentiles** of the posterior distribution for every parameter.

- **Columns:**
  - logM_p16, logM_p50, logM_p84 (Float32) – Stellar Mass.
  - logSFR_p16, logSFR_p50, logSFR_p84 (Float32) – Star Formation Rate.
  - dust_Av_p50 (Float32) – Dust Attenuation.
  - age_p50 (Float32) – Mass-weighted Age.
- **Defense:** This "1-sigma" summary allows users to assume a split-normal distribution and reconstruct the error bars dynamically. It preserves the asymmetry of the posterior (common in SFRs and dust) while compressing the data footprint by a factor of ~1000x compared to raw chains.
- **Data Type:** We use **Float32**. The systematic uncertainties in SED fitting (due to Initial Mass Function assumptions, isochrone errors) far exceed the precision difference between 32-bit and 64-bit floats. Using Float32 reduces the column width by 50%, improving scan speeds on the bandwidth-limited Samsung PM893 drives.

4.3 Empirical Truths: Kinematics and Lick Indices

Bayesian models depend on priors (e.g., assumed Star Formation Histories) which can be incorrect. The ARD must provide model-independent empirical measurements to serve as a "Reality Check."

pPXF (Penalized Pixel-Fitting):
We utilize pPXF to extract stellar kinematics and emission line fluxes. pPXF is a linear least-squares solver that is extremely fast (seconds per object) and robust.1

- **Columns:** vel_disp (σ_*), v_gas, h_alpha_flux, oiii_5007_flux.

Lick Indices:
The ARD pre-computes standard absorption line indices, specifically $D_n(4000)$ (the

4000Å break strength) and $H\delta_A$ (Balmer absorption).31

- **Scientific Use Case:** The combination of $D_n(4000)$ and $H\delta_A$ forms the classic "Quenching Diagram." It allows researchers to instantly identify "Green Valley" (transitioning) galaxies and "Post-Starburst" (E+A) galaxies (which have strong $H\delta$ but high D_n4000).[1] By providing these as pre-computed columns, the ARD allows users to perform this classification via a simple SQL query (SELECT * FROM cat WHERE Dn4000 > 1.5 AND Hdelta > 3), shielding them from the complexity of continuum normalization and feature integration.

---

5. The Topological Layer Schema: The Pre-Computed Cosmic Web

A galaxy's evolution is determined not just by its internal physics ("Nature"), but by its environment ("Nurture"). The Topological Layer embeds every object into the context of the Cosmic Web (Clusters, Filaments, Sheets, Voids).

5.1 The Memory Constraint and Domain Decomposition

The standard tool for quantifying the Cosmic Web is **DisPerSE** (Discrete Persistent Structures Extractor).[17] It uses Discrete Morse Theory to identify topological features in the density field calculated via the Delaunay Tessellation Field Estimator (DTFE).

- **The Constraint:** Calculating the Delaunay tessellation for 18 million points is memory-intensive. The number of tetrahedra scales as $\sim 6N$. Storing the full complex and filtration values for the entire DESI volume requires >500 GB of RAM.[16] This exceeds the 144 GB limit of our individual compute nodes.
- **The Solution:** The ARD pipeline employs **Domain Decomposition**. The survey volume is sliced into overlapping cubic sub-volumes (with buffer zones to mitigate edge effects). Each node processes a sub-volume within its 144 GB RAM envelope. The resulting skeletons are then stitched together.
- **Justification:** This approach is mandatory given the hardware. It validates the "One-Time Cost" philosophy because an end-user with a standard workstation (e.g., 32-64 GB RAM) simply *cannot* run DisPerSE on the full DESI catalog. The ARD must provide the output because the user cannot generate it themselves.

5.2 Schema Definition: Environmental Metrics

We distill the complex topological skeleton into three specific, galaxy-centric metrics stored as columns:

1. The Web Classification (Web_Class)

- **Type:** Int8 (Categorical).
- **Values:** 0 (Void), 1 (Sheet), 2 (Filament), 3 (Knot/Cluster).[1]
- **Defense:** An Int8 column is negligible in size (1 byte/row) but provides the highest-level environmental context. It allows for rapid subsets (e.g., "Compare the Mass Function of Void Galaxies vs. Cluster Galaxies").

2. Distance to Filament (Dist_Filament)

- **Type:** Float32 (Mpc).
- **Definition:** The Euclidean distance from the galaxy to the nearest segment of the "Persistence 3-sigma" filament skeleton.
- **Scientific Defense:** Literature from surveys like GAMA and VIPERS demonstrates that galaxy properties (like quenching) correlate strongly with distance to filaments, even at fixed local density.[32] This metric probes "anisotropic assembly bias" and gas accretion mechanisms (cold flows) that are theoretically channeled along filaments.[35]

3. Local Density (Sigma_5)

- **Type:** Float32.
- **Definition:** Surface density calculated from the distance to the 5th nearest neighbor ($k=5$).[20]
- **Defense:** Providing *both* Dist_Filament (Global Topology) and Sigma_5 (Local Environment) allows users to disentangle "Nature vs. Nurture" effects. Recent studies suggest that filamentary quenching might be distinct from cluster quenching.[33] The ARD empowers this sophisticated analysis by providing the orthogonal metrics required to test it.

---

6. Storage Architecture and File Formats

The physical realization of the schema is dictated by the I/O characteristics of the Samsung PM893 SATA SSDs.

6.1 The Metadata Catalog: Apache Parquet

- **Choice:** Apache Parquet (with Snappy or Zstd compression).
- **Justification:** The PM893 SATA drive has a sequential read limit of ~550 MB/s. This is the narrowest pipe in the system.
    - **Columnar Efficiency:** Analytical queries in astrophysics typically access only a small subset of columns (e.g., plotting M_star vs SFR uses 2 columns out of 500). A row-oriented format (CSV, FITS binary table) would require reading the entire 300+ byte row for every galaxy to access these two floats. This would waste >95% of the I/O bandwidth. Parquet allows the system to read *only* the chunks for M_star and SFR.
    - **Throughput Model:** Reading 2 columns of Float32 for 18 million objects involves transferring roughly $18M \times 8 \text{ bytes} \approx 144 \text{ MB}$. On a 550 MB/s drive, this takes **< 0.3 seconds**. A full row scan could take 10-20 seconds per query. This difference defines the user experience between "interactive" and "sluggish."
    - **Partitioning:** The data is partitioned by **HEALPix** (High-Order Sky Pixel). This aligns the file structure with the physical distribution of data on the sky, enabling "predicate pushdown"—queries restricted to a specific sky area only open the relevant files, further saving I/O.

6.2 The Spectral Data: HDF5 (Chunked)

While metadata fits in Parquet, the actual spectral flux arrays (18M objects $\times$ 12,000 pixels $\times$ Float32) are too large and multi-dimensional for efficient Parquet storage.

- **Choice:** HDF5 (Hierarchical Data Format v5).
- **Comparison to Zarr:** Zarr is often preferred for cloud object storage (S3) because it stores each chunk as a separate object. However, on a local POSIX filesystem (like the XFS/EXT4 on our SATA SSDs), creating one file per chunk for 18 million spectra would result in tens of millions of tiny files. This would exhaust the **inode limit** of the filesystem and degrade performance due to metadata overhead.[36]
- **Defense:** HDF5 consolidates these chunks into a single file (or one file per HEALPix group). This manages the inode count efficiently while still supporting fast **Random Access** to individual spectra via efficient B-tree chunk indexing. The 4-node cluster can handle the HDF5 library overhead easily, making it the superior choice for this local hardware configuration.

---

7. Operationalizing the "Swiss Army Knife"

The ARD is not just a dataset; it is a platform that enables specific, high-value scientific workflows. We defend the schema by demonstrating how it operationalizes complex queries on the specified hardware.

7.1 Workflow: Rare Object Search (The "Green Pea" Hunt)

- **Traditional Workflow:** User downloads TBs of spectra, writes a script to measure line ratios, filters for high [OIII]/Hb, visually inspects thousands of candidates. Time: Weeks.
- **ARD Workflow:**
  1. **Query:** User selects one known "Green Pea" galaxy.
  2. **Vector Search:** User queries the emb_astroclip column for the 100 nearest neighbors.
  3. **Hardware Path:** The system loads the embedding vectors (Float16) into the 144GB RAM (fitting easily). It performs a cosine similarity search in memory.
  4. **Result:** The system returns objects that *look* and *behave* like Green Peas, even if they are at different redshifts or have slightly different line ratios that a hard cut would miss.
  5. **Time:** Seconds.

7.2 Workflow: Environmental Quenching Analysis

- **Scientific Question:** "Do galaxies quench faster in filaments than in voids at fixed stellar mass?"
- **ARD Workflow:**
  1. **Query:** SELECT logM_p50, logSFR_p50, Web_Class FROM catalog
  2. **Hardware Path:** Parquet performs column projection, reading only these 3 columns from the SATA SSDs. The I/O load is negligible.
  3. **Analysis:** User groups by Web_Class (0 vs 2) and bins by logM_p50.
  4. **Time:** Sub-second.
- **Defense:** This analysis utilizes the Web_Class (Topology), logM (Physics), and logSFR (Physics) layers simultaneously. The schema's pre-computation of these values makes a

paper-quality plot an instantaneous operation.

---

8. Conclusion

The DESI ARD Schema is a rigorous architectural response to the challenge of high-dimensional spectroscopy in the era of "Edge-Cloud" science. It rejects the assumption that petascale data requires petascale hardware for analysis. Instead, it leverages the constraints of the 4-node, 144GB RAM, 16GB VRAM, SATA-based cluster to enforce a **"One-Time Cost"** design philosophy.

By carefully selecting Foundation Models that fit the GPU envelope (Universal Tokenizer, AstroCLIP), optimizing storage formats for SATA throughput (Parquet), and utilizing Domain Decomposition to circumvent RAM limits for topology, this schema transforms the 4-node cluster into a discovery engine capable of rivaling national supercomputing centers. It delivers a dataset that is not merely "available," but truly "Analysis Ready," bridging the gap between the raw photon and the physical insight.

Appendix: Data Schema Summary

Category	Column Name	Data Type	Source Tool	Description	Constraint Defense
ID	TargetID	Int64	DESI Pipeline	Unique Identifier	Primary Key for HDF5 Chunking
AI Layer	emb_shen	Array	Universal Tokenizer	768-d Physical Embedding	Float16 fits storage & GPU Tensor Cores
AI Layer	emb_astroclip	Array	AstroCLIP	512-d Morphological	Multimodal (Image+Spec)

				Embedding	alignment
Physics	logM_p16/p50/p84	Float32	Bagpipes	Stellar Mass Posterior	Percentiles compress chains 1000x
Physics	logSFR_p16/p50/p84	Float32	Bagpipes	SFR Posterior	Captures asymmetric errors
Physics	Dn4000	Float32	Lick Analysis	Age Indicator	Empirical check on models
Topology	Web_Class	Int8	DisPerSE	0=Void, 1=Sheet, 2=Fil, 3=Knot	Categorical, minimal storage
Topology	Dist_Filament	Float32	DisPerSE	Distance to Filament (Mpc)	Pre-computed via Domain Decomposition
Topology	Sigma_5	Float32	KNN ($k=5$)	Local Surface Density	Separates Local vs Global environment

Works cited

1. DESI ARD_ Embeddings, Physics, Topology.pdf
2. Dark Energy Spectroscopic Instrument | Kavli Institute for Particle Astrophysics and Cosmology (KIPAC), accessed November 23, 2025, https://kipac.stanford.edu/research/projects/dark-energy-spectroscopic-instrument
3. [2503.14745] Data Release 1 of the Dark Energy Spectroscopic Instrument - arXiv, accessed November 23, 2025, https://arxiv.org/abs/2503.14745

4. DESI Data Release 1: Stellar Catalogue - arXiv, accessed November 23, 2025, https://arxiv.org/html/2505.14787v1
5. MZ7L3960HCJR-00A07 (960 GB) | PM893 | Samsung Semiconductor Global, accessed November 23, 2025, https://semiconductor.samsung.com/ssd/datacenter-ssd/pm893/mz7l3960hcjr-00a07/
6. PM893 | Data center SSD | Samsung Semiconductor Global, accessed November 23, 2025, https://semiconductor.samsung.com/ssd/datacenter-ssd/pm893/
7. Samsung PM893 1.92TB SATA 6GB/s 2.5" Solid State Drive - Cloud Ninjas, accessed November 23, 2025, https://cloudninjas.com/products/samsung-pm893-1-92tb-sata-6gb-s-2-5-solid-state-drive
8. SAMSUNG PM893 2.5" 1.92TB SATA III V-NAND TLC Enterprise Solid State Drive - Newegg, accessed November 23, 2025, https://www.newegg.com/p/N82E16820147851
9. Samsung PM893 1.9 TB Specs - SSD Database - TechPowerUp, accessed November 23, 2025, https://www.techpowerup.com/ssd-specs/samsung-pm893-1-9-tb.d1863
10. Data center SSD | SATA SSD | Samsung Semiconductor Global, accessed November 23, 2025, https://semiconductor.samsung.com/ssd/datacenter-ssd/
11. RTX A4000 GPU Pricing & Specs (Compare 11+ Providers) | ComputePrices.com, accessed November 23, 2025, https://computeprices.com/gpus/rtxa4000
12. (PDF) Universal Spectral Tokenization via Self-Supervised Panchromatic Representation Learning - ResearchGate, accessed November 23, 2025, https://www.researchgate.net/publication/396747830_Universal_Spectral_Tokenization_via_Self-Supervised_Panchromatic_Representation_Learning
13. Navigating the High Cost of AI Compute | Andreessen Horowitz, accessed November 23, 2025, https://a16z.com/navigating-the-high-cost-of-ai-compute/
14. [2510.17959] Universal Spectral Tokenization via Self-Supervised Panchromatic Representation Learning - arXiv, accessed November 23, 2025, https://www.arxiv.org/abs/2510.17959
15. [2310.03024] AstroCLIP: A Cross-Modal Foundation Model for Galaxies - arXiv, accessed November 23, 2025, https://arxiv.org/abs/2310.03024
16. VIMOS Public Extragalactic Redshift Survey (VIPERS): galaxy segregation inside filaments at z ≈ 0.7 | Monthly Notices of the Royal Astronomical Society | Oxford Academic, accessed November 23, 2025, https://academic.oup.com/mnras/article/465/4/3817/2420717
17. probability of identifying the cosmic web environment of galaxies around clusters motivated by the Weave Wide Field Cluster Survey - Oxford Academic, accessed November 23, 2025, https://academic.oup.com/mnras/article/524/2/2148/7209904
18. Inference Performance for Data Center Deep Learning | NVIDIA Developer, accessed November 23, 2025, https://developer.nvidia.com/deep-learning-performance-training-inference/ai-inference
19. Dark Energy Spectroscopic Instrument Data Release 1 | NADC, accessed

November 23, 2025, https://nadc.china-vo.org/res/r102043/?lang=en
20. coadd-SPECTROGRAPH-TILEID-GROUPID.fits — desidatamodel 25.11.dev1530 documentation, accessed November 23, 2025, https://desidatamodel.readthedocs.io/en/latest/DESI_SPECTRO_REDUX/SPECPROD/tiles/GROUPTYPE/TILEID/GROUPID/coadd-SPECTROGRAPH-TILEID-GROUPID.html
21. Data Release 1 (DR1) - DESI, accessed November 23, 2025, https://data.desi.lbl.gov/doc/releases/dr1/
22. DESI (Dark Energy Spectroscopic Instrument) - Astro Data Lab - NOIRLab, accessed November 23, 2025, https://datalab.noirlab.edu/data/desi
23. Universal Spectral Tokenization via Self-Supervised Panchromatic Representation Learning - arXiv, accessed November 23, 2025, https://www.arxiv.org/pdf/2510.17959v2
24. Universal Spectral Tokenization via Self-Supervised Panchromatic Representation Learning, accessed November 23, 2025, https://arxiv.org/html/2510.17959v1
25. NVIDIA RTX A4000 BERT Large Fine Tuning Benchmarks in TensorFlow - Exxact Corp., accessed November 23, 2025, https://www.exxactcorp.com/blog/Benchmarks/nvidia-rtx-a4000-bert-large-fine-tuning-benchmarks-in-tensorflow
26. AstroCLIP Update: A Cross-Modal Foundation Model for Galaxies - Polymathic AI, accessed November 23, 2025, https://polymathic-ai.org/blog/astroclip_update/
27. AstroCLIP: A Cross-Modal Foundation Model for Galaxies - arXiv, accessed November 23, 2025, https://arxiv.org/html/2310.03024v2
28. AstroCLIP: a cross-modal foundation model for galaxies | by Eleventh Hour Enthusiast, accessed November 23, 2025, https://medium.com/@EleventhHourEnthusiast/astroclip-a-cross-modal-foundation-model-for-galaxies-529105285e33
29. Data Release Description - DESI Legacy Imaging Surveys, accessed November 23, 2025, https://www.legacysurvey.org/dr9/description/
30. zarr slower than npy, hdf5 etc? · Issue #519 · zarr-developers/zarr-python - GitHub, accessed November 23, 2025, https://github.com/zarr-developers/zarr-python/issues/519
31. Galaxy And Mass Assembly (GAMA): end of survey report and data release 2 | Monthly Notices of the Royal Astronomical Society | Oxford Academic, accessed November 23, 2025, https://academic.oup.com/mnras/article/452/2/2087/1069711
32. Filaments in VIPERS: galaxy quenching in the infalling regions of groups - CONICET, accessed November 23, 2025, https://ri.conicet.gov.ar/bitstream/handle/11336/123821/CONICET_Digital_Nro.2821d704-4439-4469-ab08-2442eb973a36_A.pdf?sequence=2
33. effect of cosmic web filaments on galaxy evolution | Monthly Notices of the Royal Astronomical Society | Oxford Academic, accessed November 23, 2025, https://academic.oup.com/mnras/article/534/3/1682/7756891
34. [1710.02676] Galaxy evolution in the metric of the Cosmic Web - arXiv, accessed November 23, 2025, https://arxiv.org/abs/1710.02676
35. (PDF) The effect of cosmic web filaments on galaxy evolution - ResearchGate,

accessed November 23, 2025, https://www.researchgate.net/publication/384057474_The_effect_of_cosmic_web_filaments_on_galaxy_evolution
36. *Otherwise, HDF5 offers every single advantage that zarray has and is much more ... | Hacker News, accessed November 23, 2025, https://news.ycombinator.com/item?id=22426791
37. How many files does zarr generate? - Stack Overflow, accessed November 23, 2025, https://stackoverflow.com/questions/55656962/how-many-files-does-zarr-generate
38. storage size for zarr file on disk · Issue #86 · zarr-developers/zarr-python - GitHub, accessed November 23, 2025, https://github.com/zarr-developers/zarr/issues/86