

Multi-Cloud Orchestration for Bioinformatic Pipelines

A Unified, Automated Framework for Integrating Technological Advancements in Sequencing and Analysis

Srithan Chowdarapu¹, Kirthi Moulika¹, Pratibha Chitipothu¹, Ben Habermeyer²

1 - Strand Life Sciences, Bengaluru, India

2 - Myriad Genetics, Salt Lake City, Utah, United States

Myriad genetics®

strand



Contact

Ben Habermeyer
Staff Software Engineer, Myriad

ben.habermeyer@myriad.com

Contact

Pratibha Chitipothu
Senior Director - Software

pratibha@strandls.com

Introduction

Bioinformatic analysis increasingly spans multiple cloud and sequencing platforms. Each platform offers unique strengths—from scalable storage to high-performance secondary analysis and compliant long-term archiving.

Our orchestration solution integrates **AWS S3**, **Illumina Connected Analytics (ICA)**, and **DNAexus** into a seamless data-flow framework that automates sequencing data movement, metadata management, and workflow execution across environments.

All orchestration is automated using AWS Step Functions, for robust and highly-scalable execution of pipeline components.

End-to-End Automation

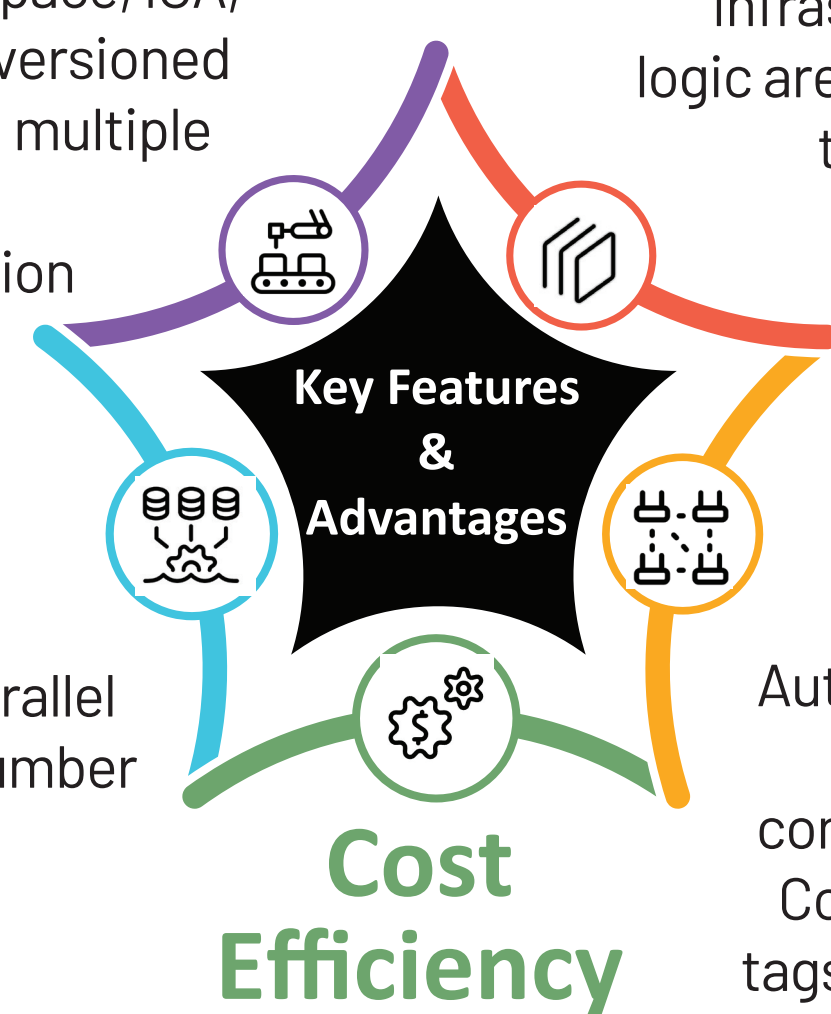
AWS Step Functions coordinate transitions across BaseSpace, ICA, S3, and DNAexus using versioned state machines, enabling multiple active pipeline versions under a single orchestration framework.

Compliance & Reproducibility

Infrastructure and pipeline logic are defined as versioned templates, providing a complete audit trail of all data transitions across systems.

Scalability

Platform-agnostic orchestration enables parallel processing across any number of samples or flow cells.



Reliability & Robustness

Automated retries, error categorisation, and controlled failure points. Consistent data lineage tags across all platforms.

Offloads expensive compute from DNAexus to ICA's DRAGEN hardware.

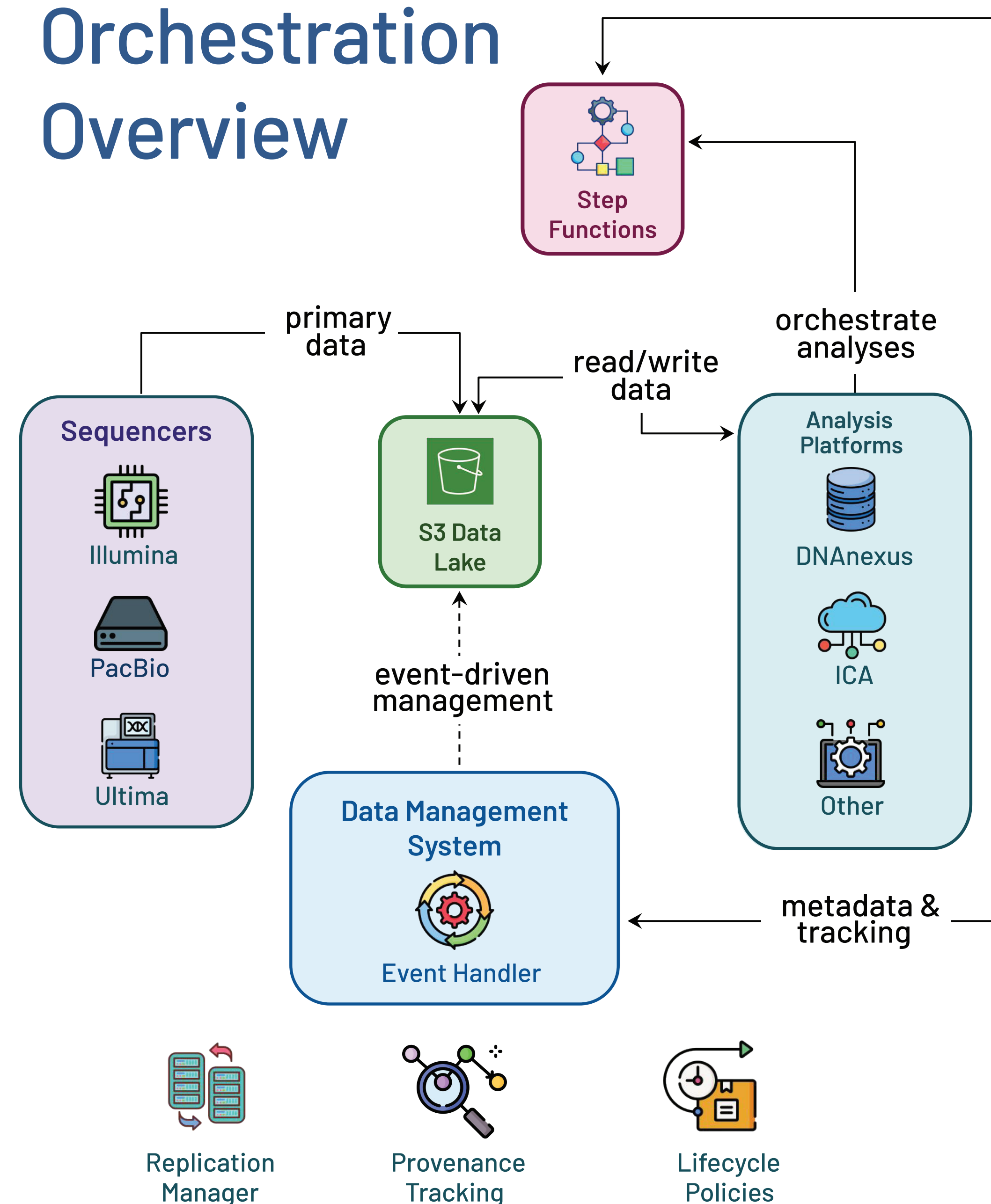
Motivation

- **Legacy approach:**
 - CPU-based analysis on DNAexus
 - High per-sample cost
 - Slow data transfers and primary analyses

Modern multi-cloud architecture:

- Labs are transitioning from targeted assays to Whole Genome Sequencing (WGS) on high-throughput next-gen sequencers (eg., NovaSeq X, Ultima), driving ~100x increases in data and compute requirements.
- WGS workflows demand accelerated compute, scalable storage, and robust automation to maintain a feasible COGS.
- **ICA** provides vendored, accelerated tools like **BCL Convert** and **DRAGEN alignment**, delivering major cost and speed advantages over CPU-based approaches.
- **DNAexus** serves as a flexible development platform for complex, custom downstream pipelines built and maintained by the computational biologists.
- Combining **S3**, **BaseSpace**, **ICA**, and **DNAexus** platforms leverages each of their strengths for cost-efficient compute and scalable automation.

Multi-Cloud Orchestration Overview



Summary

- This multi-cloud orchestration framework enables each platform to operate where it is strongest—**S3** for shared storage, **ICA** for accelerated demultiplexing and alignment, and **DNAexus** for flexible downstream analysis
- BCL Convert and DRAGEN on ICA gives **90% cheaper** and **50% faster** results, which reduces the overall cost by approximately 70%
- The framework is extensible by design, allowing new sequencers and analytics to be added without modifying the orchestration layer
- This approach maintains unified automation, reproducibility, and regulatory readiness, providing a fast, cost-effective, and fully traceable multi-cloud architecture for large-scale genomic data

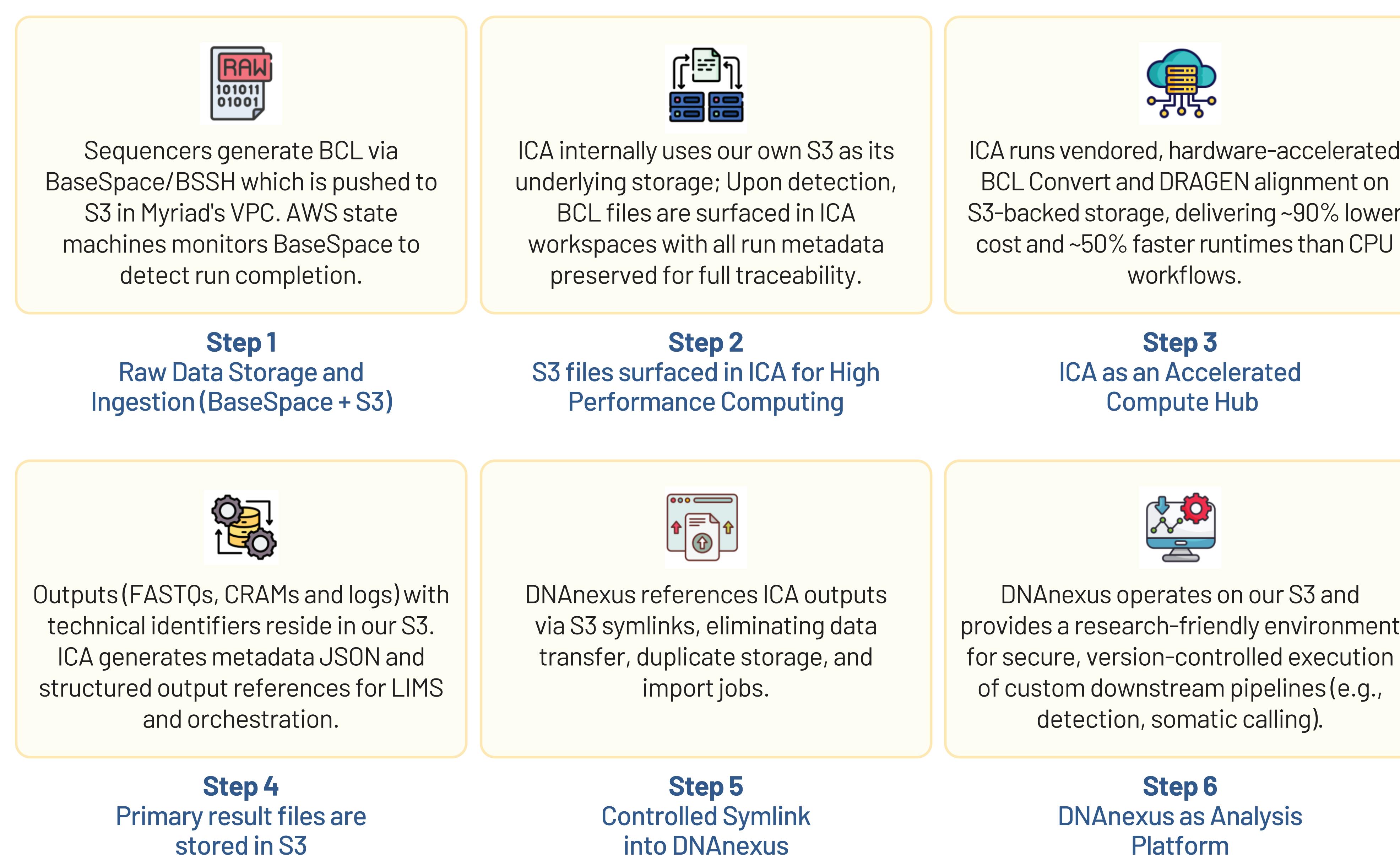
What is Multi-Cloud Orchestration?

A coordinated, automated system that monitors sequencing runs and executes analysis workflows across cloud environments. It manages data discovery, metadata tagging, cross-platform

registration, and status monitoring with robust error-handling and lineage tracking to ensure compliance.

Case study: How WGS data moves across S3, ICA, and DNAexus

Each stage below is executed and monitored using AWS Step Functions for consistency, retries, and automation.



If you're looking for similar software solutions, you can reach out to Strand here →

