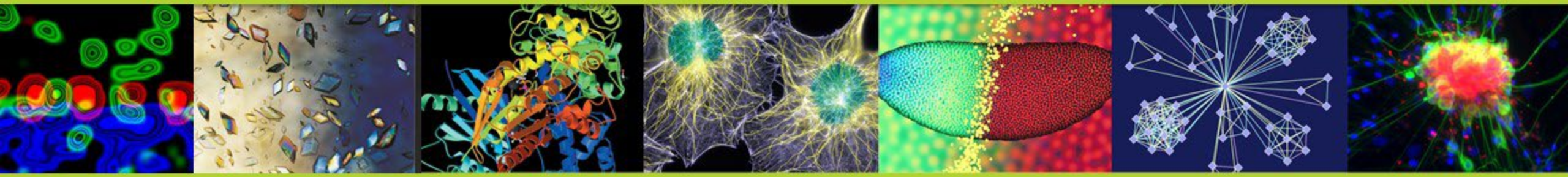
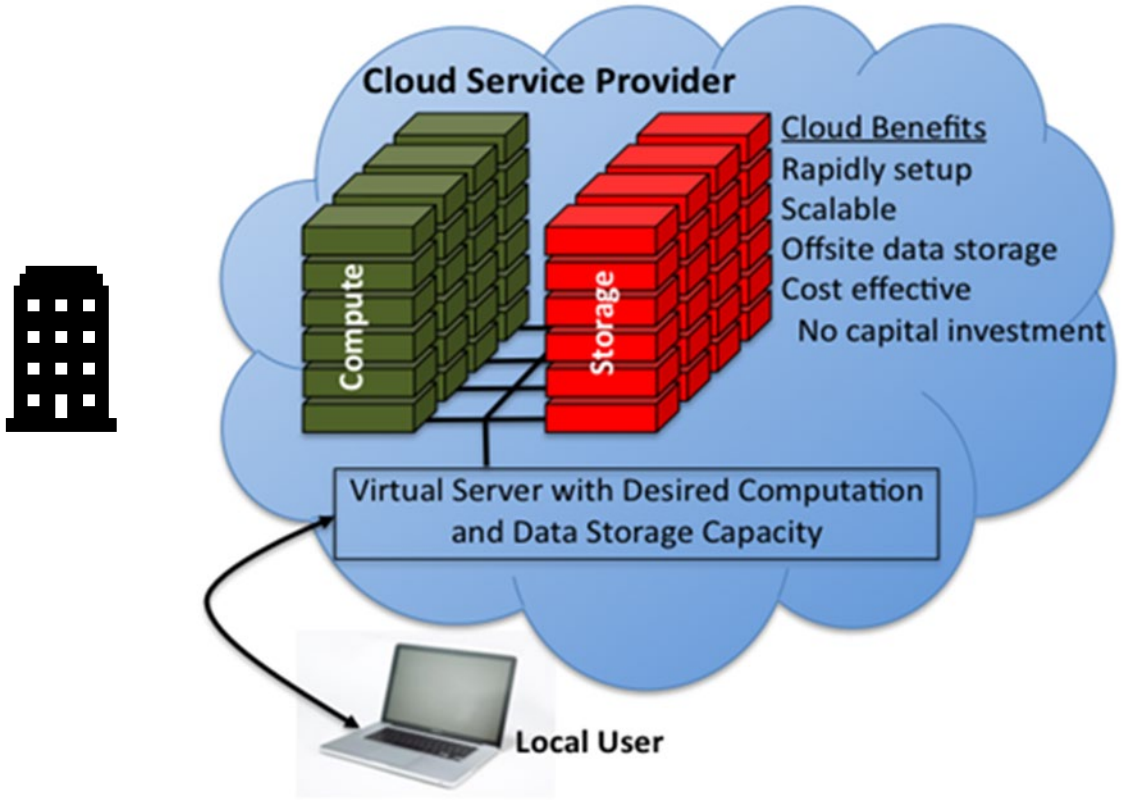
National Institute of General Medical Sciences

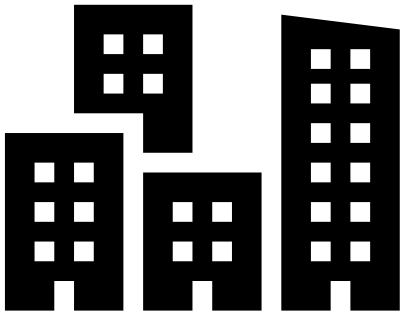
Biomedical Research Cloud-based Learning Sandbox

Lakshmi Kumar Matukumalli



Why Cloud Computing

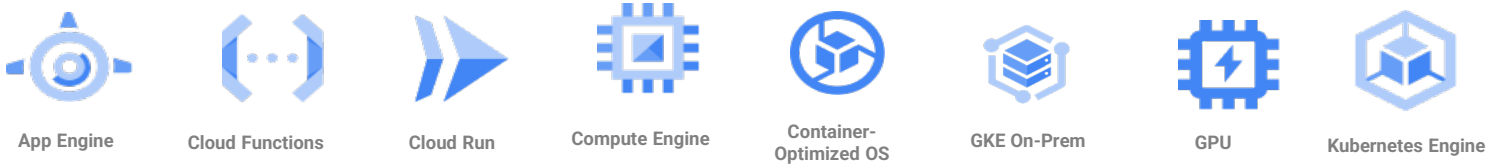




- Common Computing tools
- Hybrid Computing
- Large Biomedical Research Datasets are in the cloud

Cloud Features

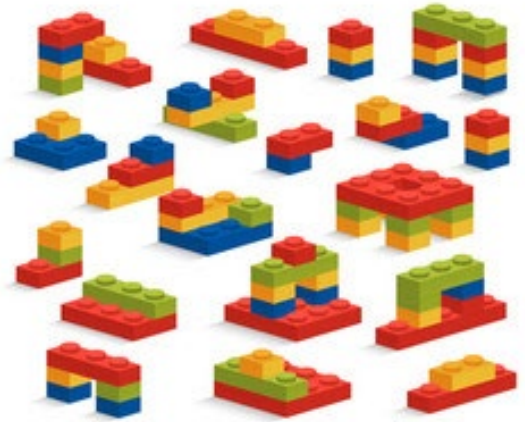
Compute



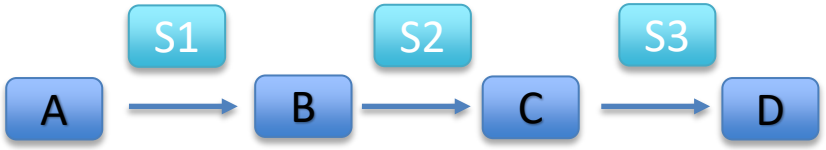
Data Analytics

AI and Machine Learning

Cloud Applications



docker



NIH National Heart, Lung, and Blood Institute | BioData CATALYST

AnVIL

NIH NATIONAL CANCER INSTITUTE Genomic Data Commons

NIH National Institute of General Medical Sciences

Cloud Capacity Building

Tools/Apps

NIH STRIDES
Cloud Account Setup and Discounts

Cloud Platform / Cloud Infrastructure

Amazon

https://bit.ly/2XgSCrD

- AWS Tech Essentials
- Practical Data Science with Amazon SageMaker
- Running Containers on Amazon Elastic Kubernetes Service
- Migrating to AWS

Google

- GCP Fundamentals: Big Data & ML
- Architecting with Google Cloud: Design and Process
- Getting Started with Google Kubernetes Engine (GKE)
- Architecting Hybrid Cloud Infrastructure with Anthos

Amazon Web Services (AWS) offering cloud credits for

- Cloud Credit for Research (CCR) - $250,000
- Diagnostic Development Initiative (DDI) - $250,000

NIGMS – ODSS partnership for supporting funding Data Science in IDeA states – COBRE – Phase I

NIGMS Pilots: Biomedical Research Cloud-Based Learning Tools

Projects:
- RNA-Seq Data Analysis – Maine INBRE
- Proteomics Data Analysis – Univ of Arkansas Medical Center

Cloud Professional Services
- Google professional Services provided technical expertise
- NIH/CIT acted as a liaison, ODSS provided cofounding.

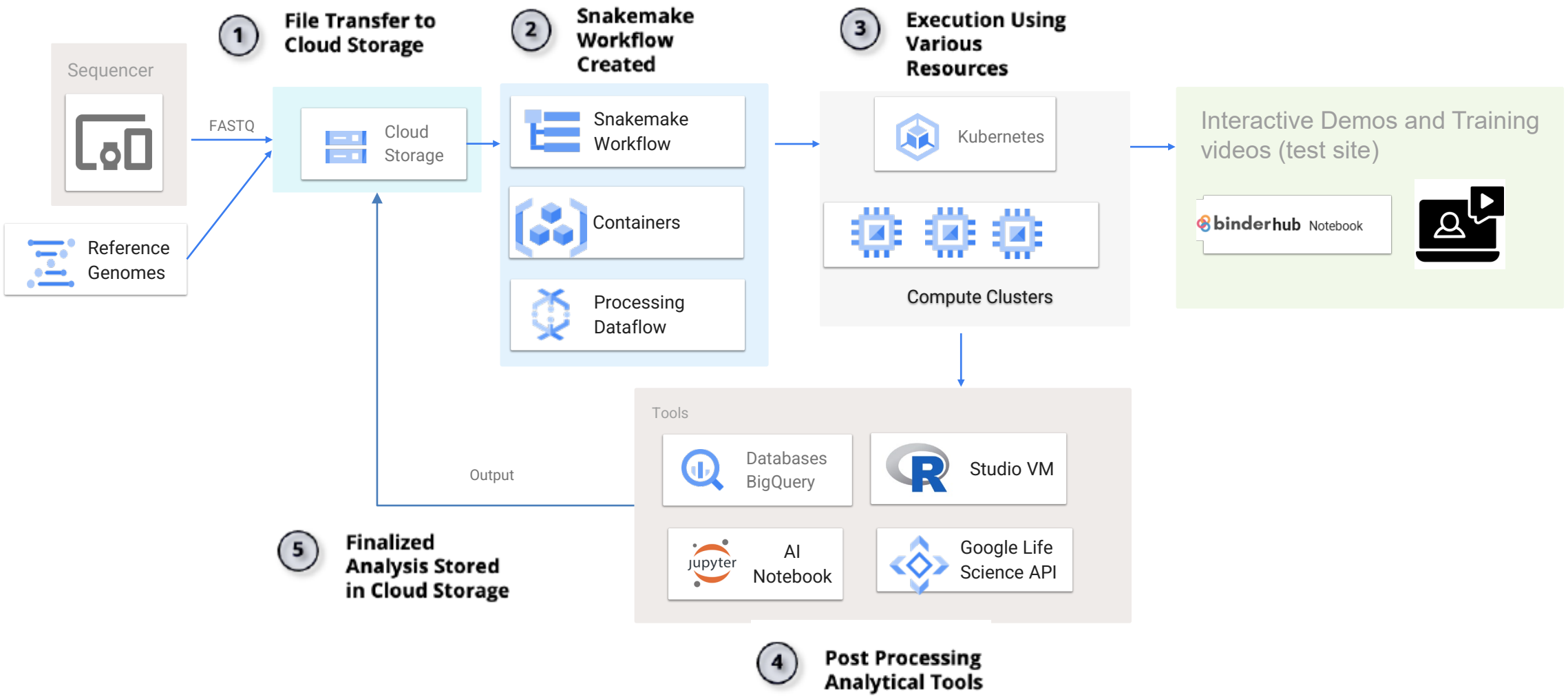
Train the Trainer
Google professional Services trained researchers from Maine and Arkansas to help develop Cloud-based workflows

Student Self Learning Modules
Maine INBRE and UAMS researchers are developing learning modules that includes videos, and interactive demos

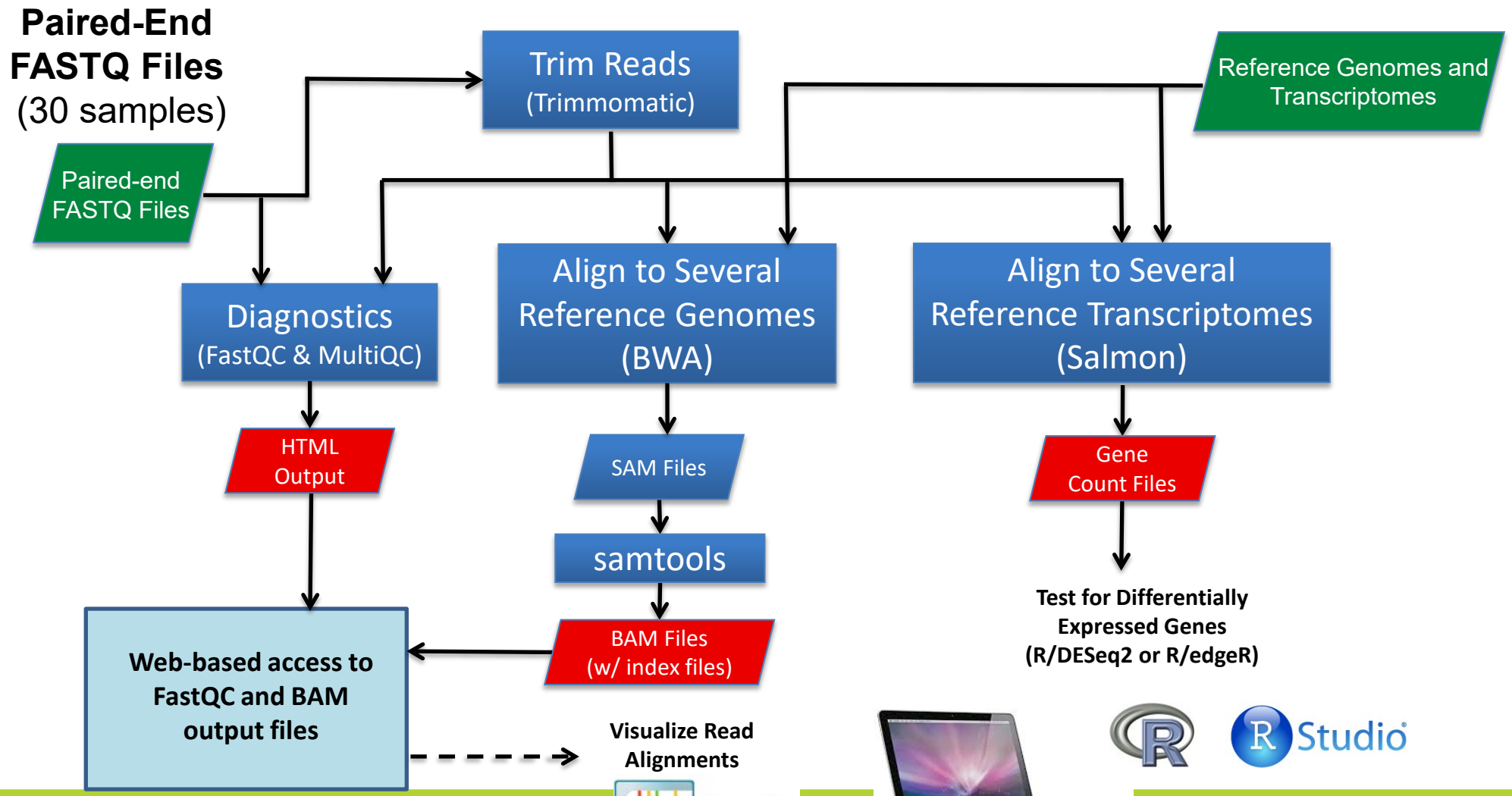
Cloud Infrastructure
MDIBL, Univ of Maine, and UAMS established cloud accounts and participating in STRIDES Discounts – University IT services

Maine INBRE - RNA-Seq Workflow

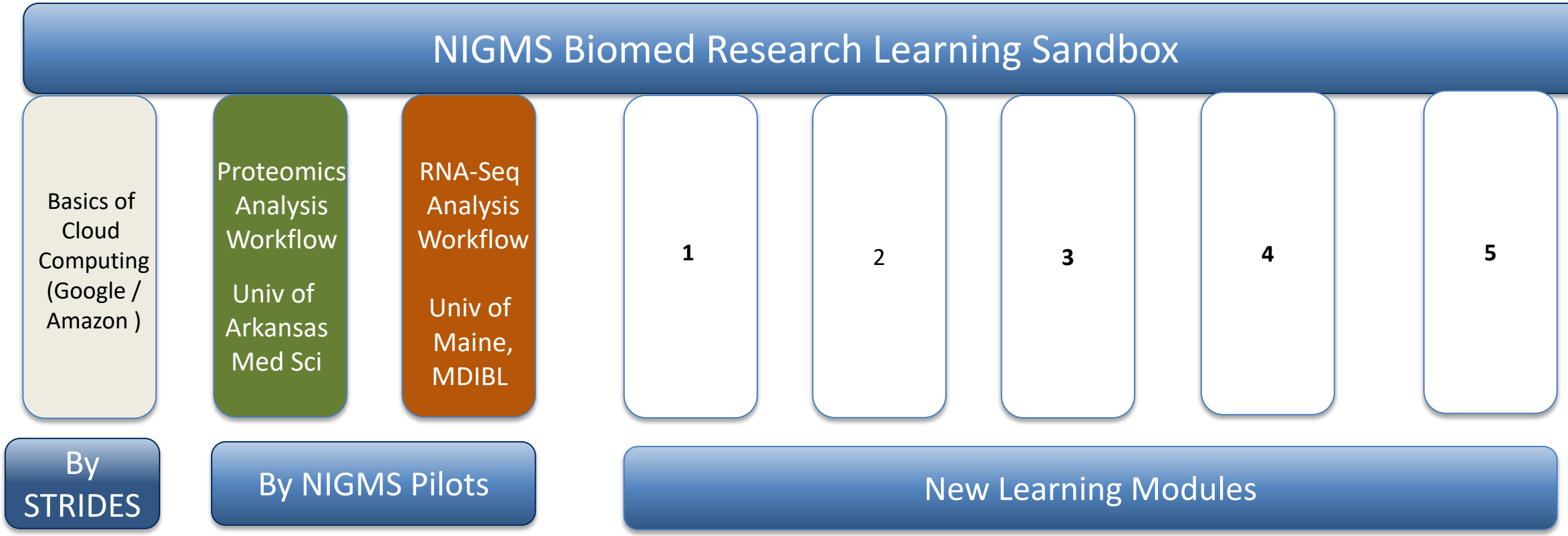
Google Cloud



RNA-Seq Workflow: Gene Expression Study of a Phage that Alters Drug Resistance in Pathogenic Mycobacteria



An NIGMS "Sandbox" with Multiple Learning Modules



Examples: Image Analysis, Metagenomics, CRISPR, GWAS, Phylogeny …

Strategic Inputs – Training and Capacity Building

- Status of currently available training for faculty and students
- Training material under NIGMS Sandbox
  - Utility, Features, Quality, Topic areas
- Integration with other NIH Cloud activities
  - Cloud Labs, AnVIL, BioData Catalyst, Cancer Genomics Commons
- Cloud Infrastructure
  - Institutional, Regional
  - Administration