



Post Exascale: Workloads and Emerging Trends

Heidi Poxon

ModSim 2025

History of Change in Supercomputing



CDC 6600

Custom Silicon



Cray-1

Vector Processors



MPP + Beowulf Clusters

Intel/AMD CPUs



Heterogeneous
Computing

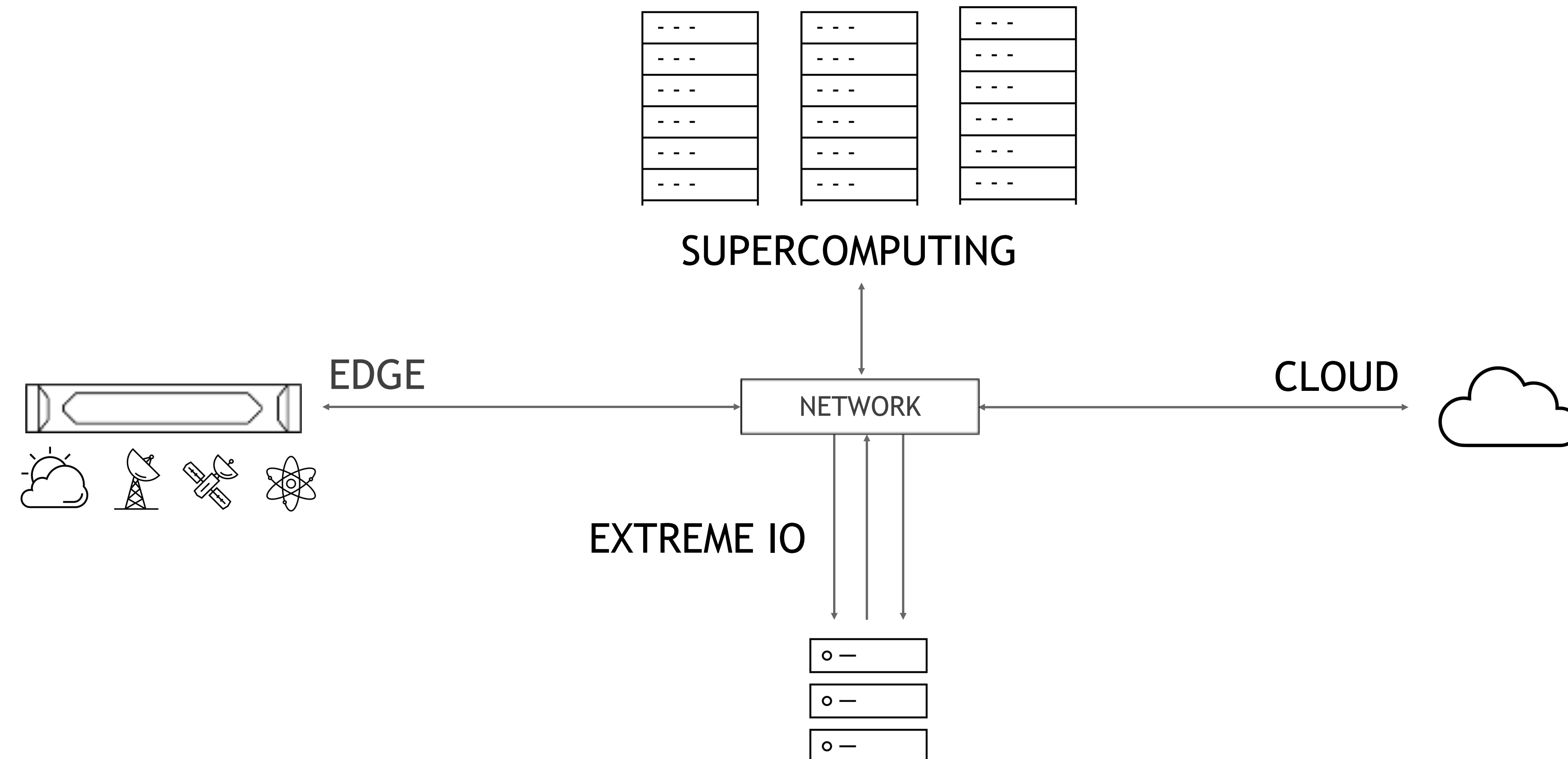
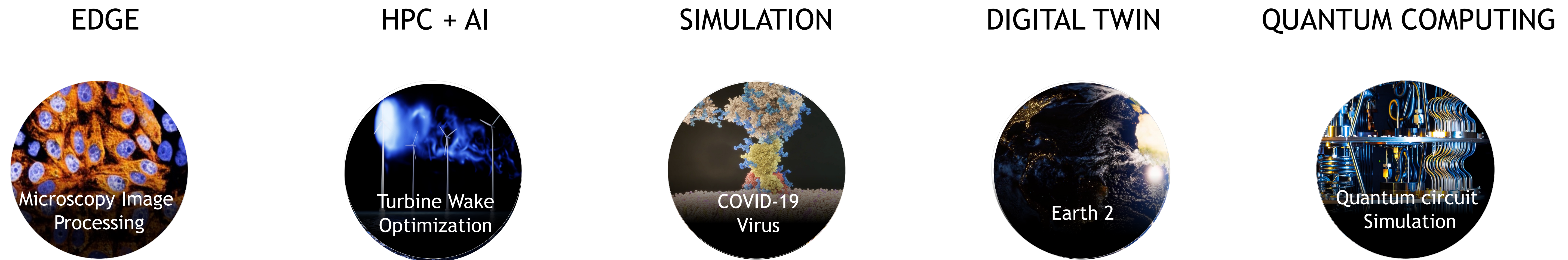
CPU + GPU



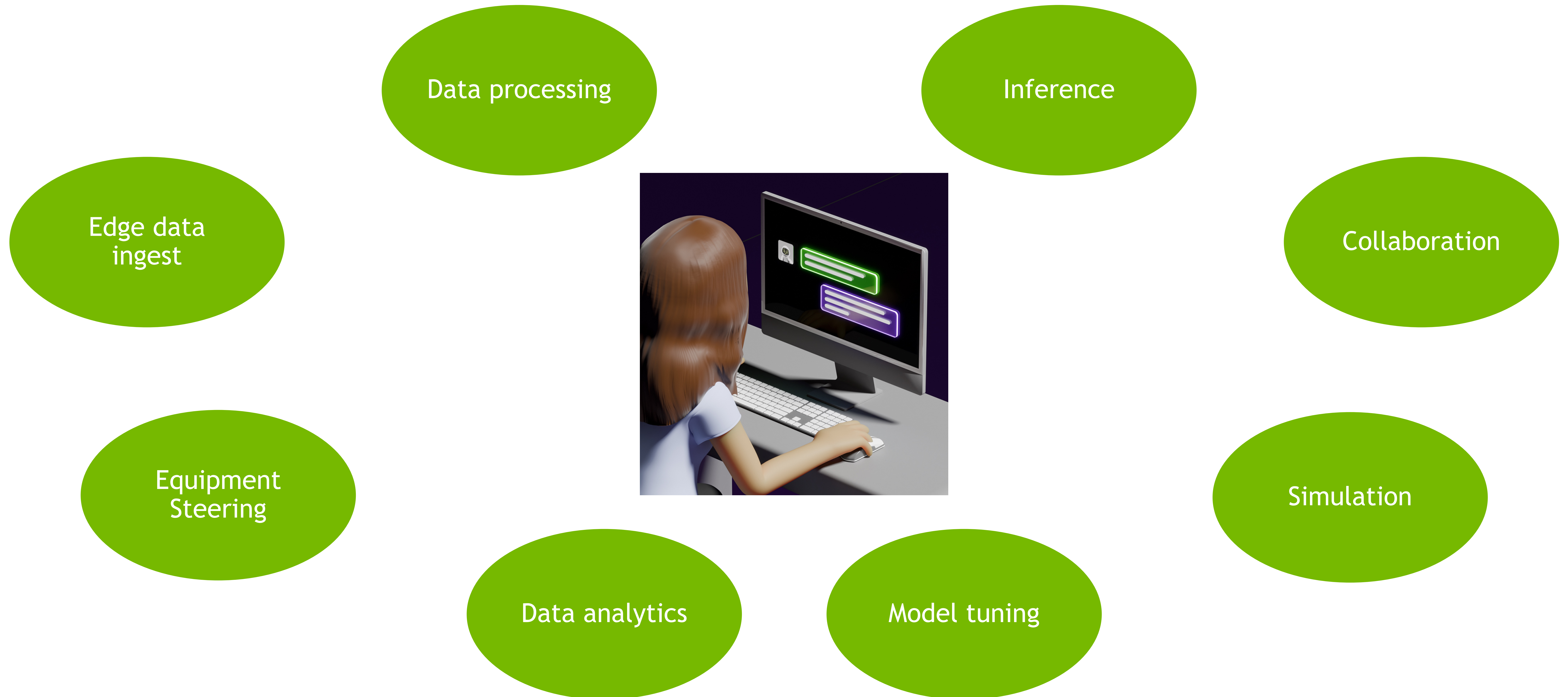
Accelerated Computing

AI + Quantum

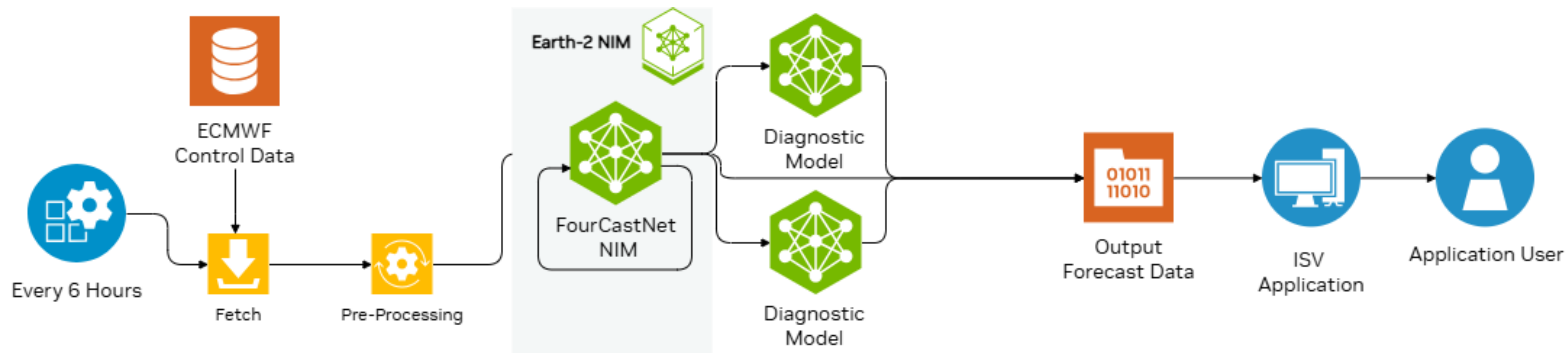
The Next Generation Supercomputing System



Augmented and New Workflows



From Workloads to Workflows

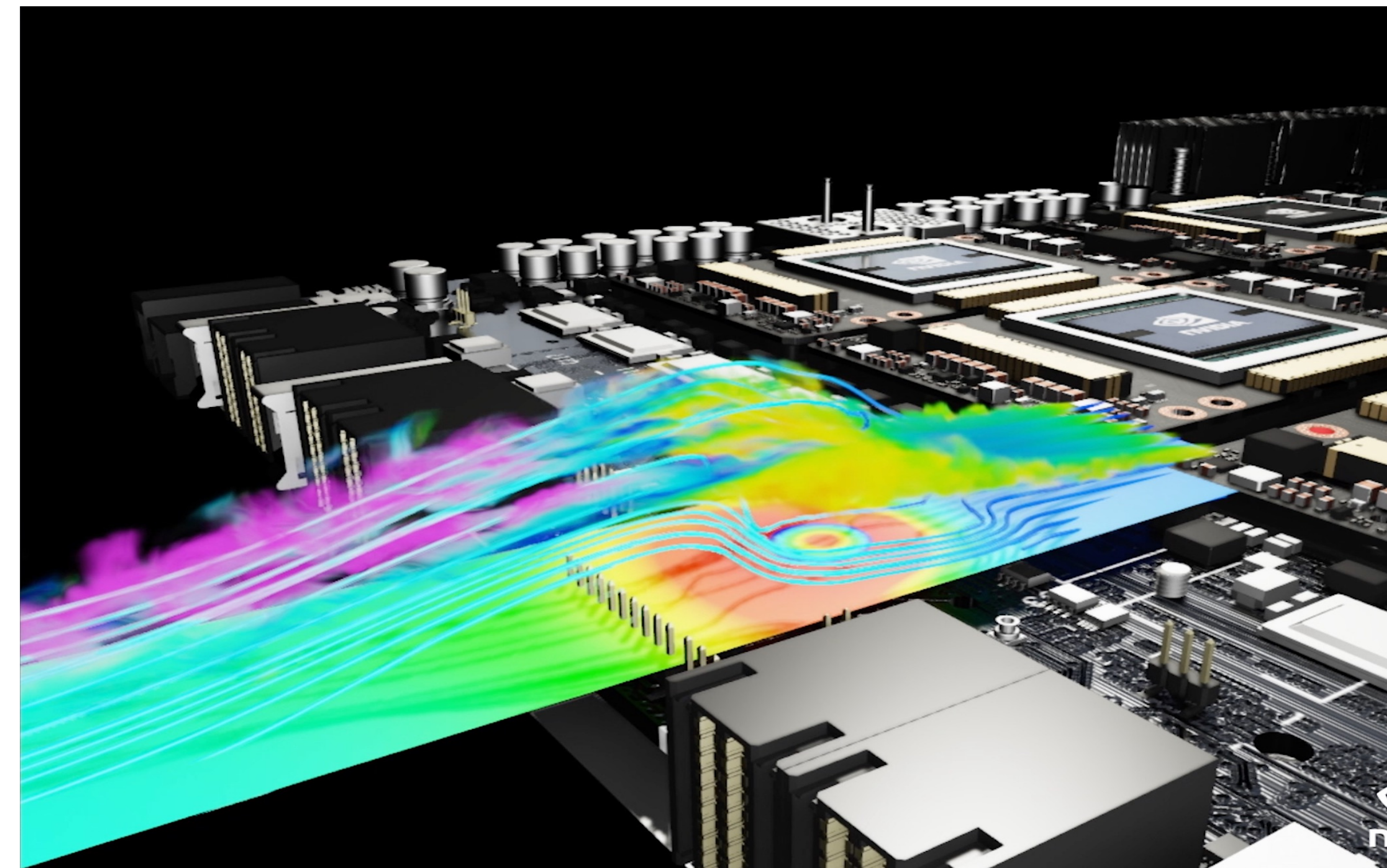


The Importance of Use Cases

How do we communicate requirements for

- system architects and designers
- modelers
- modeling tool developers

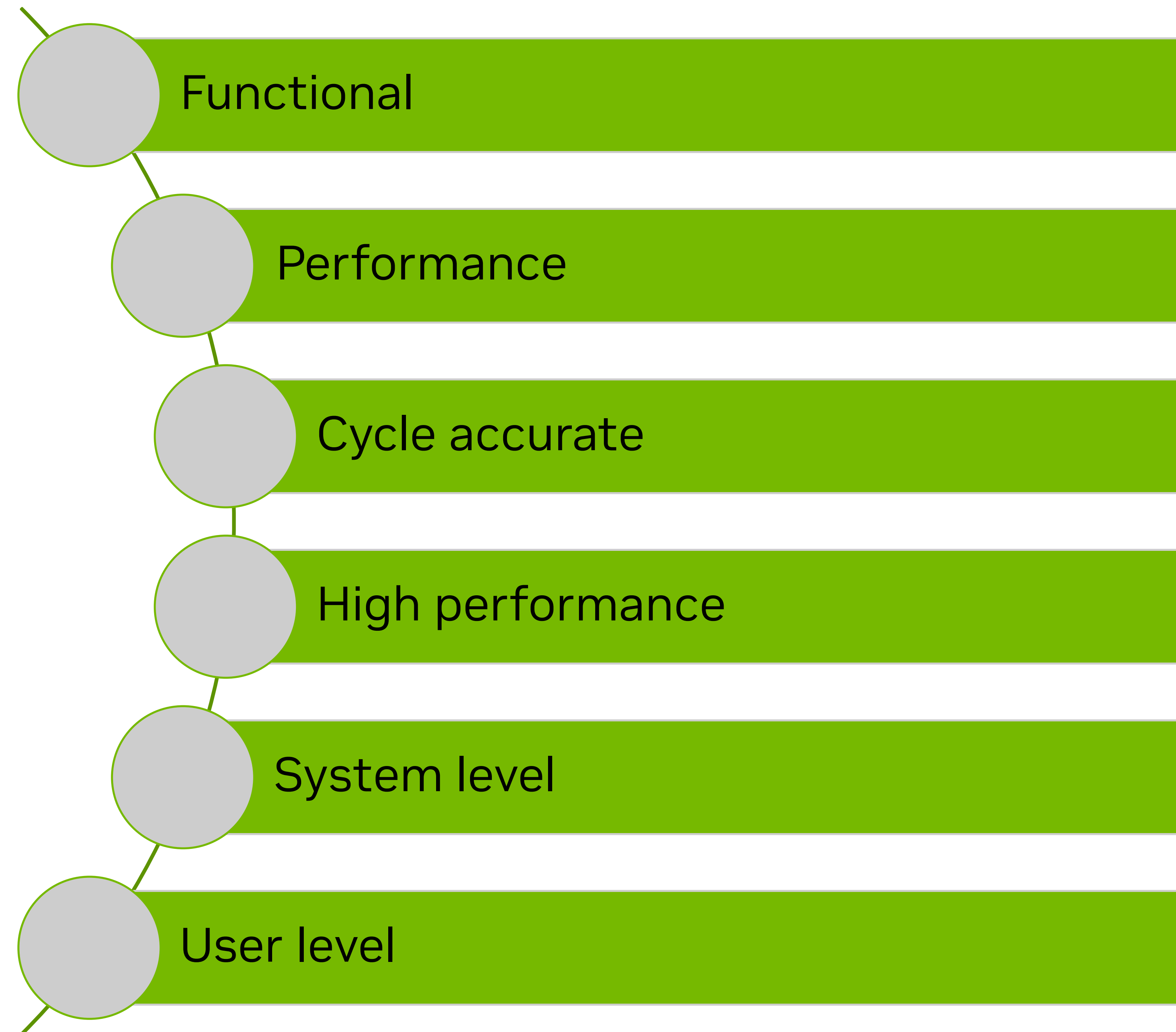
Manufacturing and Industrials Product Engineering



- Electronics and Mechanical Design
- Fluid Dynamics
- Solid Mechanics
- Chemistry Discovery
- Generative Product Engineering Workflow Automation

Modeling Options

Modeling styles



Code progression

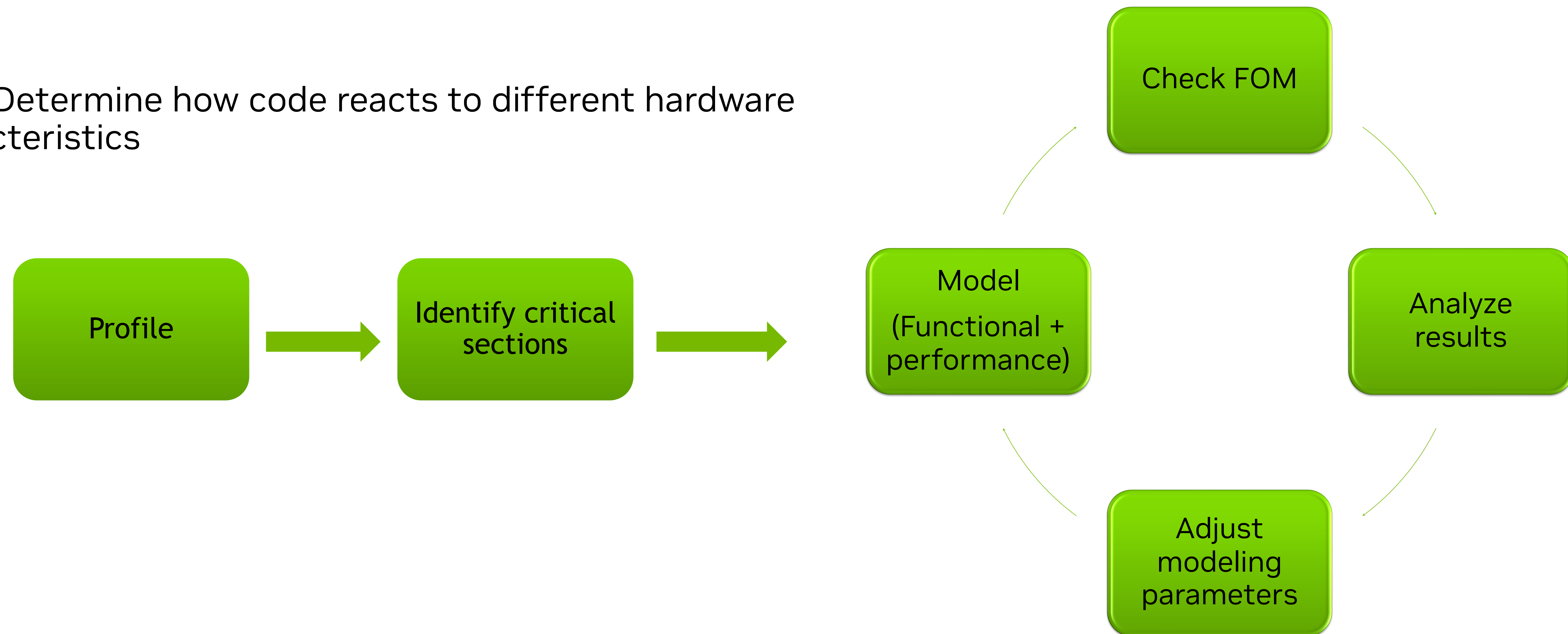
Benchmarks / diagnostics

Proxy apps

HPC applications (Gromacs,
OpenFOAM, WRF,...)

Example Performance Modeling Methodology

Goal: Determine how code reacts to different hardware characteristics



HPC Benchmarks

- NVIDIA HPC Benchmarks Collection:
<https://docs.nvidia.com/nvidia-hpc-benchmarks/>
- Grace CPU Benchmarking Guide:
<https://nvidia.github.io/grace-cpu-benchmarking-guide/index.html>



How to Use This Guide

Platform Configuration

Software Environment

Foundational Benchmarks

Fused Multiply Add

STREAM

Common Benchmarks

High Performance Linpack (HPL)

HiBench: K-means

GAP Benchmark Suite

Graph500

NAS Parallel Benchmarks



Proxy Applications for Performance Modeling

Simplified codes designed to communicate important features of larger applications

ECP Proxy Applications project: 2017-2022

<https://proxyapps.exascaleproject.org/>

Proxy App

FlexFlow – CandleUno

MLPerf-DimeNet++

miniGAN

miniRL

CRADL

MLperf-Cosmoflow

MLperf-DeepCam

Proxy App

ExaMiniMD

Quicksilver

ExaMPM

SNAP

CabanaPIC

E3SM-kernels

RIOPA

GAMESS_RI-MP2_MiniApp

HyPar

FFTX Examples

Goulash

IAMR

Proxy App

AMG

Ember

ExaMiniMD

Laghos

MACSio

miniAMR

miniQMC

miniVite

NEKbone

PICSARlite

SW4lite

SWFFT

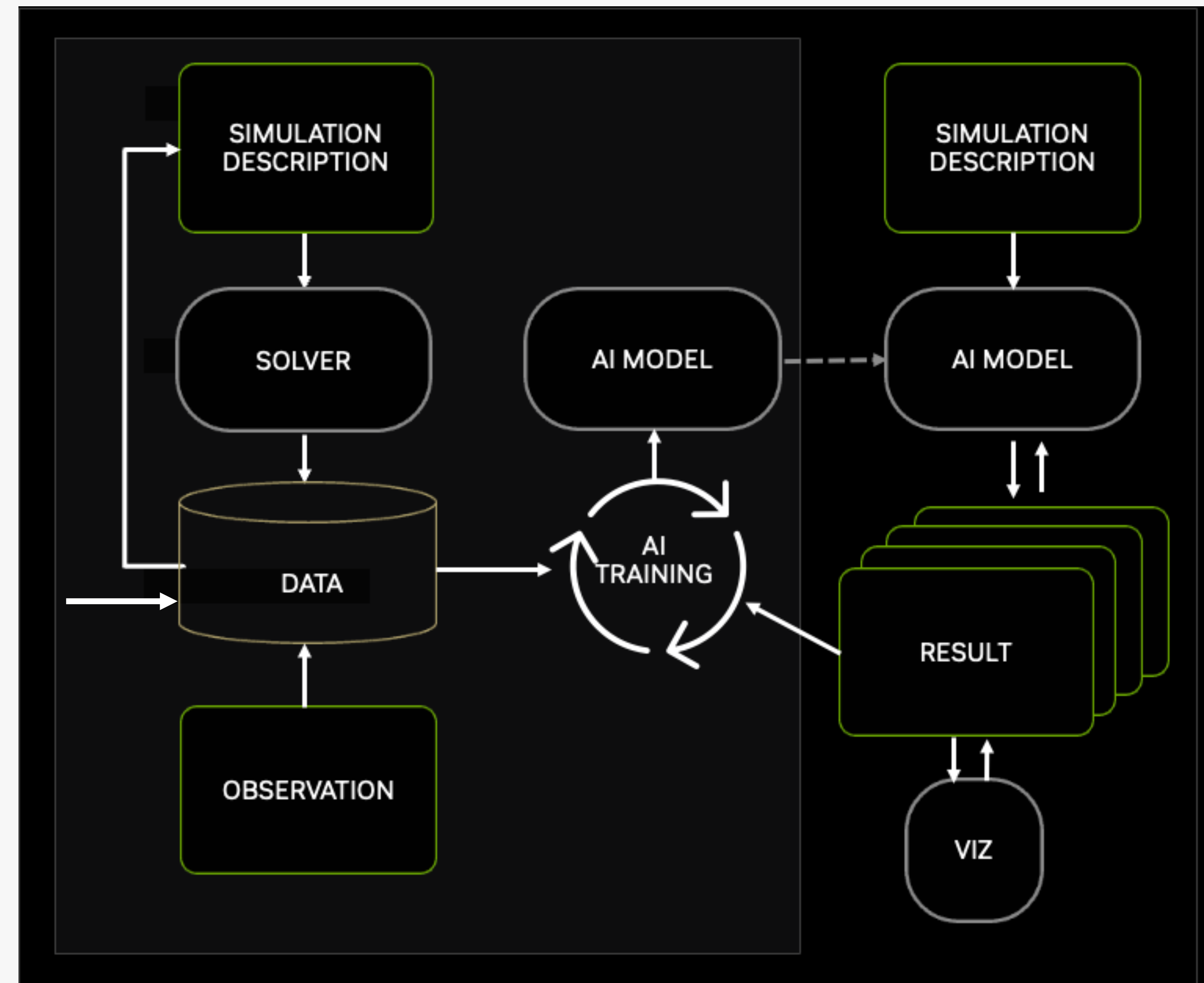
Thornado-mini

XSbench

Modeling Proxy Workflows

What should we be considering?

- Adjustable workflow with core components
- Flexibility for rapid technology changes
- Adjustable events and connectivity
- Modular
- Tolerances
- Maintainability and automation
- Assume iterative process



Some Thoughts on Next Proxies

Building use case workflows helps identify metrics

- Workflow templates or blueprints?
- Digital twins?
- Hierarchy of workflow? – layers on top of proxy-apps?
- How can AI help us with synthetic data?
- How will different modeling tools and technologies interact?



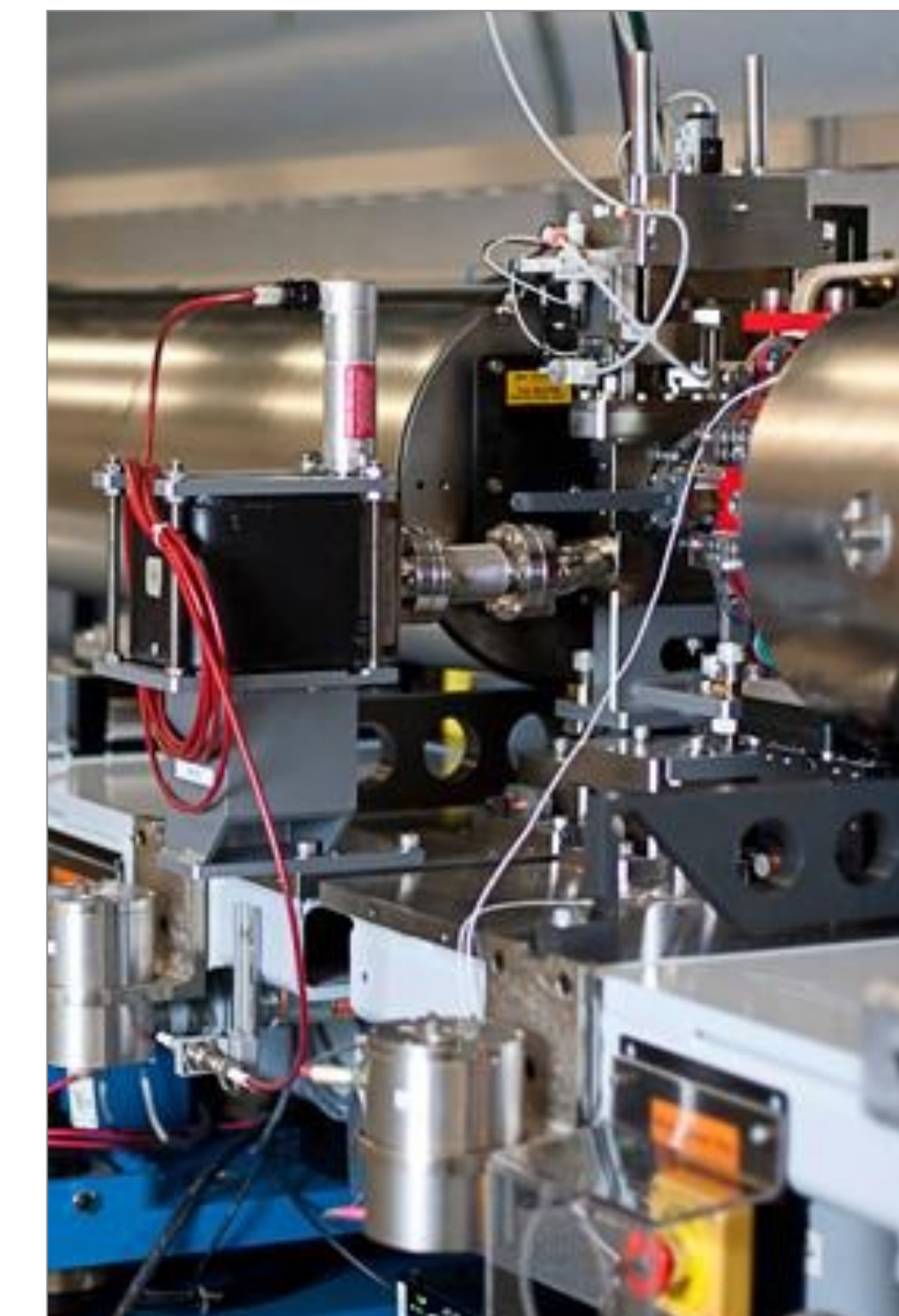
Scale out



Scale up pods



Data Integration and
Digital Twins

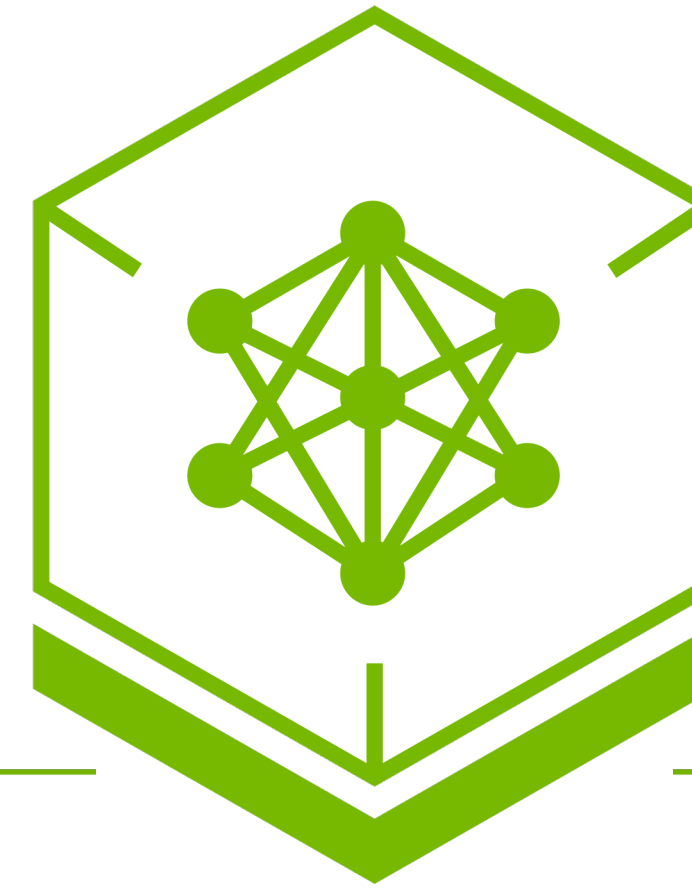


Bridge to external
instruments, digital labs

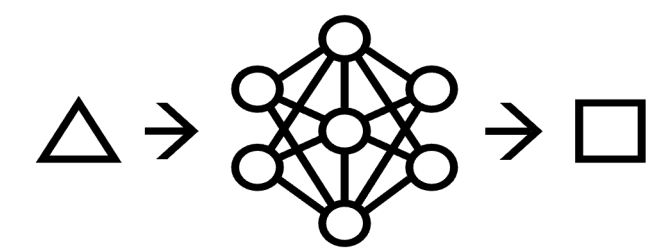
Concept of Blueprints - NVIDIA Blueprints

Available on build.nvidia.com

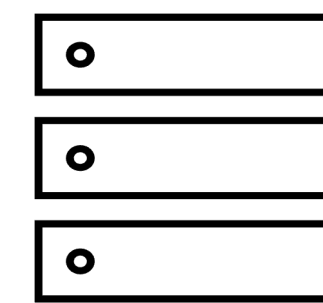
NVIDIA NIM & microservices



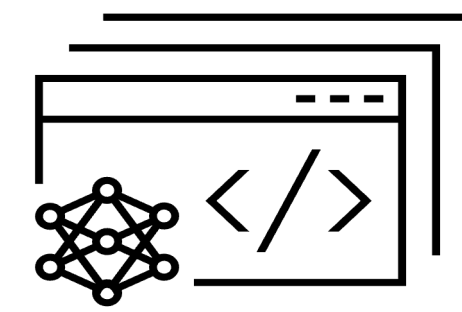
Blueprints



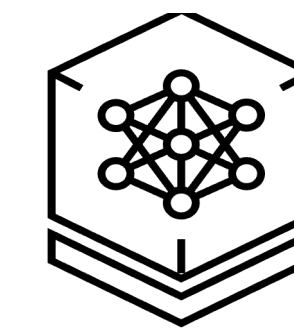
Reference Application



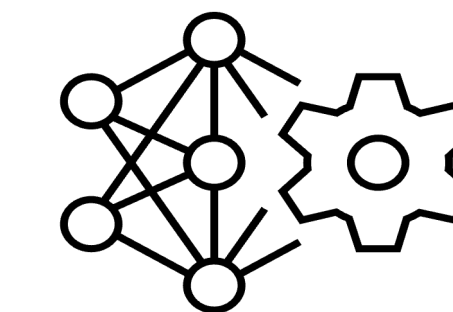
Sample Data



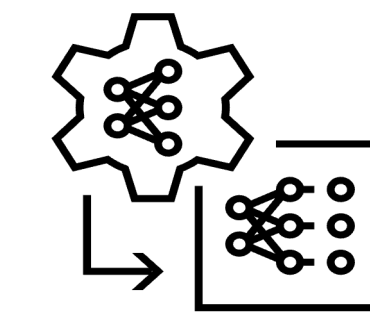
Reference Code



Architecture



Customization Tools



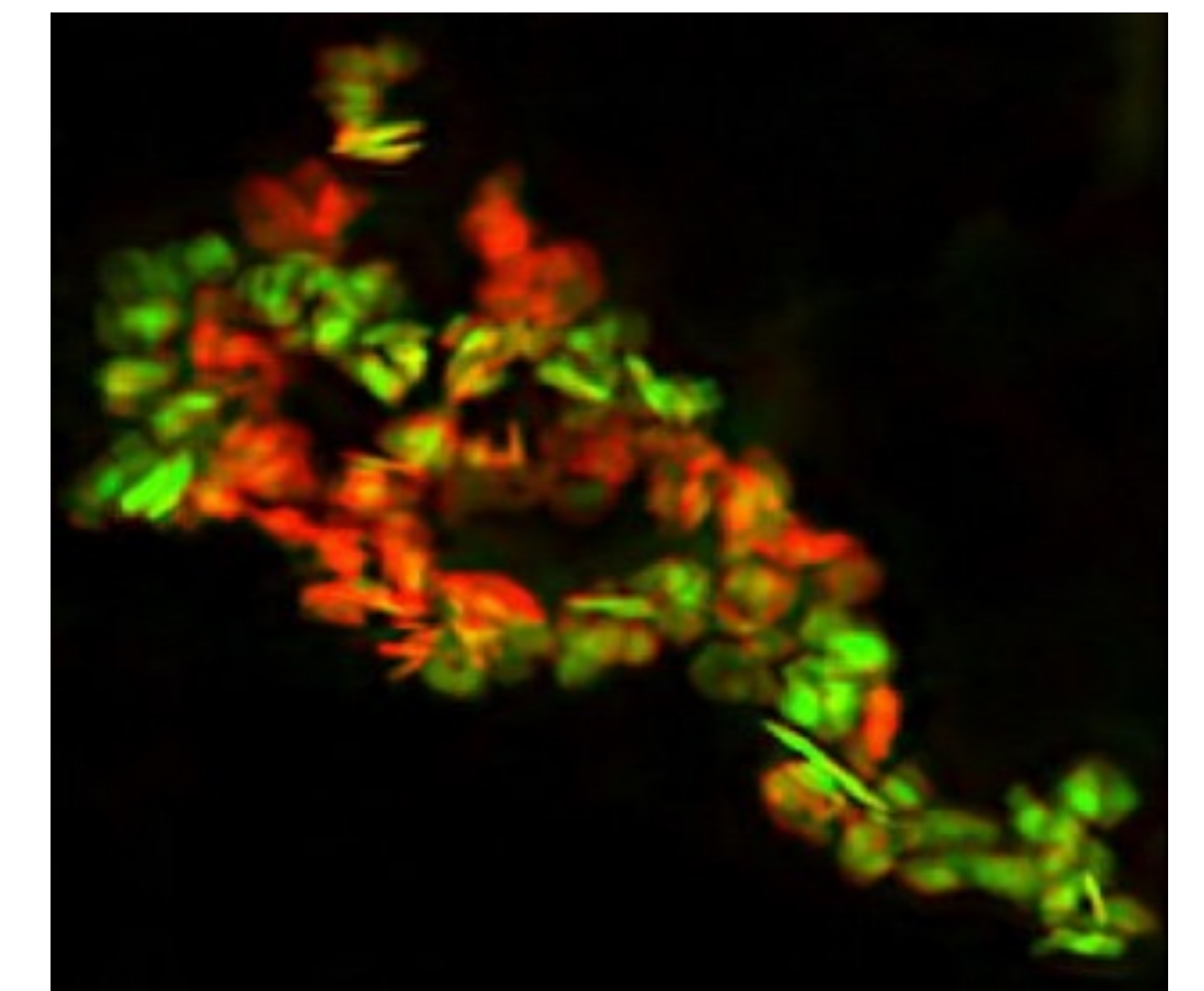
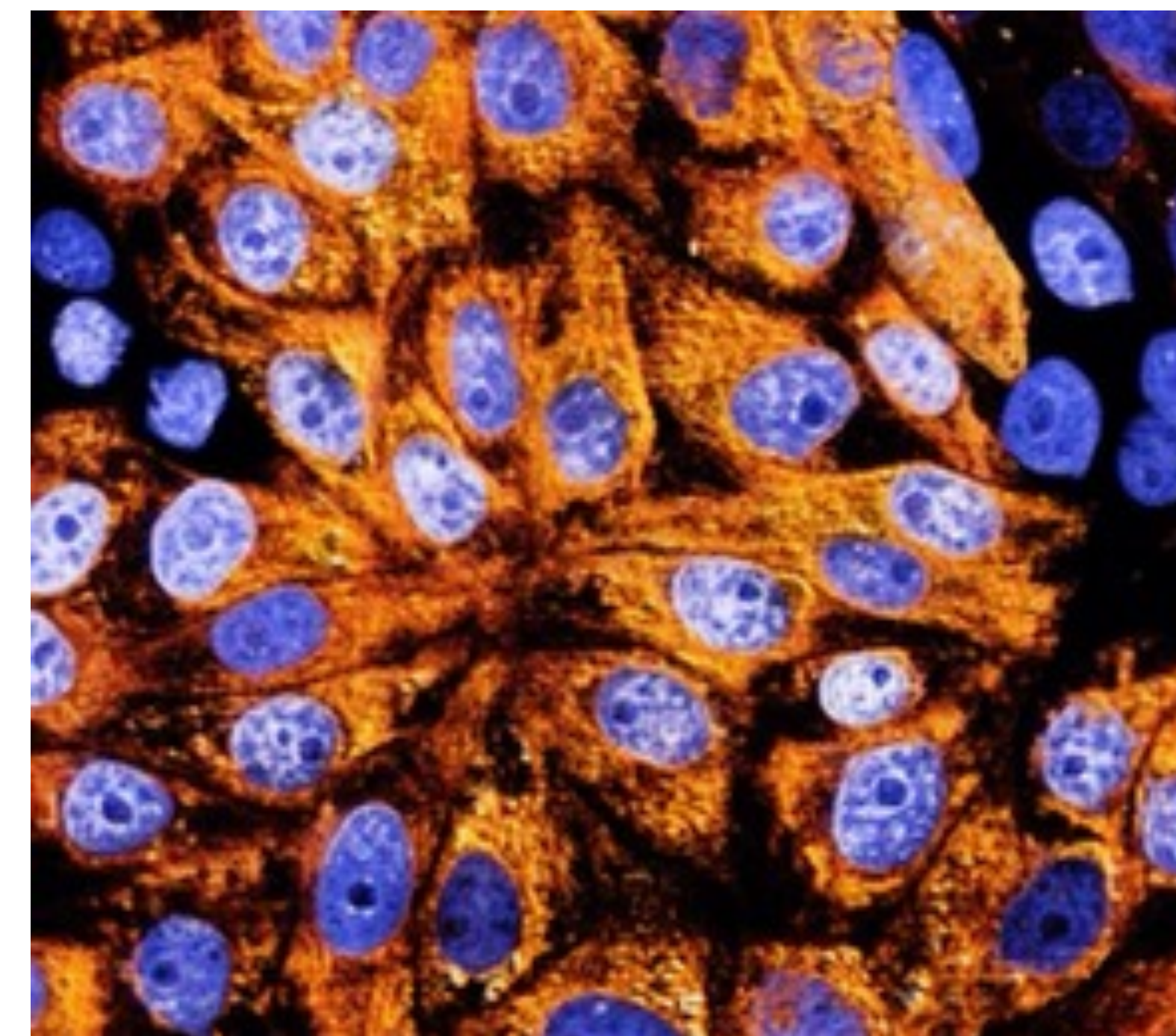
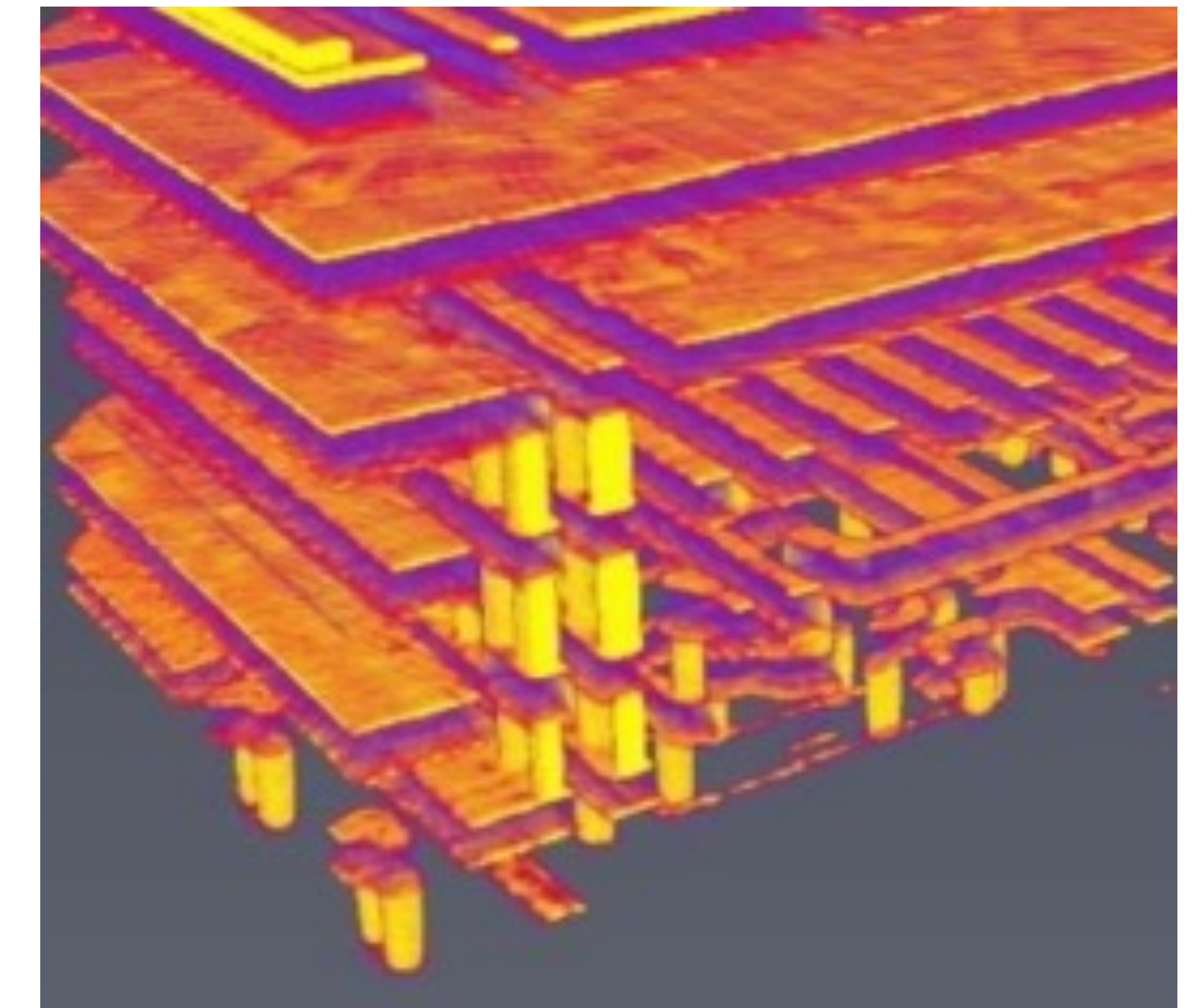
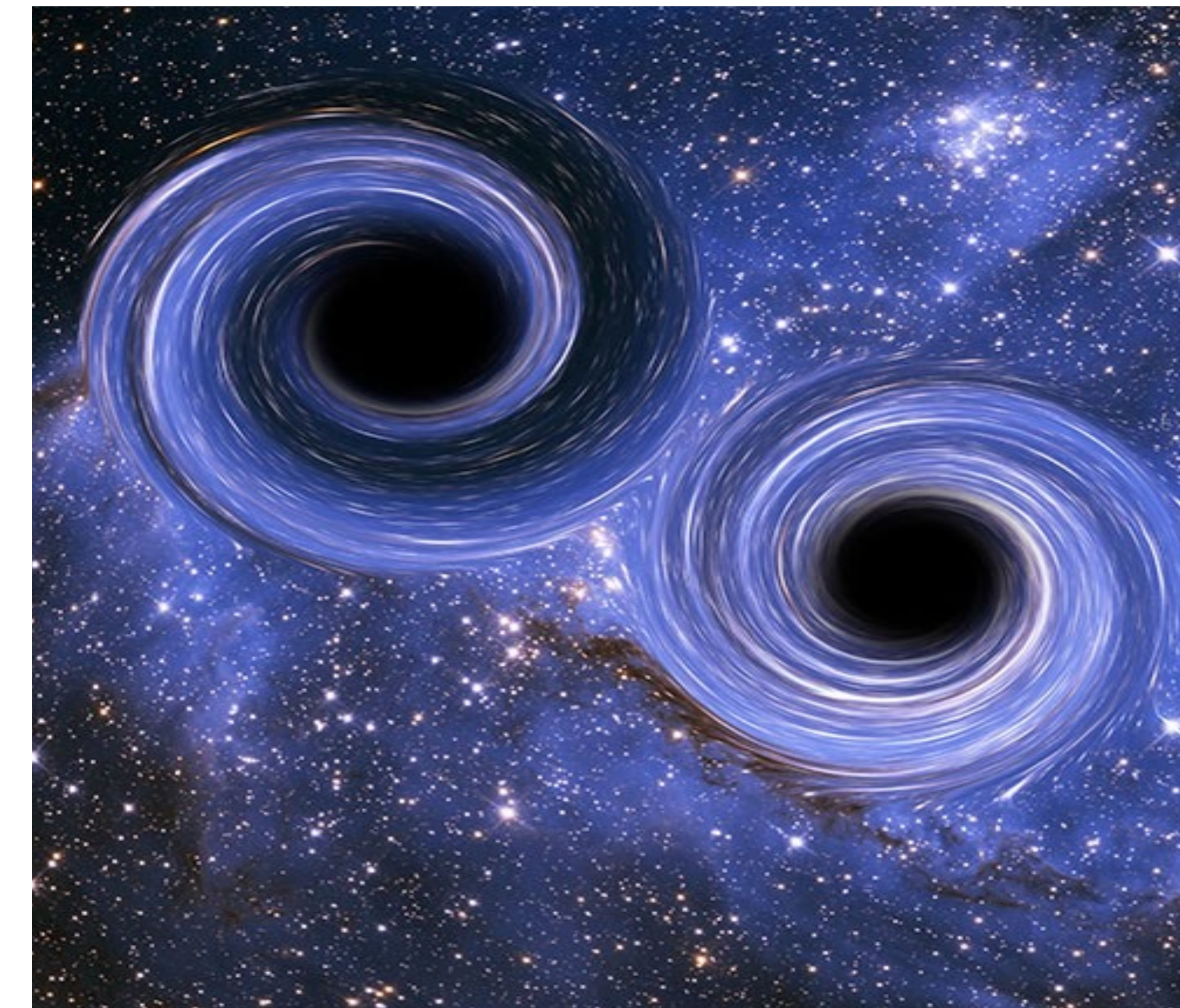
Orchestration Tools



AI Agents

Takeaways

- History of proxy apps to communicate use cases
- Proxy workflows can be used to communicate use cases and metrics to system architects
- Re-evaluate previous and current learnings, be curious about the new
- Look outside our typical environment for inspiration





Thank You