



Sodalite

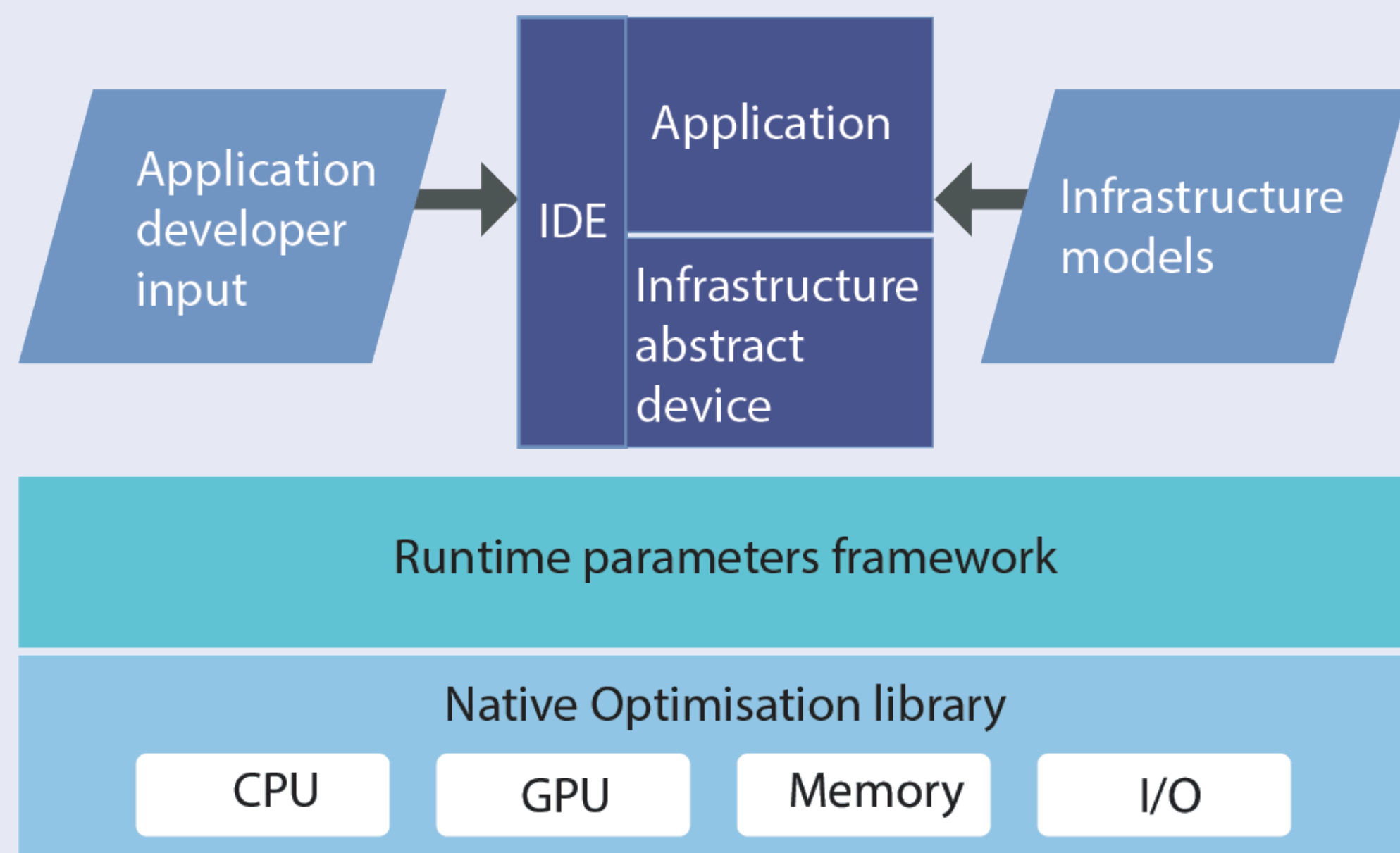
Software Defined Application Infrastructures
Management and Engineering
Feb 2019 to Jan 2022

Simplify & fully exploit benefits of heterogeneous platforms

Vision

Simpler and faster development, deployment, operation and execution of heterogeneous apps in HPC, Cloud & SW-defined computing environments.

Architecture



A **pattern-based abstraction library** that includes application, infrastructure, and performance abstractions.

An **automatic Infrastructure as Code (IaC)** engine that facilitates the development process, and also reduces deploying errors.

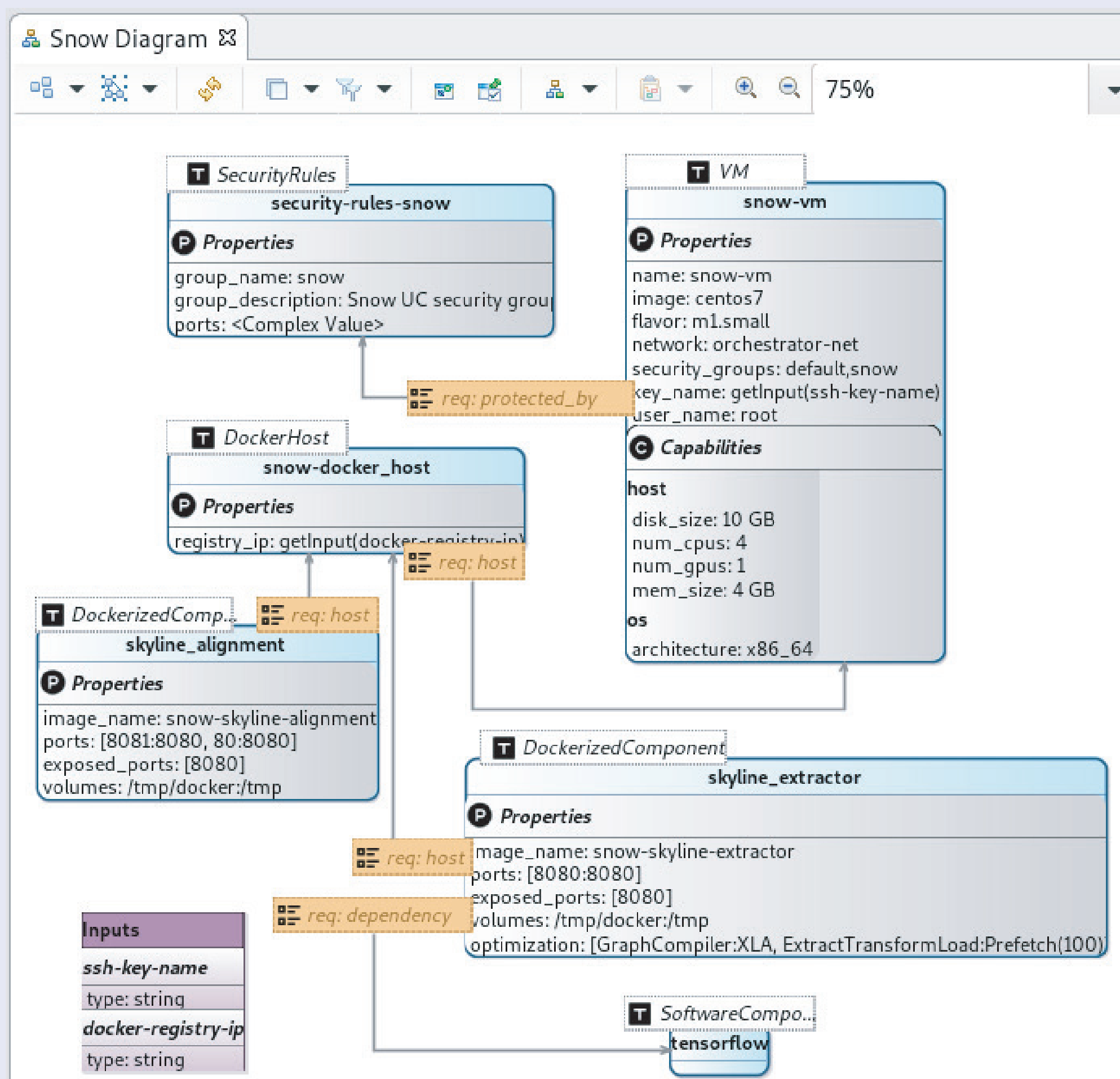
A **design and programming model** for applications and infrastructures based on the abstraction library.

A **deployment framework** that enables the static optimization of abstracted applications onto specific infrastructure.

Automated static and run-time **optimization and management of applications**.

- ✓ SODALITE allows the Application Ops Expert (AoE) to model the deployment of an optimised application on an infrastructure target using predefined application, infrastructure and performance abstractions (**SODALITE IDE**).
- ✓ The models created by the AoE are automatically translated into infrastructural code, which is then translated to an optimised deployment using state of the art container technologies (**MODAK**).
- ✓ This optimized application is then deployed by an orchestrator on multiple diverse computing platforms.
- ✓ Deployment to clusters and supercomputers with homogeneous or heterogeneous node architectures for heavy batch computations, including resources available on the Cloud and Edge devices, is supported.

SODALITE IDE



Status

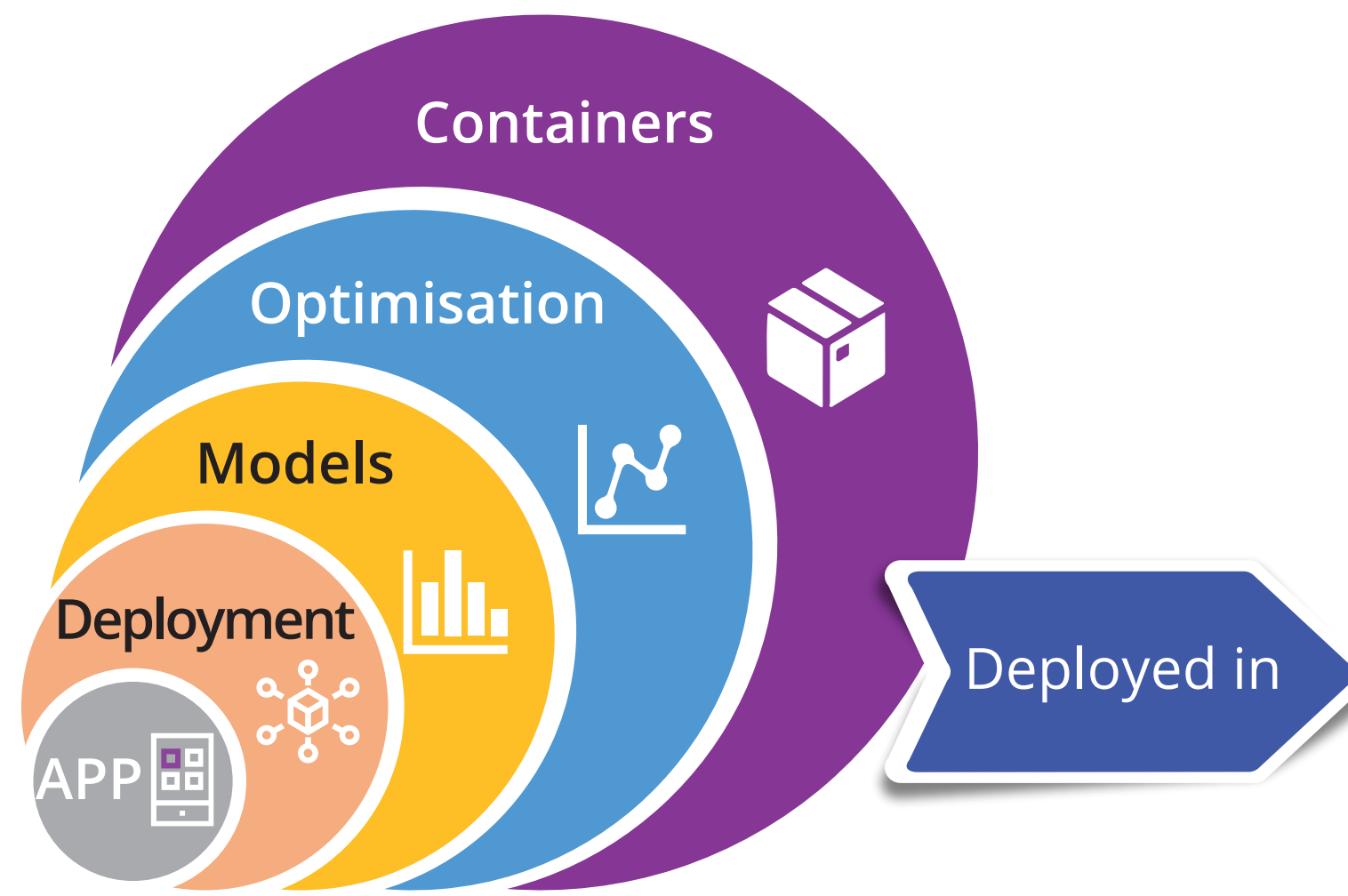
Run Time: 3-year project, started in February 2019.

Current Status: Design and development of the alpha prototype of MODAK and SODALITE framework.

Next Steps: Integrating MODAK with the SODALITE framework and demonstrating with GPU Snow and Insilico clinical trails use case.

Karthee Sivalingam (CRAY-HPE), Alfio Lazzaro (CRAY-HPE), Nina Mujkanovic (CRAY-HPE), Daniel Vladušić (XLAB), Joao Pita Costa (XLAB) Maria Carbonell (ATOS) and Yosu Gorroñoigoitia (ATOS)

Performance Requirements



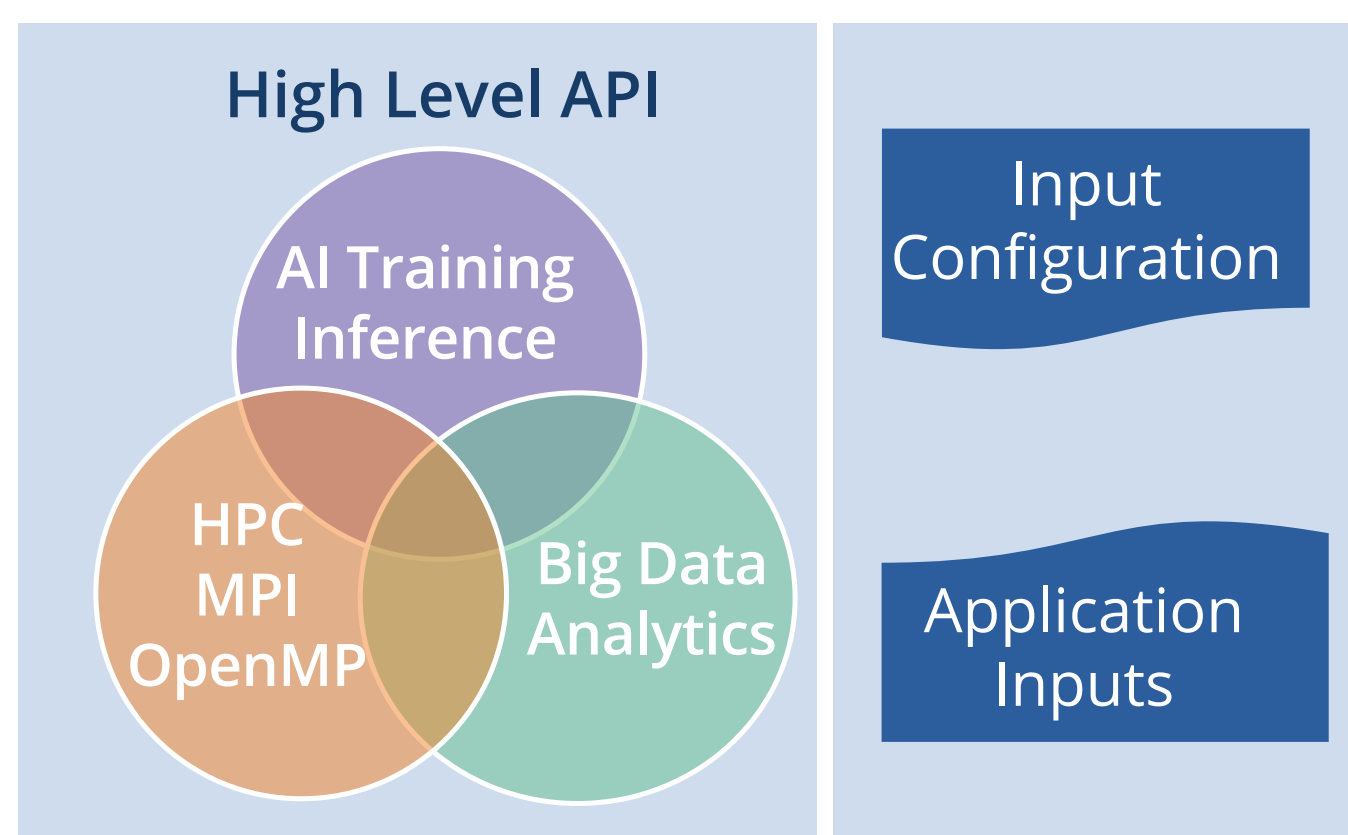
SODALITE attempts to bring the vast knowledge of performance optimisation accrued by the HPC industry over decades into the cloud computing area.

Heterogenous Infrastructure (HPC/Cloud)

COMPUTE	STORAGE	NETWORK	SCHEDULER
CPU GPU FPGA	LUSTRE SSD Object Store	Ethernet InfiniBand PCIe	Torque/Slurm Kubernetes OpenStack

MODAK Architecture

MODAK focuses on supporting three major application types for static optimisation: AI training/Inference, Big Data Analytics and Traditional HPC

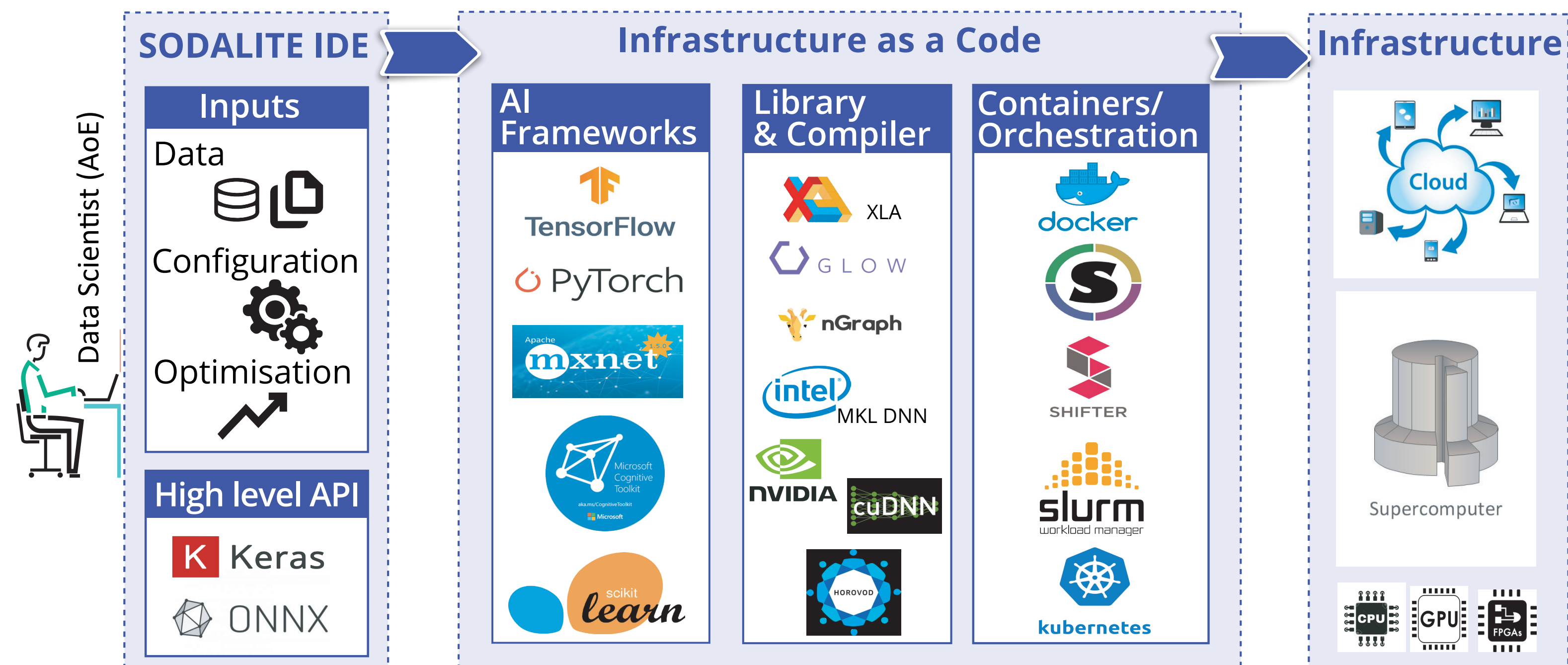


The performance models are developed by running standard benchmarks across different configurations and then building a simple linear statistical model.

Singularity, for HPC Build system, custom format for images (SIF)
Can use Docker images
Rootless installation

- ✓ MODAK, a model-based application deployment optimiser enables static optimisation.
- ✓ An application's performance can be predicted using the performance models of the application and infrastructure.
- ✓ These model will inform how the application parameters (like input data size and format) influence the performance and also the performance characteristics of the target infrastructure, such as peak performance and memory bandwidth.
- ✓ Based on AoE selected optimisations (Optimisation recipe), MODAK maps the optimal application parameters to the infrastructure target and builds an optimised container (using prebuilt images from the image registry).

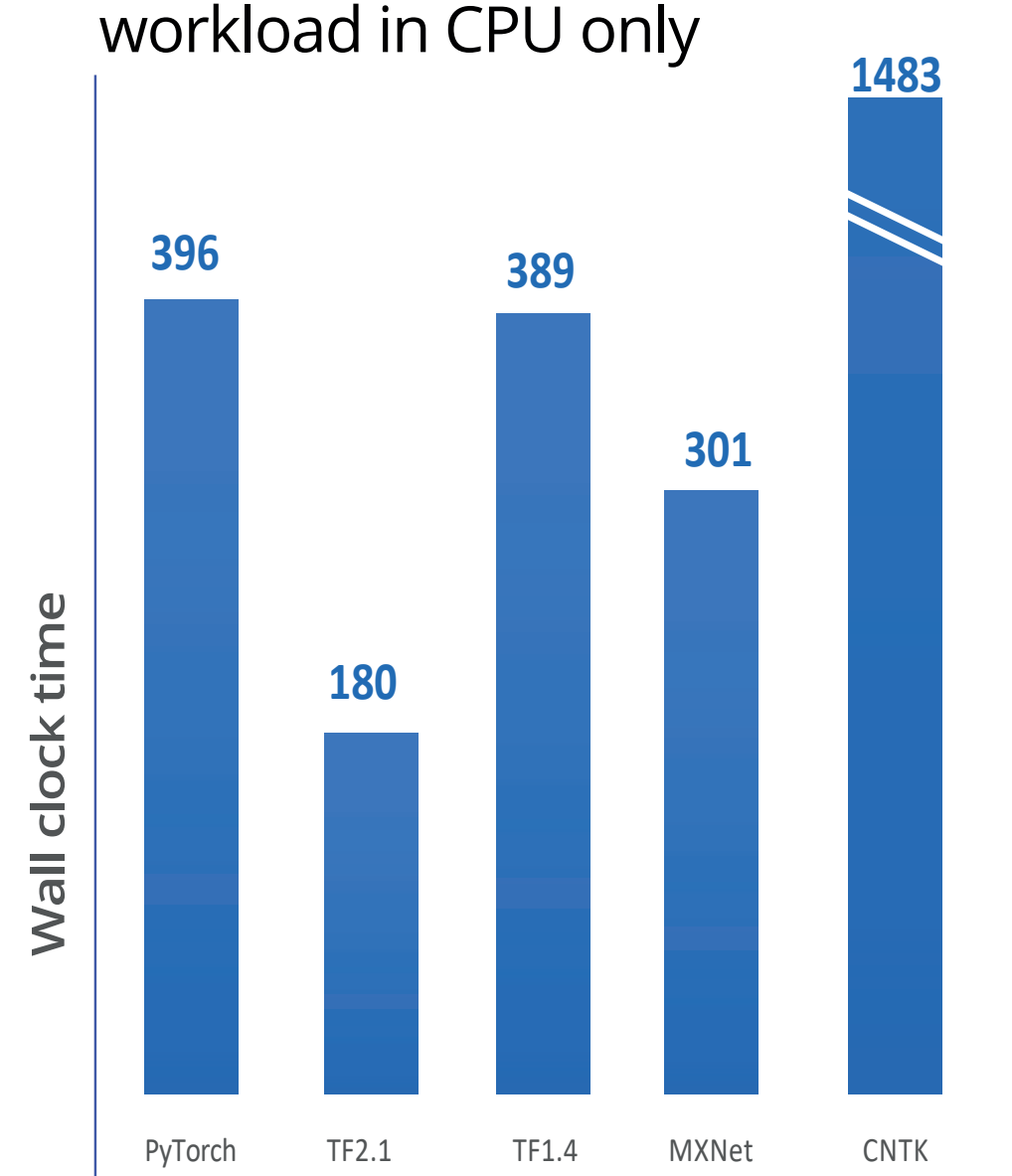
AI Example



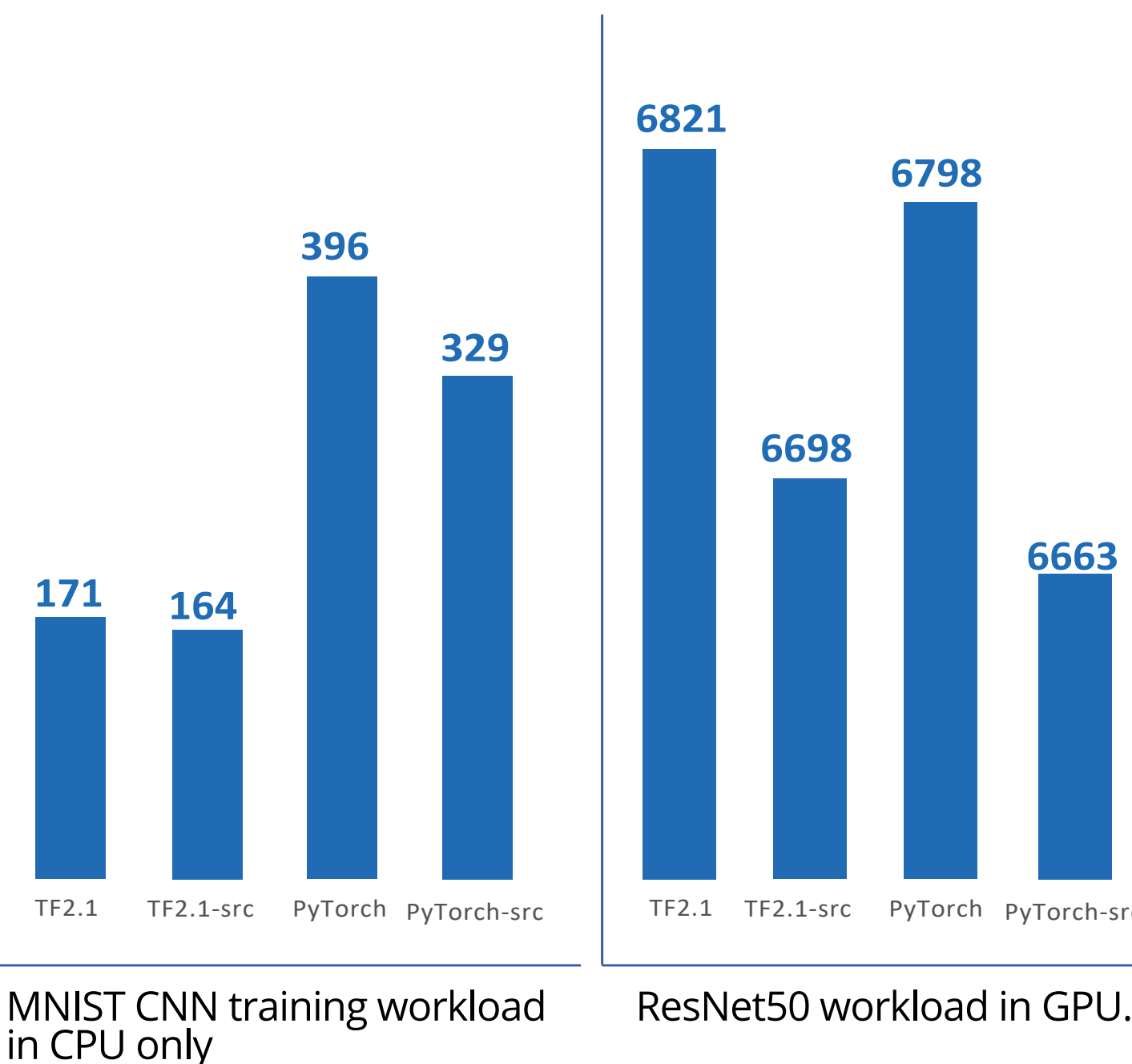
AI example shows how the data scientist will use the SODALITE framework by specifying the data, config and optimisations options while deploying an AI network written in a highlevel API. The SODALITE application optimiser will select a preferred AI framework, optimised library and compiler and then build an optimised container. This will then be deployed to an HPC or Cloud infrastructure.

Comparison of AI Frameworks

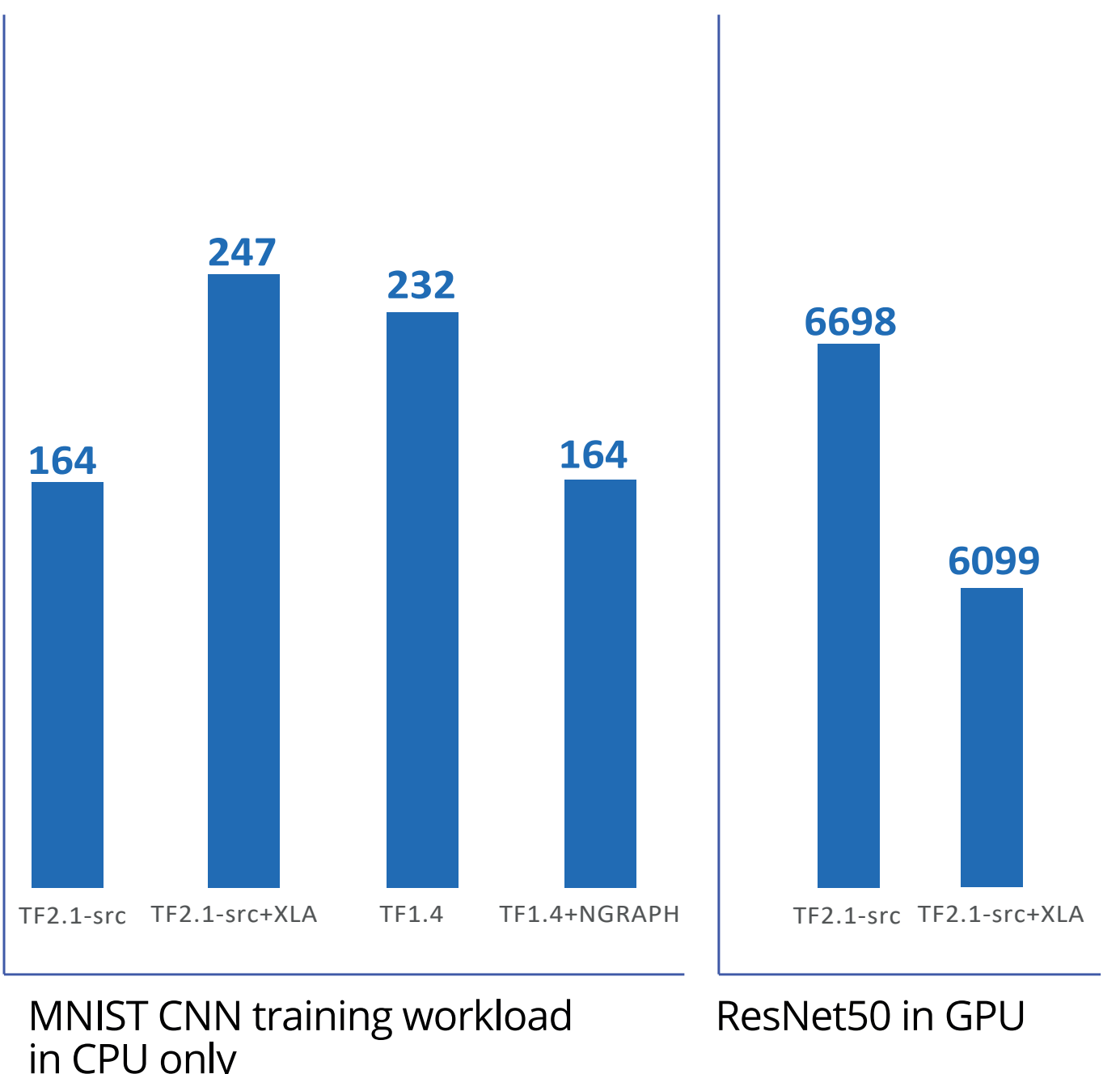
Performance of different AI frameworks (in singularity container) for a MNIST CNN training workload in CPU only



Performance of AI frameworks in singularity containers built from source



Performance of AI frameworks in singularity containers with Graph compilers



Use Cases

GPU Snow

Problem: Need for a reconfigurable workflow (CPU/GPU/IO bound), to be deployed anywhere and optimised for that infrastructure.

Solution: Optimisation and reconfiguration improve potential prediction accuracy due to improved throughput of data.

Vehicle IoT

Problem: Changing compliance, privacy, and security needs in a dynamic environment, combined with limited computational capacity at the network Edge.

Solution: Adaptive Application and Deployment Reconfiguration, leveraging heterogeneous compute resources in a multi-cloud (Cloud-to-Edge) environment.

In-silico Clinical Trials

Problem: Production-ready, complex workflow, needs to be capable to efficiently run anywhere.

Solution: Deployment optimisation, heterogeneity support and deployment reconfiguration - enabling to target any infrastructure.

@SODALITESW

@SODALITE.EU

projectinfo@sodalite.eu

www.sodalite.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825480. Privacy policy

