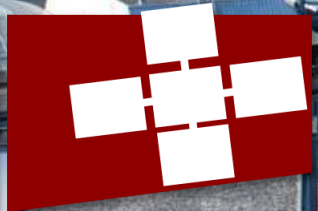


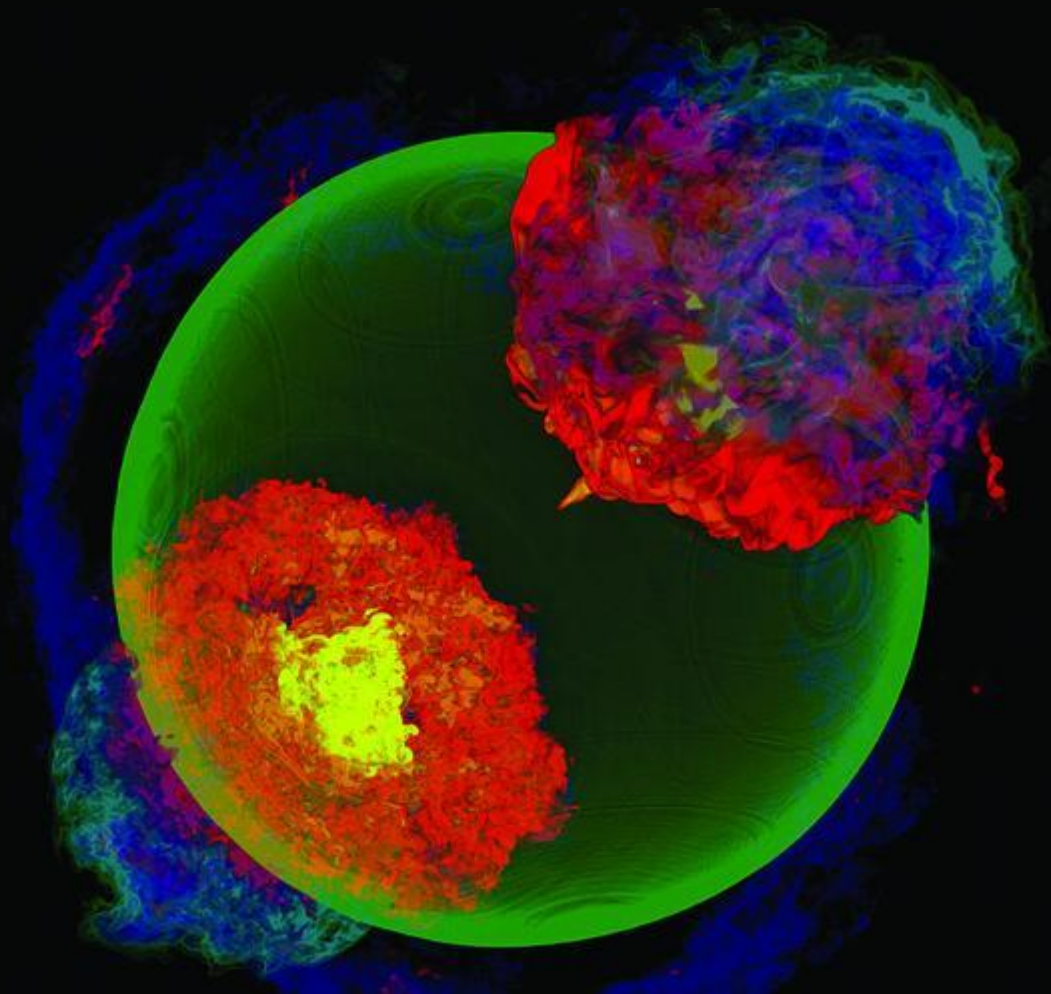
SALVATORE DI GIROLAMO, P. SCHMID, T. SCHULTHESS, T. HOEFLER

SimFS: A Simulation Data Virtualizing File System Interface

PRESENTED AT IPDPS '19

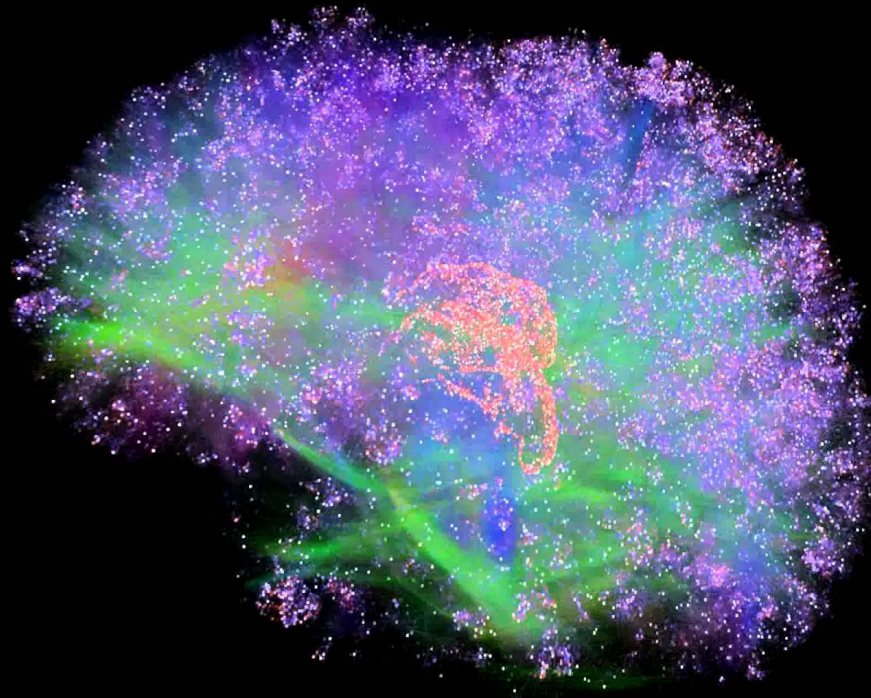


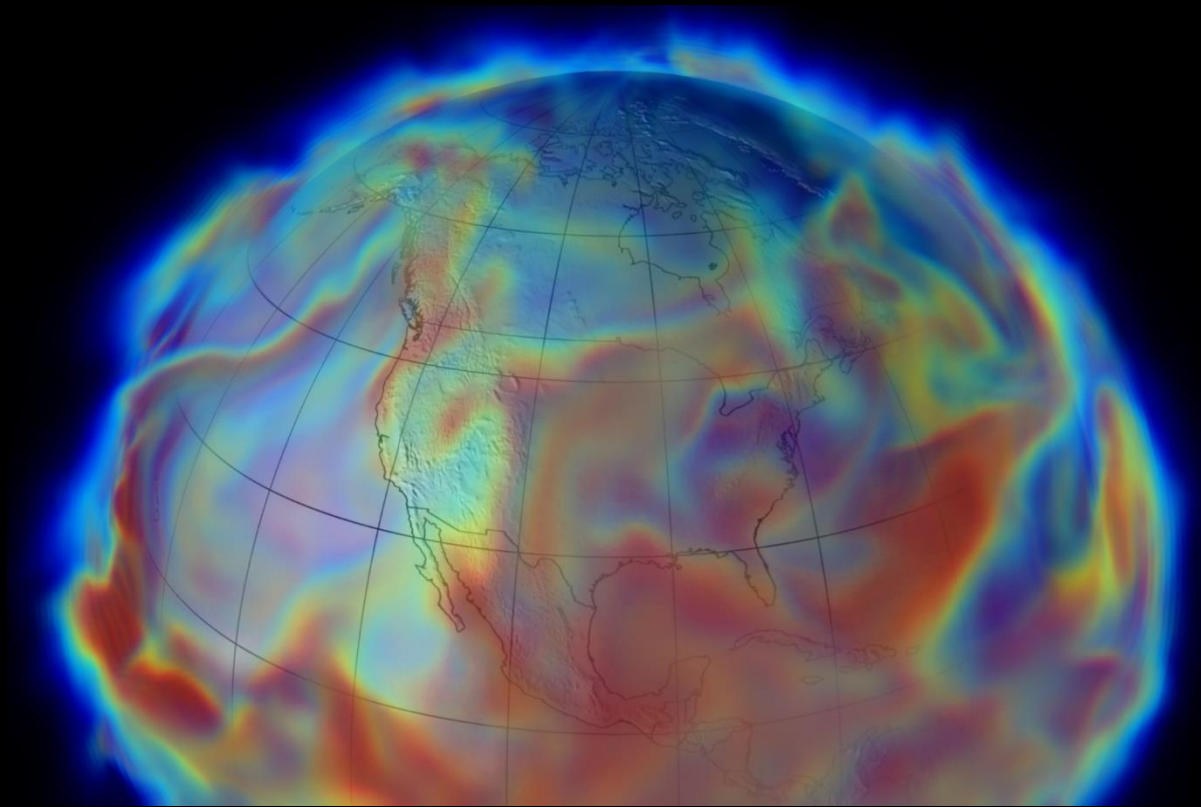
IPDPS
2019 Rio de Janeiro



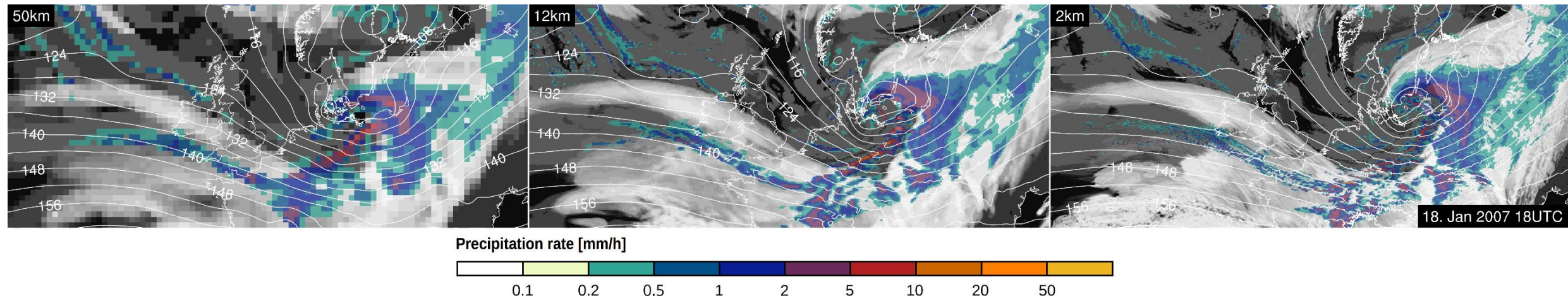
SpineSim v0.71 (64-bit) - (C) 2011-2012 Tom Dierker
#388608 Neurons with 1447276538 Synapses
Simulation time: 1792.5 ms

sp3v (TheLemon): ON

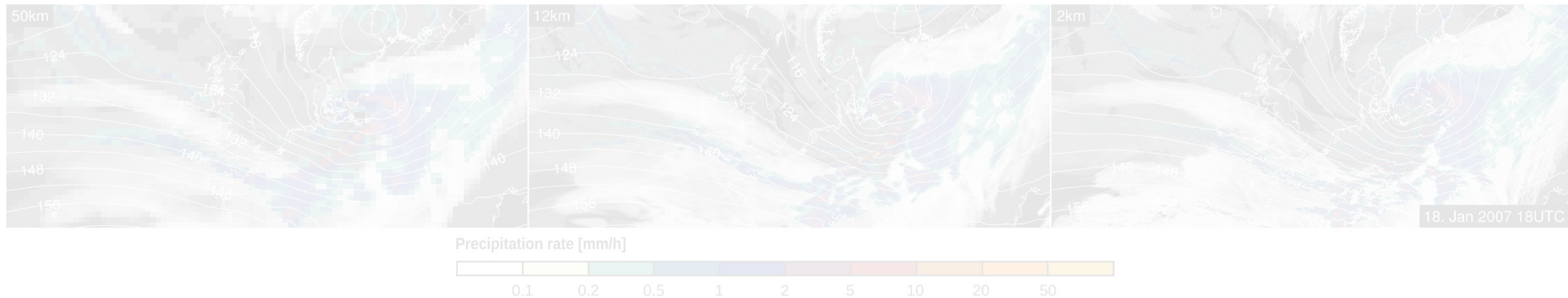




crCLIM: Convection-Resolving Climate Modeling



crCLIM: Convection-Resolving Climate Modeling



innovations report
Forum for Science, Industry and Business

Search our Site:

RSS-Feed

SCIENCE REPORTS SPECIAL TOPICS B2B AREA JOBS & OPPORTU

REPORTS AND NEWS
Home → Science Reports → Reports and News → Information Technology


- Agricultural and Forestry Science
- Architecture and Construction
- Automotive Engineering
- Business and Finance
- Communications Media
- Earth Sciences
- Ecology, The Environment and Conservation
- Health and Medicine
- Information Technology
- Interdisciplinary Research
- Life Sciences
- Machine Engineering
- Material Sciences

The world's largest archive with up to 500 Petabyte of climate data is located in Hamburg

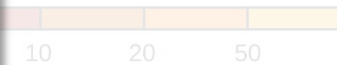
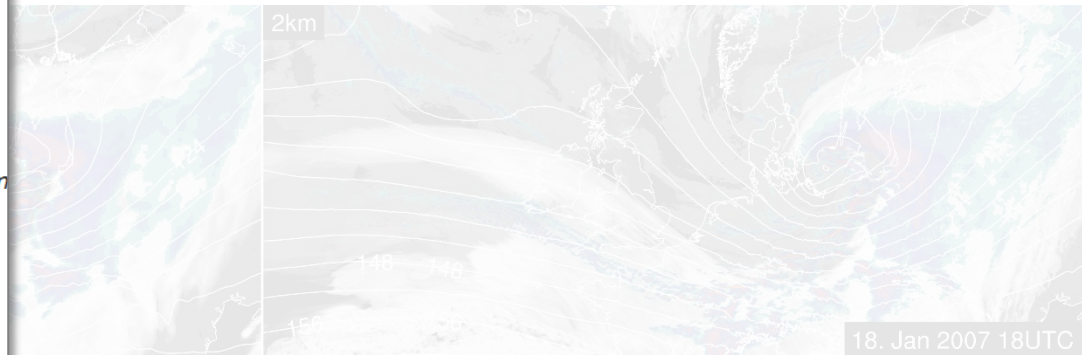
08.05.2015
DKRZ, supported by IBM, took in its new Hierarchical Storage Management System based on a High Performance Storage Systems (HPSS). Experts from IBM/Global Technology Services have implemented the latest HPSS software and the Linux operation system by Red Hat on the IBM x86 server as well as a HPSS disk cache with five Petabyte.

The new system allows currently a transfer of climate data between the high performance computer and the archive system with a bidirectional bandwidth up to 12 Gigabyte per second, increasing up to 18 Gigabyte per second in autumn 2015.

The system manages all archived simulation data at DKRZ and serves as the central in- and output



ling





One of the Largest Cosmological Simulations Ever Shows the Evolution of the Universe

Off World

RSS-Feed

- REPORTS AND NEWS
- Agricultural and Food Sciences
- Architecture and Construction
- Automotive Engineering
- Business and Finance
- Communications Media
- Earth Sciences
- Ecology, The Environment and Conservation
- Health and Medicine
- Information Technology
- Interdisciplinary Research
- Life Sciences
- Machine Engineering
- Material Sciences

IN BRIEF

Scientists run one of the largest ever simulations of the evolution of the universe in order to understand more about dark matter, generating **2.5 petabytes** of data in the process.

THE BIRTH OF THE UNIVERSE

Researchers from the U.S. Department of Energy's Argonne National Laboratory have acquired an "avalanche of data" from one of the largest cosmological simulations ever. The simulation, called the Q Continuum, ran a model of the universe's evolution from 50 million years after the Big Bang up to the present day. The goal of the simulation is to help researchers understand dark

SHARE



WRITTEN BY

Miguel Santos

November 4, 2015

#big bang #dark matter
#space #titan supercomputer



2007 18UTC

nce is knowledg

RSS-Feed

REPORTS AND N

- Agricultural and Fo
- Architecture and C
- Automotive Engine
- Business and Finan
- Communications M
- Earth Sciences
- Ecology, The Envir
- Conservation
- Health and Medicin
- Information Techno
- Interdisciplinary Re
- Life Sciences
- Machine Engineeri
- Material Sciences

Off World

One of the Largest Cosmological Simulations Ever Shows the Evolution of the Universe



IN BRIEF

Scientists run one of the largest ever simulations of the evolution of the universe in order to understand more about dark matter, generating **2.5 petabytes** of data in the process.

THE BIRTH OF THE UNIVERSE

Researchers from the U.S. Department of Energy's Argonne National Laboratory have acquired an "avalanche of data" from one of the largest cosmological simulations ever. The simulation, called the Q Continuum, ran a model of the universe's evolution from 50 million years after the Big Bang up to the present day. The goal of the simulation is to help researchers understand dark

SHARE



WRITTEN BY

Miguel Santos

November 4, 2015

#big bang #dark matter
#space #titan supercomputer

nce is knowledg

RSS-Feed

Off World

REPORTS AND N

Agricultural and Fo

Architecture and C

Automotive Engine

Business and Finar

Communications M

Earth Sciences

Ecology, The Envir

Conservation

Health and Medicin

Information Techno

Interdisciplinary Re

Life Sciences

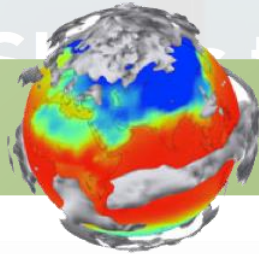
Machine Engineeri

Material Sciences

One of the Largest Cosmological Simulations

MegaBytes

1995



IN BRIEF

Scientists run one of the largest ever simulations of the evolution of the universe in order to understand more about dark matter, generating **2.5 petabytes** of data in the process.

THE BIRTH OF THE UNIVERSE

Researchers from the U.S. Department of Energy's Argonne National Laboratory have acquired an "avalanche of data" from one of the largest cosmological simulations ever. The simulation, called the Q Continuum, ran a model of the universe's evolution from 50 million years after the Big Bang up to the present day. The goal of the simulation is to help researchers understand dark

SHARE



WRITTEN BY

Miguel Santos

November 4, 2015

#big bang #dark matter
#space #titan supercomputer

One of the Largest Cosmological Simulations



IN BRIEF

Scientists run one of the largest ever simulations of the evolution of the universe in order to understand more about dark matter, generating **2.5 petabytes** of data in the process.

THE BIRTH OF THE UNIVERSE

Researchers from the U.S. Department of Energy's Argonne National Laboratory have acquired an "avalanche of data" from one of the largest cosmological simulations ever. The simulation, called the Q Continuum, ran a model of the universe's evolution from 50 million years after the Big Bang up to the present day. The goal of the simulation is to help researchers understand dark

SHARE



WRITTEN BY

Miguel Santos

November 4, 2015

#big bang #dark matter #space #titan supercomputer

innovations report

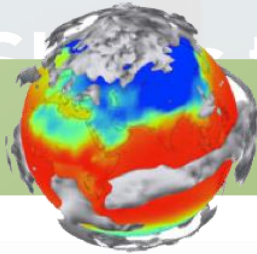
Forum for Science, Industry and Business

Off World

One of the Largest Cosmological Simulations

MegaBytes

1995



PetaBytes

2015



ExaBytes?

2025



IN BRIEF

Scientists run one of the largest ever simulations of the evolution of the universe in order to understand more about dark matter, generating **2.5 petabytes** of data in the process.

THE BIRTH OF THE UNIVERSE

Researchers from the U.S. Department of Energy's Argonne National Laboratory have acquired an "avalanche of data" from one of the largest cosmological simulations ever. The simulation, called the Q Continuum, ran a model of the universe's evolution from 50 million years after the Big Bang up to the present day. The goal of the simulation is to help researchers understand dark

SHARE



WRITTEN BY

Miguel Santos

November 4, 2015

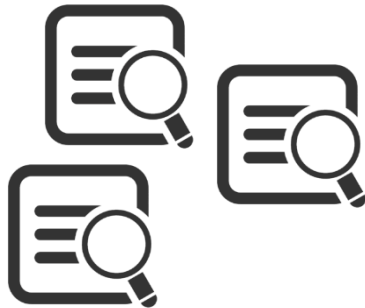
#big bang #dark matter
#space #titan supercomputer

Analyzing Data Today: On-Disk

Simulation



Analysis/Visualization Tools



Results



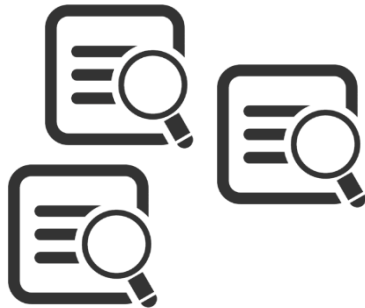
Analyzing Data Today: On-Disk



Simulation



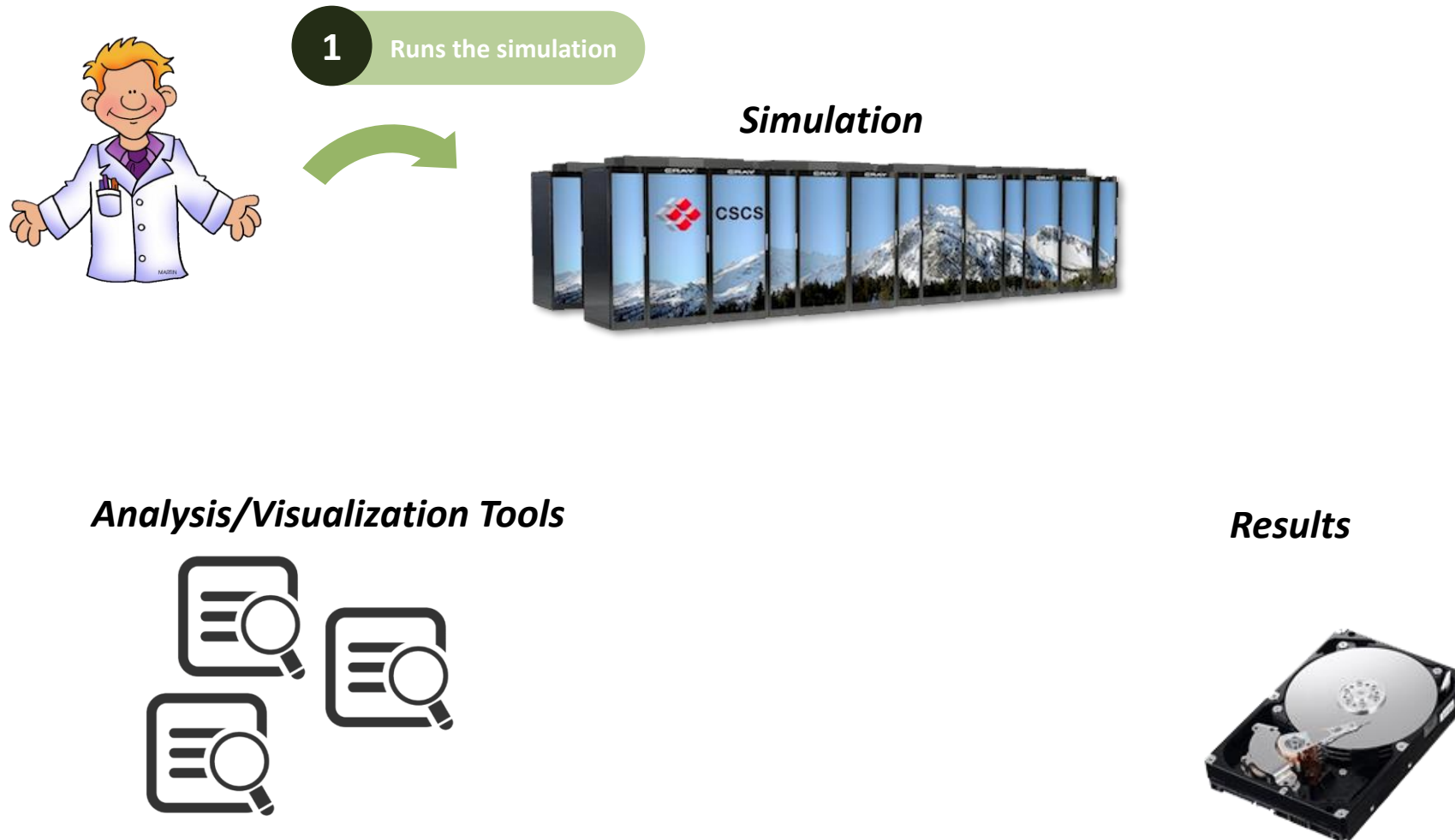
Analysis/Visualization Tools



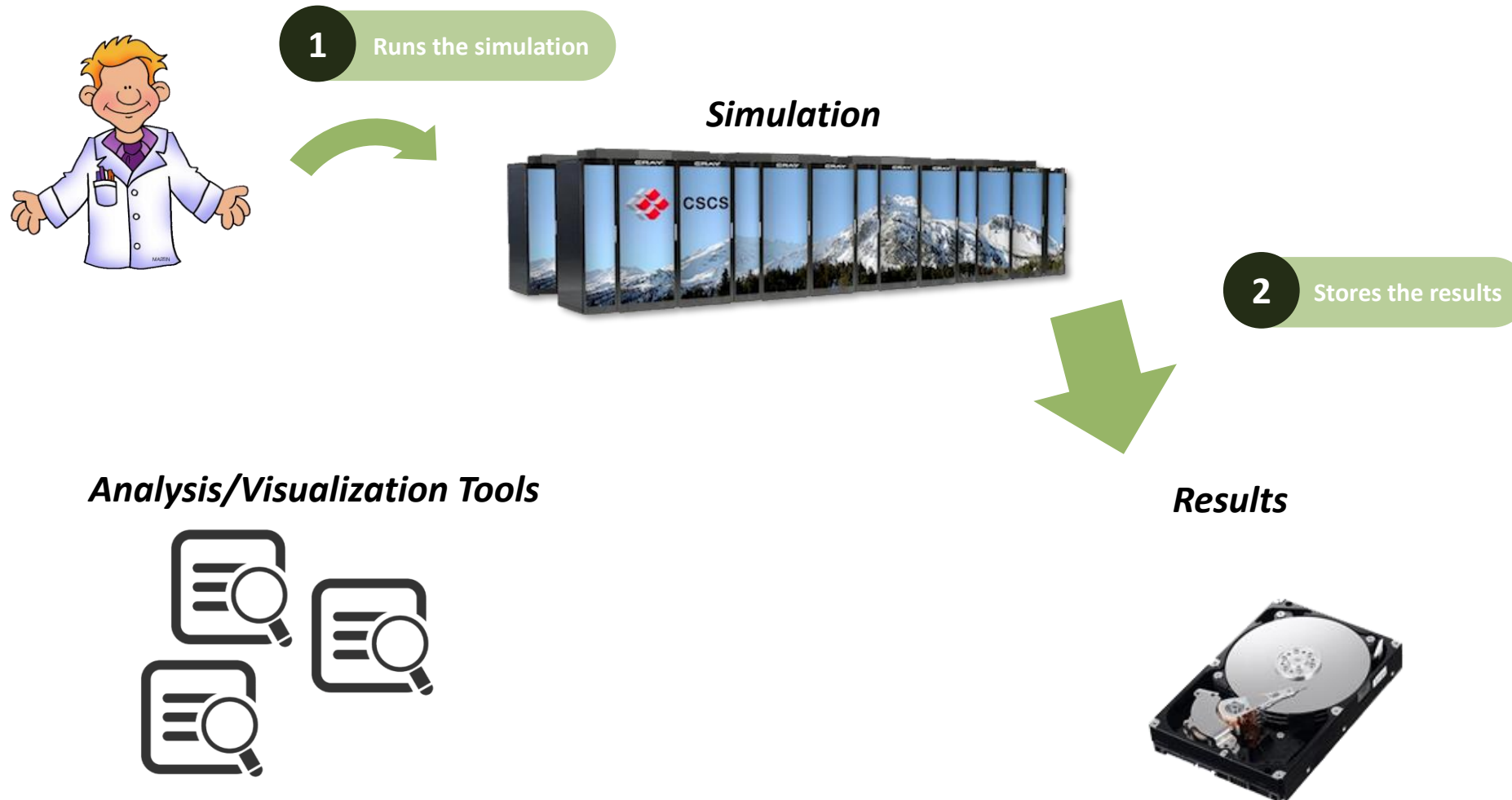
Results



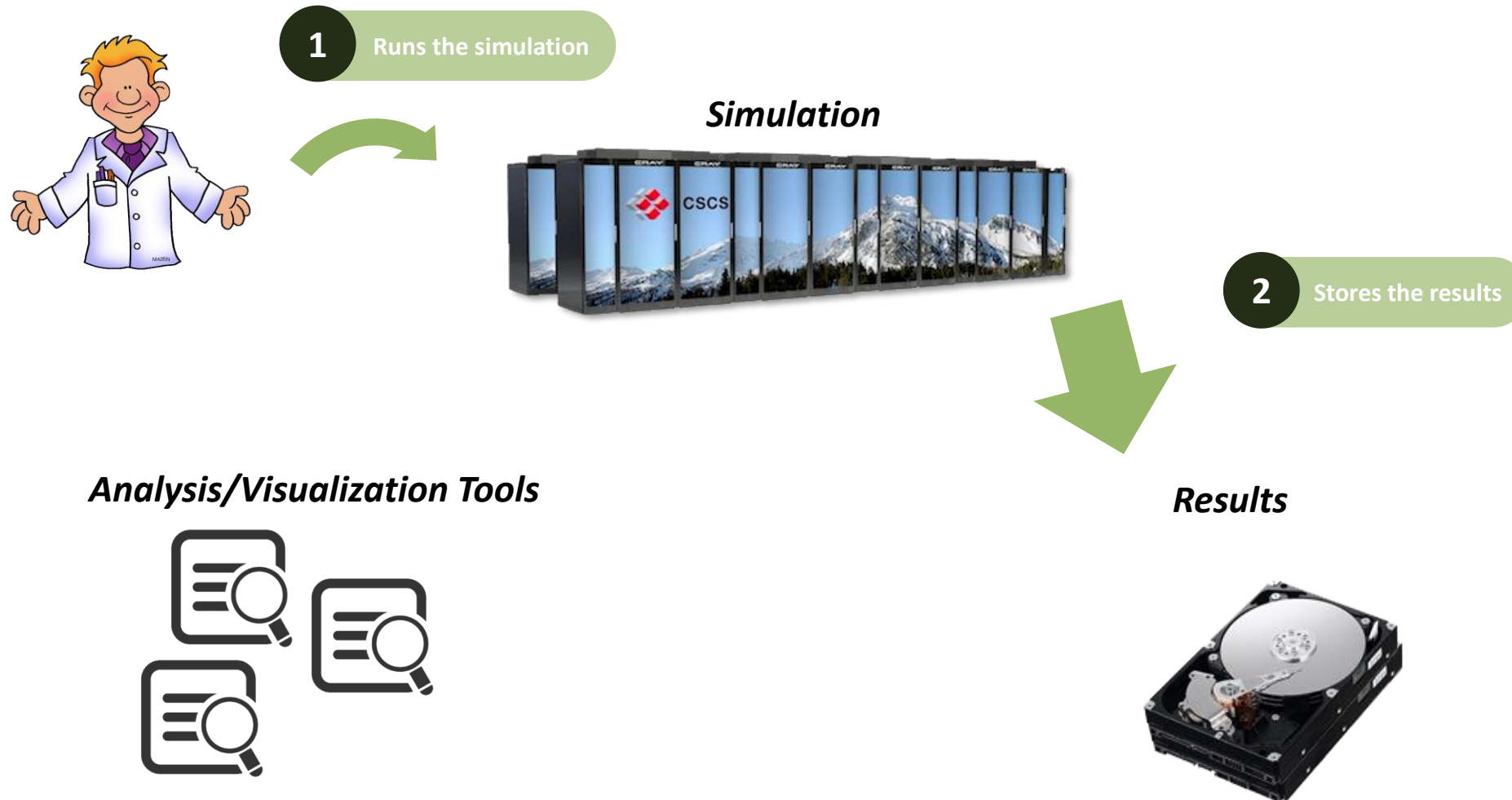
Analyzing Data Today: On-Disk



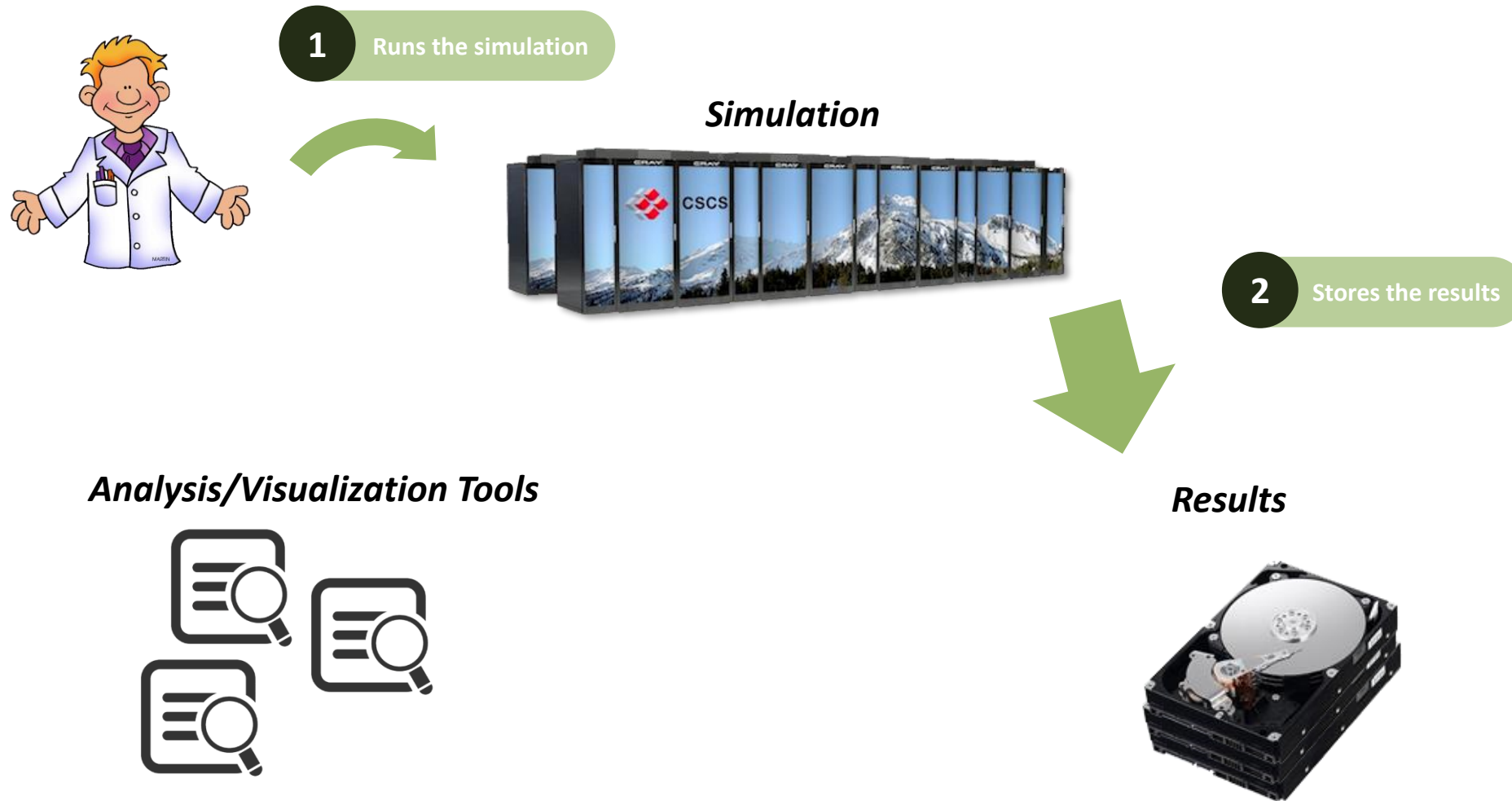
Analyzing Data Today: On-Disk



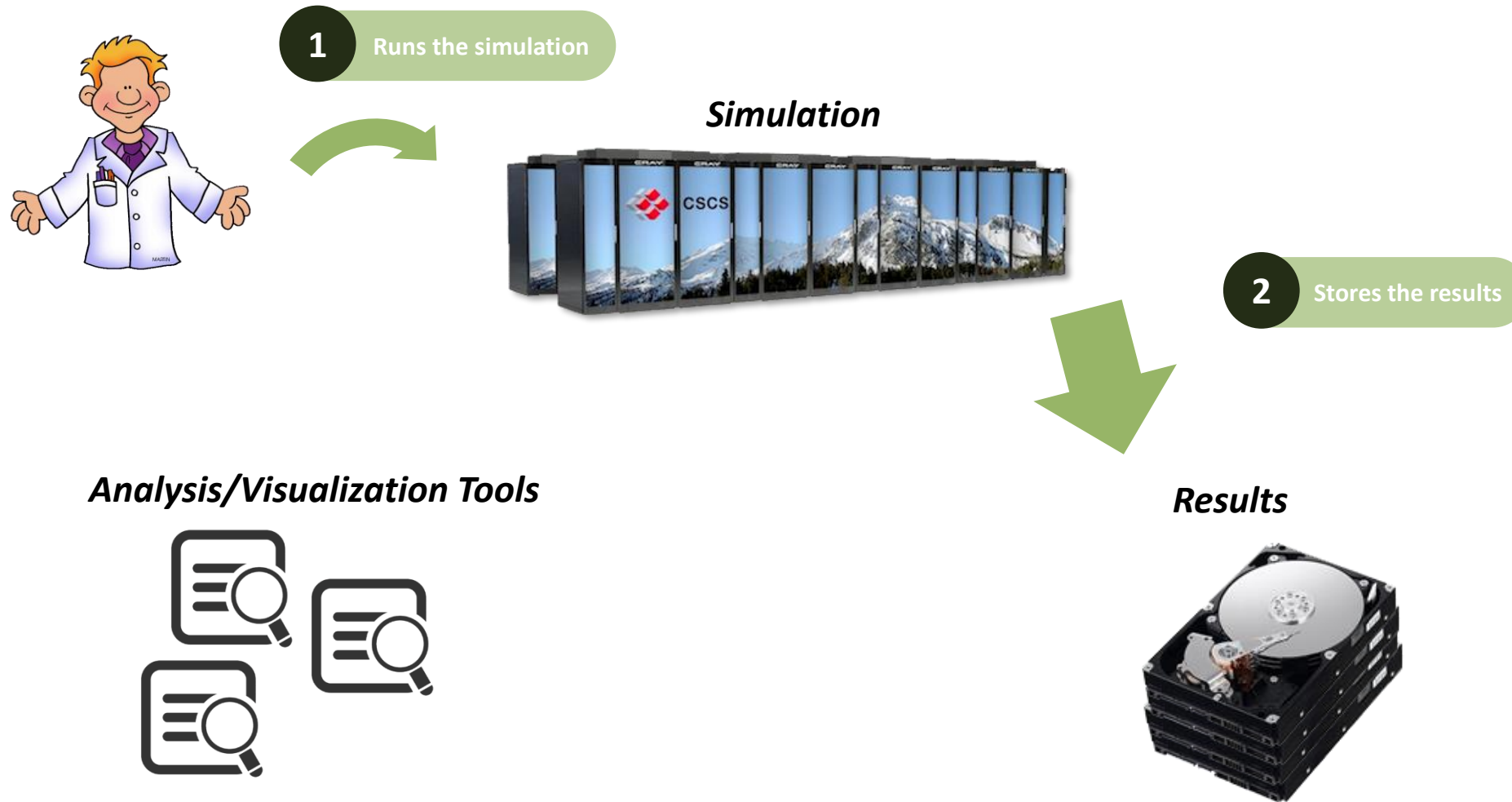
Analyzing Data Today: On-Disk



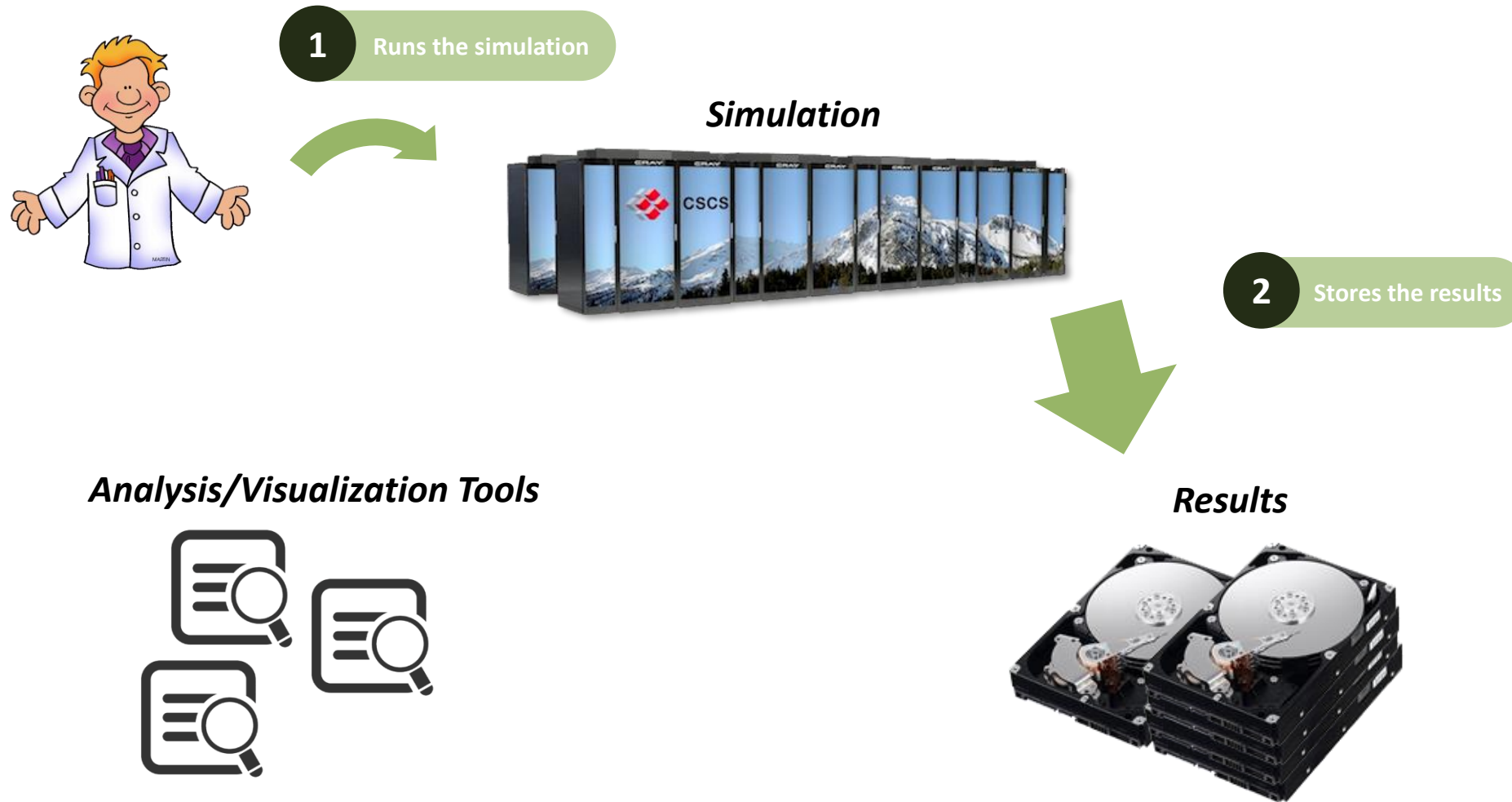
Analyzing Data Today: On-Disk



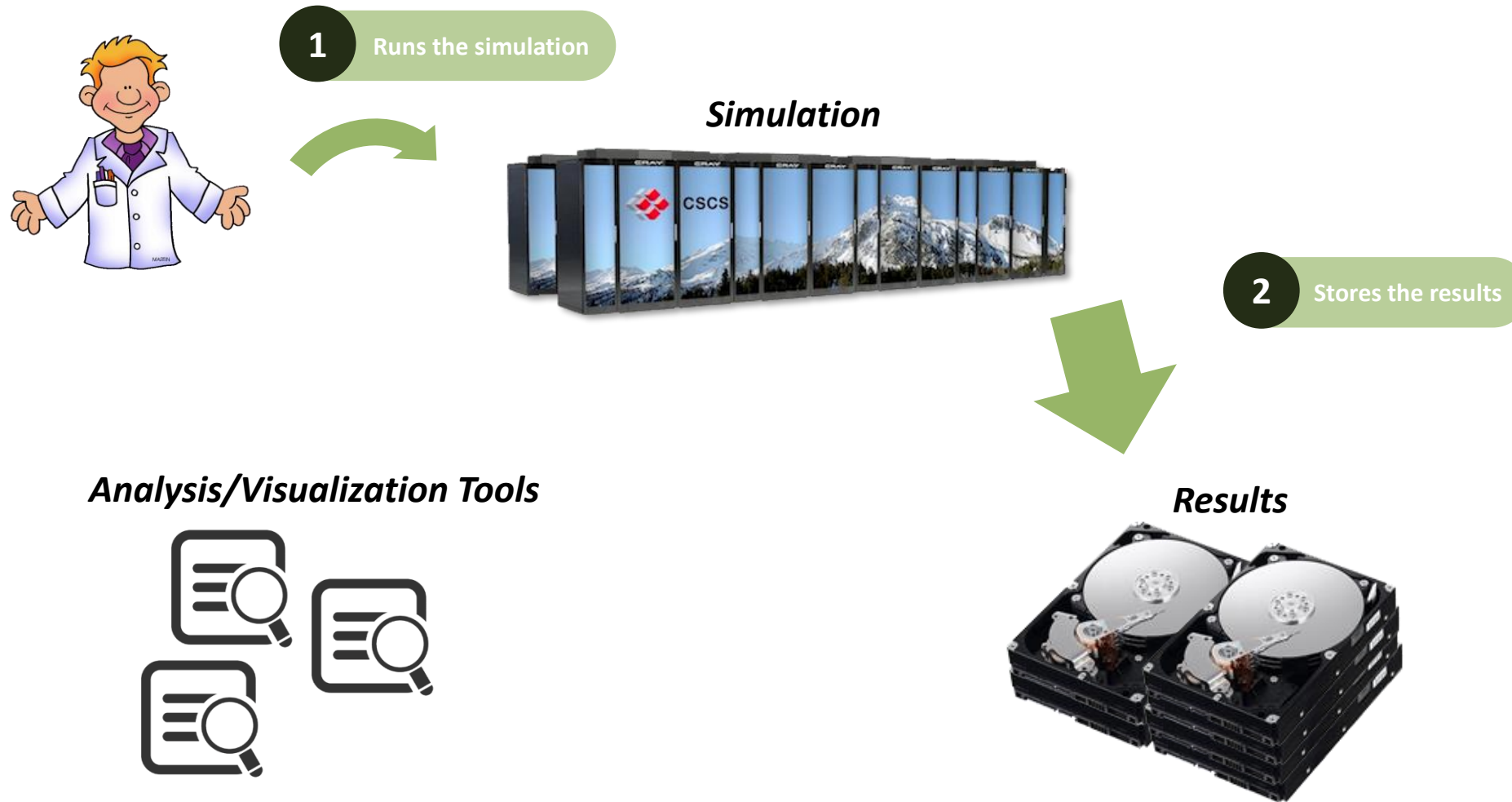
Analyzing Data Today: On-Disk



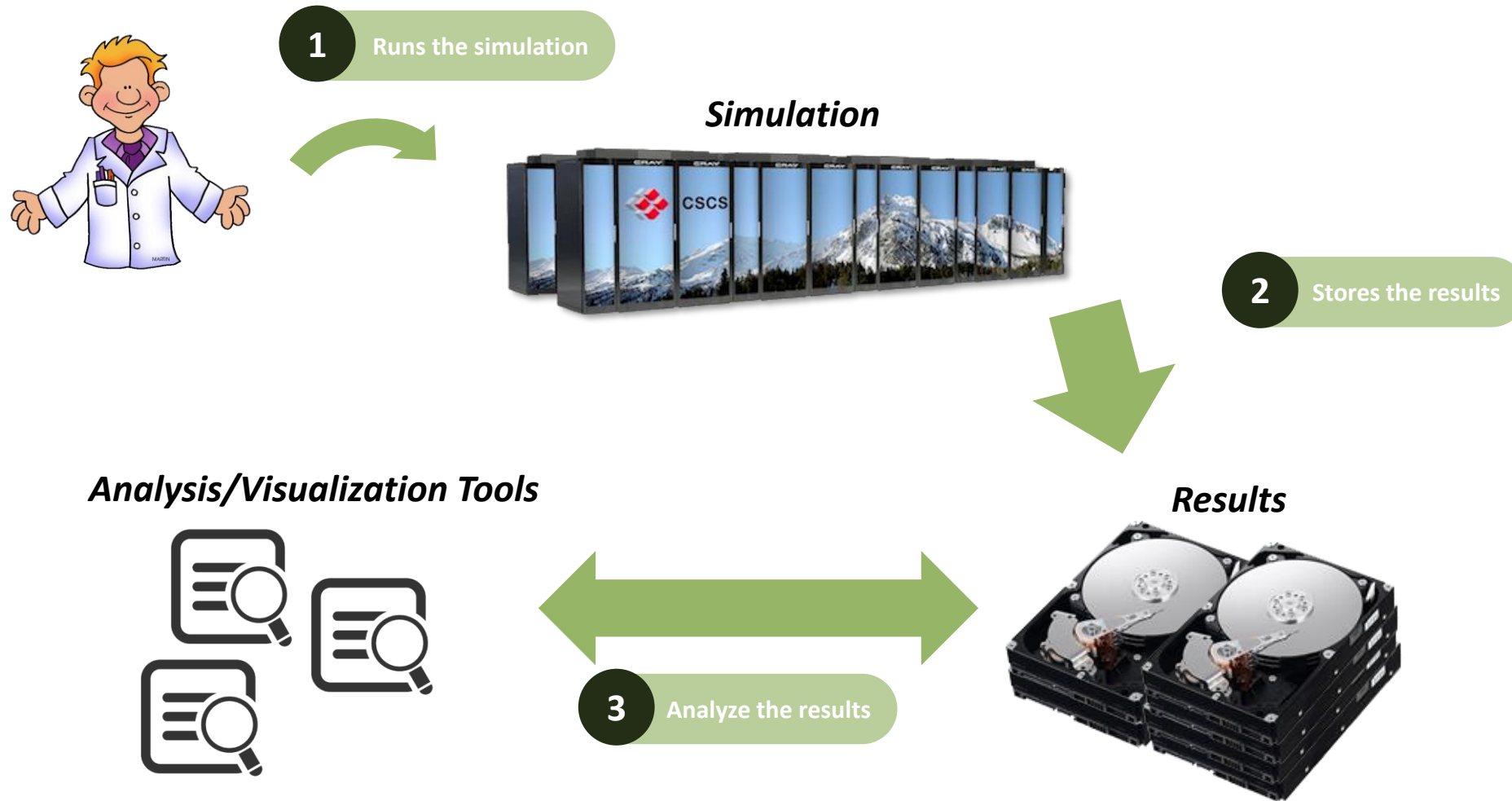
Analyzing Data Today: On-Disk



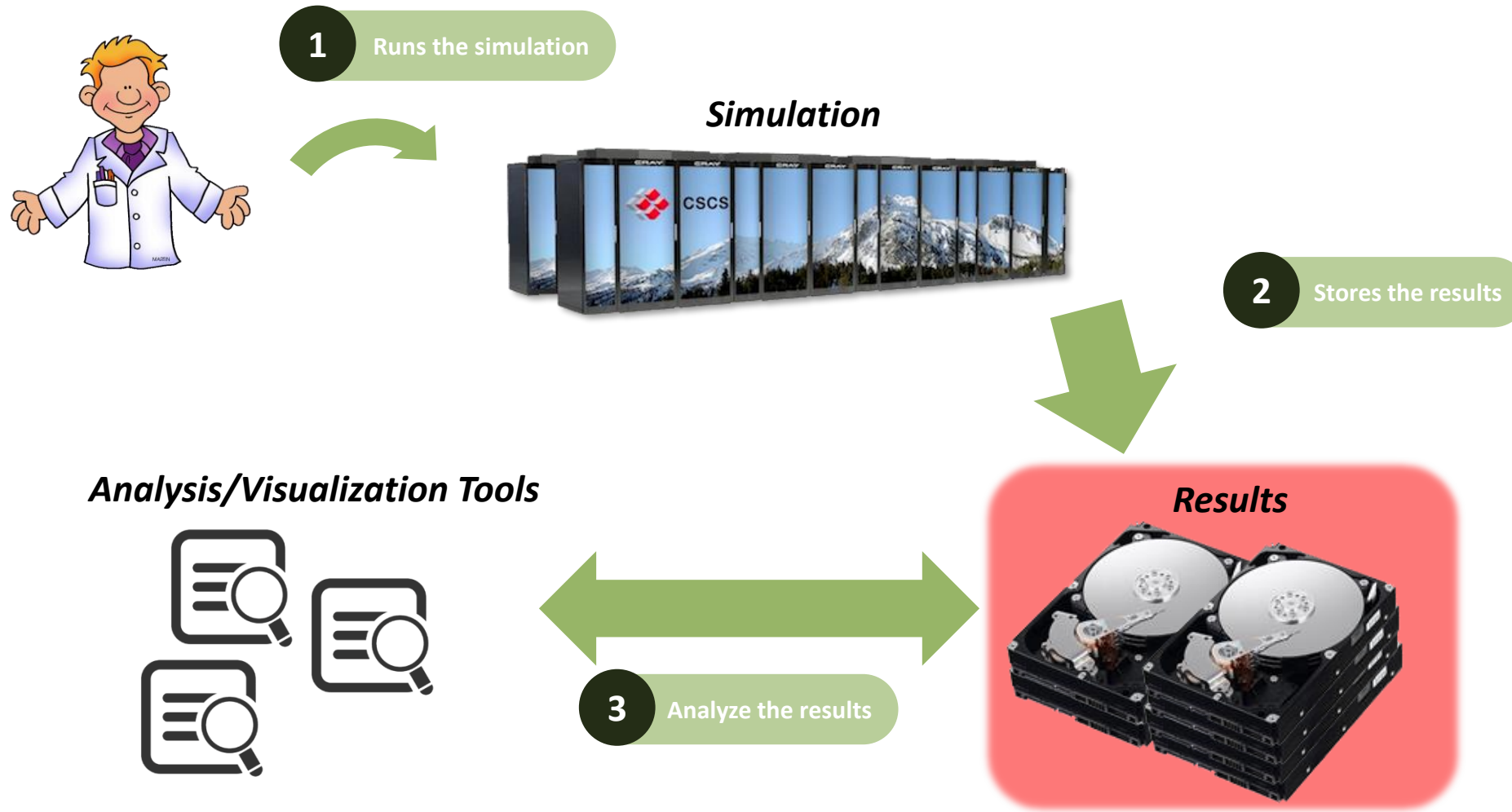
Analyzing Data Today: On-Disk



Analyzing Data Today: On-Disk



Analyzing Data Today: On-Disk

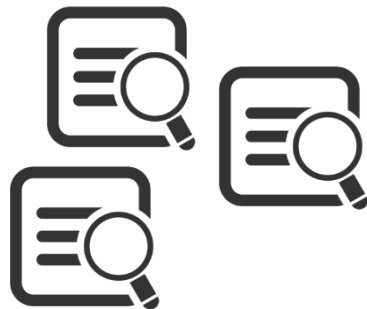


Analyzing Data Today: In Situ

Simulation



Analysis/Visualization Tools



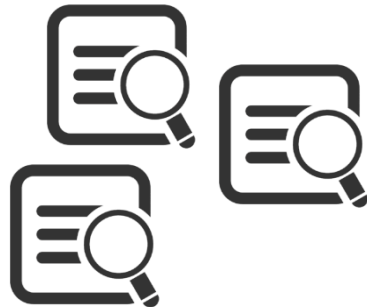
Analyzing Data Today: In Situ



Simulation



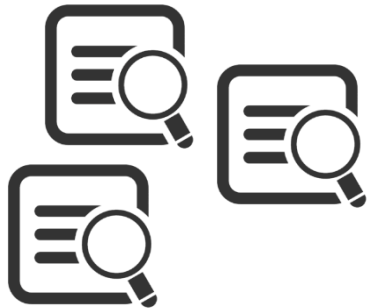
Analysis/Visualization Tools



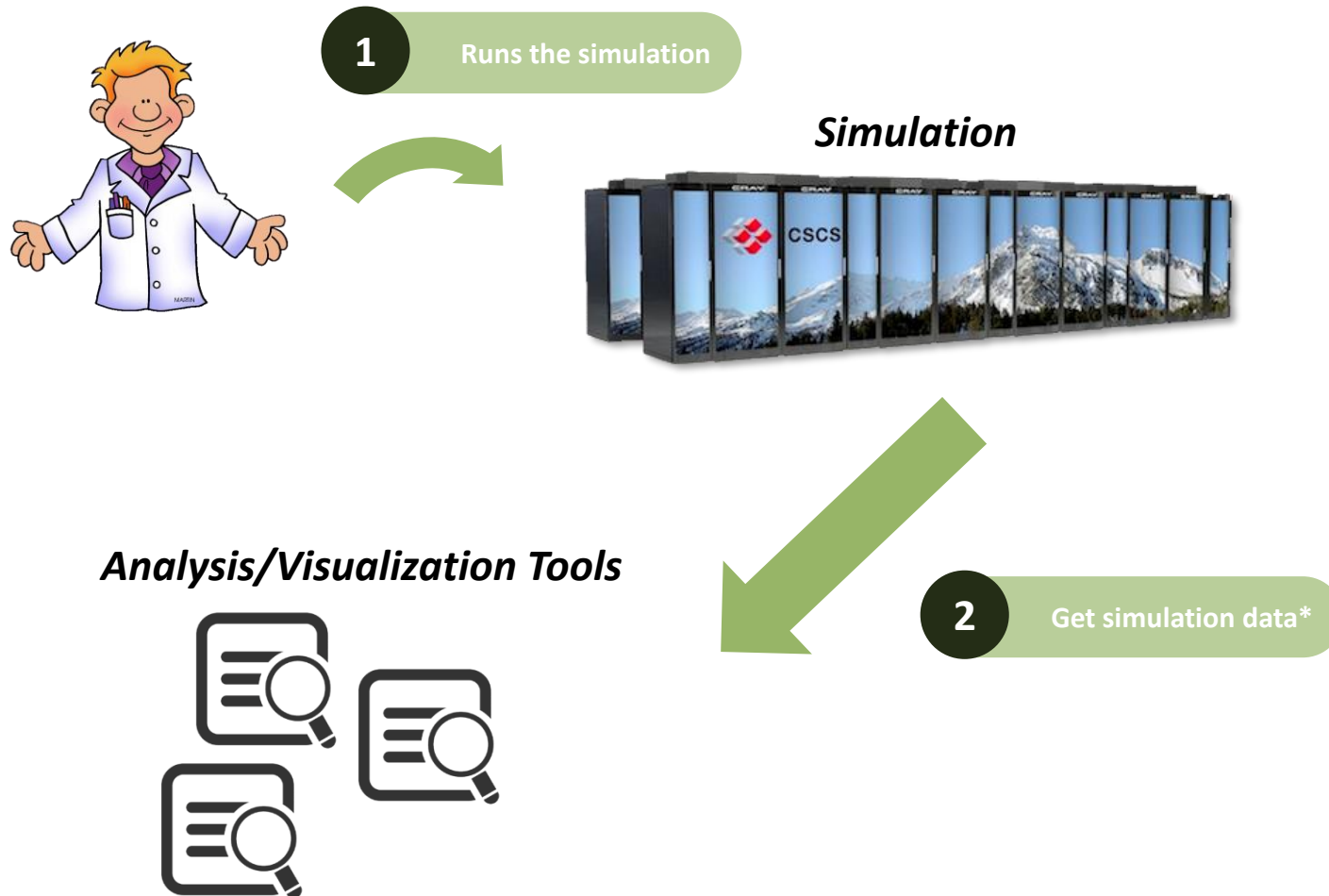
Analyzing Data Today: In Situ



Analysis/Visualization Tools



Analyzing Data Today: In Situ





ON DISK



store everything

never re-simulate



ON DISK



store everything

never re-simulate

IN SITU



store nothing

always re-simulate



ON DISK



SimFS

IN SITU



store everything

never re-simulate

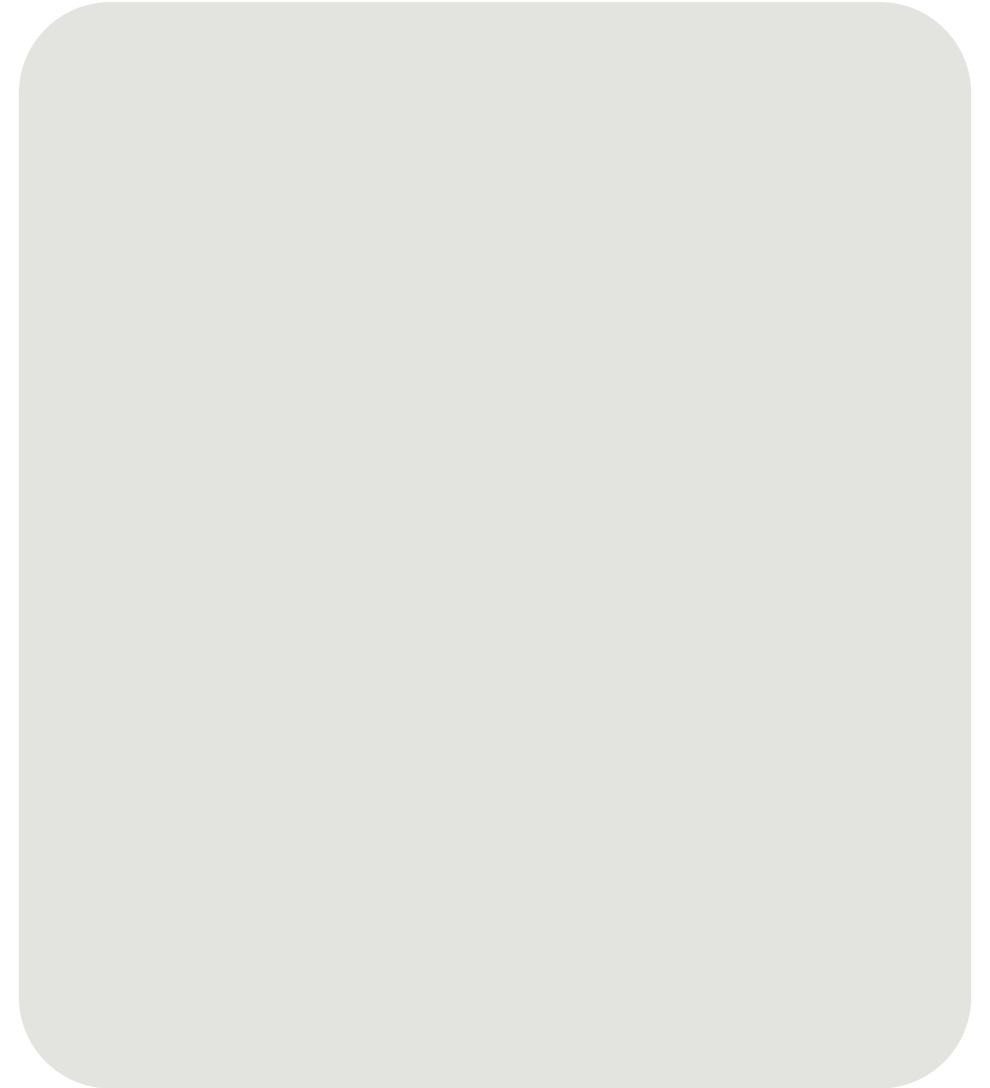
store something

re-simulate if needed

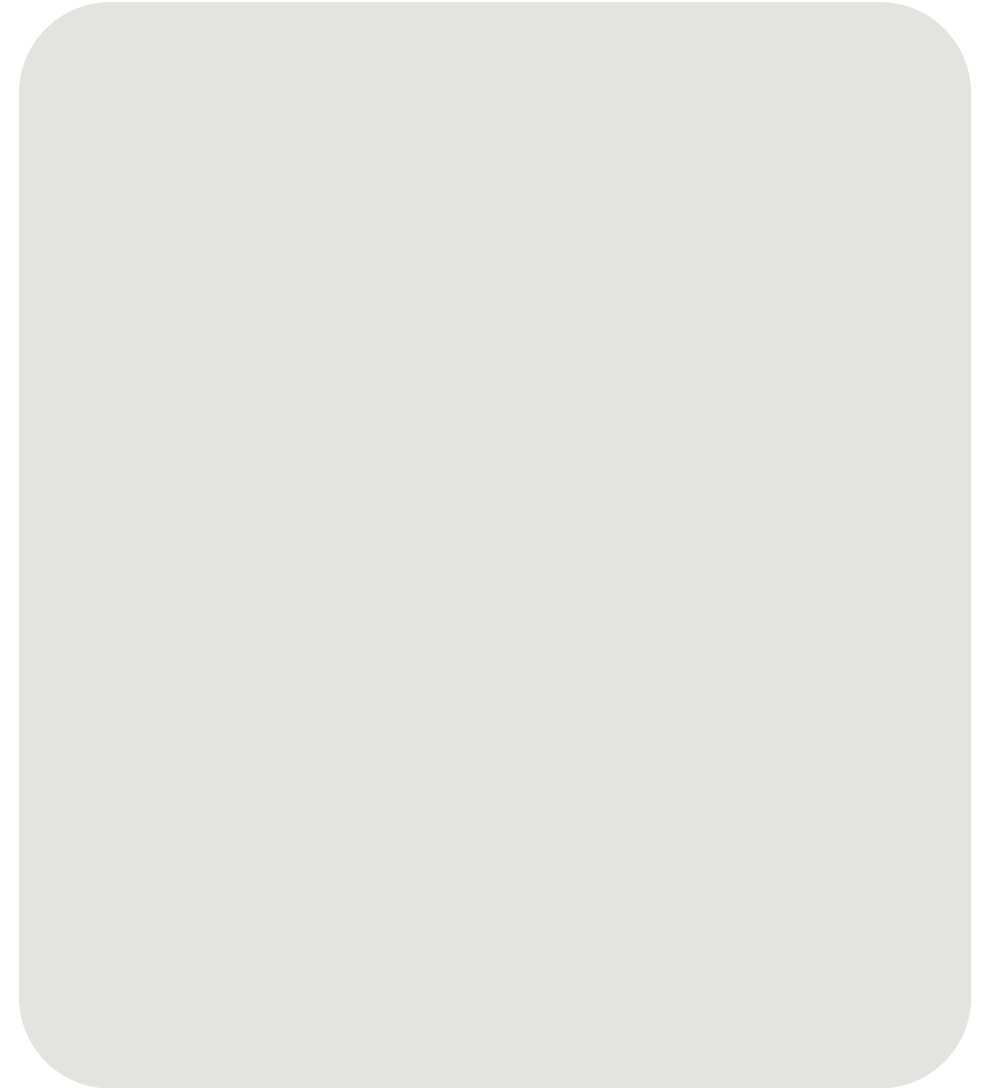
store nothing

always re-simulate

SimFS: A Simulation Data Virtualizing File System Interface



SimFS: A Simulation Data Virtualizing File System Interface



SimFS: A Simulation Data Virtualizing File System Interface

**1**

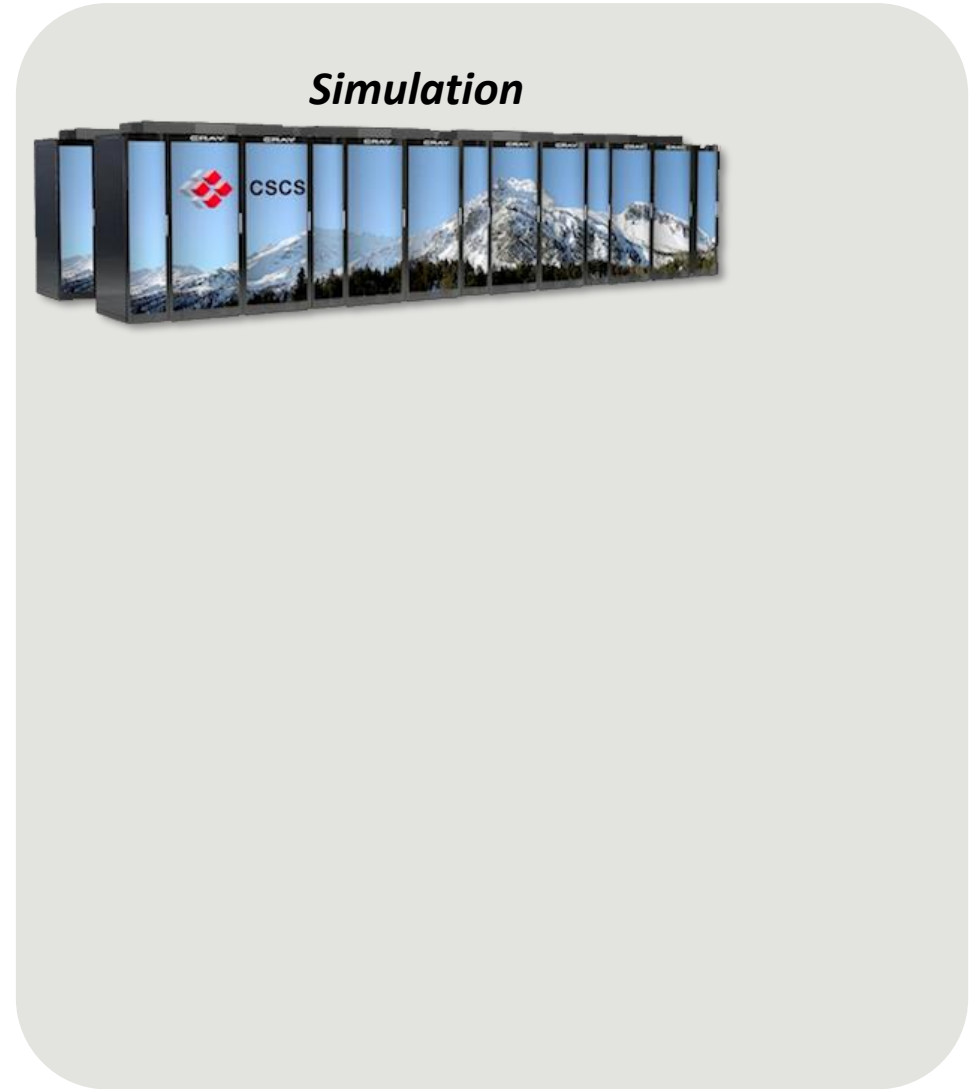
Run the simulation



SimFS: A Simulation Data Virtualizing File System Interface

**1**

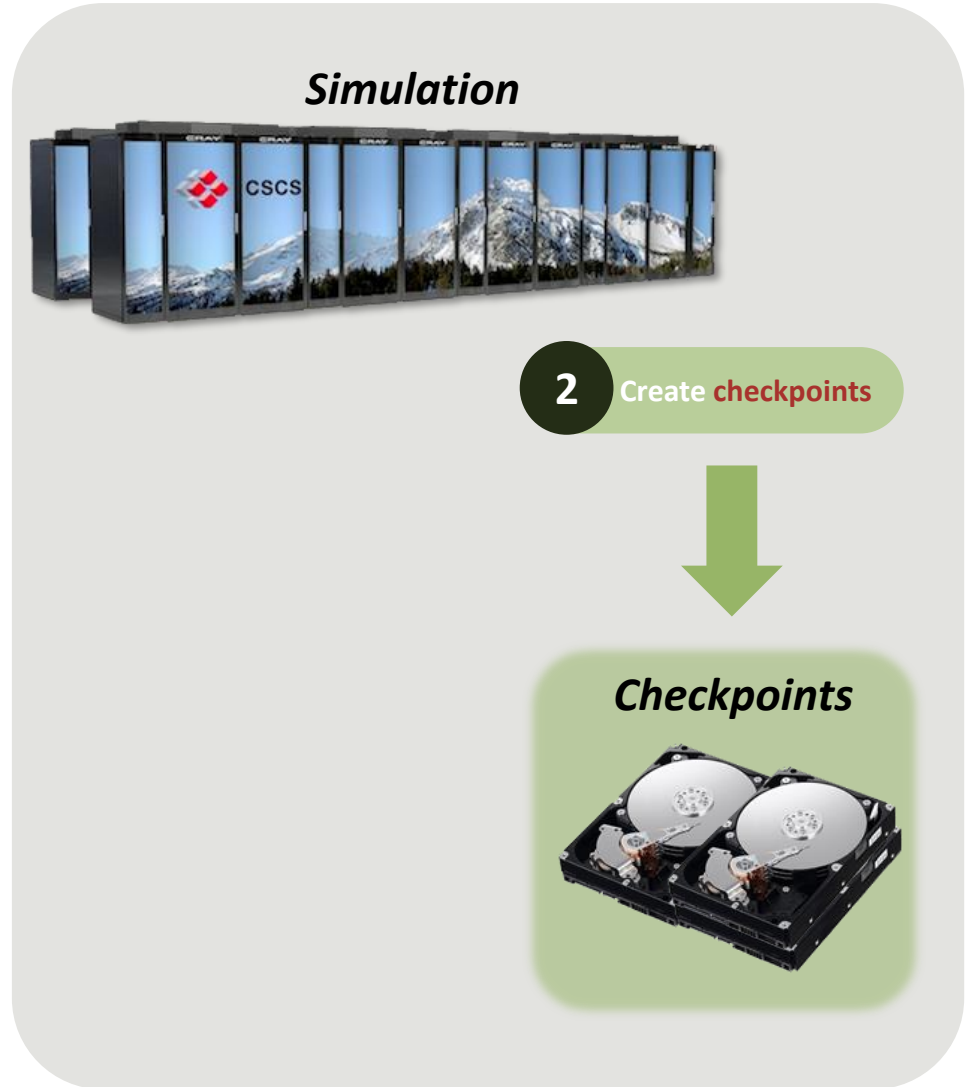
Run the simulation



SimFS: A Simulation Data Virtualizing File System Interface



1 Run the simulation



2 Create checkpoints

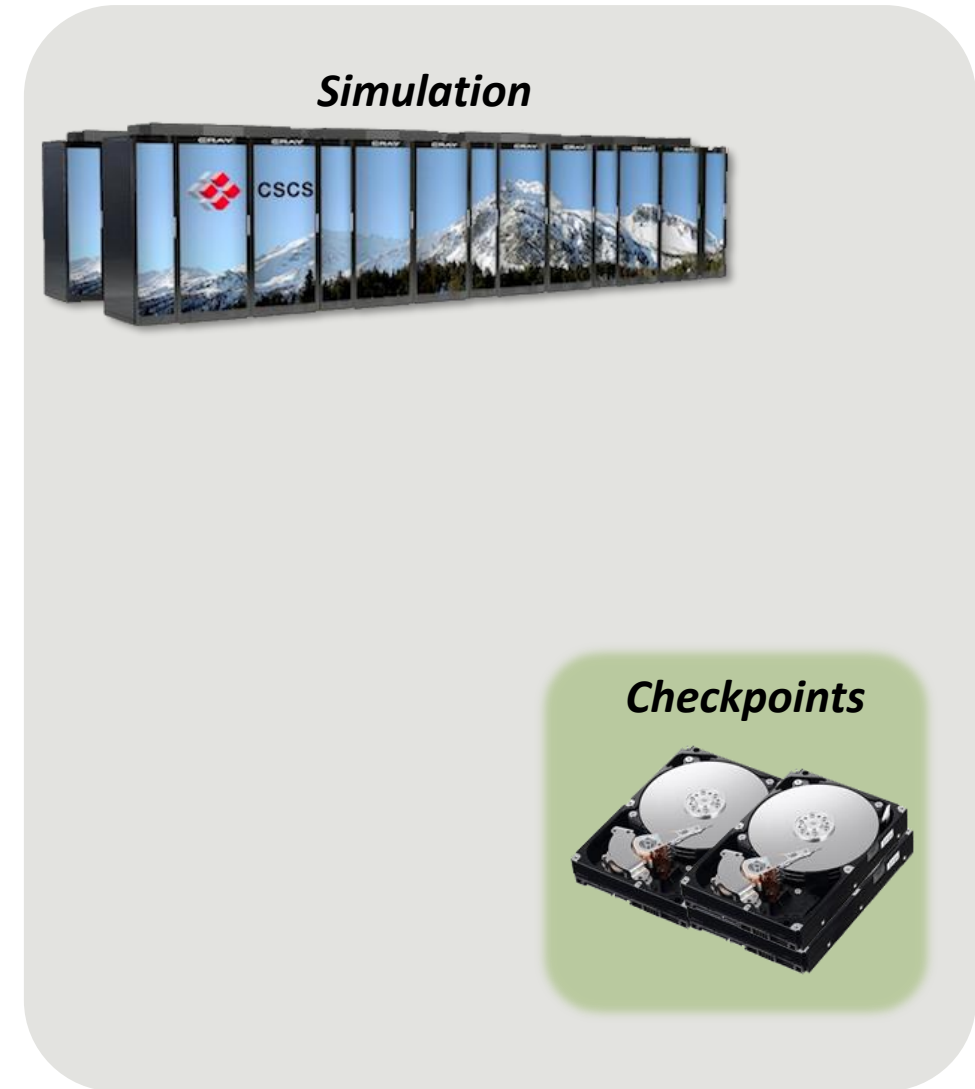
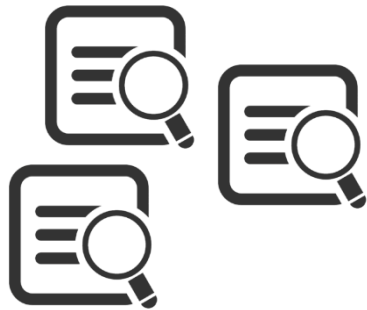


Checkpoints

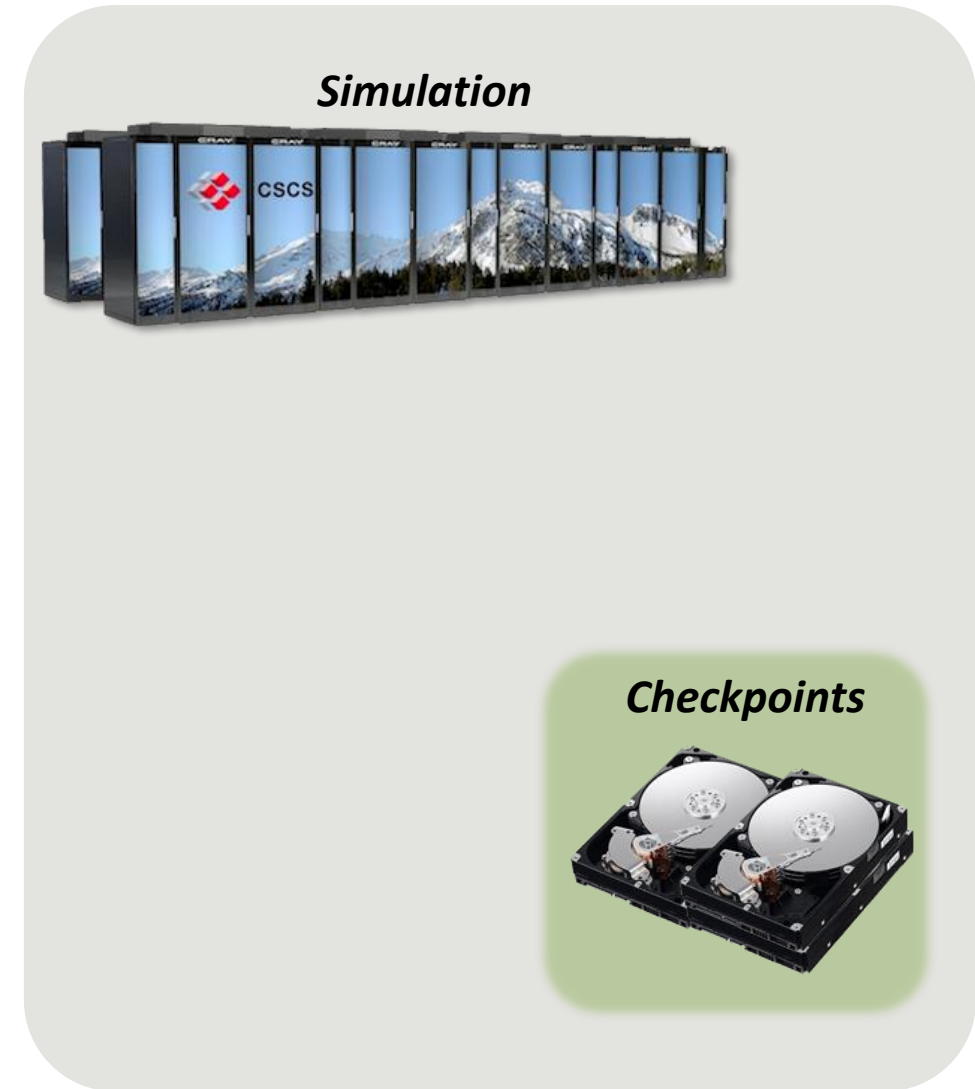


SimFS: A Simulation Data Virtualizing File System Interface

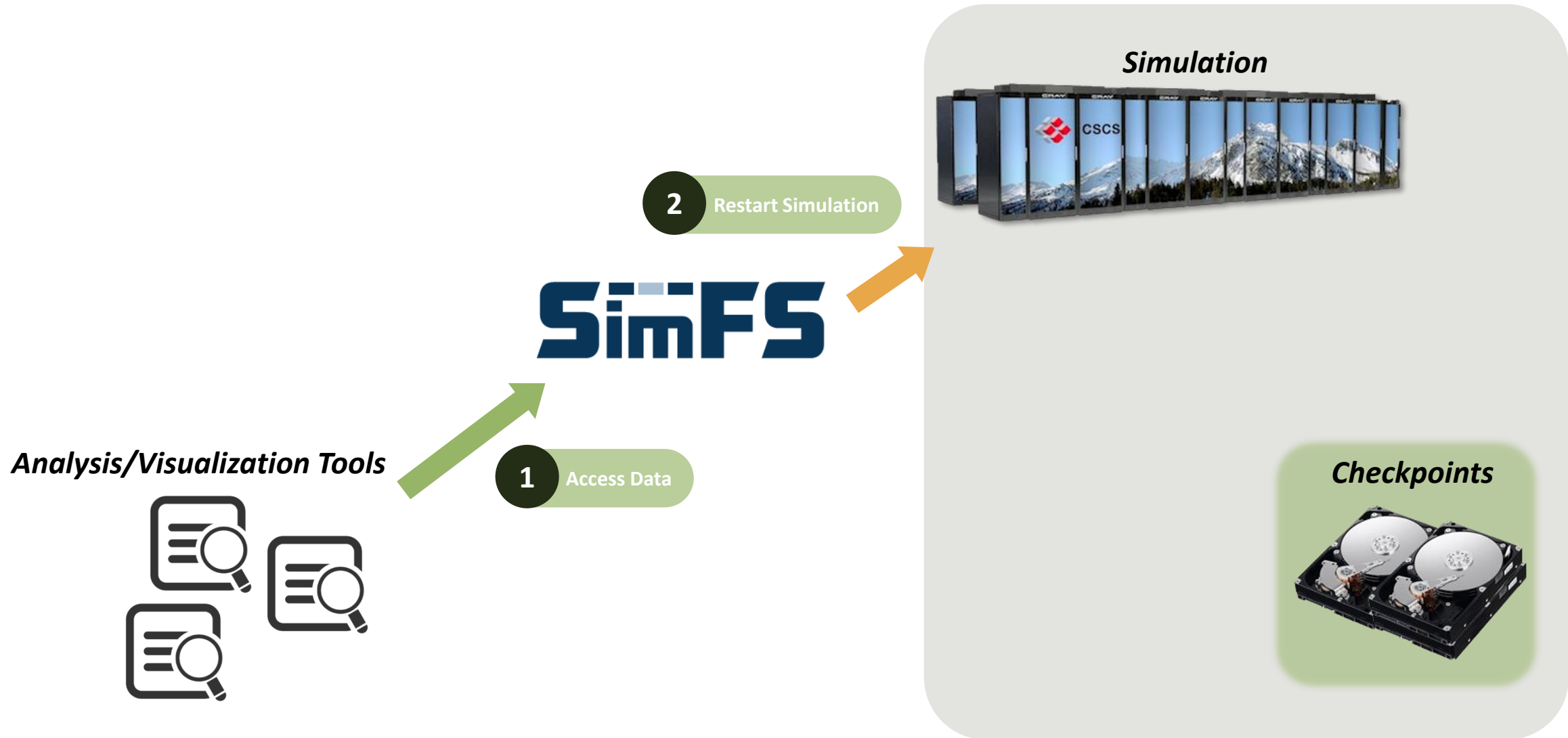
Analysis/Visualization Tools



SimFS: A Simulation Data Virtualizing File System Interface



SimFS: A Simulation Data Virtualizing File System Interface



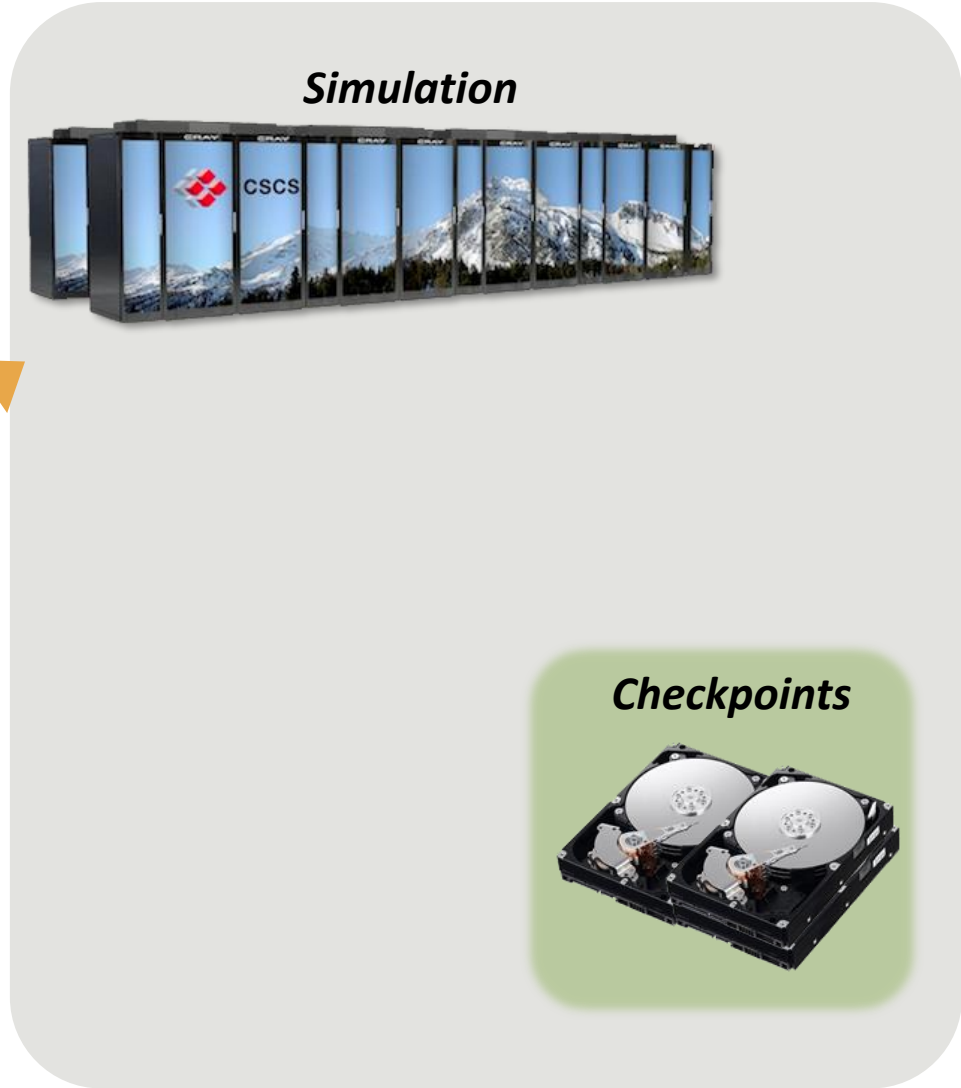
SimFS: A Simulation Data Virtualizing File System Interface



Checkpoint
 Requested
 Re-Simulated

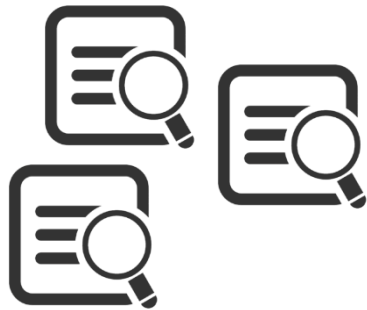
Simulation Time →

2 Restart Simulation



SimFS

Analysis/Visualization Tools



1 Access Data

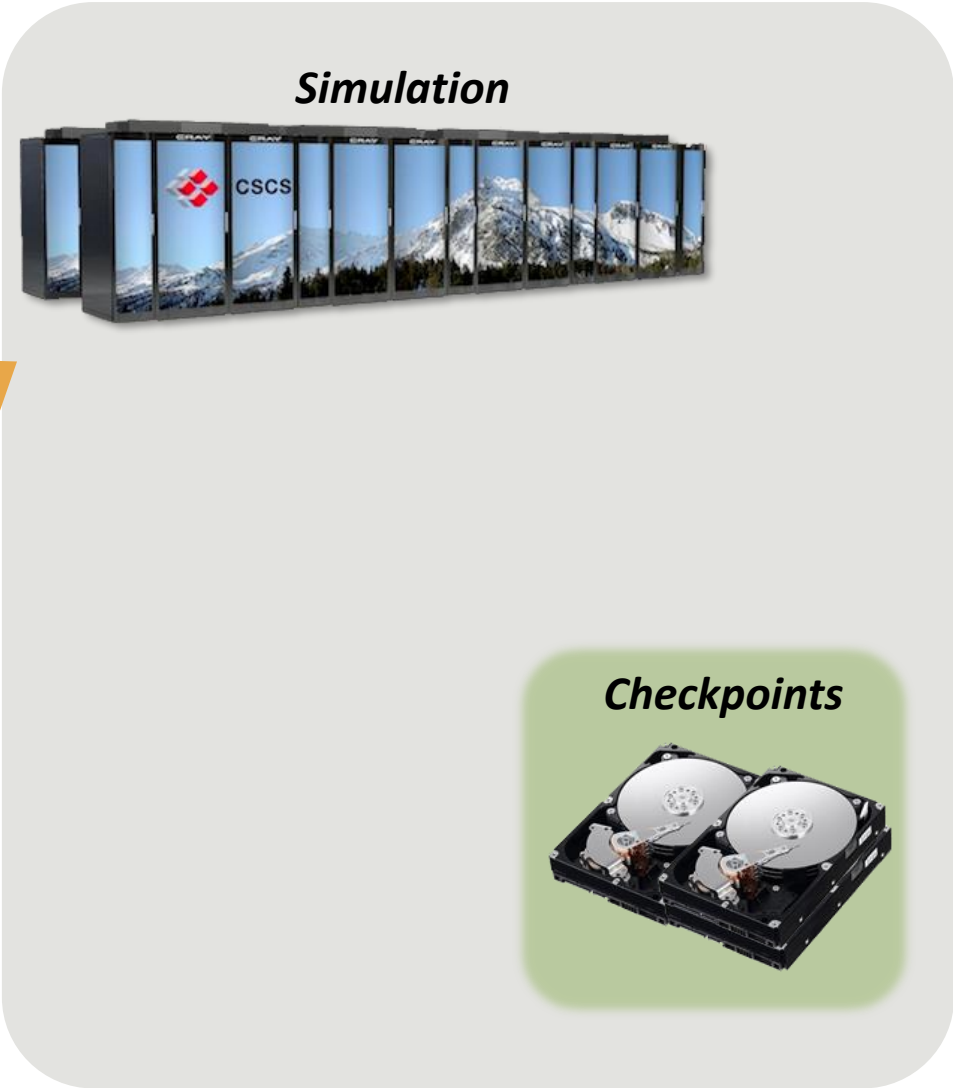
SimFS: A Simulation Data Virtualizing File System Interface



Checkpoint
 Requested
 Re-Simulated

Simulation Time →

2 Restart Simulation



SimFS

Analysis/Visualization Tools



1 Access Data

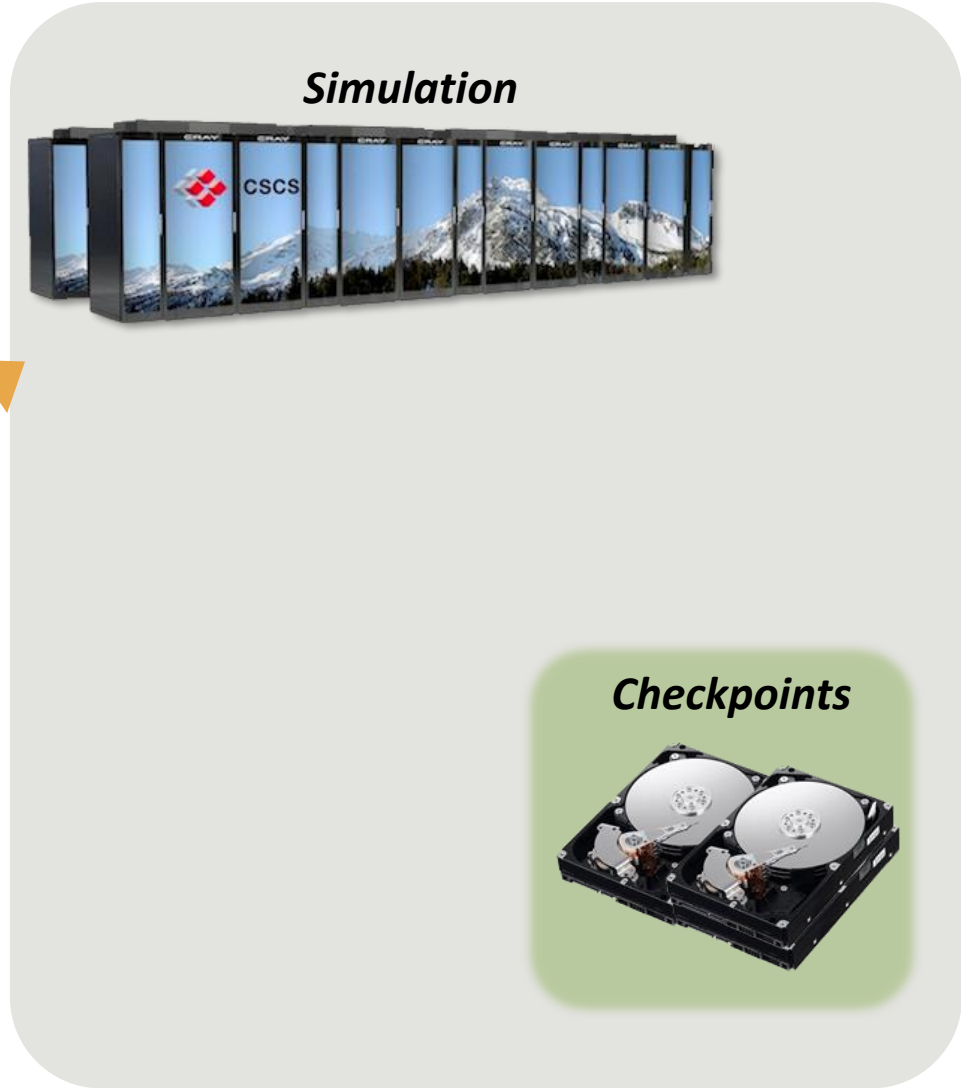
SimFS: A Simulation Data Virtualizing File System Interface



Checkpoint
 Requested
 Re-Simulated

Simulation Time →

2 Restart Simulation



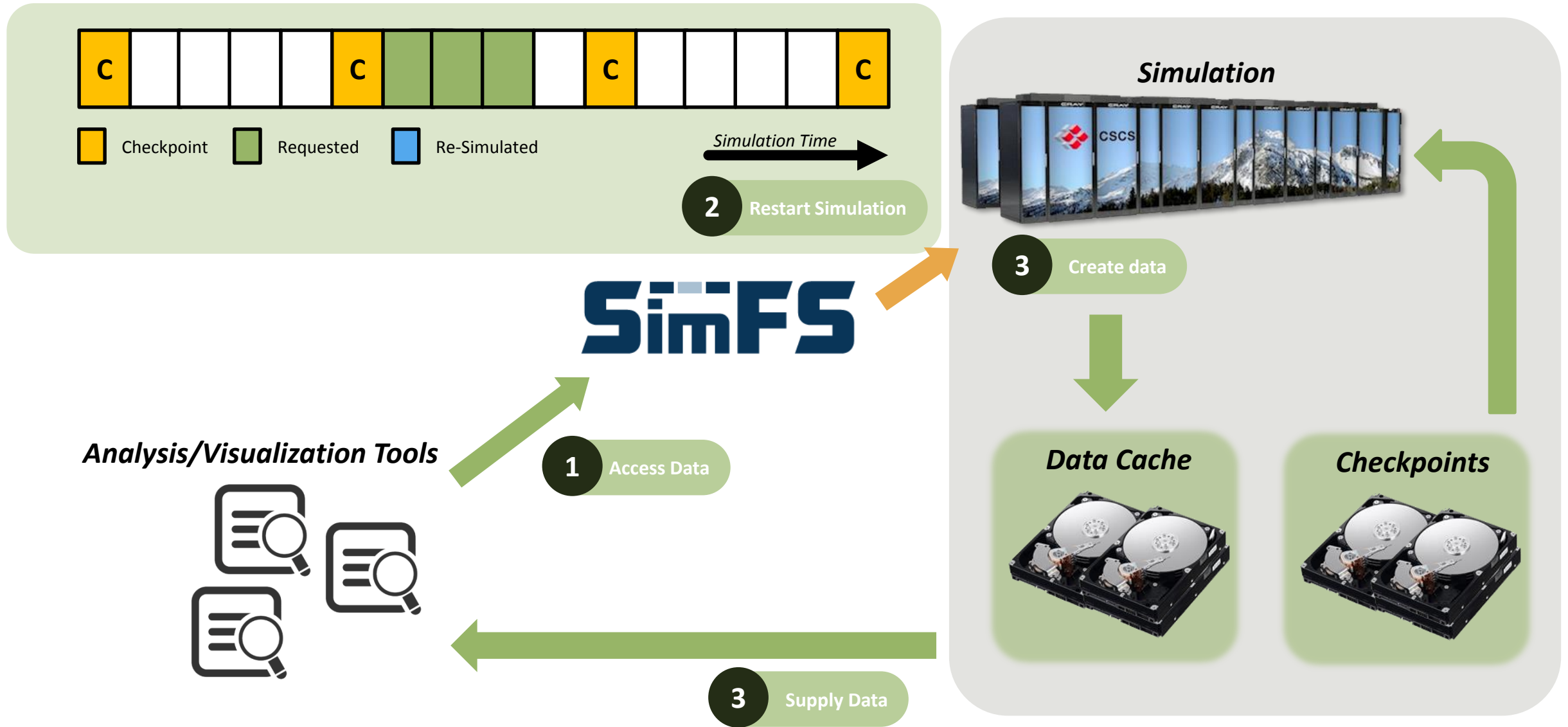
SimFS

Analysis/Visualization Tools



1 Access Data

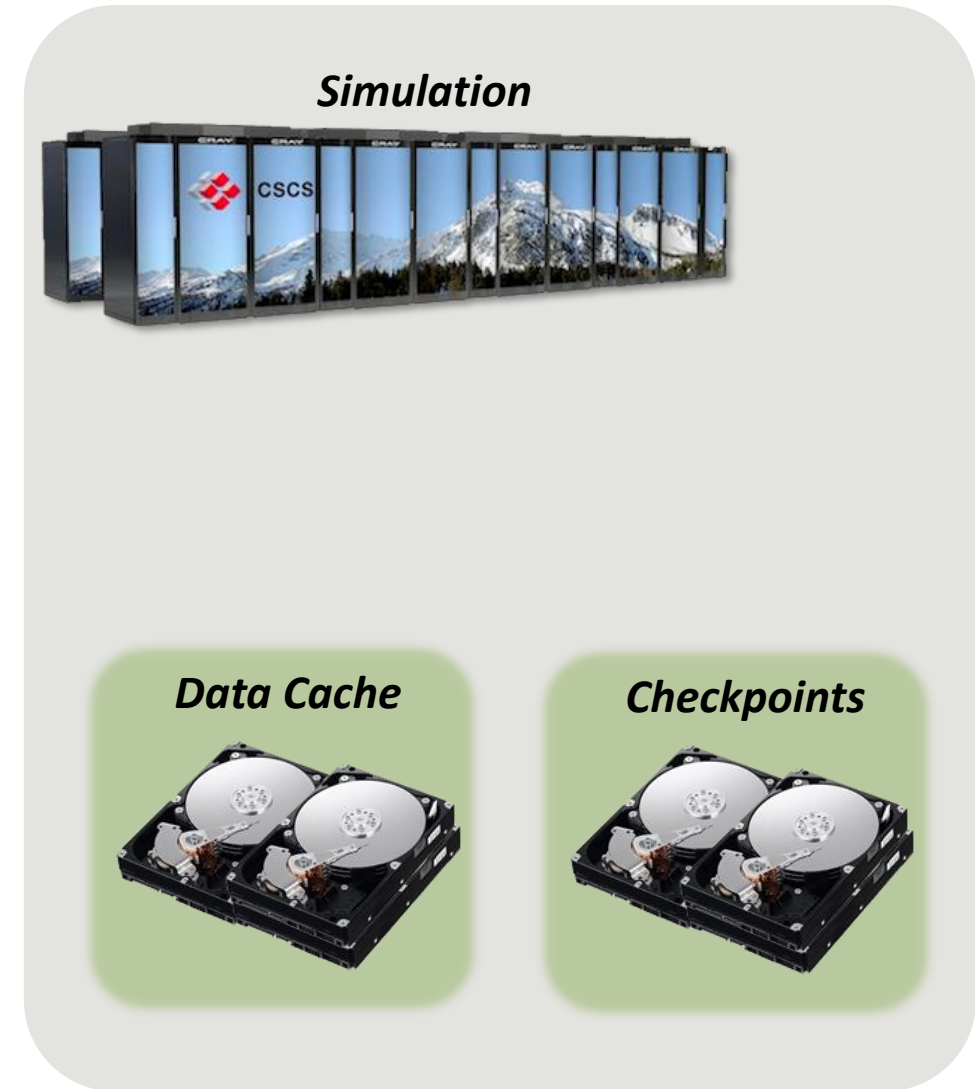
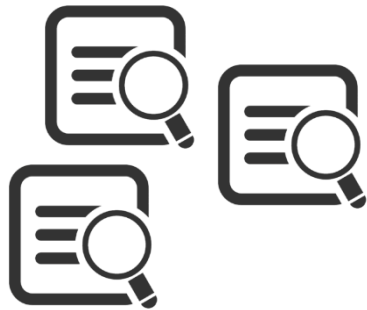
SimFS: A Simulation Data Virtualizing File System Interface



SimFS: A Simulation Data Virtualizing File System Interface

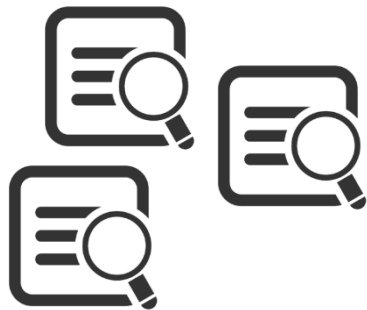


Analysis/Visualization Tools



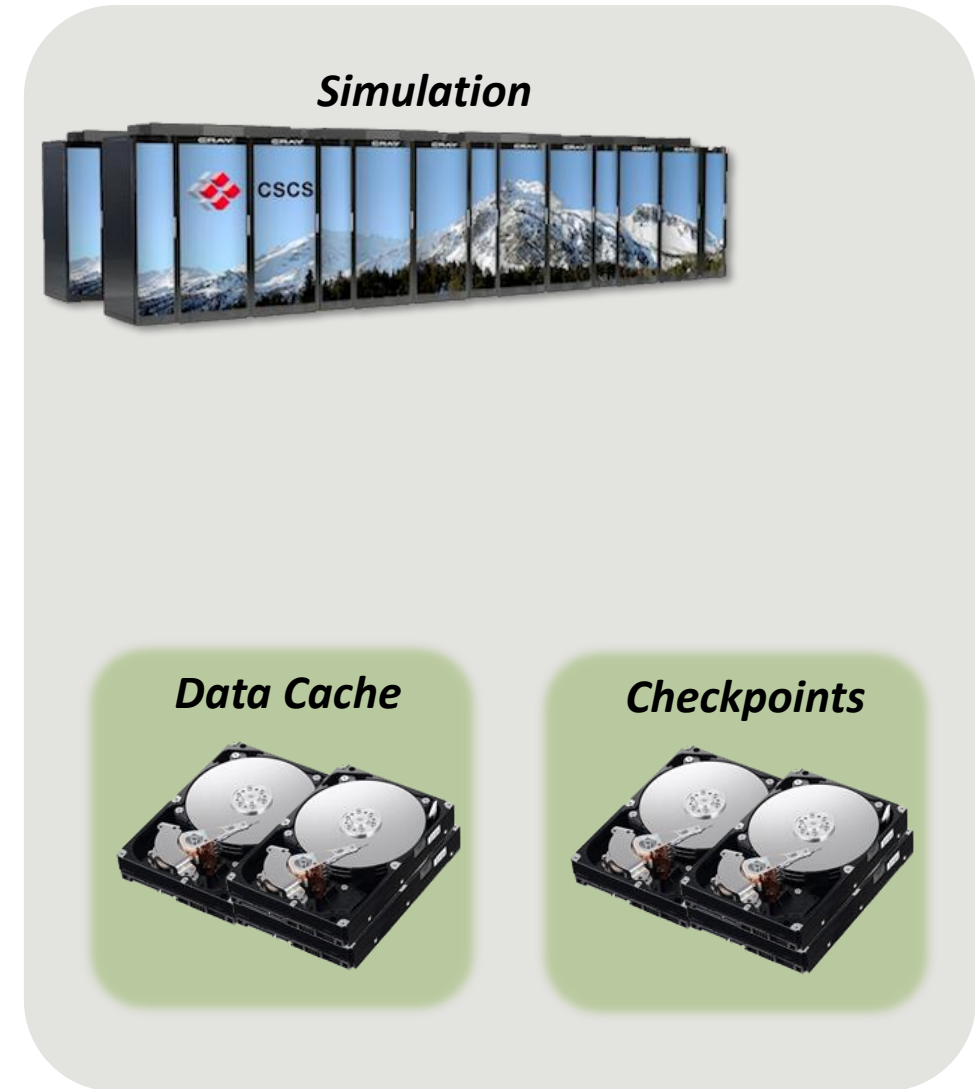
SimFS: A Simulation Data Virtualizing File System Interface

Analysis/Visualization Tools

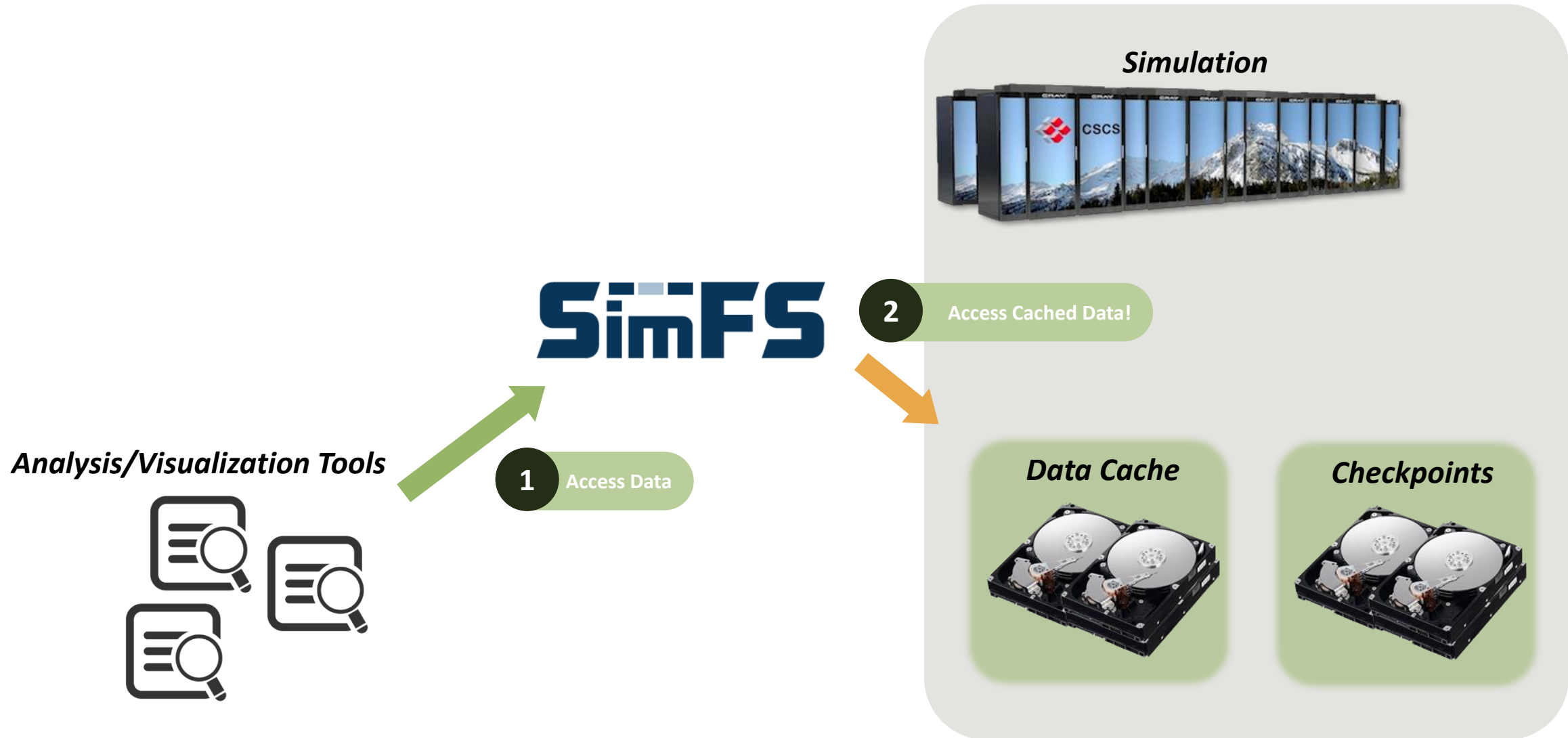


1 Access Data

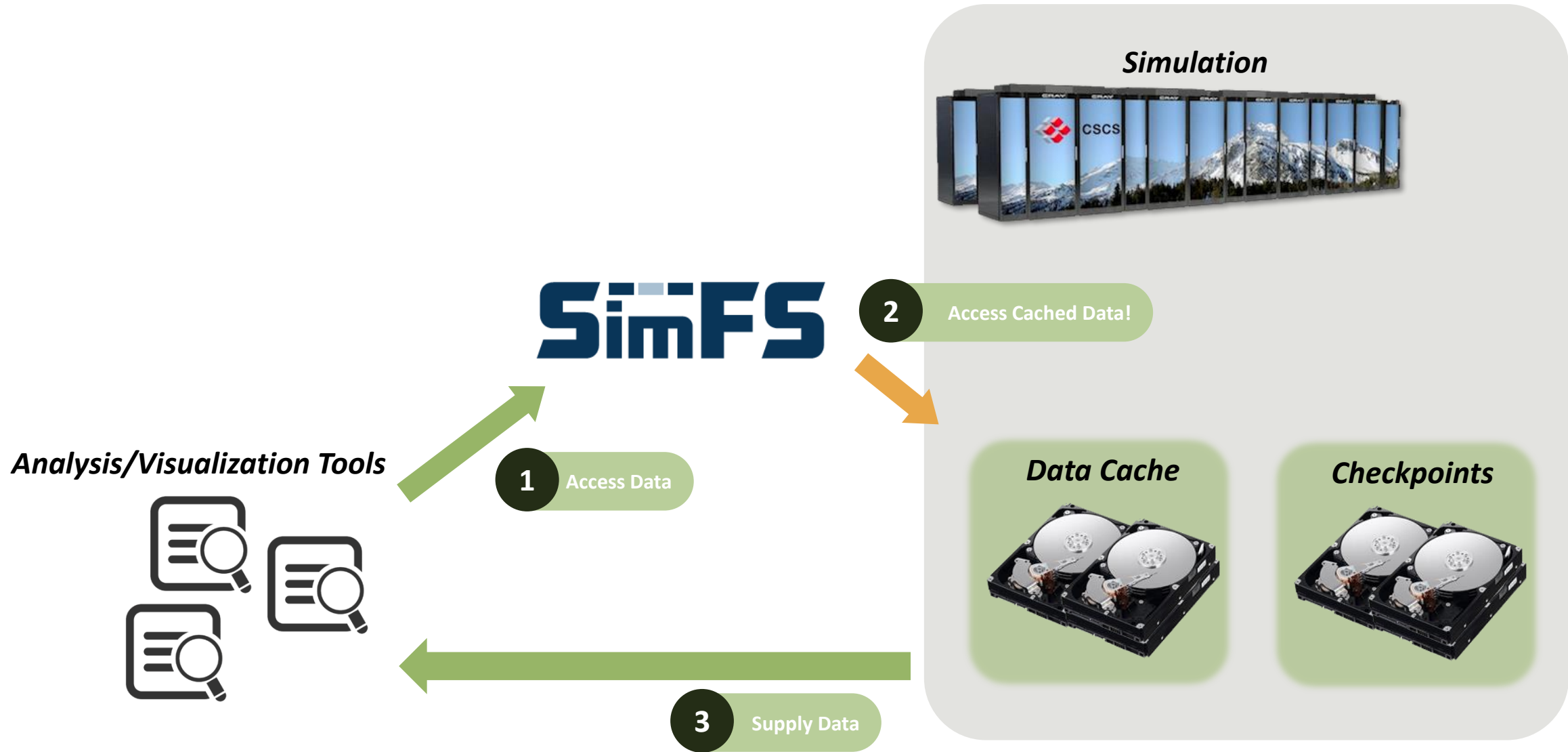
SimFS



SimFS: A Simulation Data Virtualizing File System Interface



SimFS: A Simulation Data Virtualizing File System Interface



SimFS: A Simulation Data Virtualizing File System Interface



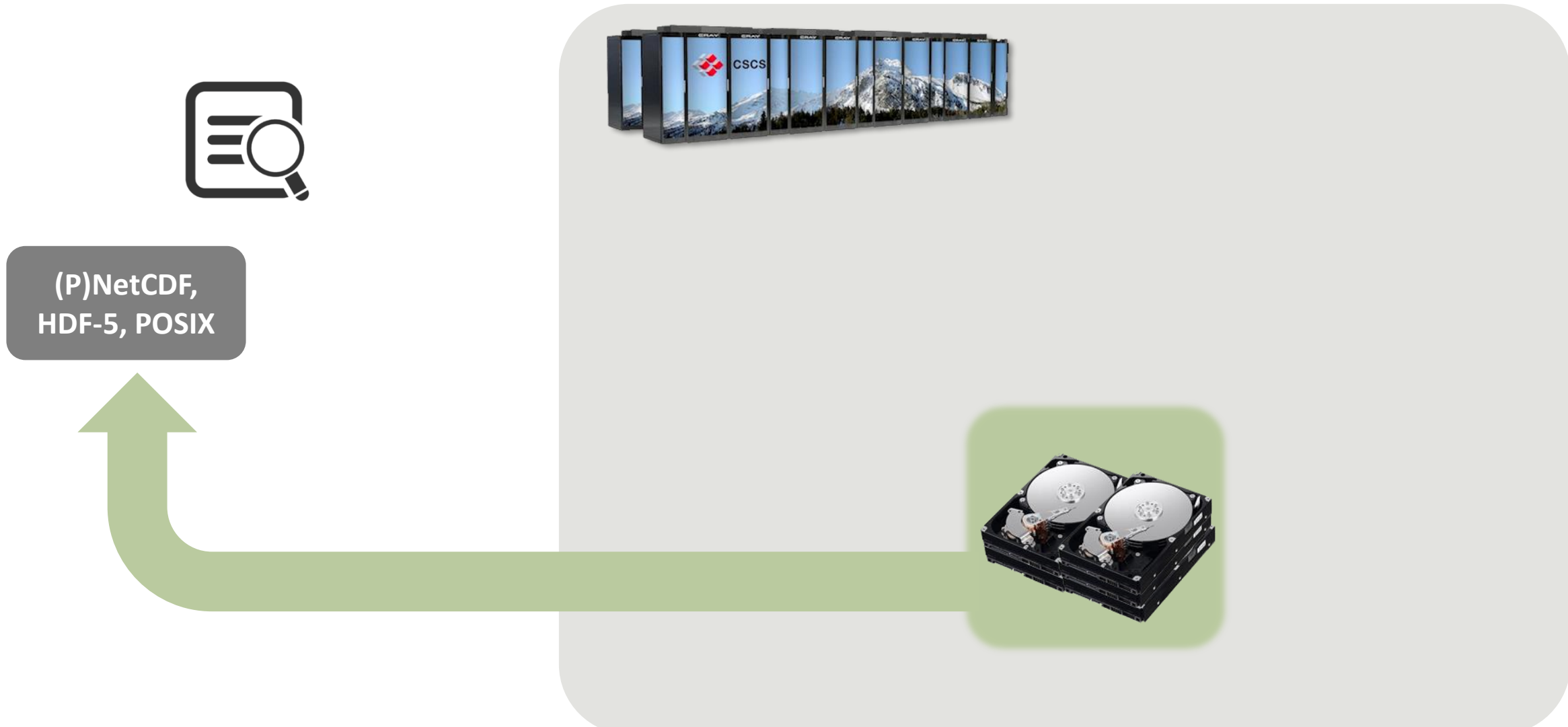
SimFS: A Simulation Data Virtualizing File System Interface



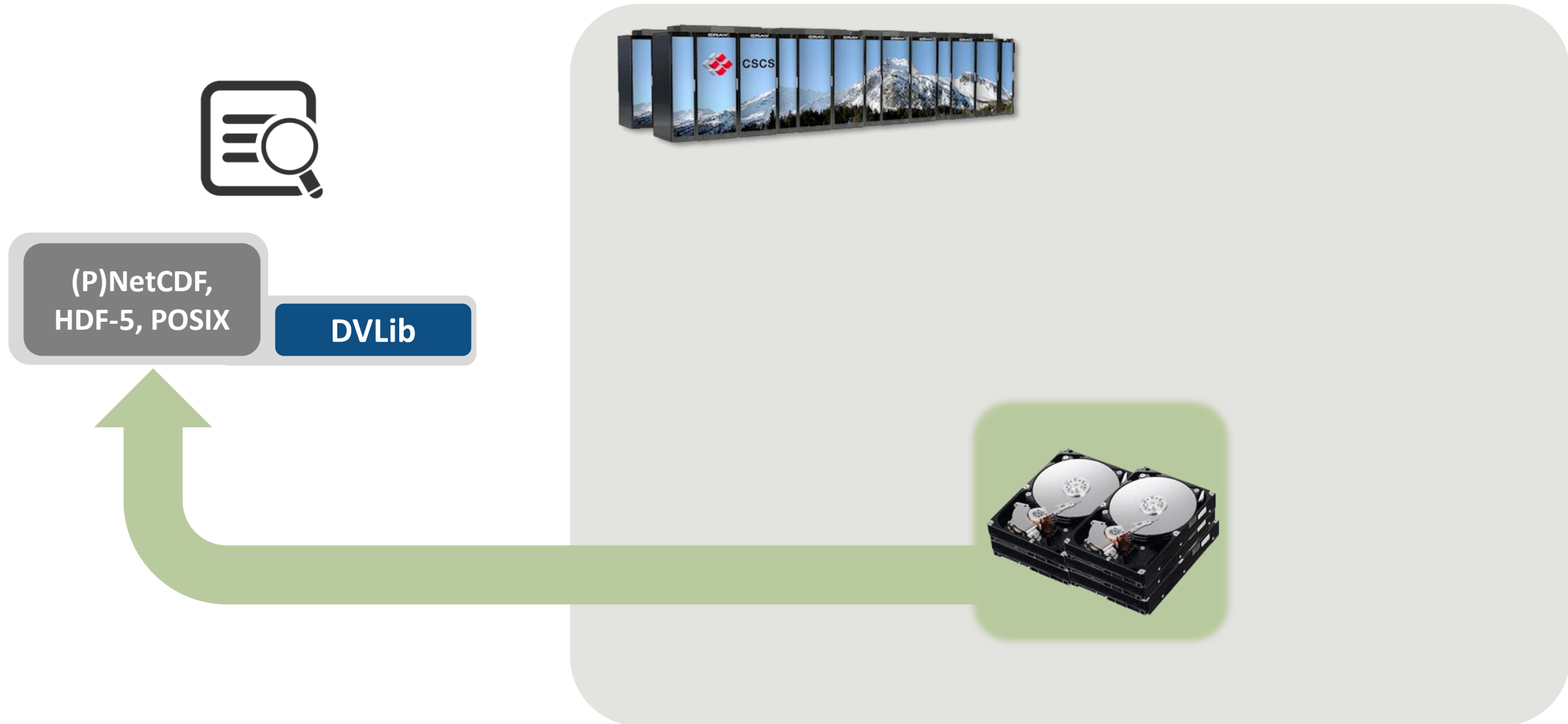
(P)NetCDF,
HDF-5, POSIX



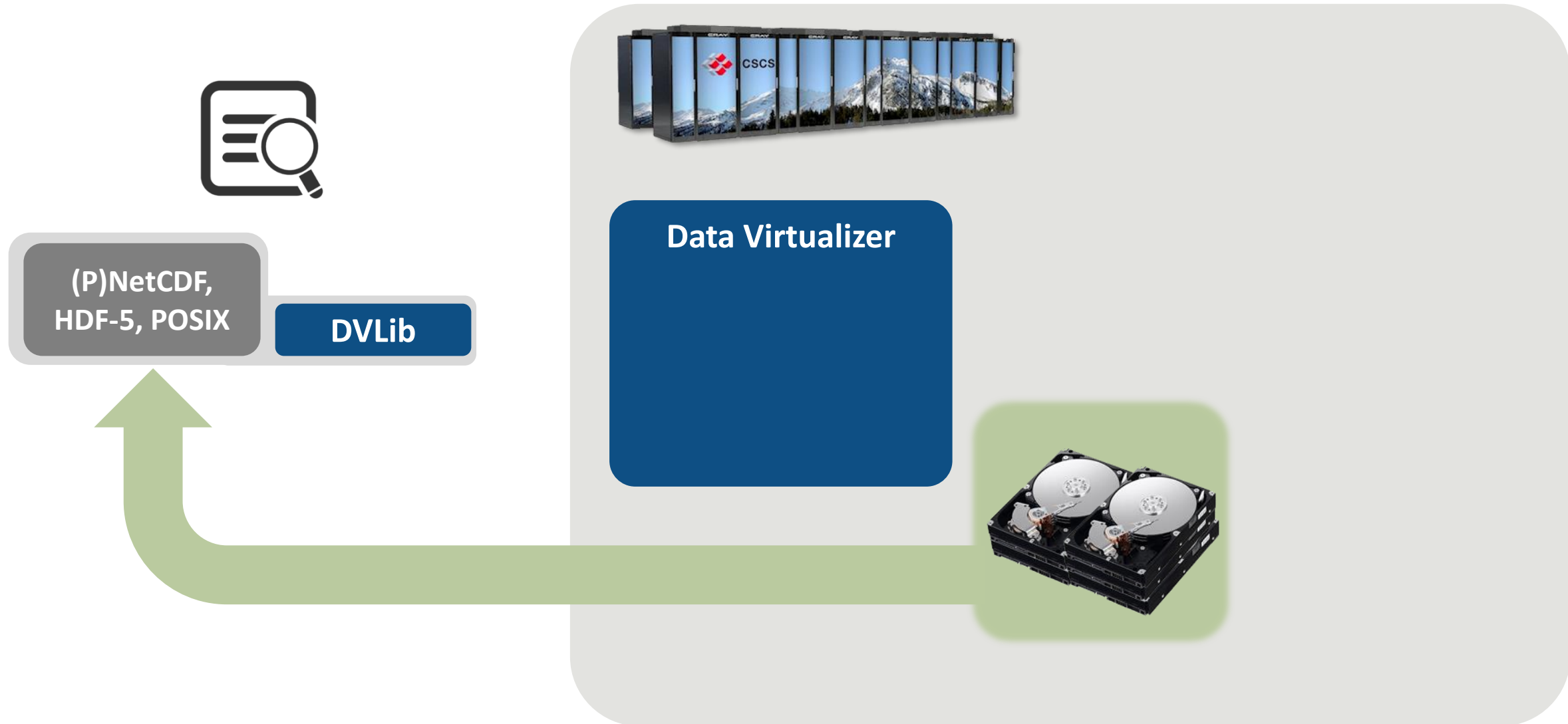
SimFS: A Simulation Data Virtualizing File System Interface



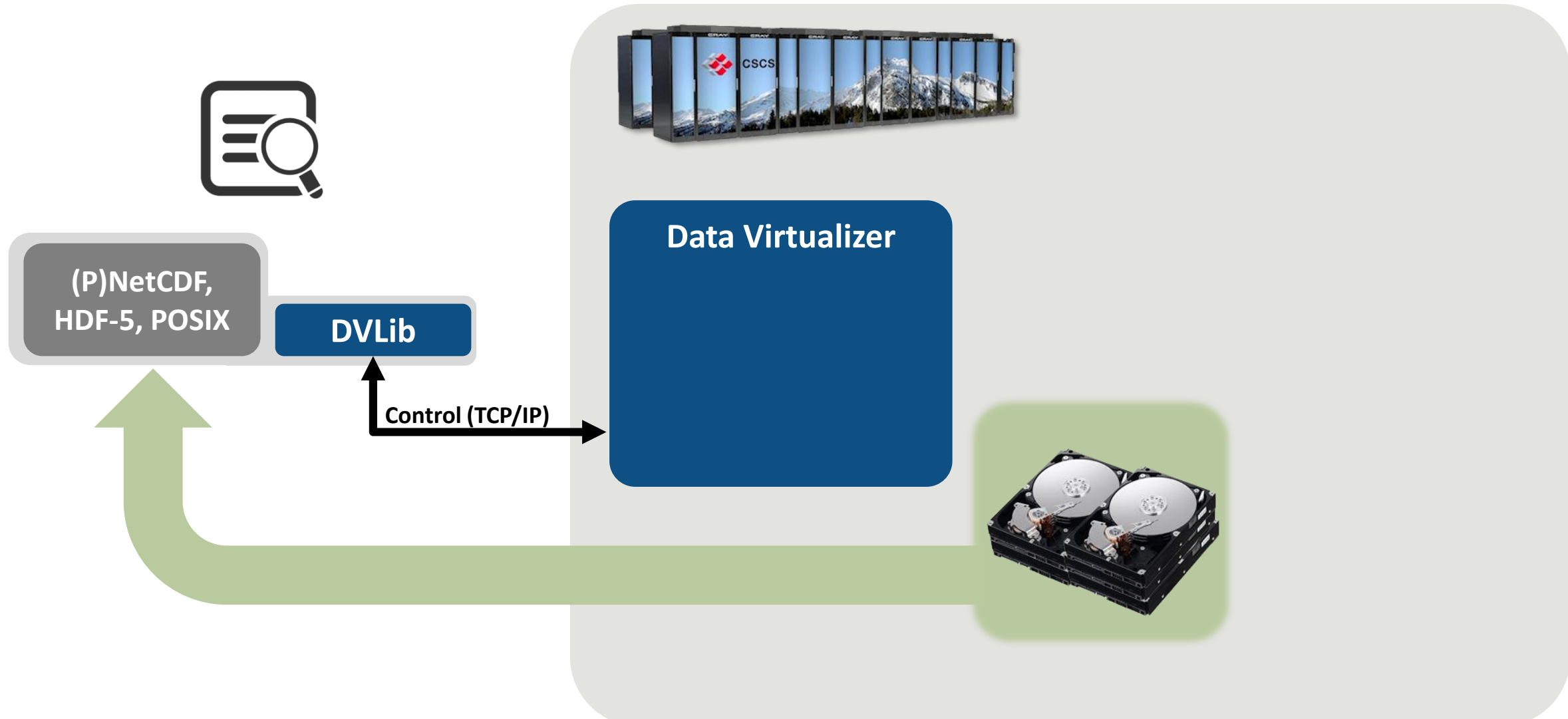
SimFS: A Simulation Data Virtualizing File System Interface



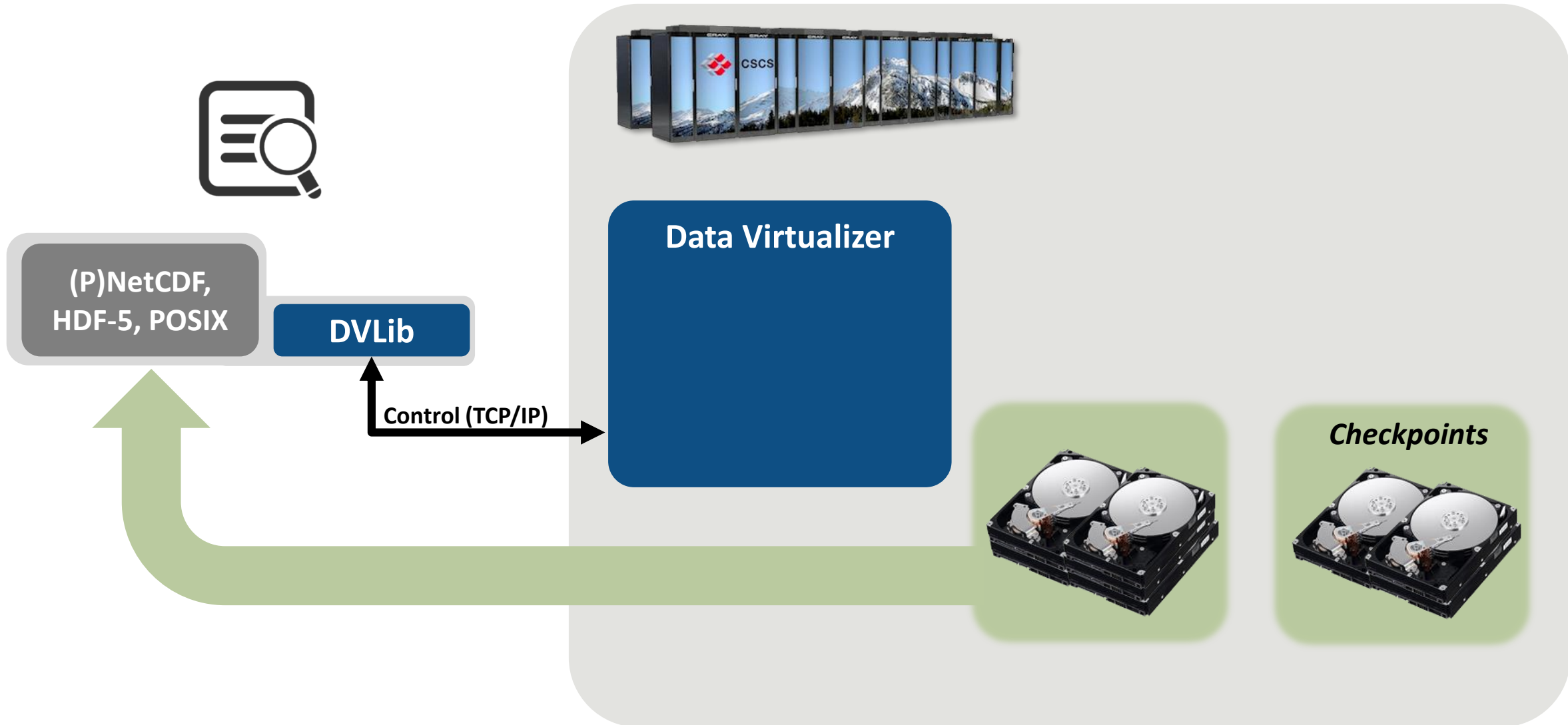
SimFS: A Simulation Data Virtualizing File System Interface



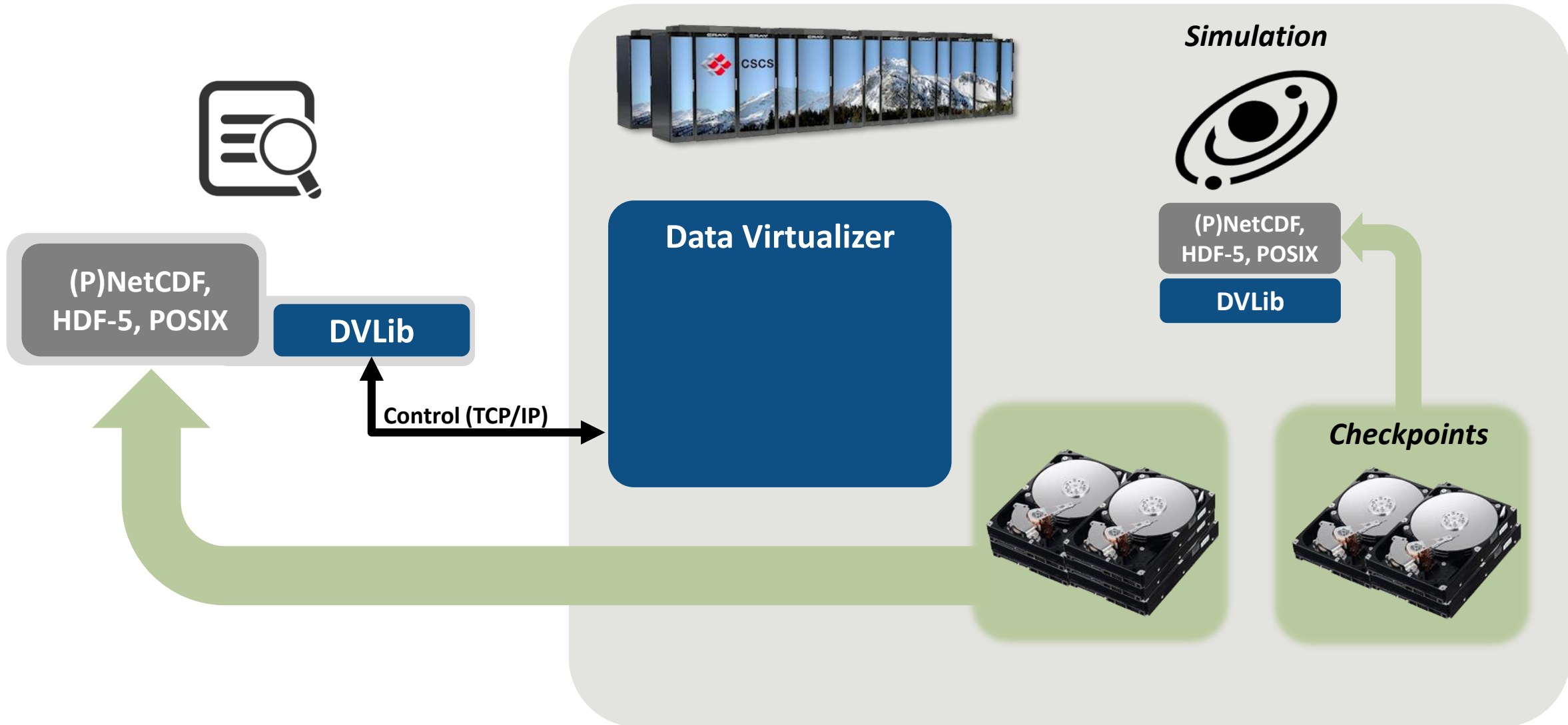
SimFS: A Simulation Data Virtualizing File System Interface



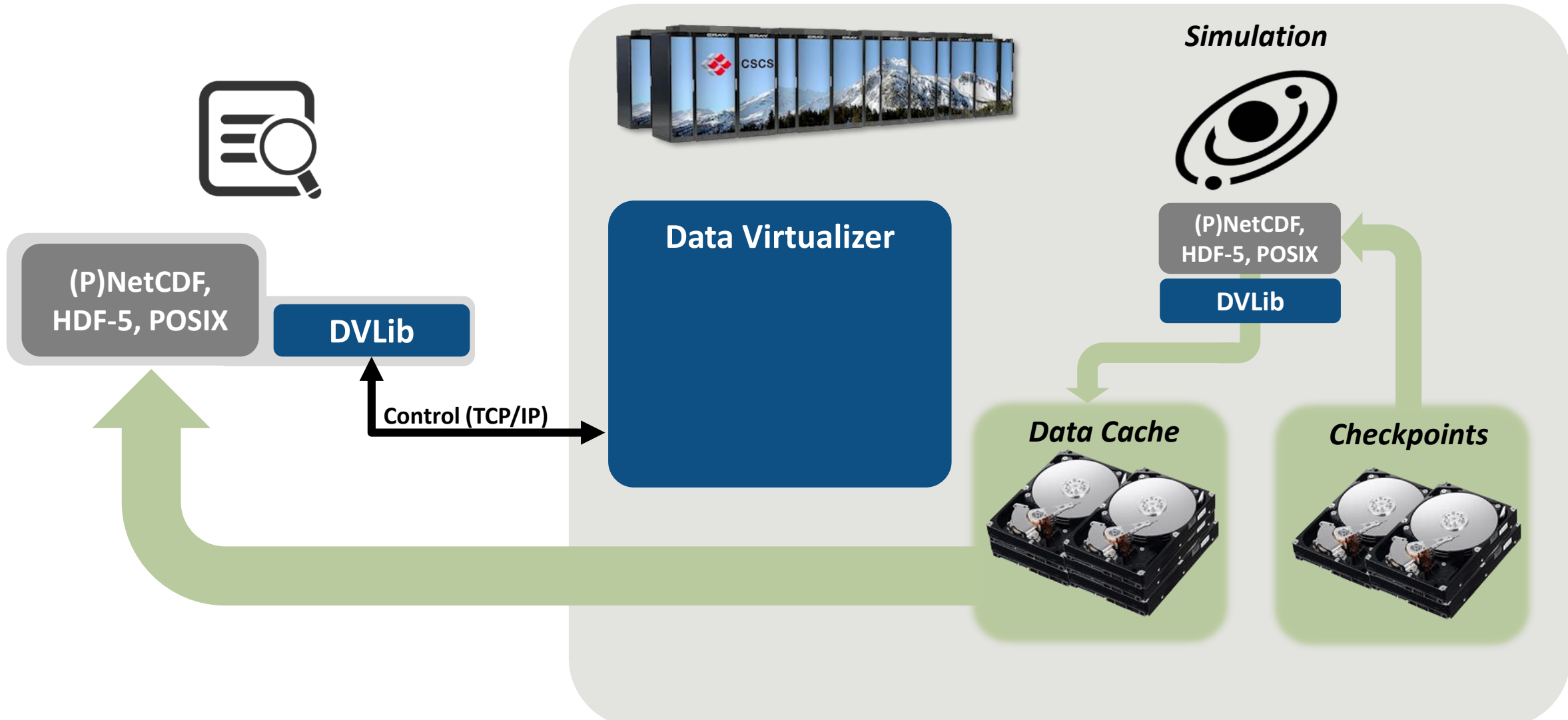
SimFS: A Simulation Data Virtualizing File System Interface



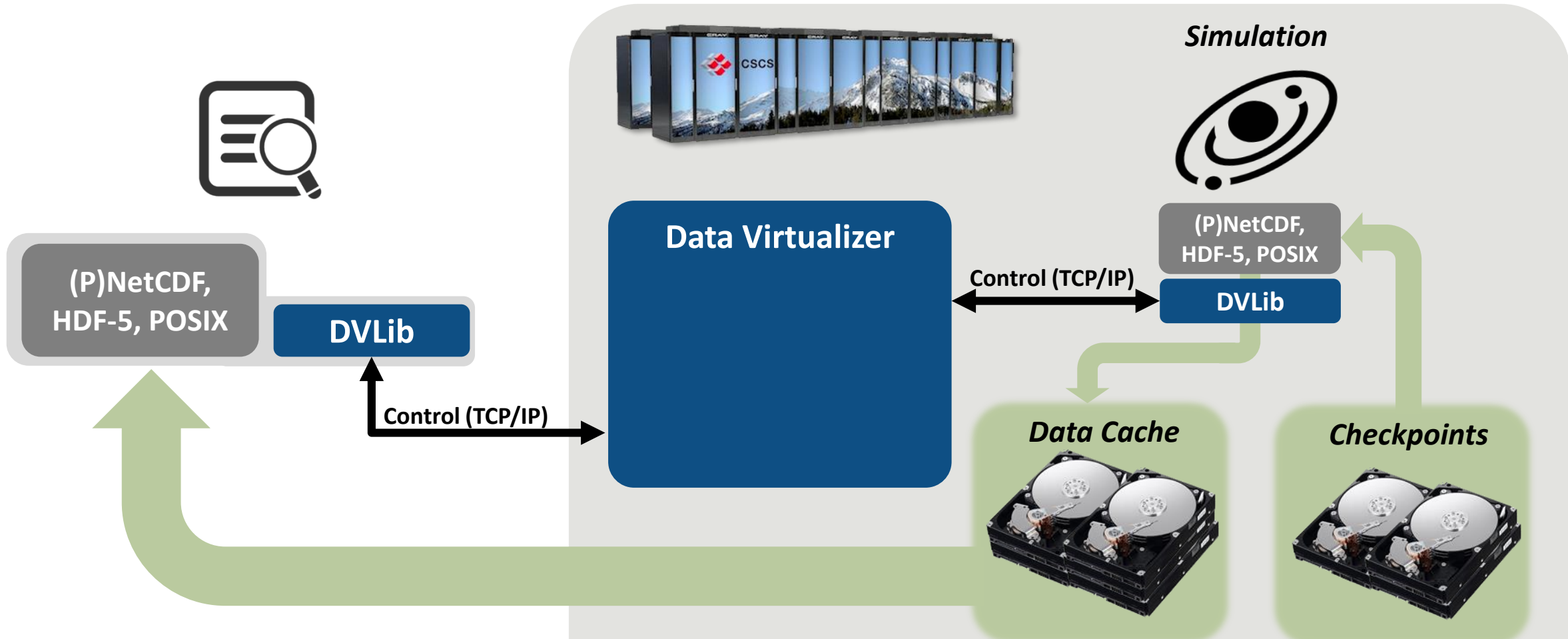
SimFS: A Simulation Data Virtualizing File System Interface



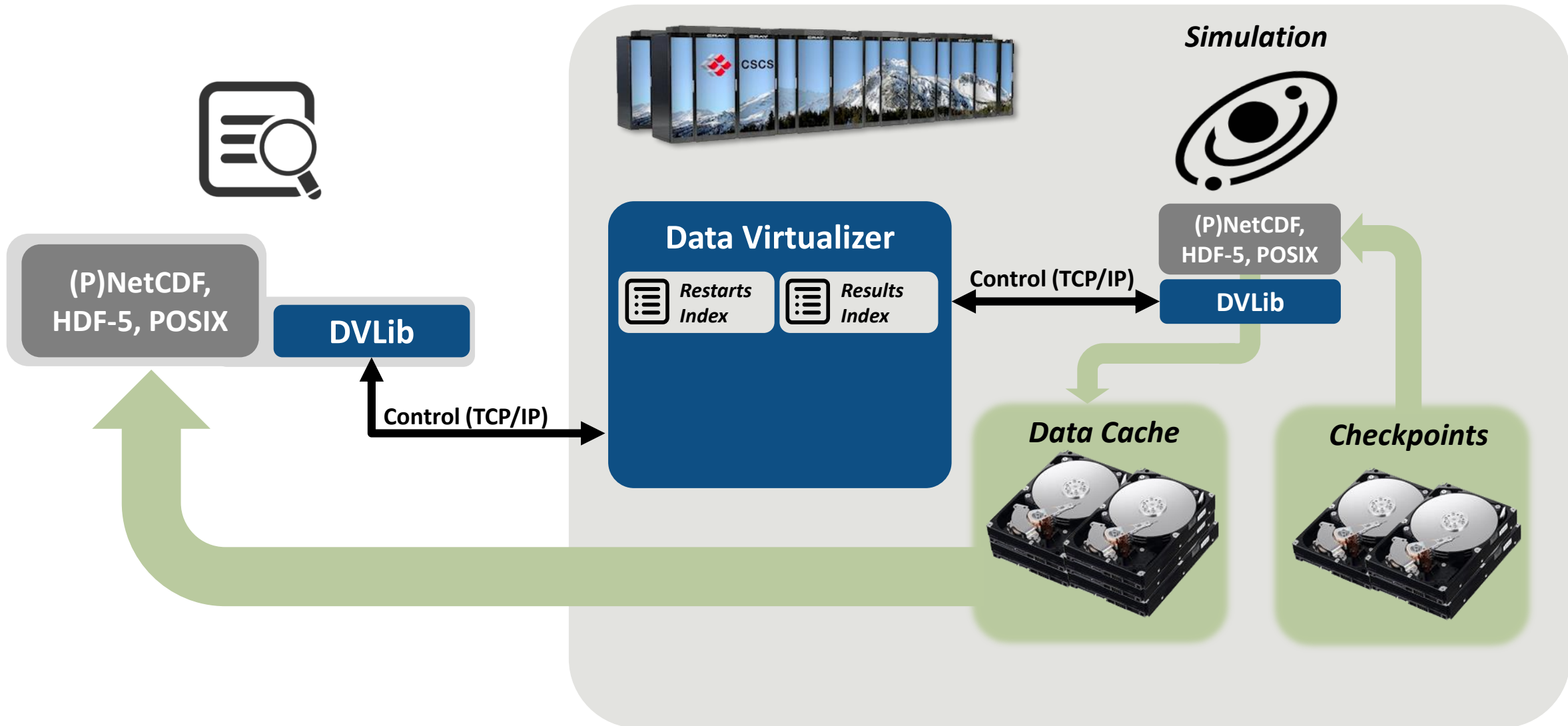
SimFS: A Simulation Data Virtualizing File System Interface



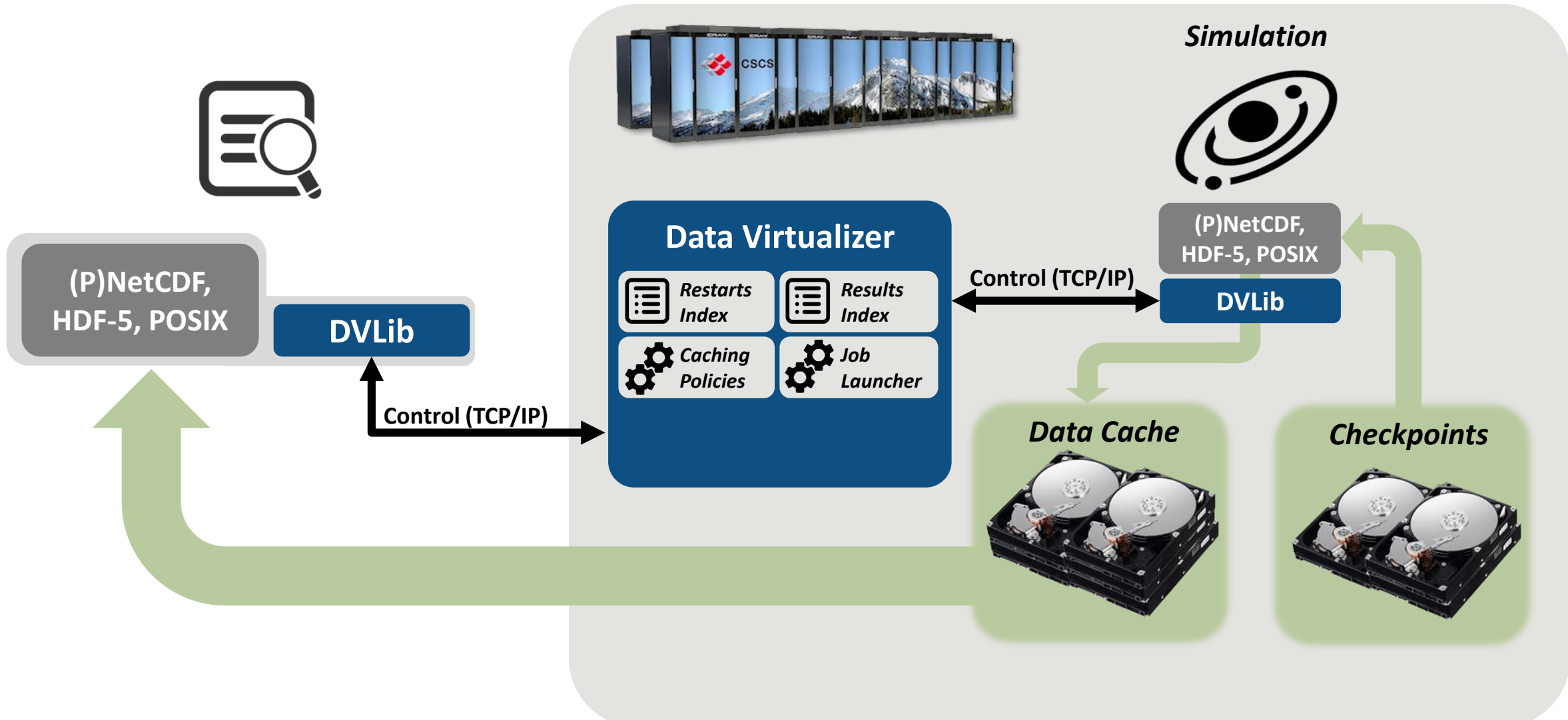
SimFS: A Simulation Data Virtualizing File System Interface



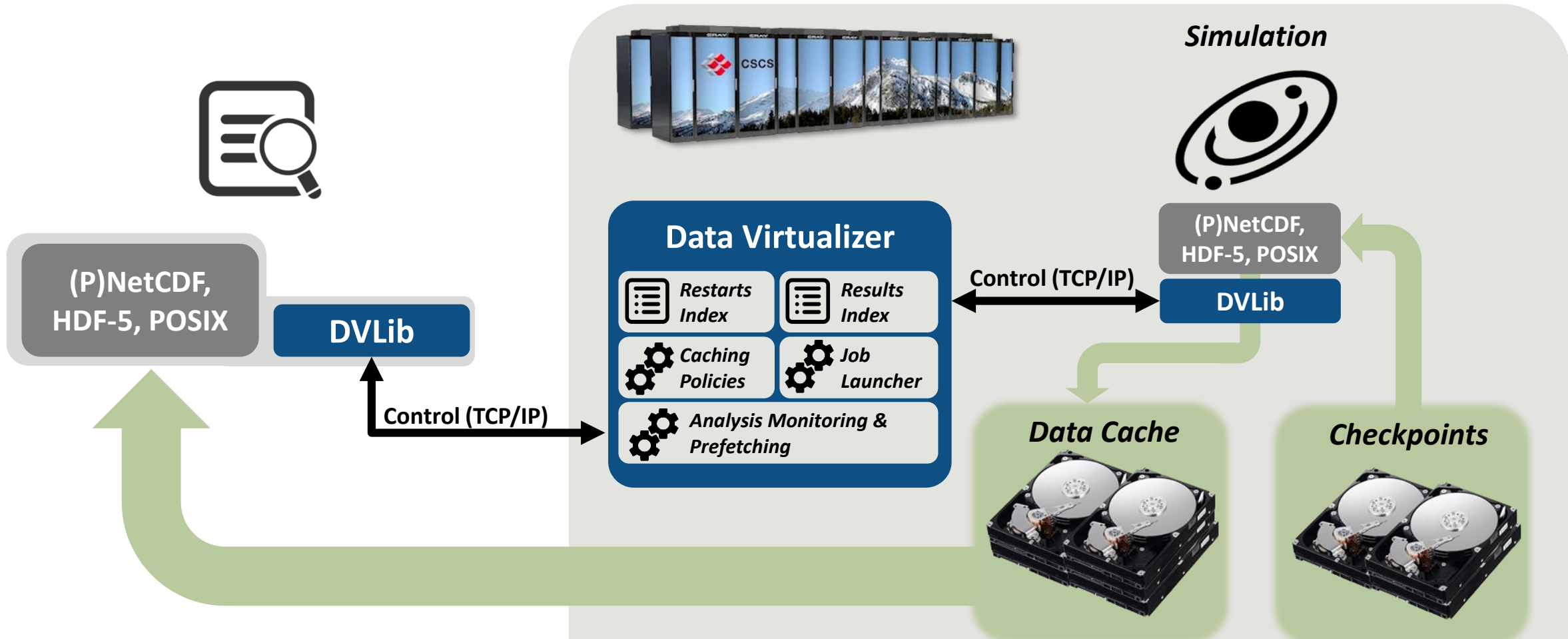
SimFS: A Simulation Data Virtualizing File System Interface



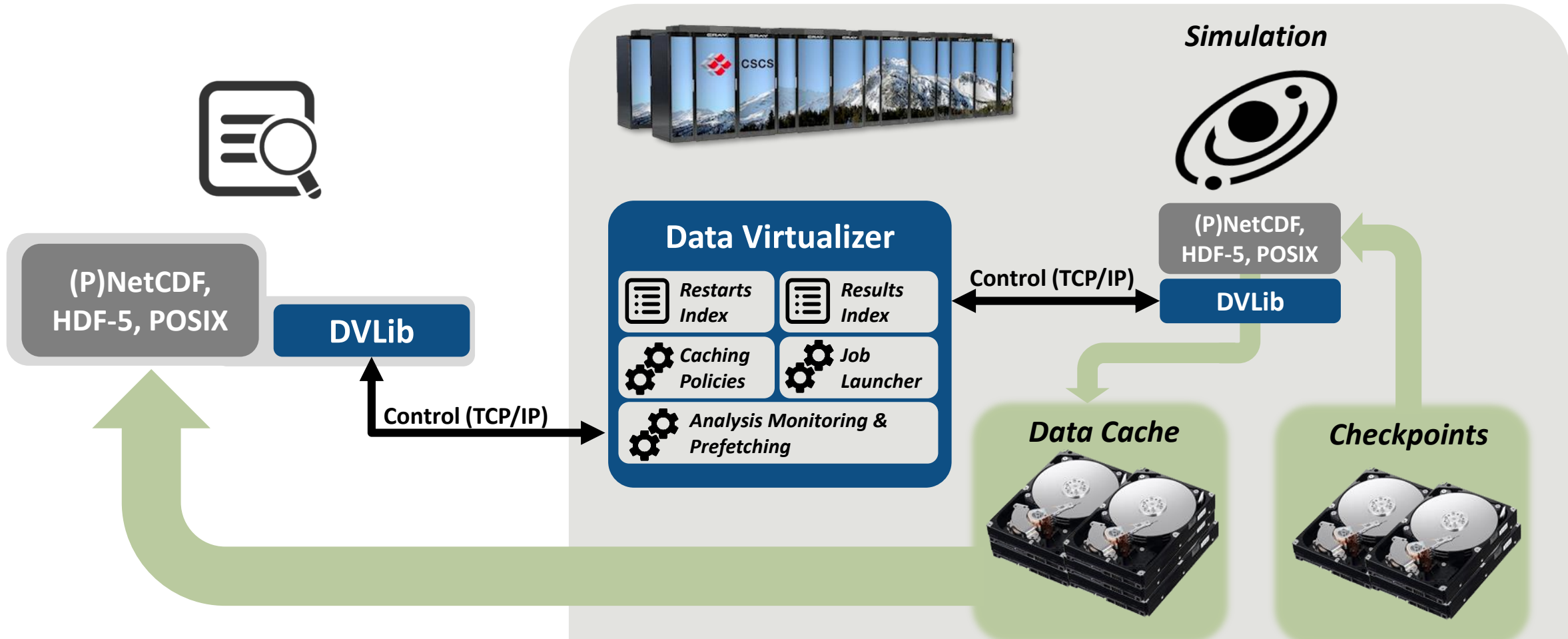
SimFS: A Simulation Data Virtualizing File System Interface



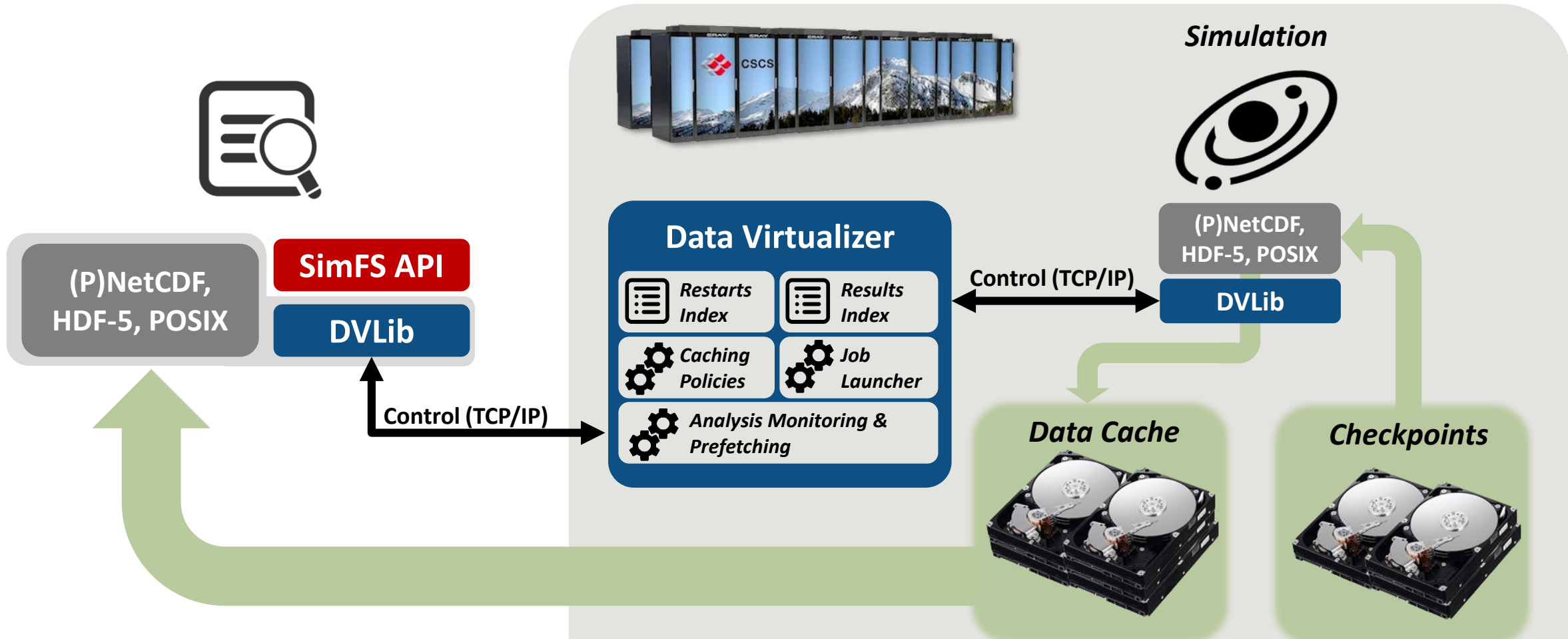
SimFS: A Simulation Data Virtualizing File System Interface



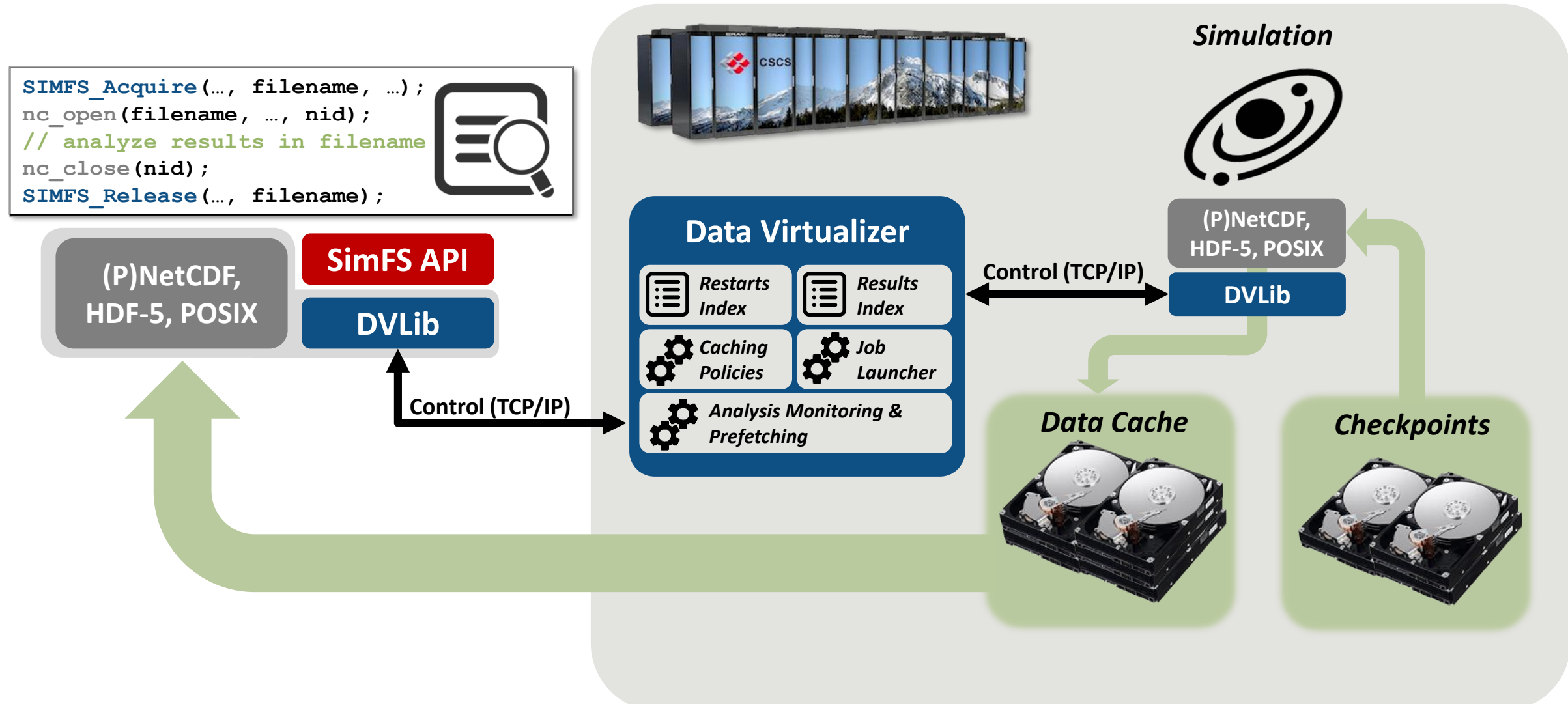
SimFS: A Simulation Data Virtualizing File System Interface



SimFS: A Simulation Data Virtualizing File System Interface



SimFS: A Simulation Data Virtualizing File System Interface

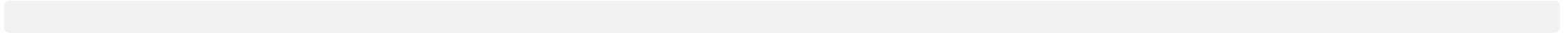


Will re-simulations slow my analysis down?

Improving Analysis Performance: Prefetching

No prefetching

time →



Improving Analysis Performance: Prefetching

No prefetching

time →

1



Improving Analysis Performance: Prefetching

No prefetching

SIM #1

1

time →

Improving Analysis Performance: Prefetching

No prefetching



1

time →

Improving Analysis Performance: Prefetching



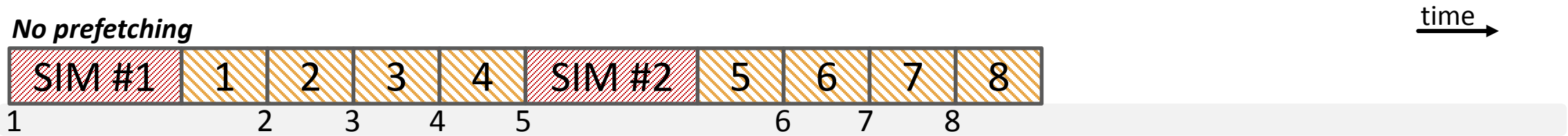
Improving Analysis Performance: Prefetching



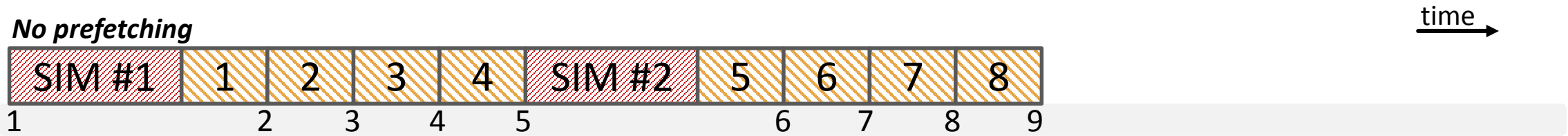
Improving Analysis Performance: Prefetching



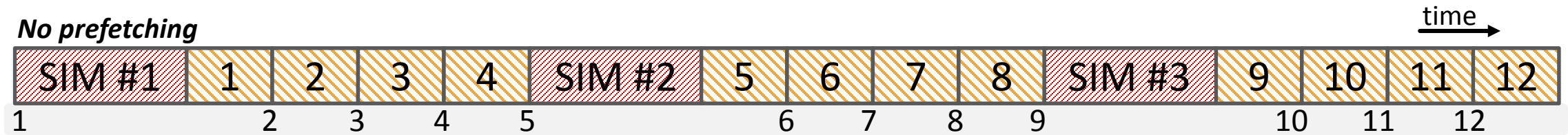
Improving Analysis Performance: Prefetching



Improving Analysis Performance: Prefetching



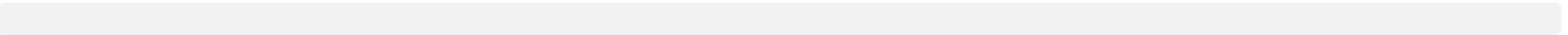
Improving Analysis Performance: Prefetching



Improving Analysis Performance: Prefetching



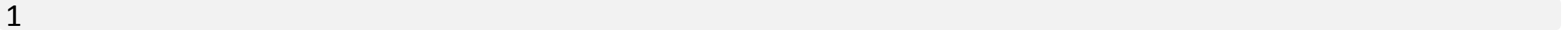
Masking Restart Latency



Improving Analysis Performance: Prefetching



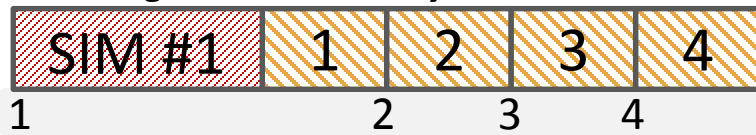
Masking Restart Latency



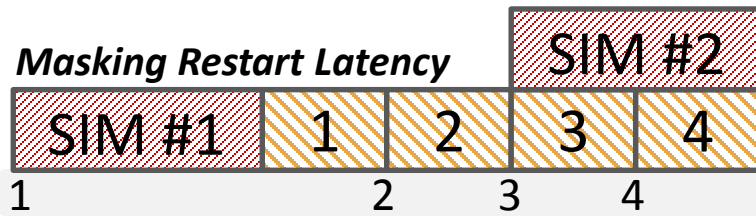
Improving Analysis Performance: Prefetching



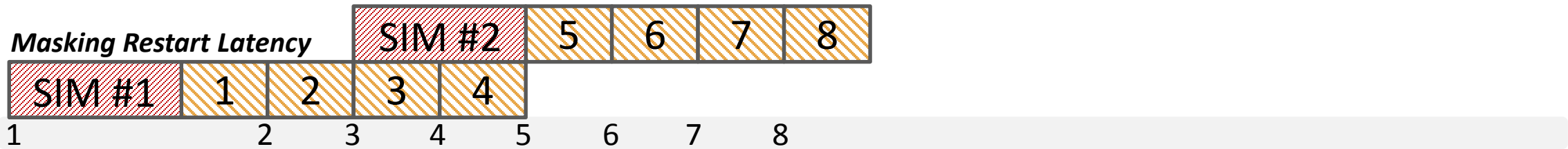
Masking Restart Latency



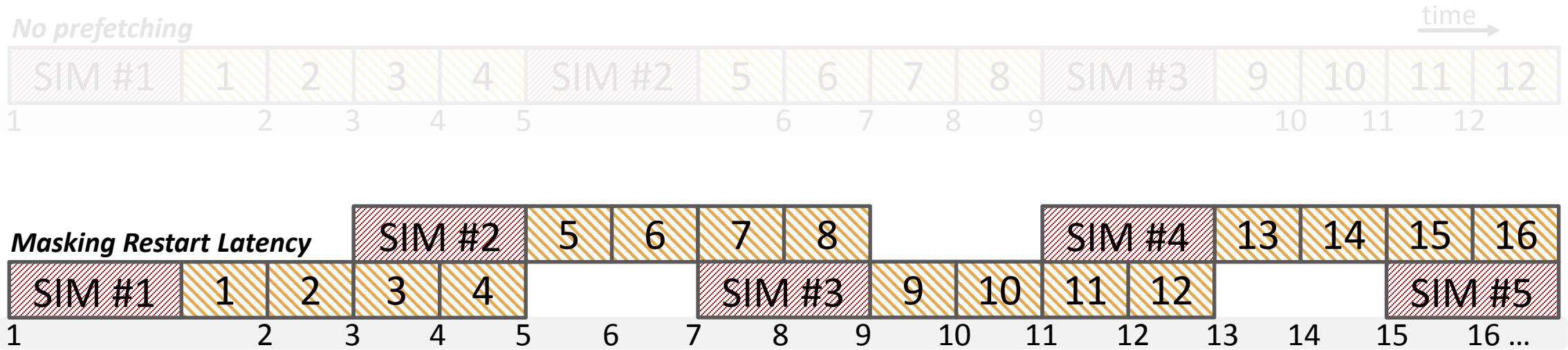
Improving Analysis Performance: Prefetching



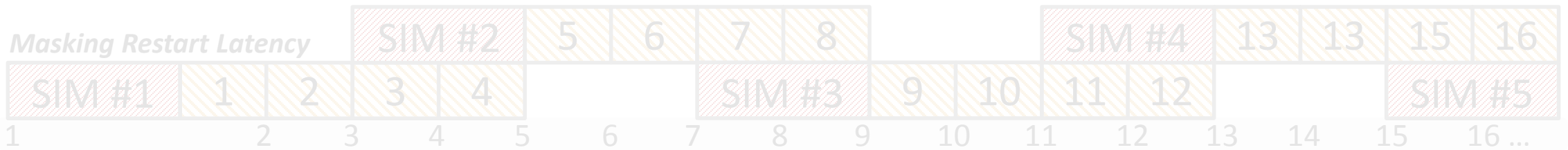
Improving Analysis Performance: Prefetching



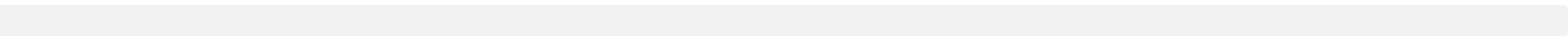
Improving Analysis Performance: Prefetching



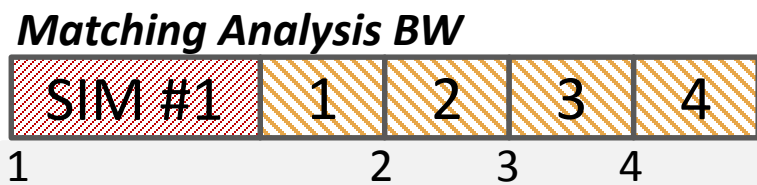
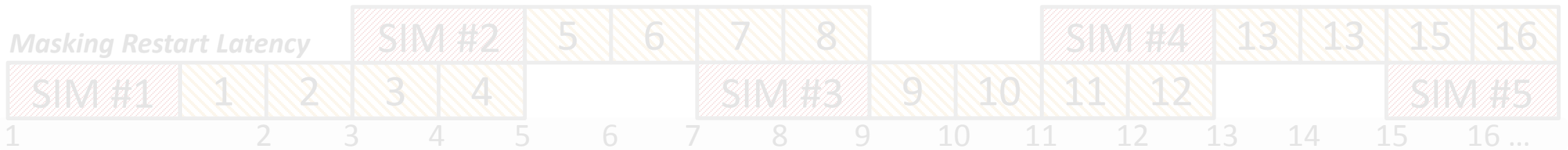
Improving Analysis Performance: Prefetching



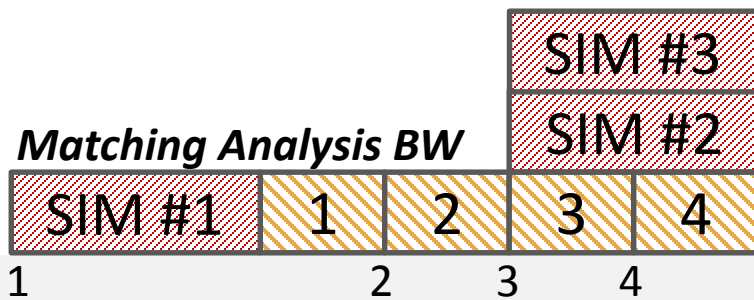
Matching Analysis BW



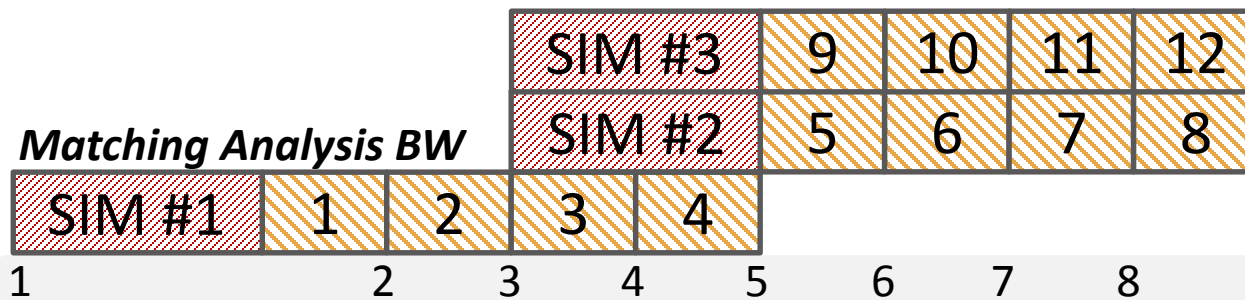
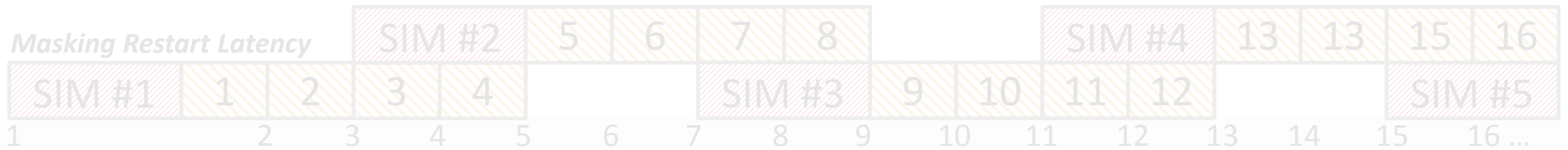
Improving Analysis Performance: Prefetching



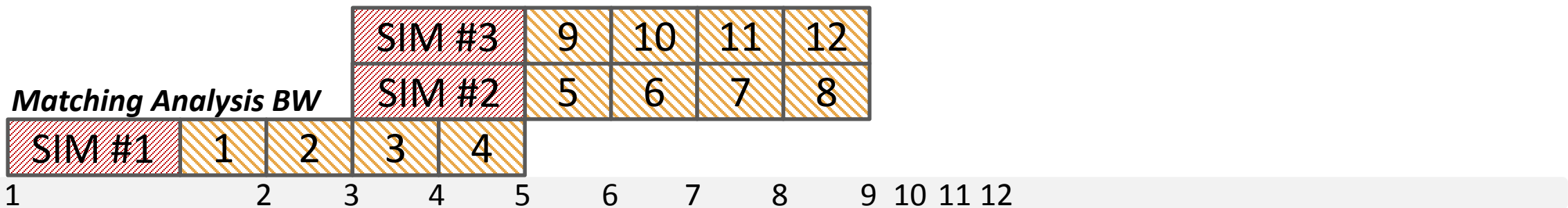
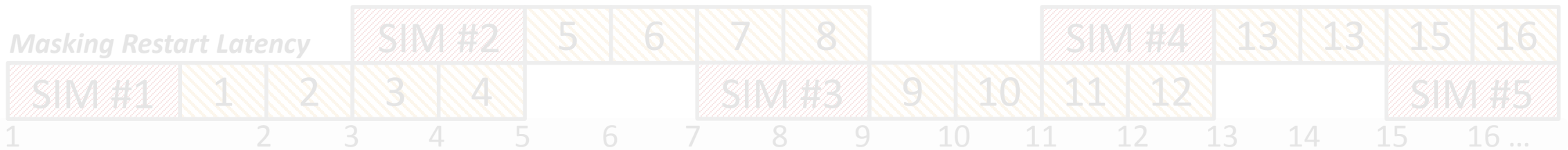
Improving Analysis Performance: Prefetching



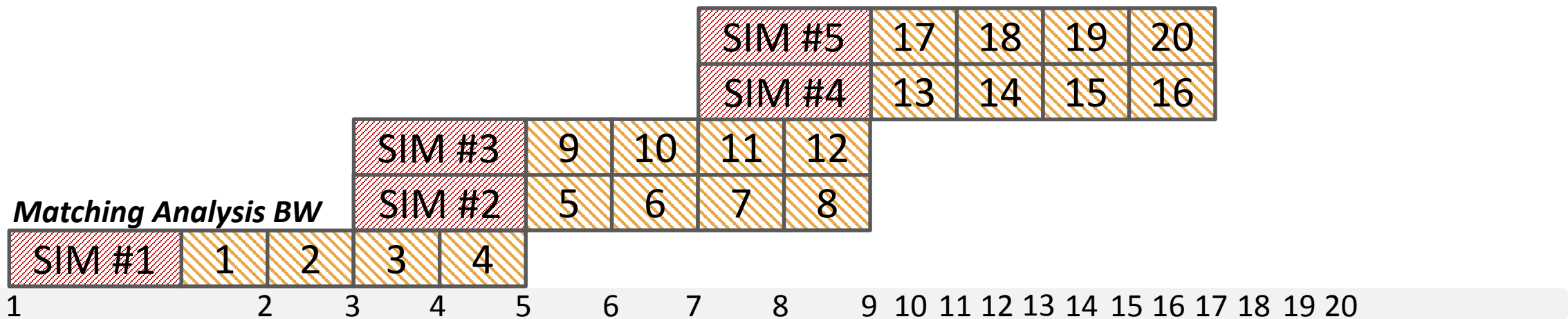
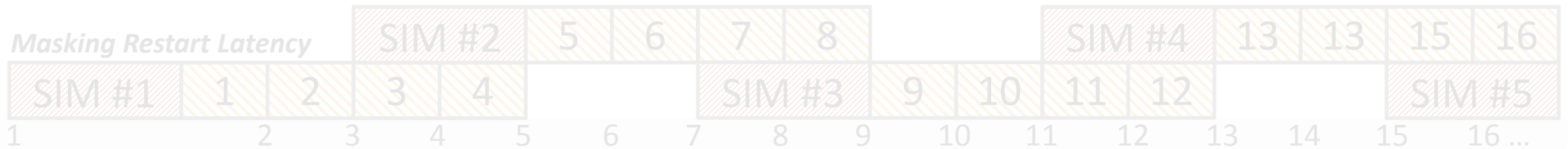
Improving Analysis Performance: Prefetching



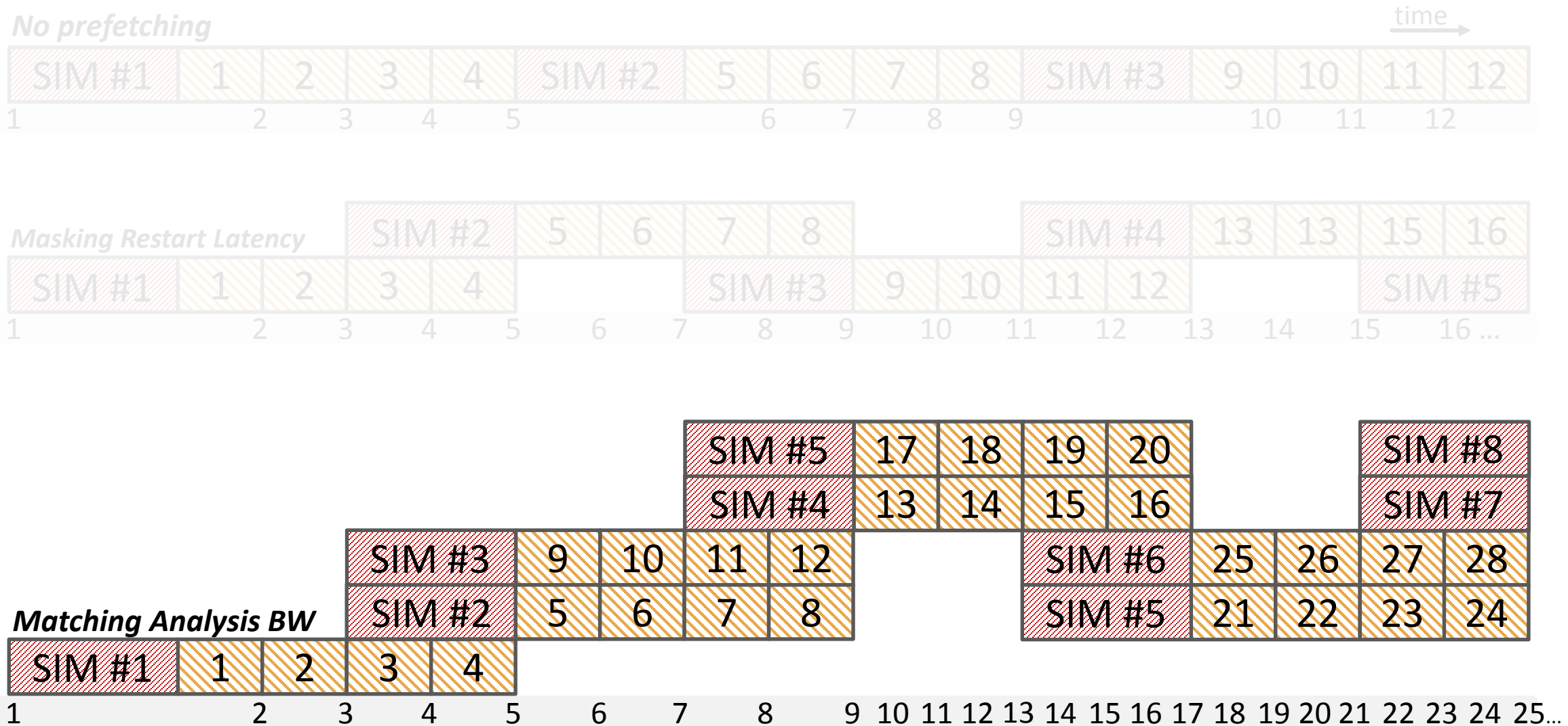
Improving Analysis Performance: Prefetching



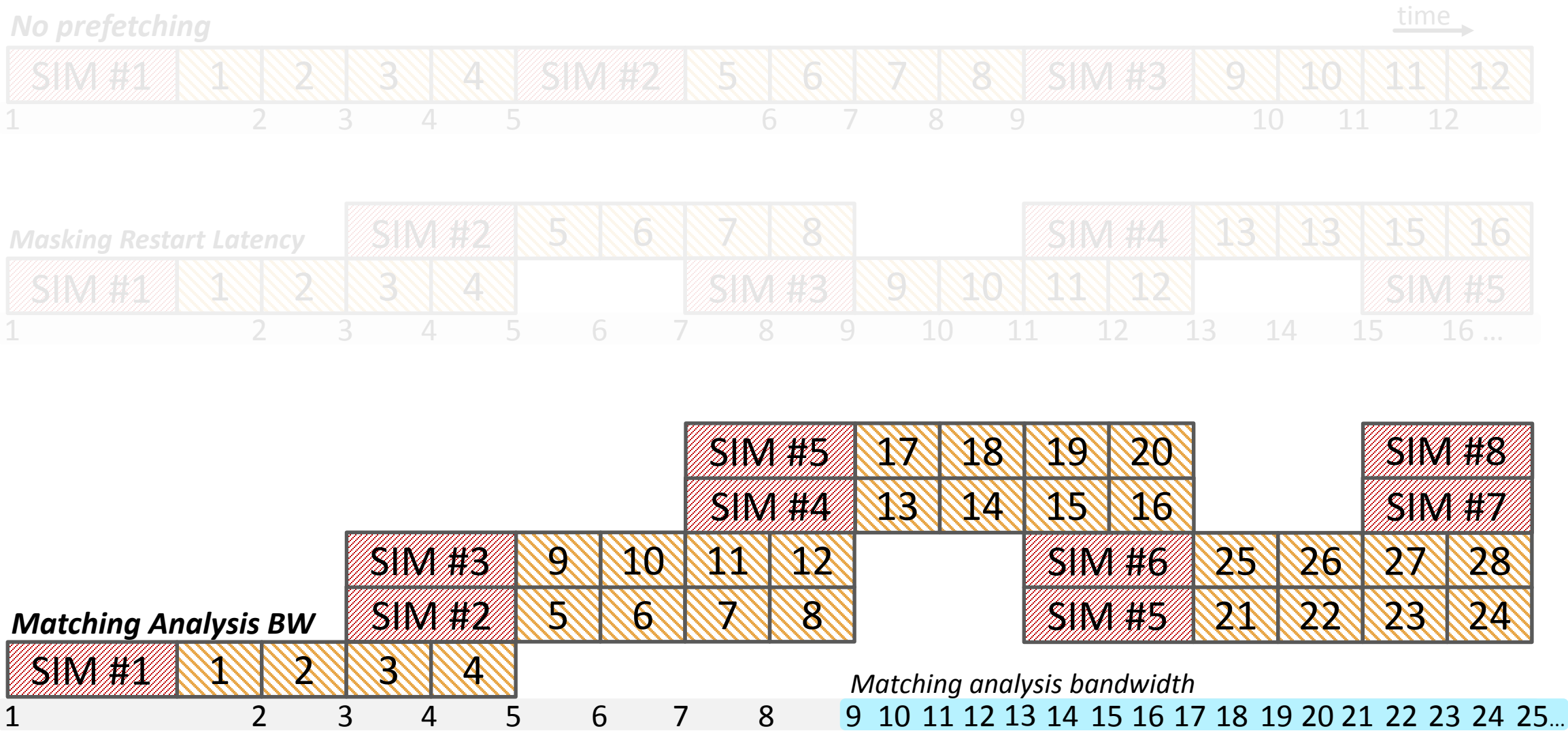
Improving Analysis Performance: Prefetching



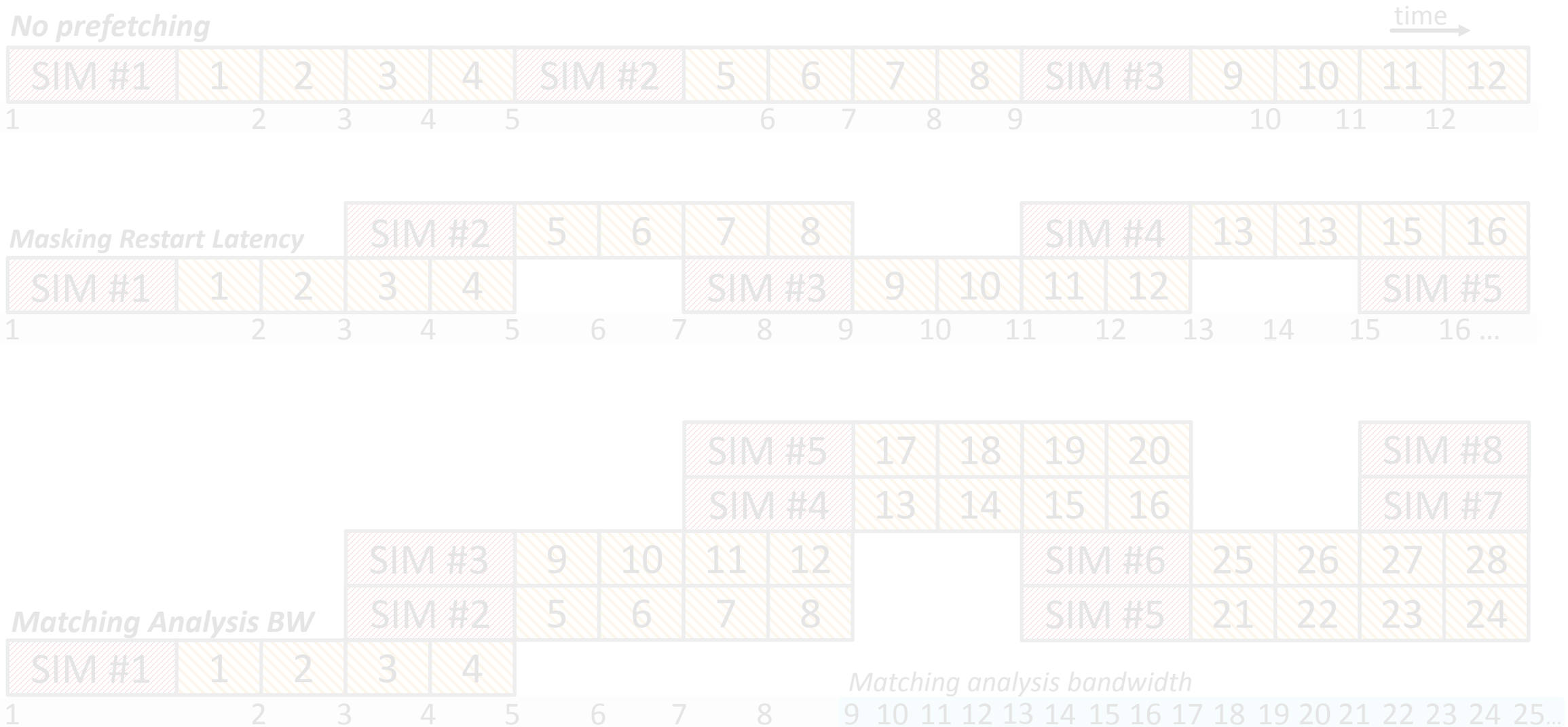
Improving Analysis Performance: Prefetching



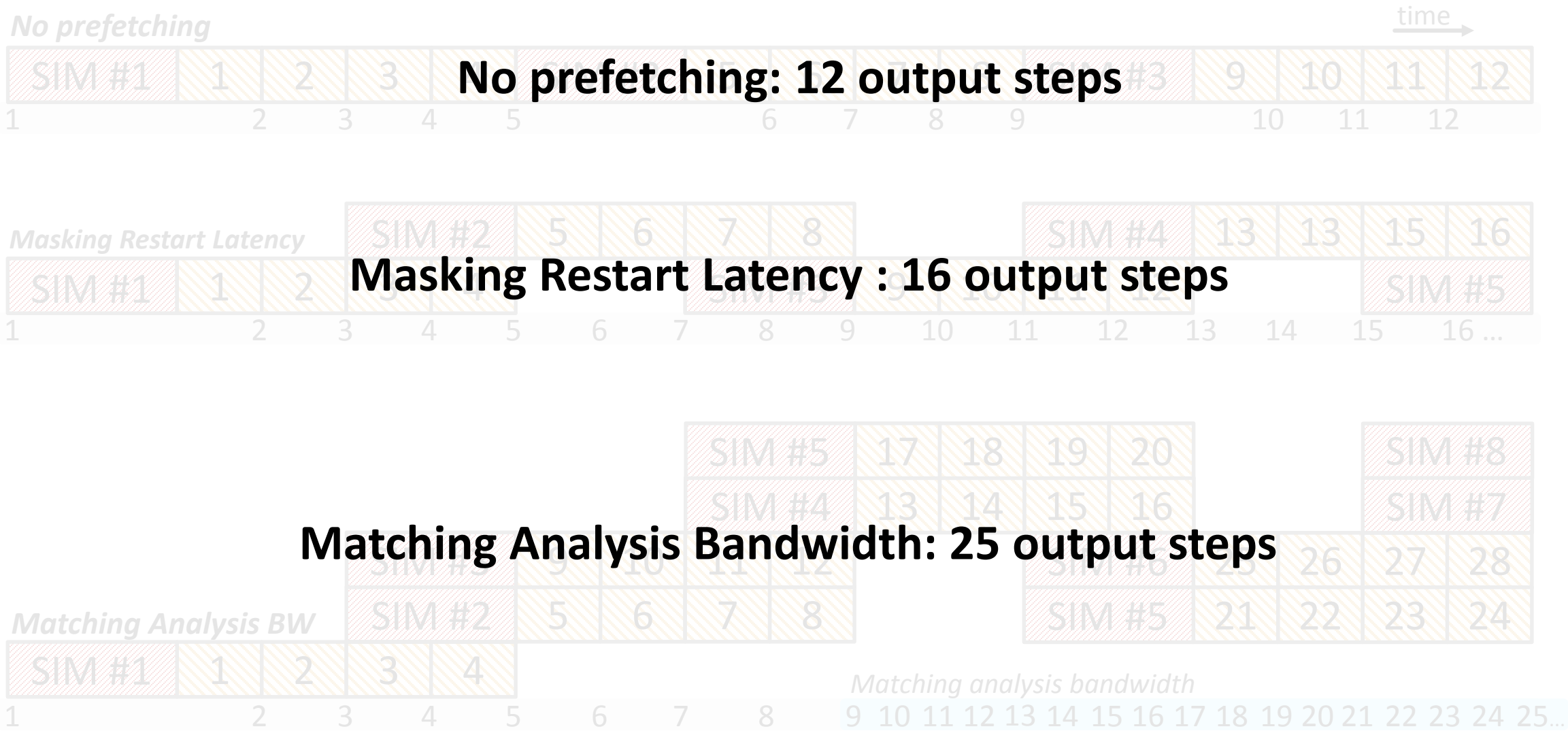
Improving Analysis Performance: Prefetching



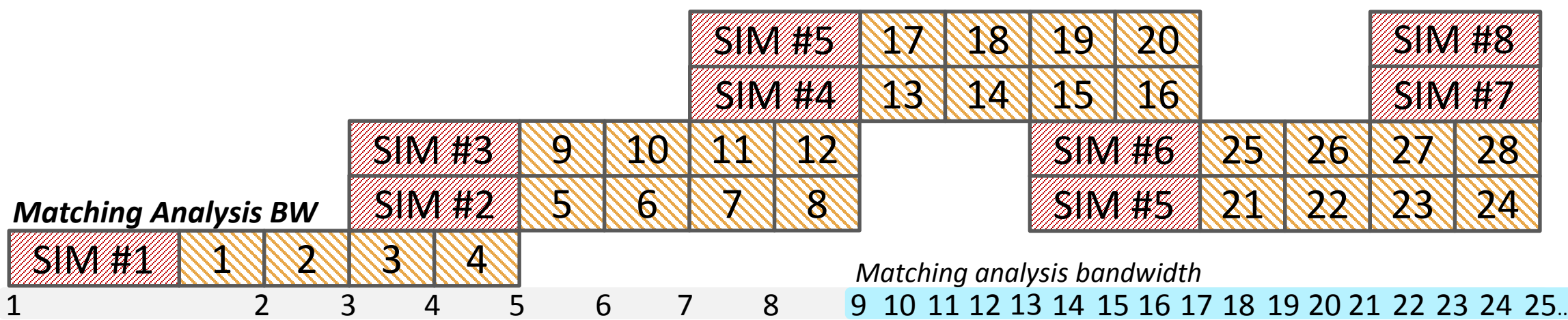
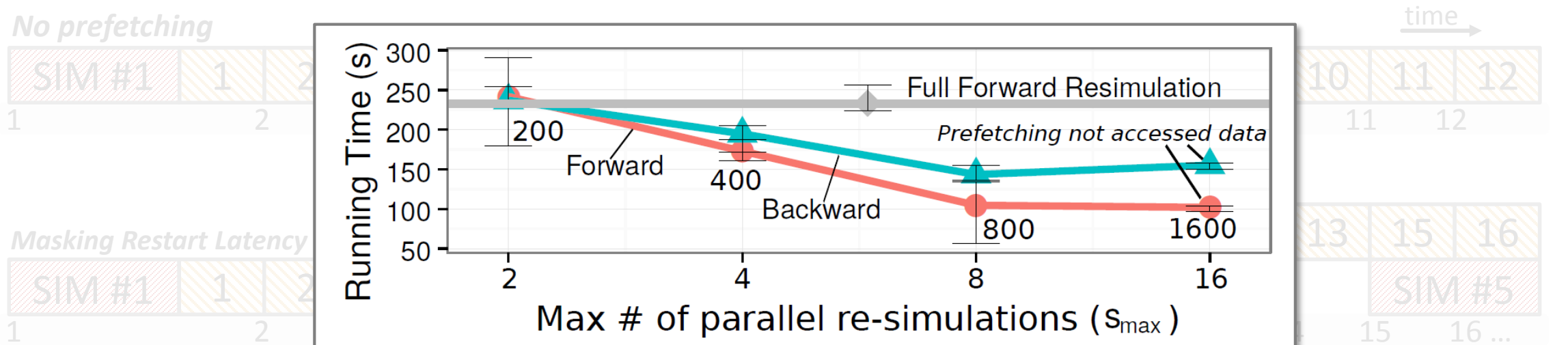
Improving Analysis Performance: Prefetching



Improving Analysis Performance: Prefetching



Improving Analysis Performance: Prefetching



Is SimFS cost-effective?

Is SimFS Cost-Effective?

Given a simulation dataset, for how long do you want to be able to access this data?

Is SimFS Cost-Effective?

Given a simulation dataset, for how long do you want to be able to access this data?

Data Availability Period

Is SimFS Cost-Effective?

Given a simulation dataset, for how long do you want to be able to access this data?

Data Availability Period

$$C_{\text{on-disk}}(\Delta t) = C_{\text{sim}}(n_o, N) + C_{\text{store}}(n_o, s_o, \Delta t)$$

$$C_{\text{SimFS}}(\Delta t) = C_{\text{sim}}(n_o, P) + C_{\text{store}}(n_r, s_r, \Delta t) + C_{\text{store}}(M, s_o, \Delta t) + C_{\text{sim}}(V(\gamma_{\Delta t}), P)$$

$$C_{\text{in-situ}}(\Delta t) = \sum_j^m C_{\text{sim}}(i_j + |\gamma_{\Delta t}(j)|, P)$$

Is SimFS Cost-Effective?

Given a simulation dataset, for how long do you want to be able to access this data?

Data Availability Period

$$C_{\text{on-disk}}(\Delta t) = C_{\text{sim}}(n_o, N) + C_{\text{store}}(n_o, s_o, \Delta t)$$

$$C_{\text{SimFS}}(\Delta t) = C_{\text{sim}}(n_o, P) + C_{\text{store}}(n_r, s_r, \Delta t) + C_{\text{store}}(M, s_o, \Delta t) + C_{\text{sim}}(V(\gamma_{\Delta t}), P)$$

$$C_{\text{in-situ}}(\Delta t) = \sum_j^m C_{\text{sim}}(i_j + |\gamma_{\Delta t}(j)|, P)$$



5min output
step every 20s

Is SimFS Cost-Effective?

Given a simulation dataset, for how long do you want to be able to access this data?

Data Availability Period

$$C_{\text{on-disk}}(\Delta t) = C_{\text{sim}}(n_o, N) + C_{\text{store}}(n_o, s_o, \Delta t)$$

$$C_{\text{SimFS}}(\Delta t) = C_{\text{sim}}(n_o, P) + C_{\text{store}}(n_r, s_r, \Delta t) + C_{\text{store}}(M, s_o, \Delta t) + C_{\text{sim}}(V(\gamma_{\Delta t}), P)$$

$$C_{\text{in-situ}}(\Delta t) = \sum_j^m C_{\text{sim}}(i_j + |\gamma_{\Delta t}(j)|, P)$$



5min output
step every 20s

5min output steps of a 1 month long simulation, output step size = 6GiB, checkpoint size = 36GiB



Is SimFS Cost-Effective?

Given a simulation dataset, for how long do you want to be able to access this data?

Data Availability Period

$$C_{\text{on-disk}}(\Delta t) = C_{\text{sim}}(n_o, N) + C_{\text{store}}(n_o, s_o, \Delta t)$$

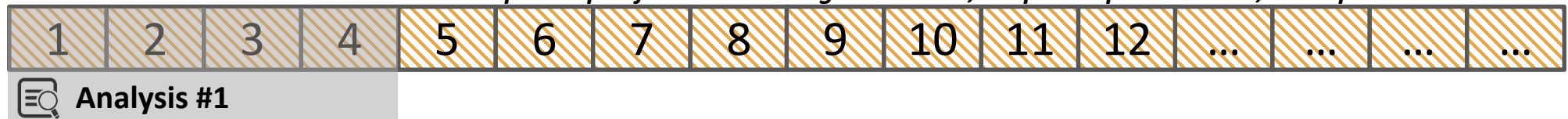
$$C_{\text{SimFS}}(\Delta t) = C_{\text{sim}}(n_o, P) + C_{\text{store}}(n_r, s_r, \Delta t) + C_{\text{store}}(M, s_o, \Delta t) + C_{\text{sim}}(V(\gamma_{\Delta t}), P)$$

$$C_{\text{in-situ}}(\Delta t) = \sum_j^m C_{\text{sim}}(i_j + |\gamma_{\Delta t}(j)|, P)$$



5min output step every 20s

5min output steps of a 1 month long simulation, output step size = 6GiB, checkpoint size = 36GiB



Is SimFS Cost-Effective?

Given a simulation dataset, for how long do you want to be able to access this data?

Data Availability Period

$$C_{\text{on-disk}}(\Delta t) = C_{\text{sim}}(n_o, N) + C_{\text{store}}(n_o, s_o, \Delta t)$$

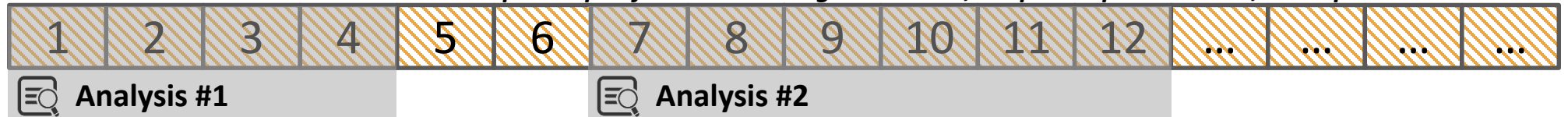
$$C_{\text{SimFS}}(\Delta t) = C_{\text{sim}}(n_o, P) + C_{\text{store}}(n_r, s_r, \Delta t) + C_{\text{store}}(M, s_o, \Delta t) + C_{\text{sim}}(V(\gamma_{\Delta t}), P)$$

$$C_{\text{in-situ}}(\Delta t) = \sum_j^m C_{\text{sim}}(i_j + |\gamma_{\Delta t}(j)|, P)$$



5min output step every 20s

5min output steps of a 1 month long simulation, output step size = 6GiB, checkpoint size = 36GiB



Is SimFS Cost-Effective?

Given a simulation dataset, for how long do you want to be able to access this data?

Data Availability Period

$$C_{\text{on-disk}}(\Delta t) = C_{\text{sim}}(n_o, N) + C_{\text{store}}(n_o, s_o, \Delta t)$$

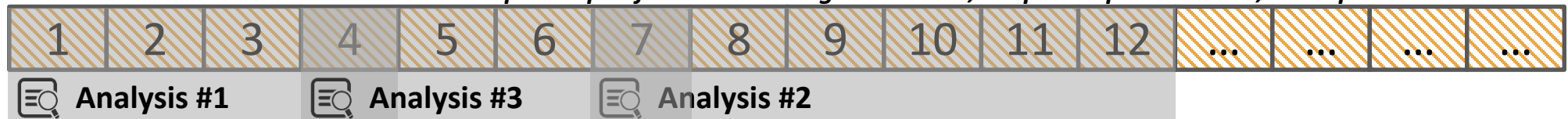
$$C_{\text{SimFS}}(\Delta t) = C_{\text{sim}}(n_o, P) + C_{\text{store}}(n_r, s_r, \Delta t) + C_{\text{store}}(M, s_o, \Delta t) + C_{\text{sim}}(V(\gamma_{\Delta t}), P)$$

$$C_{\text{in-situ}}(\Delta t) = \sum_j^m C_{\text{sim}}(i_j + |\gamma_{\Delta t}(j)|, P)$$



5min output step every 20s

5min output steps of a 1 month long simulation, output step size = 6GiB, checkpoint size = 36GiB



Number of analysis applications: 100

Is SimFS Cost-Effective?

Given a simulation dataset, for how long do you want to be able to access this data?

Data Availability Period

$$C_{\text{on-disk}}(\Delta t) = C_{\text{sim}}(n_o, N) + C_{\text{store}}(n_o, s_o, \Delta t)$$

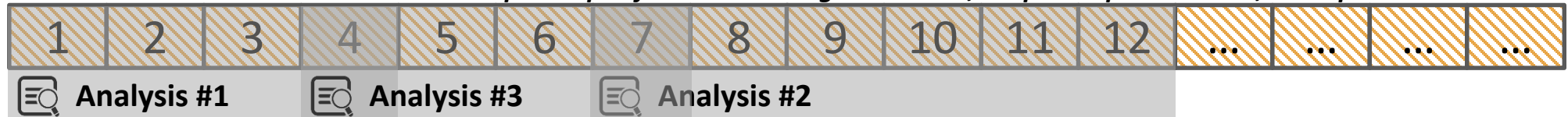
$$C_{\text{SimFS}}(\Delta t) = C_{\text{sim}}(n_o, P) + C_{\text{store}}(n_r, s_r, \Delta t) + C_{\text{store}}(M, s_o, \Delta t) + C_{\text{sim}}(V(\gamma_{\Delta t}), P)$$

$$C_{\text{in-situ}}(\Delta t) = \sum_j^m C_{\text{sim}}(i_j + |\gamma_{\Delta t}(j)|, P)$$



5min output step every 20s

5min output steps of a 1 month long simulation, output step size = 6GiB, checkpoint size = 36GiB



Number of analysis applications: 100



Compute: 2.07\$/Node/Hour (NCv2 VM)
Store: 0.06\$/GiB/Month (Azure File Share)

Is SimFS Cost-Effective?

Given a simulation dataset, for how long do you want to be able to access this data?

Data Availability Period

$$C_{\text{on-disk}}(\Delta t) = C_{\text{sim}}(n_o, N) + C_{\text{store}}(n_o, s_o, \Delta t)$$

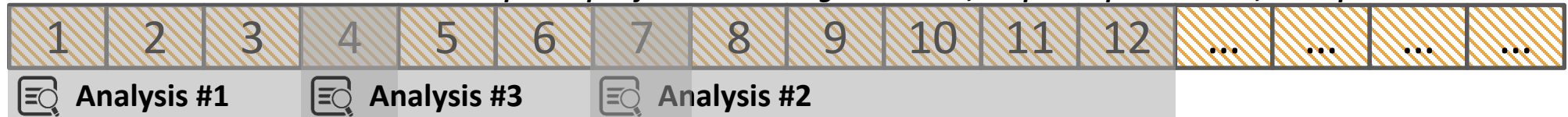
$$C_{\text{SimFS}}(\Delta t) = C_{\text{sim}}(n_o, P) + C_{\text{store}}(n_r, s_r, \Delta t) + C_{\text{store}}(M, s_o, \Delta t) + C_{\text{sim}}(V(\gamma_{\Delta t}), P)$$

$$C_{\text{in-situ}}(\Delta t) = \sum_j^m C_{\text{sim}}(i_j + |\gamma_{\Delta t}(j)|, P)$$



5min output step every 20s

5min output steps of a 1 month long simulation, output step size = 6GiB, checkpoint size = 36GiB



Number of analysis applications: 100



Compute: 2.07\$/Node/Hour (NCv2 VM)
Store: 0.06\$/GiB/Month (Azure File Share)



25% cache size
Restart interval: 8h

Is SimFS Cost-Effective?

Given a simulation dataset, for how long do you want to be able to access this data?

Data Availability Period

$$C_{\text{on-disk}}(\Delta t) = C_{\text{sim}}(n_o, N) + C_{\text{store}}(n_o, s_o, \Delta t)$$

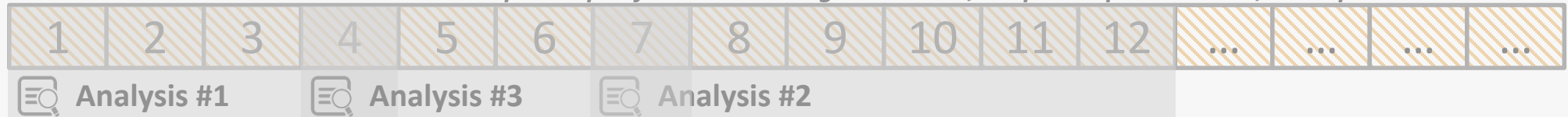
$$C_{\text{SimFS}}(\Delta t) = C_{\text{sim}}(n_o, P) + C_{\text{store}}(n_r, s_r, \Delta t) + C_{\text{store}}(M, s_o, \Delta t) + C_{\text{sim}}(V(\gamma_{\Delta t}), P)$$

$$C_{\text{in-situ}}(\Delta t) = \sum_j^m C_{\text{sim}}(i_j + |\gamma_{\Delta t}(j)|, P)$$



5min output step every 20s

5min output steps of a 1 month long simulation, output step size = 6GiB, checkpoint size = 36GiB



Number of analysis applications: 100



Compute: 2.07\$/Node/Hour (NCv2 VM)
Store: 0.06\$/GiB/Month (Azure File Share)



25% cache size
Restart interval: 8h

Is SimFS Cost-Effective?

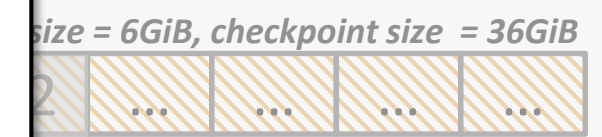
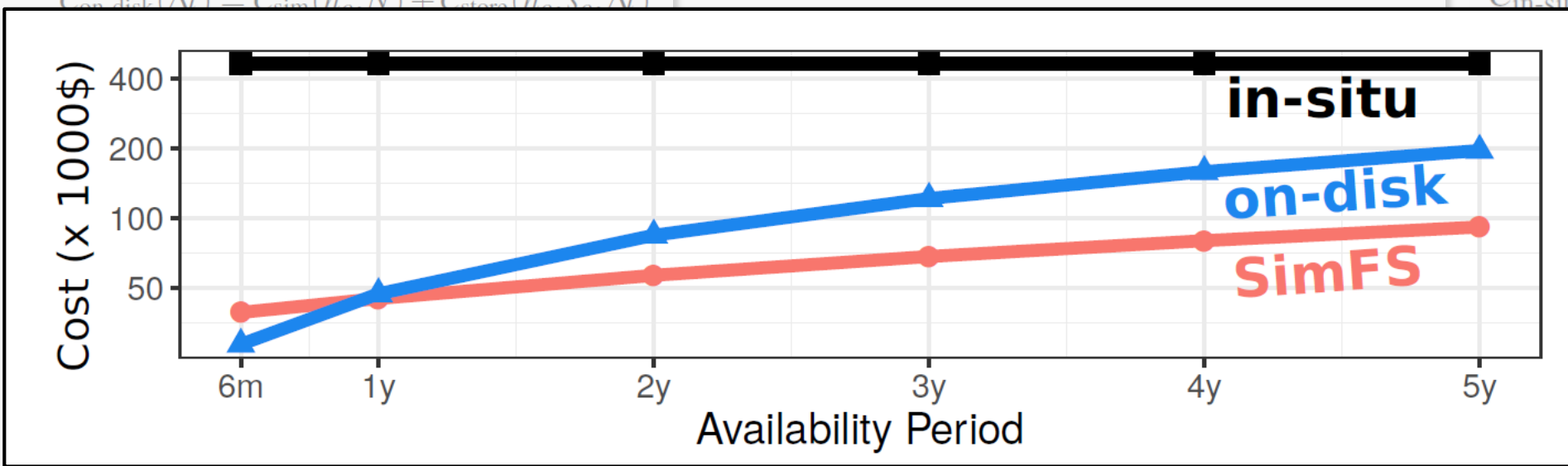
Given a simulation dataset, for how long do you want to be able to access this data?

Data Availability Period

$$C_{\text{on-disk}}(\Delta t) = C_{\text{sim}}(n_o, P) + C_{\text{store}}(n_r, s_r, \Delta t) +$$

$$C_{\text{SimFS}}(\Delta t) = C_{\text{sim}}(n_o, P) + C_{\text{store}}(n_r, s_r, \Delta t) +$$

$$C_{\text{in-situ}}(\Delta t) = \sum_j^m C_{\text{sim}}(i_j + |\gamma_{\Delta t}(j)|, P)$$



Compute: 2.07\$/Node/Hour (NCv2 VM)
Store: 0.06\$/GiB/Month (Azure File Share)

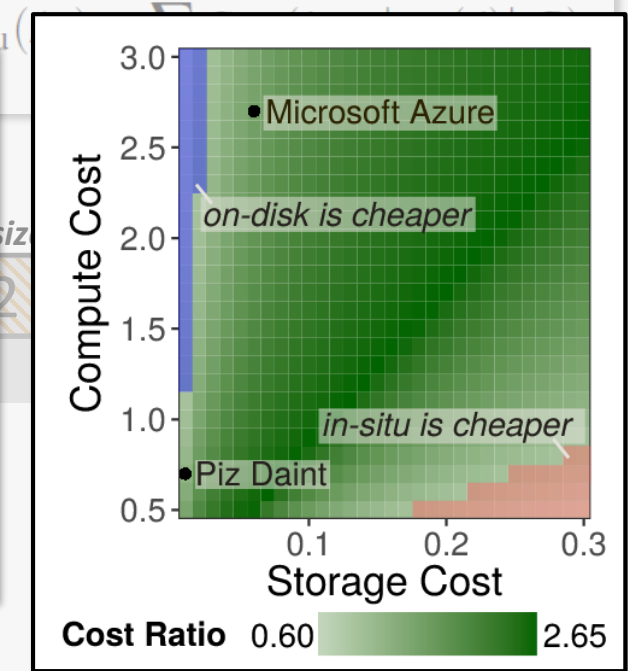
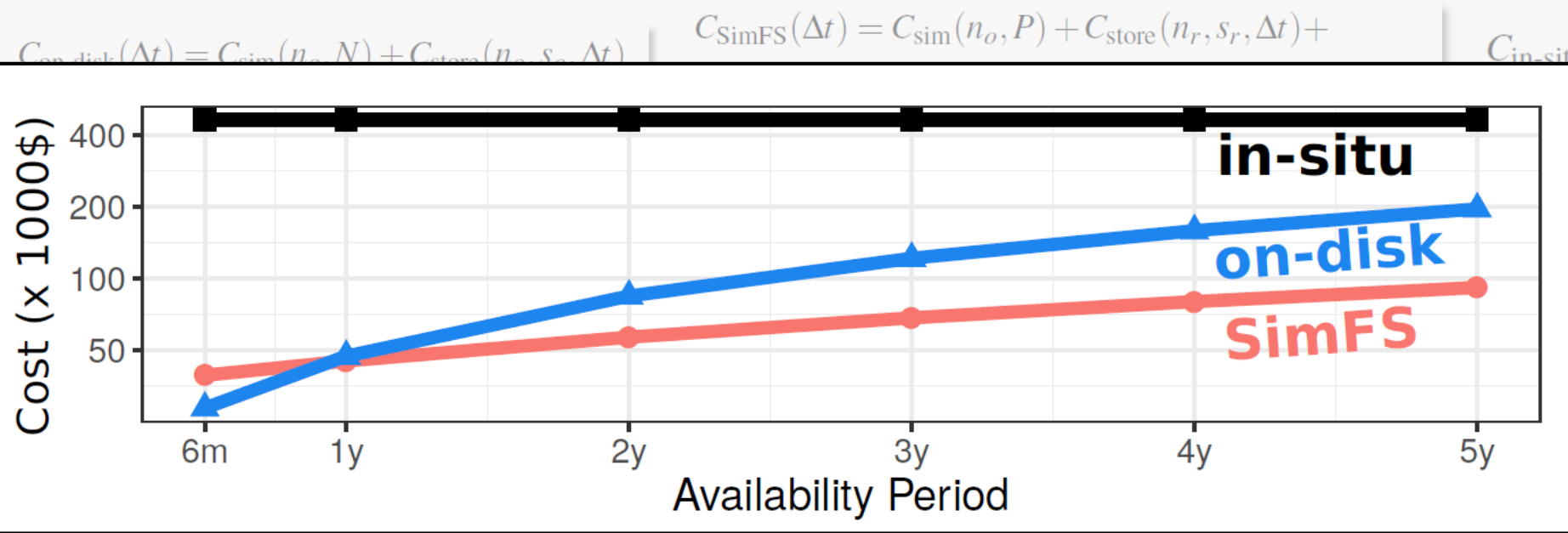


25% cache size
Restart interval: 8h

Is SimFS Cost-Effective?

Given a simulation dataset, for how long do you want to be able to access this data?

Data Availability Period



Compute: 2.07\$/Node/Hour (NCv2 VM)
Store: 0.06\$/GiB/Month (Azure File Share)



25% cache size
Restart interval: 8h

What if I need only some of the data produced by the simulation?

Simulation Contexts

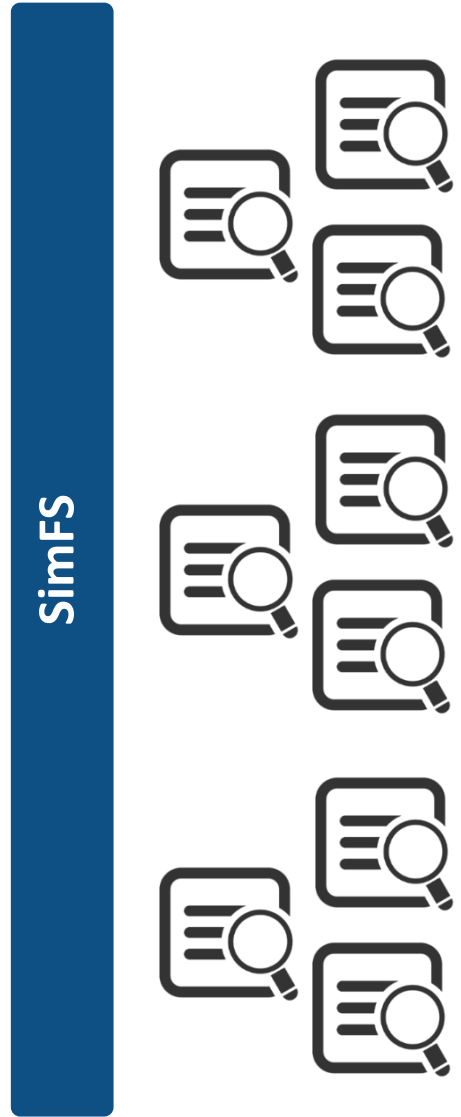
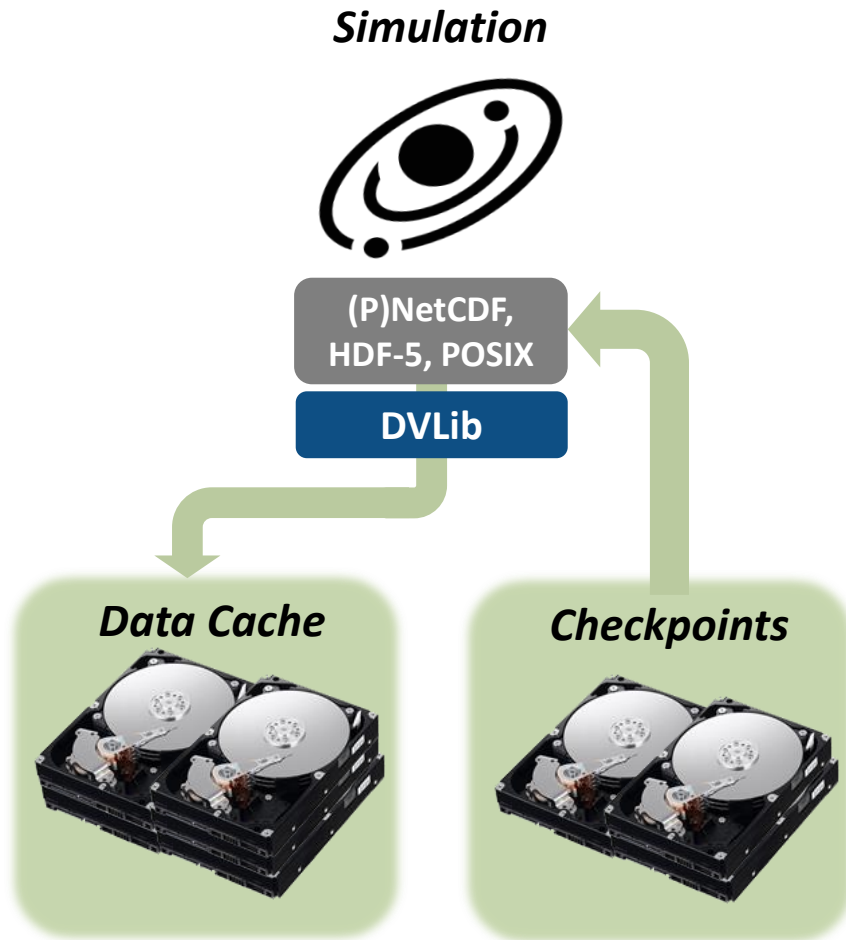
Simulation Contexts



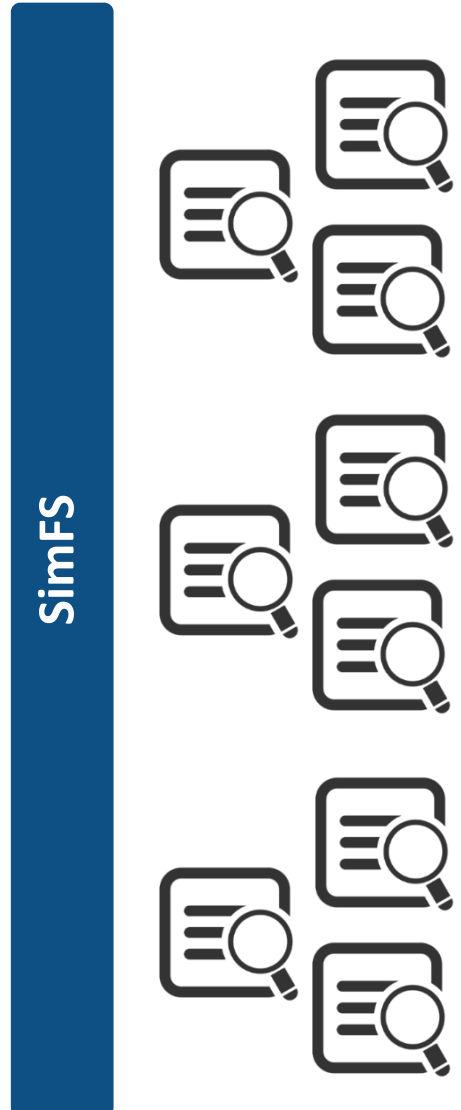
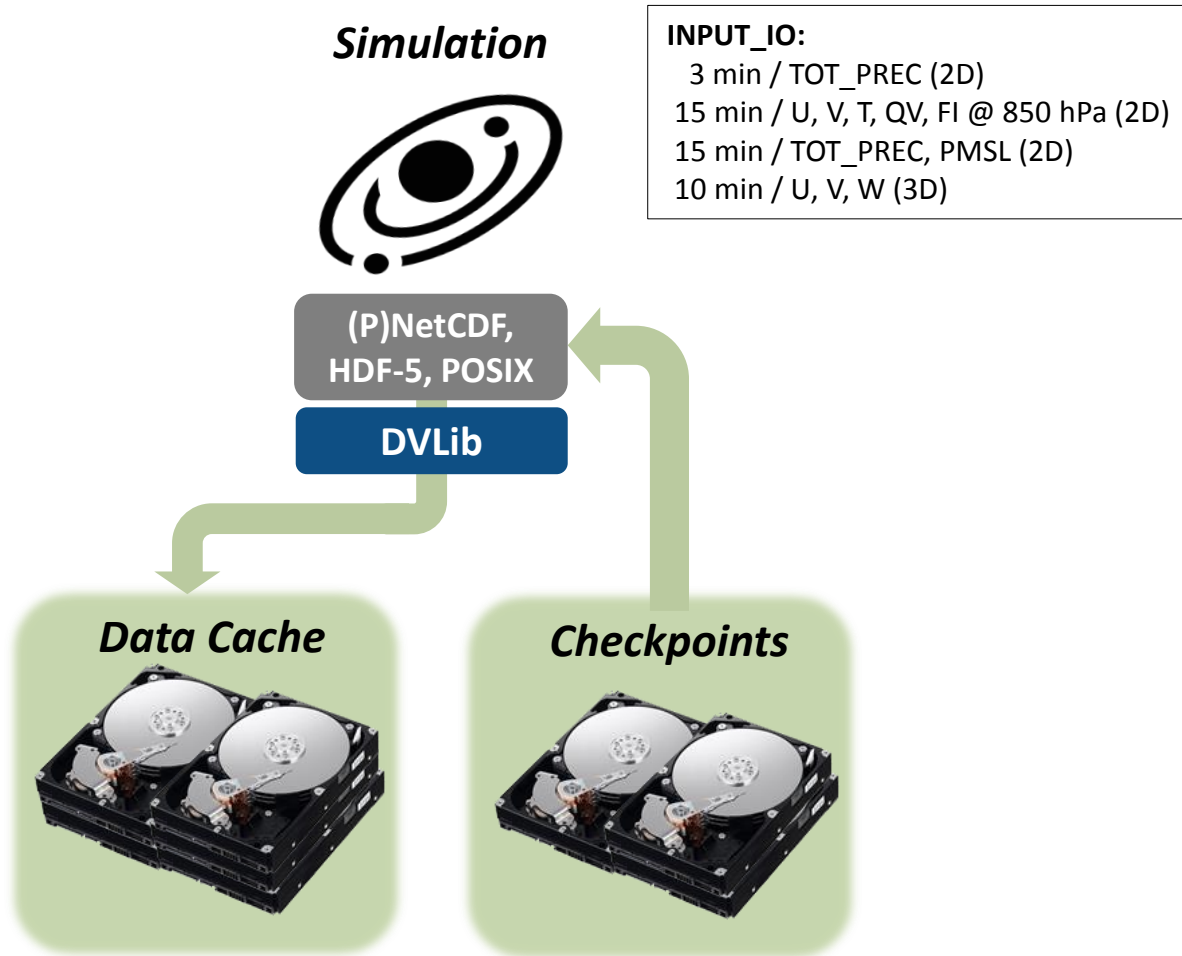
Simulation Contexts



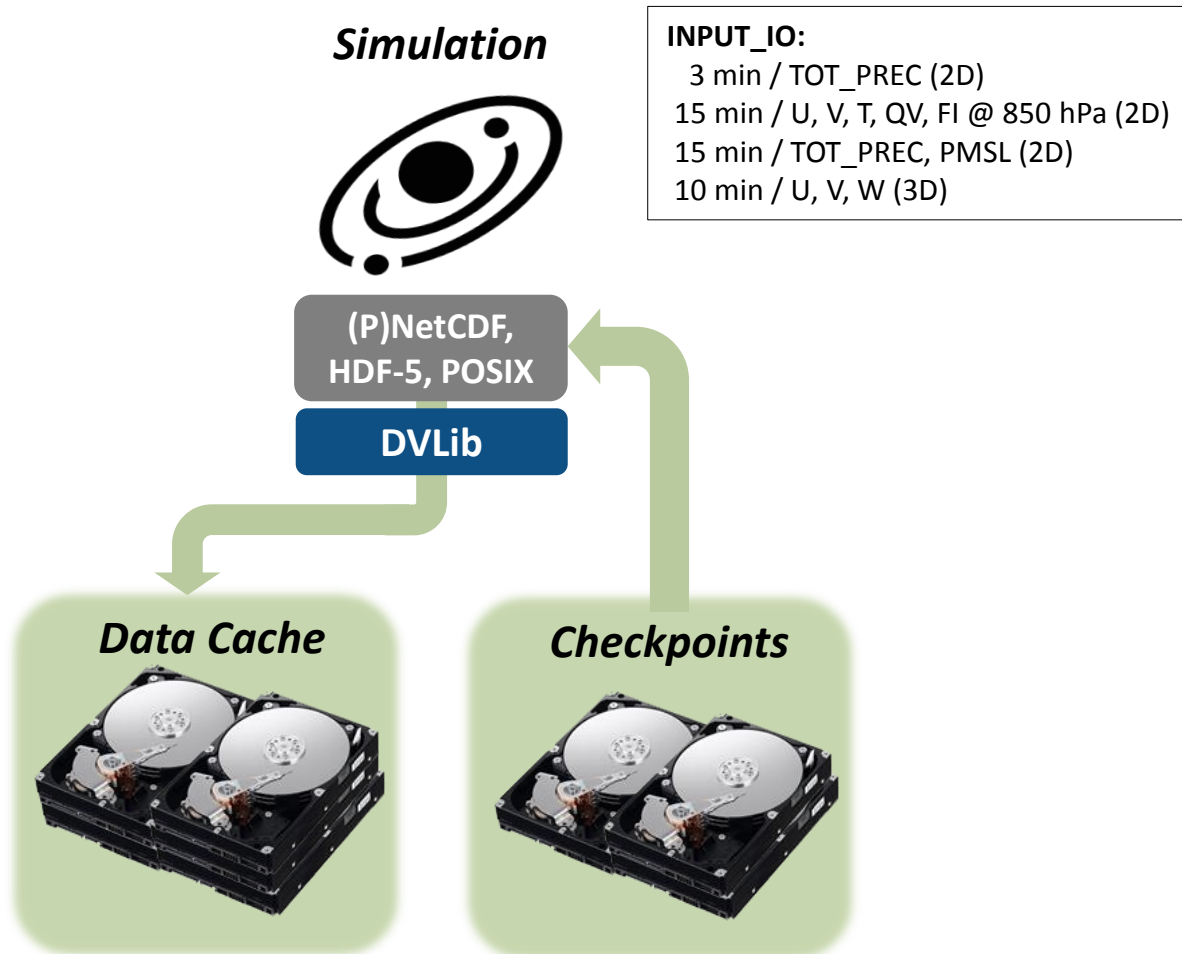
Simulation Contexts



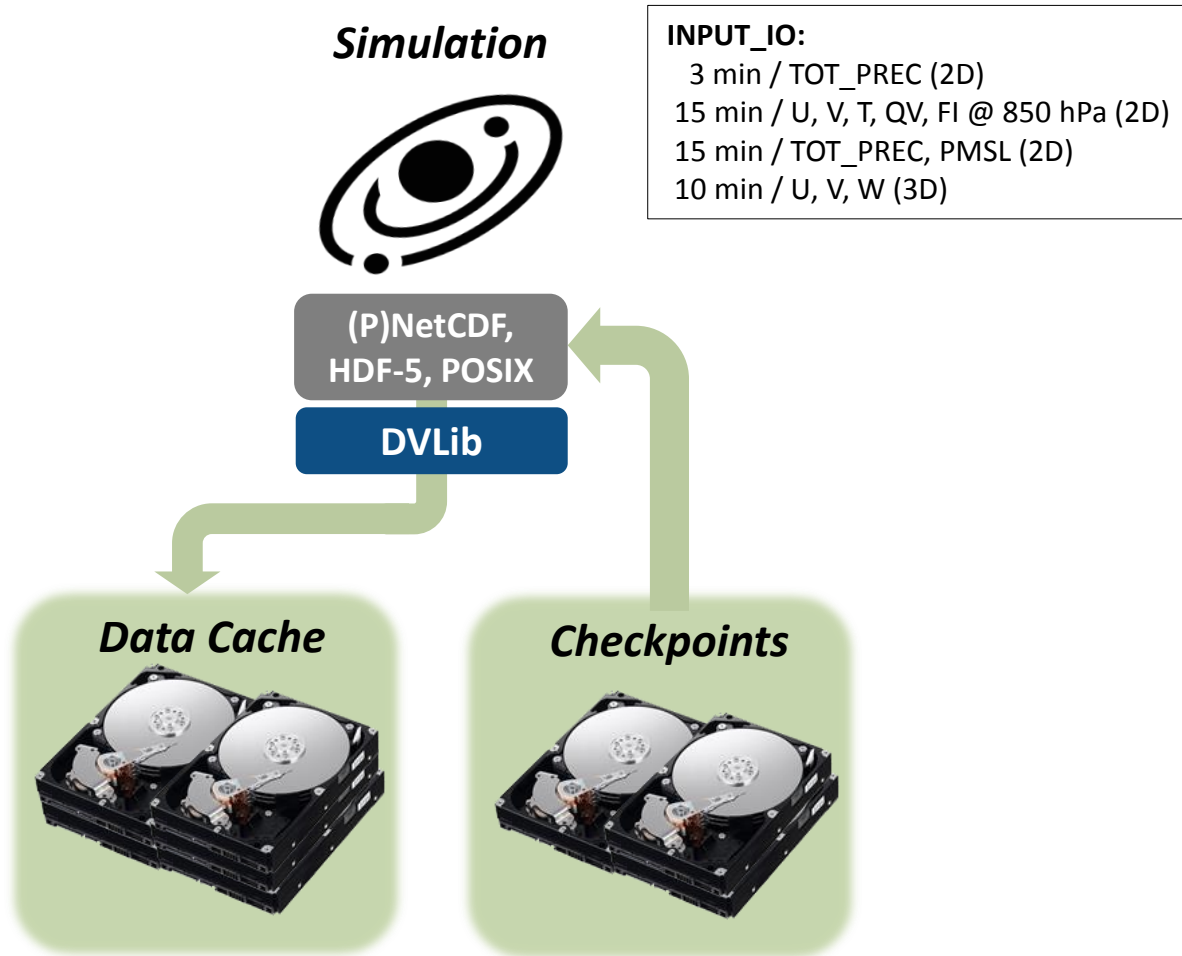
Simulation Contexts



Simulation Contexts



Simulation Contexts

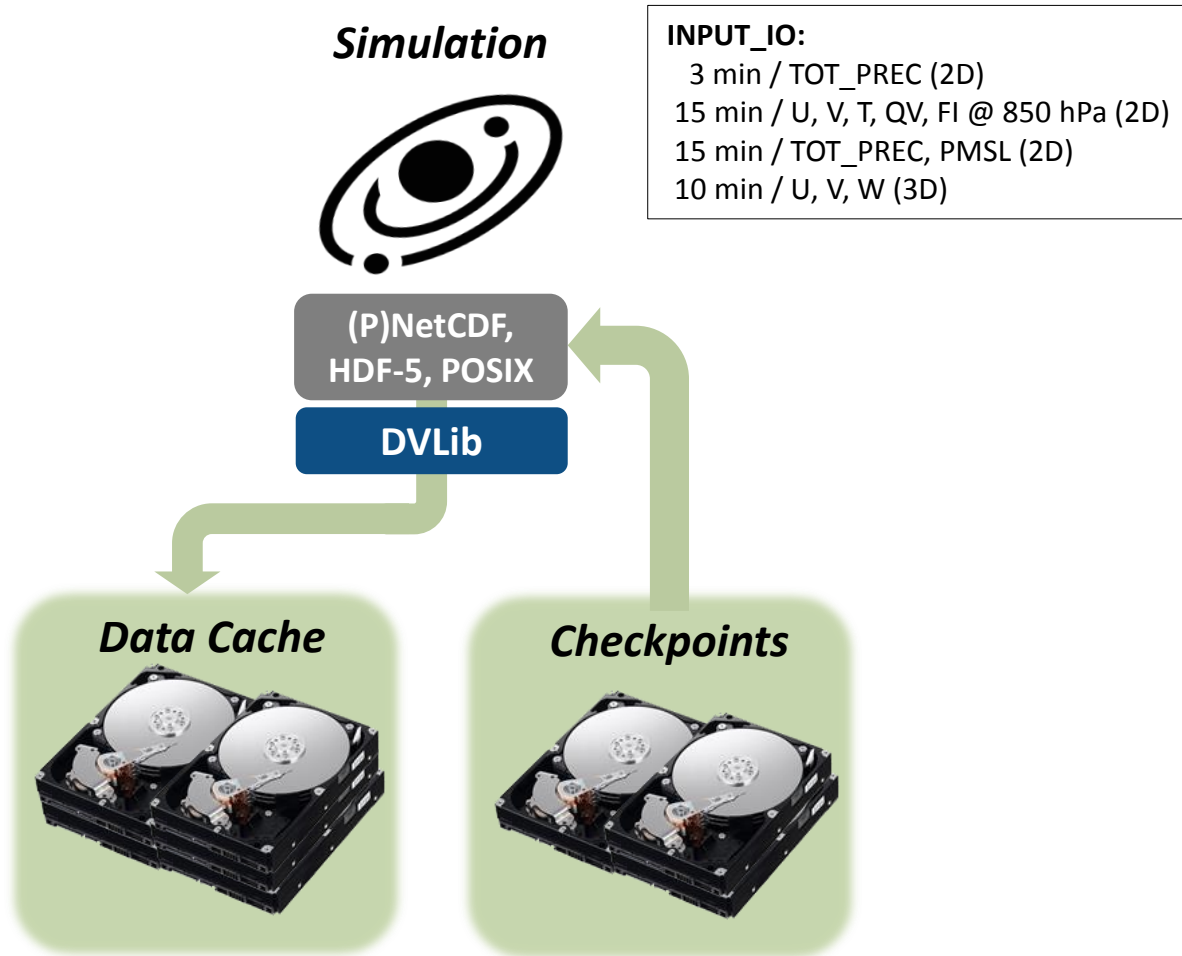


SimFS



precip_tracking:
 3min / TOT_PREC (2D)

Simulation Contexts



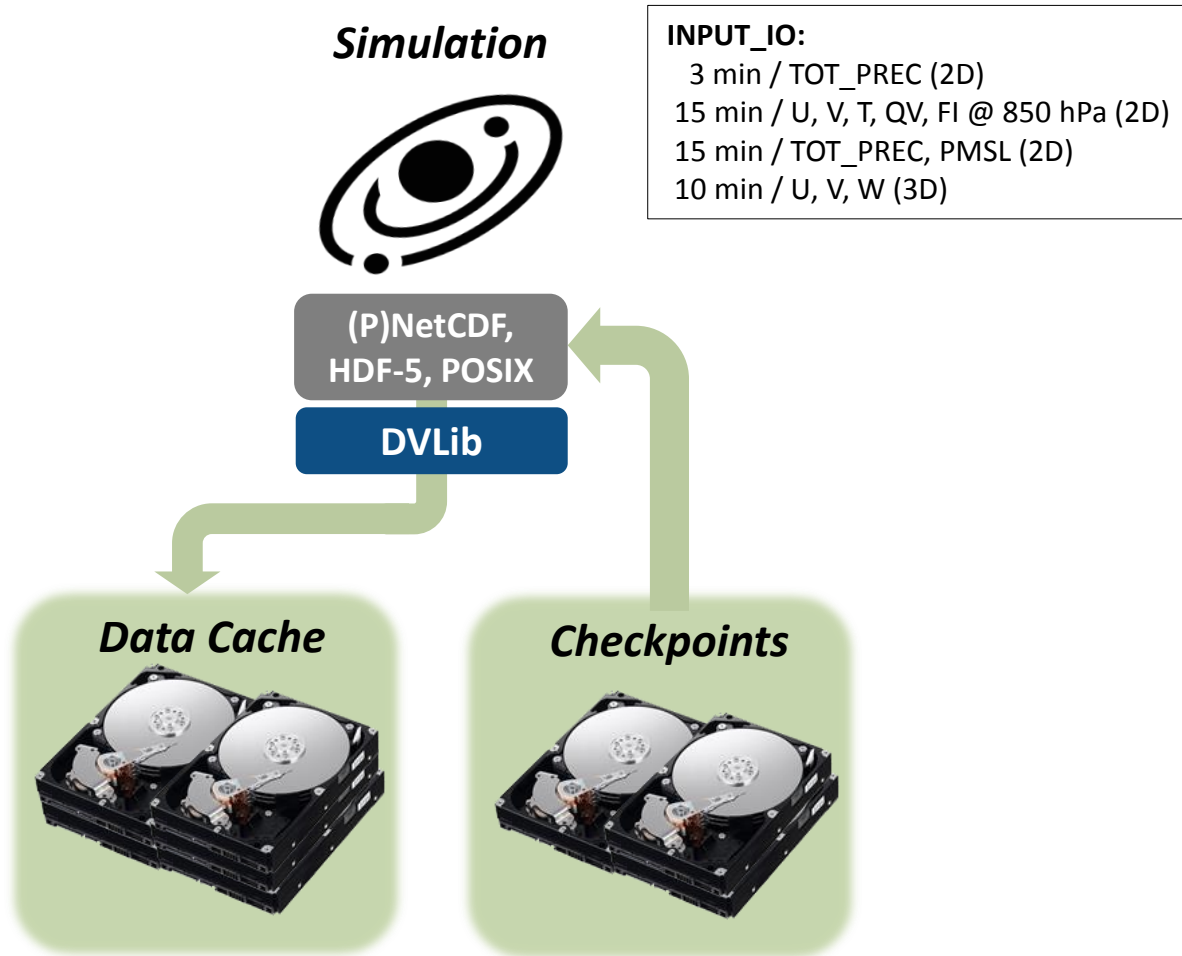
SimFS



precip_tracking:
3min / TOT_PREC (2D)

fronts_cyclones:
 15 min / U, V, T, QV, FI @ 850 hPa (2D)
 15 min / TOT_PREC, PMSL (2D)

Simulation Contexts



SimFS

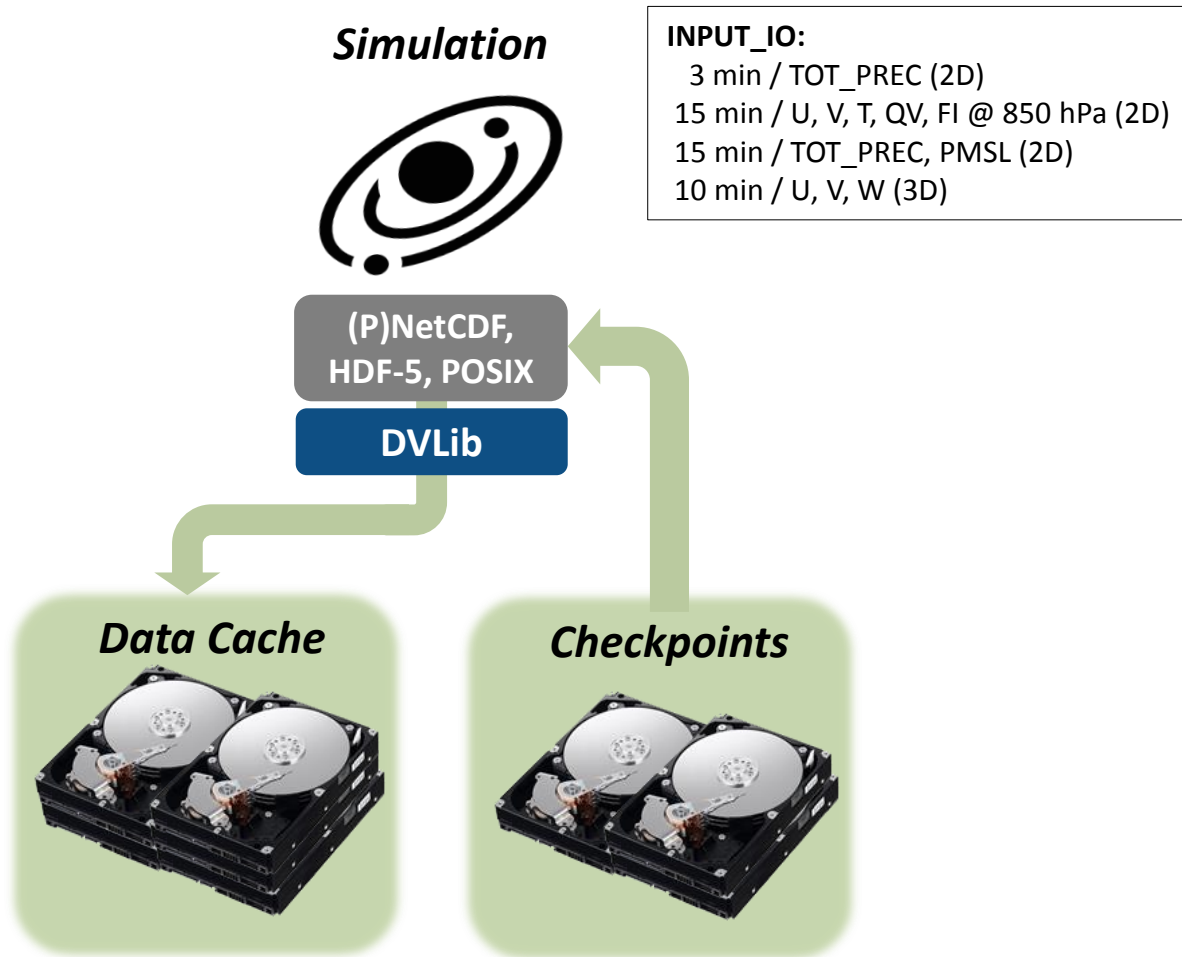


precip_tracking:
3min / TOT_PREC (2D)

fronts_cyclones:
15 min / U, V, T, QV, FI @ 850 hPa (2D)
15 min / TOT_PREC, PMSL (2D)

trajectories:
10 min / U, V, W (3D)

Simulation Contexts



SimFS

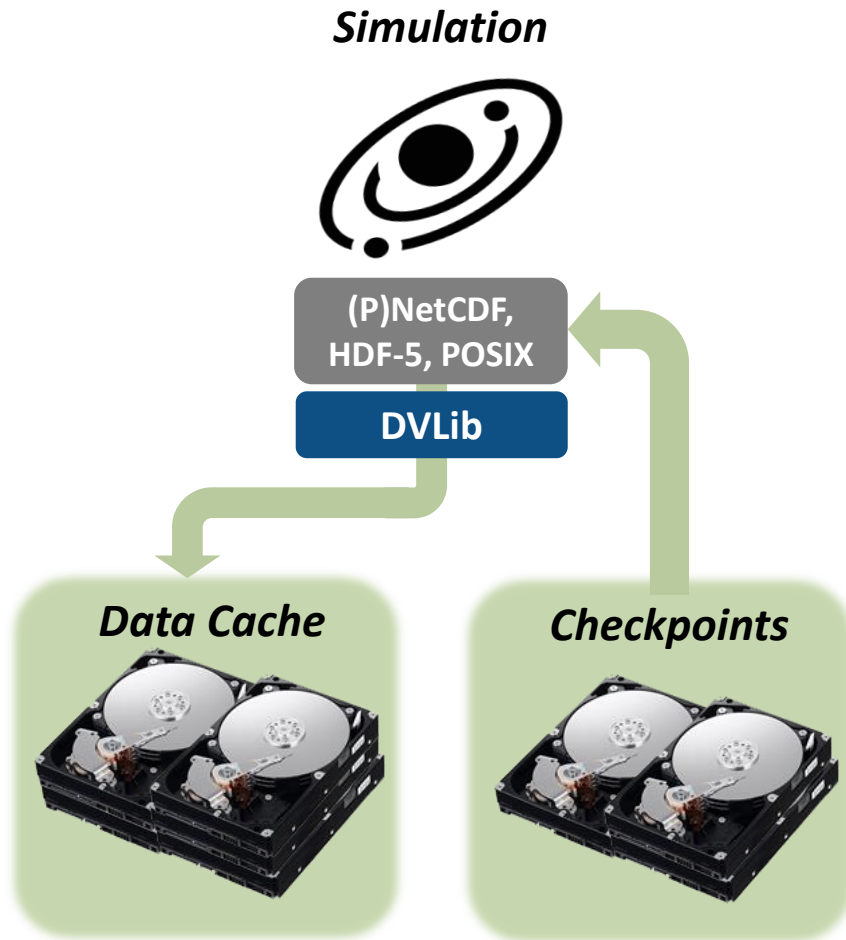


precip_tracking:
3min / TOT_PREC (2D)

fronts_cyclones:
15 min / U, V, T, QV, FI @ 850 hPa (2D)
15 min / TOT_PREC, PMSL (2D)

trajectories:
10 min / U, V, W (3D)

Simulation Contexts



SimFS

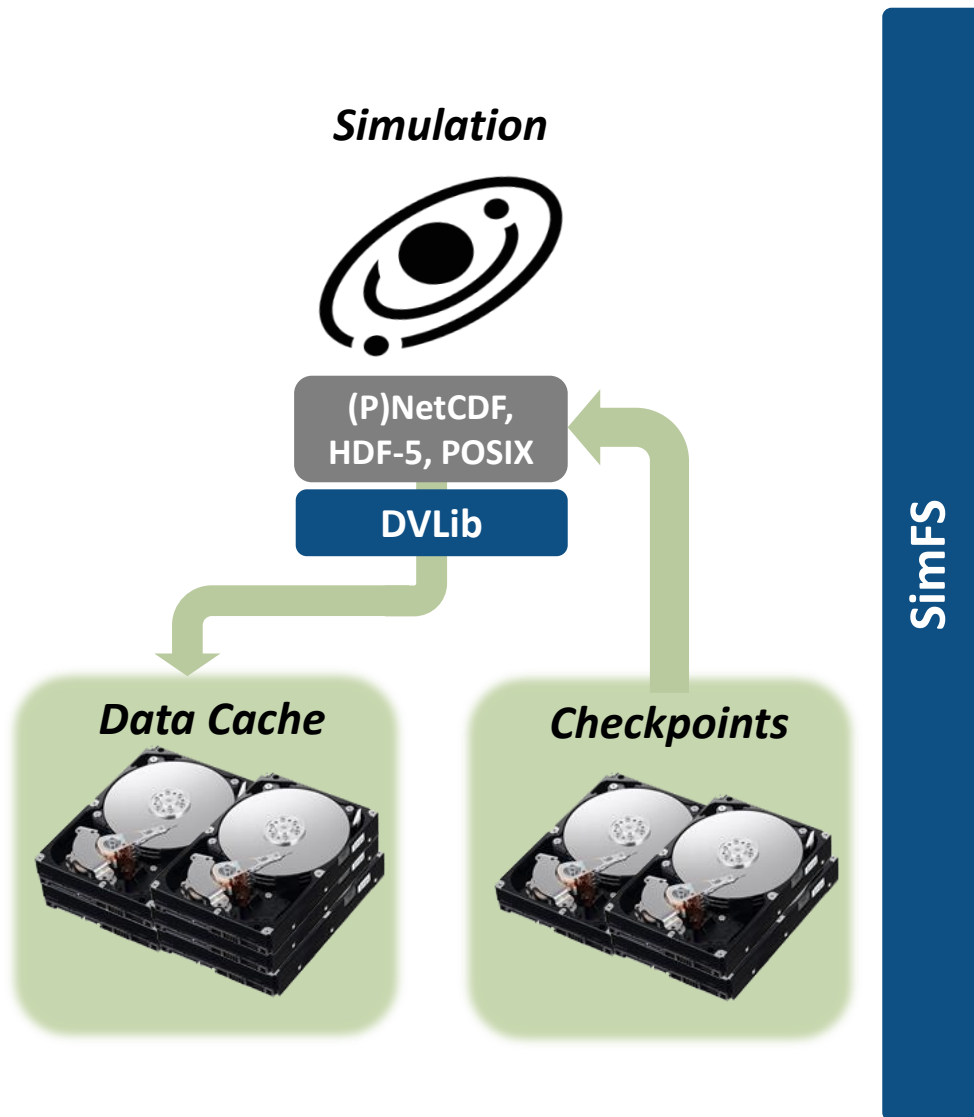


precip_tracking:
3min / TOT_PREC (2D)

fronts_cyclones:
15 min / U, V, T, QV, FI @ 850 hPa (2D)
15 min / TOT_PREC, PMSL (2D)

trajectories:
10 min / U, V, W (3D)

Simulation Contexts



precip_tracking:
3min / TOT_PREC (2D)

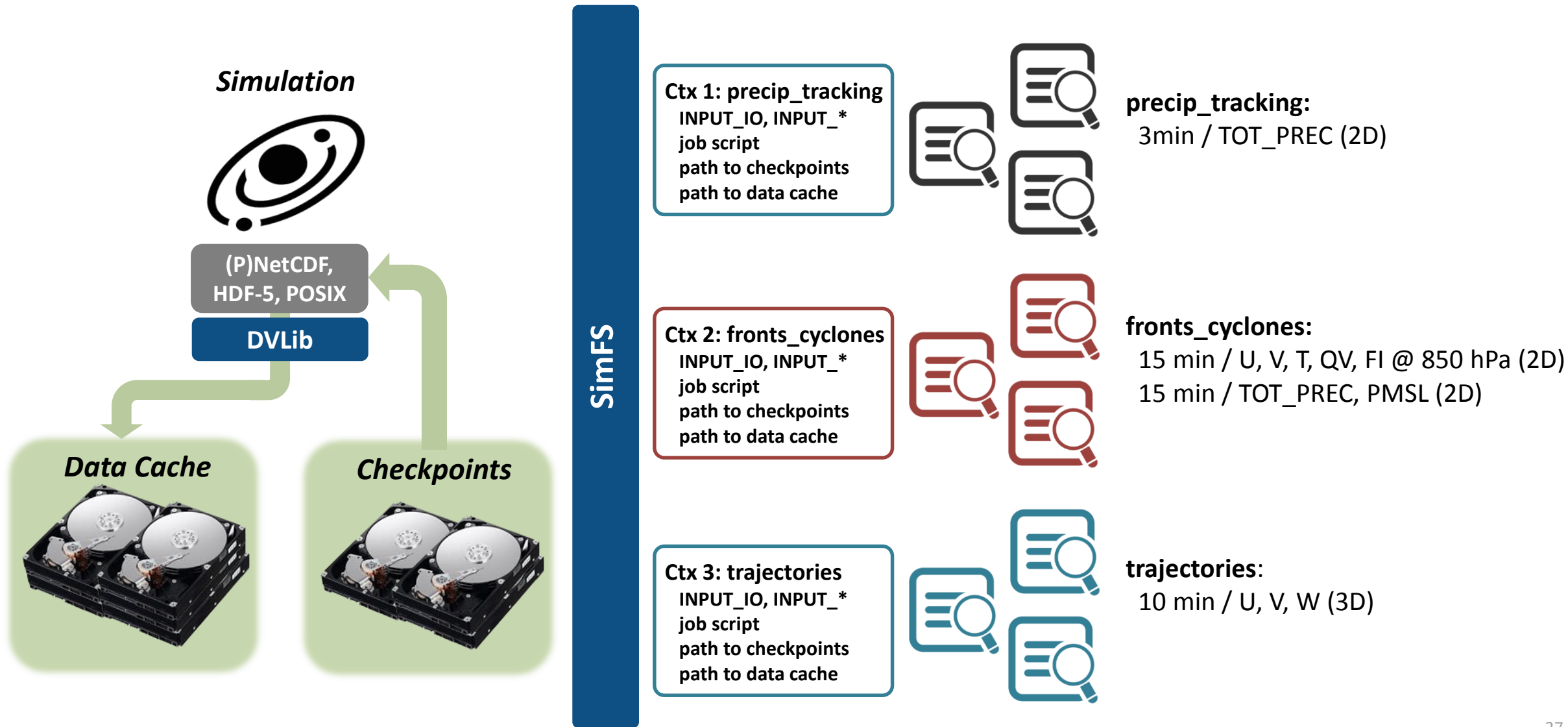


fronts_cyclones:
15 min / U, V, T, QV, FI @ 850 hPa (2D)
15 min / TOT_PREC, PMSL (2D)



trajectories:
10 min / U, V, W (3D)

Simulation Contexts

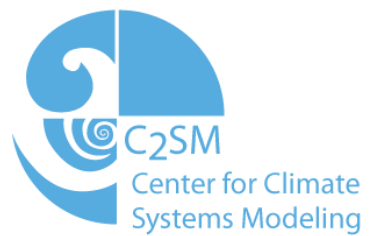
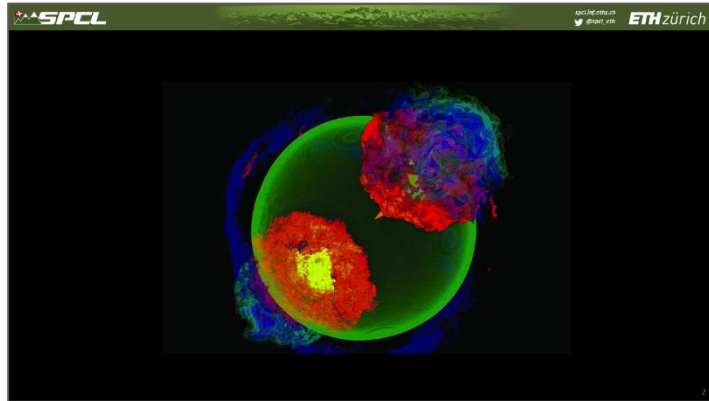




Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra



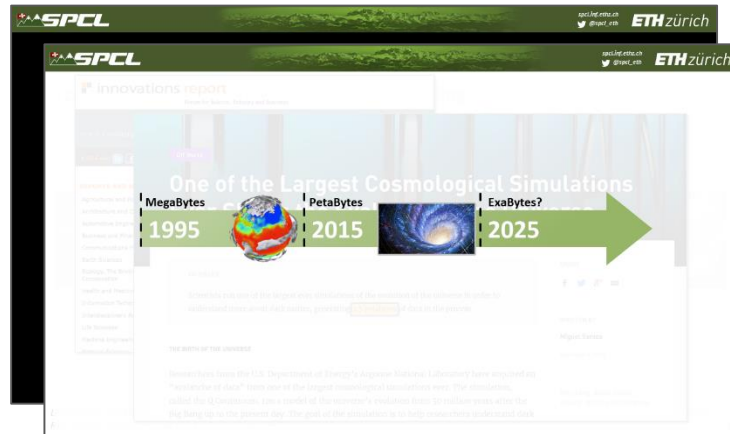
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra



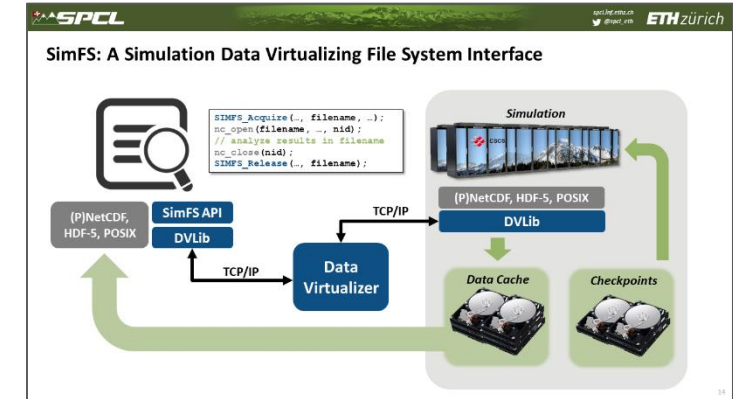
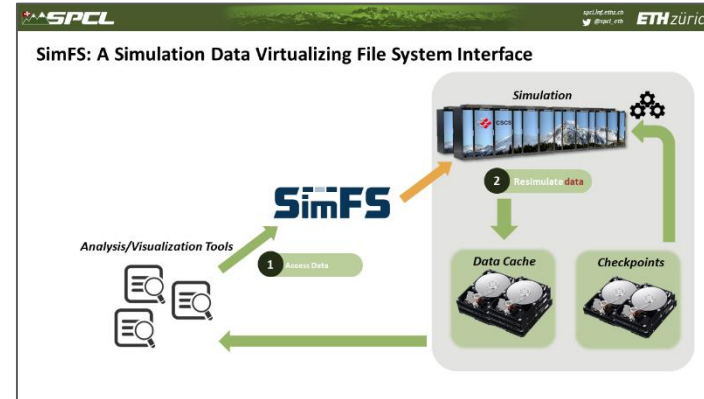
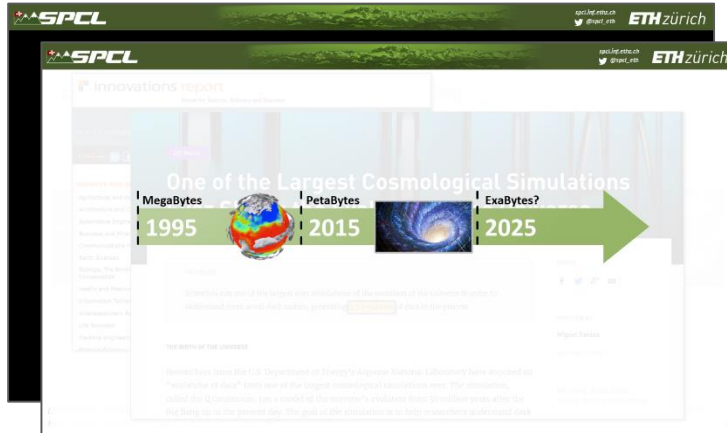
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra



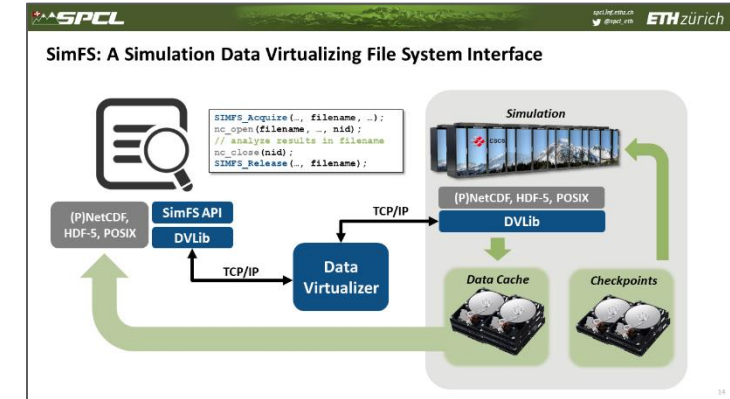
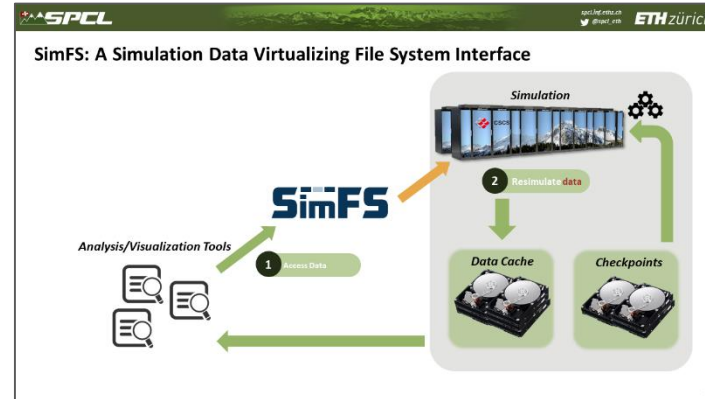
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre



innovations report

One of the Largest Cosmological Simulations

MegaBytes 1995 PetaBytes 2015 ExaBytes? 2025



When is SimFS Cost-Effective?

Data Availability Period: period of time during which analysis can be performed on the simulation data.

$$C_{on-disk}(\Delta t) = C_{sim}(n_{sim}, N) + C_{store}(n_{sim}, s_{sim}, \Delta t)$$

$$C_{SimFS}(\Delta t) = C_{sim}(n_{sim}, P) + C_{store}(n_{sim}, s_{sim}, \Delta t) + \frac{C_{store}(M_{sim}, s_{sim}, \Delta t) + C_{sim}(V(\gamma_{sim}), P)}{C_{store}(M_{sim}, s_{sim}, \Delta t) + C_{sim}(V(\gamma_{sim}), P)}$$

$$C_{in-situ}(\Delta t) = \sum_j C_{sim}(t_j + |\gamma_{sim}(j)|, P)$$

Availability Period	in-situ (x 1000\$)	on-disk (x 1000\$)	SimFS (x 1000\$)
6m	~400	~40	~30
1y	~400	~60	~40
2y	~400	~80	~50
3y	~400	~100	~60
4y	~400	~120	~70
5y	~400	~140	~80

Microsoft Azure Compute: 2.07\$/Node/Hour (NCv2 VM)
Store: 0.06\$/GiB/Month (Azure File Share)

