SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

Extending Storage to the Edge, How It Should Affect Your Storage Strategy

Live Webcast

August 25, 2021

10:00 am PT / 1:00 pm ET

Today's Presenters

Moderator:
Jim Fister
Principal
The Decision Place

Presenter:
Erin Farr
Senior Technical Staff Member
Storage CTO Office
IBM

Presenter:
Vincent Hsu
VP, IBM Fellow, and CTO for Storage and Software Defined Infrastructure
IBM

SNIA-at-a-Glance

180
industry leading organizations

2,500
active contributing members

50,000
IT end users & storage pros worldwide

Learn more: **snia.org/technical** **@SNIA**

SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

What We Do

Educate vendors and users on cloud storage, data services and orchestration

Support & promote business models and architectures: OpenStack, Software Defined Storage, Kubernetes, Object Storage

Understand Hyperscaler requirements Incorporate them into standards and programs

Collaborate with other industry associations

SNIA Legal Notice

The material contained in this presentation is copyrighted by the SNIA unless otherwise noted.

Member companies and individual members may use this material in presentations and literature under the following conditions:

- Any slide or slides used must be reproduced in their entirety without modification
- The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.

This presentation is a project of the SNIA.

Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.

The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

Agenda

- Edge Computing use cases are driving additional patterns of data movement
  - Challenges associated with these use cases
  - Considerations for distributed cloud storage architectures

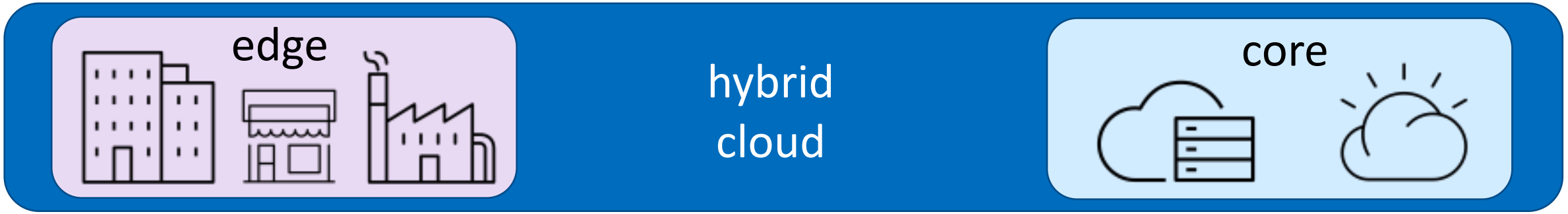
Edge Computing

Data movement pattern #1

For Today's Discussion…

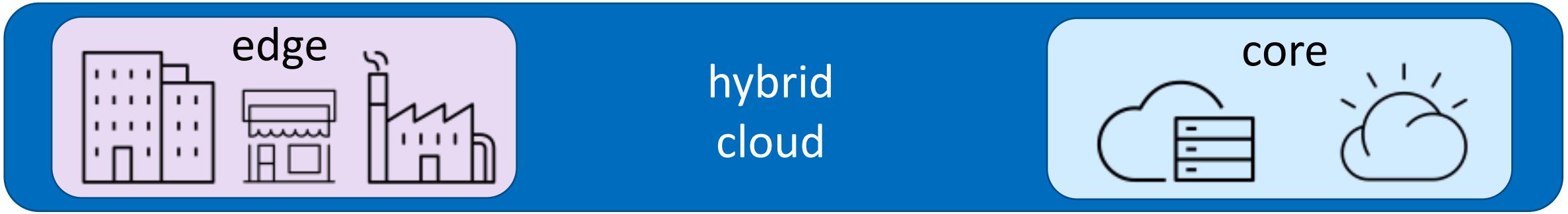
- Edge
  - Where data is being generated
- Examples include:
  - Branch offices
  - Retail stores
  - Research Labs
  - Industry/Manufacturing sites
  - Automobiles

- Core
  - Centralized location for data and business workloads
- Examples include:
  - Hybrid, Public and Private clouds
  - IT Datacenters



What's Driving Compute to the Edge?

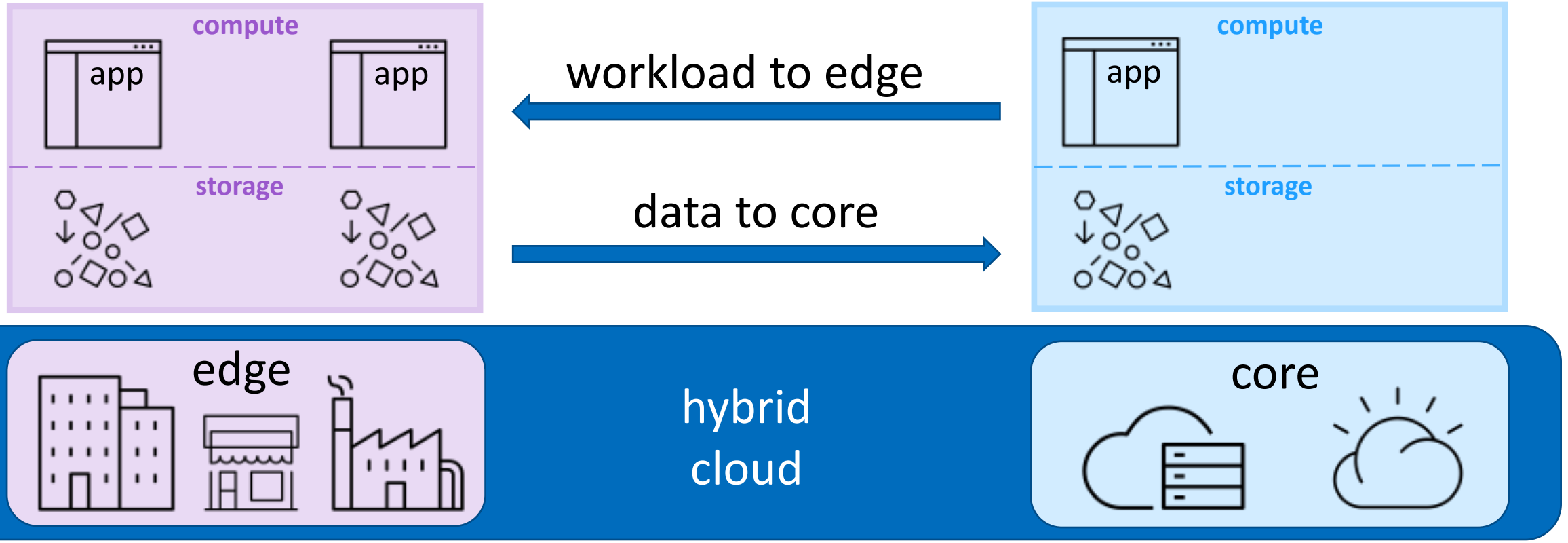
- Data gravity
  - Easier to move the compute than the large amounts of data
- Regulations/Compliance
  - Some data can't cross national or state borders
- Security
  - Less copies reduce the attack surface
- Cost
  - Moving data and making copies of data is expensive
- Performance
  - Faster response times where you need it
- Unreliable connectivity
  - Remote or portable environments can't rely on connectivity to core



Familiar Patterns for Edge

New workloads are being created to gain insights from dark data

Existing workloads on the core where data is sent from the edge… are getting migrated to the edge to avoid data movement

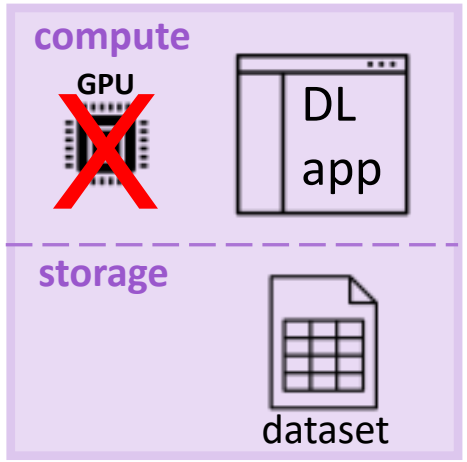


But What Happens When…

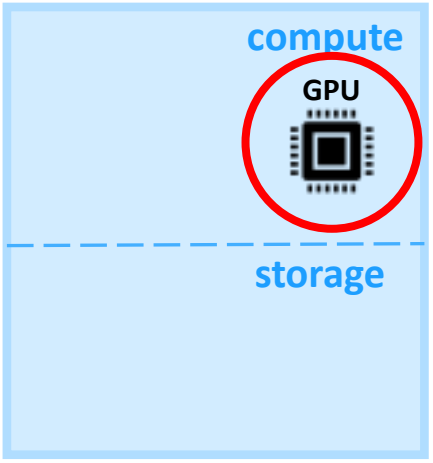
- Your data scientist wants to use deep learning against data at the edge site
- DL requires a GPU for faster training

- You don't have a GPU on the edge site
- It is cost prohibitive to put GPUs in all of your edge sites

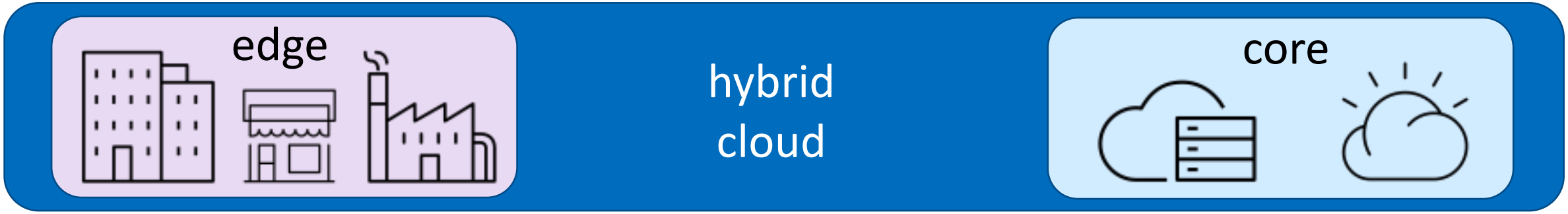
Data is here



- She decides to do training on the core site
- Inferencing needs to remain on the edge



GPU is here



So, She Copies the Dataset Herself

But first, she needs to talk to three different teams

May I make a copy of this data?

Where can I run my workload?

Where can I put my dataset?

Data governance team
- Manage permission and access to data
- Data replica management

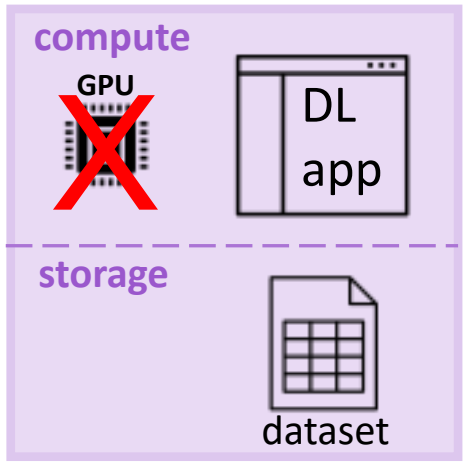
Cluster management team
- Manage the compute backbone
- Resource management
- Host compute tenants from different business units

Storage management team
- Manage the central storage
- Provision storage to compute cluster

Data Management is Manual and Error-prone

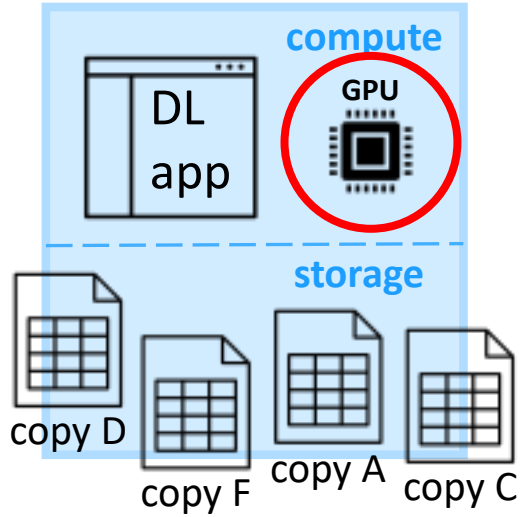
Keeps track of which datasets were used in a spreadsheet

Performs training (8x faster)



copies model to edge

she copies the data using new filenames



Data of record is here

Over time, there are multiple copies of data

- Over time, this creates data sprawl
- Data scientist has trouble recreating results because she can't find the correct dataset used.

Some copies are deleted to free space

Cloud Bursting

- An application running in a data center, private cloud or edge environment can rapidly scale compute, storage or memory by “bursting” into a public cloud if it needs more resources

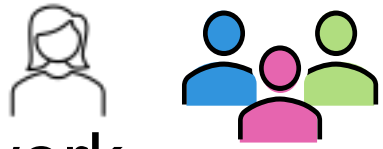
- Can burst for performance or additional capacity

Advantages

- Performance
- Cost
  - only pay when there is a demand
- Availability
  - avoids interruption of services

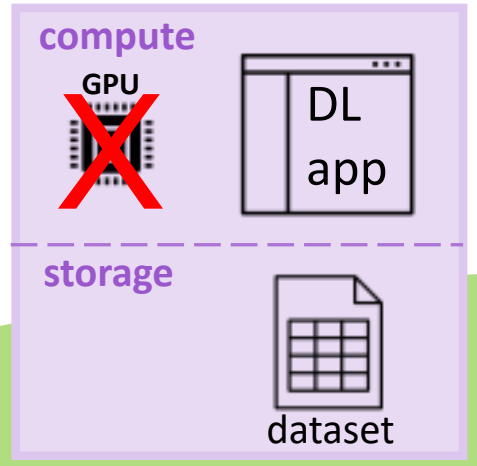
How Can Storage Help with Cloud Bursting?

- Single source of truth for your data
- Data scientist can focus on data science work

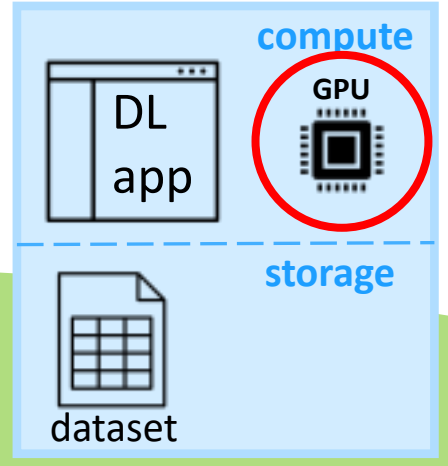


Get approvals and resources for a cluster (one time)

Create cluster with single namespace



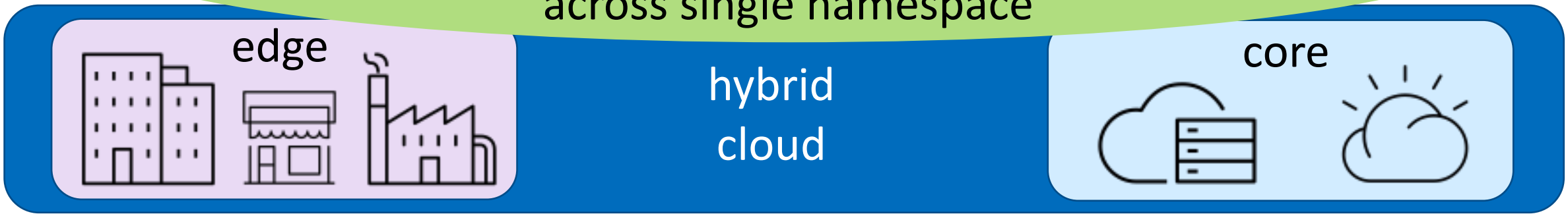
model pushed to edge



Train model (8x faster)

Storage evicts data

Storage prefetches data across single namespace



Cloud Bursting – Data and Storage Considerations

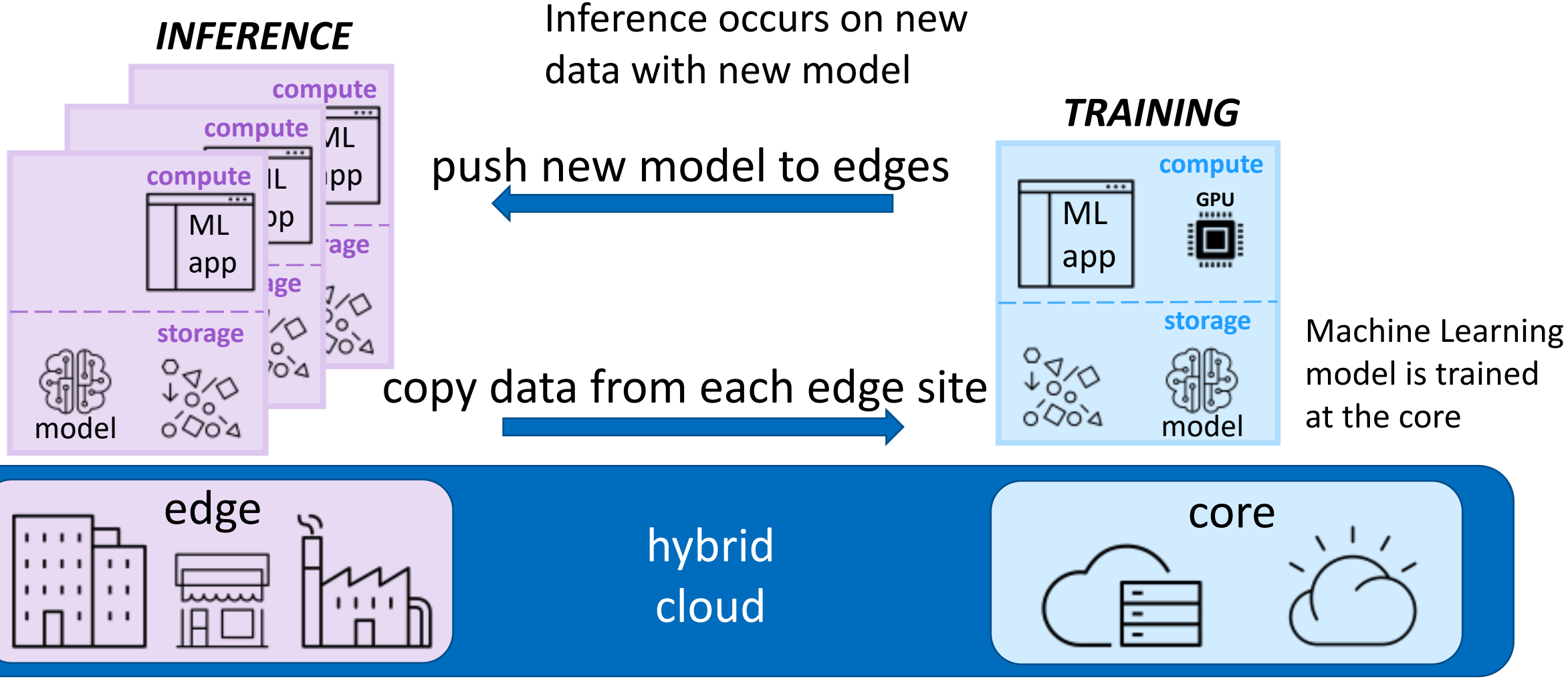
- Data Replication
  - Move the data with the applications seamlessly
  - Still need to consider data locality and compliance
- Data Consistency
  - If application is stateful, need to ensure data consistency
- Data Protection
  - The data needs secured across locations
- Data Eviction
  - Once data is used and no longer needed, clean it up
- Efficiency
  - How to do everything above with the least amount of human effort?

Edge Computing

Data movement pattern #2

Machine Learning with Data from Multiple Edge Sites

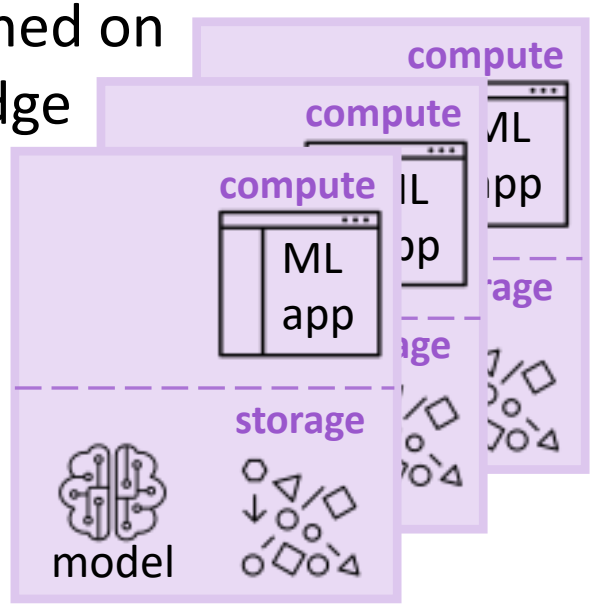
What usually happens today…



Federated Learning

Models are retrained on the edge

TRAINING** and **INFERENCE



Inference occurs on new data with new model

push new model to edges

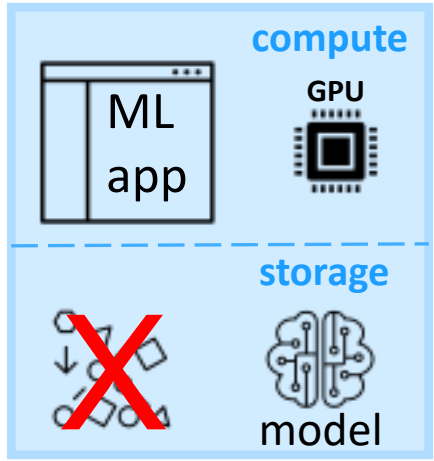
request retrain

pull models from edge

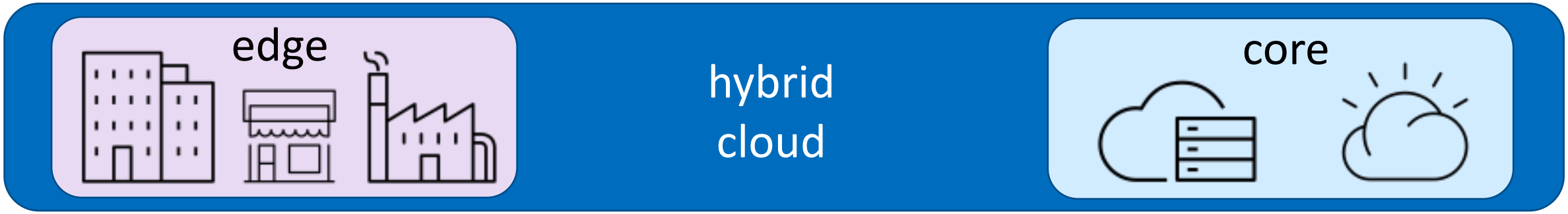
(not raw data)

Machine Learning model is trained on the core

TRAINING



Aggregator collects models from each edge site, retrains, and redistributes new models



Value of Federated Learning

- Improve model training across locations
- Address data privacy, locality and security
- Adhere to regulatory compliance
- Tackle data volumes at lower cost and risk (e.g., minimize egress charges)

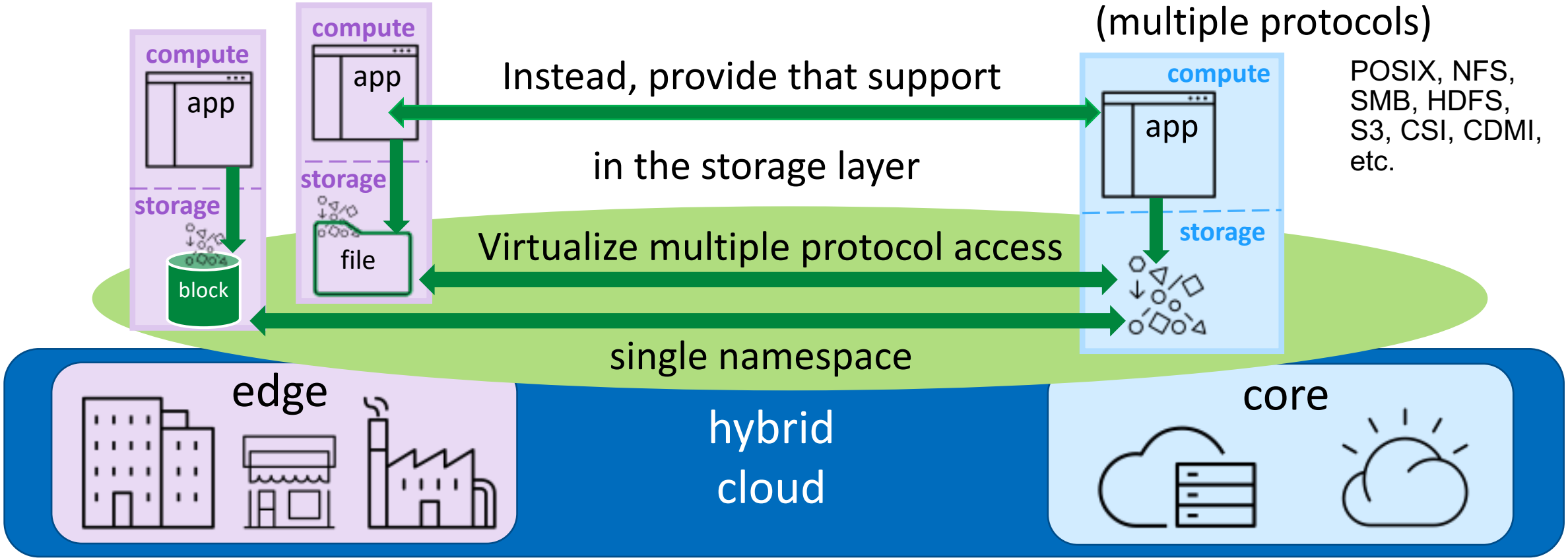
- Application spans both Edge and Core
  - Most data management is handled at the **application layer** today
  - If handled at the **storage layer**, that can simplify application complexity and improve time to market

Federated Applications – How Can Storage Help?

- Move data with the applications seamlessly
- Provide access to data regardless of how it's stored

Application itself contains logic to

- Move data across sites
- Support heterogeneous data (multiple protocols)



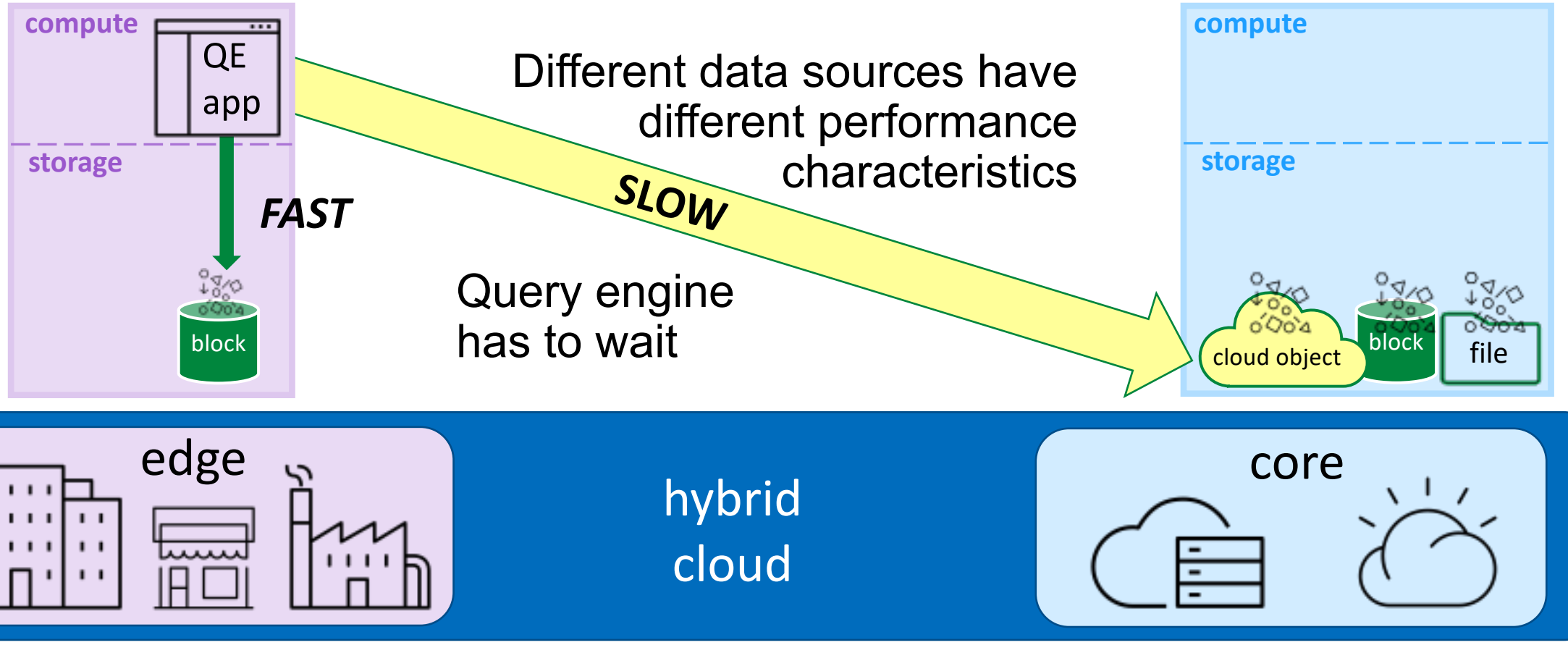
Edge Computing

Data movement pattern #3

Data is Everywhere

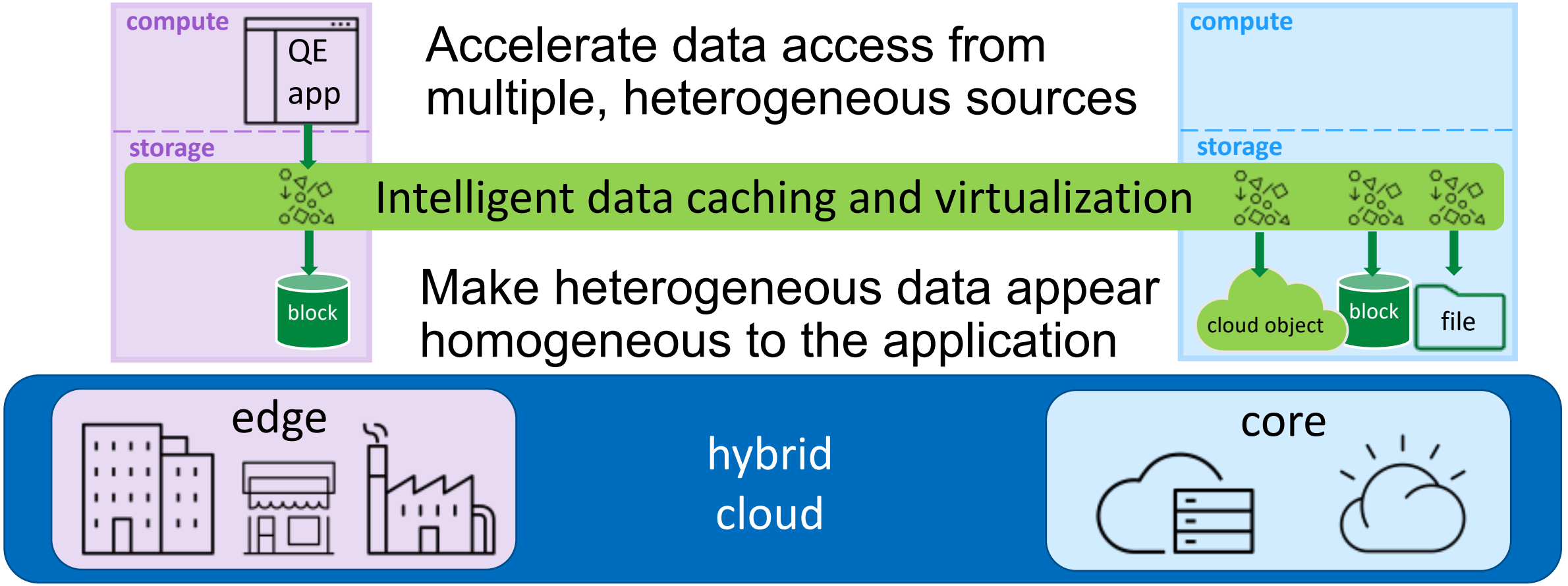
Query engine runs on the edge so the data can stay there

However, you also want to use data located on the core site

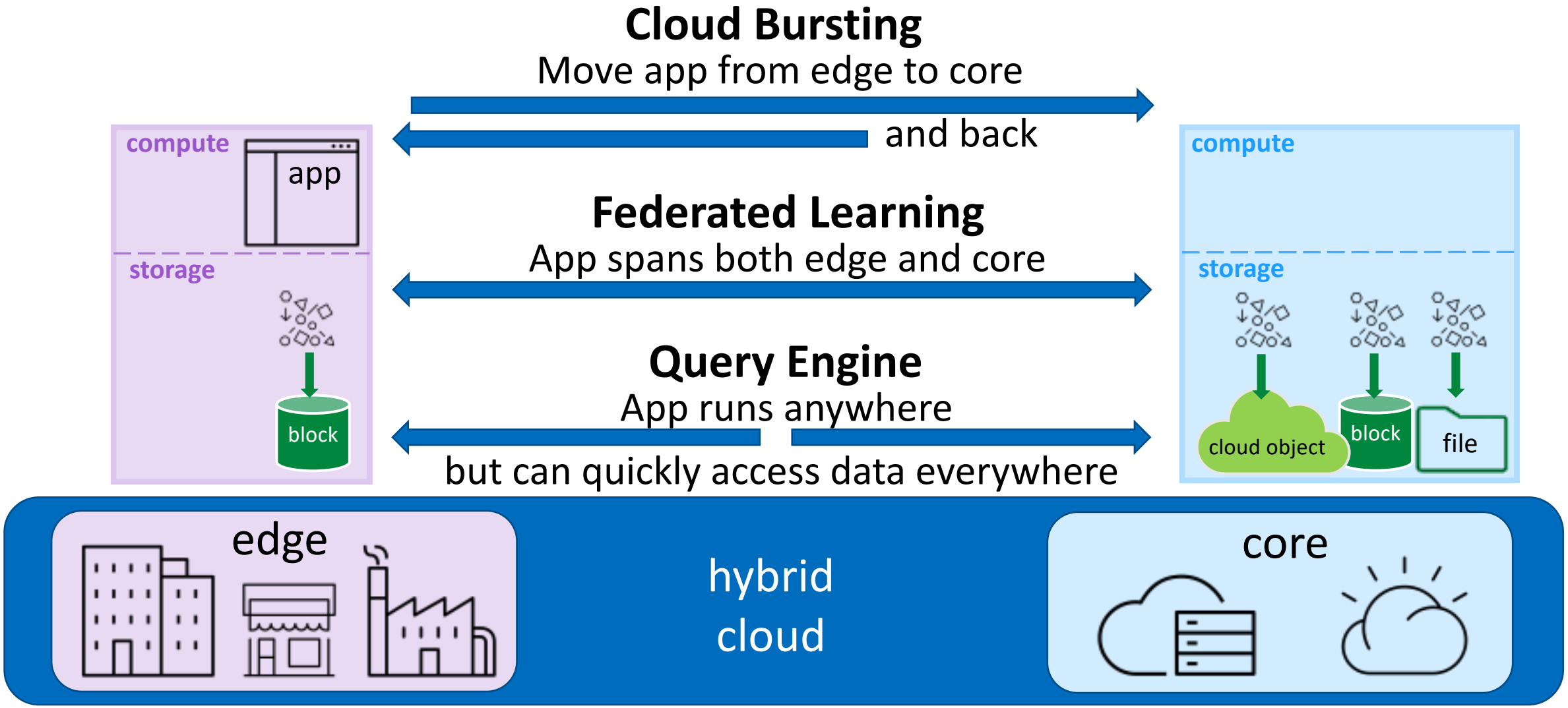


How Can Storage help?

- Provide fast access to data regardless of where it lives

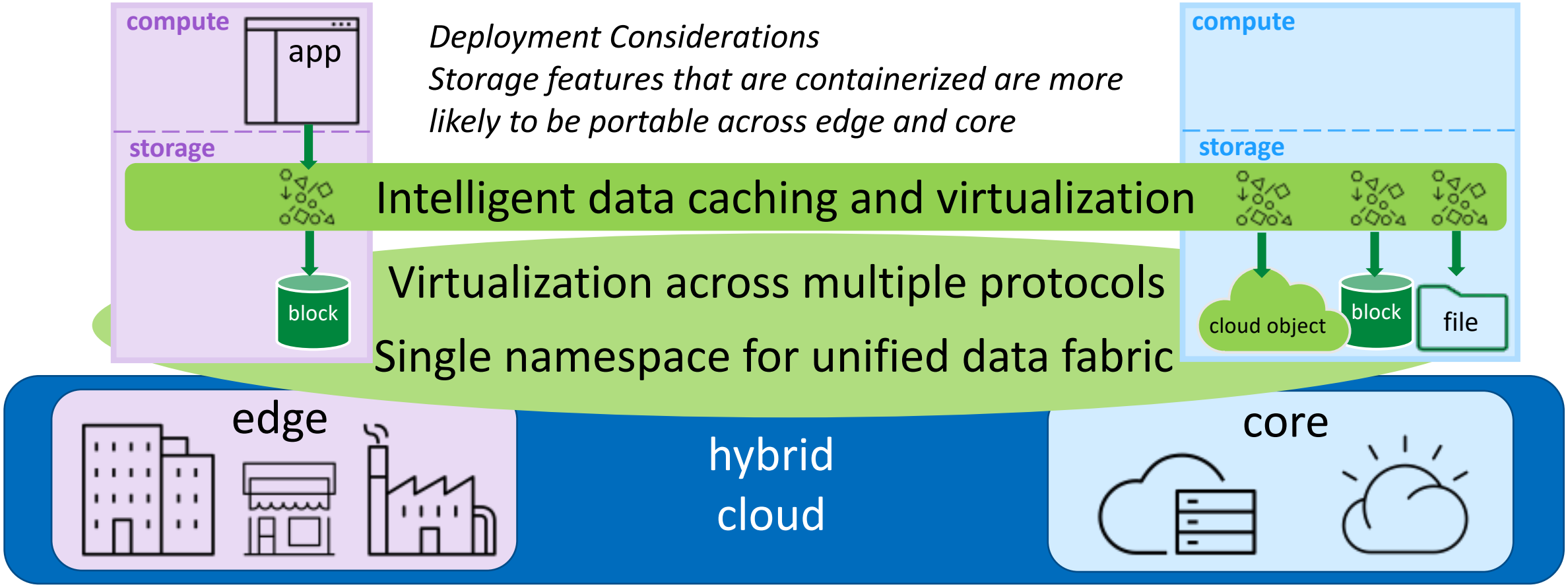


Summary of Edge Use Cases



Summary of Storage Considerations for Edge Data Patterns

- Provide a single source of truth for the data
- Fast access to data regardless of how it's stored and where it lives



Thanks for Viewing this Webcast

Please rate the webcast and provide us with feedback

This webcast and a copy of the slides will be available at the SNIA Educational Library https://www.snia.org/educational-library

A Q&A from this webcast will be posted to the SNIA Cloud blog: www.sniacloud.com/

Follow us on Twitter @SNIACloud

Thank you

Questions?