dCache - data access optimization in a hybrid cloud

Michael Schuh for dCache people

About dCache

- Joined effort between DESY(2000), FNAL(2001) and NDGF(2006)

- Provides storage solution for scientific data

- Supports standard and HEP specific access protocols

- Supports standard and HEP specific authentication mechanisms

- Developed for HERA and Tevatron, used for LHC and others

  - WLCG, Belle II, LOFAR, CTA, IceCUBE, EU-XFEL, Petra3, DUNE, And many more ...

dCache: Motivation

- Data never fits into a single server
  - Multiple servers
  - Off-load to tape
- Growing number of client hosts
  - Main frame vs. Linux cluster
- Control over HW/OS selection
  - Better offers
  - Local expertise

Key design elements

dCache.org

- Single-rooted namespace, distributed data

- Client talks to namespace for metadata operations only

- Bandwidth and performance grow with number of Pool nodes (data servers)

- Standard clients (OS native or experiment framework)

- Same data can be provided by any access protocol and security flavor
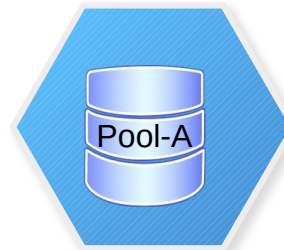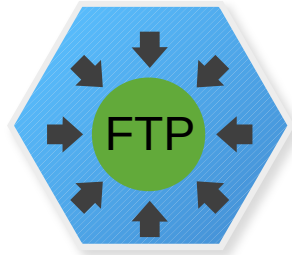
Four main components

- **DOOR**
  - protocol specific user entry points (NFS, FTP, DCAP, XROOT)
- **POOL**
  - data storage nodes, talk all protocols
- **Namespace**
  - metadata DB, POSIX layer
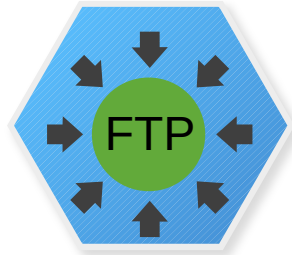- **PoolManager**
  - request distribution unit
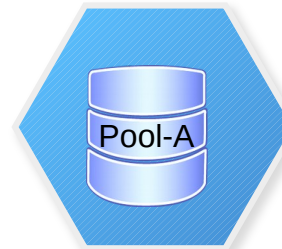
Minimal Setup

FTP

Namespace

PoolMgr

Pool-A
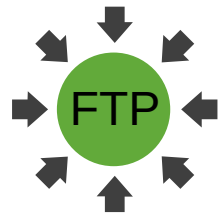
Minimal Setup

Minimal Setup

FTP

Namespace

PoolMgr

All components are **CELL**s :
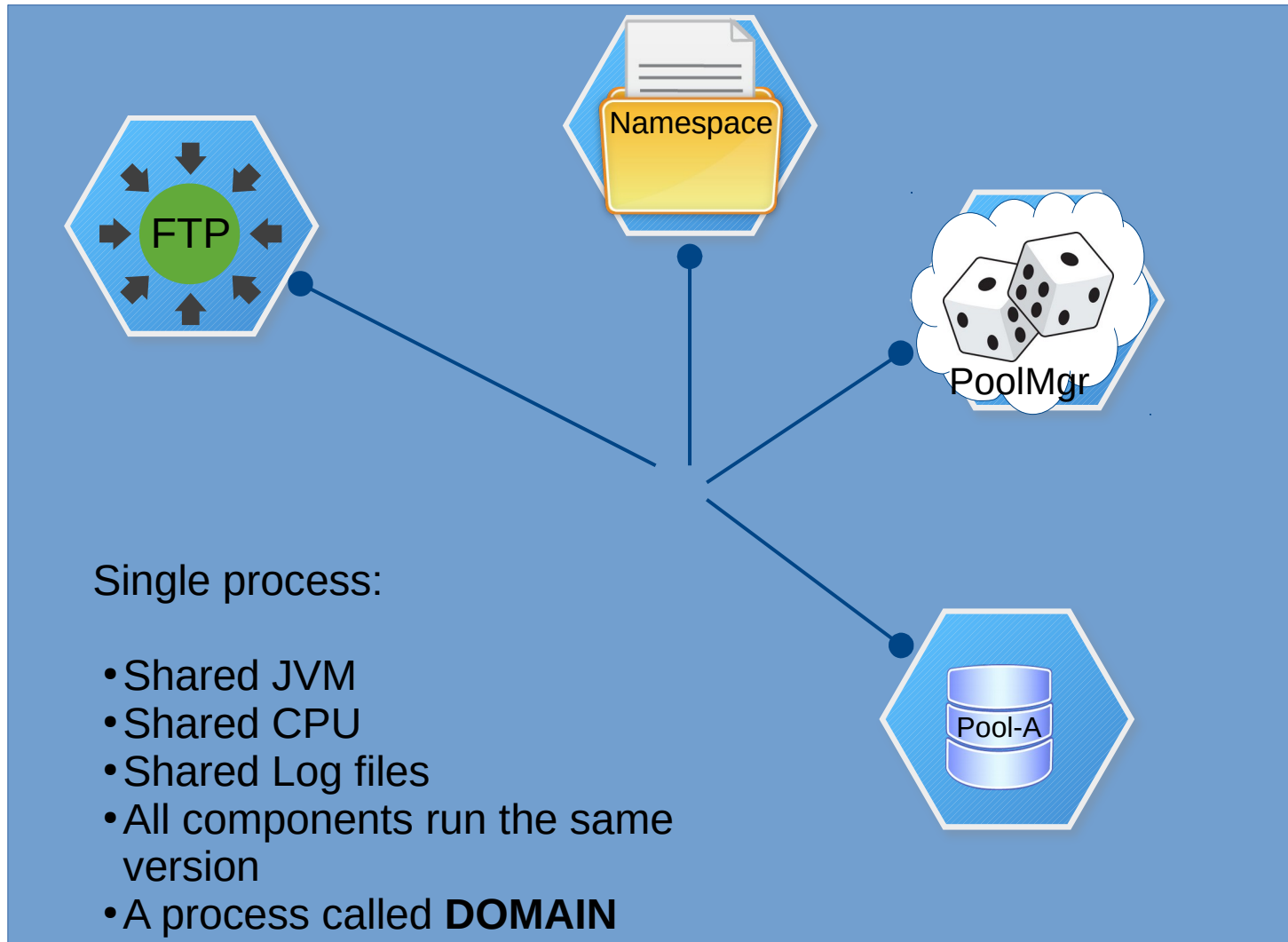they are independent and can interact with
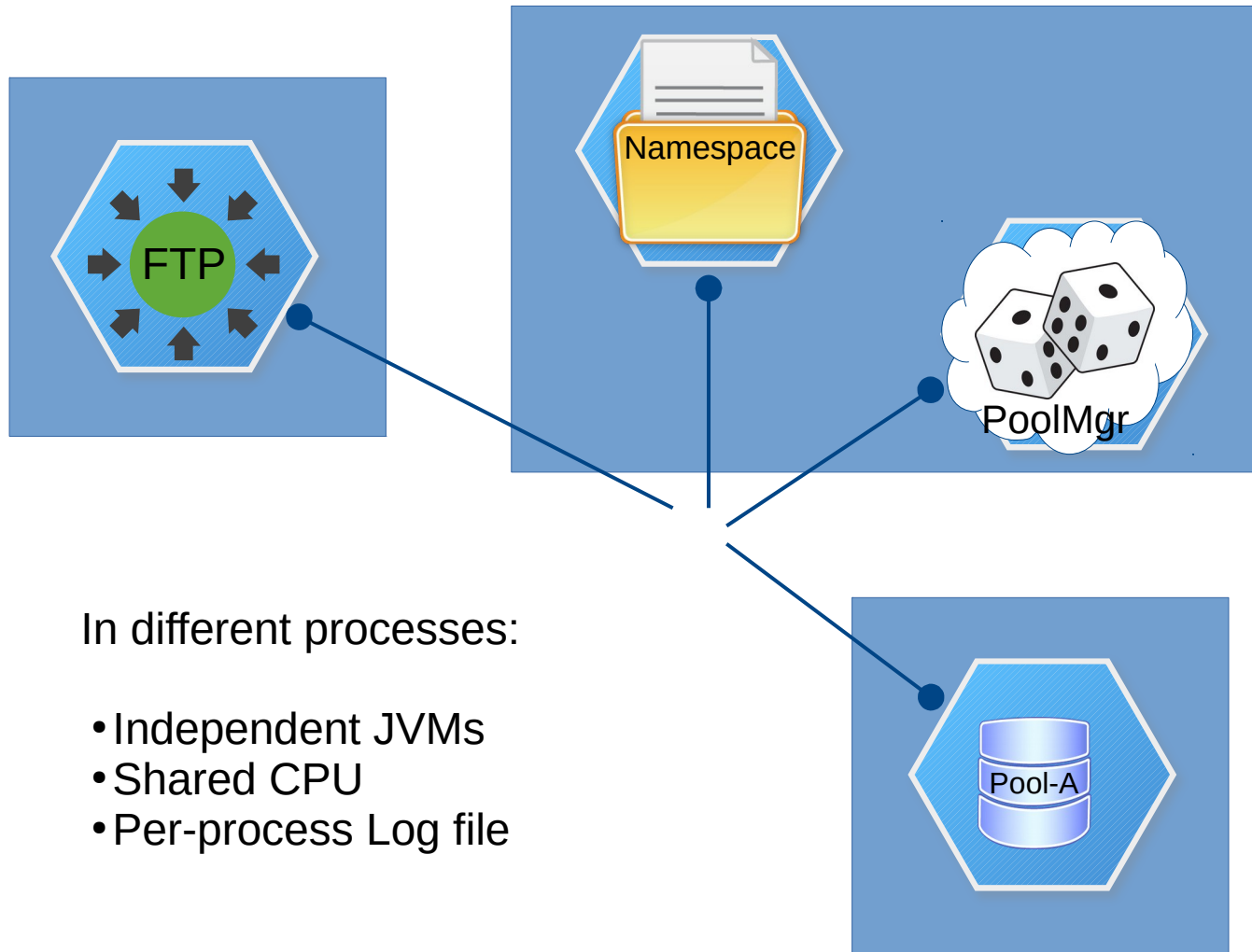each other (send messages).

Pool-A

Grouping CELLs

Grouping CELLs
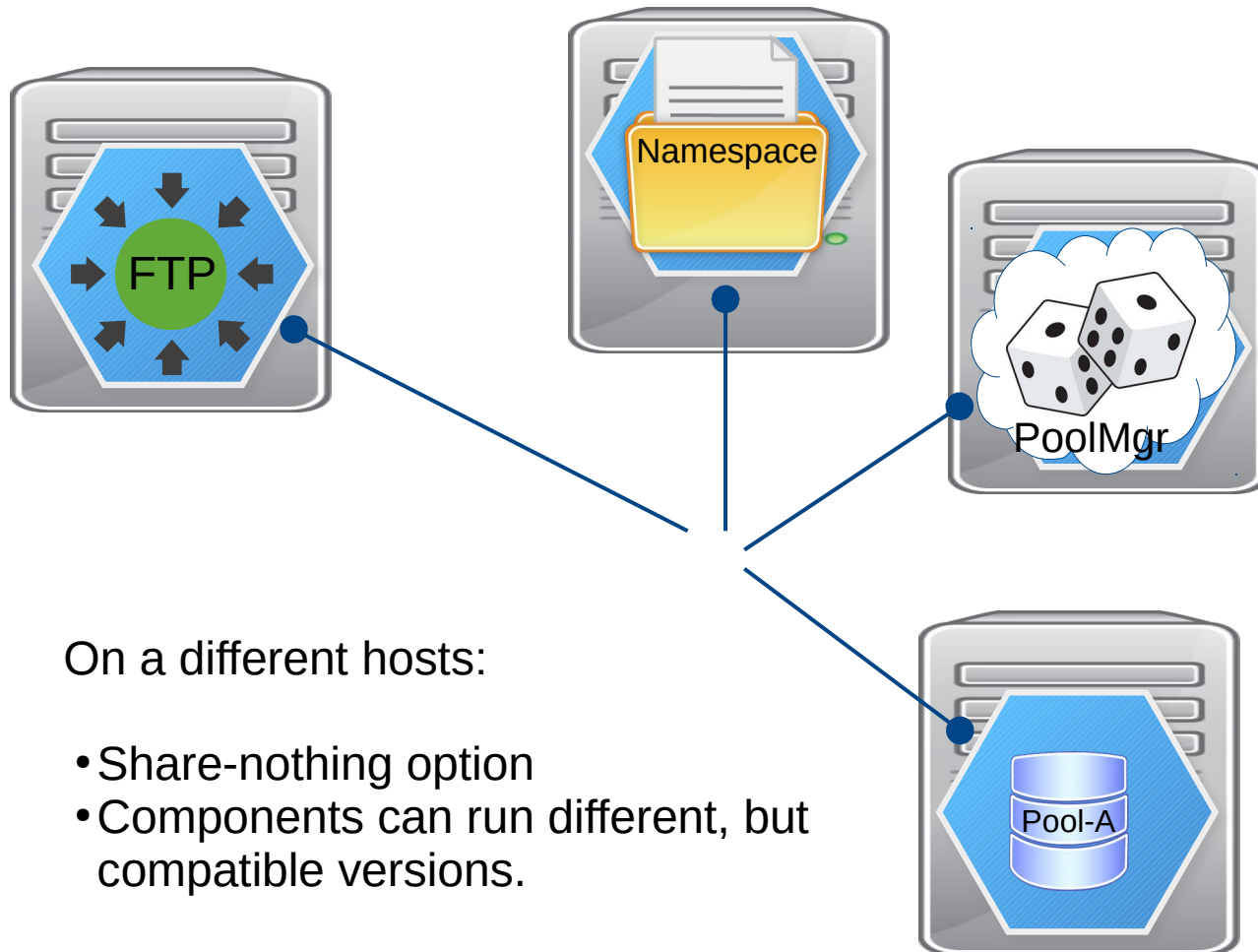
Namespace

FTP

PoolMgr

Pool-A

Single process:

• Shared JVM
• Shared CPU
• Shared Log files
• All components run the same version
• A process called **DOMAIN**

Grouping CELLs

In different processes:

- Independent JVMs
- Shared CPU
- Per-process Log file

Labels in figure: FTP, Namespace, PoolMgr, Pool-A

Grouping CELLs

FTP

Namespace

PoolMgr

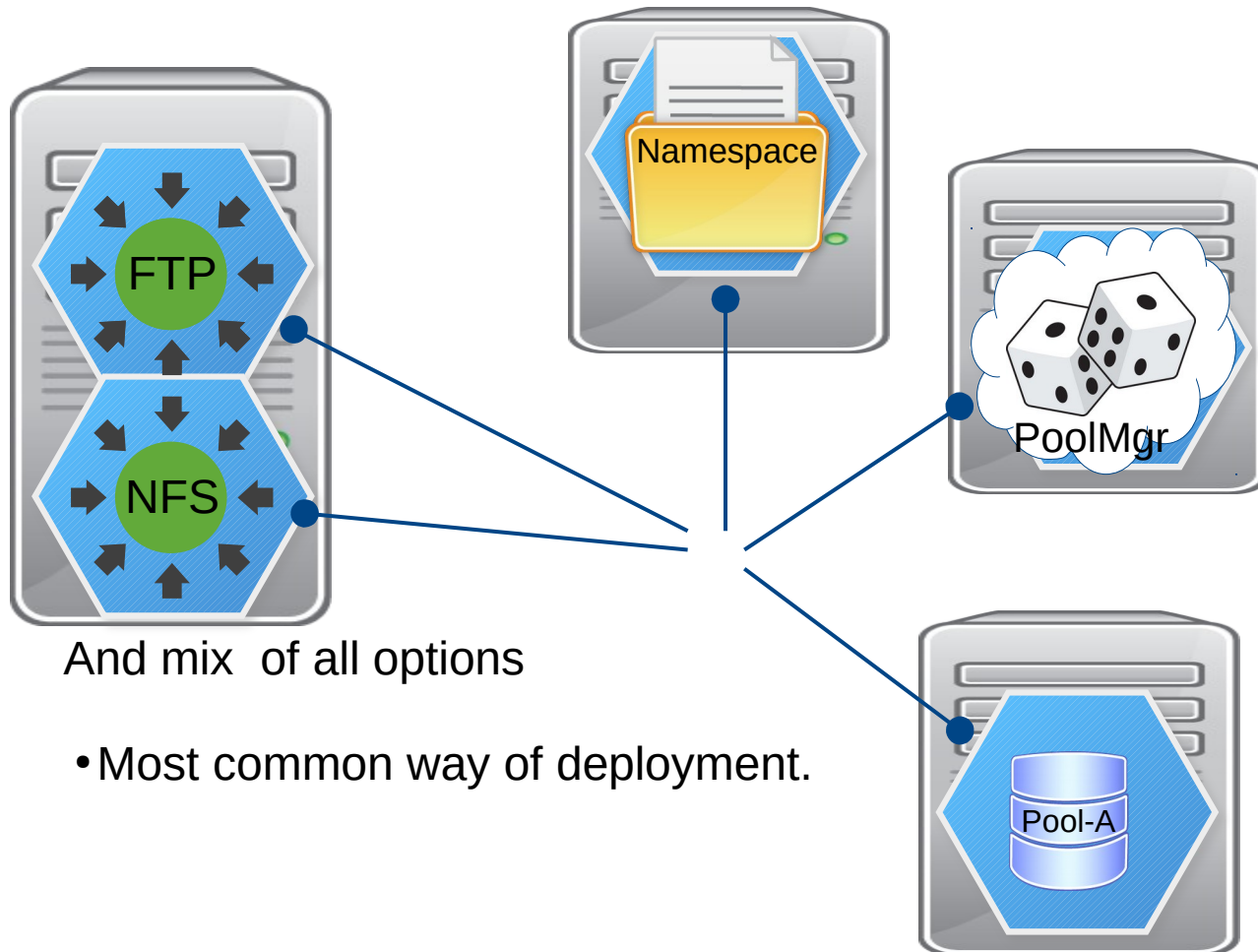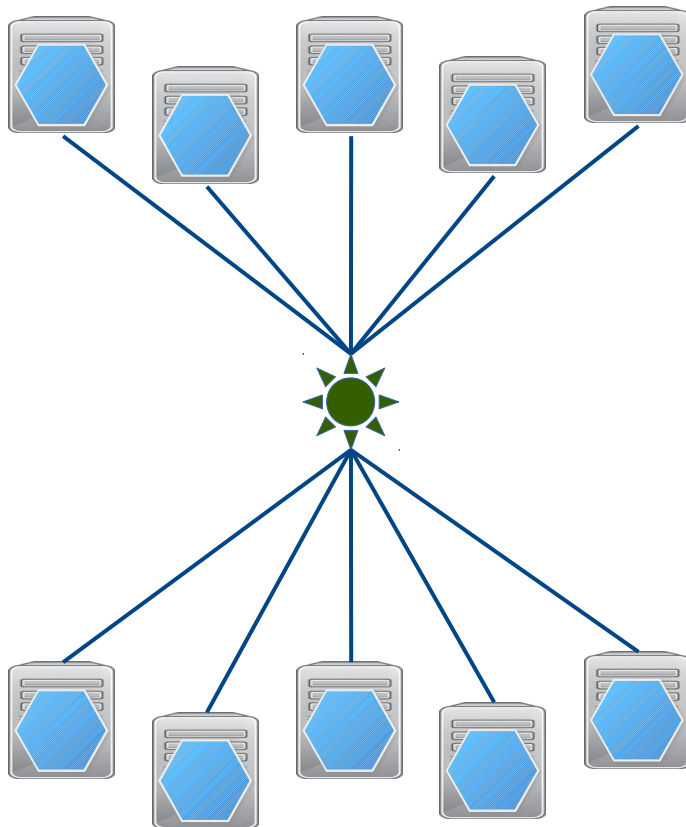Pool-A

On a different hosts:

- Share-nothing option
- Components can run different, but compatible versions.

Grouping CELLs

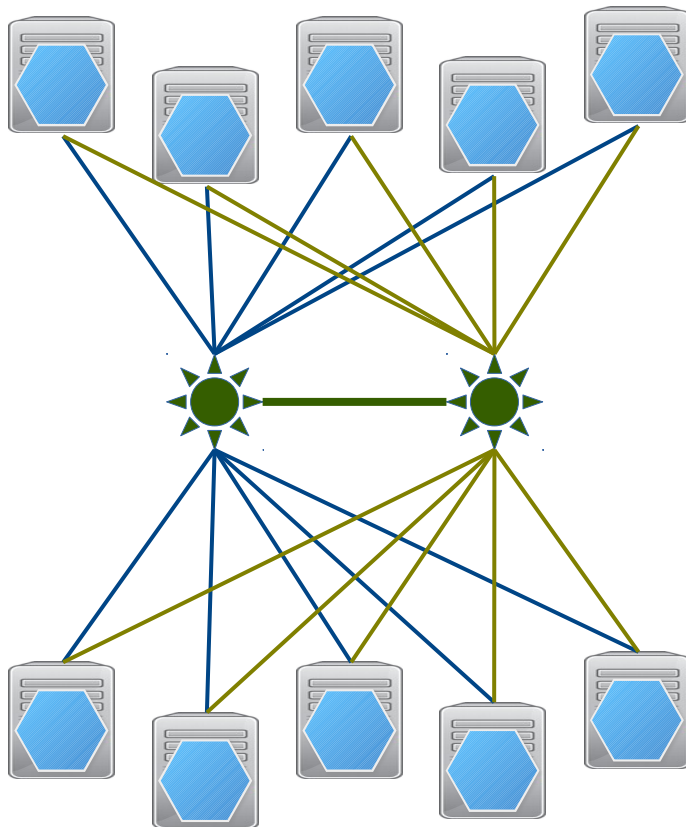And mix of all options

- Most common way of deployment.

Cell messaging 101

- Star like topology
- Selected node configured as a hub called **CORE** domain
- All communication goes through CORE domain
- Other domains called **SATELLITE**
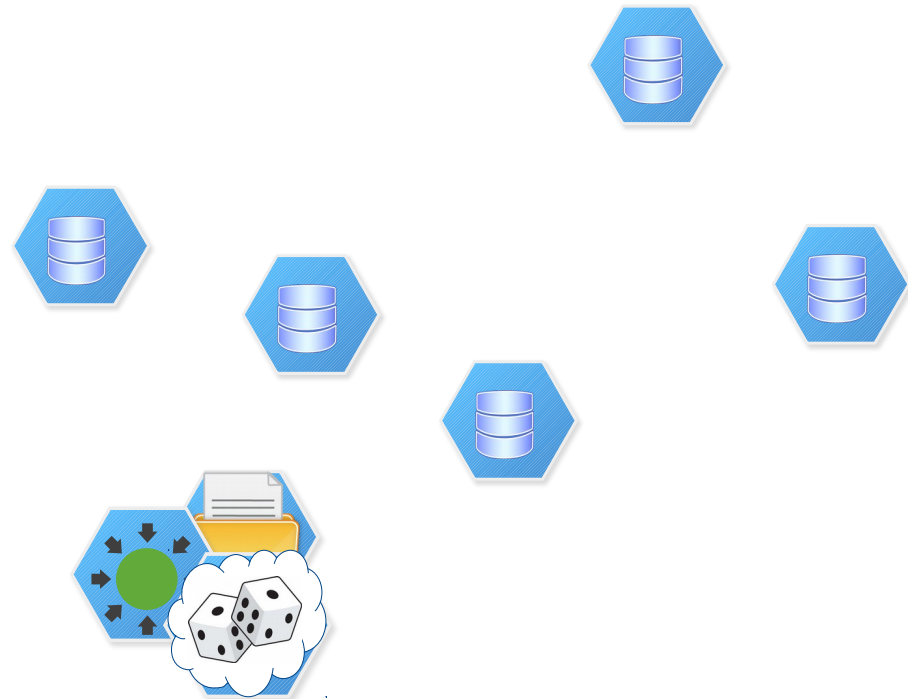
Cell messaging 101

- Star like topology
- Selected nodes configured as hubs called **CORE** domains
- All communication goes through **CORE** domains
- Multiple **CORE** domains make communication fault tolerant

Multi-Site deployment
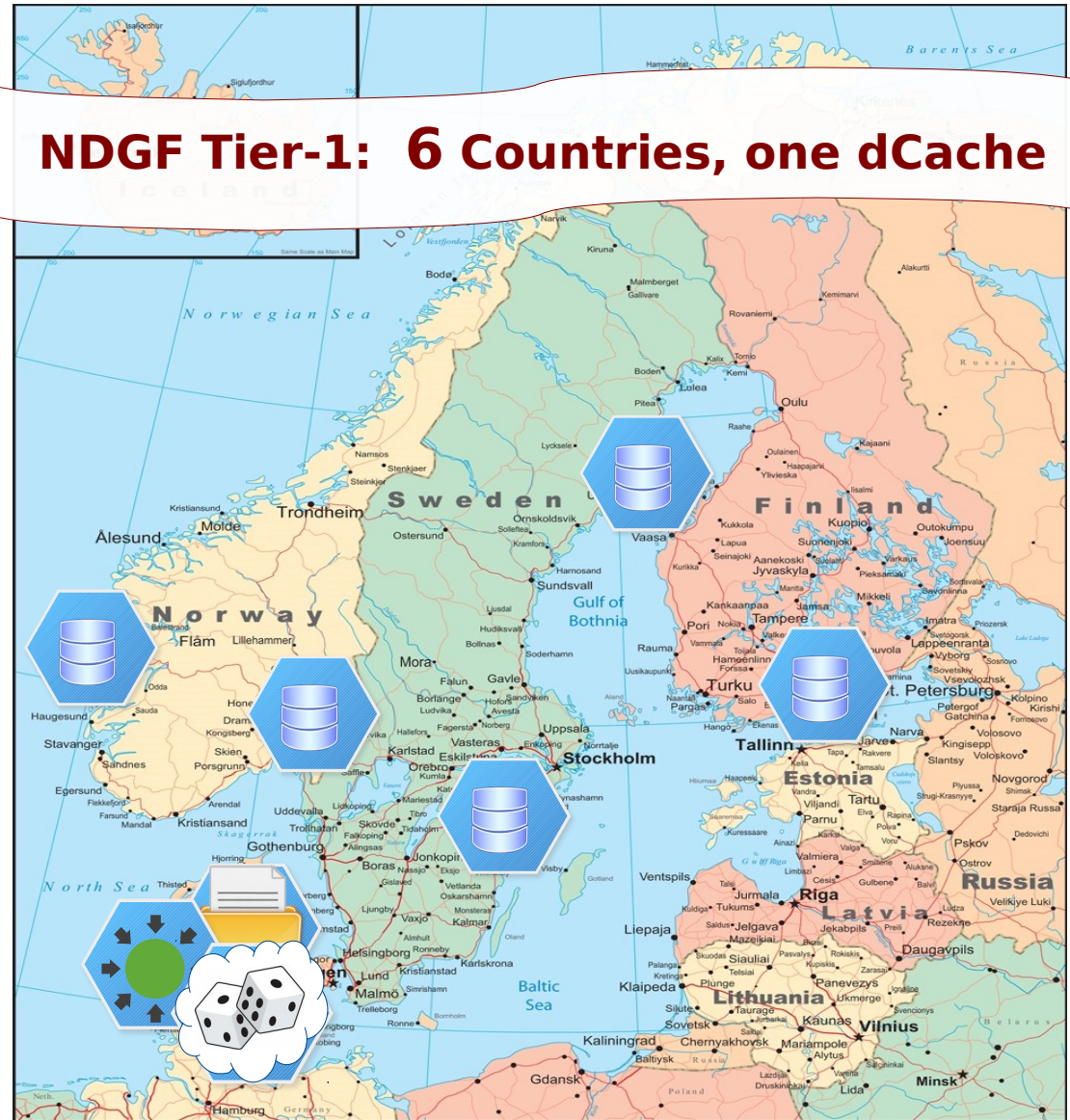
Multi-Site deployment

- Distribute data over multiple locations

- Multiple administrative domains

- Use available resources

Multi-Site deployment

- Distribute data over multiple locations

- Multiple administrative domains

- Use available resources

NDGF Tier-1: **6 Countries, one dCache**

Multi-Site deployment

- Works for all protocols

- Support HSM connectivity

  - Each site/pool may have it's own tape system

- Pools may run different major versions

  - Site has two years to upgrade pools

Multi-Site deployment

- Preferred write location depending on IP (location) or directory path (if requested)

- Preferred 'local' read access if data is available

- Replication

  - **On Demand**, when requested from remote site

  - **Permanent**, data protection, location adjustment

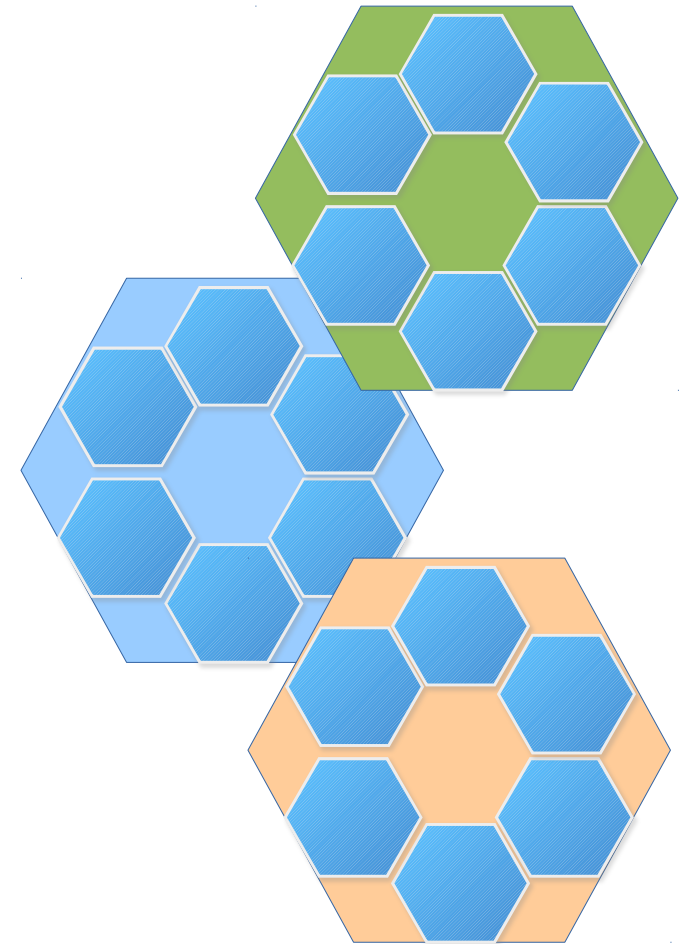 - **Manual**, for data location optimization, maintenance
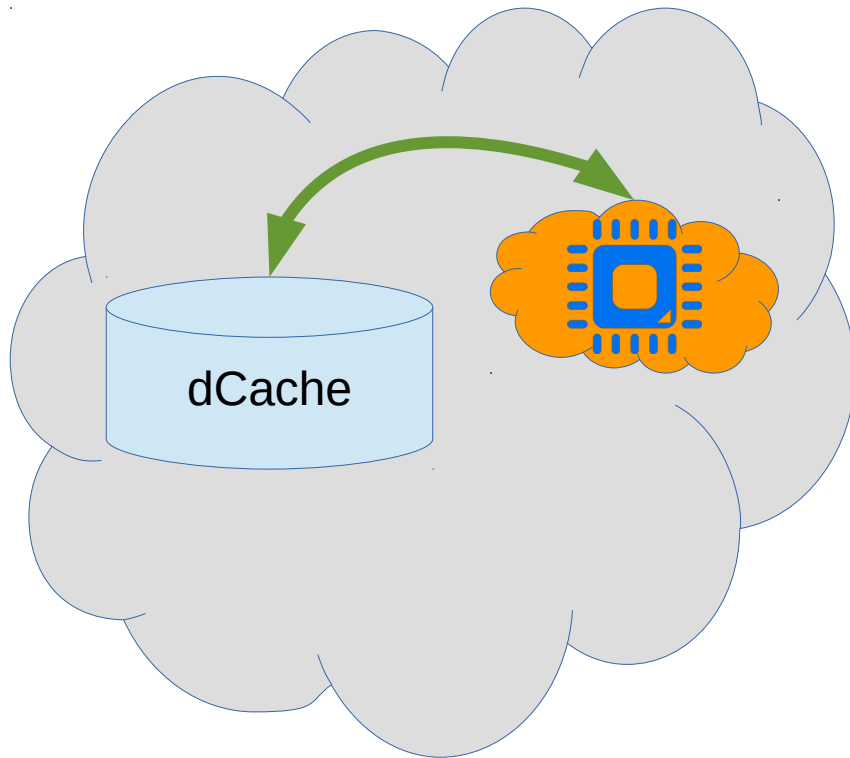
Pool selection internals

Pool Groups

Request Parameters
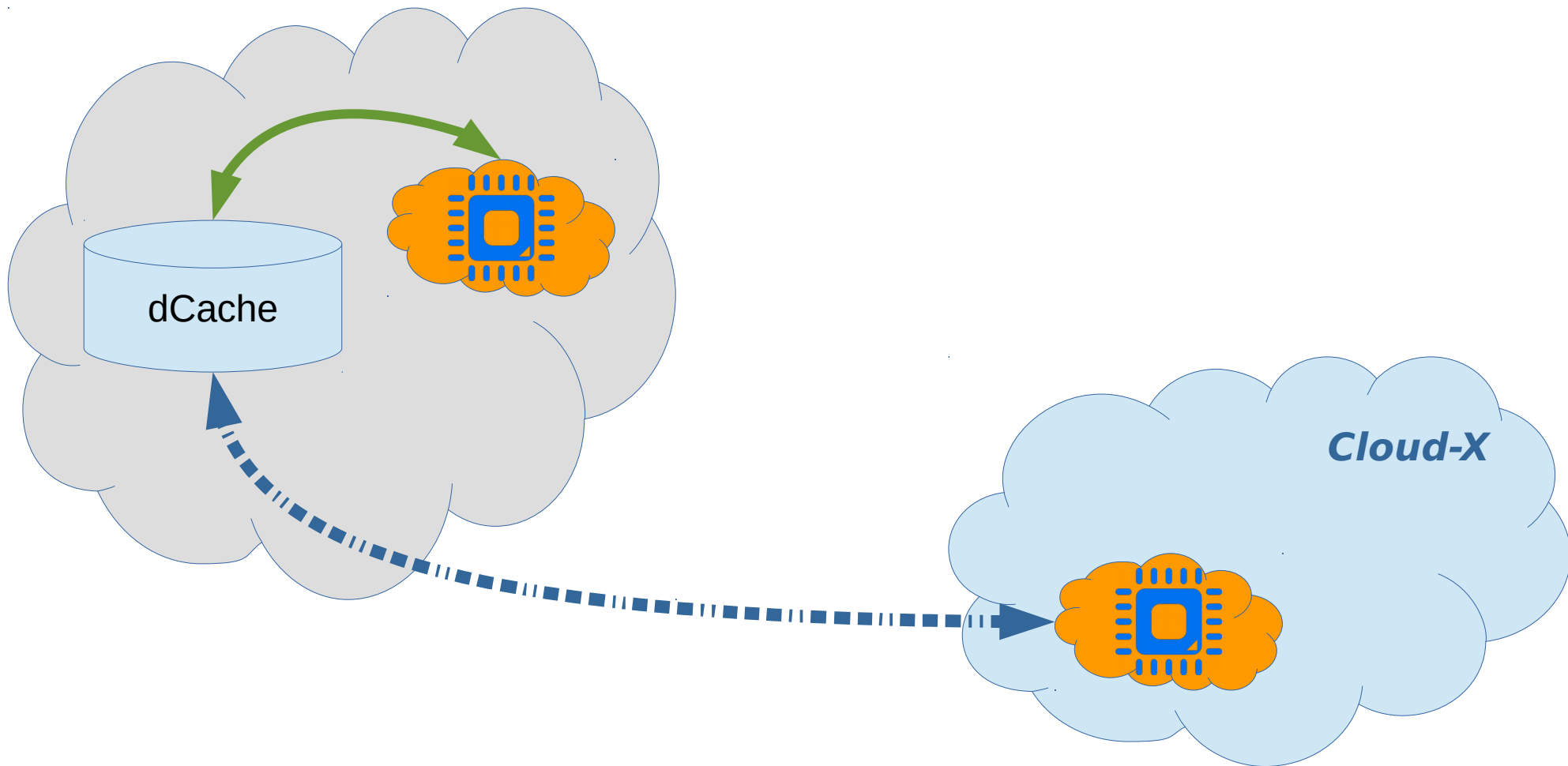
- Protocol
- Path
- Client IP

dCache in the cloud

Icon made by Freepik from www.flaticon.com

dCache in the cloud

dCache

Cloud-X

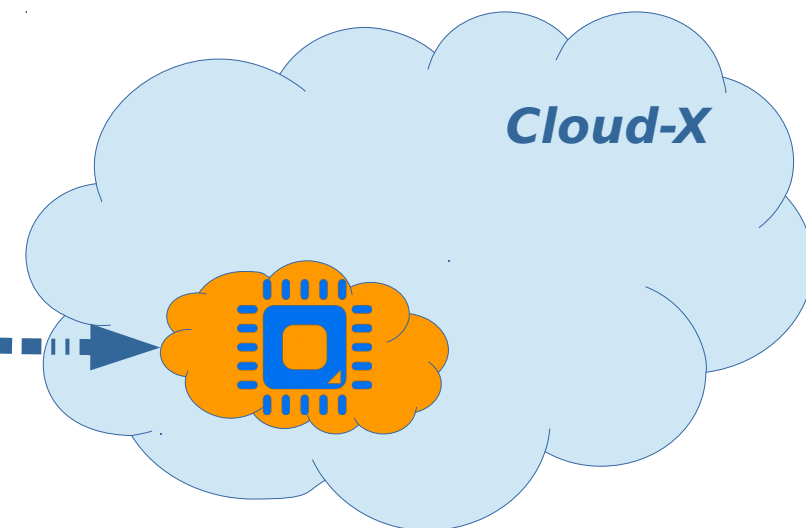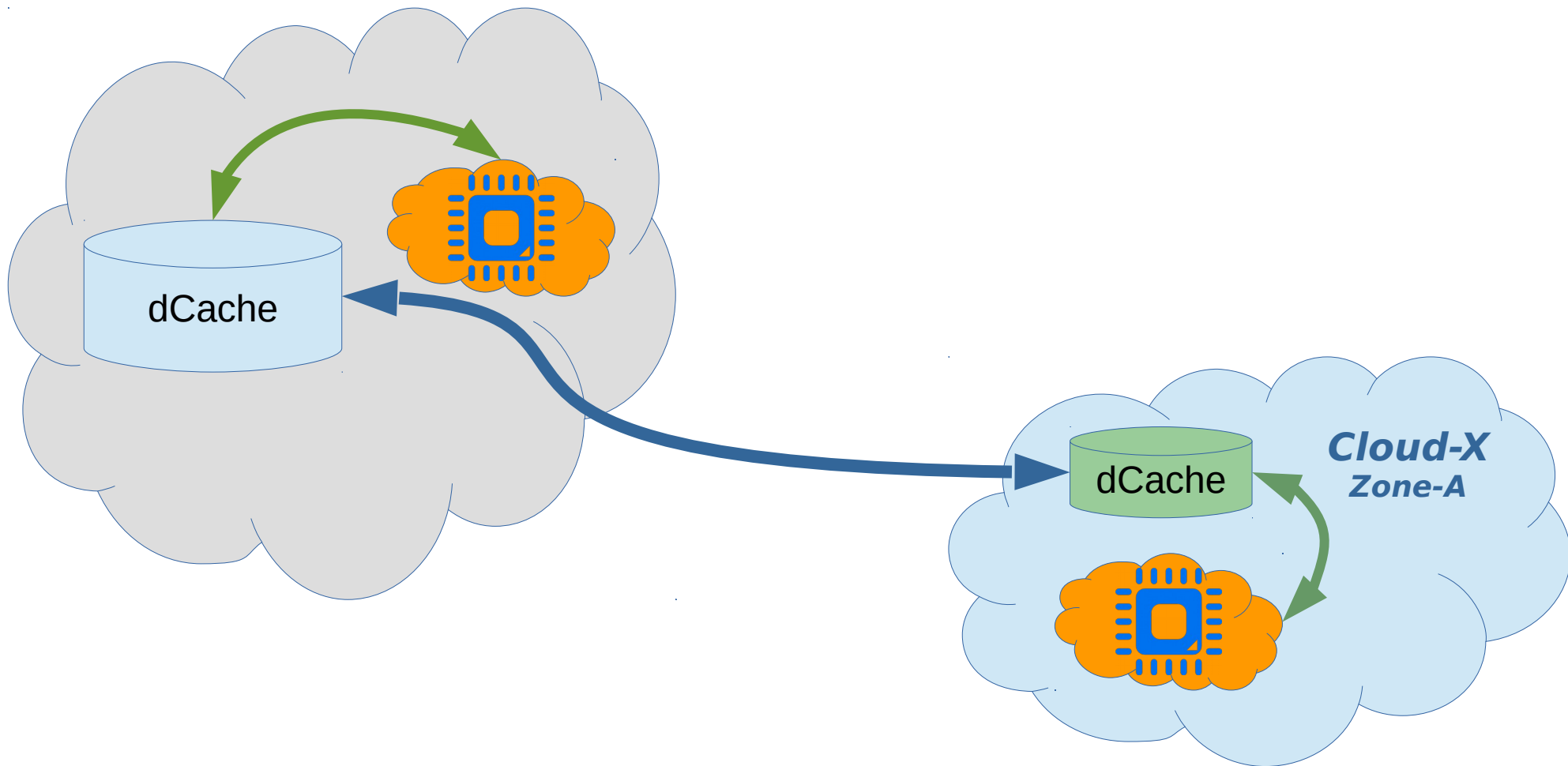Icon made by Freepik from www.flaticon.com
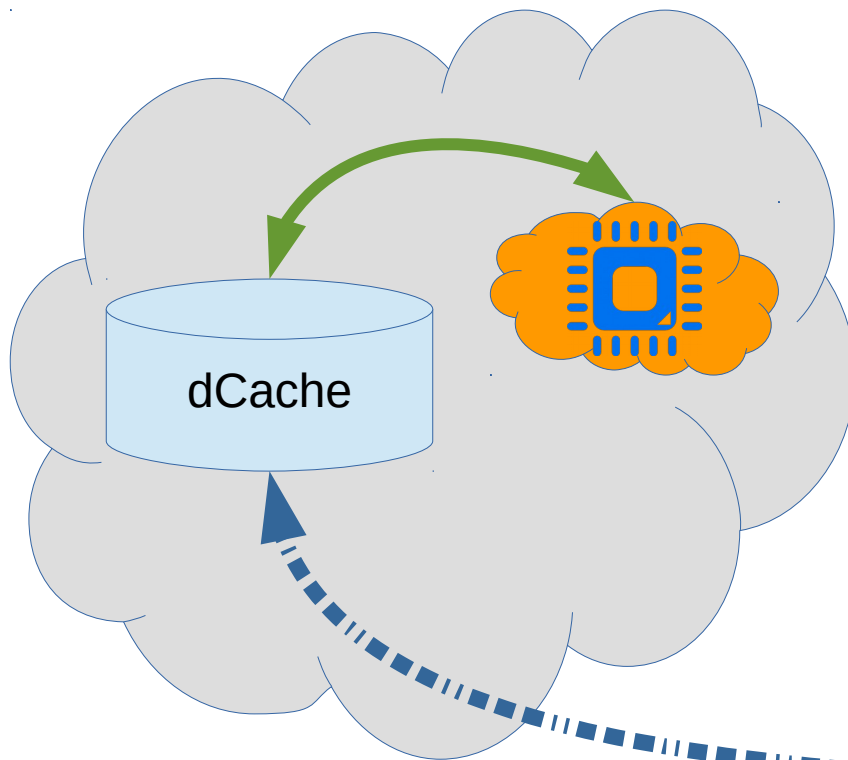
dCache in the cloud

Remote Data access:

- High latency for random access (poor CPU efficiency)
- Hot data transferred multiple times
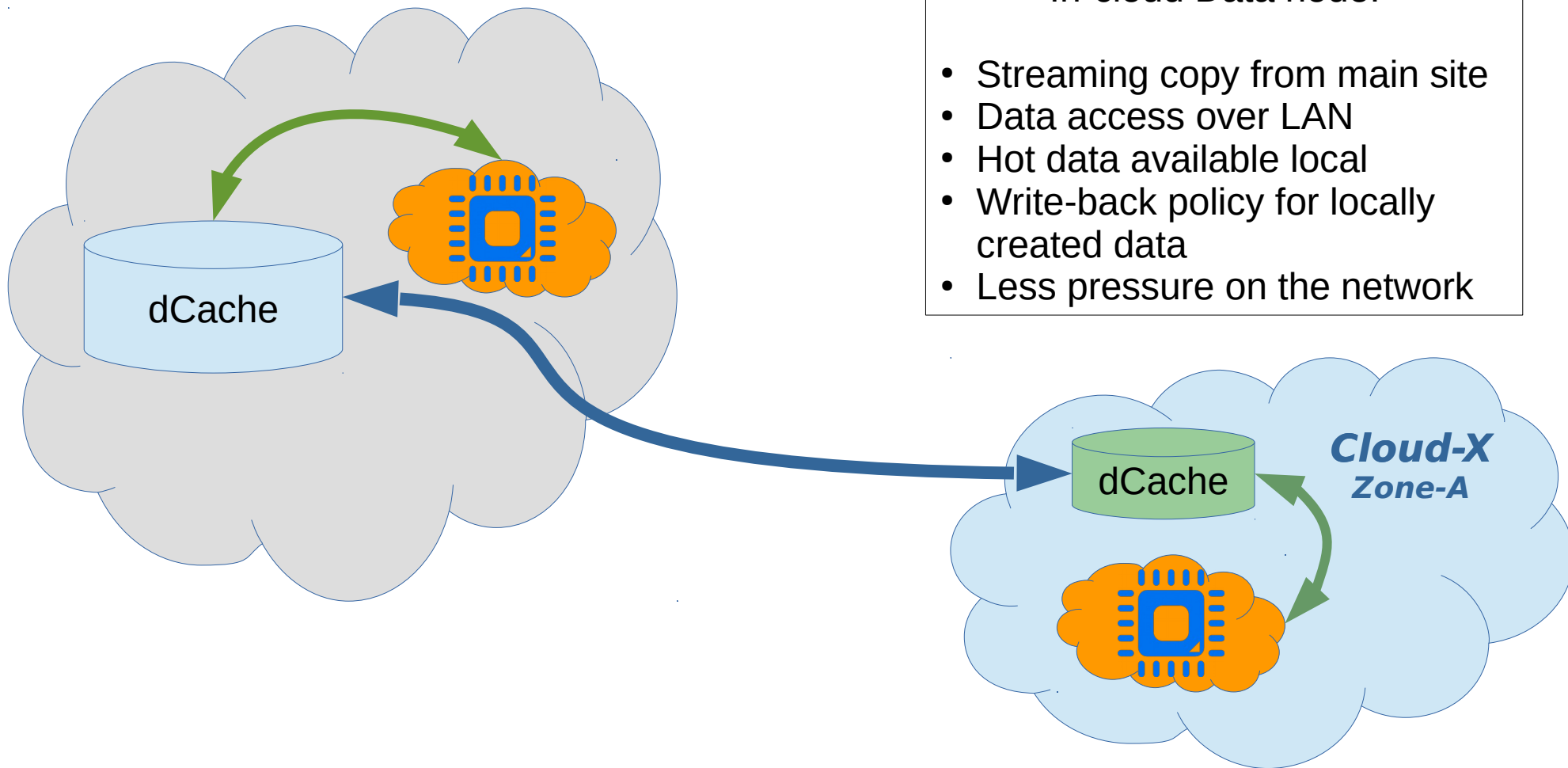- Network bandwidth can be a bottleneck.

dCache

Cloud-X

Icon made by Freepik from www.flaticon.com

dCache in the cloud

Cloud-X
Zone-A

Icon made by Freepik from www.flaticon.com

dCache in the cloud

In-cloud Data node:

- Streaming copy from main site
- Data access over LAN
- Hot data available local
- Write-back policy for locally created data
- Less pressure on the network

dCache

dCache

Cloud-X
Zone-A

Icon made by Freepik from www.flaticon.com

dCache in the cloud

Icon made by Freepik from www.flaticon.com

Summary and Conclusions dCache.org
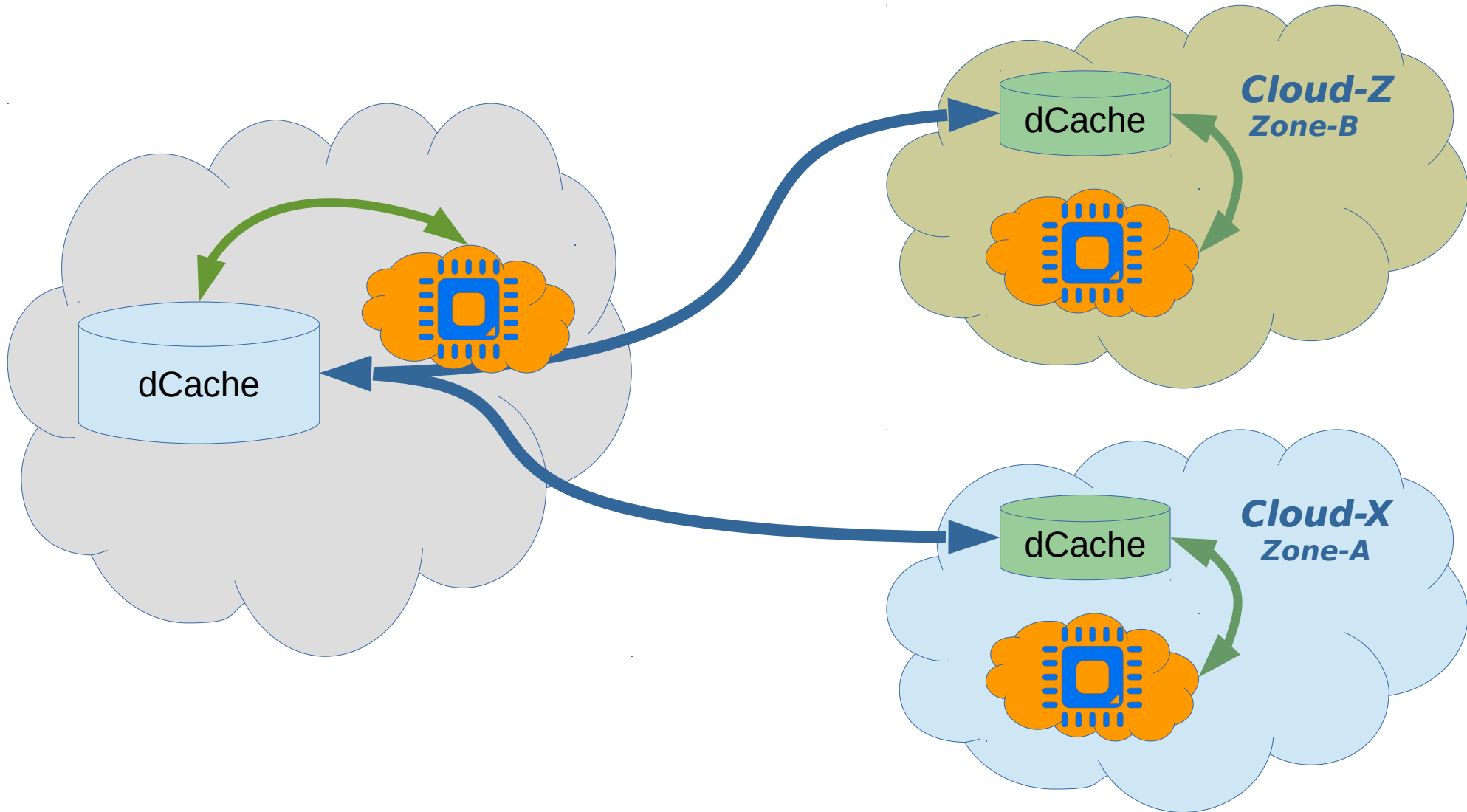
- dCache has a long tradition in providing federated storage for WLCG

- The system flexibility allows to control data placement and replication

- dCache's smart caching and dynamic pool configuration optimize data access in hybrid cloud deployments.

- **We help experiments to manage data, by taking away burden of managing storage!**

Thank You!

13'th dCache user workshop:
May 21-22, Madrid

More info: https://www.dcache.org