dCache.org

dCache, agile adoption of storage technology

Paul Millar

CHEP-2012 New York, 2012-05-24

Fermilab NDGF Nordic DataGrid Facility European Middleware Initiative DESY Helmholtz Association

Overview

- Some news
- Flexibility
- Future directions
- Summary

Funding

- dCache is our **contribution to WLCG**:

  from Germany, the Nordic countries and USA/Fermilab,

- has been funded (independently from WLCG) for **over 10 years**

- Funding for dCache **is secure** for after EMI:

  Without EMI, funding only drops by ~20–25%

Community

- 3rd **International workshop**:
  - 57 participants, from 13 countries
  - New user-communities presented how they wish to use dCache

- Forging **links with industry**:

  DESY and IBM form "large data" strategic partnership based on dCache storage competence (CeBIT)

- Establishing a **Stack Exchange site**

  `http://area51.stackexchange.com/proposals/40050/dcache`

Evolution

- Within **WLCG**:
  - Strong involvement with **TEG groups**
  - Working in collaboration on federated storage

    (both xrootd and HTTP)
- Outside WLCG:
  - OGF standardisation
  - Engaging new communities
- Improve dCache **modularity**:

  Allow dCache to be easily adapted to novel environments

  Agility is a process, not a target

News: under the hood

- Splitting the code into smaller, **reusable** pieces:
  - **Chimera**: enstore

    See ***Enstore with Chimera namespace provider*** by **D. Litvintsev**
  - **jrpc**: BACNET, a Swiss Bank, ...

    See ***dCache: Implementing a high-end NFSv4.1 server using a Java NIO framework*** by **T. Mkrtchyan.**
  - **xrootd4j**: (ALICE?)
- dCache is adopting Free/Open-source license
  - Mostly **AGPLv3**, the rest is LGPL or BSD
  - Needed to get dCache into distributions

AGPL v3
Free Software
Free as in Freedom

News: NFSv4.1 / pNFS

- **Industry standard protocol**
- Client availability:
  - RHEL/**SL 6.x**,
  - RHEL/**SL 5.x** (with Oracle kernel + `nfs-utils` upgrade),
  - **Fedora 15**,
  - **Debian 7.0** ("Wheezy"),
  - **Windows 7** (with driver from CITI),
  - **Windows 8**,
  - Solaris "Oracle (..) will deliver implementations of (a client and server) in future releases of Solaris" (1)
- **Hardware vendor** support:
  - **NetApp OnTap 8.1**
  - Panasas "in 2012" (1)
  - BlueArc,
  - IBM "key part of SONAS Active Cloud Engine" (1)

(1) Source is "FAST 2012 pNFS BoF" 2012-02-15

News: dCache & pNFS

- NFS v4.1 / pNFS has been supported since 2009.
- Deployed **in production** (at DESY) for over a year.
- Fermilab's REX dept. evaluated dCache NFSv4.1 for their Intensity Frontier experiments:

  "**Results look promising**, throughput scales well with number of pool nodes"
- Supports:
  - authn: trusted-host and Kerberos
  - all three GSS security modes.

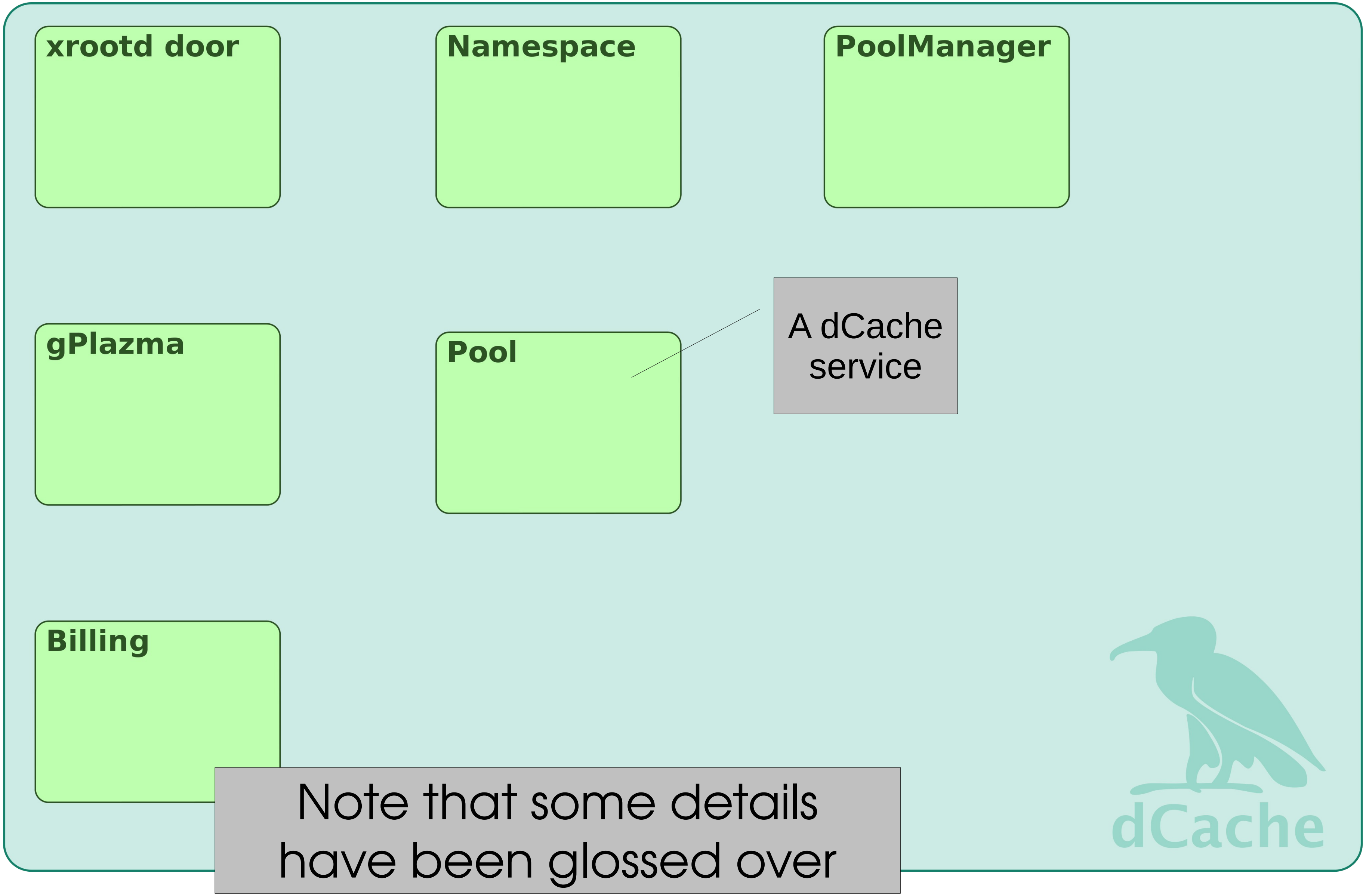
Flexibility

(plugins and extension points)

Plugins: who should be interested & why

- **Core developers**:
  - New functionality can be added as a plugin
  - Backwards compatibility by keeping old plugins
  - Can test-deploy new features at friendly sites
- **dCache sites**:
  - integrating with **local, site-specific services**
- **User-communities**:
  - Add some **experiment-specific behaviour**
- **External developers** / **trail-blazer sites**:
  - Experiment with **exciting new features**

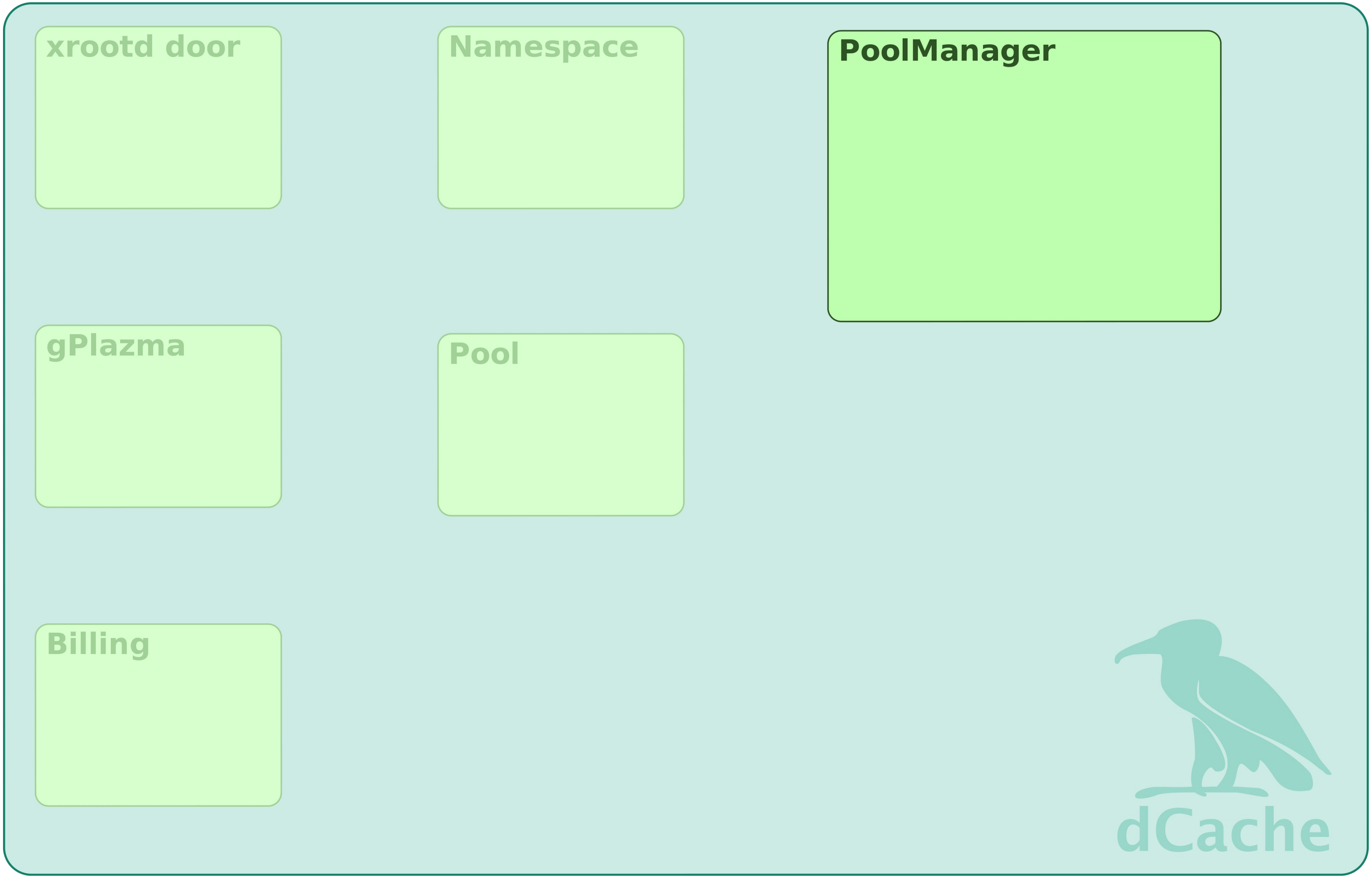
What can I enhance?



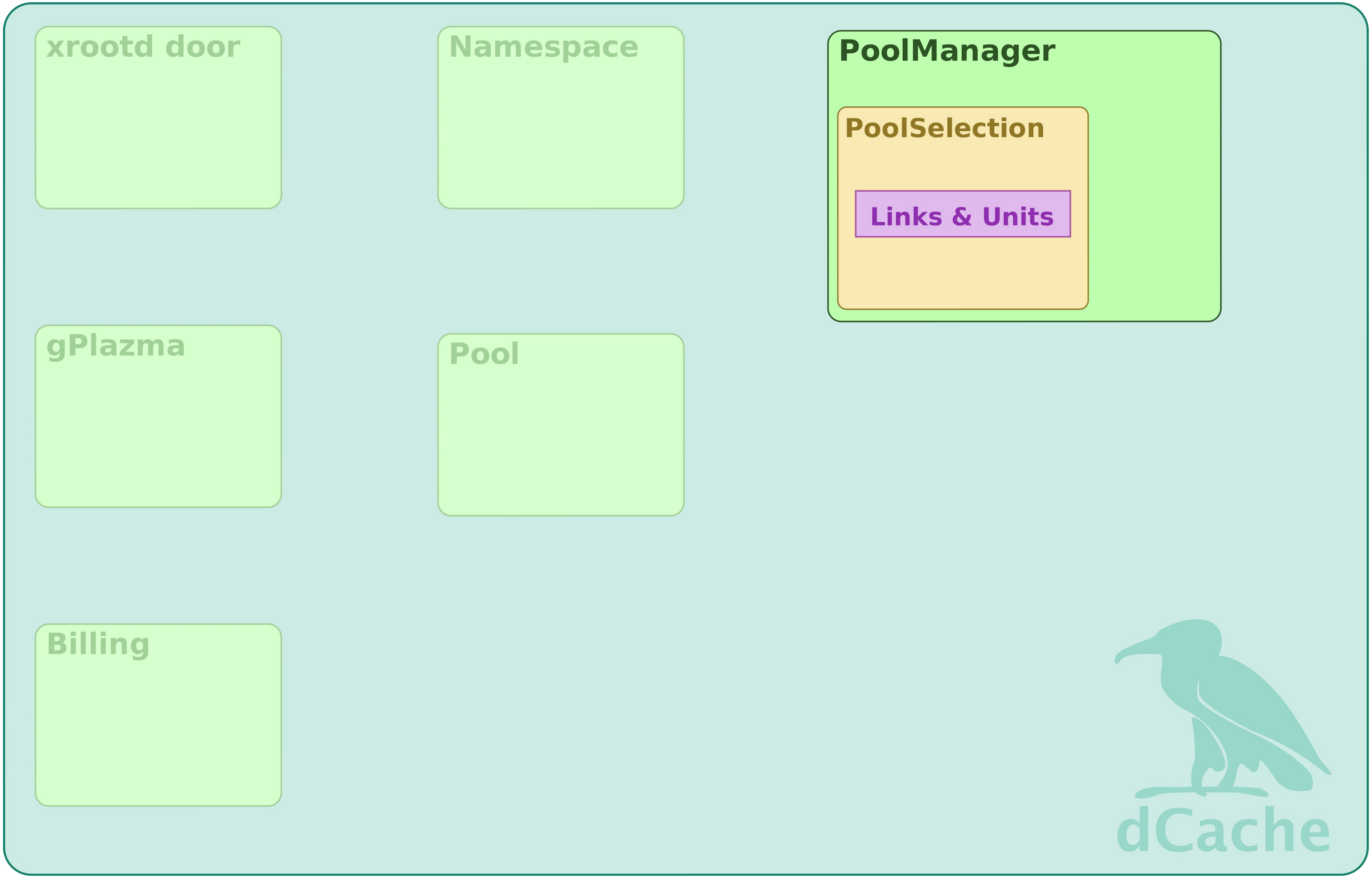
A dCache service

Note that some details have been glossed over

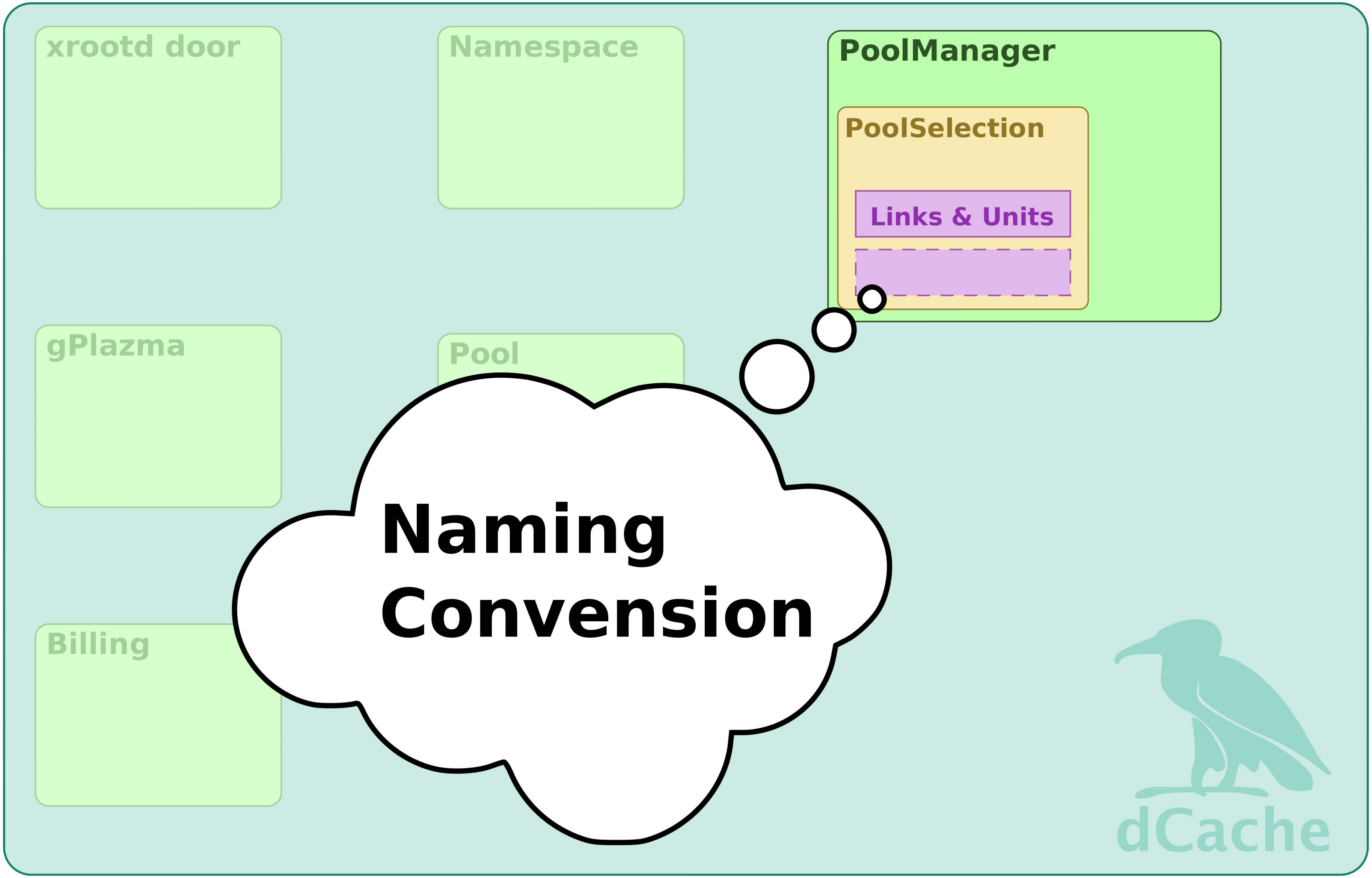
What can I enhance?



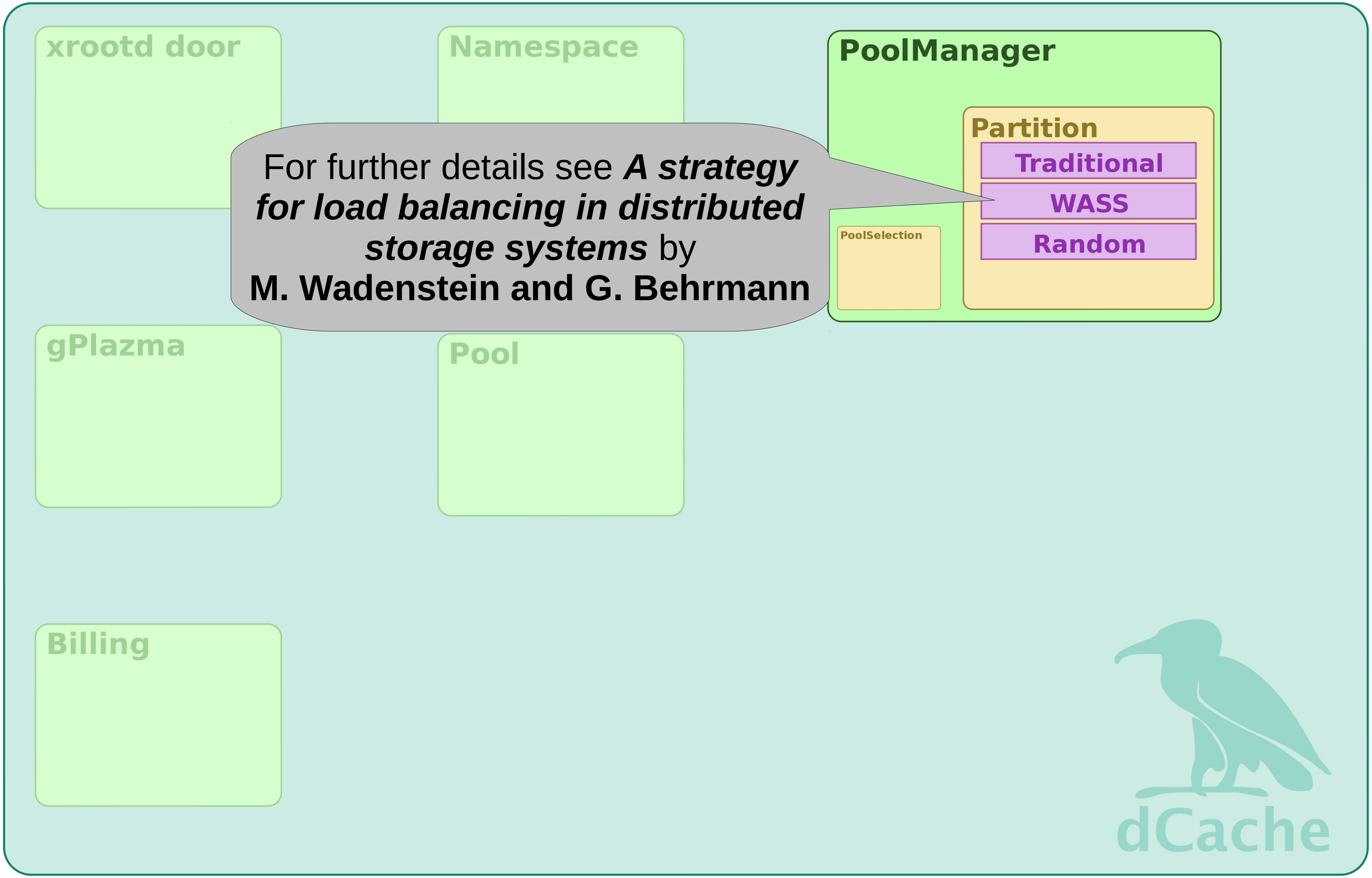
What can I enhance?



What can I enhance?

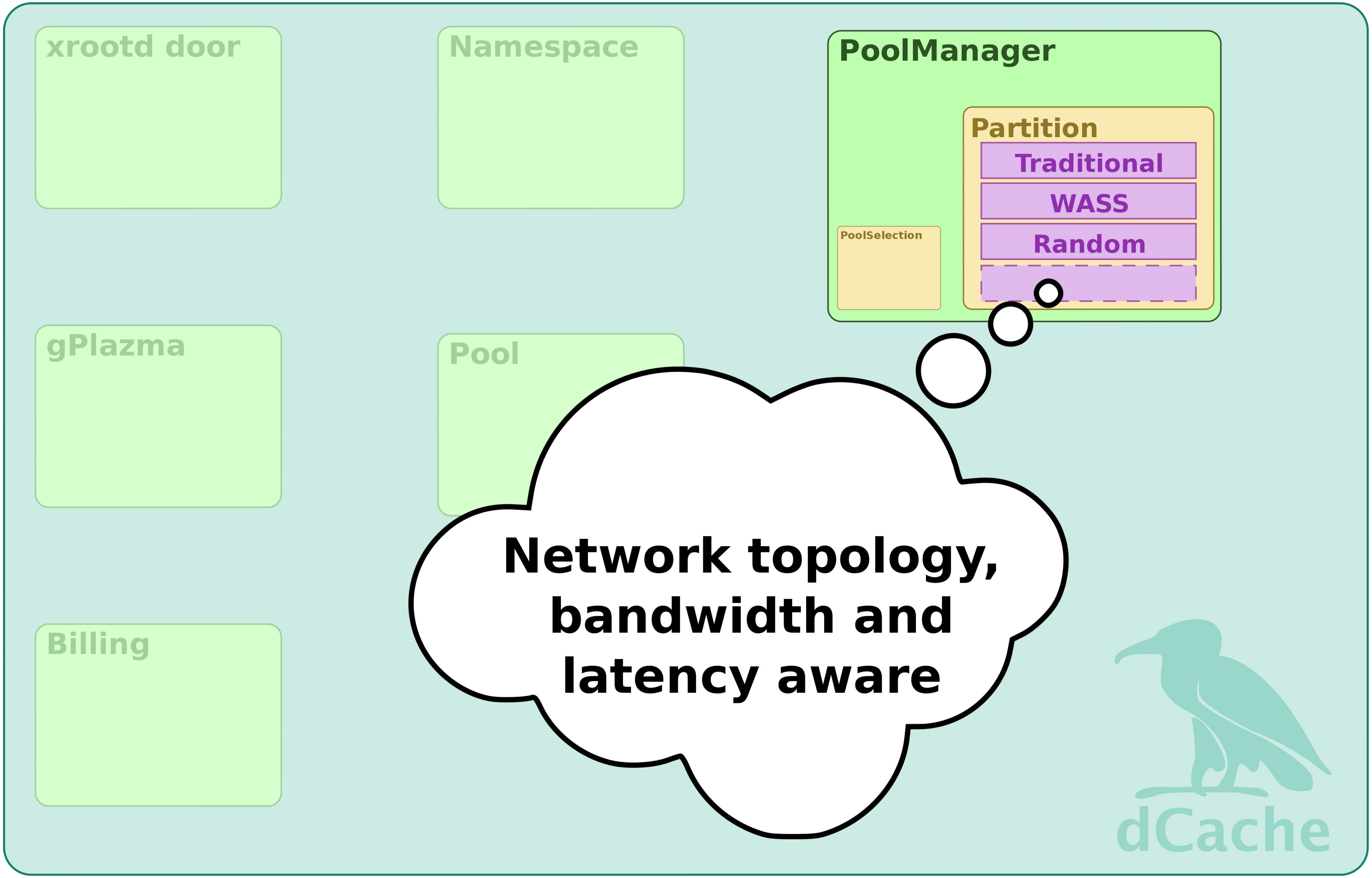


What can I enhance?



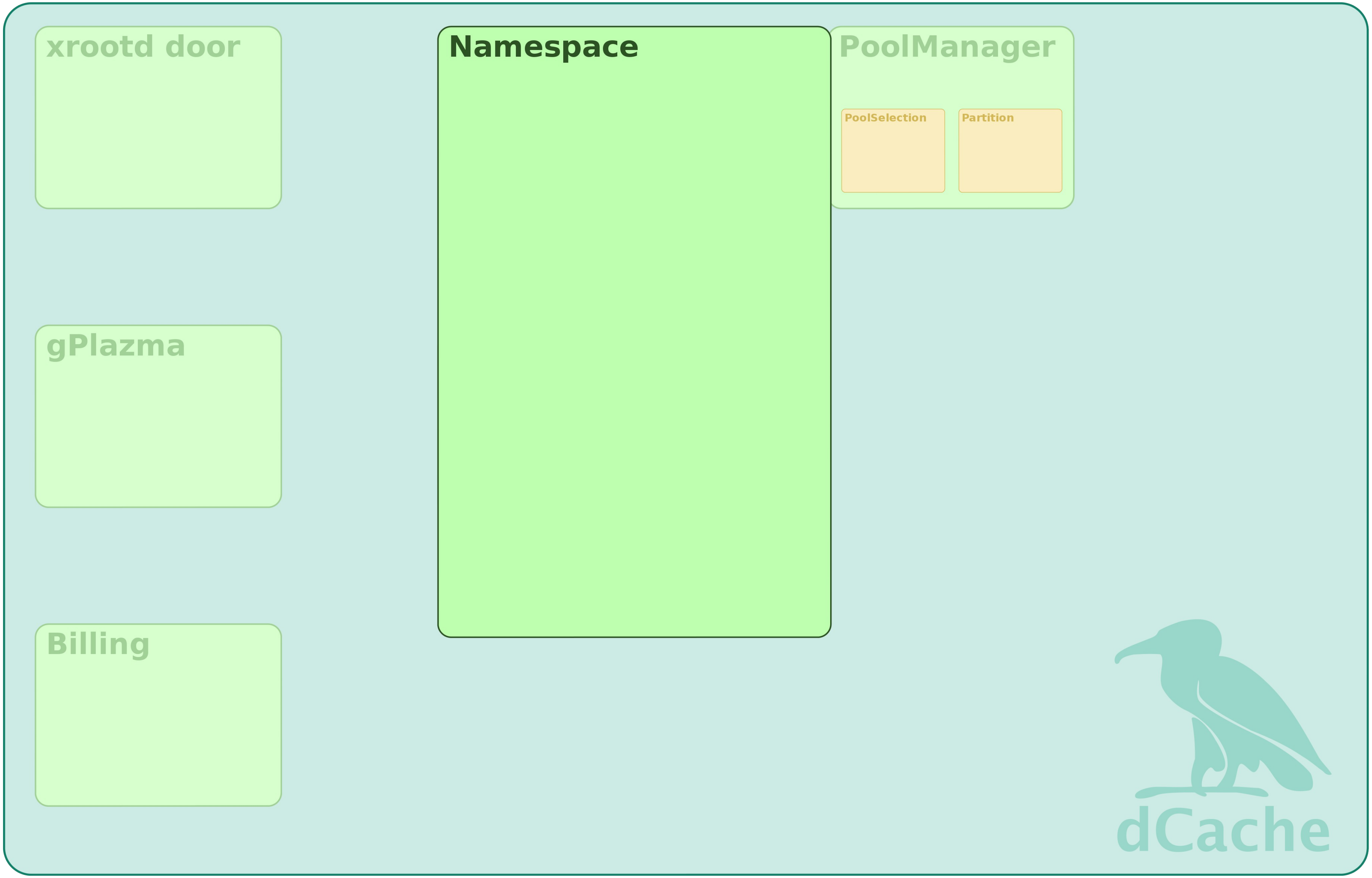
For further details see ***A strategy for load balancing in distributed storage systems*** by **M. Wadenstein and G. Behrmann**

What can I enhance?

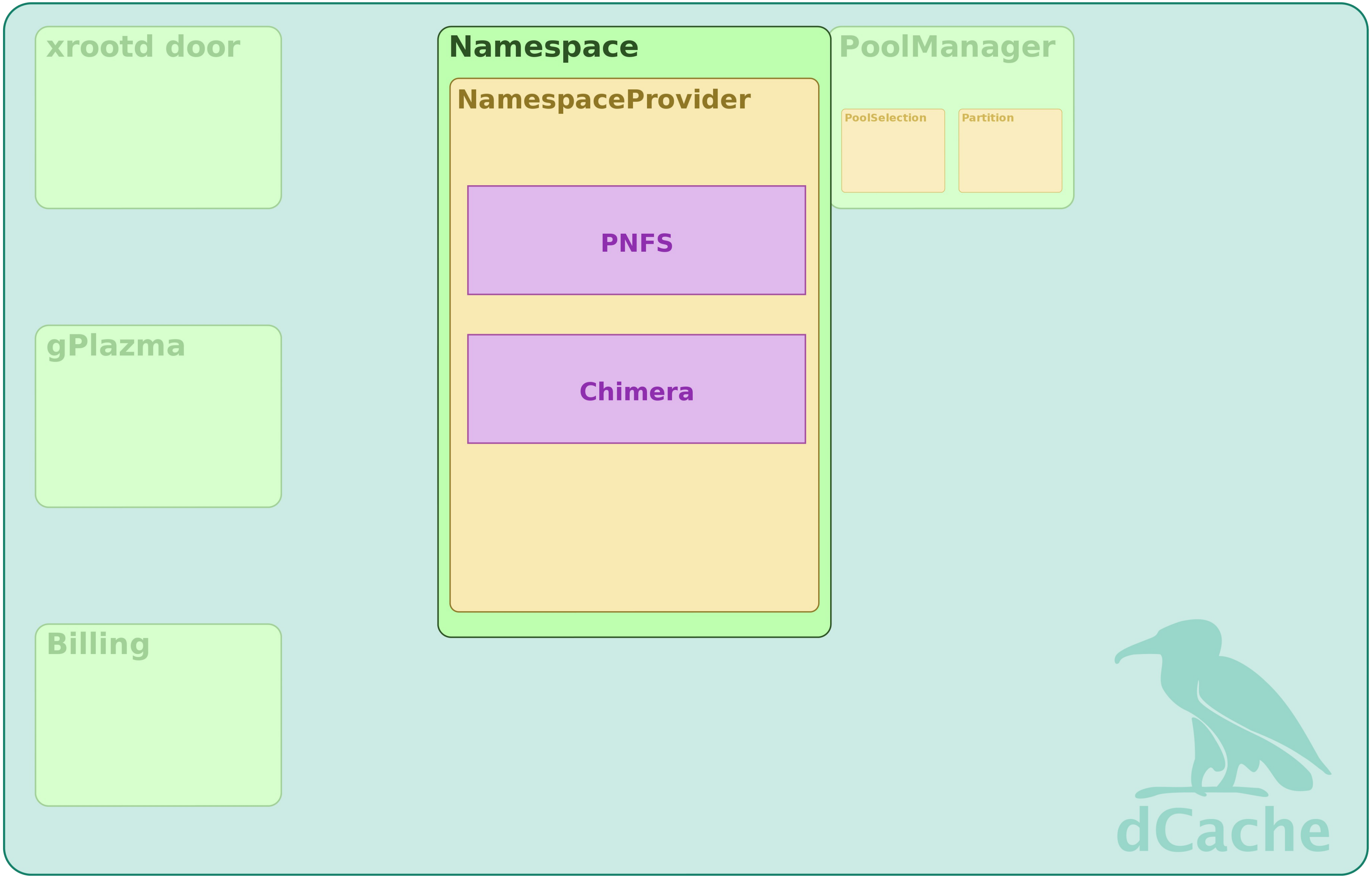


Network topology, bandwidth and latency aware

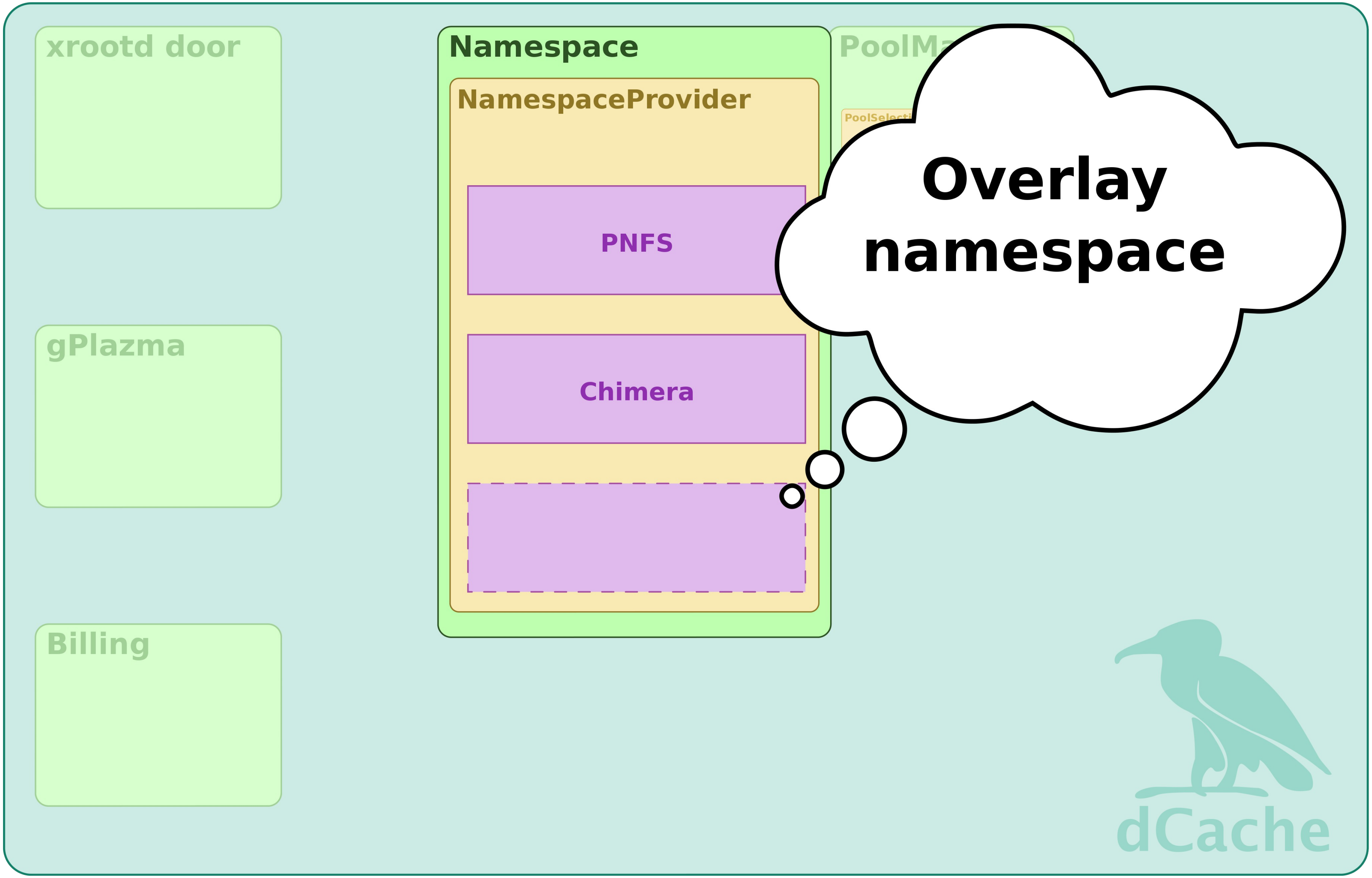
What can I enhance?



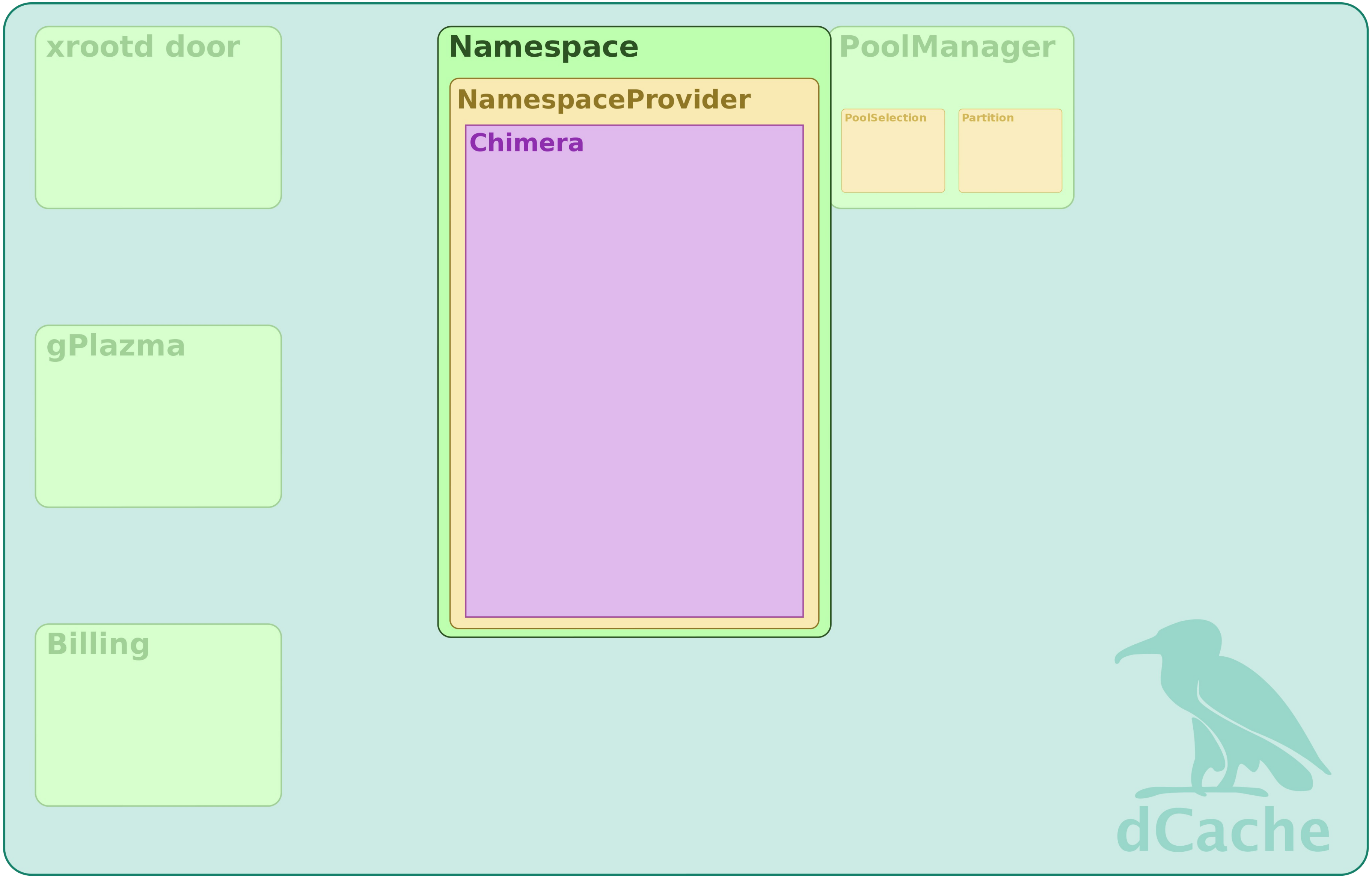
What can I enhance?



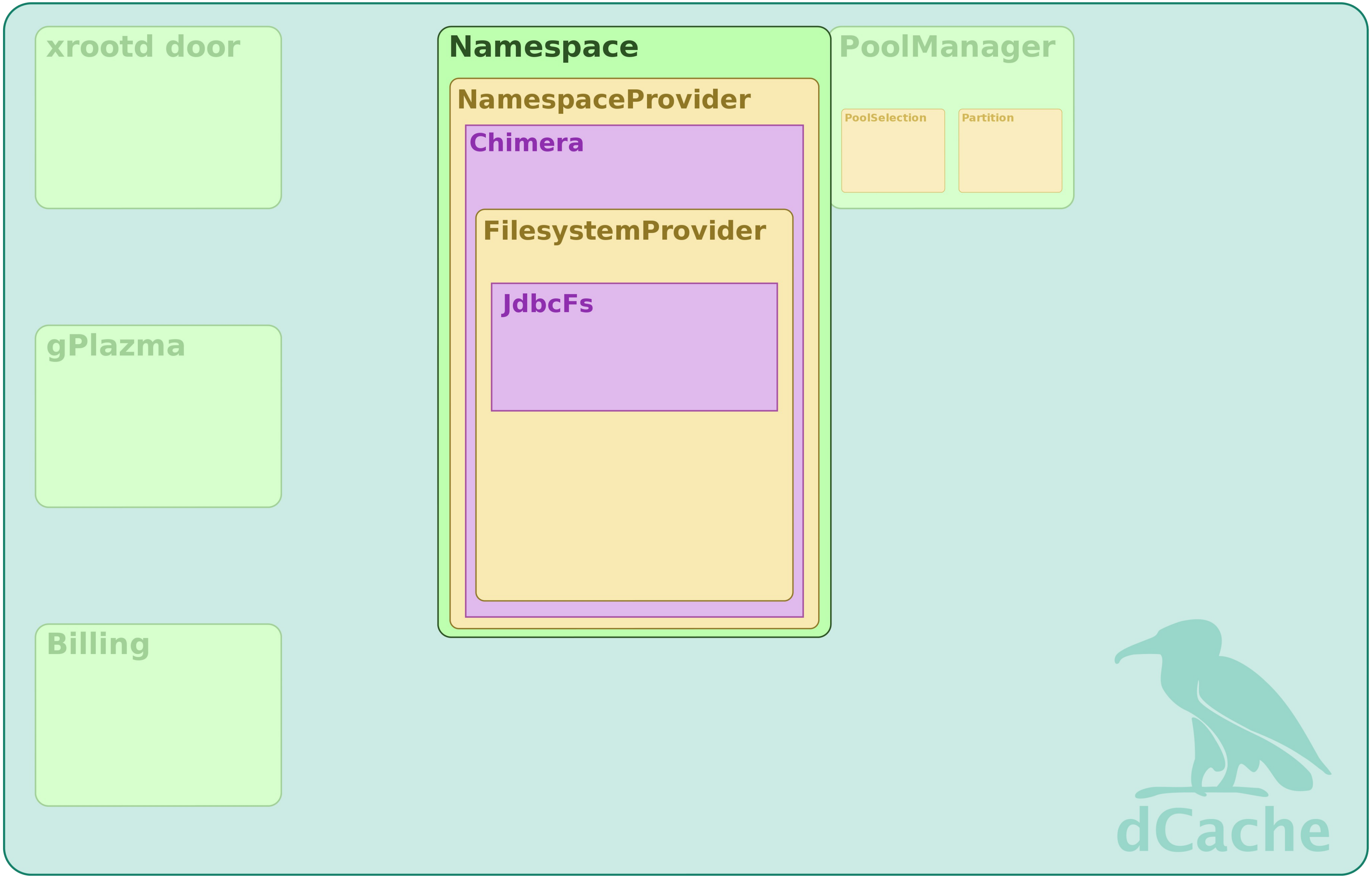
What can I enhance?



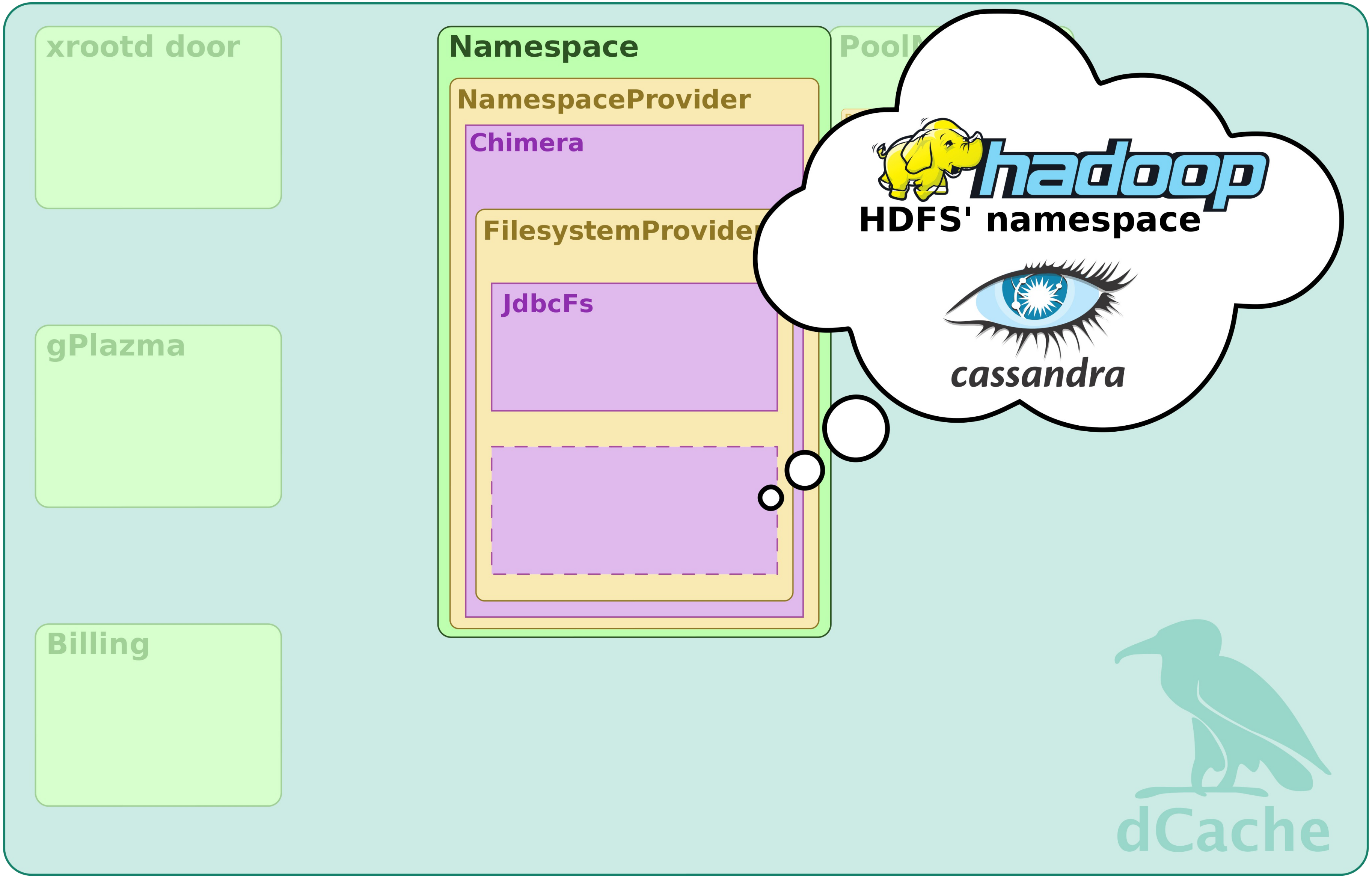
What can I enhance?



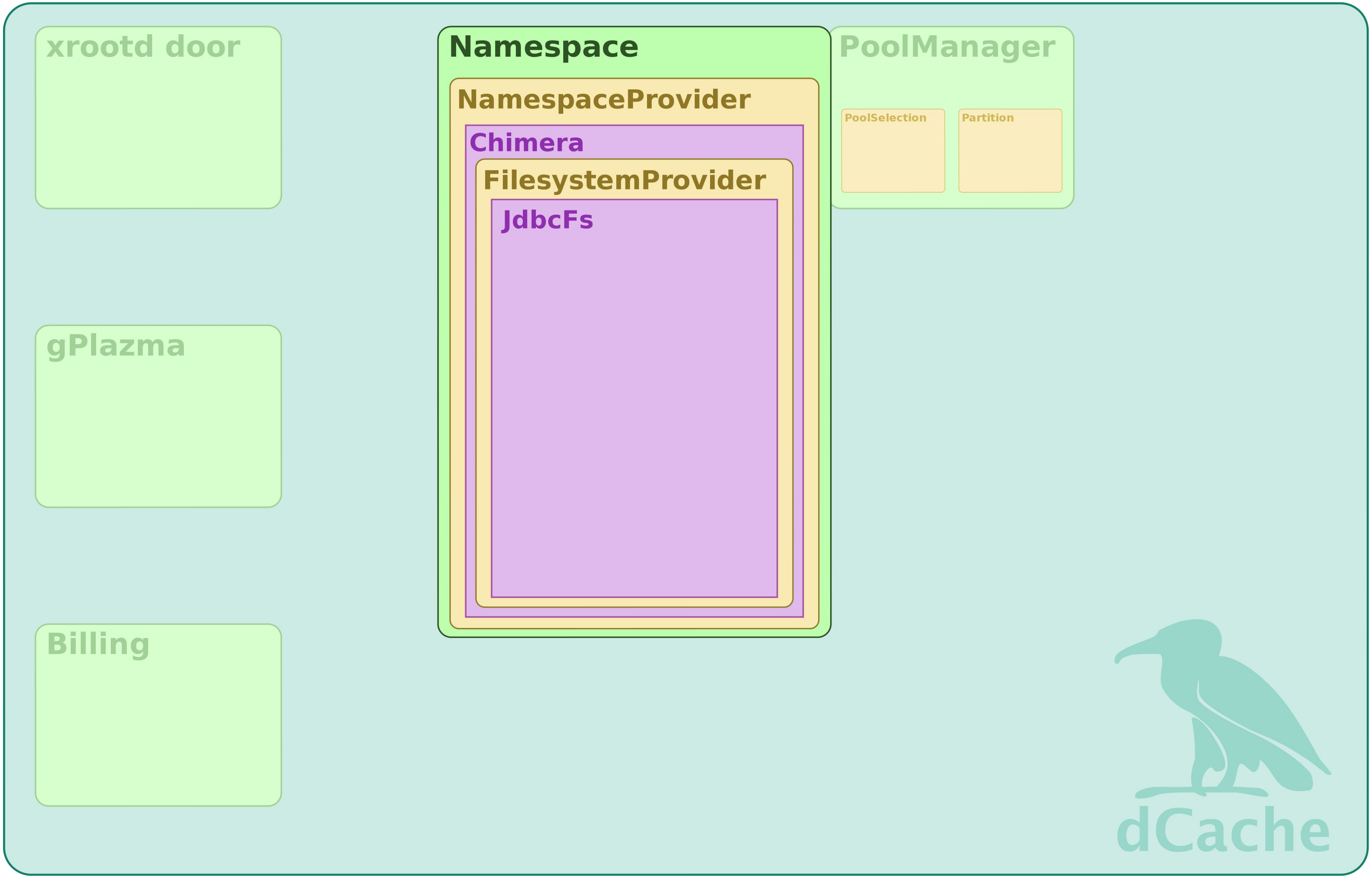
What can I enhance?



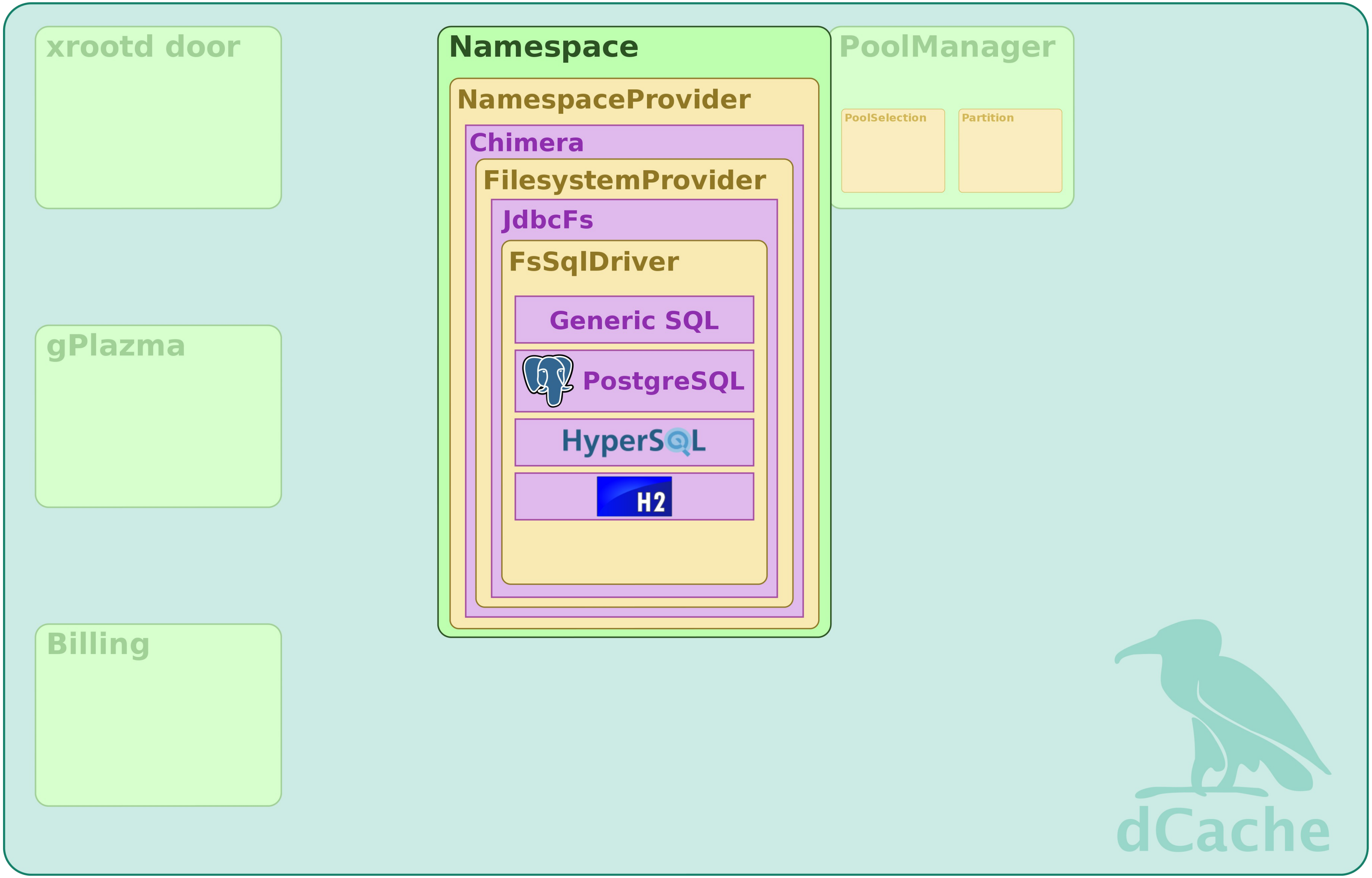
What can I enhance?



What can I enhance?



What can I enhance?



xrootd door

gPlazma

Billing

Namespace

NamespaceProvider

Chimera

FilesystemProvider

JdbcFs

FsSqlDriver

Generic SQL

PostgreSQL

HyperSQL

H2

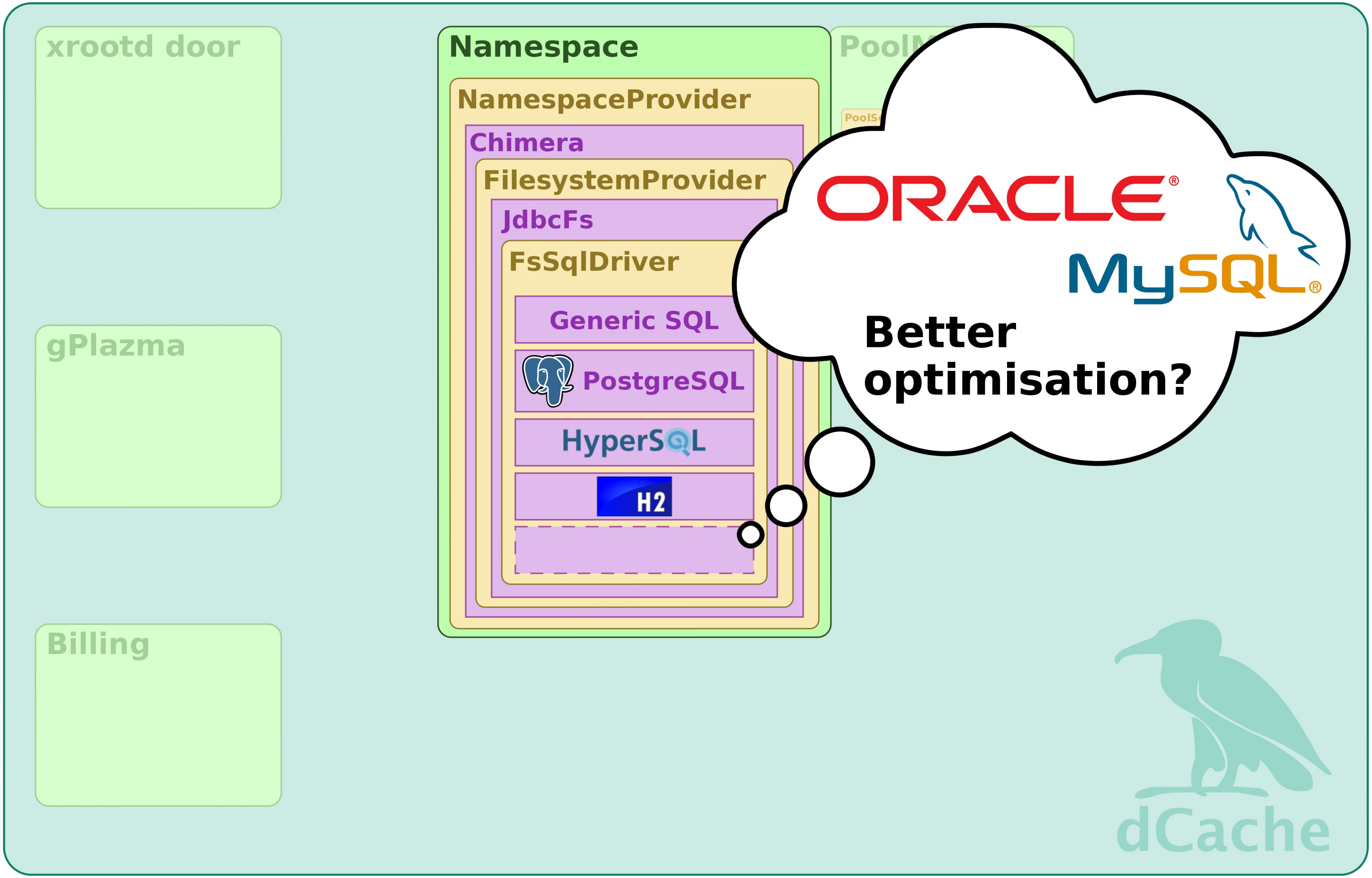
PoolManager

PoolSelection

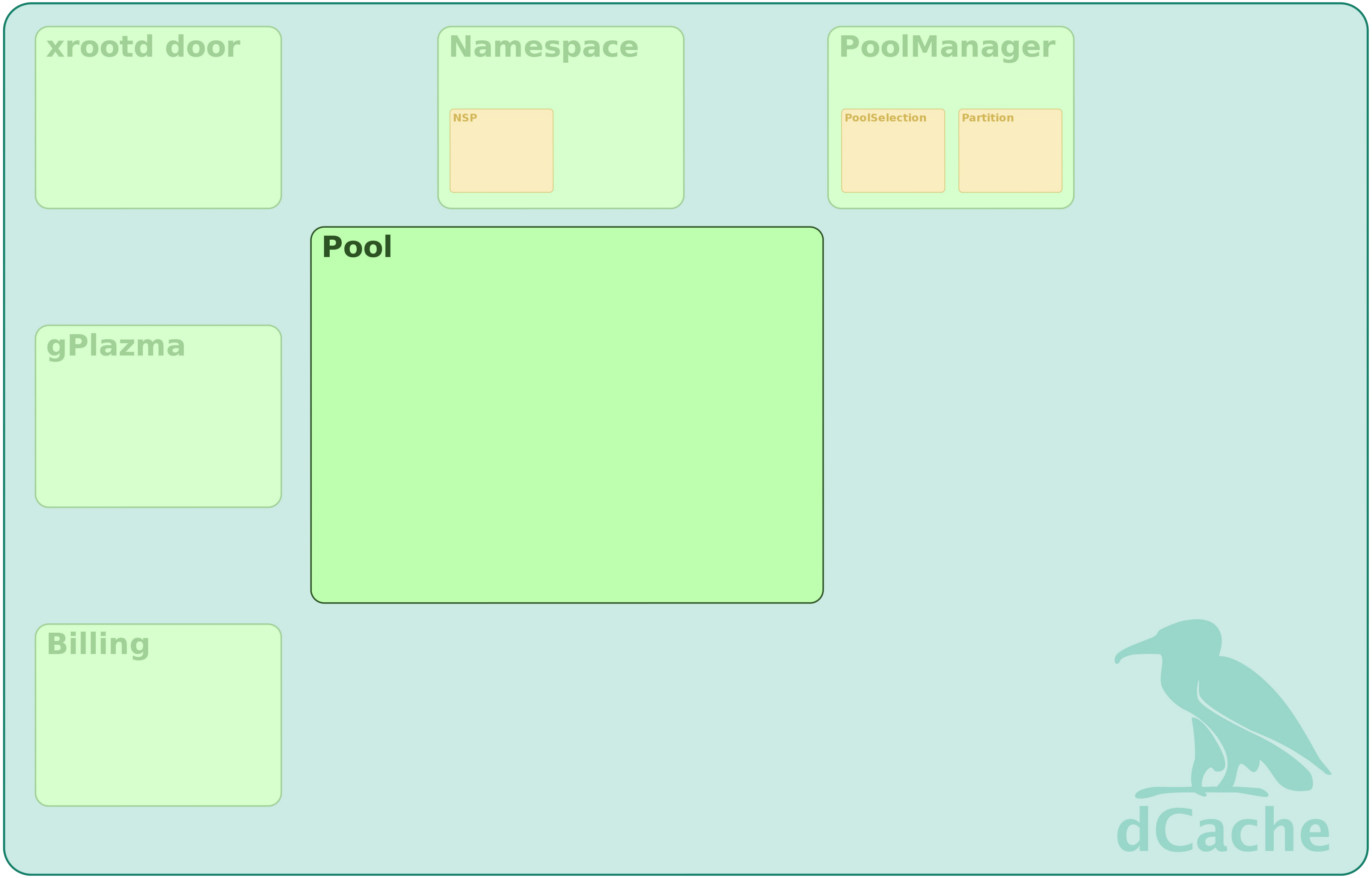
Partition

dCache

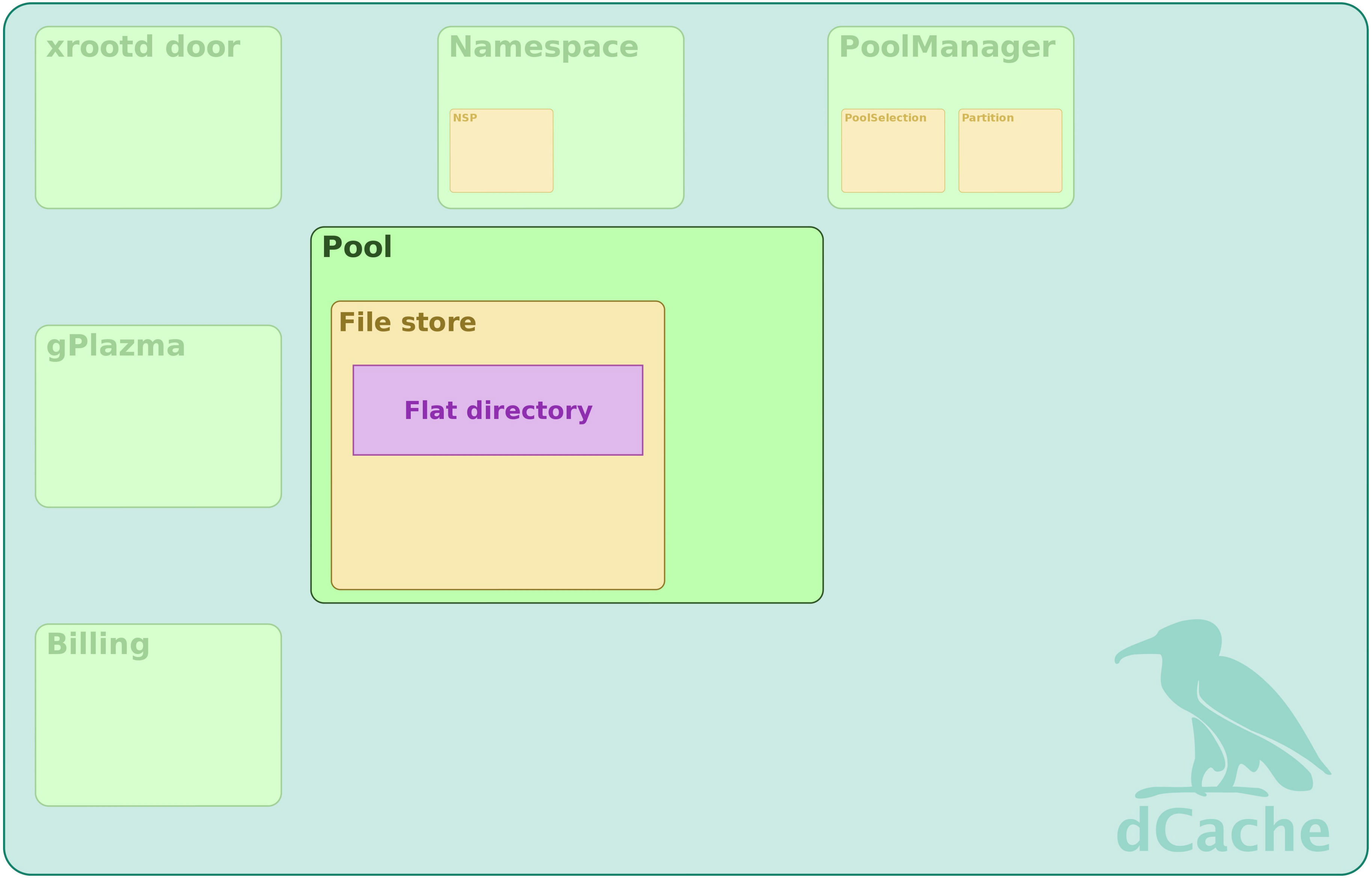
What can I enhance?



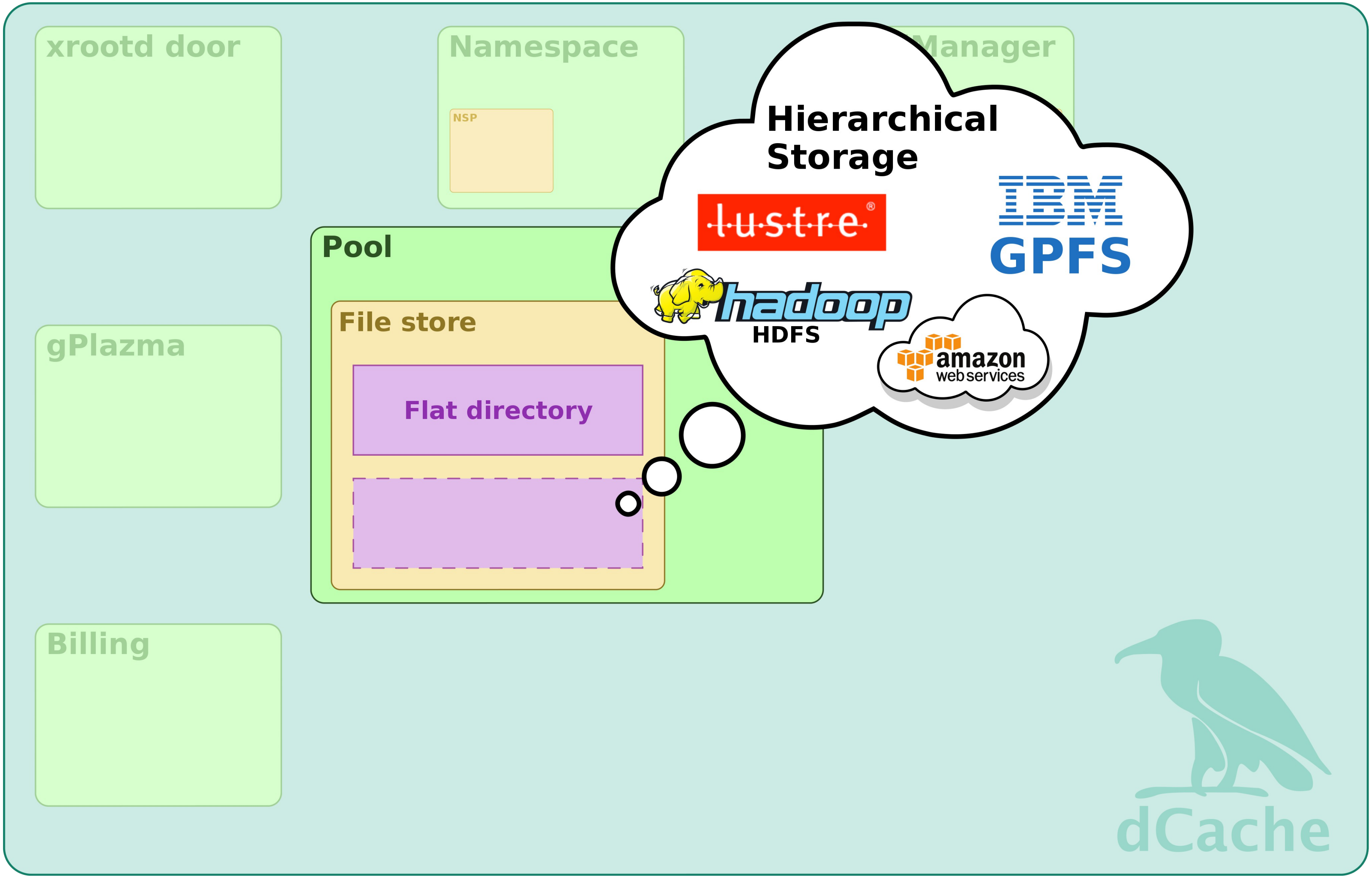
What can I enhance?



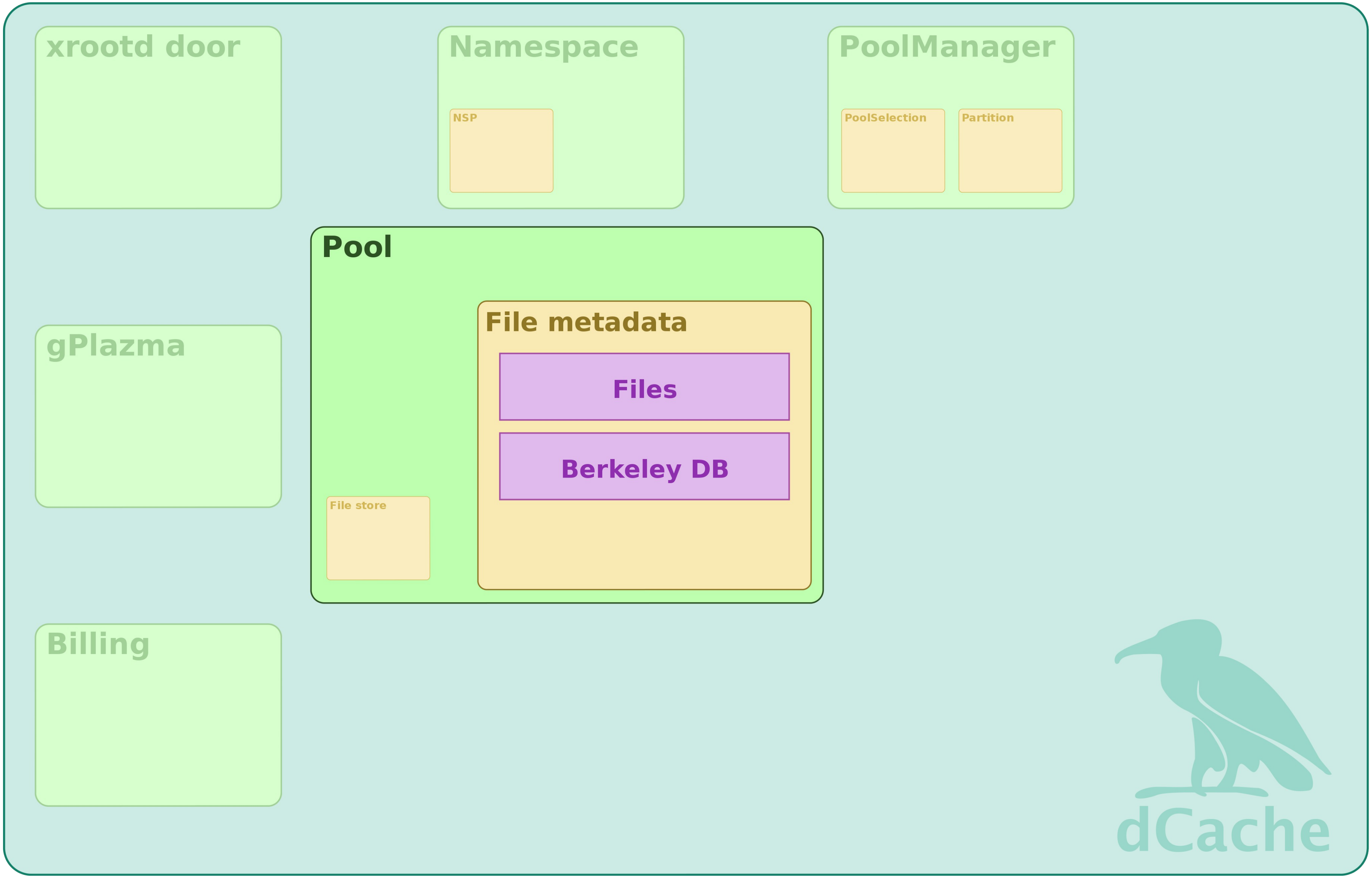
What can I enhance?



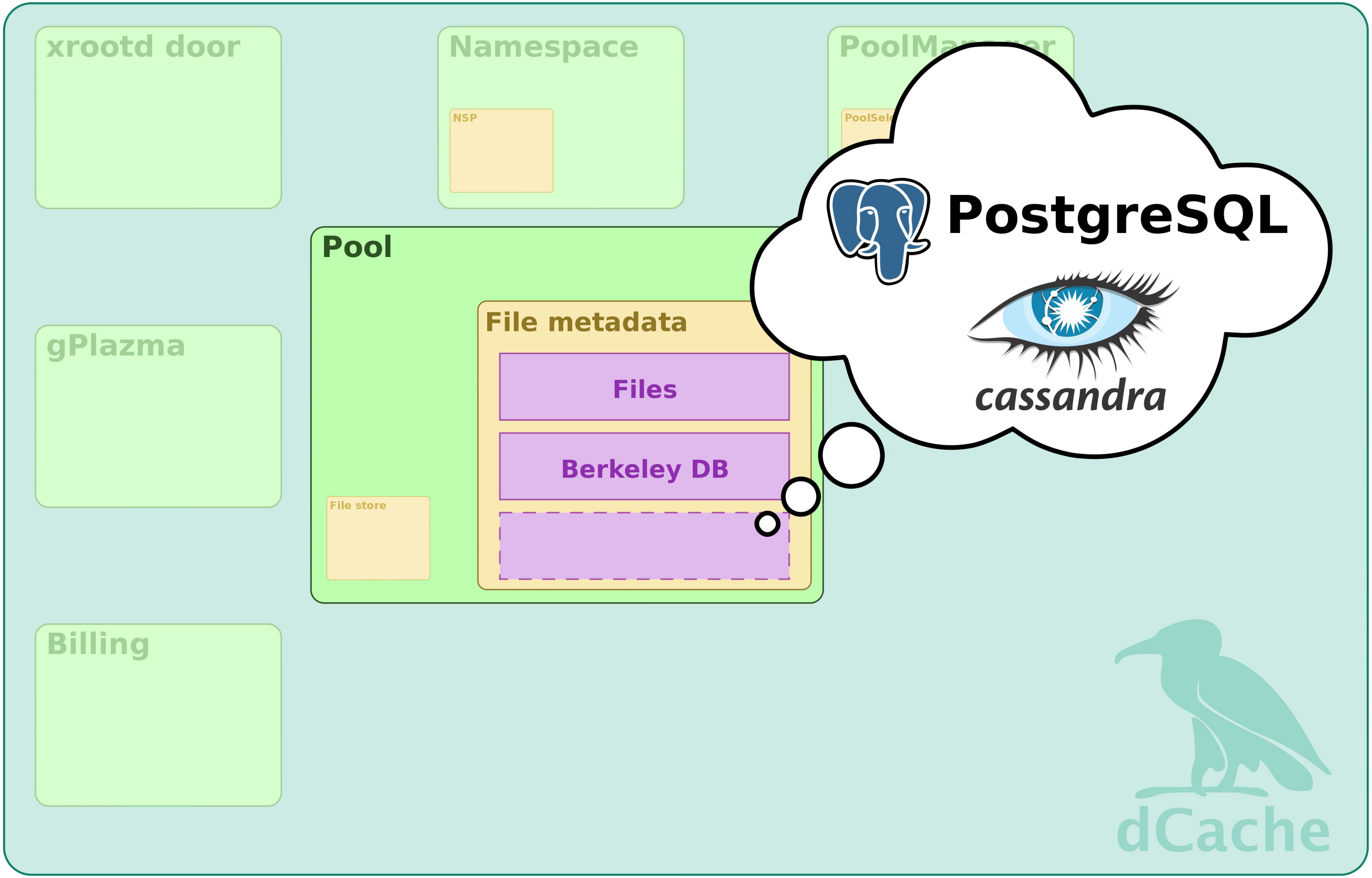
What can I enhance?



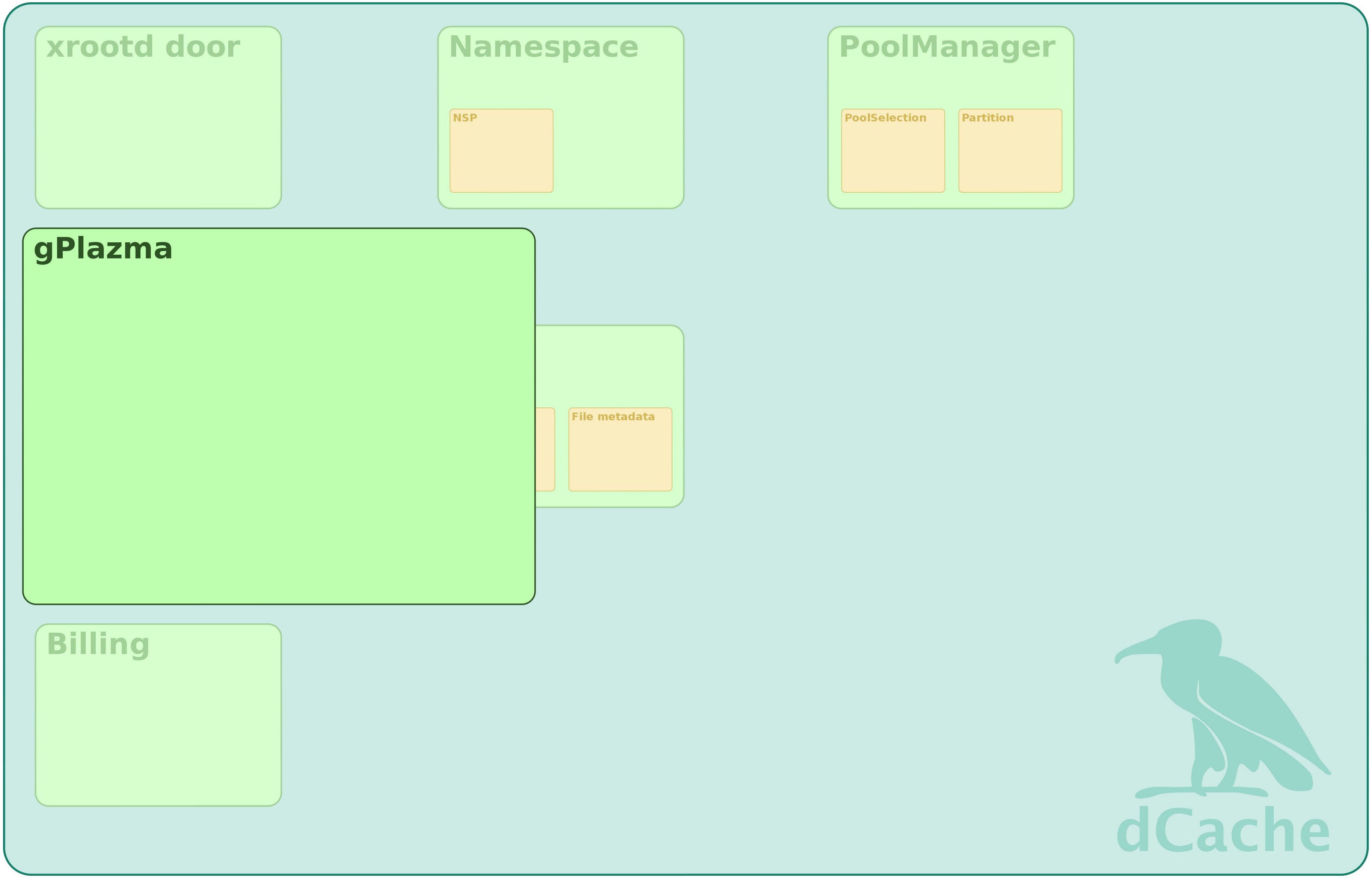
What can I enhance?



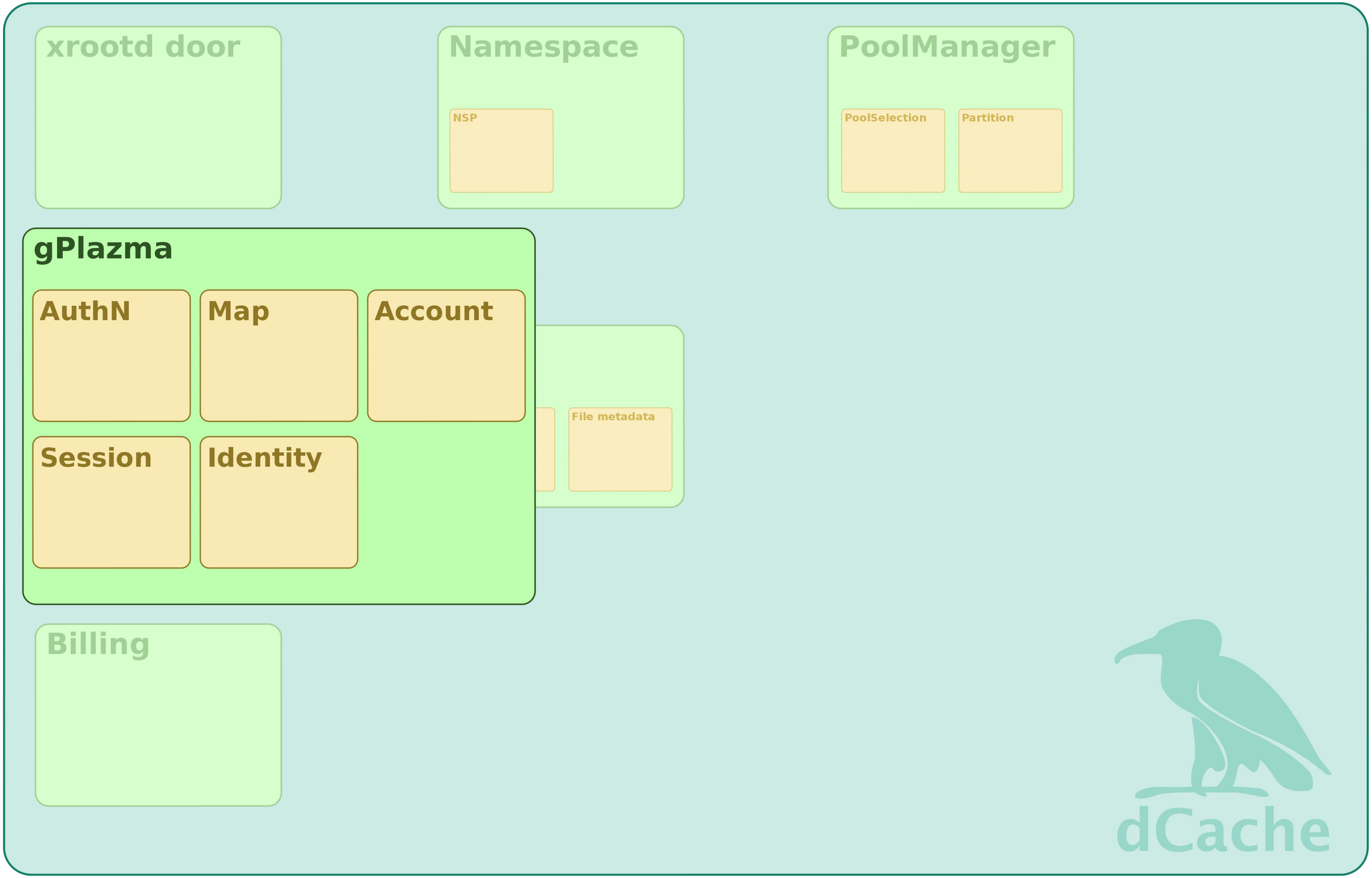
What can I enhance?



What can I enhance?



What can I enhance?



xrootd door

Namespace

NSP

PoolManager

PoolSelection

Partition

gPlazma

AuthN

Map

Account

Session

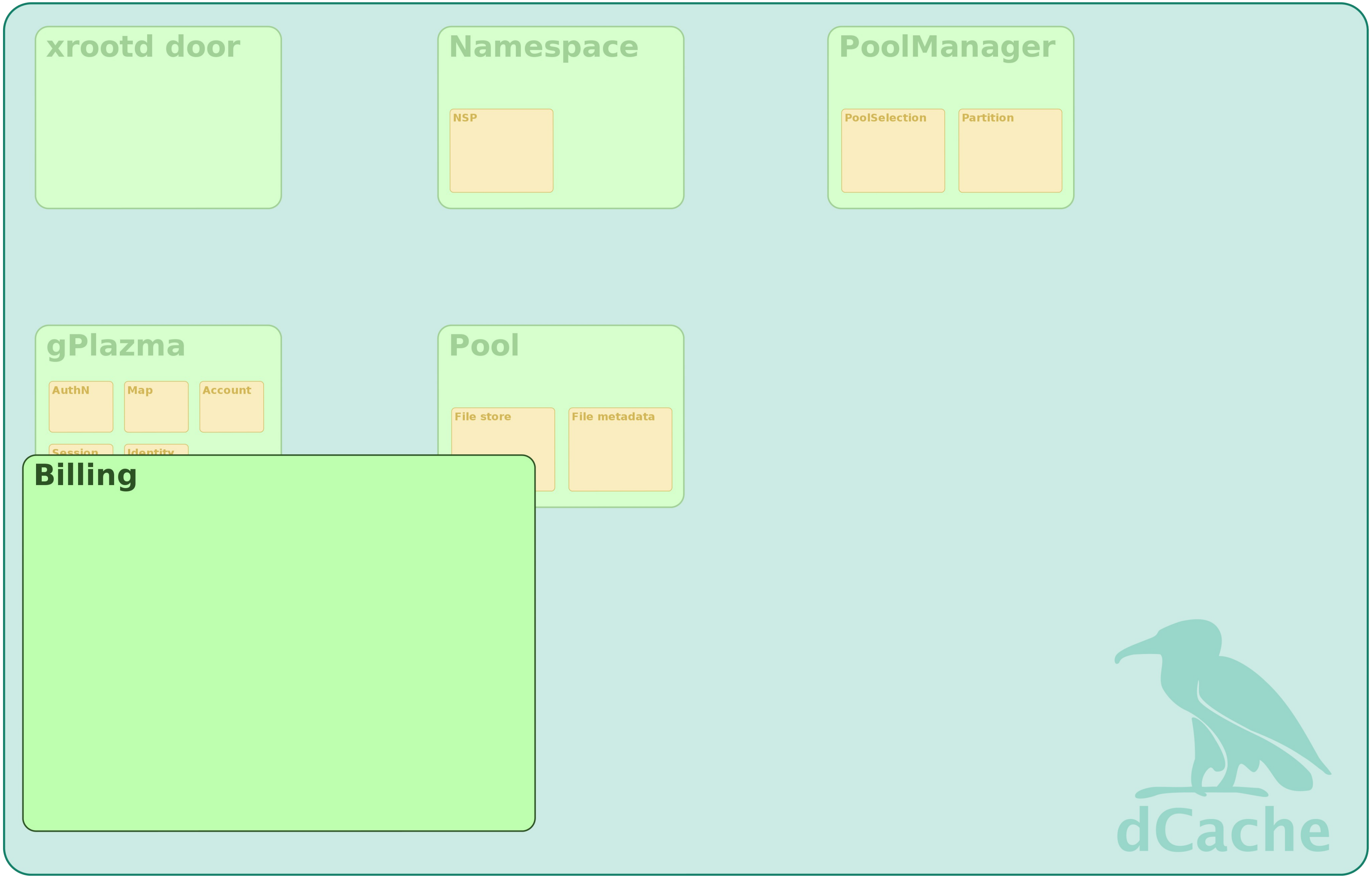
Identity

File metadata

Billing

dCache

What can I enhance?



xrootd door

Namespace

NSP

PoolManager

PoolSelection

Partition

gPlazma

AuthN

Map

Account

Session

Identity

Pool

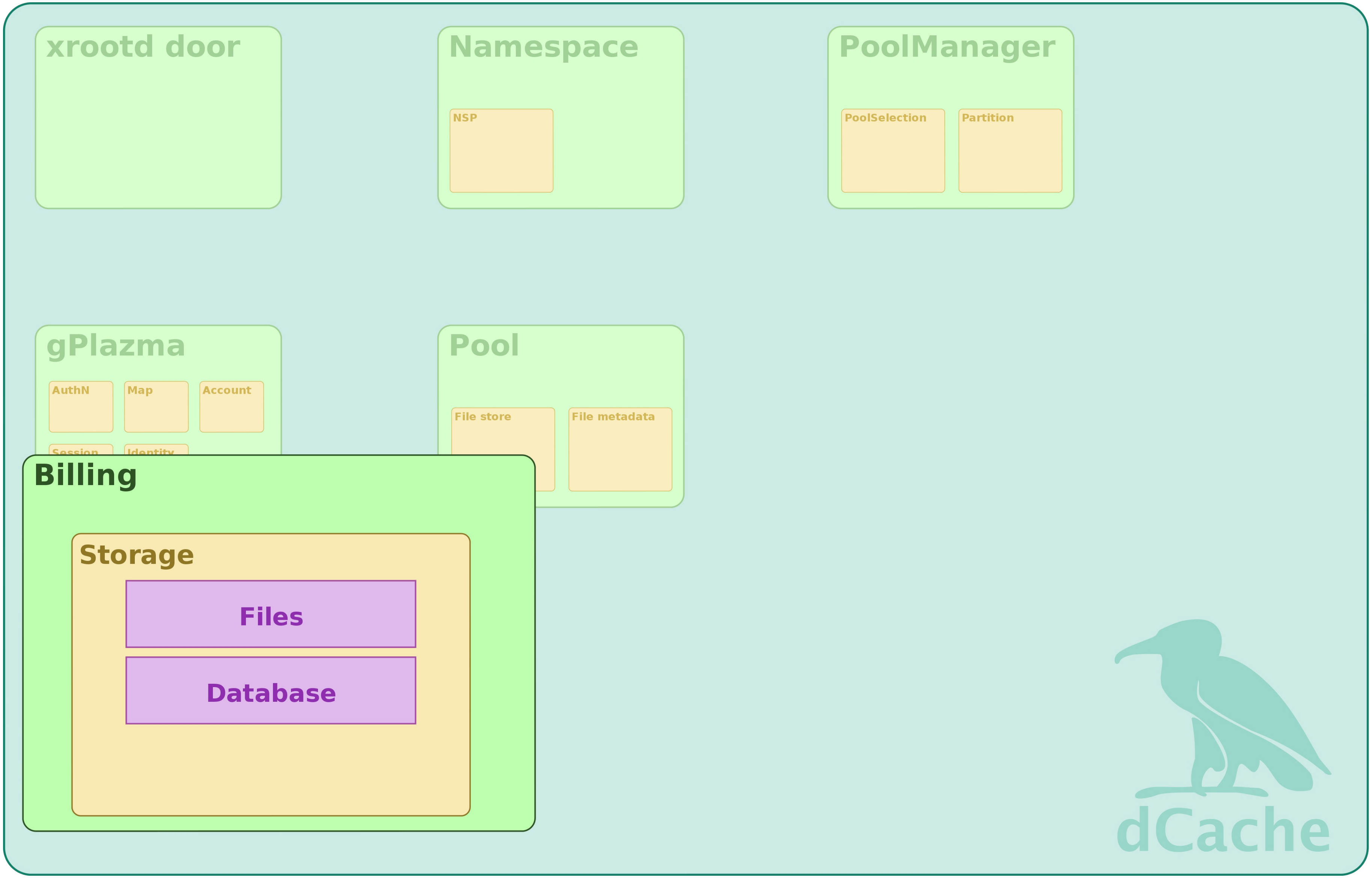
File store

File metadata

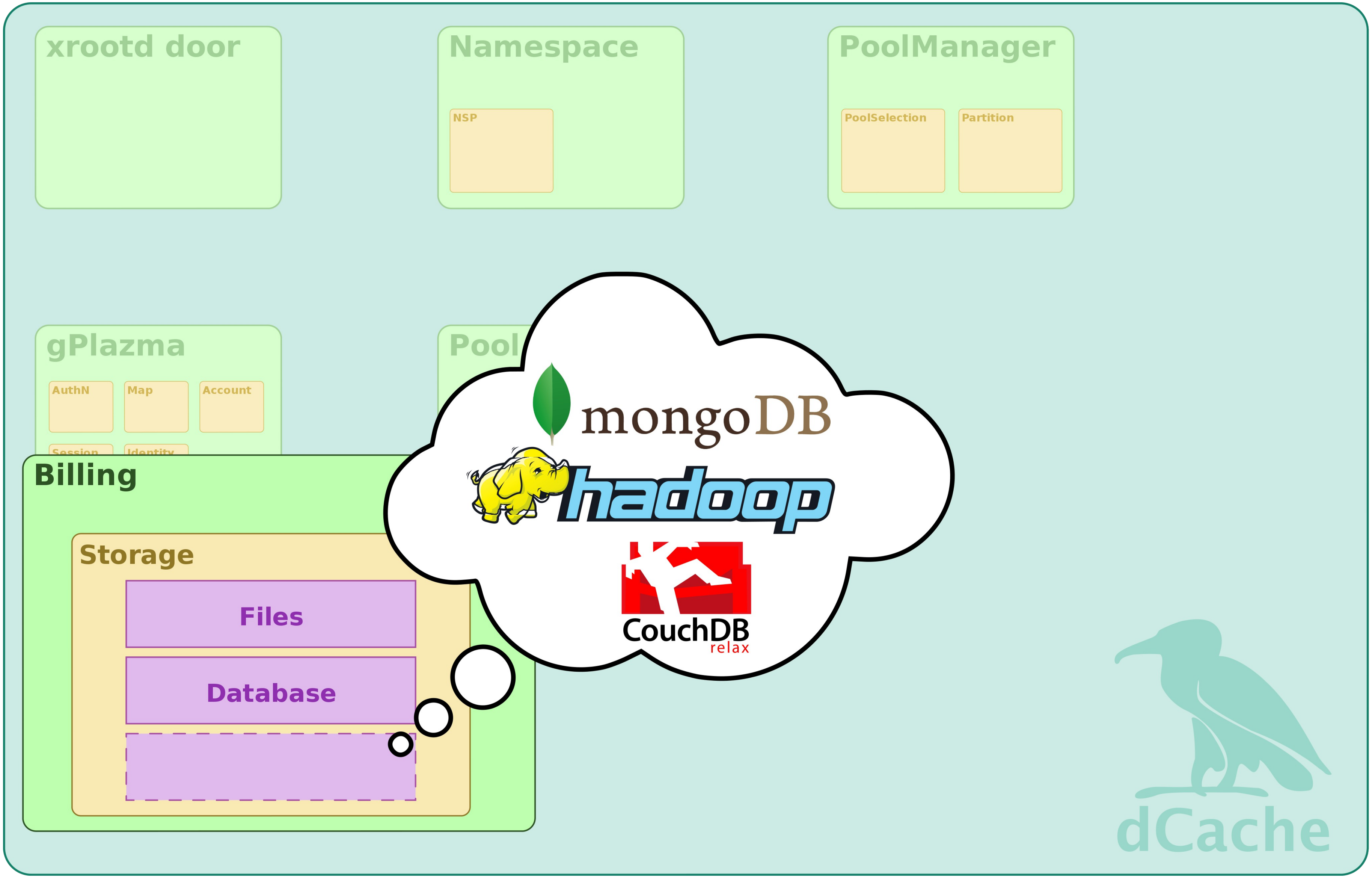
Billing

dCache

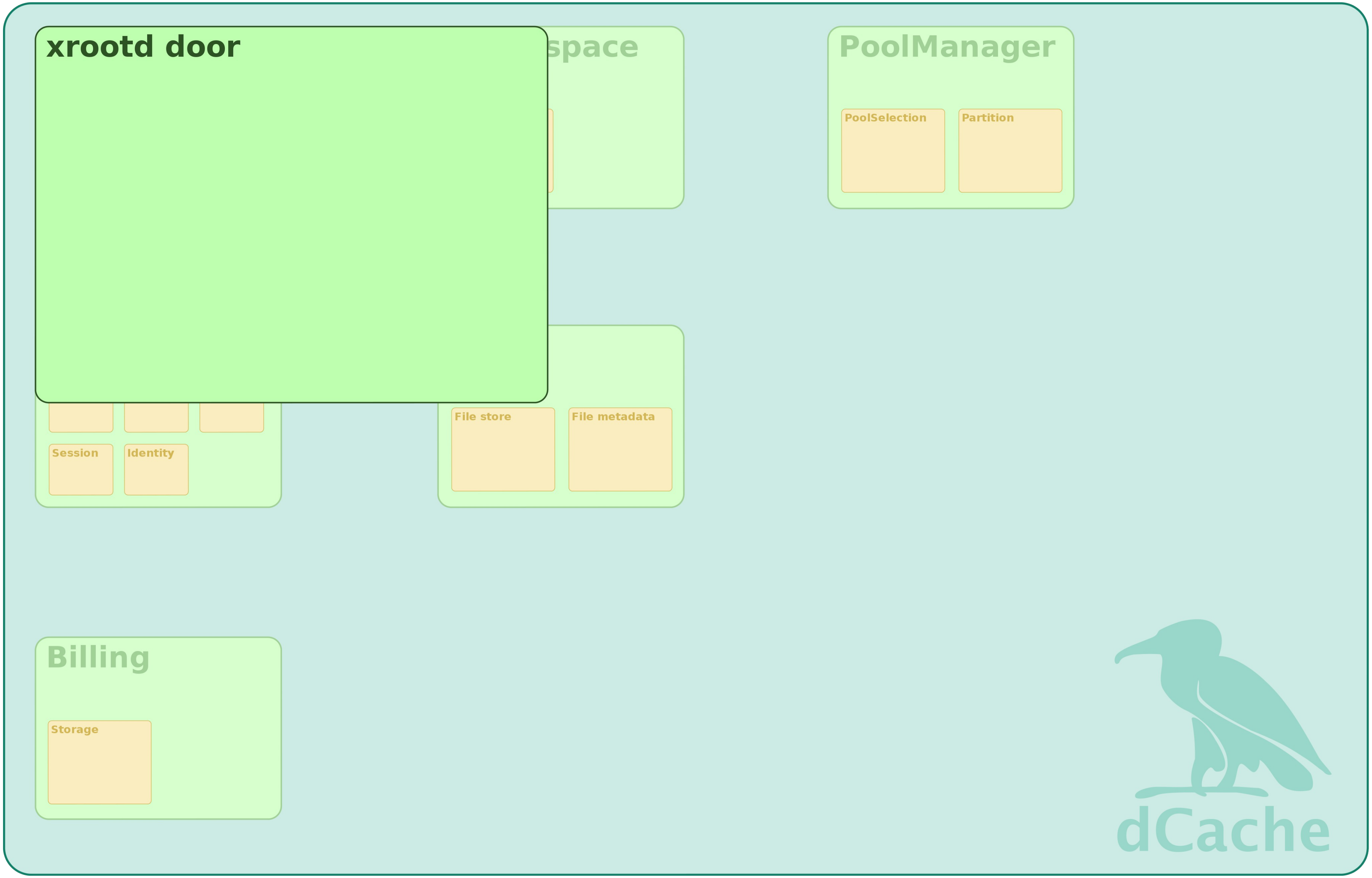
What can I enhance?



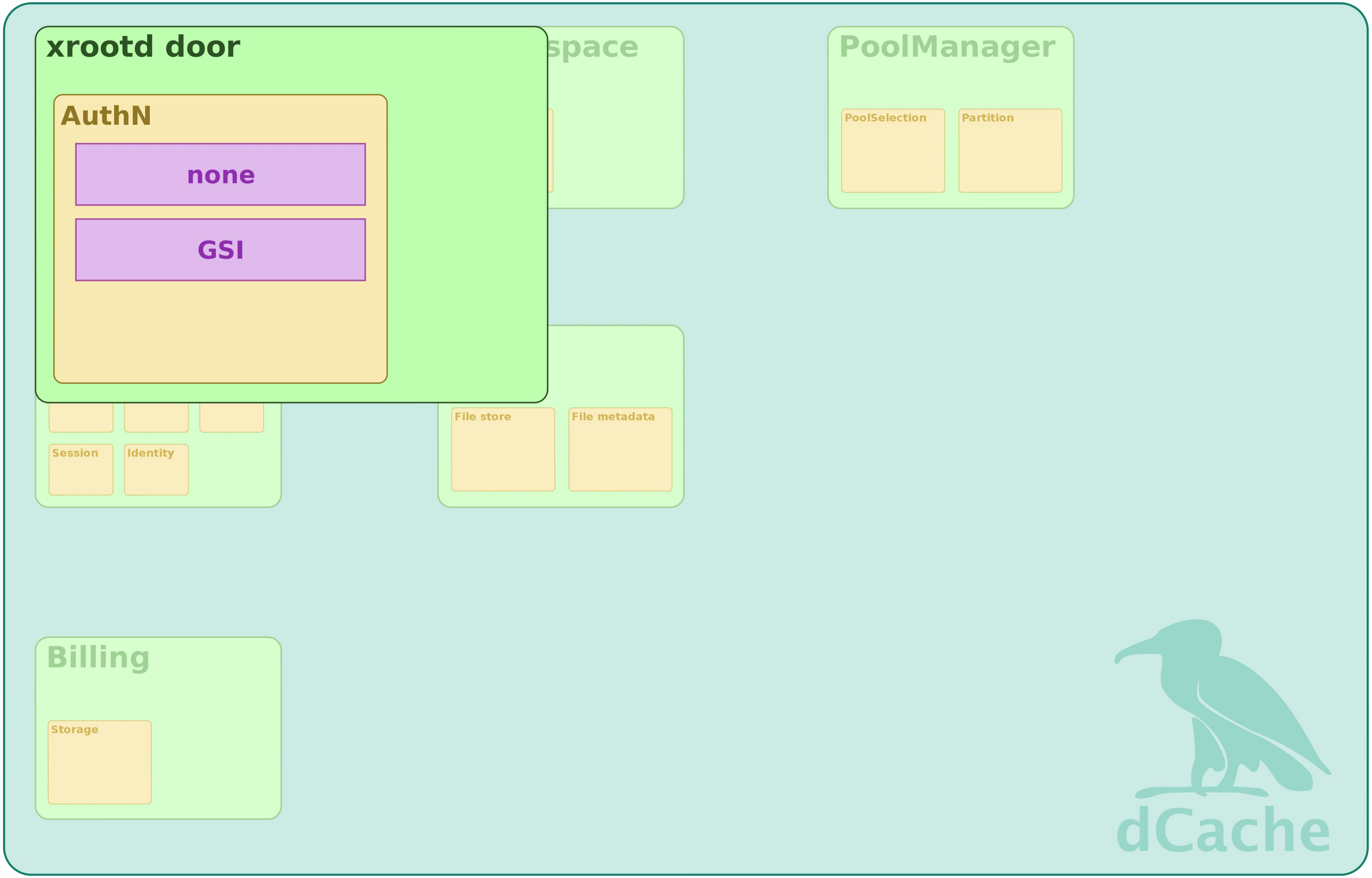
What can I enhance?



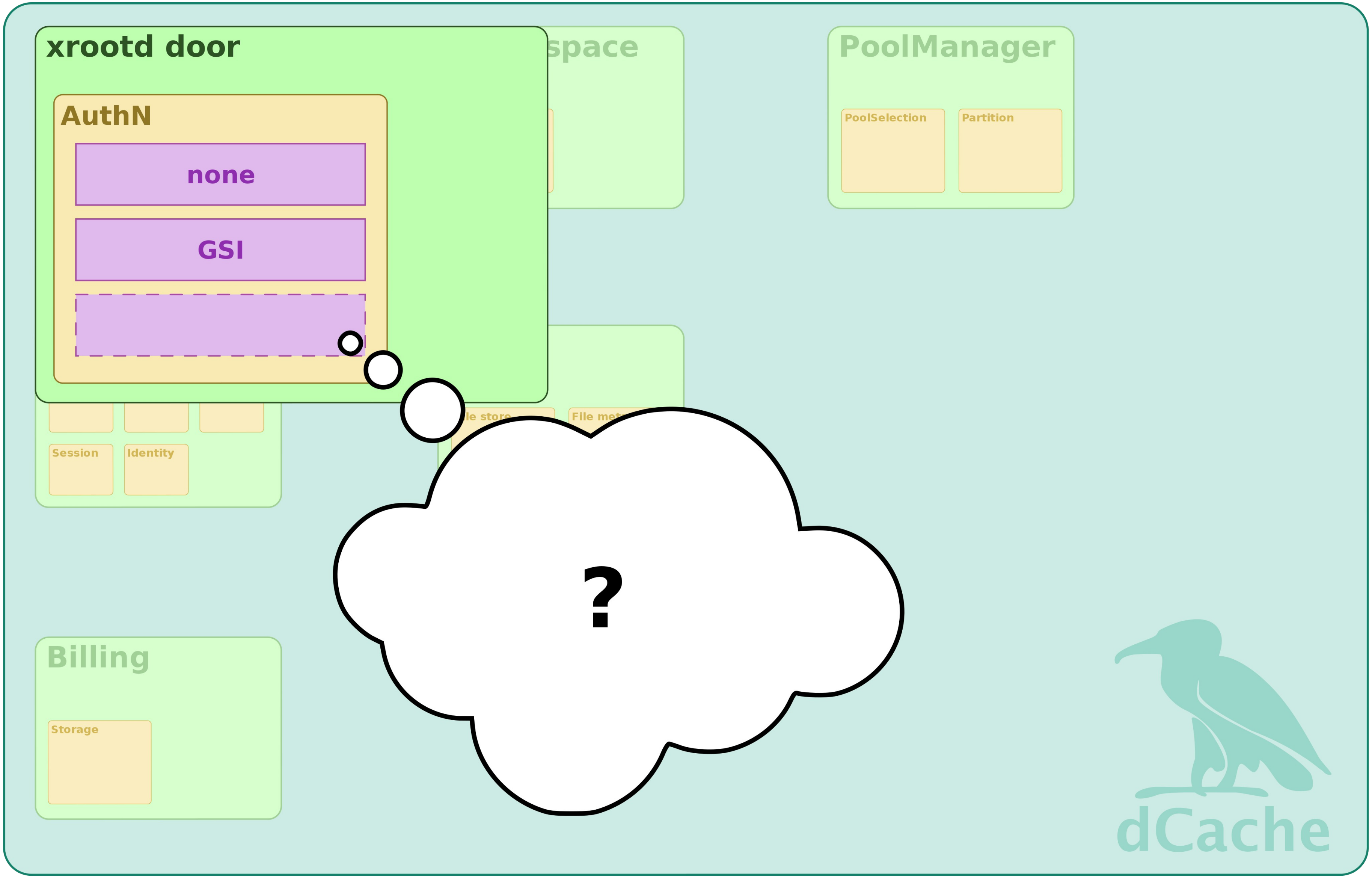
What can I enhance?



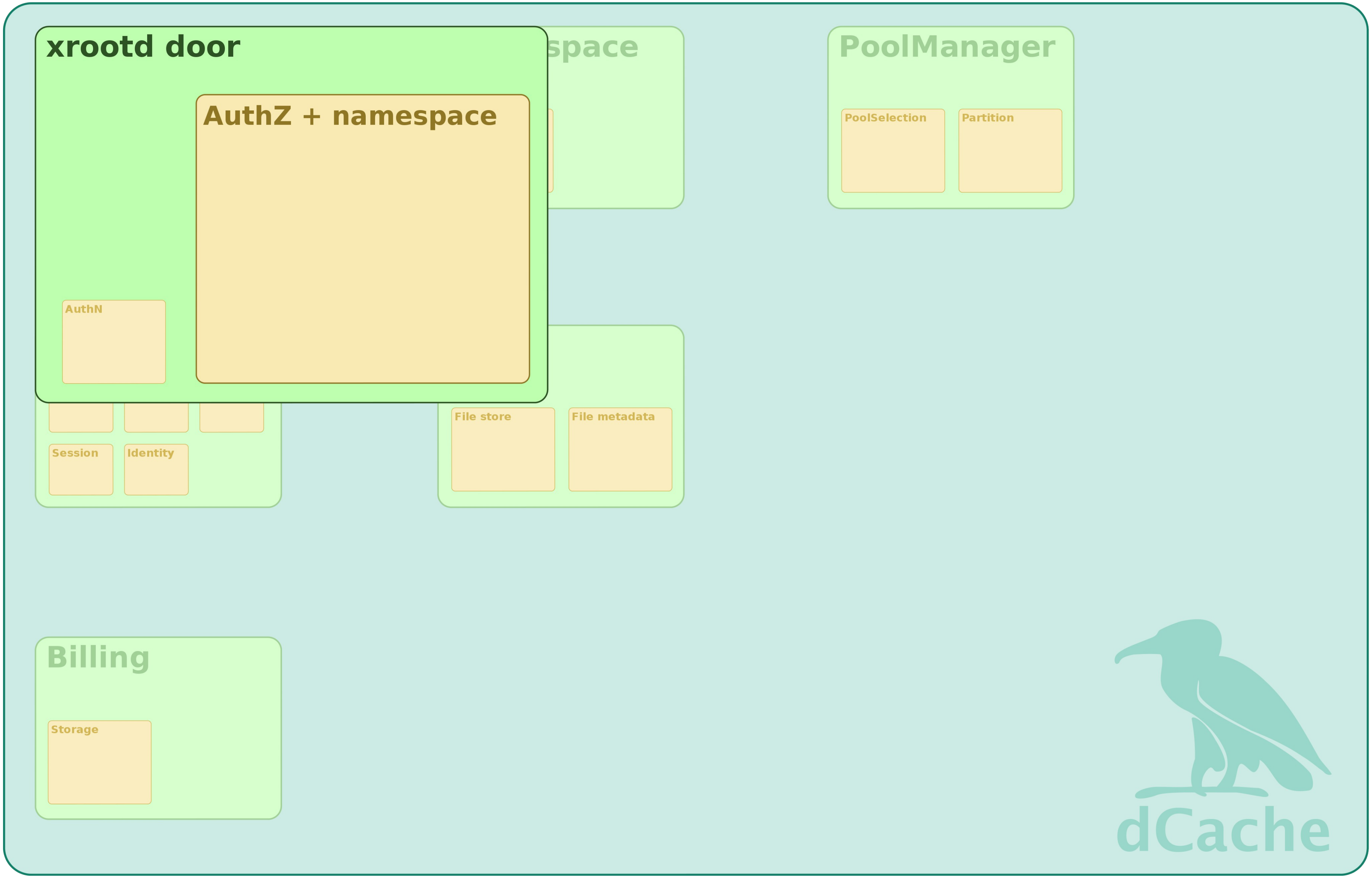
What can I enhance?



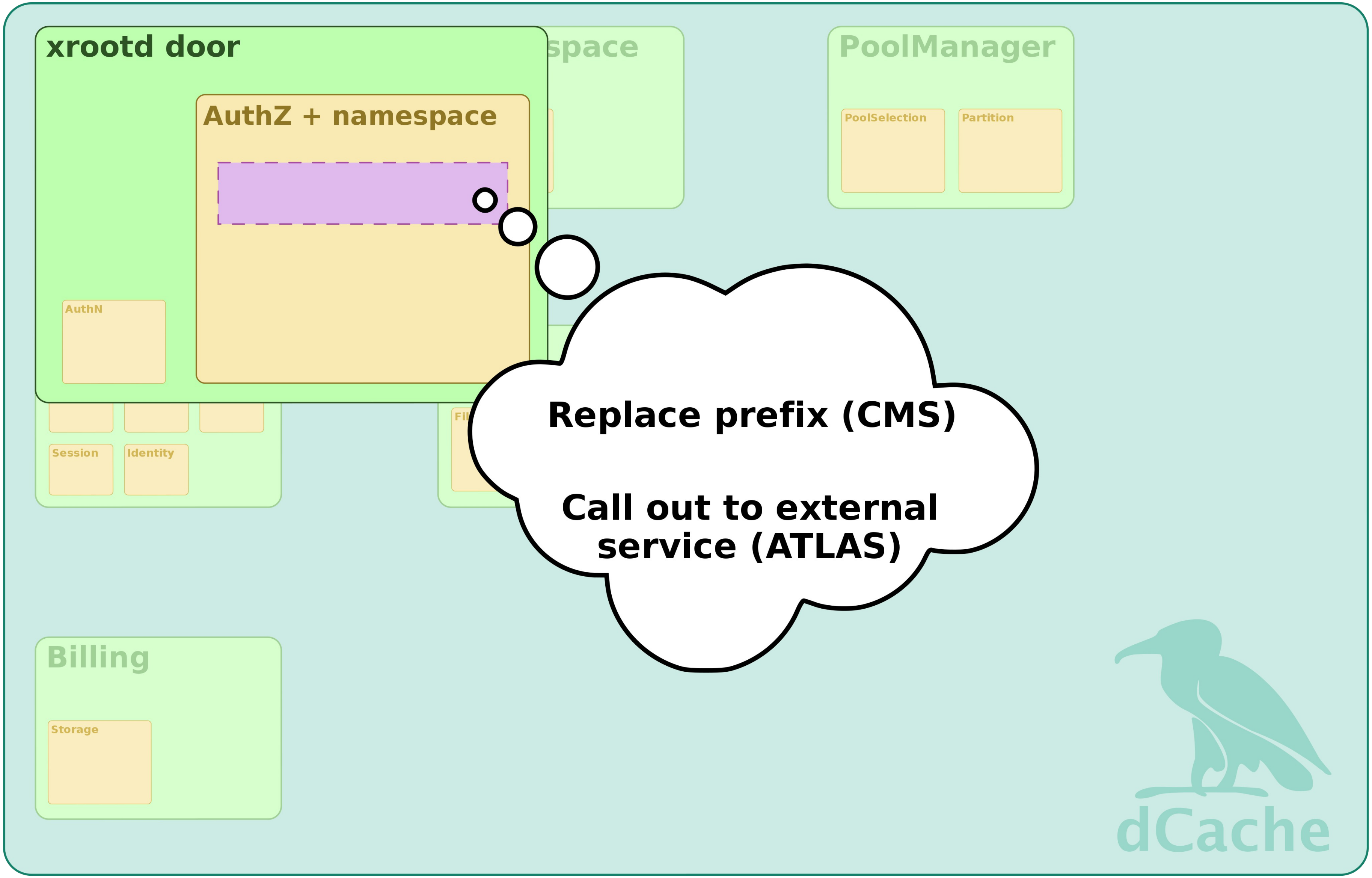
What can I enhance?



What can I enhance?



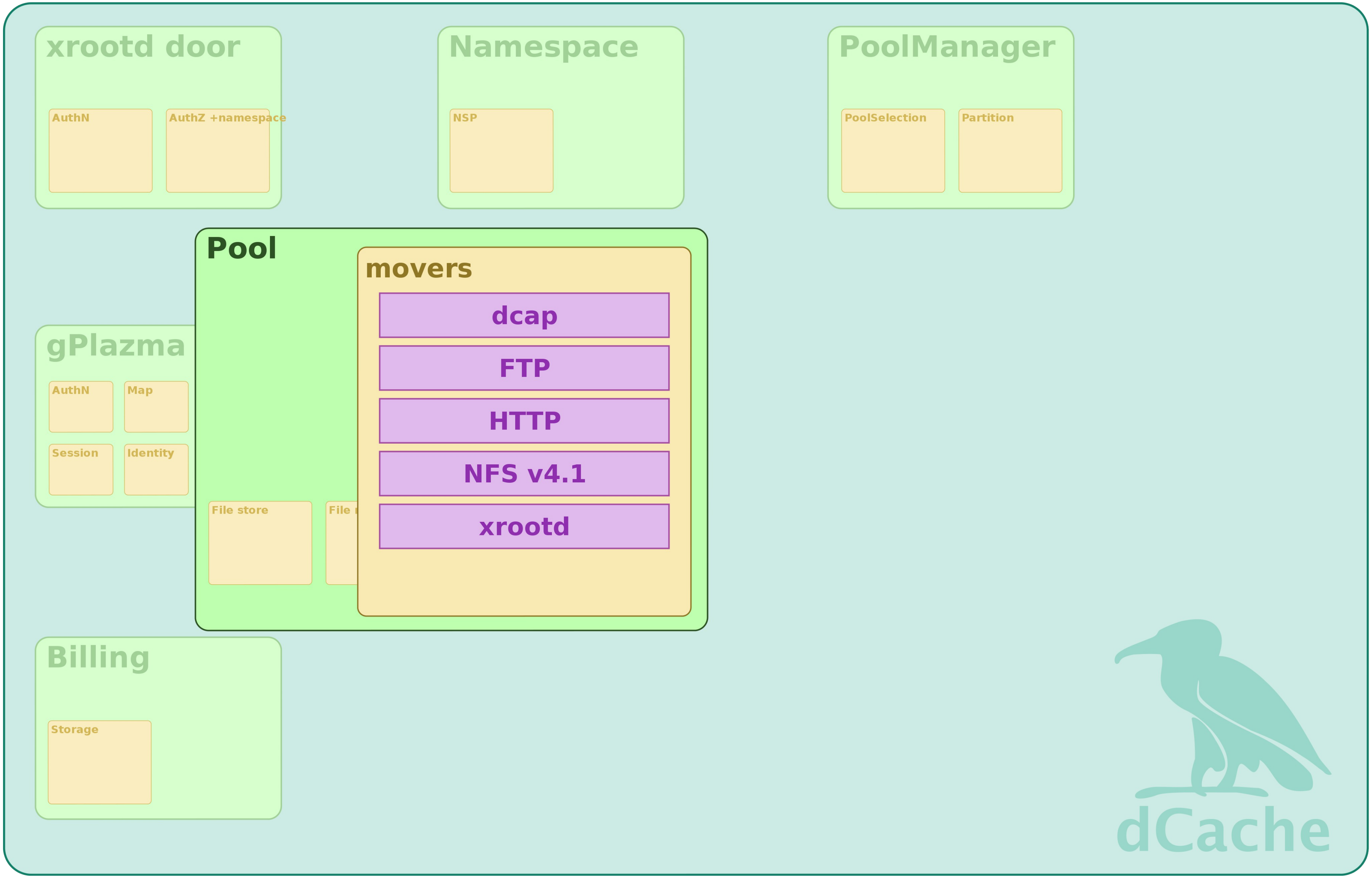
What can I enhance?



Replace prefix (CMS)

Call out to external service (ATLAS)

What can I enhance?



xrootd door

AuthN

AuthZ +namespace

Namespace

NSP

PoolManager

PoolSelection

Partition

gPlazma

AuthN

Map

Session

Identity

Pool

File store

File r

movers

dcap

FTP

HTTP

NFS v4.1

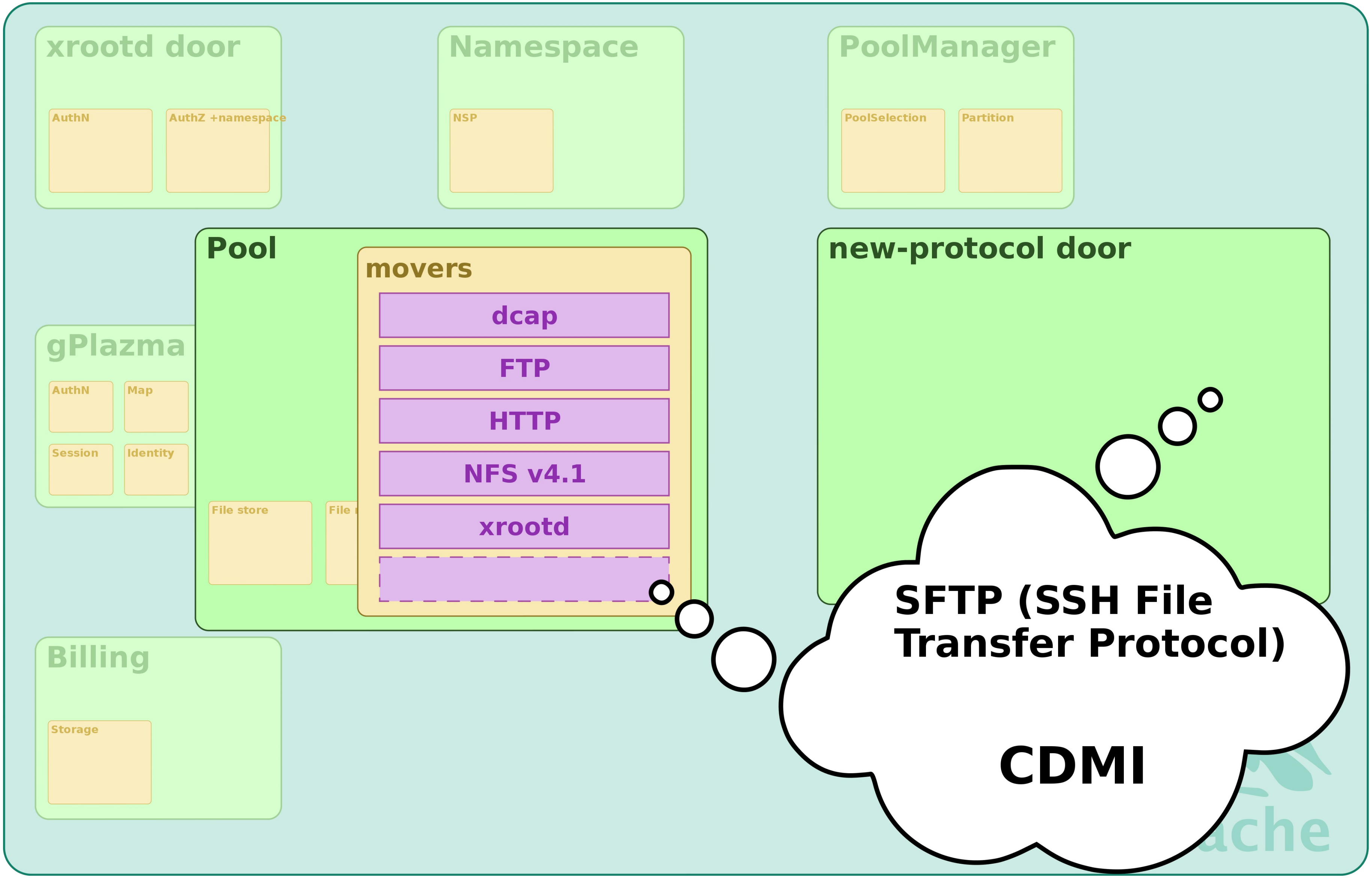
xrootd

Billing

Storage

dCache

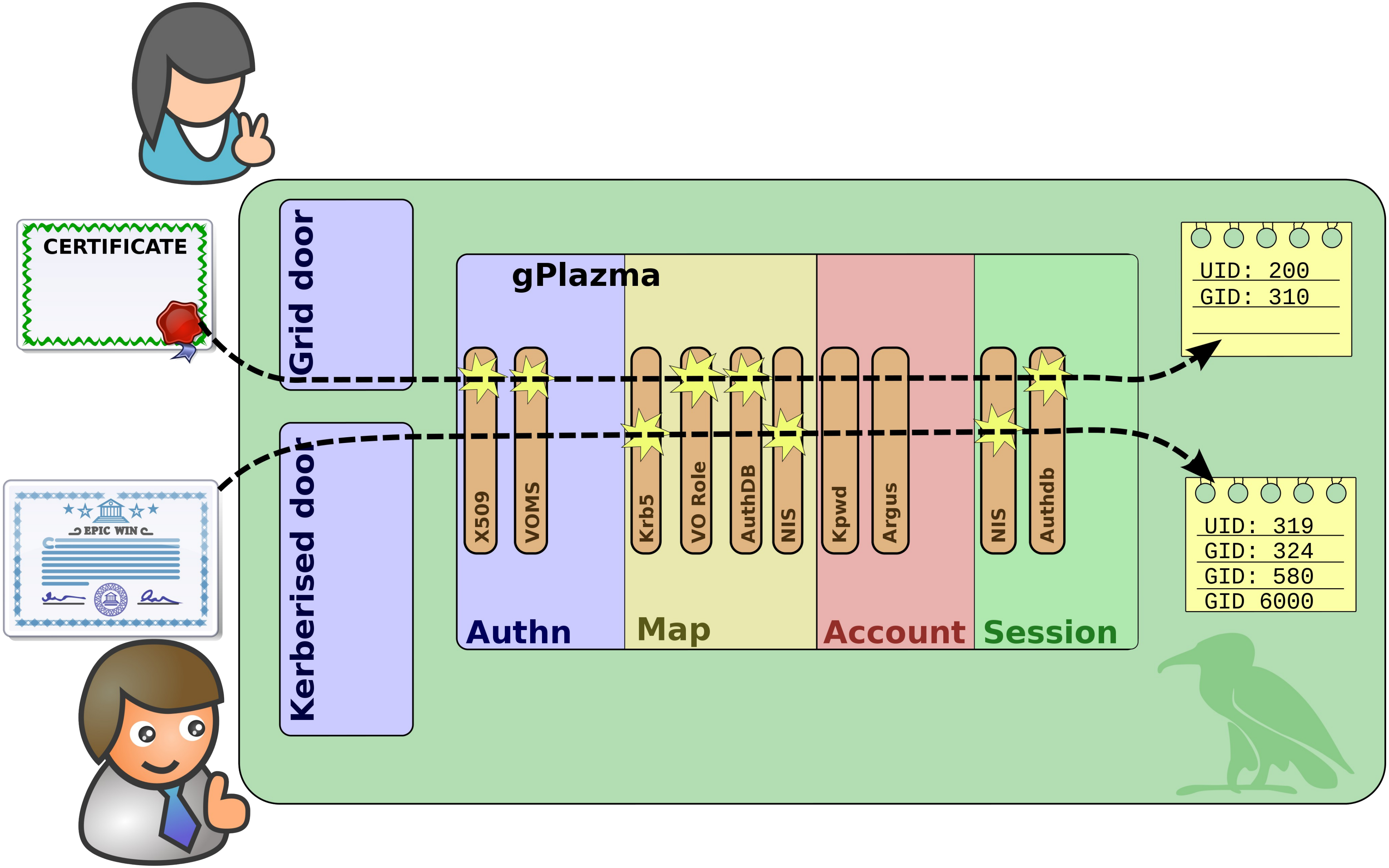
What can I enhance?



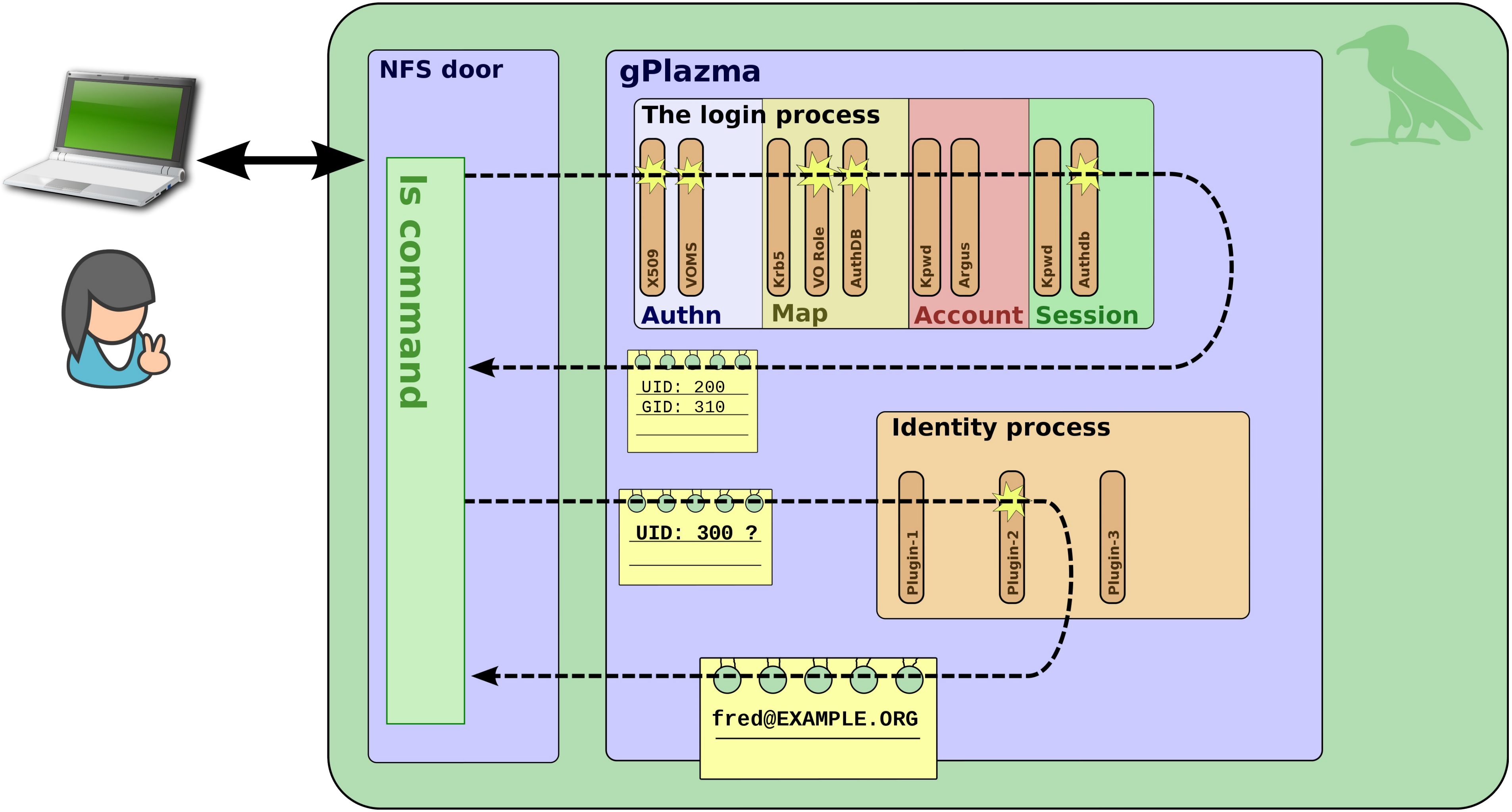
SFTP (SSH File Transfer Protocol)

CDMI

gPlazma: logging in



gPlazma: identities



Future directions

HTTP and WebDAV

- How do we support **non-HEP users**?
- Dcap, SRM, rfio, xrootd

  Nobody outside of HEP has heard of these (HEP is 1% of scientists)
- **HTTP** & **WebDAV**
  - Everyone has a web-browser
  - WebDAV is commonly available on platforms
  - Used by some cloud storage providers (Microsoft SkyDrive, Deutscher Telekom, ..)
- Deployed **in production**: DESY, PIC, BNL, ...

Federating storage

“Collection of disparate storage resources managed by co-operating but independent administrative domains transparently accessible via a common name space.”

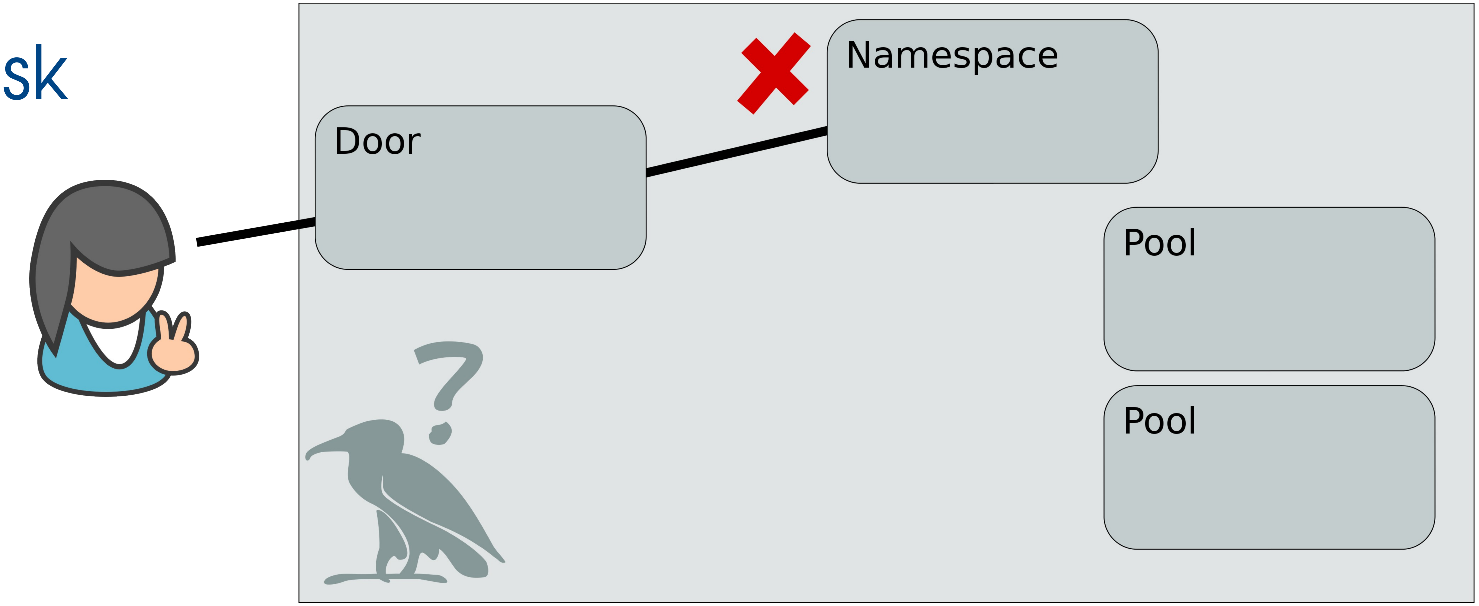
Hey, we can do this with a **standard protocol: HTTP**!

- Benefits:
  - Get **high-performance client** for free,
  - **Loads of free software** (Apache, Squid, Varnish, …)
- Two stage approach:
  - **Web front-end** to existing catalogues (LFC, …)
  - **Dynamically** discovering available data using WebDAV
    - All replicas of a file are discoverable (c.f. **dark data** problem)

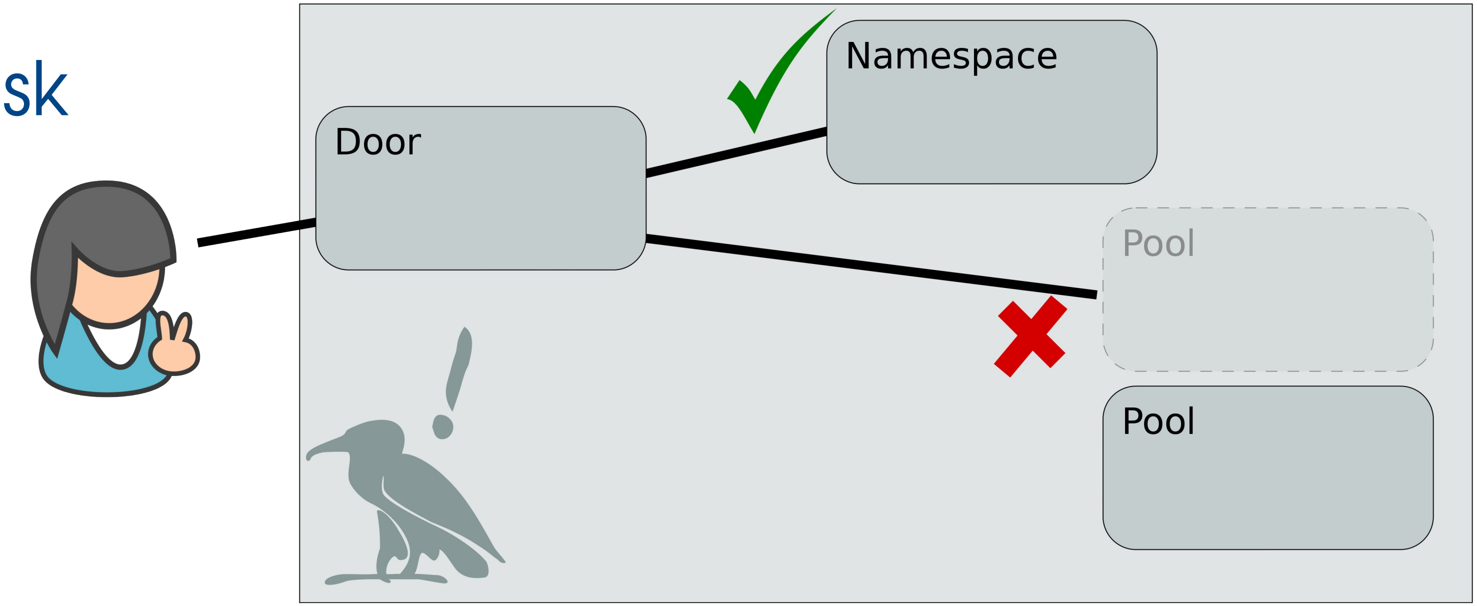
For further details, see ***Dynamic federations: storage aggregation using open tools and protocols*** by **F. Furano**

Missing files

A user may ask for a file that doesn't exist

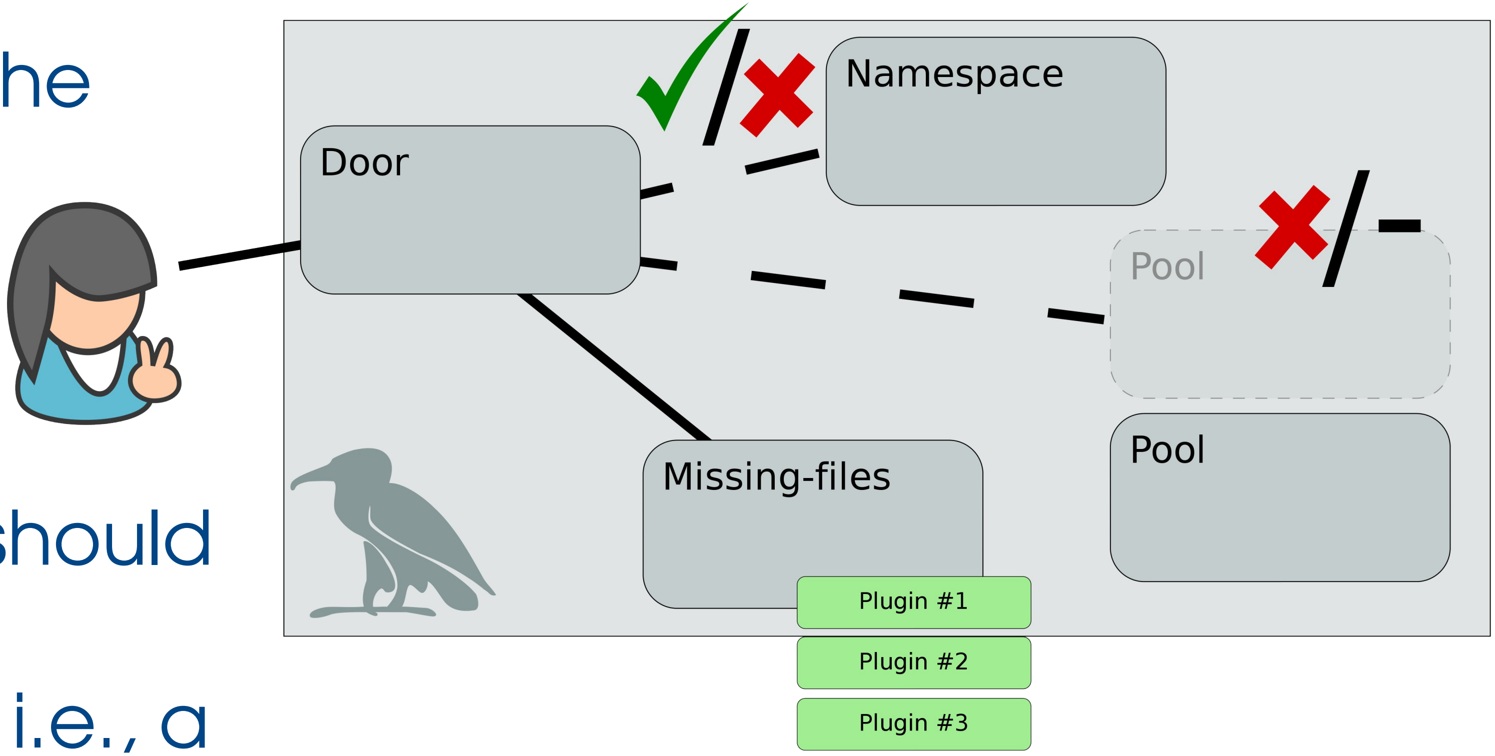


A user may ask for a file that should exist, but the pool is broken



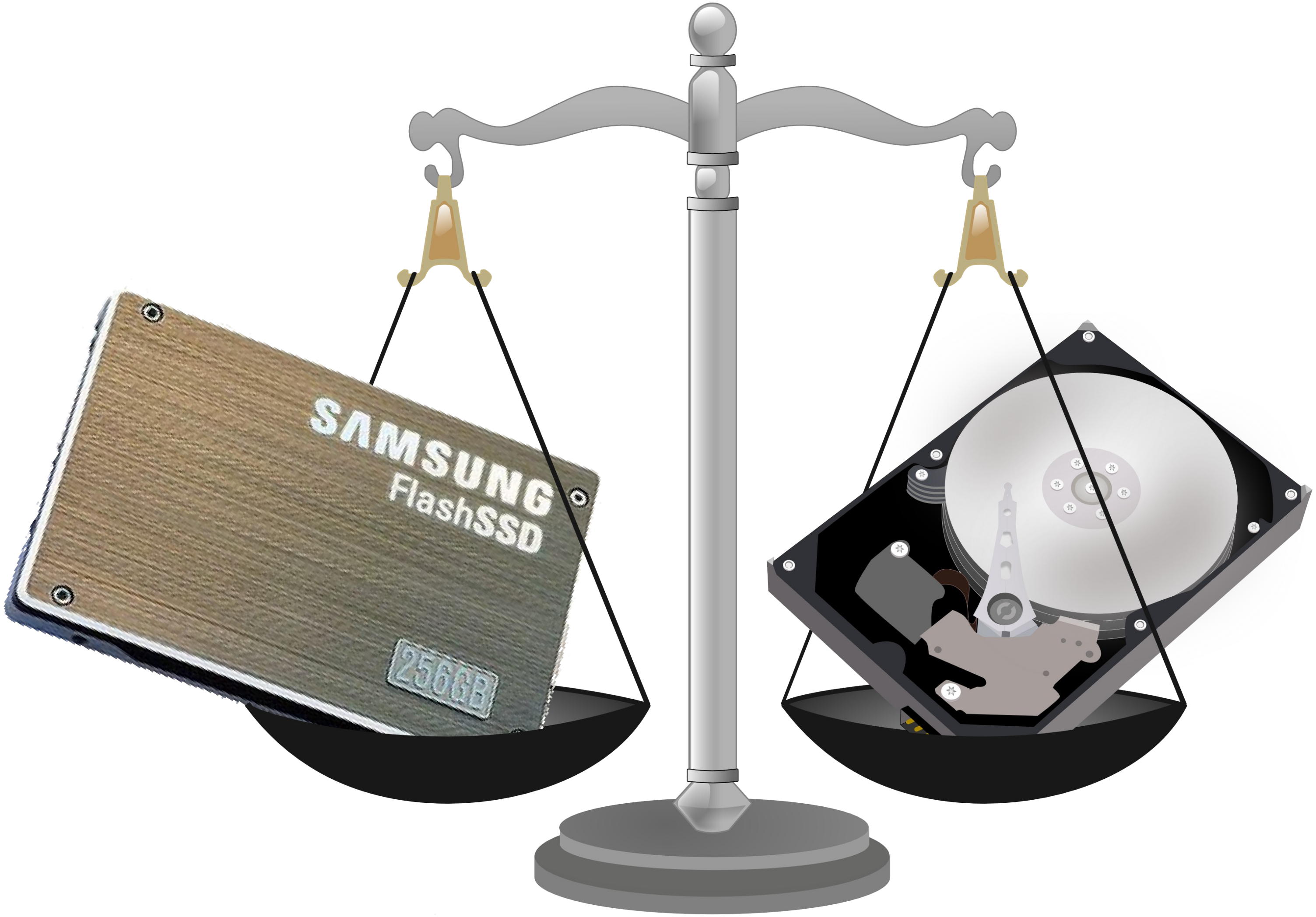
Missing files

Maybe dCache should do "something" in these cases. That "something" should be highly configurable; i.e., a plugin.

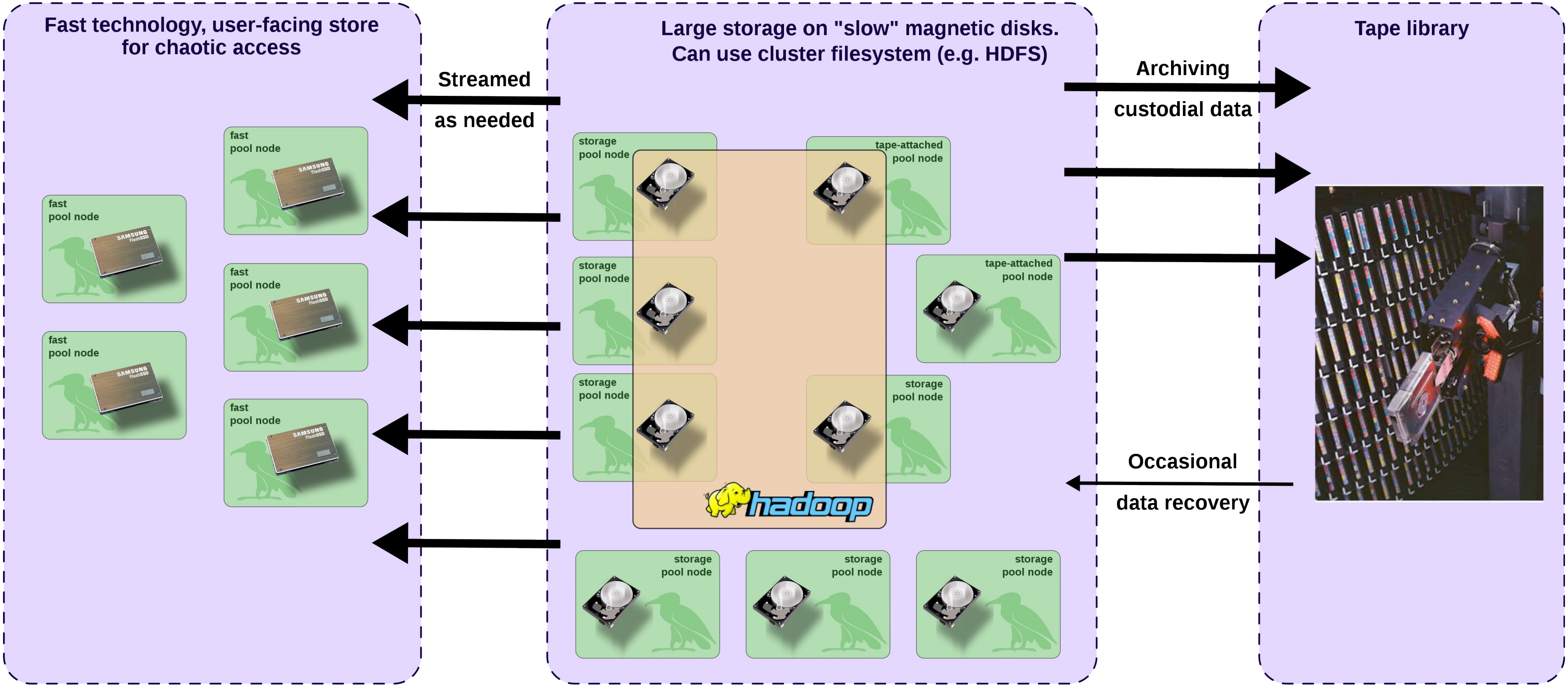


For further details, see ***SYNCAT – Storage Catalogue Consistency*** by **F. Furano**

Faster storage



3 Tier Model



For further details see ***Evaluation of benefits of a three tier data model for WLCG analysis*** by **D. Ozerov** and **P. Fuhrmann**

Summary

The dCache project is **independent** of WLCG and EMI funding.

dCache has the **flexibility** to adapt to new deployments, scenarios and technology.

The dCache community is **growing**.

Thanks for listening