DE LA RECHERCHE À L'INDUSTRIE

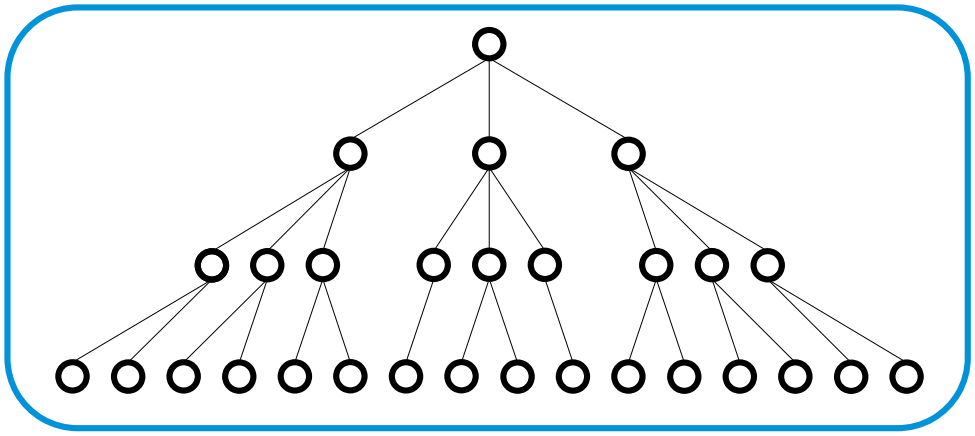
RobinHood-v4 progress report

Commissariat à l'énergie atomique et aux énergies alternatives – www.cea.fr

What is RobinHood?

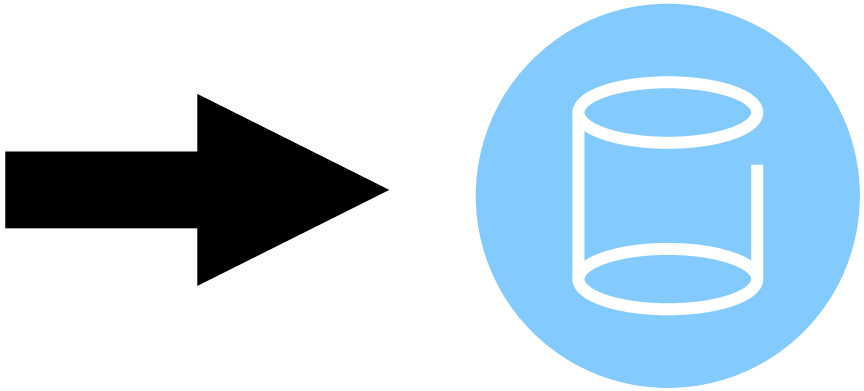
RobinHood provides efficient and easy to use means to replicate and query any filesystem's metadata.

Filesystem



search: O(n)

Database

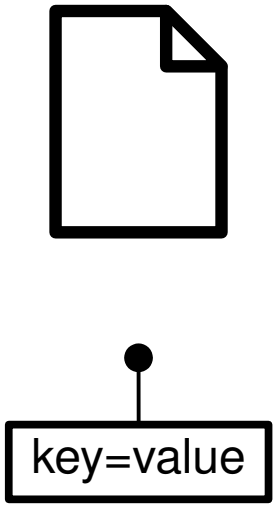


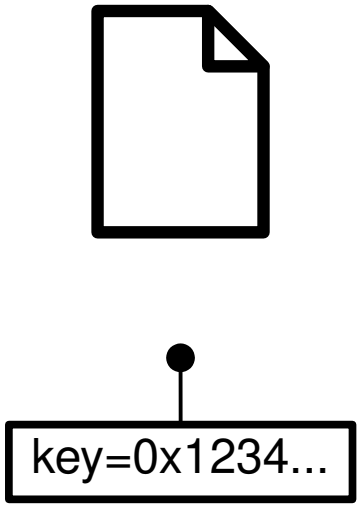
search: O(1) (with an index)

What’s new?

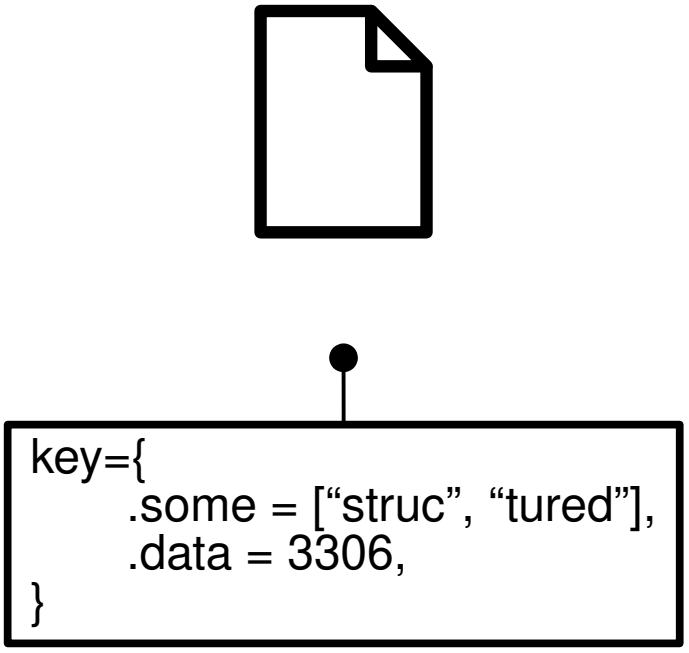
Extended attributes (xattrs)

Custom metadata attributes

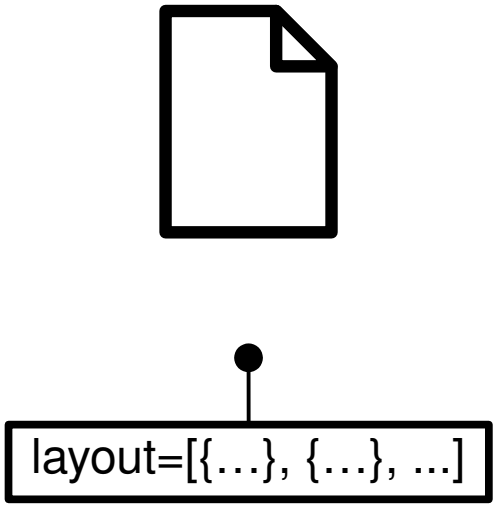

key=value

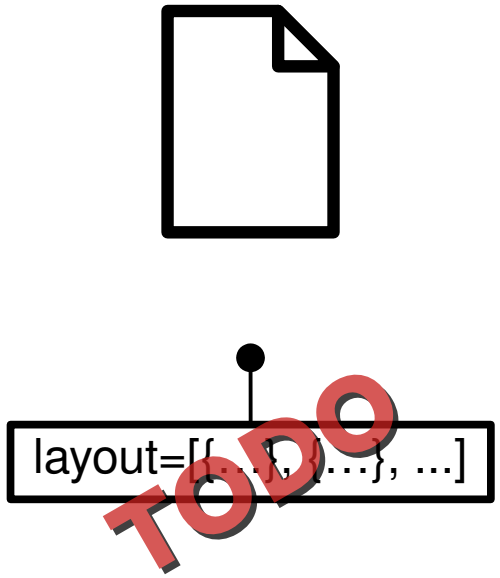

key=0x1234...

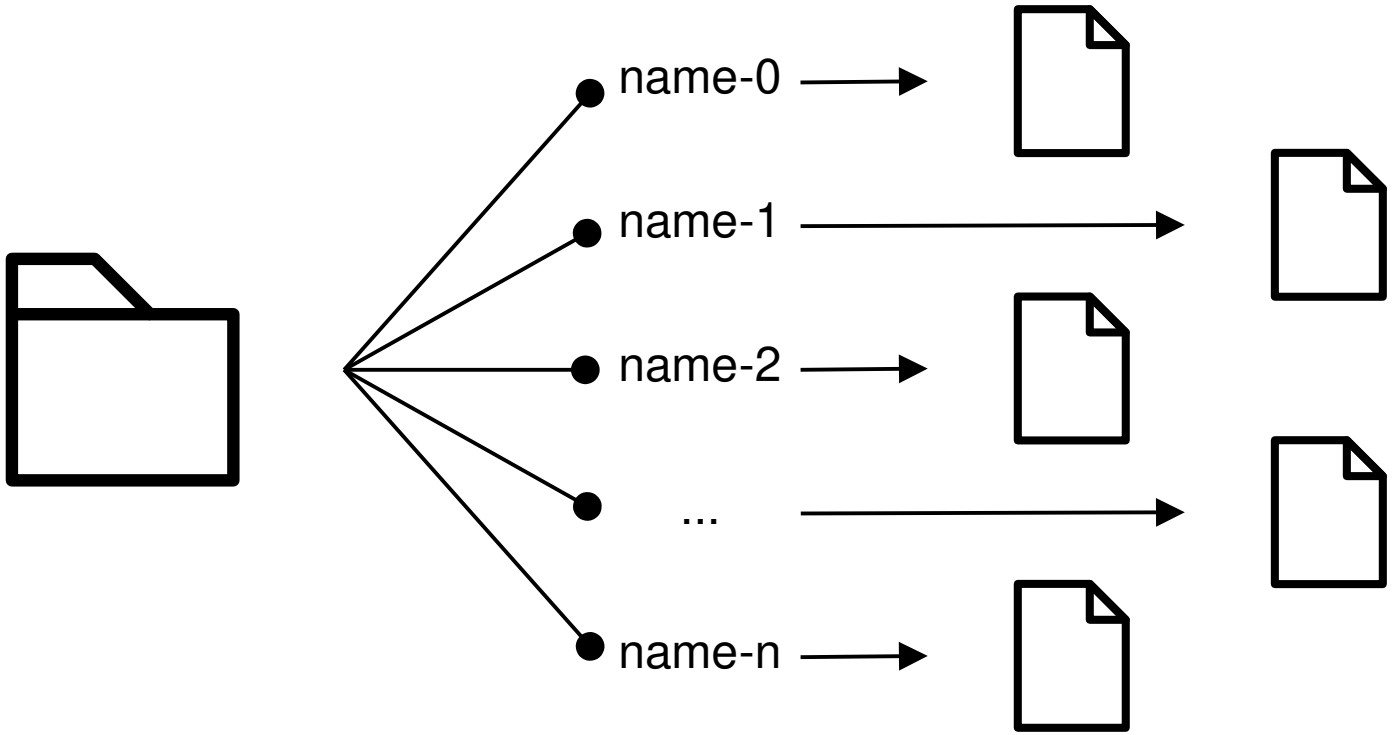
xattrs

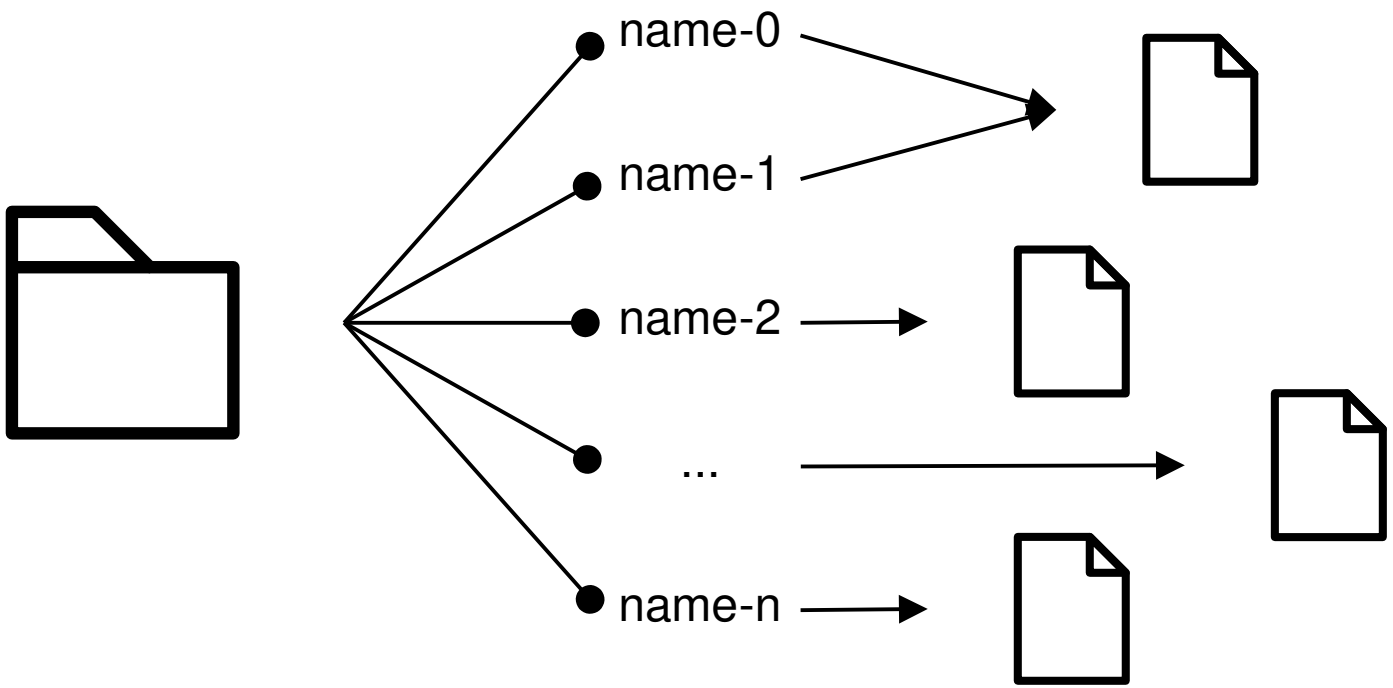


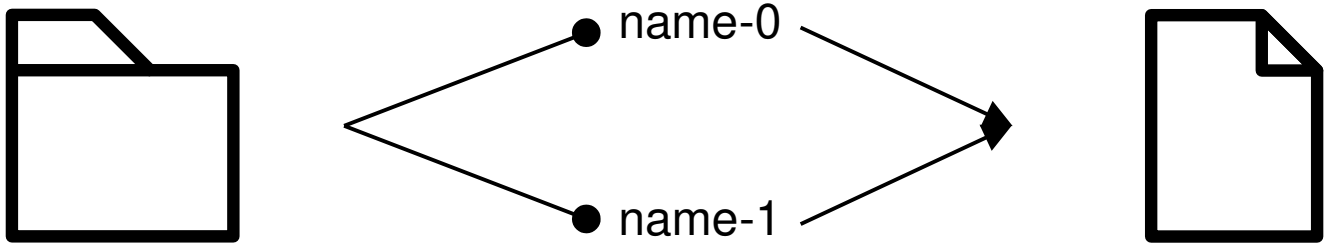
- string-indexed mappings
- sequences
- scalars (sized integers, strings, binary blobs)

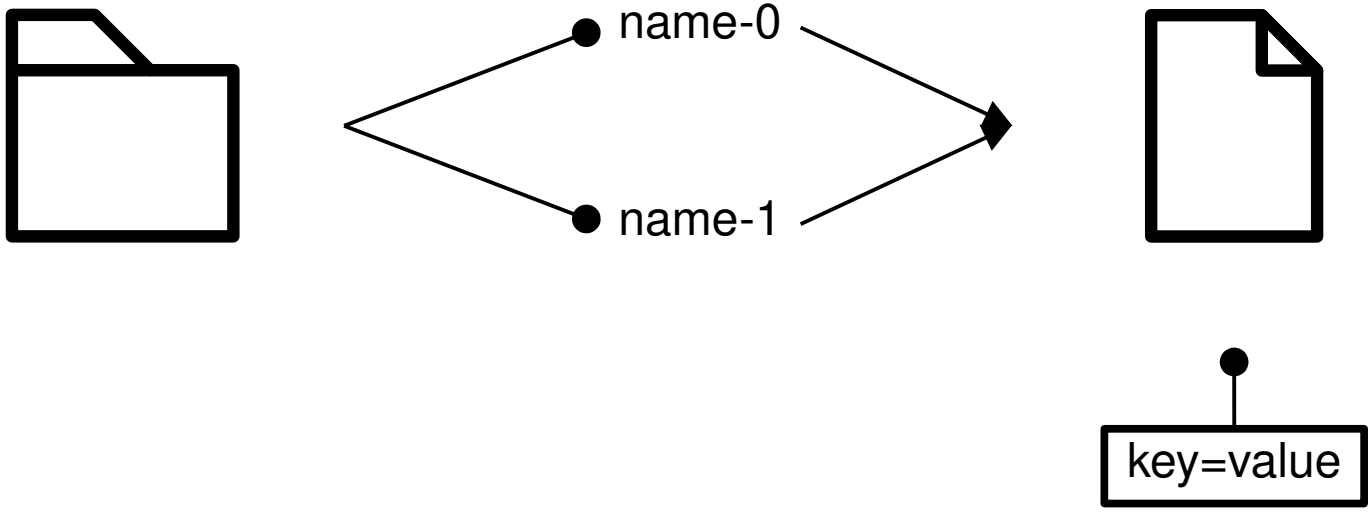

layout=[{…}, {…}, ...]

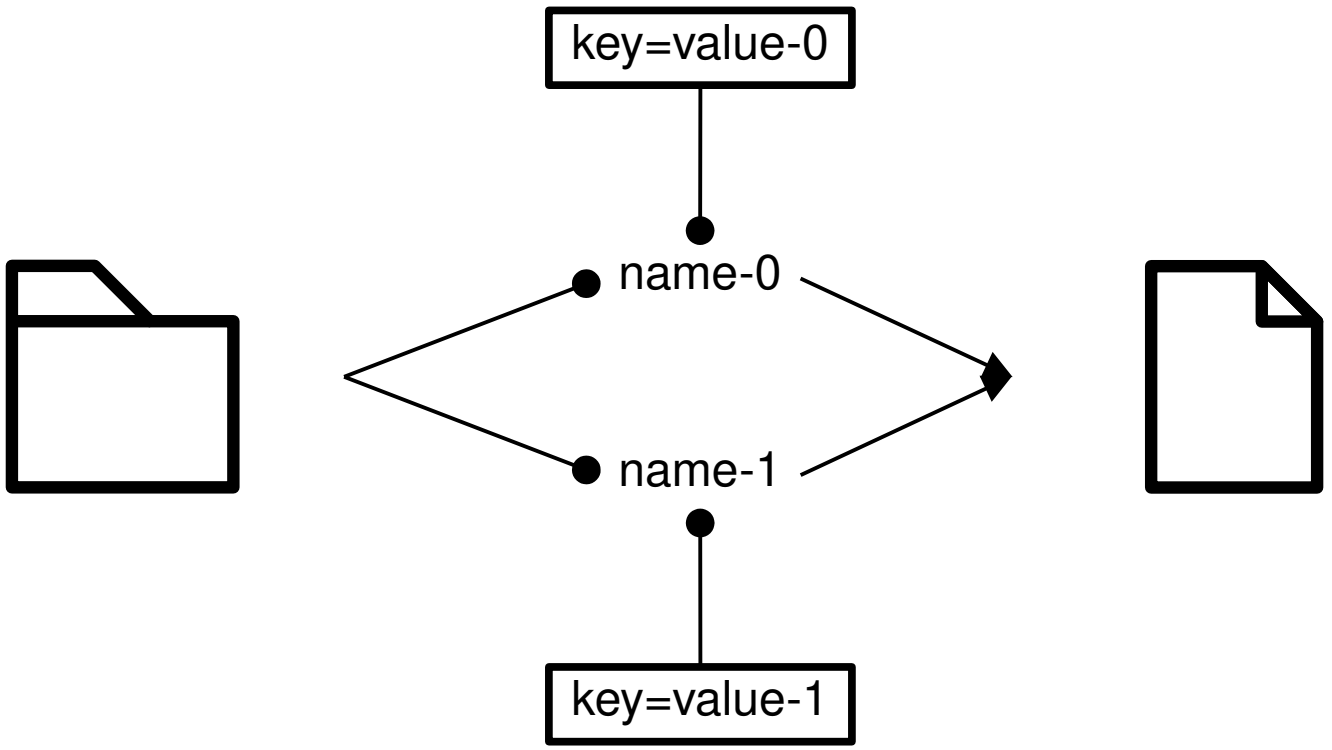

layout=[{...}, {...}, ...]
TODO

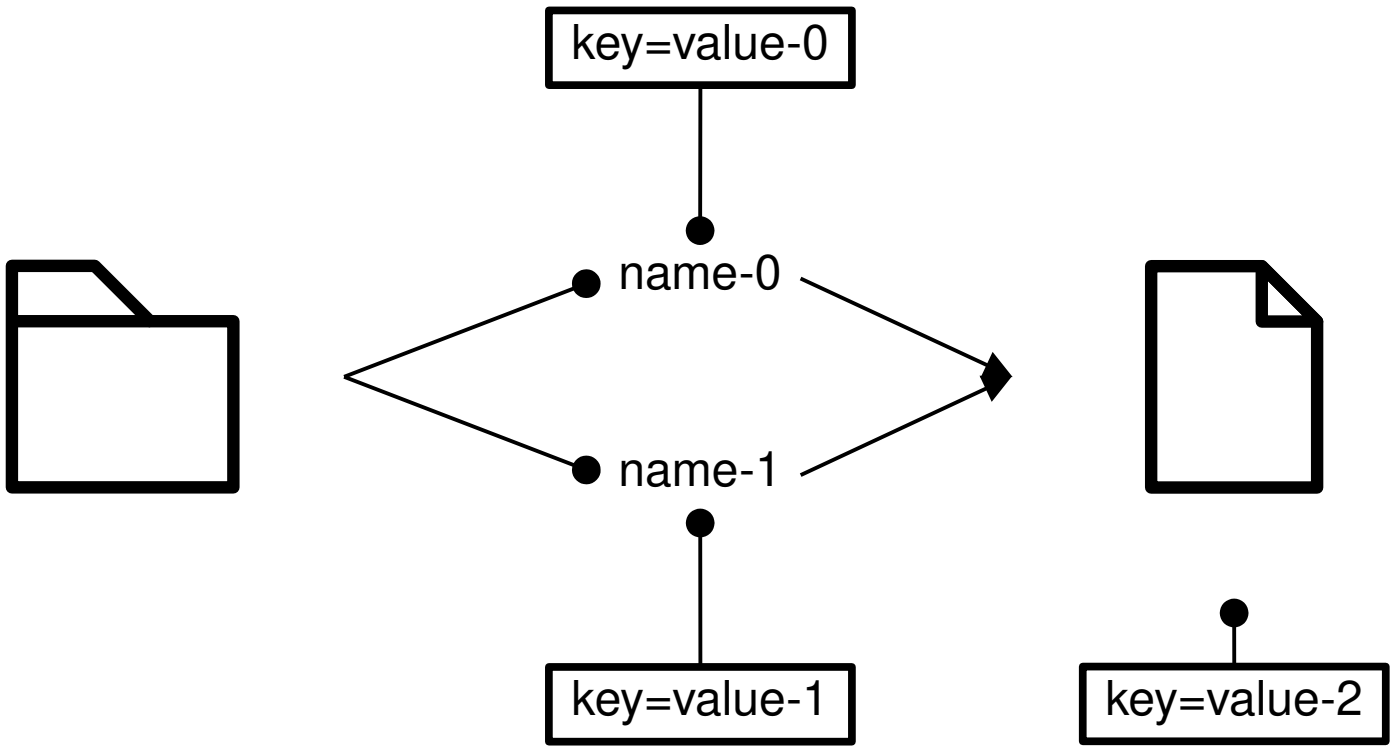

name-0
name-1
name-2
...
name-n

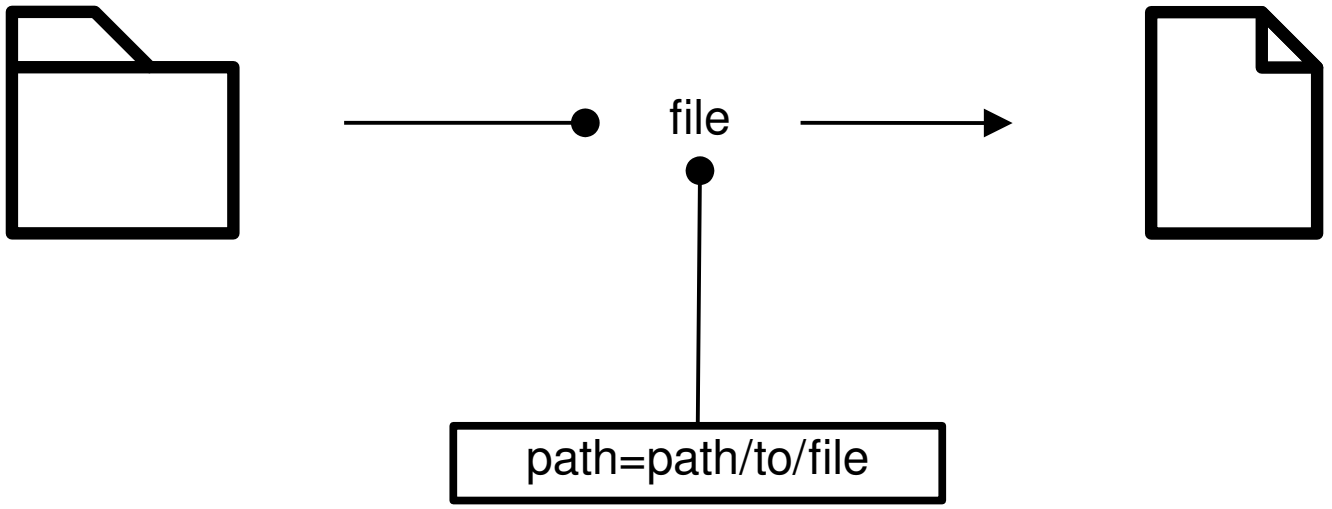

name-0
name-1
name-2
...
name-n


name-0
name-1


name-0
name-1
key=value


key=value-0
name-0
name-1
key=value-1

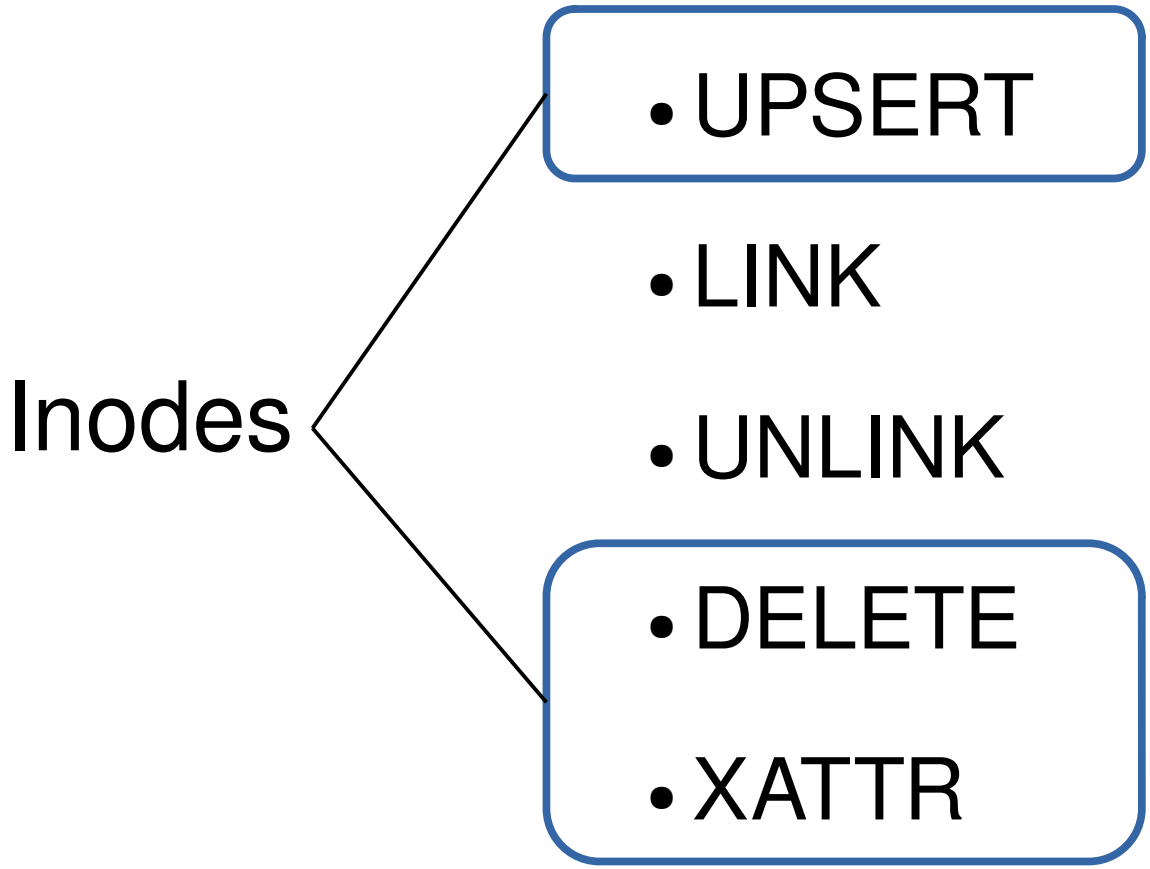

key=value-0
name-0
name-1
key=value-1
key=value-2

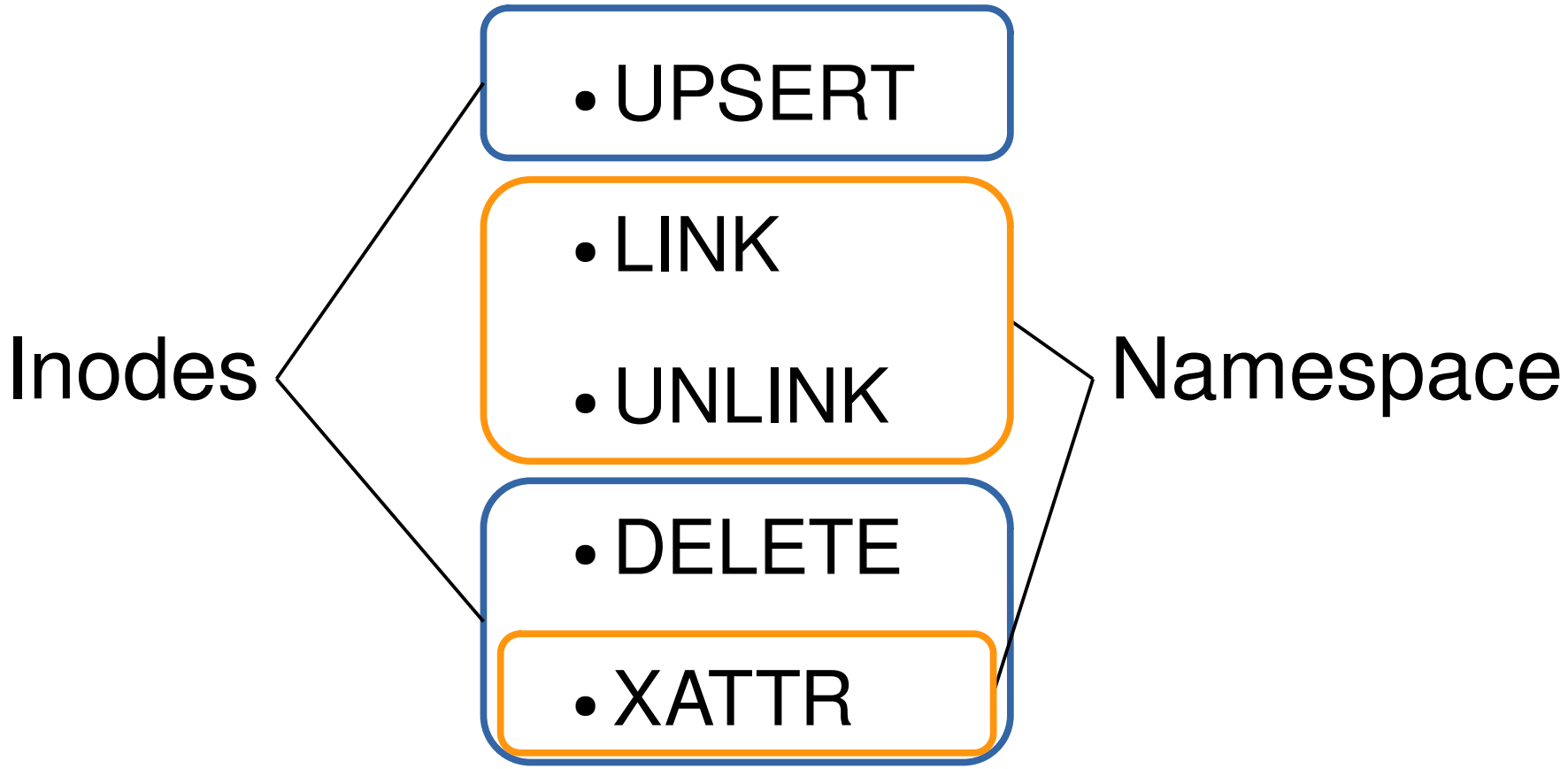

file
path=path/to/file

FSEvents

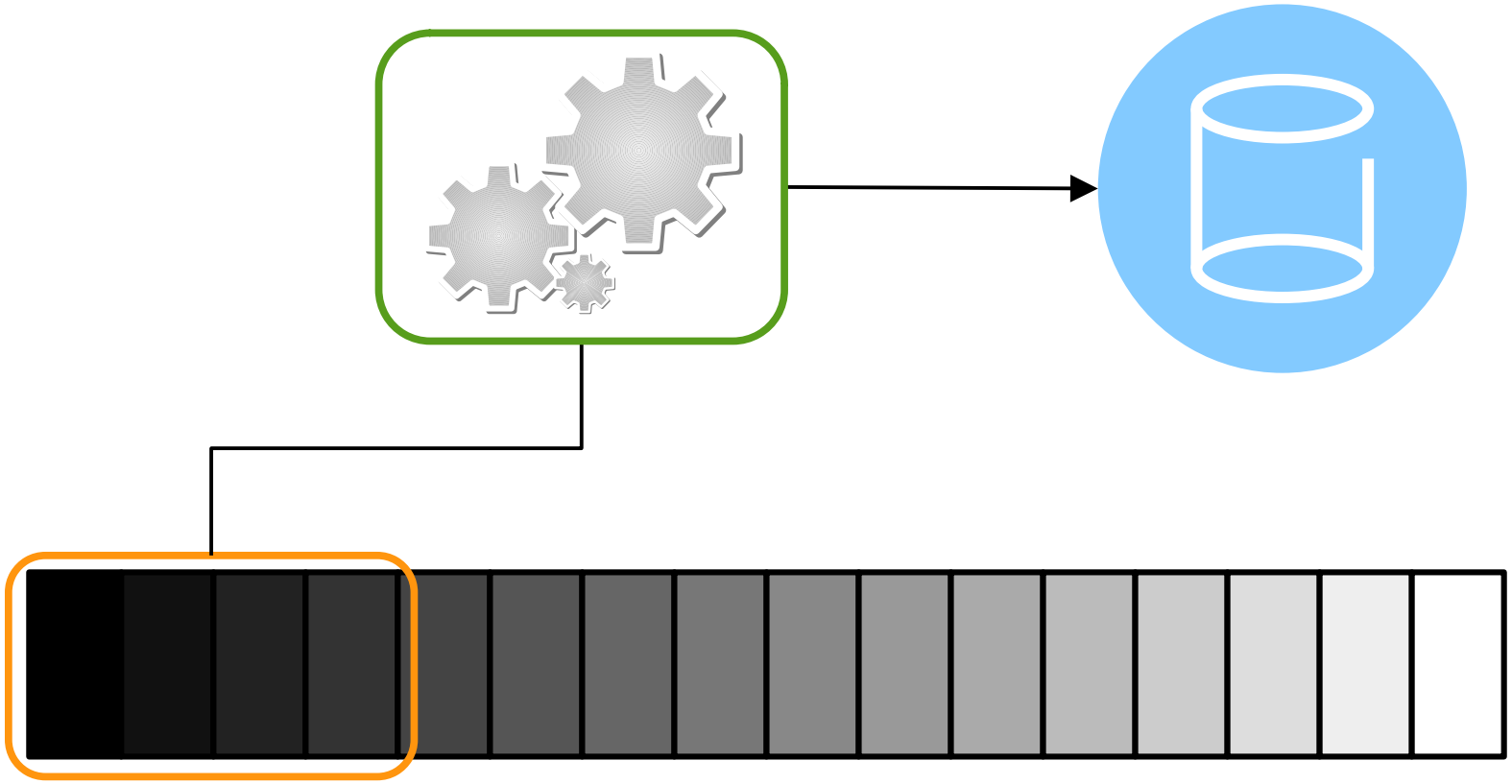
Metadata update records

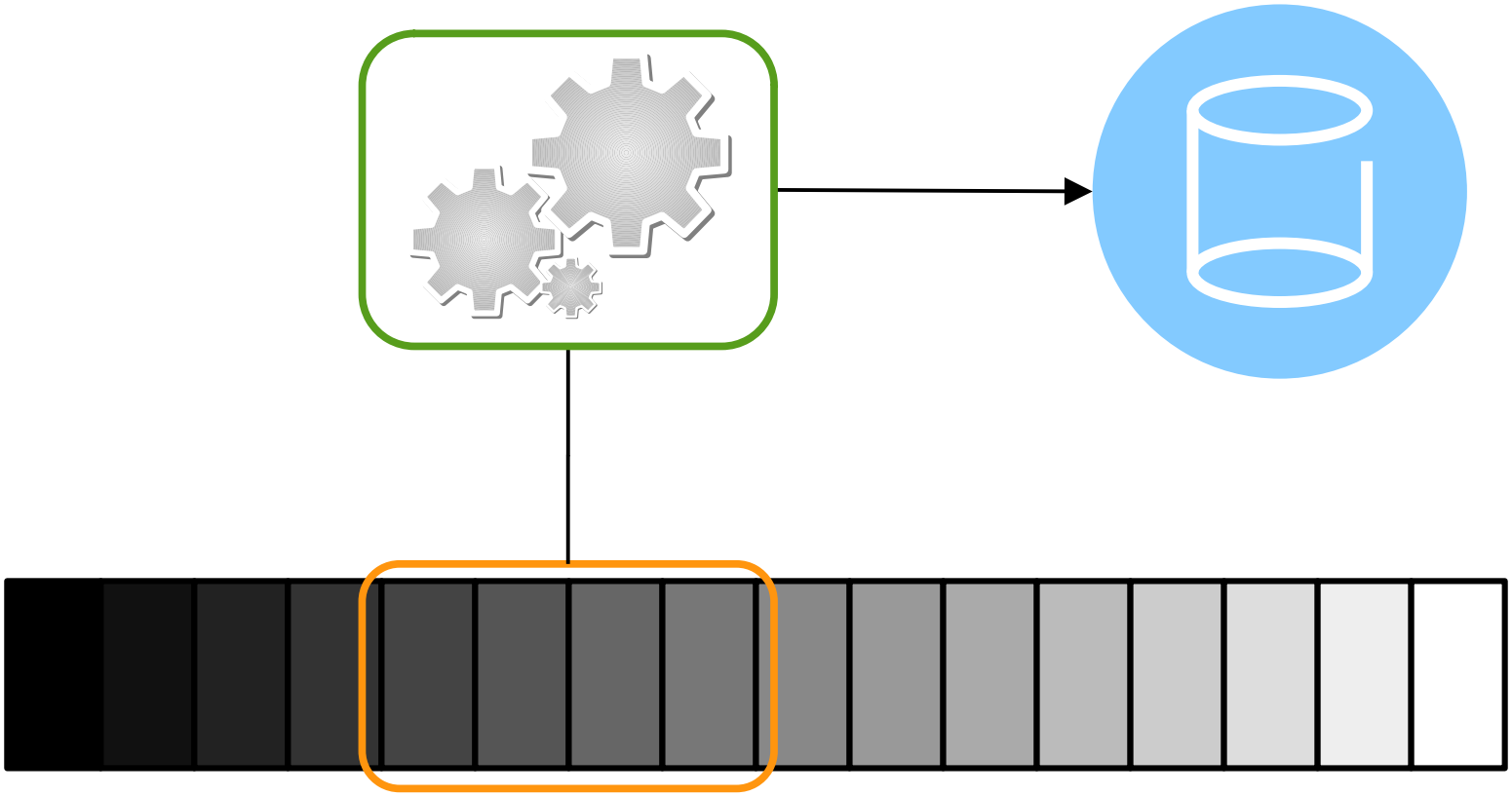
- UPSERT
- LINK
- UNLINK
- DELETE
- XATTR

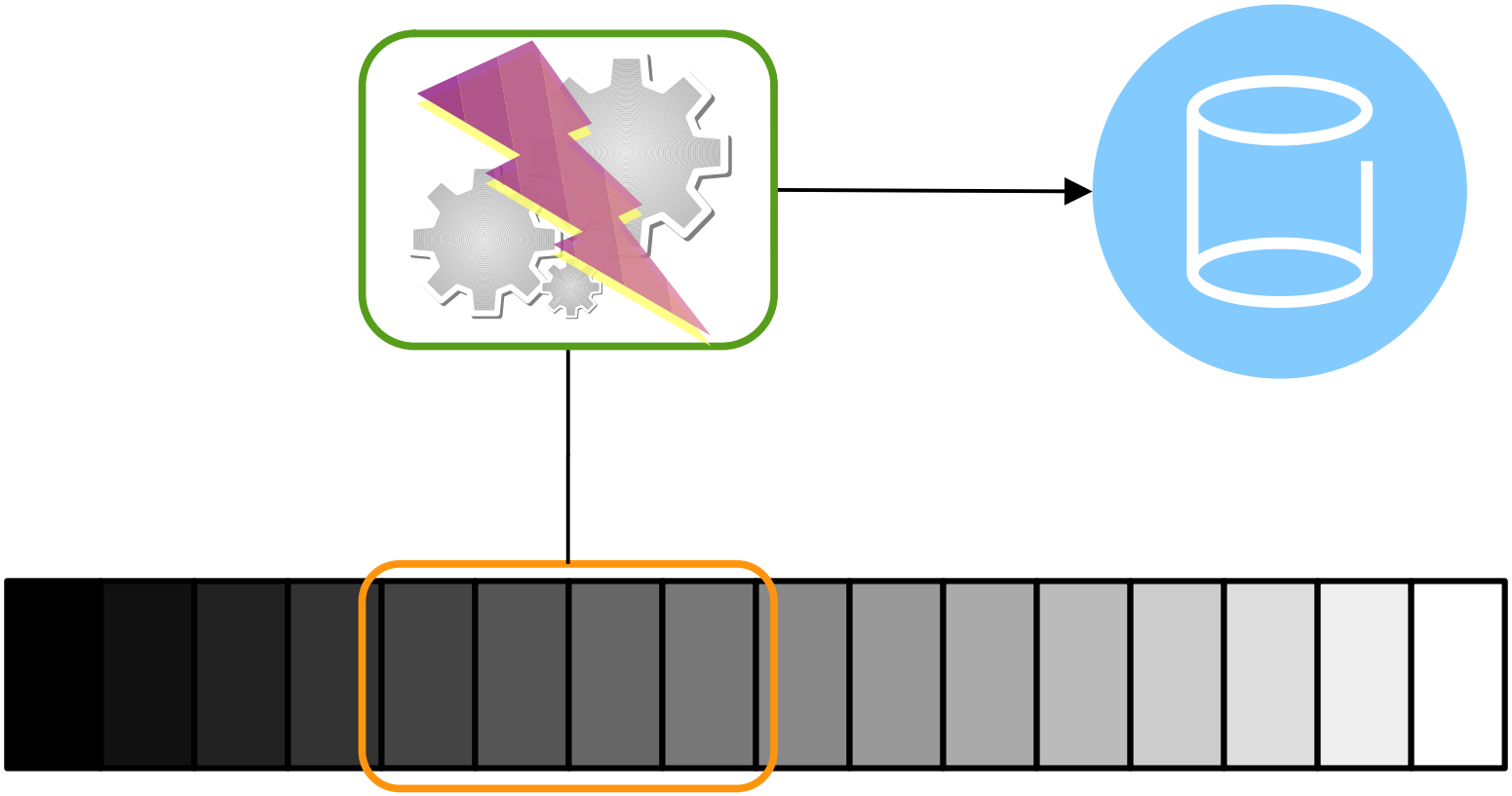

Inodes
• UPSERT
• LINK
• UNLINK
• DELETE
• XATTR

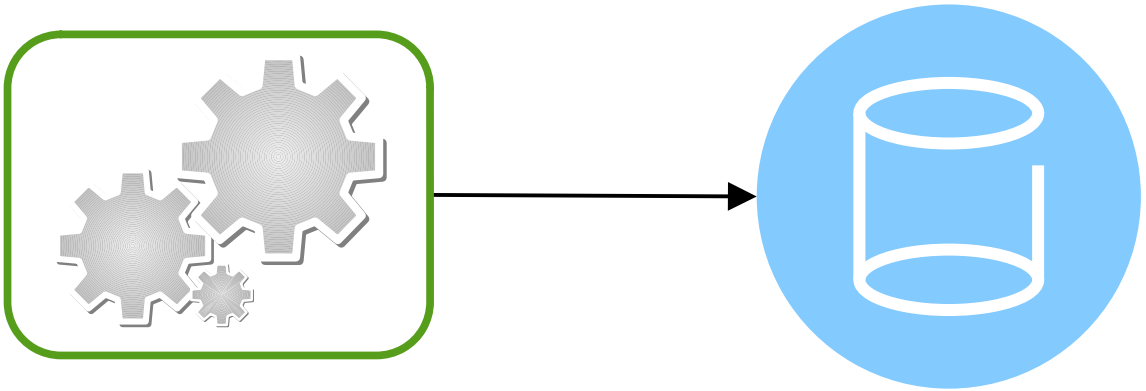

Inodes
• UPSERT
• LINK
• UNLINK
• DELETE
• XATTR
Namespace

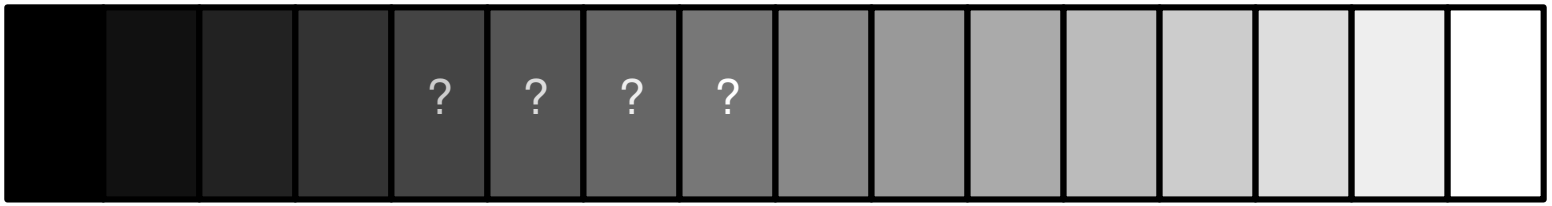
FSEvents are replayable

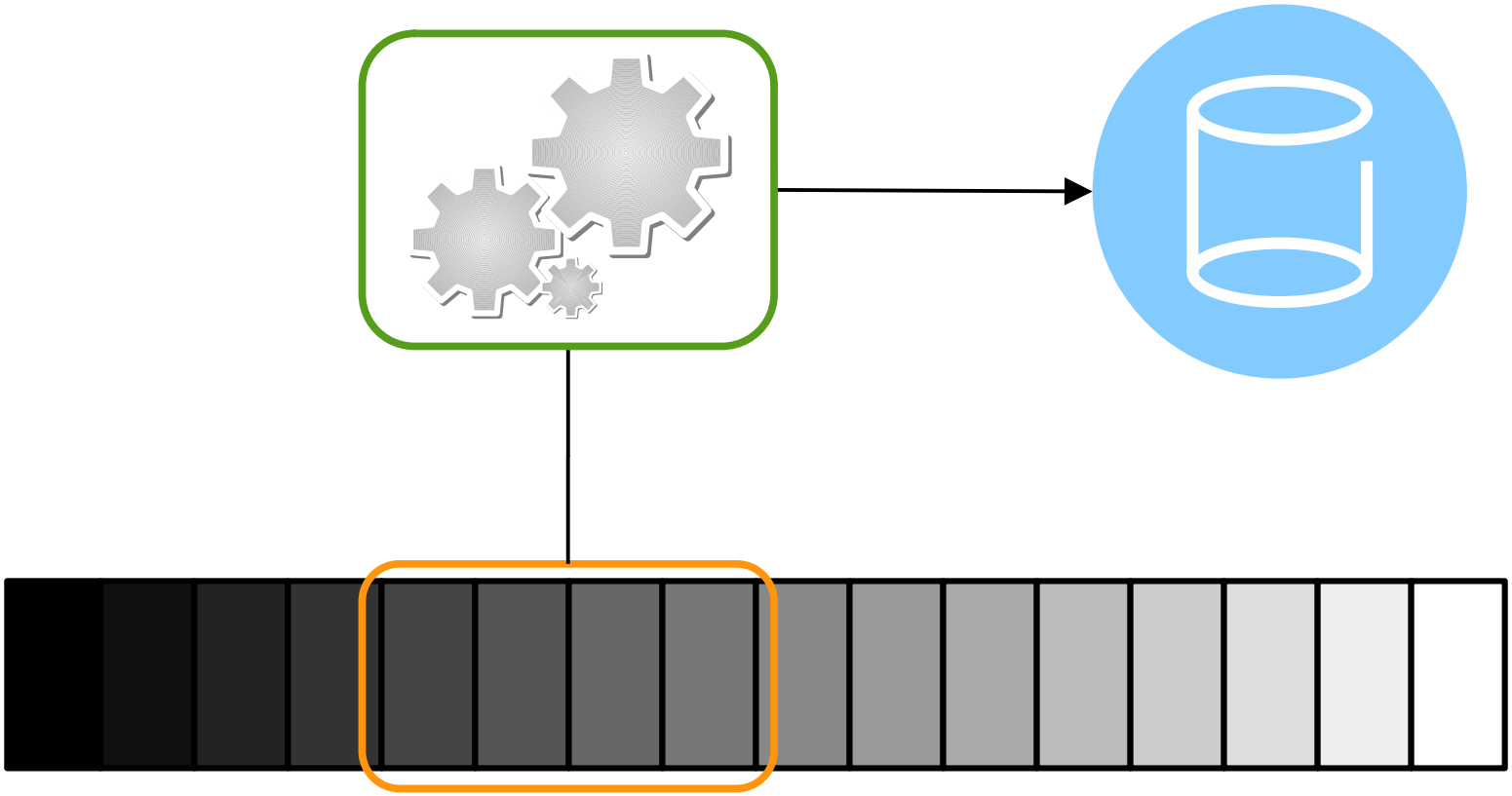


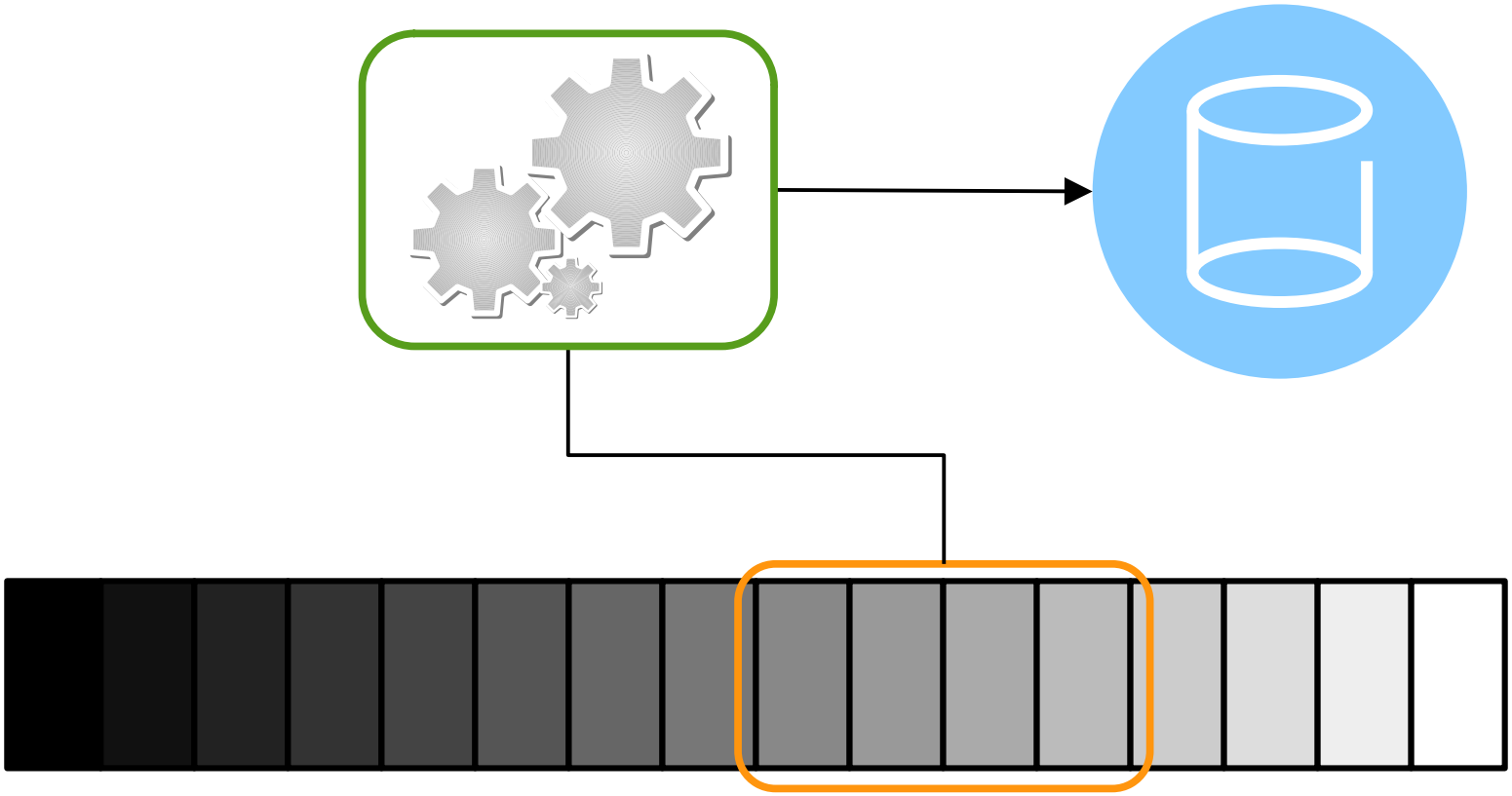


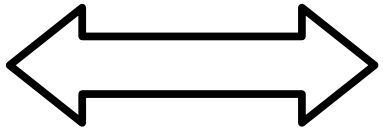


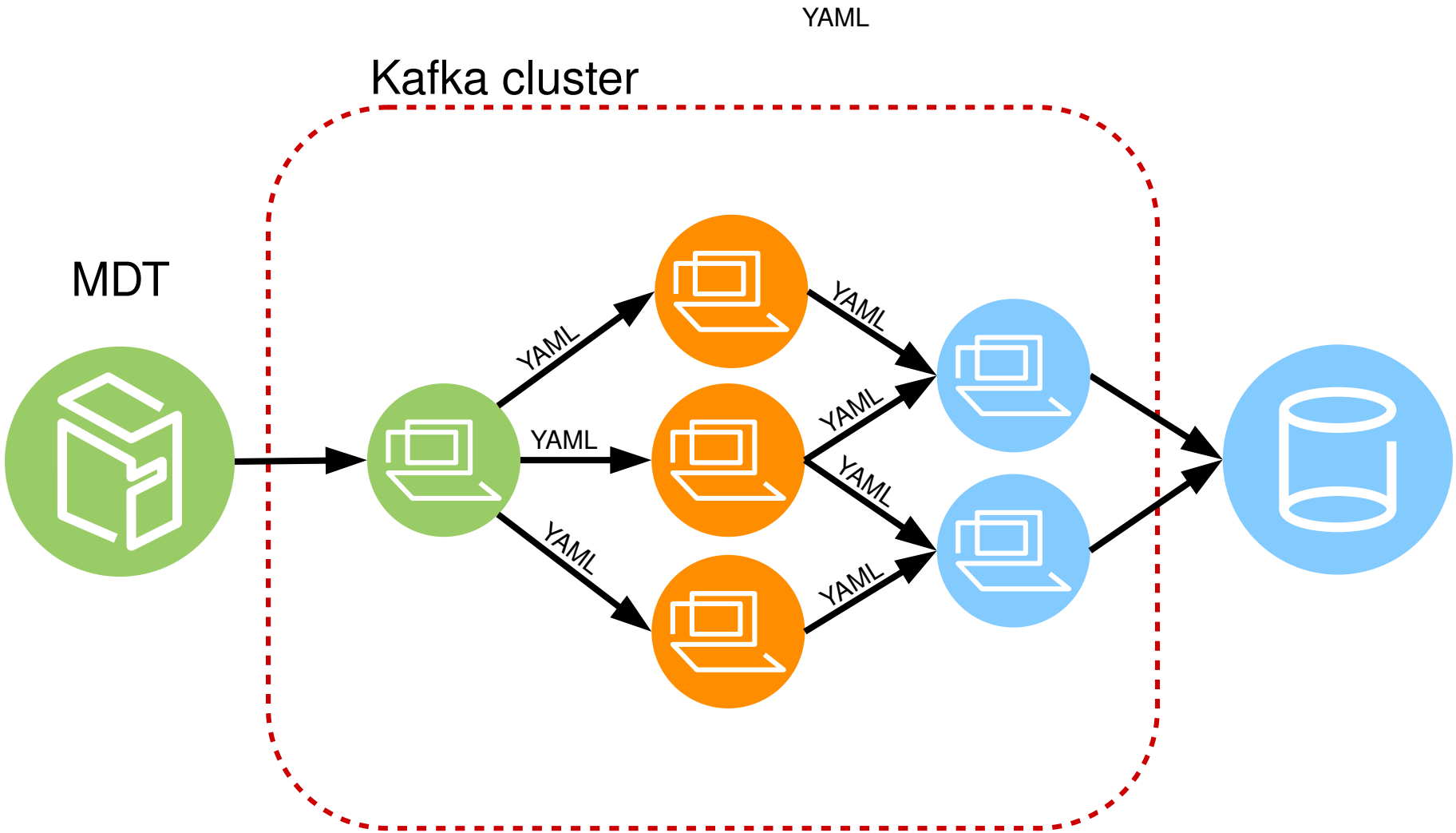








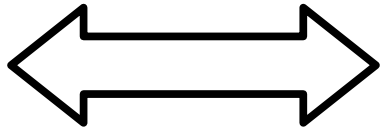
FSEvents  YAML


YAML
Kafka cluster
MDT
YAML
YAML
YAML
YAML
YAML
YAML
YAML

https://github.com/cea-hpc/miniyaml

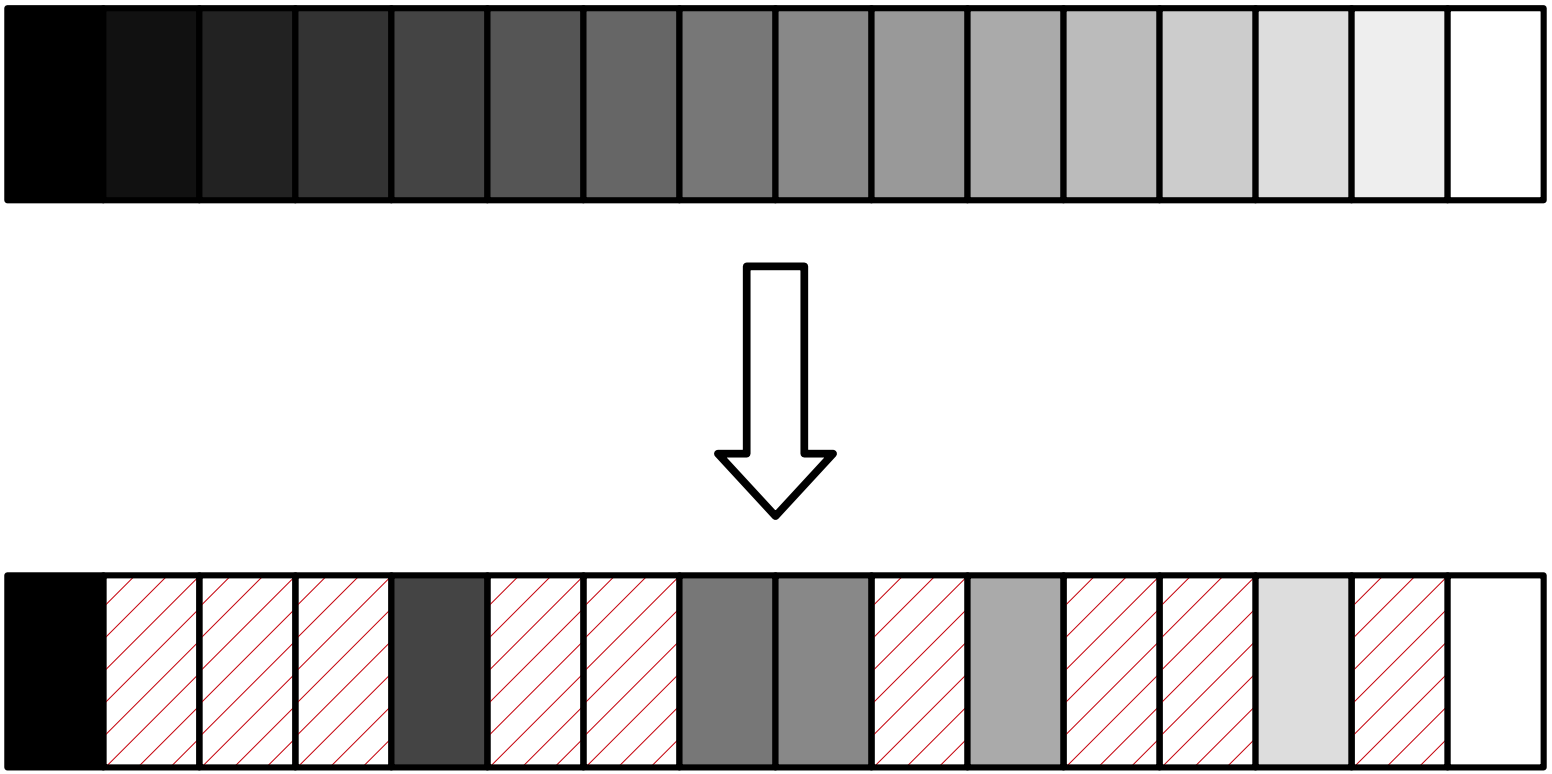
What’s next?

Lustre changelog support

Lustre changelog records  FSEvents

FSEvent deduplication mechanism

FSEvent deduplication mechanism



Python binding

Python binding

```python
from datetime import datetime, timedelta

from robinhood.backends import MongoBackend
from robinhood.filter import FilterField

ninety_days_ago = datetime.now() - timedelta(days=90)
filter_ = FilterField.STATX.ATIME >= ninety_days_ago.timestamp()

for fsentry in Backend("my-filesystem").filter(filter_):
    fsentry.unlink()
```

Demo time!

Thank you