
# Forensic Examination of Ceph

Florian Bausch
28.06.2023, Troopers 2023

ERNW
RESEARCH
pursuing knowledge.

# Agenda

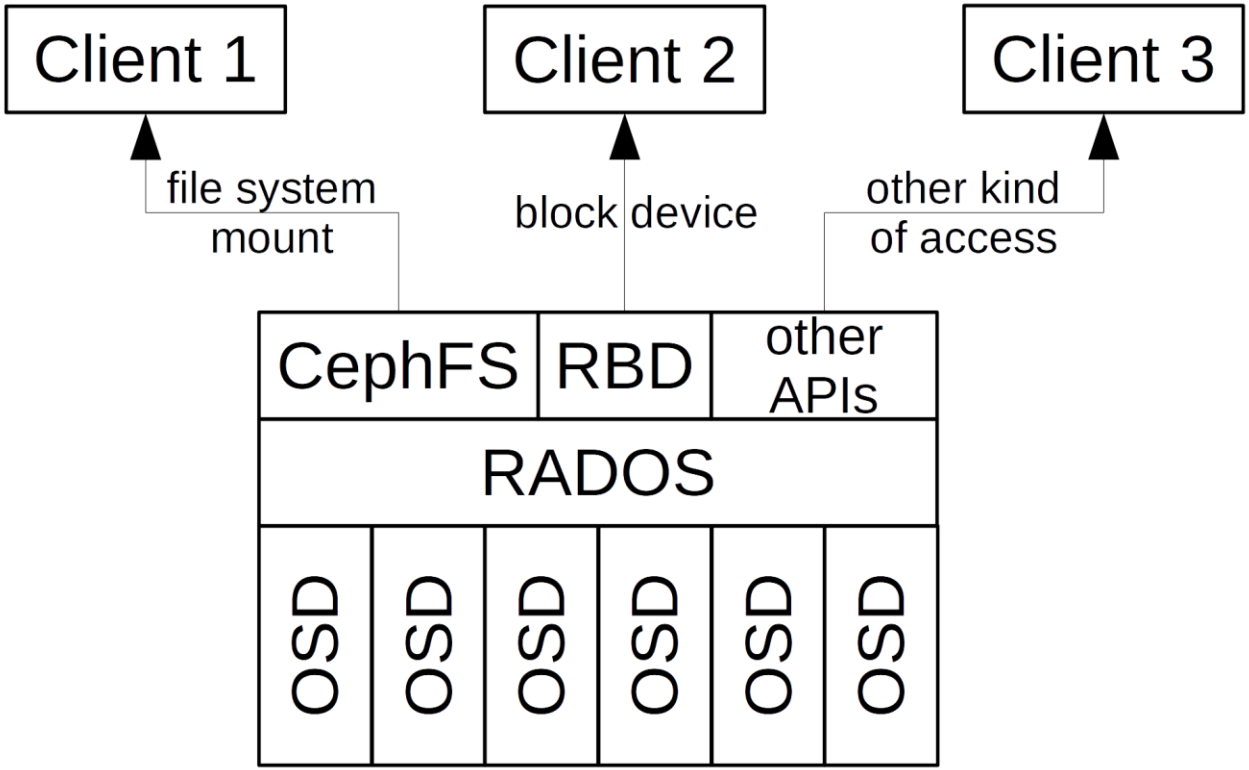
- Motivation and Background
- Forensic Examination of Ceph
- Implementation and Evaluation of *Vampyr*
- Conclusion and Outlook

ERNW RESEARCH
pursuing knowledge.

# Motivation and Background

- Master Thesis
  - 2018
  - FAU Erlangen / Hochschule Albstadt-Sigmaringen
- Software-defined storage with growing importance
  - Clusters of commodity hardware logically joined by software to a large storage with redundancy
  - For example: OpenStack

# Motivation and Background – Ceph RADOS



ERNW RESEARCH
pursuing knowledge.

# Motivation and Background

- Ceph released *BlueStore* storage format in 2017
  - Documentation of data structures
  - Categorization into Carrier categories
    - Filesystem category
    - Content Category
    - Metadata Category
    - Filename Category
    - Application Category
  - Implementation of a forensic software tool (Vampyr)

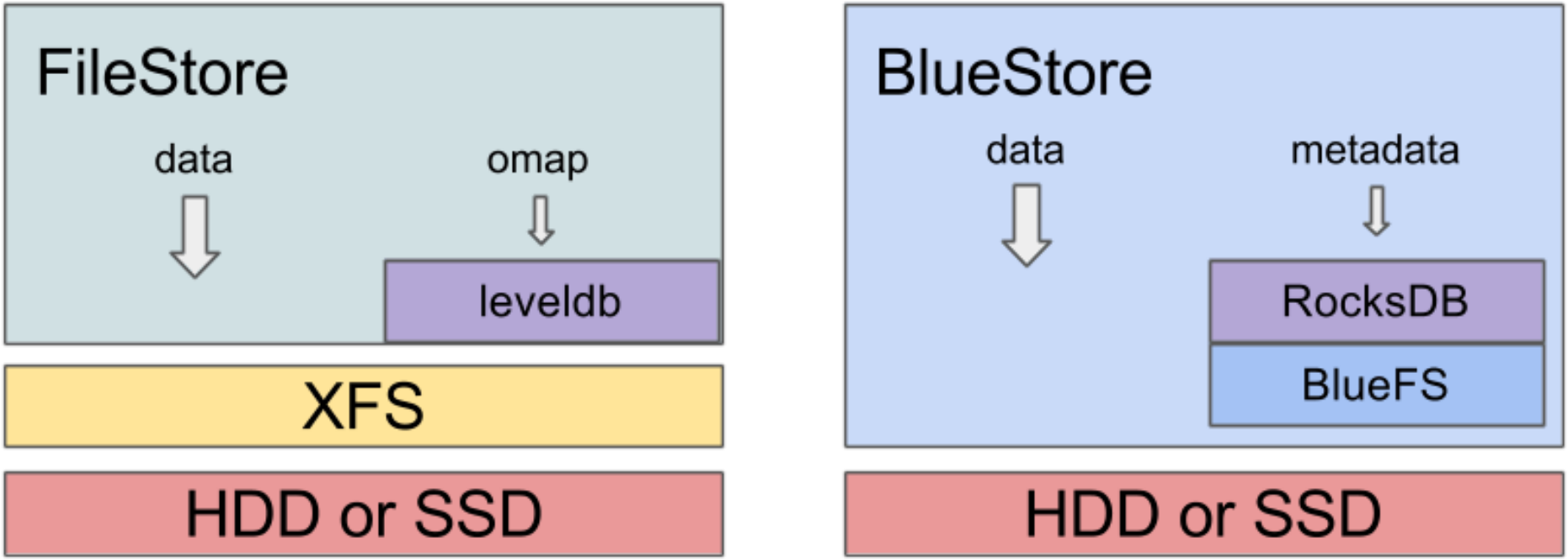
ERNW RESEARCH
pursuing knowledge.

# Motivation and Background

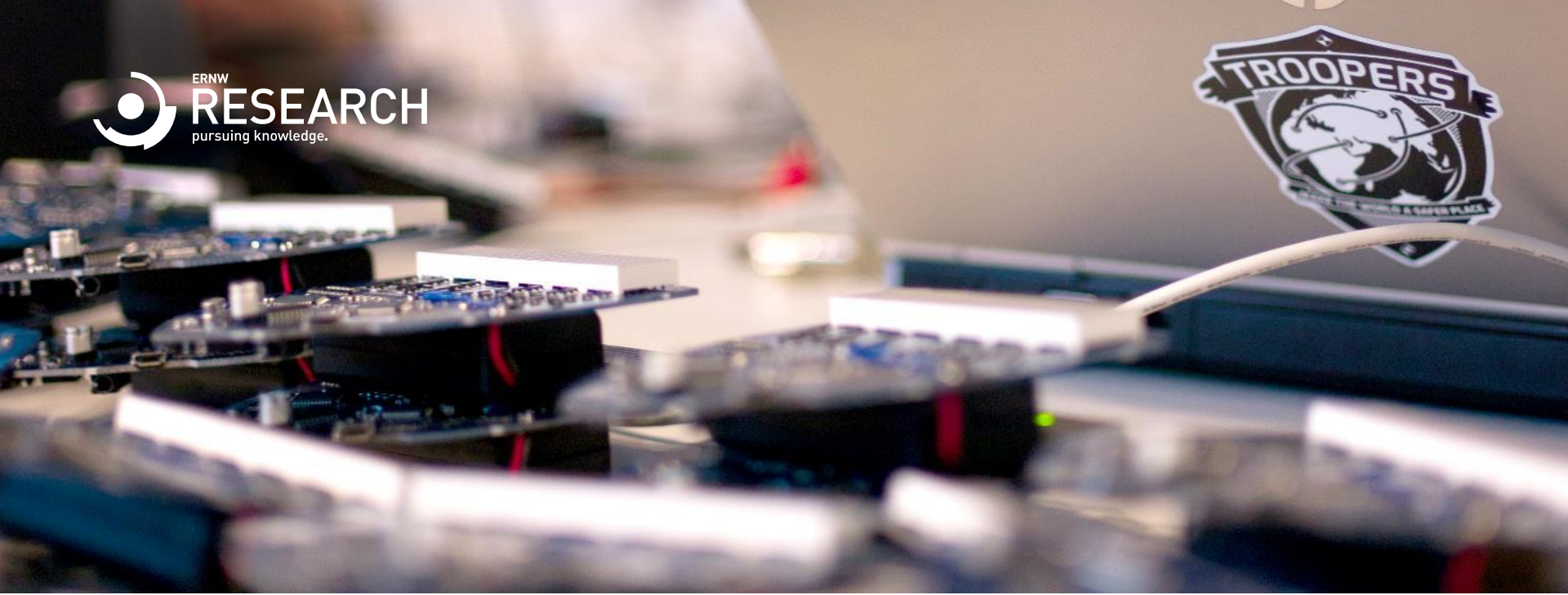
- Not every OSD (object storage device, i.e., HDD / SSD) stores every object of the RADOS.
- CRUSH algorithm determines OSDs of an object.
  - Deterministic
  - But seems random

ERNW RESEARCH pursuing knowledge.

# Motivation and Background - OSDs




https://i0.wp.com/ceph.com/wp-content/uploads/2017/08/filestore-vs-bluestore-2.png




# Forensic Examination of Ceph

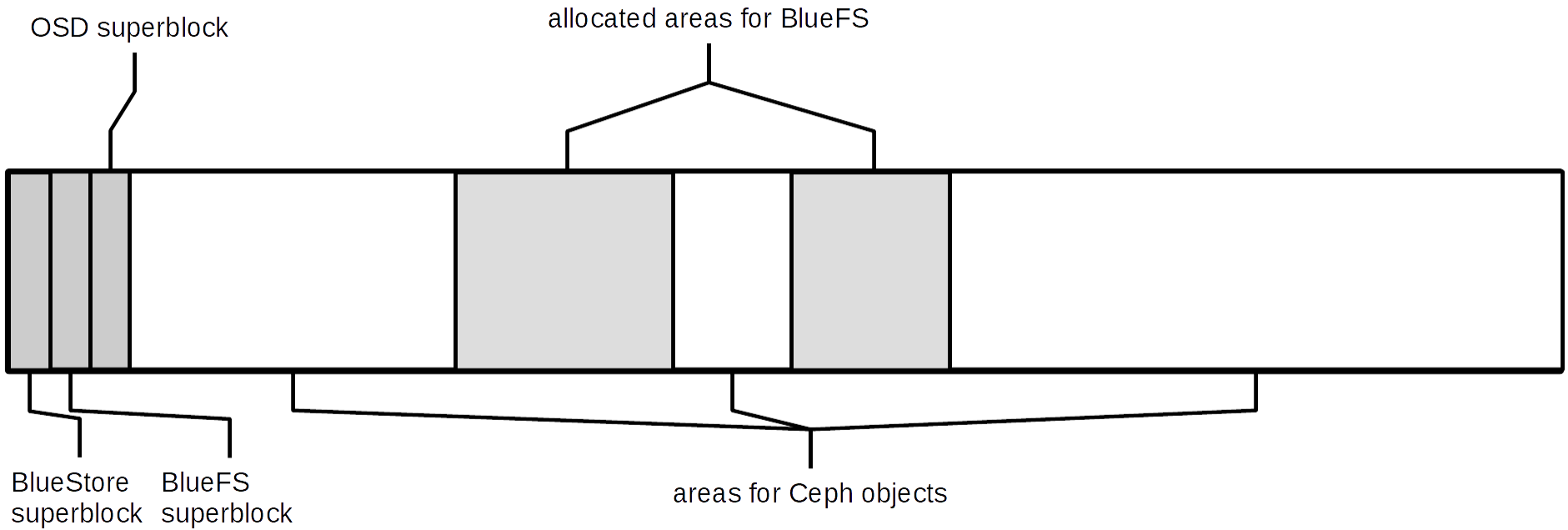
ERNW RESEARCH
pursuing knowledge.

# Forensic Examination

- BlueStore OSDs with RocksDB Key-Value Store
- Test cluster with three VMs
  - Defined states
- OSDs from clusters with SAP HANA installed
- Reading source code
- Reading hex dumps

ERNW RESEARCH pursuing knowledge.

# Content of OSDs



ERNW RESEARCH
pursuing knowledge.

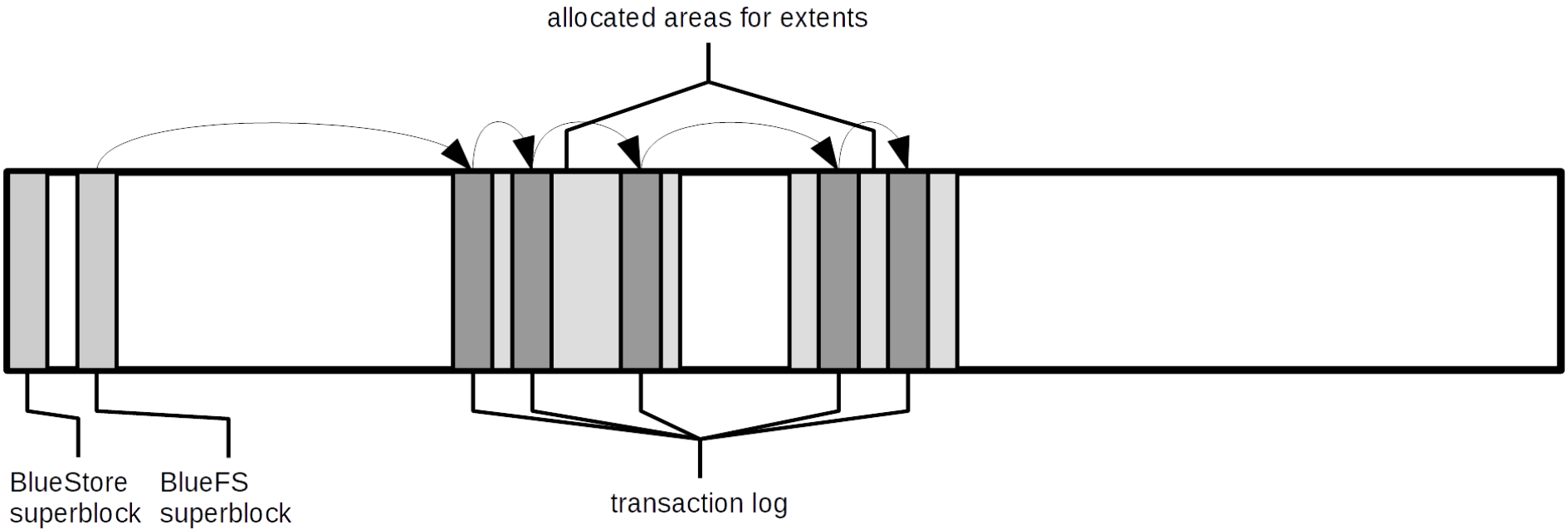
# Key-Value Store

- Stores metadata of objects
- Uses a RocksDB database
  - Located in BlueFS
- Restore / extract KV store
  - For metadata analysis
  - To find / restore objects on OSD


# BlueFS

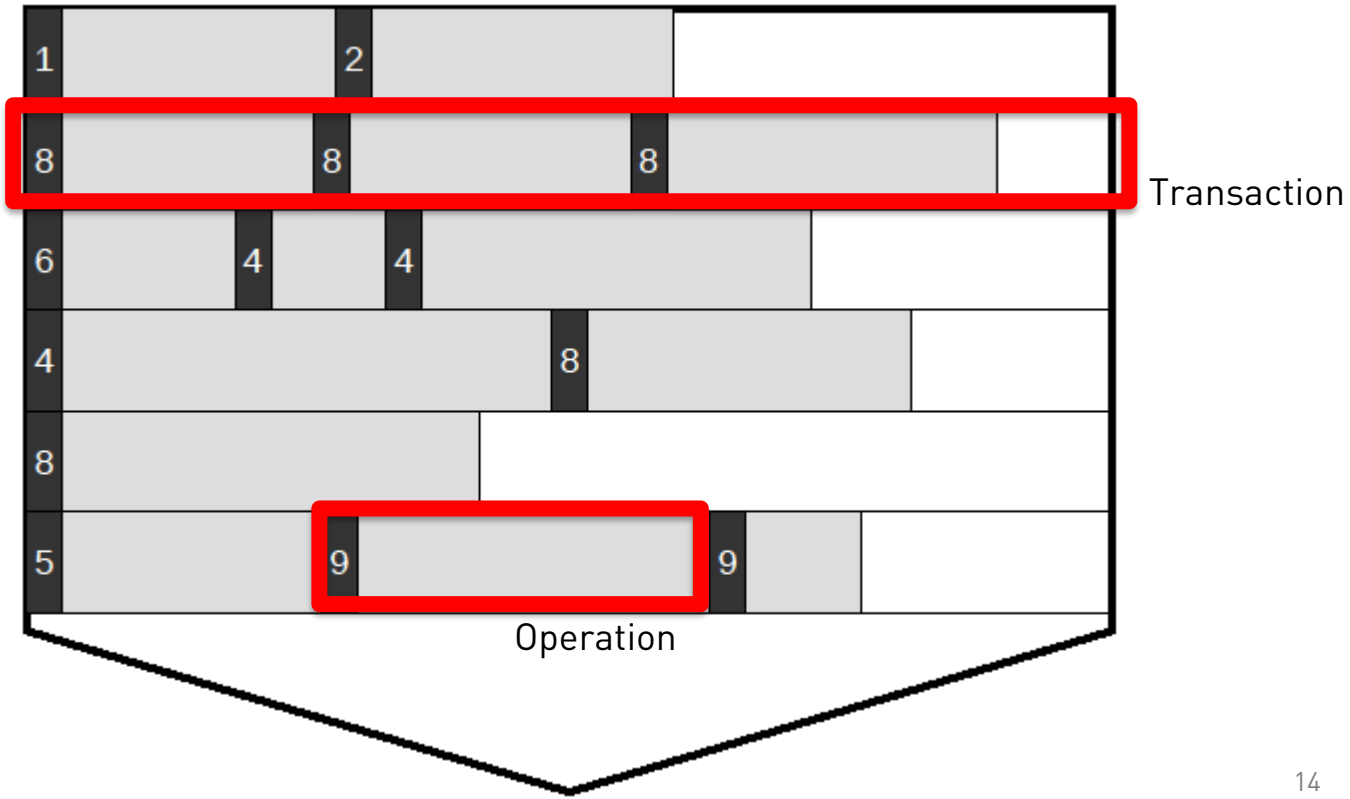
ERNW RESEARCH pursuing knowledge.

# BlueFS



ERNW RESEARCH pursuing knowledge.

# BlueFS Transaction Log



ERNW RESEARCH
pursuing knowledge.

# Get the Key-Value Store

1. Read BlueFS superblock
2. Jump to transaction log
3. Read and interpret all the transactions sequentially
4. Get list of files
5. Read all extents of files and write to analysis machine
6. Run the ldb command from RocksDB


# Object Metadata

ERNW RESEARCH pursuing knowledge.

# Key-Value Store

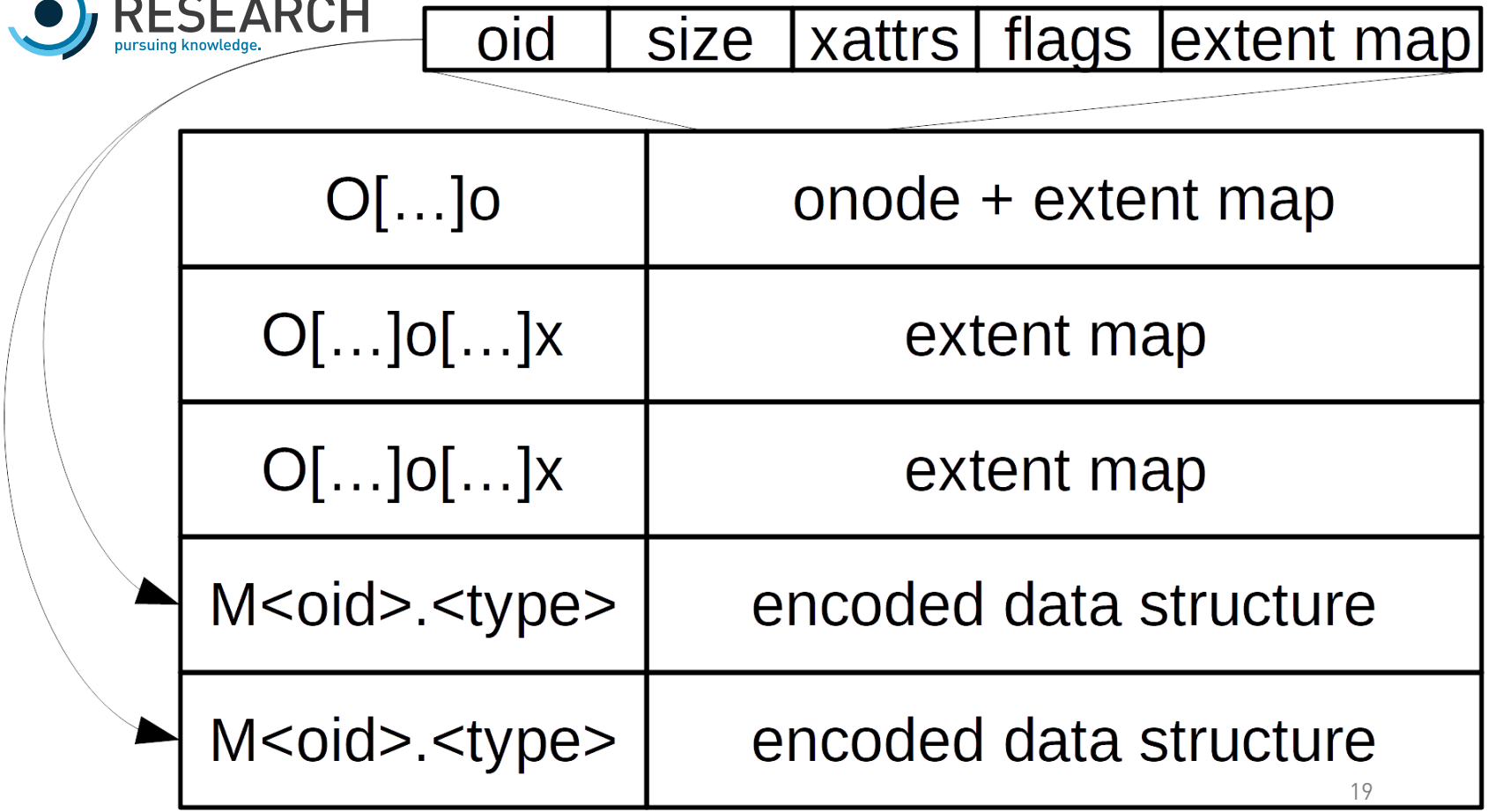
| Prefix | Description |
|---|---|
| S | OSD metadata (S = super) |
| T | OSD statistics (statfs data structure) |
| C | Collection information |
| O | Object metadata |
| M | Additional object metadata |
| P | Placement Group metadata |
| L | Deferred transactions |
| B | Bitmap metadata |
| b | Bitmaps |
| X | Shared blob information |

ERNW RESEARCH pursuing knowledge.

# Key-Value Store

| Prefix | Description |
|---|---|
| S | OSD metadata (S = super) |
| T | OSD statistics (statfs data structure) |
| C | Collection information |
| O | Object metadata |
| M | Additional object metadata |
| P | Placement Group metadata |
| L | Deferred transactions |
| B | Bitmap metadata |
| b | Bitmaps |
| X | Shared blob information |

ERNW RESEARCH
pursuing knowledge.




# RBD and CephFS

ERNW RESEARCH
pursuing knowledge.

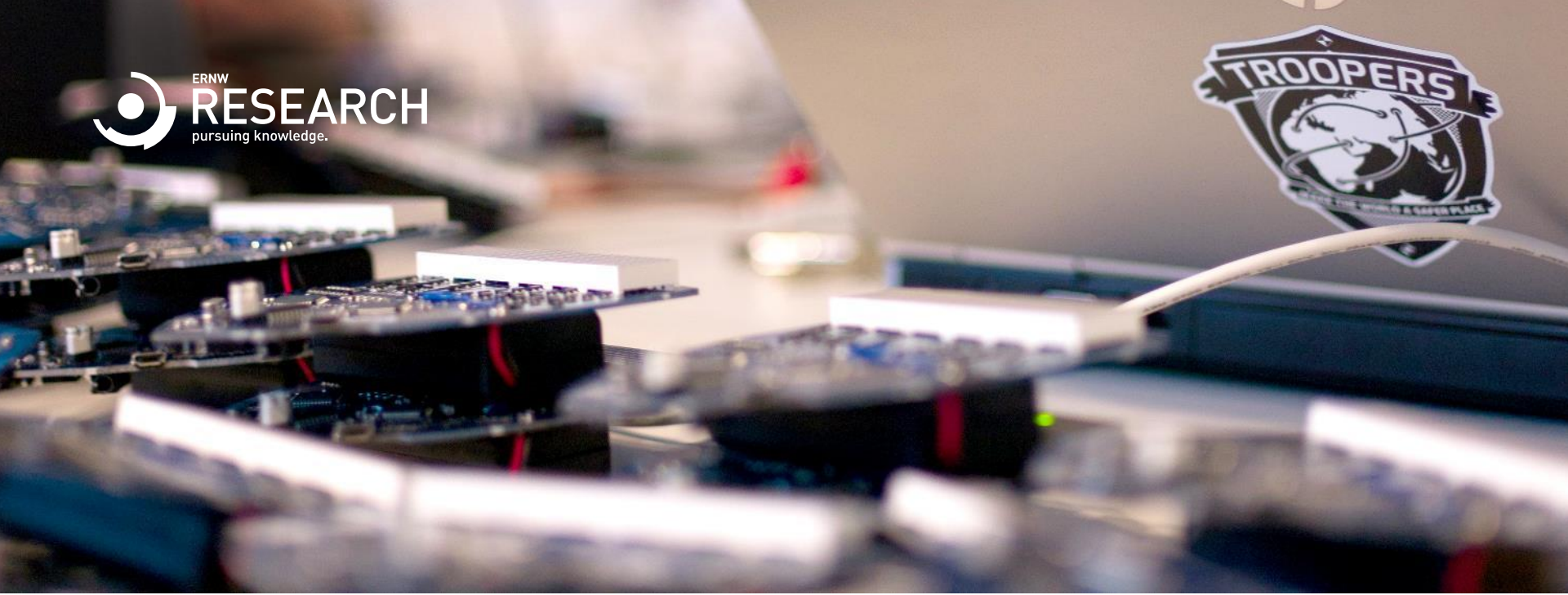
# RBD – RADOS Block Device

- Each RBD has a unique ID
- Object name rbd_data.<ID>.<Block>
  - Content of block device
  - Lazy allocation
- Object name rbd_header.<ID>
  - Creation time
  - Features
  - Size
  - Object prefix
  - Lock

ERNW RESEARCH
pursuing knowledge.

# CephFS

- Not BlueFS!
- Directories / files → one or more objects
  - Names of objects: <Inode>.<Chunk>
- 0-row in KV store:
  - Complete path as extended attribute _parent
- M-row in KV store:
  - -: directory metadata (mtime, num. of files)
  - <filename>_head: metadata of file (inode, permission)
- Application category:
  - Object prefixes 100., 200., 300., 400., 500., 600.




# Implementation and Evaluation

ERNW RESEARCH
pursuing knowledge.

# Implementation and Evaluation

- Vampyrotheutis Infernalis
- https://github.com/fbausch/vampyr
  - Python 3, no Ceph libraries used
- Vampyr can
  - Show data of the filesystem category
  - Reconstruct/extract BlueFS and load KV store
  - Decode and carve for osdmaps (topology info)
  - Decode object metadata / extract object contents

Citron / CC-BY-SA 3.0

ERNW RESEARCH
pursuing knowledge.

# Implementation and Evaluation

- Objects that belong together can be joined
  - rbd_data objects: reconstruct (parts of) RBD
  - CephFS objects: reconstruct (parts of) files
- CephFS file names and directory names can be reconstructed

ERNW RESEARCH
pursuing knowledge.

# Implementation and Evaluation

- Vampyr can
  - Extract slack spaces
  - Combine objects of several OSDs
  - Determine unallocated areas of OSDs
  - Extract unallocated areas
  - Determine which unallocated 512kB blocks are actually empty

ERNW RESEARCH
pursuing knowledge.

# Implementation and Evaluation

- Training data
  - Test setup
  - Database system (SAP HANA)

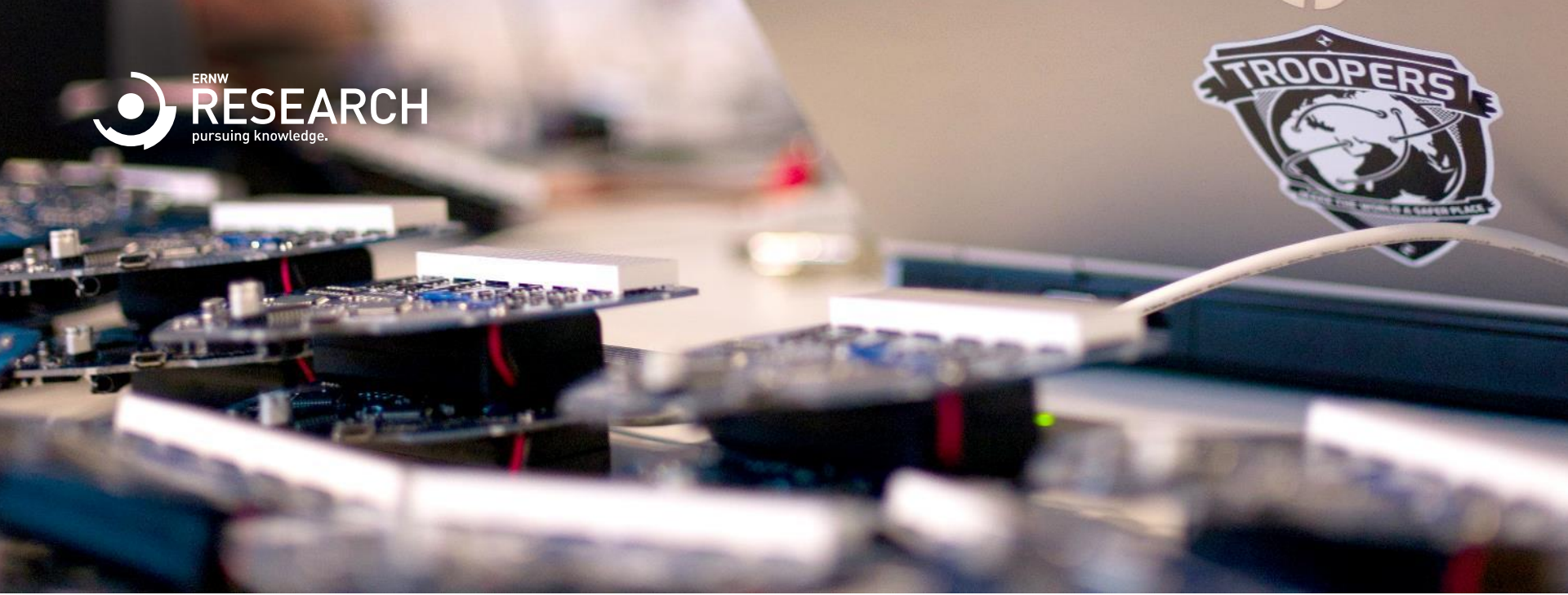
```
---------------------------------
BlueStore Superblock Information:
---------------------------------
Start at: 0x0
End at:   0x126
---------------------------------
bluestore block device
1406c680-5d99-4406-a1e5-bf7aeca64b36

OSD UUID: 1406c680-5d99-4406-a1e5-bf7aeca64b36
OSD length: 0x2e8d931000 B = ~ 186 GiB
Last used at: 2018-03-07 14:00:34.148527627
Description:  main
Metadata information:
- bluefs: 1
- ceph_fsid: 585de5f3-bdc9-3160-8b5a-b72620af43e4
- kv_backend: rocksdb
- magic: ceph osd volume v026
- mkfs_done: yes
- ready: ready
- whoami: 12
CRC32 checksum: 0xdf2c722b
---------------------------------
Volume slack starts at offset 0x2e93e31000 of image file
```

```
fbausch@vm901:~$ ./vampyr.py --ldb rocksdb/build/tools/ldb --image /datengrab/be
rn/bern.sda.img --offset 0x6500000 --lsobjects --objfilter "rbd.*"
-----------------------
Object List:
-----------------------
Prefix        -> Object
-----------------------
rbd_data.2fee2ae8944a -> 0000000000010015
rbd_data.30572ae8944a -> 0000000000010029, 0000000000010031, 0000000000010035, 0
00000000001fff0
rbd_data.305f2ae8944a -> 0000000000000027, 00000000000000bf, 00000000000000d8, 0
0000000000000100, 000000000000019b, 00000000000001a2, 00000000000001e6, 000000000
0000200, 0000000000000284, 00000000000002b7, 00000000000002ea, 0000000000000321,
 0000000000000042b
rbd_data.30e52ae8944a -> 000000000000000e, 0000000000000051, 0000000000000098, 0
0000000000000fe, 0000000000000166, 0000000000000189, 00000000000001b2, 000000000
00001d6, 0000000000000274, 00000000000002a3, 00000000000002ea, 000000000000033f,
 0000000000003d3, 0000000000003e4, 0000000000003f6, 0000000000000405, 0000000
00000044e, 000000000000049e, 00000000000004c4, 0000000000000524, 000000000000052
f, 0000000000000540, 0000000000002031
-----------------------
```

```
Key: shard: 0x-1, ns: , key: 10000000010.00000000, name: 10000000010.00000000, p
oolid: 0x4, snap: 0xfffffffffffffffe, gen: 0xffffffffffffffff
Value:
oid: 12338, object_size: 1082325, shards:
         _: size: 0x1083d5, mtime: 2018-08-06 16:49:11.501443970, soid: key: , o
id: 10000000010.00000000, nspace: , pool: 0x4
   snapset: snapid: 0x1, snaps: Number of elements: 0 (0x0), clones: Number of e
lements: 0 (0x0)
   _layout: objectsize: 0x400000, poolid: 0x4, pool_ns:
   _parent: inode: 0x10000000010 -> ancestors: ino: 0x10000000002, dname: kap-ag
ulhas_20236318375_o.jpg, ver: 0x5e->ino: 0x10000000001, dname: southafrica, ver:
 0x61->ino: 0x10000000000, dname: images, ver: 0x3b->ino: 0x1, dname: test_files
, ver: 0x36, pool: 0x4
Filename: kap-agulhas_20236318375_o.jpg
Fullpath: <CephFSroot>/test_files/images/southafrica/kap-agulhas_20236318375_o.j
pg
Own inode: 0x10000000010
Inodes in path: 0x1/0x10000000000/0x10000000001/0x10000000002

Logical extents:
Logical offset: 0x0, length: 0x80000, Physical extents: 0x3400000-0x80000
Logical offset: 0x80000, length: 0x80000, Physical extents: 0x3480000-0x80000
Logical offset: 0x100000, length: 0x83d5, Physical extents: 0x3500000-0x10000
```




# Conclusion and Outlook

ERNW RESEARCH
pursuing knowledge.

# Conclusion and Outlook

- Documentation of data structures
- Categorization
- Vampyr

ERNW RESEARCH
pursuing knowledge.

# Conclusion and Outlook

- Future work
  - SeaStore
    - Successor of BlueStore
    - Under development
  - WAL + DB devices
  - CephFS journal
  - RGW, librados
  - Compression
  - Checksums

ERNW RESEARCH
pursuing knowledge.

# Thank you for you attention


fbausch@ernw.de

@WeAreTroopers

www.ernw.de

www.insinuator.net