DB

block.db and block.wal

```
The DB stores BlueStore's internal metadata and the WAL is BlueStore's
internal journal or write-ahead log.
It is recommended to use a fast SSD or NVRAM for better performance.
```

```
Important
Since Ceph has to write all data to the journal (or WAL+DB) before it can
ACK writes,
having this metadata and OSD performance in balance is really important!
```

For hosts with multiple HDDs (multiple OSDs), it is possible to use one SSD for all OSDS DB/WAL (one partition per OSD).

NOTE: The recommended scenario for mixed setup for one host is to use

- multiple HDDS (one OSD per HDD)
- one fast SSD/NVMe drive for DB/WAL (i.e. 4GB or 30GB per 2TB HDD only needed).

Proxmox UI and CLI expects only whole device as DB device, not partition!. It will not destroy existing drive. It expect LVM volume with free space and it will create new LVM volume for DB/WAK.

Ceph native CLI can work with partition specified as DB (it also works with whole drive or LVM).

MORE INFO:

- https://docs.ceph.com/en/latest/rados/configuration/bluestore-config-ref/#block-and-block-db
- https://docs.ceph.com/en/latest/rados/operations/add-or-rm-osds/#adding-removing-osds
- https://www.reddit.com/r/ceph/comments/jnyxgm/how_do_you_create_multiple_osds_per_disk_with/

DB/WAL sizes

- If there is <1GB of fast storage, the best is to use it as WAL only (without DB).
- if a DB device is specified but an explicit WAL device is not, the WAL will be implicitly colocated with the DB on the faster device.

DB size:

- (still true for Octopus 15.2.6) DB should be 30GB. And this doesn't depend on the size of the data partition.
    - all block.db sizes except **4, 30, 286 GB** are pointless,
        - see: About block.db sizing
        - Leveled Compaction
- should have as large as possible logical volumes
- for RGW (Rados Gateway) workloads: min 4% of block device size
- for RBD (Rados Block Device) workloads: 1-2% is enough (2% from 2TB is 40GB)

- according to `ceph daemon osd.0 perf dump | jq .bluefs` 80GB was reserved on HDD for DB, where 1.6-2.4GB is used.