Ceph BlueStore Write Analyse

笔记本:	bluestore素材		
创建时间:	2020/9/2 11:49	**更新时间:**	2020/9/2 11:56

问题:
(1) lextent和pextent之间是怎么建立映射关系的? blob设计意图?
Onode是逻辑对象，对象空间的管理本质是对逻辑空间的管理，即是lextent。在

> onode是对象，对象是逻辑的，对象的空间管理本质上是对逻辑空间的管理，也就是Extent。Extent从何而来，就是从某个blob上来。在Extent看来，它不关心blob底层是逻辑空间还是物理空间，它只知道自己是属于哪个blob的一部分就够了。对象的逻辑空间管理，到blob这儿就终止了。 blob就像一个容器，把底层物理空间也就是bluestore_pextent_t（包括那些bitmap allocator，freelist manager等模块和算法）这些复杂的东西，全都给包起来了，对上层屏蔽掉了。这样一来，逻辑空间的管理和物理空间的管理，就完全解耦了。

> 快照空间共享，只是blob这一层中间层引入之后的一个红利而已，还有其他很多好处。

(2) blob的unused位图的用途，small write初次申请blob时才会标记unused，并且只有最开头的区间是标记为未使用的?
(3)
(1) 为啥blob split? 什么场景split?

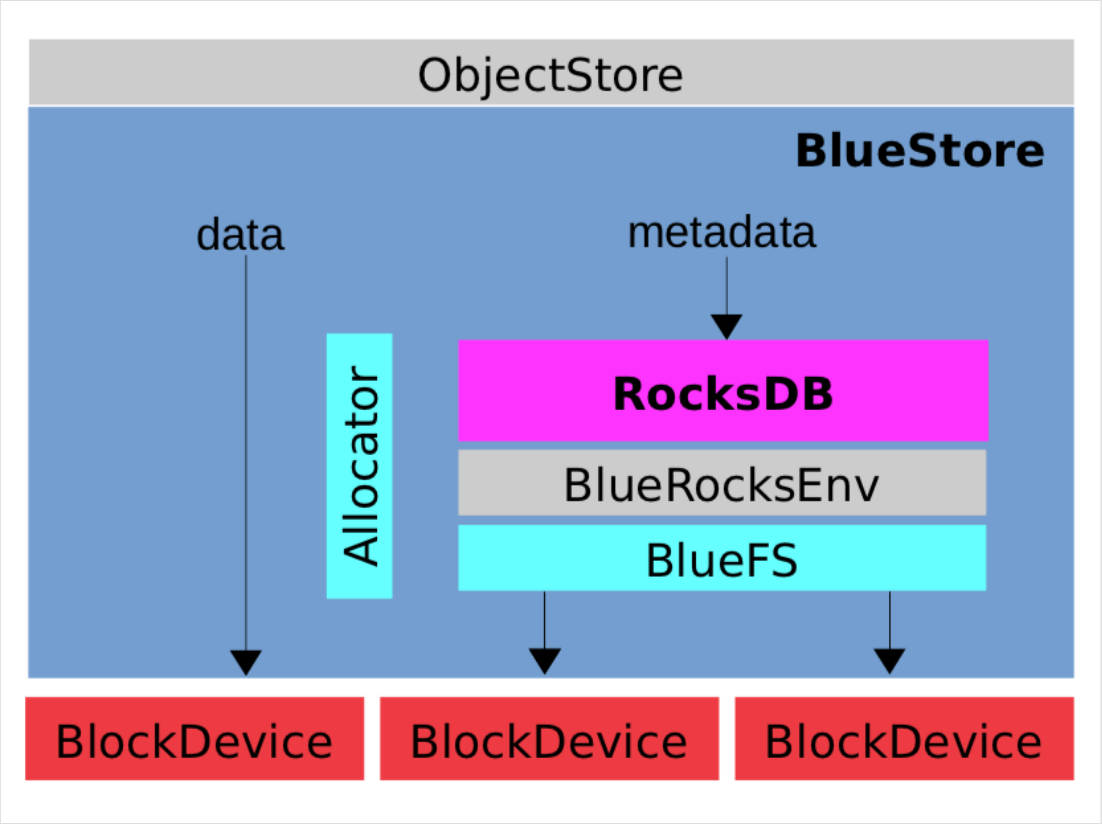
Ceph BlueStore Write Analyse

一、概述

Ceph从Luminous开始，默认使用了全新的存储引擎BlueStore，来替代之前提供存储服务的FileStore，与FileStore不一样的是，BlueStore直接管理裸盘，分配数据块来存储RADOS里的objects。Bluestore比较复杂，学习BlueStore的关键就是查看其实现的“ object → block device ”映射，所以本文先重点分析下该部分：一个Object写到Bluestore的处理过程。

二、Bluestore整体架构

下面借用*Sage Weil*的图描述下BlueStore的整体架构:



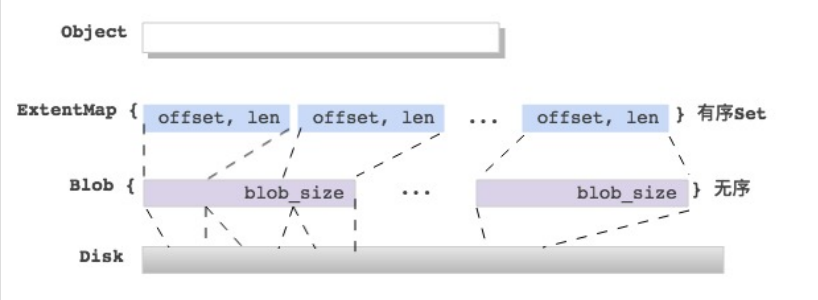
里面的几个关键组件介绍：

1. RocksDB：存储元数据信息
2. BlueRocksEnv：提供RocksDB的访问接口
3. BlueFS：实现BlueRocksEnv里的访问接口
4. Allocator：磁盘分配器

针对BlueStore的整体架构不做展开，大家只需对其几大组件有些了解即可，本文继续介绍Object写操作到底层BlockDevice的过程。

三、BlueStore中Object到底层Device的映射关系

这里先从整体上给出在BlueStore中，一个Object的数据映射到底层BlockDevice的实现，如下图：



首先详细介绍下与上诉映射关系相关的数据结构。

四、Bluestore相关配置参数

```
// 压缩模式
bluestore_d_mode = none // 参数取值范围modes: "none", "passive", "aggressive", "force"
                                // 也可以对pool单独设置, ceph osd pool set <poolname> compression_mode <modes>
// compression blob size
bluestore_compression_min_blob_size = 0
bluestore_compression_min_blob_size_hdd = 128_K
bluestore_compression_min_blob_size_ssd = 8_K
bluestore_compression_max_blob_size = 0
bluestore_compression_max_blob_size_hdd = 512_K
bluestore_compression_max_blob_size_ssd = 64_K
// 另外也可以对pool单独设置, ceph osd pool set <poolname> compression_max_blob_size/compression_min_blob_size <size>

// 压缩算法
```

```
bluestore_compression_algorithm = snappy // 参数取值范围algorithms：""，"snappy"，"zlib"，"zstd"，"lz4"
                                        // 也可以对pool单独设置，ceph osd pool set <poolname> compression_algorithm <algorithms>

// 压缩比率
bluestore_compression_required_ratio = 0.875 // 参数取值范围[0.0-1.0]
                                             // 也可以对pool单独设置，ceph osd pool set <poolname> compression_required_ratio <float[0.0-
1.0]>




//校验和
bluestore_csum_type =crc32c // 参数取值范围csum_types：none、xxhash32、xxhash64、crc32c、crc32c_16、crc32c_8，可动态修改参数
                            // 也可以对pool单独设置，ceph osd pool set <poolname> csum_type <csum_types>，如果设置则以pool的为准,否则取默认参
数值


bluestore_min_alloc_size_hdd = 64_K  // 最小分配单元区分大写or小写，
bluestore_min_alloc_size_ssd = 16_K
bluestore_max_blob_size_hdd =512_K
bluestore_max_blob_size_ssd =64_K
bluestore_prefer_deferred_size_hdd = 32_K
bluestore_prefer_deferred_size_ssd = 0
bluestore_deferred_batch_ops_hdd = 64
bluestore_deferred_batch_ops_ssd = 16
```

五、Bluestore相关数据结构

•Collection：PG在内存的数据结构。
•bluestore_cnode_t：PG在磁盘的数据结构。
•Onode：对象在内存的数据结构。
•bluestore_onode_t：对象在磁盘的数据结构。
•Extent：一段对象逻辑空间(lextent)。
•extent_map_t：一个对象包含多段逻辑空间。
•bluestore_pextent_t：一段连续磁盘物理空间。
•bluestore_blob_t一片不一定连续的磁盘物理空间，包含多段pextent。
•Blob：包含一个bluestore_blob_t、引用计数、共享blob等信息。解耦对象的逻辑空间和物理空间的管理

BlueStore里与Object 数据映射相关的数据结构罗列如下：

5.1 Onode

任何RADOS里的一个Object都对应Bluestore里的一个Onode（内存结构），定义如下：

```
struct Onode {
  Collection *c; // 对应的Collection，对应PG
  ghobject_t oid; // Object信息
  /// key under PREFIX_OBJ where we are stored
  mempool::bluestore_cache_other::string key;

  bluestore_onode_t onode; // Object存到kv DB的元数据信息
  ExtentMap extent_map; // 映射lextents到blobs
};
```

通过Onode里的ExtentMap来查询Object数据到底层的映射。

5.2 ExtentMap

ExtentMap是Extent的set集合，是有序的，定义如下：

```
/// a sharded extent map, mapping offsets to lextents to blobs
struct ExtentMap {
    Onode *onode; // 指向Onode指针
    extent_map_t extent_map; // Extents到Blobs的map
    blob_map_t spanning_blob_map; // 跨越shards的blobs，spanning id即是Blob::id
    struct Shard {
        bluestore_onode_t::shard_info *shard_info = nullptr;
        unsigned extents = 0; ///< count extents in this shard
        bool loaded = false; ///< true if shard is loaded
        bool dirty = false; ///< true if shard is dirty and needs reencoding
    };
    mempool::bluestore_cache_other::vector<Shard> shards; ///< shards
};
```

ExtentMap还提供了分片功能，防止在文件碎片化严重，ExtentMap很大时，影响写RocksDB的性能。
ExtentMap会随着写入数据的变化而变化；
ExtentMap的连续小段会合并为大；
覆盖写也会导致ExtentMap分配新的Blob；

5.3 Extent

Extent是实现object的数据映射的关键数据结构，定义如下：

```
/// a logical extent, pointing to (some portion of) a blob
```

```
struct Extent : public ExtentBase {
    uint32_t logical_offset = 0; // 对应Object的逻辑偏移
    uint32_t blob_offset = 0;    // 对应Blob上的逻辑偏移(0~max_blob_size)
    uint32_t length = 0;         // 数据段长度
    BlobRef blob;                // 指向对应Blob的指针
};
typedef boost::intrusive::set<Extent> extent_map_t;
```

每个Extent都会映射到下一层的Blob上，Extent会依据 block_size 对齐，没写的地方填充全零。
Extent中的length值，最小：block_size，最大：max_blob_size。

```
ostream& operator<<(ostream& out, const BlueStore::Extent& e)
{
  return out << std::hex << "0x" << e.logical_offset << "~" << e.length
             << ": 0x" << e.blob_offset << "~" << e.length << std::dec
             << " " << *e.blob;
}
```

5.4 Blob

Blob是Bluestore里引入的处理块设备数据映射的中间层，定义如下：

```
struct Blob {
    int16_t id = -1;  ///< id, for spanning blobs only, >= 0
    SharedBlobRef shared_blob; // 共享的blob状态
    mutable bluestore_blob_t blob; // blob的持久化元数据
    bluestore_blob_use_tracker_t used_in_blob; /// refs from this shard.  ephemeral if id<0, persisted if spanning.
};


struct bluestore_blob_t {
private:
    PExtentVector extents; // 对应磁盘上的一组数据段
    uint32_t logical_length = 0; // blob的原始数据逻辑长度,大于等于extents的长度总和
    uint32_t compressed_length = 0; // 压缩的数据长度

public:
    uint32_t flags = 0;
    typedef uint16_t unused_t;
    unused_t unused = 0;    ///< 16位的bitmap位图记录了pextent的使用情况, chunk_size = logical_length / 16
    // csum内存大小是: len = csum_value_size * (logical_length / csum_chunk_size)
    bufferptr csum_data;    ///< opaque vector of csum data
};


typedef boost::intrusive_ptr<Blob> BlobRef;
typedef mempool::bluestore_cache_other::map<int,BlobRef> blob_map_t;

// 以au_size为单元trace每个单元已使用的空间的情况，记录已使用空间的大小
struct bluestore_blob_use_tracker_t {
  uint32_t au_size; // Allocation (=tracking) unit size, == 0 if uninitialized
  uint32_t num_au;  // Amount of allocation units tracked, == 0 if single unit or the whole blob is tracked
  union {
    uint32_t* bytes_per_au; // 数组指针，num_au大于1的情况使用，数组元素记录了每个au_size为单元已使用空间大小
    uint32_t total_bytes; // 当num_au只有1个的情况下使用，bytes_per_au指针为空，表示已使用的空间大小
  };
}

SHA-1: 2df9aa8e793d2450771e7ca509baa53bc92f9202
* os/bluestore: make blob_t unused helpers use logical length
These were taking min_alloc_size, but this can change
across mounts; better to use the logical blob length
instead (that's what we want anyway!).

SHA-1: 9ee134e2950d9770427886c6edab4c3fb59b8d37
* os/bluestore: replace bluestore_blob_t::unused from interval to bitmap

SHA-1: b159dd770c184d3b51ba4e93c99e61be6f306eef
* os/bluestore: move add_unused call after csum initialization as it should depend  on chunk size.
```

Blob对应一组磁盘空间PExtentVector，它就是bluestore_pextent_t 的一个数组，指向从Disk中分配的物理空间。
Blob里可能对应一个或多个lextent；
Blob里可能对应一个磁盘pextent，也可能对应多个pextent；
Blob里的pextent个数最多为：max_blob_size / min_alloc_size,
hdd：512_K / 64_K = 8
ssd：64_K / 16_K = 4
Blob里的多个pextent映射的Blob offset可能不连续，中间有空洞；

```
ostream& operator<<(ostream& out, const BlueStore::Blob& b)
{
  out << "Blob(" << &b;
  if (b.is_spanning()) {
    out << " spanning " << b.id;
  }
  out << " " << b.get_blob() << " " << b.get_blob_use_tracker();
  if (b.shared_blob) {
    out << " " << *b.shared_blob;
  } else {
    out << " (shared_blob=NULL)";
  }
  out << ")";
  return out;
}
ostream& operator<<(ostream& out, const bluestore_blob_t& o)
{
  out << "blob(" << o.get_extents();
  if (o.is_compressed()) {
    out << " clen 0x" << std::hex
    << o.get_logical_length()
    << " -> 0x"
    << o.get_compressed_payload_length()
    << std::dec;
  }
  if (o.flags) {
    out << " " << o.get_flags_string();
  }
  if (o.has_csum()) {
    out << " " << Checksummer::get_csum_type_string(o.csum_type)
    << "/0x" << std::hex << (1ull << o.csum_chunk_order) << std::dec;
  }
  if (o.has_unused())
    out << " unused=0x" << std::hex << o.unused << std::dec;
  out << ")";
  return out;
}

ostream& operator<<(ostream& out, const bluestore_blob_use_tracker_t& m)
{
  out << "use_tracker(" << std::hex;
  if (!m.num_au) {
    out << "0x" << m.au_size
        << " "
        << "0x" << m.total_bytes;
  } else {
    out << "0x" << m.num_au
        << "*0x" << m.au_size
    << " 0x[";
    for (size_t i = 0; i < m.num_au; ++i) {
      if (i != 0)
    out << ",";
      out << m.bytes_per_au[i];
    }
    out << "]";
  }
  out << std::dec << ")";
  return out;
}
```

5.5 SharedBlob

SharedBlob是

```
/// in-memory shared blob state (incl cached buffers)
struct SharedBlob {
  std::atomic_int nref = {0}; ///< reference count
  bool loaded = false;


  CollectionRef coll;
  union {
    uint64_t sbid_unloaded;              ///< sbid if persistent isn't loaded
    bluestore_shared_blob_t *persistent; ///< persistent part of the shared blob if any
  };
  BufferSpace bc;             ///< buffer cache
};
typedef boost::intrusive_ptr<SharedBlob> SharedBlobRef;
```

```
/// shared blob state
struct bluestore_shared_blob_t {
  uint64_t sbid;                        ///> shared blob id
  bluestore_extent_ref_map_t ref_map;  ///< shared blob extents
}

/// extent_map: a map of reference counted extents
struct bluestore_extent_ref_map_t {
  struct record_t {
    uint32_t length; // pextent的长度
    uint32_t refs;
    record_t(uint32_t l=0, uint32_t r=0) : length(l), refs(r) {}
    DENC(bluestore_extent_ref_map_t::record_t, v, p) {
      denc_varint_lowz(v.length, p);
      denc_varint(v.refs, p);
    }
  };

  typedef mempool::bluestore_cache_other::map<uint64_t,record_t> map_t;
  map_t ref_map; // poffset-><plength,refs>
}

/// a lookup table of SharedBlobs
struct SharedBlobSet {
  std::mutex lock;   ///< protect lookup, insertion, removal

  // we use a bare pointer because we don't want to affect the ref count
  mempool::bluestore_cache_other::unordered_map<uint64_t,SharedBlob*> sb_map;
}
```

```
ostream& operator<<(ostream& out, const BlueStore::SharedBlob& sb)
{
  out << "SharedBlob(" << &sb;

  if (sb.loaded) {
    out << " loaded " << *sb.persistent;
  } else {
    out << " sbid 0x" << std::hex << sb.sbid_unloaded << std::dec;
  }
  return out << ")";
}
ostream& operator<<(ostream& out, const bluestore_shared_blob_t& sb)
{
  out << "(sbid 0x" << std::hex << sb.sbid << std::dec;
  out << " " << sb.ref_map << ")";
  return out;
}
ostream& operator<<(ostream& out, const bluestore_extent_ref_map_t& m)
{
  out << "ref_map(";
  for (auto p = m.ref_map.begin(); p != m.ref_map.end(); ++p) {
    if (p != m.ref_map.begin())
      out << ",";
    out << std::hex << "0x" << p->first << "~" << p->second.length << std::dec
    << "=" << p->second.refs;
  }
  out << ")";
  return out;
}
```

5.5 write_item

每次aio有一个write_item 定义如下：

```
struct write_item {
    uint64_t logical_offset;  ///< write logical offset, 对象内的便宜
    BlobRef b;               // blob指针
    uint64_t blob_length;    // blob的长度, min_alloc_size对齐
    uint64_t b_off;          // after padding, block_size对齐
    bufferlist bl;
    uint64_t b_off0;        ///< original offset in a blob prior to padding
    uint64_t length0;       ///< original data length prior to padding

    bool mark_unused;
    bool new_blob; ///< whether new blob was created

    bool compressed = false;
    bufferlist compressed_bl;
    size_t compressed_len = 0;
};
```

举例：

写[62k, 4k]，此时会拆成两个small write：生成新的lextent1[0xf800~800: 0xf800~800]和lextent2[0x10000~800: 0x10000~800]

write_item：logical_offset=62k，blob_length=64k，b_off=60k, b_off0=62k，length0=2k，logical_length=64k

Blob(0x55973c3a7ea0 blob([0xd20000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x10000 0x800) SharedBlob(0x55973c6cfe30 sbid 0x0))

lextent1：0xf800~800: 0xf800~800 // 新增

unused : chuck_size = logical_length / 16 = 64k / 16 = 4k, 0x7fff表明0~60k都没写过

write_item：logical_offset=64k，blob_length=64k，b_off=0, b_off0=0，length0=2k，logical_length=128k

Blob(0x55973c6cff10 blob([!~10000,0xd30000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,800]) SharedBlob(0x55973c6cff80 sbid 0x0))
lextent2：0x10000~800: 0x10000~800 // 新增
unused : chuck_size = logical_length / 16 = 128k / 16 = 8k, 0xff表明0~64k都没写过

写[130k, 4k]，此时有一个small write：生成新的lextent[0x20800~1000: 0x20800~1000]
write_item：logical_offset=130k，blob_length=64k，b_off=0, b_off0=2k，length0=4k，logical_length=192k
Blob(0x55973c728850 blob([!~20000,0xd80000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55973c7288c0 sbid 0x0))
lextent3：0x20800~1000: 0x20800~1000 // 新增
unused : chuck_size = logical_length / 16 = 192k / 16 = 12k, 0x3ff表明0~120k都没写过

写[192k,64k]，此时有一个big write：生成新的lextent[0x30000~10000: 0x30000~10000]
write_item：logical_offset=192k，blob_length=64k，b_off=0, b_off0=0，length0=64k，logical_length=256k
Blob(0x55973c729030 blob([!~30000,0xd90000~10000] csum crc32c/0x1000) use_tracker(0x4*0x10000 0x[0,0,0,10000]) SharedBlob(0x55973c7290a0 sbid 0x0))
lextent4：0x30000~10000: 0x30000~10000 // 新增

写[256k, 128k]，此时有一个big write(复用前面的blob并扩充)：生成新的lextent[0x40000~20000: 0x40000~20000]，跟前面的lextent在同一个blob中并且是连续的则可以进行合并
write_item：logical_offset=256k，blob_length=128k，b_off=256k, b_off0=256k，length0=128k，logical_length=384k
Blob(0x55973c729030 blob([!~30000,0xd90000~10000,0xda0000~20000] csum crc32c/0x1000) use_tracker(0x6*0x10000 0x[0,0,0,10000,10000,10000]) SharedBlob(0x55973c7290a0 sbid 0x0))
lextent5：0x40000~20000: 0x40000~20000 // 新增，可合并
合并后的lextent4：0x30000~30000: 0x30000~30000

写[578k, 4k]，此时有一个small write：生成新的lextent[0x90800~1000: 0x10800~1000]
write_item：logical_offset=578k，blob_length=64k，b_off=0, b_off0=2k，length0=4k，logical_length=128k
Blob(0x55973c728d90 blob([!~10000,0xdc0000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,1000]) SharedBlob(0x55973c6c7c70 sbid 0x0))
lextent6：0x90800~1000: 0x10800~1000 // 新增
unused : chuck_size = logical_length / 16 = 128k / 16 = 8k, 0xff表明512k~576k都没写过

写[562k, 4k]，此时有一个small write(复用前面的blob)：
write_item：logical_offset=562k，blob_length=64k，b_off=48k, b_off0=50k，length0=4k，logical_length=128k
Blob(0x55973c728d90 blob([0xdd0000~10000,0xdc0000~10000] csum+has_unused crc32c/0x1000 unused=0xbf) use_tracker(0x2*0x10000 0x[1000,1000]) SharedBlob(0x55973c6c7c70 sbid 0x0))
lextent7：0x8c800~1000: 0xc800~1000 // 新增

写[6k，4k]，此时有一个small write：
Blob(0x55973c3a7ea0 blob([0xd20000~10000] csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55973c6cfe30 sbid 0x0))
lextent8：0x1800~1000: 0x1800~1000 // 新增

总结：
在small write流程中，write_item的blob_length即是min_alloc_size，b_off0即是min_alloc_size内的偏移(取余)，b_off是b_off0对block_size向下对齐的大小，是对blob内的逻辑偏移。在big write流程中，write_item的blob_length即是数据的长度(min_alloc_size整数倍)；如果是新的blob，那么b_off和b_off0都为0；如果是复用blob，那么b_off和b_off0是对blob内的逻辑偏移。

5.6 bluestore_pextent_t

bluestore_pextent_t是管理物理磁盘上的数据段的，定义如下：

```
struct bluestore_pextent_t : public bluestore_interval_t<uint64_t, uint32_t> {
    ...
};

template <typename OFFS_TYPE, typename LEN_TYPE>
struct bluestore_interval_t
{
  static const uint64_t INVALID_OFFSET = ~0ull;

  OFFS_TYPE offset = 0; // 磁盘上的物理偏移
  LEN_TYPE length = 0; // 数据段的长度
  ...
};


typedef mempool::bluestore_cache_other::vector<bluestore_pextent_t> PExtentVector;
```

AllocExtent的 length值，最小：min_alloc_size，最大：max_blob_size

5.7 Sequencer

```
//
struct DeferredBatch : public AioContext {
    OpSequencer *osr; // pg内保序
    struct deferred_io {
      bufferlist bl;    ///< data
      uint64_t seq;     ///< deferred transaction seq
    };
    map<uint64_t,deferred_io> iomap; ///< map of ios in this batch
    deferred_queue_t txcs;           ///< txcs in this batch
    IOContext ioc;                   ///< our aios
    /// bytes of pending io for each deferred seq (may be 0)
```

```
    map<uint64_t,int> seq_bytes;
}

// db中记录需要写的op和需要释放的物理地址空间
struct bluestore_deferred_transaction_t {
  uint64_t seq = 0;
  list<bluestore_deferred_op_t> ops;
  interval_set<uint64_t> released;  ///< allocations to release after tx
}

// deferred write的op
struct bluestore_deferred_op_t {
  typedef enum {
    OP_WRITE = 1,
  } type_t;
  __u8 op = 0;

  PExtentVector extents;  // 需要写的物理地址区间
  bufferlist data;
}


small write --> bluestore_deferred_op_t --> TransContext::bluestore_deferred_transaction_t(txc里可能有两个deferred op, 作为一个事务整体持久化到db) --> DeferredBatch::deferred_io(一个deferred op对应一个) --> aio_t(对于物理地址连续的deferred io可以进行合并)
```

5.7 Sequencer

```
struct Sequencer {
    string name;
    spg_t shard_hint;
    Sequencer_implRef p; /// 实际是指向不同Store类型实现的子类型实例
};
在PG实例创建时根据spg_t(即是pgid)时实例化。


struct Sequencer_impl : public RefCountedObject {
    CephContext* cct;
};
class OpSequencer : public Sequencer_impl {
    q_list_t q;  ///< transactions
    Sequencer *parent; ///指向PG层的Sequencer
    BlueStore *store;
    uint64_t last_seq = 0;
};
Store层处理第一个事务时进行实例化，里面包含了一个transactions的链表，事务的处理必须按照链表的先后顺序进行。


class PG : public DoutPrefixProvider {
    // for ordering writes
    ceph::shared_ptr<ObjectStore::Sequencer> osr;
};
每个pg都对应有有一个osr，用于PG内操作的保序。在调用ObjectStore的queue_transactions()接口时以参数的形式传递给ObjectStore层。
```

5.8 TransContext

```
struct TransContext : public AioContext {
    state_t state = STATE_PREPARE;
    KeyValueDB::Transaction t;
    IOContext ioc;
};
```

事务上下文，包含了事务状态，回调函数，操作的onode集合，分配/释放的空间信息等等。每次执行事务时会将事务相关的操作封装到该上下文结构中，事务处理完成后调用_txc_finish()释放该结构内存。

六、BlueStore接口分析

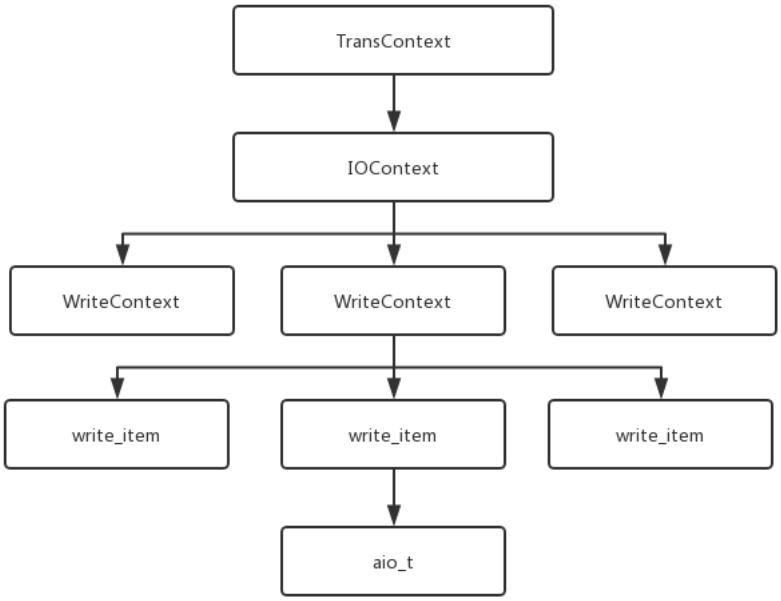
6.1 queue_transactions

queue_transactions是Store层处理事务的接口，对上层提供对某个PG的事务的写操作。

```
BlueStore::queue_transactions(Sequencer *posr,vector<Transaction>& tls,TrackedOpRef op,ThreadPool::TPHandle *handle)
    (1) 根据osr创建事务上下文TransContext，并将其加入osr的链表中。整个事务的处理都是围绕txc
    TransContext *txc = _txc_create(osr);

    (2) 遍历tls逐个事务进行处理
    _txc_add_transaction(TransContext *txc, Transaction *t)
        // 获取操作对象的Onode实例
        BlueStore::Collection::get_onode(const ghobject_t& oid,bool create)

    (3) 封装db事务
    _txc_write_nodes(txc, txc->t)

    (4) 封装wal db事务
    if (txc->deferred_txn)
        txc->t->set(PREFIX_DEFERRED, key, bl);

    (5) 对象物理空间在内存中fm中申请/释放
    _txc_finalize_kv(txc, txc->t)
```

```
    (6) 流控
    throttle_bytes.get(txc->cost)

    (7) 对封装好的txc开始执行状态机
    _txc_state_proc(txc)
```

一句话总结：queue_transactions()接口可能包含了多个Transaction，每个Transaction中可能包含多个write(对应多个WriteContext)，而一个WriteContext中根据off和len是否对min_alloc_size对齐可能包含一个或多个write_item，而一个write_item对应一个aio请求。所有的aio请求都封装到TransContext的IOContext中，所以在向内核提交aio请求时需要对IOContext层面的aio请求进行去重处理。另外对于不同的txc中如果存在覆盖写或者写的范围overlapped的aio，这必须通过PG的osr进行严格保序，排在后面的overlapped aio需要等待前面txc的aio完成后才能提交到内核。



6.2 Transaction::OP_TOUCH

touch对象首先会申请一个bluestore层唯一 的nid，由BlueStore::nid_last递增获得，作为对象的元数据需要持久化。同时将对象标记为已存在。另外BlueStore::nid_max也是需要持久化的，OSD重启后赋值给BlueStore::nid_last。

```
BlueStore::_touch(TransContext *txc,CollectionRef& c,OnodeRef &o)
    BlueStore::_assign_nid(TransContext *txc, OnodeRef o)
```

6.3 Transaction::OP_WRITE

BlueStore里的写数据入口是BlueStore::_do_write()，它会根据 min_alloc_size 来切分 [offset, length] 的写，然后分别依据 small write 和 big write 来处理。将构造好的write_item挂到WriteContext中，如下：

```
// 按照min_alloc_size大小切分，把写数据映射到不同的块上
[offset, length]
|==p1==|=======p2=======|==p3==|
|----------------|----------------|----------------|
| min_alloc_size | min_alloc_size | min_alloc_size |
|----------------|----------------|----------------|
small write: p1, p3
big write: p2
BlueStore::_do_write()
|-- BlueStore::_do_write_data()  // 将构造好的write_item挂到WriteContext中
| // 依据`min_alloc_size`把写切分为`small/big`写
| | -- BlueStore::_do_write_small()
| | | -- BlueStore::ExtentMap::seek_lextent()
| | | | -- BlueStore::Blob::get_ref()
| | | | | -- bluestore_blob_use_tracker_t::init(blob_logical_length, min_release_size)   // min_release_size = blob_logical_length 或 min_alloc_size
| | | | | -- bluestore_blob_use_tracker_t::get(blob_offset, length) //
| | | | -- BlueStore::ExtentMap::punch_hole()
| | | -- BlueStore::Blob::can_reuse_blob()
| | reuse blob? or new blob?
| | | -- insert to struct WriteContext {};
| | -- BlueStore::_do_write_big()
| | | -- BlueStore::ExtentMap::punch_hole()
| | | -- BlueStore::Blob::can_reuse_blob()
| | reuse blob? or new blob?
| | | -- insert to struct WriteContext {};
|-- BlueStore::_do_alloc_write()
| | -- StupidAllocator::allocate()
| | -- BlueStore::ExtentMap::set_lextent()
| | | -- BlueStore::Blob::get_ref()
| | | -- BlueStore::ExtentMap::punch_hole()
| | -- BlueStore::_buffer_cache_write()
|-- BlueStore::_wctx_finish()
```

deferred write

defer write是将要写数据封装成一个PREFIX_DEFERRED前缀的key，与元数据一块先写到db中，db写完成即可向客户端应答。然后内存中构造deferred_io，待txc数量积攒到设置的额度时开始apply，即是根据deferred_io构造aio(如果物理地址连续会进行合并)，然后就是向内核提交aio。aio完成后执行_deferred_aio_finish回调，状态机变成STATE_DEFERRED_CLEANUP，将DeferredBatch加入deferred_done_queue队列由_kv_sync_thread线程做后续的处理，删除PREFIX_DEFERRED前缀key。

defer write apply的触发条件：

(1) deferred_queue_size队列中积攒的defer write txc数量超过额度值

(2) umount时，调用_osr_drain_all()主动触发apply

(3) osd启动时进行deferred replay，先从db中读取PREFIX_DEFERRED前缀的k/v值

(4) deferred writes超过流控时主动触发apply, 流控大小是bluestore_throttle_bytes + bluestore_throttle_deferred_bytes = 192M

```
bluestore_debug_omit_block_device_write = false // 默认是false表示支持deferred write模式
bluestore_prefer_deferred_size = 0
bluestore_prefer_deferred_size_hdd = 32768  // hdd类型的osd小于32k的io则会采用deferred write模式
bluestore_prefer_deferred_size_ssd = 0      // ssd类型的osd不采用deferred write模式

bluestore_deferred_batch_ops = 0
bluestore_deferred_batch_ops_hdd = 64  // hdd类型的osd批量提交deferred write的txc数量
bluestore_deferred_batch_ops_ssd = 16  // ssd类型的osd批量提交deferred write的txc数量

bluestore_throttle_bytes = 64_M           //
bluestore_throttle_deferred_bytes = 128_M // osd_op_tp线程调用queue_transactions()时会检查deferred_bytes是否超过192_M(64_M+128_M),如果超过则首先会将此前的deferred write提交aio，当前线程会等待。待aio完成后线程被唤醒才开始处理当前io的状态机流程。不过这种场景很少出现，因为deferred write还收ops数量的条件限
                                                    // 制，通常ops会先达到最大值就开始提交aio了


bluestore_max_deferred_txc = 32 // osr里面的txc数量超过设置值并且第一个
```

哪些场景需要defered write?

(1) hdd场景小写并且io大小小于bluestore_prefer_deferred_size_hdd=32k

(2) 小写覆盖写场景，为保证数据写原子性

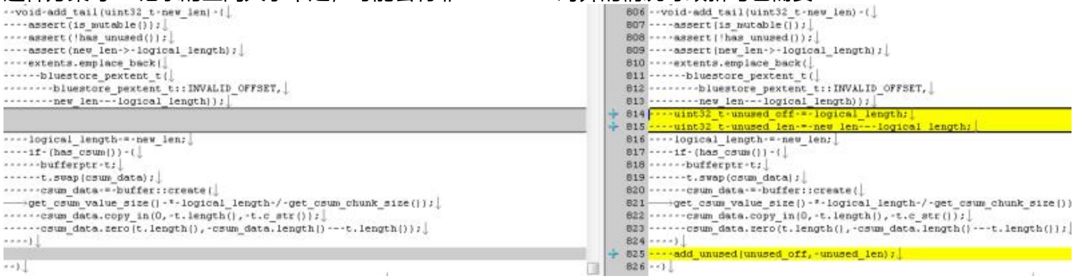
(3) 小写扩充blob场景，即使是对对象append操作，此时也会被按照覆盖写的场景来处理(这看似有点不太合理)

避免的办法：最小分配单元MAS和Blob的最大大小配置相同大小，这样pextent和lextent就成一一对应关系了

比如：以4k的io大小连续写1M，日志如下：

修改方案一：沿用原来的用16bit的bitmap来记录空间是否使用，blob首次申请空间末尾部分和扩充blob后将新申请的空间标记都为unused。

这种方案每bit记录的空间大小不定，可能会有非block size对齐的情况导致新写也需要deferred write



修改方案二：动态申请bitmap来记录空间是否使用，每bit记录的大小是block size，blob首次申请空间末尾部分和扩充blob后将新申请的空间标记都为unused。

```
[root@node3]# /var/lib/ceph/bin/ceph-kvstore-tool rocksdb db/ list L
L     %00%00%00%00%00%00%00%01
L     %00%00%00%00%00%00%00%02
L     %00%00%00%00%00%00%00%03
L     %00%00%00%00%00%00%00%04
L     %00%00%00%00%00%00%00%05
L     %00%00%00%00%00%00%00%06
L     %00%00%00%00%00%00%00%07
L     %00%00%00%00%00%00%00%08
L     %00%00%00%00%00%00%00%09
L     %00%00%00%00%00%00%00%0a  // seq是10
[root@node3]# /var/lib/ceph/bin/ceph-kvstore-tool rocksdb db/ get L %00%00%00%00%00%00%00%0a out L.bin
[root@node3]# /var/lib/ceph/bin/ceph-dencoder import L.bin type bluestore_deferred_transaction_t decode dump_json
{
    "seq": 10, //
    "ops": [
        {
            "op": 1,
            "data_len": 8192,
            "extents": [
                {
                    "offset": 24264704,
                    "length": 8192
                }
            ]
        }
    ],
    "released extents": []
}
```

6.4 Transaction::OP_ZERO

zero接口是对对象某部分空间进行释放

```
BlueStore::_zero(TransContext *txc, CollectionRef& c, OnodeRef& o, uint64_t offset, size_t length)
    BlueStore::_do_zero()
```

void BlueStore::_wctx_finish(TransContext *txc, CollectionRef& c, OnodeRef o, WriteContext *wctx,
set<SharedBlob*> *maybe_unshared_blobs)

一个事务中可能包含了多个write、zero、truncate操作，每个操作会生成相应的上下文WriteContext。该接口是对每种操作做最后的处理，将相关的信息封装到事务上下文TransContext中。

void BlueStore::_txc_finish(TransContext *txc)

事务在store层处理完成的最后处理，比如是释放事务上下文TransContext结构内存，处理wal写

BlueStore small write Log分析

可以通过开启Ceph bluestore debug来抓取其写过程中对数据的映射，具体步骤如下：
1、多副本场景

```
///// 写0~2k
2019-11-25 17:18:41.202313 7f36e7035700 15 bluestore(/sandstone-data/ceph-0) _write 2.7e_head
#2:7f7e33d5:::rbd_data.101f241b25c.0000000000000001:head# 0x0~800
2019-11-25 17:18:41.202320 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_write
#2:7f7e33d5:::rbd_data.101f241b25c.0000000000000001:head# 0x0~800 - have 0x0 (0) bytes fadvise_flags 0x0
2019-11-25 17:18:41.202327 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_onode 0x559ba0368fc0
#2:7f7e33d5:::rbd_data.101f241b25c.0000000000000001:head# nid 1435 size 0x0 (0) expected_object_size 1048576 expected_write_size 1048576
in 0 shards, 0 spanning blobs
2019-11-25 17:18:41.202335 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_onode attr _ len 292
2019-11-25 17:18:41.202338 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_onode attr snapset len 35
2019-11-25 17:18:41.202344 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2019-11-25 17:18:41.202348 7f36e7035700 30 bluestore.extentmap(0x559ba03690b0) fault_range 0x0~800
2019-11-25 17:18:41.202367 7f36e7035700 10 bluestore(/sandstone-data/ceph-0) _do_write_small 0x0~800
2019-11-25 17:18:41.202376 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _pad_zeros 0x0~800 chunk_size 0x1000
2019-11-25 17:18:41.202380 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _pad_zeros pad 0x0 + 0x800 on front/back, now 0x0~1000
2019-11-25 17:18:41.202388 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_alloc_write txc 0x559b9daca580 1 blobs
2019-11-25 17:18:41.202392 7f36e7035700 10 stupidalloc 0x0x559b9d748720 allocate_int want_size 0x10000 alloc_unit 0x10000 hint 0x0
2019-11-25 17:18:41.202394 7f36e7035700 30 stupidalloc 0x0x559b9d748720 _choose_bin len 0x10000 -> 5
2019-11-25 17:18:41.202397 7f36e7035700 30 stupidalloc 0x0x559b9d748720 allocate_int got 0x59800000~10000 from bin 9
2019-11-25 17:18:41.202401 7f36e7035700 30 stupidalloc 0x0x559b9d748720 _choose_bin len 0x5a5ef0000 -> 9
2019-11-25 17:18:41.202405 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_alloc_write prealloc [0x59800000~10000]
2019-11-25 17:18:41.202407 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_alloc_write forcing csum_order to block_size_order 12
// 创建blob
2019-11-25 17:18:41.202409 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_alloc_write initialize csum setting for new blob
Blob(0x559ba06ca000 blob([]) use_tracker(0x0 0x0) SharedBlob(0x559ba06ca1c0 sbid 0x0)) csum_type crc32c csum_order 12 csum_length 0x10000
2019-11-25 17:18:41.202432 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_alloc_write blob Blob(0x559ba06ca000
blob([0x59800000~10000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x559ba06ca1c0 sbid 0x0))
2019-11-25 17:18:41.202458 7f36e7035700 20 bluestore.blob(0x559ba06ca000) get_ref 0x0~800 Blob(0x559ba06ca000 blob([0x59800000~10000]
csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x0 0x0) SharedBlob(0x559ba06ca1c0 sbid 0x0))
2019-11-25 17:18:41.202467 7f36e7035700 20 bluestore.blob(0x559ba06ca000) get_ref init 0x10000, 10000

// 创建lextent[0x0~800], logical_offset=0x0, blob_offset=0x0, length=0x800
2019-11-25 17:18:41.202472 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_alloc_write lex 0x0~800: 0x0~800 Blob(0x559ba06ca000
blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x800) SharedBlob(0x559ba06ca1c0 sbid 0x0))

2019-11-25 17:18:41.202484 7f36e7035700 20 bluestore.BufferSpace(0x559ba06ca1d8 in 0x559b9d74e0e0) _discard 0x0~1000
2019-11-25 17:18:41.202491 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_alloc_write deferring small 0x1000 write via deferred
2019-11-25 17:18:41.202516 7f36e7035700 30 estimate gc range(hex): [0, 800)
2019-11-25 17:18:41.202522 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_write extending size to 0x800
2019-11-25 17:18:41.202526 7f36e7035700 30 bluestore.extentmap(0x559ba03690b0) dirty_range 0x0~800
2019-11-25 17:18:41.202530 7f36e7035700 20 bluestore.extentmap(0x559ba03690b0) dirty_range mark inline shard dirty
2019-11-25 17:18:41.202534 7f36e7035700 10 bluestore(/sandstone-data/ceph-0) _write 2.7e_head
#2:7f7e33d5:::rbd_data.101f241b25c.0000000000000001:head# 0x0~800 = 0

///// 写2~2k
2019-11-25 17:18:41.211629 7f36e7035700 15 bluestore(/sandstone-data/ceph-0) _write 2.7e_head
#2:7f7e33d5:::rbd_data.101f241b25c.0000000000000001:head# 0x800~800
2019-11-25 17:18:41.211632 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_write
#2:7f7e33d5:::rbd_data.101f241b25c.0000000000000001:head# 0x800~800 - have 0x800 (2048) bytes fadvise_flags 0x0
2019-11-25 17:18:41.211636 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_onode 0x559ba0368fc0
#2:7f7e33d5:::rbd_data.101f241b25c.0000000000000001:head# nid 1435 size 0x800 (2048) expected_object_size 1048576 expected_write_size
1048576 in 0 shards, 0 spanning blobs
2019-11-25 17:18:41.211657 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_onode attr _ len 292
2019-11-25 17:18:41.211660 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_onode attr snapset len 35
2019-11-25 17:18:41.211663 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_extent_map 0x0~800: 0x0~800 Blob(0x559ba06ca000
blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x800) SharedBlob(0x559ba06ca1c0 sbid 0x0))
2019-11-25 17:18:41.211712 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_extent_map csum:
[6706be76,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
2019-11-25 17:18:41.211719 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_extent_map 0x0~1000 buffer(0x559ba36c8510 space
0x559ba06ca1d8 0x0~1000 writing nocache)
2019-11-25 17:18:41.211727 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2019-11-25 17:18:41.211732 7f36e7035700 30 bluestore.extentmap(0x559ba03690b0) fault_range 0x800~800
2019-11-25 17:18:41.211736 7f36e7035700 10 bluestore(/sandstone-data/ceph-0) _do_write_small 0x800~800
2019-11-25 17:18:41.211741 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_write_small considering Blob(0x559ba06ca000
blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x800) SharedBlob(0x559ba06ca1c0 sbid 0x0))
bstart 0x0
2019-11-25 17:18:41.211749 7f36e7035700 30 bluestore.extentmap(0x559ba03690b0) fault_range 0x0~1000

// 复用blob[0x559ba06ca000], 需要进行block size对齐读
2019-11-25 17:18:41.211756 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_write_small reading head 0x800 and tail 0x0
2019-11-25 17:18:41.211760 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_read 0x0~800 size 0x800 (2048)
2019-11-25 17:18:41.211765 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_read defaulting to buffered read
2019-11-25 17:18:41.211768 7f36e7035700 30 bluestore.extentmap(0x559ba03690b0) fault_range 0x0~800
2019-11-25 17:18:41.211771 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_onode 0x559ba0368fc0
#2:7f7e33d5:::rbd_data.101f241b25c.0000000000000001:head# nid 1435 size 0x800 (2048) expected_object_size 1048576 expected_write_size
1048576 in 0 shards, 0 spanning blobs
2019-11-25 17:18:41.211780 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_onode attr _ len 292
2019-11-25 17:18:41.211782 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_onode attr snapset len 35
2019-11-25 17:18:41.211784 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_extent_map 0x0~800: 0x0~800 Blob(0x559ba06ca000
blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x800) SharedBlob(0x559ba06ca1c0 sbid 0x0))
2019-11-25 17:18:41.211794 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_extent_map csum:
[6706be76,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
2019-11-25 17:18:41.211799 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_extent_map 0x0~1000 buffer(0x559ba36c8510 space
0x559ba06ca1d8 0x0~1000 writing nocache)
2019-11-25 17:18:41.211810 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_read blob Blob(0x559ba06ca000 blob([0x59800000~10000]
csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x800) SharedBlob(0x559ba06ca1c0 sbid 0x0)) need 0x0~800 cache has
0x[0~800]
2019-11-25 17:18:41.211821 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _do_read use cache 0x0: 0x0~800
2019-11-25 17:18:41.211826 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _do_read assemble 0x0: data from 0x0~800
2019-11-25 17:18:41.211833 7f36e7035700 20 bluestore.BufferSpace(0x559ba06ca1d8 in 0x559b9d74e0e0) _discard 0x0~1000
```

```
2019-11-25 17:18:41.211844 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_write_small deferred write 0x0~1000 of mutable Blob(0x559ba06ca000 blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x800) SharedBlob(0x559ba06ca1c0 sbid 0x0)) at [0x59800000~1000]
2019-11-25 17:18:41.211853 7f36e7035700 20 bluestore.blob(0x559ba06ca000) get_ref 0x800~800 Blob(0x559ba06ca000 blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x800) SharedBlob(0x559ba06ca1c0 sbid 0x0))

// 创建lextent[0x800~800], logical_offset=0x800, blob_offset=0x800, length=0x800.
2019-11-25 17:18:41.211862 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_write_small lex 0x800~800: 0x800~800 Blob(0x559ba06ca000 blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000) SharedBlob(0x559ba06ca1c0 sbid 0x0))

2019-11-25 17:18:41.211870 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_alloc_write txc 0x559ba081f600 0 blobs
2019-11-25 17:18:41.211874 7f36e7035700 30 estimate gc range(hex): [800, 1000)
2019-11-25 17:18:41.211877 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_write extending size to 0x1000

// 注意与前面的lextent会进行合并lextent[0x0~1000]
2019-11-25 17:18:41.211880 7f36e7035700 20 bluestore.extentmap(0x559ba03690b0) compress_extent_map 0x800~800 next shard 0xffffffff merging 0x0~800: 0x0~800 Blob(0x559ba06ca000 blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000) SharedBlob(0x559ba06ca1c0 sbid 0x0)) and 0x800~800: 0x800~800 Blob(0x559ba06ca000 blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000) SharedBlob(0x559ba06ca1c0 sbid 0x0))
2019-11-25 17:18:41.211895 7f36e7035700 30 bluestore.extentmap(0x559ba03690b0) dirty_range 0x800~800
2019-11-25 17:18:41.211898 7f36e7035700 20 bluestore.extentmap(0x559ba03690b0) dirty_range mark inline shard dirty
2019-11-25 17:18:41.211902 7f36e7035700 10 bluestore(/sandstone-data/ceph-0) _write 2.7e_head #2:7f7e33d5:::rbd_data.101f241b25c.0000000000000001:head# 0x800~800 = 0

// 写4k~2k
2019-11-25 17:18:41.218859 7f36e7035700 15 bluestore(/sandstone-data/ceph-0) _write 2.7e_head #2:7f7e33d5:::rbd_data.101f241b25c.0000000000000001:head# 0x1000~800
2019-11-25 17:18:41.218865 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_write #2:7f7e33d5:::rbd_data.101f241b25c.0000000000000001:head# 0x1000~800 - have 0x1000 (4096) bytes fadvise_flags 0x0
2019-11-25 17:18:41.218872 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_onode 0x559ba0368fc0 #2:7f7e33d5:::rbd_data.101f241b25c.0000000000000001:head# nid 1435 size 0x1000 (4096) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2019-11-25 17:18:41.218880 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_onode attr _ len 292
2019-11-25 17:18:41.218882 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_onode attr snapset len 35
2019-11-25 17:18:41.218884 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_extent_map 0x0~1000: 0x0~1000 Blob(0x559ba06ca000 blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000) SharedBlob(0x559ba06ca1c0 sbid 0x0))
2019-11-25 17:18:41.218897 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_extent_map csum: [6706be76,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
2019-11-25 17:18:41.218903 7f36e7035700 30 bluestore(/sandstone-data/ceph-0) _dump_extent_map 0x0~1000 buffer(0x559ba0ab2240 space 0x559ba06ca1d8 0x0~1000 writing nocache)
2019-11-25 17:18:41.218909 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _choose_write_options prefer csum_order 12 target_blob_size 0x80000 compress=0 buffered=0
2019-11-25 17:18:41.218913 7f36e7035700 30 bluestore.extentmap(0x559ba03690b0) fault_range 0x1000~800
2019-11-25 17:18:41.218916 7f36e7035700 10 bluestore(/sandstone-data/ceph-0) _do_write_small 0x1000~800
2019-11-25 17:18:41.218920 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_write_small considering Blob(0x559ba06ca000 blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000) SharedBlob(0x559ba06ca1c0 sbid 0x0)) bstart 0x0
2019-11-25 17:18:41.218928 7f36e7035700 30 bluestore.extentmap(0x559ba03690b0) fault_range 0x1000~1000
2019-11-25 17:18:41.218933 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _apply_padding can pad head 0x0 tail 0x800
2019-11-25 17:18:41.218936 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_write_small write to unused 0x1000~1000 pad 0x0 + 0x800 of mutable Blob(0x559ba06ca000 blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000) SharedBlob(0x559ba06ca1c0 sbid 0x0))
2019-11-25 17:18:41.218947 7f36e7035700 20 bluestore.BufferSpace(0x559ba06ca1d8 in 0x559b9d74e0e0) _discard 0x1000~1000
2019-11-25 17:18:41.218953 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_write_small deferring small 0x1000 unused write via deferred
2019-11-25 17:18:41.218961 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_write_small lex old 0x0~1000: 0x0~1000 Blob(0x559ba06ca000 blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000) SharedBlob(0x559ba06ca1c0 sbid 0x0))
2019-11-25 17:18:41.218971 7f36e7035700 20 bluestore.blob(0x559ba06ca000) get_ref 0x1000~800 Blob(0x559ba06ca000 blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000) SharedBlob(0x559ba06ca1c0 sbid 0x0))

// 创建lextent[0x1000~800], logical_offset=0x1000, blob_offset=0x1000, length=0x800.
2019-11-25 17:18:41.218980 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_write_small lex 0x1000~800: 0x1000~800 Blob(0x559ba06ca000 blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffc) use_tracker(0x10000 0x1800) SharedBlob(0x559ba06ca1c0 sbid 0x0))

2019-11-25 17:18:41.218989 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_alloc_write txc 0x559ba074bb80 0 blobs
2019-11-25 17:18:41.218992 7f36e7035700 30 estimate gc range(hex): [1000, 1800)
2019-11-25 17:18:41.218995 7f36e7035700 20 bluestore(/sandstone-data/ceph-0) _do_write extending size to 0x1800

// 注意与前面的lextent会进行合并lextent[0x0~1800]
2019-11-25 17:18:41.218998 7f36e7035700 20 bluestore.extentmap(0x559ba03690b0) compress_extent_map 0x1000~800 next shard 0xffffffff merging 0x0~1000: 0x0~1000 Blob(0x559ba06ca000 blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffc) use_tracker(0x10000 0x1800) SharedBlob(0x559ba06ca1c0 sbid 0x0)) and 0x1000~800: 0x1000~800 Blob(0x559ba06ca000 blob([0x59800000~10000] csum+has_unused crc32c/0x1000 unused=0xfffc) use_tracker(0x10000 0x1800) SharedBlob(0x559ba06ca1c0 sbid 0x0))
2019-11-25 17:18:41.219013 7f36e7035700 30 bluestore.extentmap(0x559ba03690b0) dirty_range 0x1000~800
2019-11-25 17:18:41.219016 7f36e7035700 20 bluestore.extentmap(0x559ba03690b0) dirty_range mark inline shard dirty
2019-11-25 17:18:41.219019 7f36e7035700 10 bluestore(/sandstone-data/ceph-0) _write 2.7e_head #2:7f7e33d5:::rbd_data.101f241b25c.0000000000000001:head# 0x1000~800 = 0
```

1、fio以4k大小的io，写1M数据

```
fio -filename=/dev/sdd -direct=1 -iodepth 1 -thread -rw=write -ioengine=libaio -bs=4k -size=1M -numjobs=1 -runtime=180 -group_reporting -name=sqe_100write_4k
```

2、查看osd日志

///// 写0~4k

```
2019-11-20 17:06:43.888543 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) queue_transactions existing 0x558e2d5b3420 osr(2.b 0x558e2d5ffbf0)
2019-11-20 17:06:43.888564 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _txc_create osr 0x558e2d5b3420 = 0x558e2debedc0 seq 1297
2019-11-20 17:06:43.888570 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) transaction dump:
{
    "ops": [
        {
            "op_num": 0,
            "op_name": "touch",
            "collection": "2.b_head",
            "oid": "#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head#"
        },
        {
            "op_num": 1,
            "op_name": "setattrs",
```

```
            "collection": "2.b_head",
            "oid": "#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head#",
            "attr_lens": {
                "_": 292,
                "snapset": 35
            }
        },
        {
            "op_num": 2,
            "op_name": "op_setallochint",
            "collection": "2.b_head",
            "oid": "#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head#",
            "expected_object_size": "1048576",
            "expected_write_size": "1048576"
        },
        {
            "op_num": 3,
            "op_name": "write",
            "collection": "2.b_head",
            "oid": "#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head#",
            "length": 4096,
            "offset": 0,
            "bufferlist length": 4096
        },
        {
            "op_num": 4,
            "op_name": "omap_setkeys",
            "collection": "2.b_head",
            "oid": "#2:d0000000:::head#",
            "attr_lens": {
                "0000000035.00000000000000001282": 181,
                "_fastinfo": 186
            }
        }
    ]
}


2019-11-20 17:06:43.888660 7ff197c0a700 30 bluestore.OnodeSpace(0x558e2d574f48 in 0x558e2b06f420) lookup
2019-11-20 17:06:43.888663 7ff197c0a700 30 bluestore.OnodeSpace(0x558e2d574f48 in 0x558e2b06f420) lookup
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# hit 0x558e2dc0a6c0
2019-11-20 17:06:43.888667 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _touch 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head#
2019-11-20 17:06:43.888670 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _assign_nid 1278
2019-11-20 17:06:43.888672 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _touch 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# = 0
2019-11-20 17:06:43.888686 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _setattrs 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys
2019-11-20 17:06:43.888695 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _setattrs 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys = 0
2019-11-20 17:06:43.888699 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _set_alloc_hint 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags -
2019-11-20 17:06:43.888702 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _set_alloc_hint 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags - = 0
2019-11-20 17:06:43.888706 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _write 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x0~1000
2019-11-20 17:06:43.888709 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x0~1000 - have 0x0 (0) bytes fadvise_flags 0x0
2019-11-20 17:06:43.888712 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode 0x558e2dc0a6c0
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# nid 1278 size 0x0 (0) expected_object_size 1048576 expected_write_size 1048576
in 0 shards, 0 spanning blobs
2019-11-20 17:06:43.888717 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode  attr _ len 292
2019-11-20 17:06:43.888718 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode  attr snapset len 35
2019-11-20 17:06:43.888720 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2019-11-20 17:06:43.888722 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) fault_range 0x0~1000
2019-11-20 17:06:43.888728 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _do_write_small 0x0~1000
2019-11-20 17:06:43.888732 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _pad_zeros 0x0~1000 chunk_size 0x1000
2019-11-20 17:06:43.888734 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _pad_zeros pad 0x0 + 0x0 on front/back, now 0x0~1000
2019-11-20 17:06:43.888737 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write txc 0x558e2debedc0 1 blobs
2019-11-20 17:06:43.888741 7ff197c0a700 10 stupidalloc 0x0x558e2b07c720 allocate_int want_size 0x10000 alloc_unit 0x10000 hint 0x0
2019-11-20 17:06:43.888749 7ff197c0a700 30 stupidalloc 0x0x558e2b07c720 _choose_bin len 0x10000 -> 5
2019-11-20 17:06:43.888758 7ff197c0a700 30 stupidalloc 0x0x558e2b07c720 allocate_int got 0x990000~10000 from bin 9
2019-11-20 17:06:43.888762 7ff197c0a700 30 stupidalloc 0x0x558e2b07c720 _choose_bin len 0x5fed60000 -> 9
2019-11-20 17:06:43.888763 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write prealloc [0x990000~10000]
2019-11-20 17:06:43.888765 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write forcing csum_order to block_size_order 12

// 第一次写创建blob
2019-11-20 17:06:43.888766 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write initialize csum setting for new blob
Blob(0x558e2dd77650 blob([]) use_tracker(0x0 0x0) SharedBlob(0x558e2dd76f50 sbid 0x0)) csum_type crc32c csum_order 12 csum_length 0x10000

2019-11-20 17:06:43.888773 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write blob Blob(0x558e2dd77650
blob([0x990000~10000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.888800 7ff197c0a700 20 bluestore.blob(0x558e2dd77650) get_ref 0x0~1000 Blob(0x558e2dd77650 blob([0x990000~10000]
csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x0 0x0) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.888805 7ff197c0a700 20 bluestore.blob(0x558e2dd77650) get_ref init 0x10000, 10000

// 第一个lextent[0x0~1000]：logical_offset=0x0, blob_offset=0x0, length=0x1000
2019-11-20 17:06:43.888808 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write lex 0x0~1000: 0x0~1000 Blob(0x558e2dd77650
blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000) SharedBlob(0x558e2dd76f50 sbid 0x0))

2019-11-20 17:06:43.888812 7ff197c0a700 20 bluestore.BufferSpace(0x558e2dd76f68 in 0x558e2b06f420) _discard 0x0~1000




//// 写4k~4k
2019-11-20 17:06:43.895479 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _setattrs 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys
2019-11-20 17:06:43.895482 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _setattrs 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys = 0
2019-11-20 17:06:43.895486 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _set_alloc_hint 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags -
2019-11-20 17:06:43.895489 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _set_alloc_hint 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags - = 0
```

```
2019-11-20 17:06:43.895491 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _write 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x1000~1000
2019-11-20 17:06:43.895494 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x1000~1000 - have 0x1000 (4096) bytes fadvise_flags 0x0
2019-11-20 17:06:43.895497 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode 0x558e2dc0a6c0
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# nid 1278 size 0x1000 (4096) expected_object_size 1048576 expected_write_size
1048576 in 0 shards, 0 spanning blobs
2019-11-20 17:06:43.895500 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode attr _ len 292
2019-11-20 17:06:43.895501 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode attr snapset len 35
2019-11-20 17:06:43.895501 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x0~1000: 0x0~1000 Blob(0x558e2dd77650
blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.895536 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map csum:
[e8ba7e74,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
2019-11-20 17:06:43.895540 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x0~1000 buffer(0x558e2dfb3830 space
0x558e2dd76f68 0x0~1000 writing nocache)
2019-11-20 17:06:43.895543 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2019-11-20 17:06:43.895546 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) fault_range 0x1000~1000
2019-11-20 17:06:43.895547 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _do_write_small 0x1000~1000

// 复用blob[0x558e2dd77650]中未使用的空间[0x1000~1000]
2019-11-20 17:06:43.895549 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small considering Blob(0x558e2dd77650
blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000) SharedBlob(0x558e2dd76f50 sbid 0x0)) bstart
0x0
2019-11-20 17:06:43.895557 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small write to unused 0x1000~1000 pad 0x0 + 0x0 of
mutable Blob(0x558e2dd77650 blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000)
SharedBlob(0x558e2dd76f50 sbid 0x0))

2019-11-20 17:06:43.895561 7ff197c0a700 20 bluestore.BufferSpace(0x558e2dd76f68 in 0x558e2b06f420) _discard 0x1000~1000
2019-11-20 17:06:43.895563 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small deferring small 0x1000 unused write via
deferred

2019-11-20 17:06:43.895567 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small lex old 0x0~1000: 0x0~1000
Blob(0x558e2dd77650 blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000)
SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.895571 7ff197c0a700 20 bluestore.blob(0x558e2dd77650) get_ref 0x1000~1000 Blob(0x558e2dd77650 blob([0x990000~10000]
csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000) SharedBlob(0x558e2dd76f50 sbid 0x0))

// 新生成一个lextent[0x1000~1000]：logical_offset=0x1000, blob_offset=0x1000, length=0x1000. 注意与前面的lextent会进行合并lextent[0x0~2000]
2019-11-20 17:06:43.895574 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small lex 0x1000~1000: 0x1000~1000
Blob(0x558e2dd77650 blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0xfffc) use_tracker(0x10000 0x2000)
SharedBlob(0x558e2dd76f50 sbid 0x0))

2019-11-20 17:06:43.895578 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write txc 0x558e2ded6b00 0 blobs
2019-11-20 17:06:43.895580 7ff197c0a700 30 estimate gc range(hex): [1000, 2000)
2019-11-20 17:06:43.895581 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write extending size to 0x2000
2019-11-20 17:06:43.895582 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) compress_extent_map 0x1000~1000 next shard 0xffffffff
merging 0x0~1000: 0x0~1000 Blob(0x558e2dd77650 blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0xfffc) use_tracker(0x10000
0x2000) SharedBlob(0x558e2dd76f50 sbid 0x0)) and 0x1000~1000: 0x1000~1000 Blob(0x558e2dd77650 blob([0x990000~10000] csum+has_unused
crc32c/0x1000 unused=0xfffc) use_tracker(0x10000 0x2000) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.895590 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) dirty_range 0x1000~1000
2019-11-20 17:06:43.895591 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) dirty_range mark inline shard dirty
2019-11-20 17:06:43.895592 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _write 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x1000~1000 = 0

// 写8k~4k
2019-11-20 17:06:43.899798 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _write 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x2000~1000
2019-11-20 17:06:43.899801 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x2000~1000 - have 0x2000 (8192) bytes fadvise_flags 0x0
2019-11-20 17:06:43.899803 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode 0x558e2dc0a6c0
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# nid 1278 size 0x2000 (8192) expected_object_size 1048576 expected_write_size
1048576 in 0 shards, 0 spanning blobs
2019-11-20 17:06:43.899806 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode attr _ len 292
2019-11-20 17:06:43.899807 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode attr snapset len 35
2019-11-20 17:06:43.899808 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x0~2000: 0x0~2000 Blob(0x558e2dd77650
blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0xfffc) use_tracker(0x10000 0x2000) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.899813 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map csum:
[e8ba7e74,2cab4bfd,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
2019-11-20 17:06:43.899816 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x0~1000 buffer(0x558e2dfb3830 space
0x558e2dd76f68 0x0~1000 writing nocache)
2019-11-20 17:06:43.899818 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x1000~1000 buffer(0x558e2e11f950
space 0x558e2dd76f68 0x1000~1000 writing nocache)
2019-11-20 17:06:43.899820 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2019-11-20 17:06:43.899821 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) fault_range 0x2000~1000
2019-11-20 17:06:43.899823 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _do_write_small 0x2000~1000
2019-11-20 17:06:43.899824 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small considering Blob(0x558e2dd77650
blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0xfffc) use_tracker(0x10000 0x2000) SharedBlob(0x558e2dd76f50 sbid 0x0)) bstart
0x0
2019-11-20 17:06:43.899828 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small write to unused 0x2000~1000 pad 0x0 + 0x0 of
mutable Blob(0x558e2dd77650 blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0xfffc) use_tracker(0x10000 0x2000)
SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.899832 7ff197c0a700 20 bluestore.BufferSpace(0x558e2dd76f68 in 0x558e2b06f420) _discard 0x2000~1000
2019-11-20 17:06:43.899834 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small deferring small 0x1000 unused write via
deferred
2019-11-20 17:06:43.899837 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small lex old 0x0~2000: 0x0~2000
Blob(0x558e2dd77650 blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0xfffc) use_tracker(0x10000 0x2000)
SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.899841 7ff197c0a700 20 bluestore.blob(0x558e2dd77650) get_ref 0x2000~1000 Blob(0x558e2dd77650 blob([0x990000~10000]
csum+has_unused crc32c/0x1000 unused=0xfffc) use_tracker(0x10000 0x2000) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.899845 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small lex 0x2000~1000: 0x2000~1000
Blob(0x558e2dd77650 blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0xfff8) use_tracker(0x10000 0x3000)
SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.899848 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write txc 0x558e2decf080 0 blobs
2019-11-20 17:06:43.899849 7ff197c0a700 30 estimate gc range(hex): [2000, 3000)
2019-11-20 17:06:43.899850 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write extending size to 0x3000
2019-11-20 17:06:43.899851 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) compress_extent_map 0x2000~1000 next shard 0xffffffff
merging 0x0~2000: 0x0~2000 Blob(0x558e2dd77650 blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0xfff8) use_tracker(0x10000
0x3000) SharedBlob(0x558e2dd76f50 sbid 0x0)) and 0x2000~1000: 0x2000~1000 Blob(0x558e2dd77650 blob([0x990000~10000] csum+has_unused
crc32c/0x1000 unused=0xfff8) use_tracker(0x10000 0x3000) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.899856 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) dirty_range 0x2000~1000
2019-11-20 17:06:43.899857 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) dirty_range mark inline shard dirty
2019-11-20 17:06:43.899859 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _write 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x2000~1000 = 0
```

```
///// 写60k~4k
2019-11-20 17:06:43.954383 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _setattrs 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys
2019-11-20 17:06:43.954387 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _setattrs 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys = 0
2019-11-20 17:06:43.954390 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _set_alloc_hint 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags -
2019-11-20 17:06:43.954393 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _set_alloc_hint 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags - = 0
2019-11-20 17:06:43.954396 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _write 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0xf000~1000
2019-11-20 17:06:43.954399 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0xf000~1000 - have 0xf000 (61440) bytes fadvise_flags 0x0
2019-11-20 17:06:43.954402 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode 0x558e2dc0a6c0
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# nid 1278 size 0xf000 (61440) expected_object_size 1048576 expected_write_size
1048576 in 0 shards, 0 spanning blobs
2019-11-20 17:06:43.954405 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode attr _ len 292
2019-11-20 17:06:43.954406 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode attr snapset len 35
2019-11-20 17:06:43.954407 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x0~f000: 0x0~f000 Blob(0x558e2dd77650
blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0x8000) use_tracker(0x10000 0xf000) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.954412 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map csum:
[e8ba7e74,2cab4bfd,f8e16d5e,906def58,8c267066,70148deb,74666a62,b6857eae,5eaa119a,d59b091a,1c153456,aa5d89d6,89244de0,51d7ed58,7f3b84e2,0]
2019-11-20 17:06:43.954415 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x0~1000 buffer(0x558e2dfb3830 space
0x558e2dd76f68 0x0~1000 writing nocache)
2019-11-20 17:06:43.954417 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x1000~1000 buffer(0x558e2e11f950
space 0x558e2dd76f68 0x1000~1000 writing nocache)
2019-11-20 17:06:43.954418 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x2000~1000 buffer(0x558e2e11f710
space 0x558e2dd76f68 0x2000~1000 writing nocache)
2019-11-20 17:06:43.954419 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x3000~1000 buffer(0x558e2e14e000
space 0x558e2dd76f68 0x3000~1000 writing nocache)
2019-11-20 17:06:43.954421 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x4000~1000 buffer(0x558e2dfb26c0
space 0x558e2dd76f68 0x4000~1000 writing nocache)
2019-11-20 17:06:43.954422 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x5000~1000 buffer(0x558e2dfb2480
space 0x558e2dd76f68 0x5000~1000 writing nocache)
2019-11-20 17:06:43.954424 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x6000~1000 buffer(0x558e2e11f560
space 0x558e2dd76f68 0x6000~1000 writing nocache)
2019-11-20 17:06:43.954425 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x7000~1000 buffer(0x558e2e1685a0
space 0x558e2dd76f68 0x7000~1000 writing nocache)
2019-11-20 17:06:43.954427 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x8000~1000 buffer(0x558e2e14ee10
space 0x558e2dd76f68 0x8000~1000 writing nocache)
2019-11-20 17:06:43.954428 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x9000~1000 buffer(0x558e2e0030e0
space 0x558e2dd76f68 0x9000~1000 writing nocache)
2019-11-20 17:06:43.954430 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0xa000~1000 buffer(0x558e2b3cb4d0
space 0x558e2dd76f68 0xa000~1000 writing nocache)
2019-11-20 17:06:43.954431 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0xb000~1000 buffer(0x558e2e168ea0
space 0x558e2dd76f68 0xb000~1000 writing nocache)
2019-11-20 17:06:43.954432 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0xc000~1000 buffer(0x558e2e0d9050
space 0x558e2dd76f68 0xc000~1000 writing nocache)
2019-11-20 17:06:43.954434 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0xd000~1000 buffer(0x558e2b3caf30
space 0x558e2dd76f68 0xd000~1000 writing nocache)
2019-11-20 17:06:43.954435 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0xe000~1000 buffer(0x558e2e0d9950
space 0x558e2dd76f68 0xe000~1000 writing nocache)
2019-11-20 17:06:43.954437 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2019-11-20 17:06:43.954439 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) fault_range 0xf000~1000
2019-11-20 17:06:43.954440 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _do_write_small 0xf000~1000
2019-11-20 17:06:43.954442 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small considering Blob(0x558e2dd77650
blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0x8000) use_tracker(0x10000 0xf000) SharedBlob(0x558e2dd76f50 sbid 0x0)) bstart
0x0
2019-11-20 17:06:43.954445 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small write to unused 0xf000~1000 pad 0x0 + 0x0 of
mutable Blob(0x558e2dd77650 blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0x8000) use_tracker(0x10000 0xf000)
SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.954449 7ff197c0a700 20 bluestore.BufferSpace(0x558e2dd76f68 in 0x558e2b06f420) _discard 0xf000~1000
2019-11-20 17:06:43.954451 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small deferring small 0x1000 unused write via
deferred
2019-11-20 17:06:43.954454 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small lex old 0x0~f000: 0x0~f000
Blob(0x558e2dd77650 blob([0x990000~10000] csum+has_unused crc32c/0x1000 unused=0x8000) use_tracker(0x10000 0xf000)
SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.954458 7ff197c0a700 20 bluestore.blob(0x558e2dd77650) get_ref 0xf000~1000 Blob(0x558e2dd77650 blob([0x990000~10000]
csum+has_unused crc32c/0x1000 unused=0x8000) use_tracker(0x10000 0xf000) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.954461 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small lex 0xf000~1000: 0xf000~1000
Blob(0x558e2dd77650 blob([0x990000~10000] csum crc32c/0x1000) use_tracker(0x10000 0x10000) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.954465 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write txc 0x558e2e0c62c0 0 blobs
2019-11-20 17:06:43.954466 7ff197c0a700 30 estimate gc range(hex): [f000, 10000)
2019-11-20 17:06:43.954467 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write extending size to 0x10000
2019-11-20 17:06:43.954469 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) compress_extent_map 0xf000~1000 next shard 0xffffffff
merging 0x0~f000: 0x0~f000 Blob(0x558e2dd77650 blob([0x990000~10000] csum crc32c/0x1000) use_tracker(0x10000 0x10000)
SharedBlob(0x558e2dd76f50 sbid 0x0)) and 0xf000~1000: 0xf000~1000 Blob(0x558e2dd77650 blob([0x990000~10000] csum crc32c/0x1000)
use_tracker(0x10000 0x10000) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.954474 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) dirty_range 0xf000~1000
2019-11-20 17:06:43.954476 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) dirty_range mark inline shard dirty
2019-11-20 17:06:43.954477 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _write 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0xf000~1000 = 0



/// 写64k~4k
2019-11-20 17:06:43.959614 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _setattrs 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys
2019-11-20 17:06:43.959618 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _setattrs 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys = 0
2019-11-20 17:06:43.959621 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _set_alloc_hint 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags -
2019-11-20 17:06:43.959624 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _set_alloc_hint 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags - = 0
2019-11-20 17:06:43.959627 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _write 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x10000~1000
2019-11-20 17:06:43.959630 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x10000~1000 - have 0x10000 (65536) bytes fadvise_flags 0x0
2019-11-20 17:06:43.959633 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode 0x558e2dc0a6c0
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# nid 1278 size 0x10000 (65536) expected_object_size 1048576 expected_write_size
1048576 in 0 shards, 0 spanning blobs
2019-11-20 17:06:43.959636 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode attr _ len 292
```

```
2019-11-20 17:06:43.959637 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode  attr snapset len 35
2019-11-20 17:06:43.959638 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map   0x0~10000: 0x0~10000 Blob(0x558e2dd77650
blob([0x990000~10000] csum crc32c/0x1000) use_tracker(0x10000 0x10000) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20    17:06:43.959643    7ff197c0a700    30    bluestore(/sandstone-data/ceph-2)    _dump_extent_map                          csum:
[e8ba7e74,2cab4bfd,f8e16d5e,906def58,8c267066,70148deb,74666a62,b6857eae,5eaa119a,d59b091a,1c153456,aa5d89d6,89244de0,51d7ed58,7f3b84e2,78fa831a]
2019-11-20 17:06:43.959646 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map        0x0~1000 buffer(0x558e2dfb3830 space
0x558e2dd76f68 0x0~1000 writing nocache)
2019-11-20 17:06:43.959647 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x1000~1000 buffer(0x558e2e11f950
space 0x558e2dd76f68 0x1000~1000 writing nocache)
2019-11-20 17:06:43.959649 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x2000~1000 buffer(0x558e2e11f710
space 0x558e2dd76f68 0x2000~1000 writing nocache)
2019-11-20 17:06:43.959651 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x3000~1000 buffer(0x558e2e14e000
space 0x558e2dd76f68 0x3000~1000 writing nocache)
2019-11-20 17:06:43.959652 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x4000~1000 buffer(0x558e2dfb26c0
space 0x558e2dd76f68 0x4000~1000 writing nocache)
2019-11-20 17:06:43.959653 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x5000~1000 buffer(0x558e2dfb2480
space 0x558e2dd76f68 0x5000~1000 writing nocache)
2019-11-20 17:06:43.959655 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x6000~1000 buffer(0x558e2e11f560
space 0x558e2dd76f68 0x6000~1000 writing nocache)
2019-11-20 17:06:43.959656 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x7000~1000 buffer(0x558e2e1685a0
space 0x558e2dd76f68 0x7000~1000 writing nocache)
2019-11-20 17:06:43.959658 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x8000~1000 buffer(0x558e2e14ee10
space 0x558e2dd76f68 0x8000~1000 writing nocache)
2019-11-20 17:06:43.959659 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x9000~1000 buffer(0x558e2e0030e0
space 0x558e2dd76f68 0x9000~1000 writing nocache)
2019-11-20 17:06:43.959661 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xa000~1000 buffer(0x558e2b3cb4d0
space 0x558e2dd76f68 0xa000~1000 writing nocache)
2019-11-20 17:06:43.959662 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xb000~1000 buffer(0x558e2e168ea0
space 0x558e2dd76f68 0xb000~1000 writing nocache)
2019-11-20 17:06:43.959663 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xc000~1000 buffer(0x558e2e0d9050
space 0x558e2dd76f68 0xc000~1000 writing nocache)
2019-11-20 17:06:43.959665 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xd000~1000 buffer(0x558e2b3caf30
space 0x558e2dd76f68 0xd000~1000 writing nocache)
2019-11-20 17:06:43.959666 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xe000~1000 buffer(0x558e2e0d9950
space 0x558e2dd76f68 0xe000~1000 writing nocache)
2019-11-20 17:06:43.959668 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xf000~1000 buffer(0x558e2e0d9d40
space 0x558e2dd76f68 0xf000~1000 writing nocache)
2019-11-20 17:06:43.959669 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2019-11-20 17:06:43.959671 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) fault_range 0x10000~1000
2019-11-20 17:06:43.959673 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _do_write_small 0x10000~1000

//复用Blob[0x558e2dd77650],因为blob已分配的空间已使用完可以申请新的物理空间扩充blob的大小
2019-11-20   17:06:43.959674   7ff197c0a700   20   bluestore(/sandstone-data/ceph-2)   _do_write_small   considering   Blob(0x558e2dd77650
blob([0x990000~10000] csum crc32c/0x1000) use_tracker(0x10000 0x10000) SharedBlob(0x558e2dd76f50 sbid 0x0)) bstart 0x0
2019-11-20 17:06:43.959680 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _pad_zeros 0x10000~1000 chunk_size 0x1000
2019-11-20 17:06:43.959682 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _pad_zeros pad 0x0 + 0x0 on front/back, now 0x10000~1000
2019-11-20   17:06:43.959683   7ff197c0a700   20   bluestore(/sandstone-data/ceph-2)   _do_write_small   reuse   blob   Blob(0x558e2dd77650
blob([0x990000~10000,!~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,0]) SharedBlob(0x558e2dd76f50 sbid 0x0)) (0x10000~1000)
(0x10000~1000)

2019-11-20 17:06:43.959689 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write txc 0x558e2de93600 1 blobs
2019-11-20 17:06:43.959692 7ff197c0a700 10 stupidalloc 0x0x558e2b07c720 allocate_int want_size 0x10000 alloc_unit 0x10000 hint 0x0
2019-11-20 17:06:43.959693 7ff197c0a700 30 stupidalloc 0x0x558e2b07c720 _choose_bin len 0x10000 -> 5
2019-11-20 17:06:43.959696 7ff197c0a700 30 stupidalloc 0x0x558e2b07c720 allocate_int got 0x9a0000~10000 from bin 9
2019-11-20 17:06:43.959699 7ff197c0a700 30 stupidalloc 0x0x558e2b07c720 _choose_bin len 0x5fed50000 -> 9
2019-11-20 17:06:43.959701 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write prealloc [0x9a0000~10000]
2019-11-20    17:06:43.959704    7ff197c0a700    20    bluestore(/sandstone-data/ceph-2)    _do_alloc_write    blob    Blob(0x558e2dd77650
blob([0x990000~10000,0x9a0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,0]) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20    17:06:43.959709    7ff197c0a700    20    bluestore.blob(0x558e2dd77650)    get_ref    0x10000~1000    Blob(0x558e2dd77650
blob([0x990000~10000,0x9a0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,0]) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20   17:06:43.959712   7ff197c0a700   20   bluestore(/sandstone-data/ceph-2)   _do_alloc_write     lex   0x10000~1000:   0x10000~1000
Blob(0x558e2dd77650     blob([0x990000~10000,0x9a0000~10000]     csum     crc32c/0x1000)     use_tracker(0x2*0x10000     0x[10000,1000])
SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.959716 7ff197c0a700 20 bluestore.BufferSpace(0x558e2dd76f68 in 0x558e2b06f420) _discard 0x10000~1000
2019-11-20 17:06:43.959718 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write deferring small 0x1000 write via deferred
2019-11-20 17:06:43.959720 7ff197c0a700 30 estimate gc range(hex): [10000, 11000)
2019-11-20 17:06:43.959722 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write extending size to 0x11000
2019-11-20 17:06:43.959723 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) compress_extent_map 0x10000~1000 next shard 0xffffffff
merging  0x0~10000:  0x0~10000  Blob(0x558e2dd77650  blob([0x990000~10000,0x9a0000~10000]  csum  crc32c/0x1000)  use_tracker(0x2*0x10000
0x[10000,1000])      SharedBlob(0x558e2dd76f50      sbid      0x0))      and      0x10000~1000:      0x10000~1000      Blob(0x558e2dd77650
blob([0x990000~10000,0x9a0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,1000]) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.959729 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) dirty_range 0x10000~1000
2019-11-20 17:06:43.959730 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) dirty_range mark inline shard dirty
2019-11-20        17:06:43.959731        7ff197c0a700        10        bluestore(/sandstone-data/ceph-2)        _write        2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x10000~1000 = 0



// 68k~4k //overwrite流程
2019-11-20        17:06:43.963751        7ff197c0a700        15        bluestore(/sandstone-data/ceph-2)        _setattrs        2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys
2019-11-20        17:06:43.963754        7ff197c0a700        10        bluestore(/sandstone-data/ceph-2)        _setattrs        2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys = 0
2019-11-20       17:06:43.963757       7ff197c0a700       15       bluestore(/sandstone-data/ceph-2)       _set_alloc_hint       2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags -
2019-11-20       17:06:43.963760       7ff197c0a700       10       bluestore(/sandstone-data/ceph-2)       _set_alloc_hint       2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags - = 0
2019-11-20        17:06:43.963763        7ff197c0a700        15        bluestore(/sandstone-data/ceph-2)        _write        2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x11000~1000
2019-11-20           17:06:43.963766           7ff197c0a700           20           bluestore(/sandstone-data/ceph-2)           _do_write
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x11000~1000 - have 0x11000 (69632) bytes fadvise_flags 0x0
2019-11-20       17:06:43.963769       7ff197c0a700       30       bluestore(/sandstone-data/ceph-2)       _dump_onode       0x558e2dc0a6c0
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# nid 1278 size 0x11000 (69632) expected_object_size 1048576 expected_write_size
1048576 in 0 shards, 0 spanning blobs
2019-11-20 17:06:43.963772 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode  attr _ len 292
2019-11-20 17:06:43.963773 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode  attr snapset len 35
2019-11-20 17:06:43.963774 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map   0x0~11000: 0x0~11000 Blob(0x558e2dd77650
blob([0x990000~10000,0x9a0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,1000]) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20    17:06:43.963779    7ff197c0a700    30    bluestore(/sandstone-data/ceph-2)    _dump_extent_map                          csum:
[e8ba7e74,2cab4bfd,f8e16d5e,906def58,8c267066,70148deb,74666a62,b6857eae,5eaa119a,d59b091a,1c153456,aa5d89d6,89244de0,51d7ed58,7f3b84e2,78fa831a,999ce1c,0,
2019-11-20 17:06:43.963783 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map        0x0~1000 buffer(0x558e2dfb3830 space
0x558e2dd76f68 0x0~1000 writing nocache)
2019-11-20 17:06:43.963785 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x1000~1000 buffer(0x558e2e11f950
space 0x558e2dd76f68 0x1000~1000 writing nocache)
```

```
2019-11-20 17:06:43.963787 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x2000~1000 buffer(0x558e2e11f710
space 0x558e2dd76f68 0x2000~1000 writing nocache)
2019-11-20 17:06:43.963788 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x3000~1000 buffer(0x558e2e14e000
space 0x558e2dd76f68 0x3000~1000 writing nocache)
2019-11-20 17:06:43.963790 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x4000~1000 buffer(0x558e2dfb26c0
space 0x558e2dd76f68 0x4000~1000 writing nocache)
2019-11-20 17:06:43.963791 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x5000~1000 buffer(0x558e2dfb2480
space 0x558e2dd76f68 0x5000~1000 writing nocache)
2019-11-20 17:06:43.963793 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x6000~1000 buffer(0x558e2e11f560
space 0x558e2dd76f68 0x6000~1000 writing nocache)
2019-11-20 17:06:43.963794 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x7000~1000 buffer(0x558e2e1685a0
space 0x558e2dd76f68 0x7000~1000 writing nocache)
2019-11-20 17:06:43.963796 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x8000~1000 buffer(0x558e2e14ee10
space 0x558e2dd76f68 0x8000~1000 writing nocache)
2019-11-20 17:06:43.963798 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x9000~1000 buffer(0x558e2e0030e0
space 0x558e2dd76f68 0x9000~1000 writing nocache)
2019-11-20 17:06:43.963799 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xa000~1000 buffer(0x558e2b3cb4d0
space 0x558e2dd76f68 0xa000~1000 writing nocache)
2019-11-20 17:06:43.963800 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xb000~1000 buffer(0x558e2e168ea0
space 0x558e2dd76f68 0xb000~1000 writing nocache)
2019-11-20 17:06:43.963802 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xc000~1000 buffer(0x558e2e0d9050
space 0x558e2dd76f68 0xc000~1000 writing nocache)
2019-11-20 17:06:43.963803 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xd000~1000 buffer(0x558e2b3caf30
space 0x558e2dd76f68 0xd000~1000 writing nocache)
2019-11-20 17:06:43.963804 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xe000~1000 buffer(0x558e2e0d9950
space 0x558e2dd76f68 0xe000~1000 writing nocache)
2019-11-20 17:06:43.963806 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xf000~1000 buffer(0x558e2e0d9d40
space 0x558e2dd76f68 0xf000~1000 writing nocache)
2019-11-20 17:06:43.963807 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x10000~1000 buffer(0x558e2e11f8c0
space 0x558e2dd76f68 0x10000~1000 writing nocache)
2019-11-20 17:06:43.963809 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2019-11-20 17:06:43.963811 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) fault_range 0x11000~1000
2019-11-20 17:06:43.963812 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _do_write_small 0x11000~1000
2019-11-20 17:06:43.963814 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small considering Blob(0x558e2dd77650
blob([0x990000~10000,0x9a0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,1000]) SharedBlob(0x558e2dd76f50 sbid 0x0))
bstart 0x0
2019-11-20 17:06:43.963818 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small reading head 0x0 and tail 0x0
2019-11-20 17:06:43.963820 7ff197c0a700 20 bluestore.BufferSpace(0x558e2dd76f68 in 0x558e2b06f420) _discard 0x11000~1000
2019-11-20 17:06:43.963823 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small deferred write 0x11000~1000 of mutable
Blob(0x558e2dd77650 blob([0x990000~10000,0x9a0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,1000])
SharedBlob(0x558e2dd76f50 sbid 0x0)) at [0x9a1000~1000]
2019-11-20 17:06:43.963829 7ff197c0a700 20 bluestore.blob(0x558e2dd77650) get_ref 0x11000~1000 Blob(0x558e2dd77650
blob([0x990000~10000,0x9a0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,1000]) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.963833 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small lex 0x11000~1000: 0x11000~1000
Blob(0x558e2dd77650 blob([0x990000~10000,0x9a0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,2000])
SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.963836 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write txc 0x558e2e1138c0 0 blobs
2019-11-20 17:06:43.963838 7ff197c0a700 30 estimate gc range(hex): [11000, 12000)
2019-11-20 17:06:43.963839 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write extending size to 0x12000
2019-11-20 17:06:43.963840 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) compress_extent_map 0x11000~1000 next shard 0xffffffff
merging 0x0~11000: 0x0~11000 Blob(0x558e2dd77650 blob([0x990000~10000,0x9a0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000
0x[10000,2000]) SharedBlob(0x558e2dd76f50 sbid 0x0)) and 0x11000~1000: 0x11000~1000 Blob(0x558e2dd77650
blob([0x990000~10000,0x9a0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,2000]) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:43.963846 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) dirty_range 0x11000~1000
2019-11-20 17:06:43.963848 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) dirty_range mark inline shard dirty
2019-11-20 17:06:43.963849 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _write 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x11000~1000 = 0
```

```
// 512k~4k
2019-11-20 17:06:44.507034 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _setattrs 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys
2019-11-20 17:06:44.507038 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _setattrs 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys = 0
2019-11-20 17:06:44.507041 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _set_alloc_hint 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags -
2019-11-20 17:06:44.507044 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _set_alloc_hint 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags - = 0
2019-11-20 17:06:44.507047 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _write 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x80000~1000
2019-11-20 17:06:44.507050 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x80000~1000 - have 0x80000 (524288) bytes fadvise_flags 0x0
2019-11-20 17:06:44.507053 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode 0x558e2dc0a6c0
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# nid 1278 size 0x80000 (524288) expected_object_size 1048576 expected_write_size
1048576 in 0 shards, 0 spanning blobs
2019-11-20 17:06:44.507056 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode  attr _ len 292
2019-11-20 17:06:44.507057 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode  attr snapset len 35
2019-11-20 17:06:44.507058 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map   0x0~80000: 0x0~80000 Blob(0x558e2dd77650
blob([0x990000~10000,0x9a0000~10000,0x9b0000~10000,0x9c0000~10000,0x9d0000~10000,0x9e0000~10000,0x9f0000~10000,0xa00000~10000] csum
crc32c/0x1000) use_tracker(0x8*0x10000 0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:44.507068 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map csum:
[e8ba7e74,2cab4bfd,f8e16d5e,906def58,8c267066,70148deb,74666a62,b6857eae,5eaa119a,d59b091a,1c153456,aa5d89d6,89244de0,51d7ed58,7f3b84e2,78fa831a,999ce1c,ec
2019-11-20 17:06:44.507078 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x6a000~1000 buffer(0x558e2e169170
space 0x558e2dd76f68 0x6a000~1000 writing nocache)
2019-11-20 17:06:44.507080 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x6b000~1000 buffer(0x558e2e14efc0
space 0x558e2dd76f68 0x6b000~1000 writing nocache)
2019-11-20 17:06:44.507082 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x6c000~1000 buffer(0x558e2dfb3a70
space 0x558e2dd76f68 0x6c000~1000 writing nocache)
2019-11-20 17:06:44.507083 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x6d000~1000 buffer(0x558e2dfb3710
space 0x558e2dd76f68 0x6d000~1000 writing nocache)
2019-11-20 17:06:44.507085 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x6e000~1000 buffer(0x558e2e14fb00
space 0x558e2dd76f68 0x6e000~1000 writing nocache)
2019-11-20 17:06:44.507086 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x6f000~1000 buffer(0x558e2e14e240
space 0x558e2dd76f68 0x6f000~1000 writing nocache)
2019-11-20 17:06:44.507088 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x70000~1000 buffer(0x558e2b3cb680
space 0x558e2dd76f68 0x70000~1000 writing nocache)
2019-11-20 17:06:44.507090 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x71000~1000 buffer(0x558e2e14e750
space 0x558e2dd76f68 0x71000~1000 writing nocache)
2019-11-20 17:06:44.507091 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0x72000~1000 buffer(0x558e2e14e870
space 0x558e2dd76f68 0x72000~1000 writing nocache)
```

```
2019-11-20 17:06:44.507093 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x73000~1000 buffer(0x558e2dec2480
space 0x558e2dd76f68 0x73000~1000 writing nocache)
2019-11-20 17:06:44.507094 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x74000~1000 buffer(0x558e2e11e870
space 0x558e2dd76f68 0x74000~1000 writing nocache)
2019-11-20 17:06:44.507096 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x75000~1000 buffer(0x558e2dec2360
space 0x558e2dd76f68 0x75000~1000 writing nocache)
2019-11-20 17:06:44.507097 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x76000~1000 buffer(0x558e2dec2120
space 0x558e2dd76f68 0x76000~1000 writing nocache)
2019-11-20 17:06:44.507099 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x77000~1000 buffer(0x558e2b3cb5f0
space 0x558e2dd76f68 0x77000~1000 writing nocache)
2019-11-20 17:06:44.507100 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x78000~1000 buffer(0x558e2e11e2d0
space 0x558e2dd76f68 0x78000~1000 writing nocache)
2019-11-20 17:06:44.507102 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x79000~1000 buffer(0x558e2e0d9320
space 0x558e2dd76f68 0x79000~1000 writing nocache)
2019-11-20 17:06:44.507103 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7a000~1000 buffer(0x558e2e0026c0
space 0x558e2dd76f68 0x7a000~1000 writing nocache)
2019-11-20 17:06:44.507104 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7b000~1000 buffer(0x558e2dfb25a0
space 0x558e2dd76f68 0x7b000~1000 writing nocache)
2019-11-20 17:06:44.507106 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7c000~1000 buffer(0x558e2dfb27e0
space 0x558e2dd76f68 0x7c000~1000 writing nocache)
2019-11-20 17:06:44.507107 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7d000~1000 buffer(0x558e2e14f050
space 0x558e2dd76f68 0x7d000~1000 writing nocache)
2019-11-20 17:06:44.507109 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7e000~1000 buffer(0x558e2e11e1b0
space 0x558e2dd76f68 0x7e000~1000 writing nocache)
2019-11-20 17:06:44.507110 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7f000~1000 buffer(0x558e2dfb3830
space 0x558e2dd76f68 0x7f000~1000 writing nocache)
2019-11-20 17:06:44.507113 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2019-11-20 17:06:44.507114 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) fault_range 0x80000~1000
2019-11-20 17:06:44.507115 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _do_write_small 0x80000~1000
2019-11-20   17:06:44.507117   7ff197c0a700   20   bluestore(/sandstone-data/ceph-2)   _do_write_small   considering   Blob(0x558e2dd77650
blob([0x990000~10000,0x9a0000~10000,0x9b0000~10000,0x9c0000~10000,0x9d0000~10000,0x9e0000~10000,0x9f0000~10000,0xa00000~10000]        csum
crc32c/0x1000) use_tracker(0x8*0x10000 0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x558e2dd76f50 sbid 0x0)) bstart
0x0
2019-11-20 17:06:44.507123 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _pad_zeros 0x0~1000 chunk_size 0x1000
2019-11-20 17:06:44.507125 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _pad_zeros pad 0x0 + 0x0 on front/back, now 0x0~1000
2019-11-20 17:06:44.507127 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write txc 0x558e2deb78c0 1 blobs
2019-11-20 17:06:44.507130 7ff197c0a700 10 stupidalloc 0x0x558e2b07c720 allocate_int want_size 0x10000 alloc_unit 0x10000 hint 0x0
2019-11-20 17:06:44.507131 7ff197c0a700 30 stupidalloc 0x0x558e2b07c720 _choose_bin len 0x10000 -> 5
2019-11-20 17:06:44.507134 7ff197c0a700 30 stupidalloc 0x0x558e2b07c720 allocate_int got 0xa10000~10000 from bin 9
2019-11-20 17:06:44.507137 7ff197c0a700 30 stupidalloc 0x0x558e2b07c720 _choose_bin len 0x5fece0000 -> 9
2019-11-20 17:06:44.507138 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write prealloc [0xa10000~10000]
2019-11-20 17:06:44.507140 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write forcing csum_order to block_size_order 12
2019-11-20  17:06:44.507141  7ff197c0a700  20  bluestore(/sandstone-data/ceph-2)  _do_alloc_write  initialize csum setting for new blob
Blob(0x558e30a7aa80 blob([]) use_tracker(0x0 0x0) SharedBlob(0x558e30a7a2a0 sbid 0x0)) csum_type crc32c csum_order 12 csum_length 0x10000
// 重新创建blob
2019-11-20    17:06:44.507145    7ff197c0a700    20    bluestore(/sandstone-data/ceph-2)    _do_alloc_write    blob    Blob(0x558e30a7aa80
blob([0xa10000~10000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20  17:06:44.507149  7ff197c0a700  20  bluestore.blob(0x558e30a7aa80)  get_ref  0x0~1000  Blob(0x558e30a7aa80  blob([0xa10000~10000]
csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x0 0x0) SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.507153 7ff197c0a700 20 bluestore.blob(0x558e30a7aa80) get_ref init 0x10000, 10000

// 第二个blob[0x558e30a7aa80], 创建lextent[0x80000~1000]：logical_offset=0x80000, blob_offset=0x0, length=0x1000. 新生成lextent所属的blob是
新的，与前面不相同不能合并.
2019-11-20   17:06:44.507154   7ff197c0a700   20   bluestore(/sandstone-data/ceph-2)   _do_alloc_write      lex   0x80000~1000:   0x0~1000
Blob(0x558e30a7aa80     blob([0xa10000~10000]     csum+has_unused     crc32c/0x1000     unused=0xfffe)     use_tracker(0x10000     0x1000)
SharedBlob(0x558e30a7a2a0 sbid 0x0))

2019-11-20 17:06:44.507158 7ff197c0a700 20 bluestore.BufferSpace(0x558e30a7a2b8 in 0x558e2b06f420) _discard 0x0~1000
2019-11-20 17:06:44.507160 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write deferring small 0x1000 write via deferred
2019-11-20 17:06:44.507163 7ff197c0a700 30 estimate gc range(hex): [80000, 81000)
2019-11-20 17:06:44.507164 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write extending size to 0x81000
2019-11-20 17:06:44.507166 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) dirty_range 0x80000~1000
2019-11-20 17:06:44.507167 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) dirty_range mark inline shard dirty
2019-11-20         17:06:44.507168         7ff197c0a700         10         bluestore(/sandstone-data/ceph-2)         _write         2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x80000~1000 = 0



// 516k~4k
2019-11-20         17:06:44.511610         7ff197c0a700         15         bluestore(/sandstone-data/ceph-2)         _setattrs         2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys
2019-11-20         17:06:44.511614         7ff197c0a700         10         bluestore(/sandstone-data/ceph-2)         _setattrs         2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys = 0
2019-11-20         17:06:44.511617         7ff197c0a700         15         bluestore(/sandstone-data/ceph-2)         _set_alloc_hint         2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags -
2019-11-20         17:06:44.511620         7ff197c0a700         10         bluestore(/sandstone-data/ceph-2)         _set_alloc_hint         2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags - = 0
2019-11-20         17:06:44.511623         7ff197c0a700         15         bluestore(/sandstone-data/ceph-2)         _write         2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x81000~1000
2019-11-20           17:06:44.511626           7ff197c0a700           20           bluestore(/sandstone-data/ceph-2)           _do_write
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x81000~1000 - have 0x81000 (528384) bytes fadvise_flags 0x0
2019-11-20         17:06:44.511628         7ff197c0a700         30         bluestore(/sandstone-data/ceph-2)         _dump_onode         0x558e2dc0a6c0
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# nid 1278 size 0x81000 (528384) expected_object_size 1048576 expected_write_size
1048576 in 0 shards, 0 spanning blobs
2019-11-20 17:06:44.511631 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode  attr _ len 292
2019-11-20 17:06:44.511632 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode  attr snapset len 35
2019-11-20 17:06:44.511633 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map   0x0~80000: 0x0~80000 Blob(0x558e2dd77650
blob([0x990000~10000,0x9a0000~10000,0x9b0000~10000,0x9c0000~10000,0x9d0000~10000,0x9e0000~10000,0x9f0000~10000,0xa00000~10000]        csum
crc32c/0x1000) use_tracker(0x8*0x10000 0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20    17:06:44.511642    7ff197c0a700    30    bluestore(/sandstone-data/ceph-2)    _dump_extent_map                          csum:
[e8ba7e74,2cab4bfd,f8e16d5e,906def58,8c267066,70148deb,74666a62,b6857eae,5eaa119a,d59b091a,1c153456,aa5d89d6,89244de0,51d7ed58,7f3b84e2,78fa831a,999ce1c,ec
2019-11-20 17:06:44.511652 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x6a000~1000 buffer(0x558e2e169170
space 0x558e2dd76f68 0x6a000~1000 writing nocache)
2019-11-20 17:06:44.511654 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x6b000~1000 buffer(0x558e2e14efc0
space 0x558e2dd76f68 0x6b000~1000 writing nocache)
2019-11-20 17:06:44.511656 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x6c000~1000 buffer(0x558e2dfb3a70
space 0x558e2dd76f68 0x6c000~1000 writing nocache)
2019-11-20 17:06:44.511657 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x6d000~1000 buffer(0x558e2dfb3710
space 0x558e2dd76f68 0x6d000~1000 writing nocache)
2019-11-20 17:06:44.511659 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x6e000~1000 buffer(0x558e2e14fb00
space 0x558e2dd76f68 0x6e000~1000 writing nocache)
2019-11-20 17:06:44.511661 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x6f000~1000 buffer(0x558e2e14e240
space 0x558e2dd76f68 0x6f000~1000 writing nocache)
```

```
2019-11-20 17:06:44.511662 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x70000~1000 buffer(0x558e2b3cb680
space 0x558e2dd76f68 0x70000~1000 writing nocache)
2019-11-20 17:06:44.511664 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x71000~1000 buffer(0x558e2e14e750
space 0x558e2dd76f68 0x71000~1000 writing nocache)
2019-11-20 17:06:44.511665 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x72000~1000 buffer(0x558e2e14e870
space 0x558e2dd76f68 0x72000~1000 writing nocache)
2019-11-20 17:06:44.511667 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x73000~1000 buffer(0x558e2dec2480
space 0x558e2dd76f68 0x73000~1000 writing nocache)
2019-11-20 17:06:44.511668 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x74000~1000 buffer(0x558e2e11e870
space 0x558e2dd76f68 0x74000~1000 writing nocache)
2019-11-20 17:06:44.511670 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x75000~1000 buffer(0x558e2dec2360
space 0x558e2dd76f68 0x75000~1000 writing nocache)
2019-11-20 17:06:44.511671 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x76000~1000 buffer(0x558e2dec2120
space 0x558e2dd76f68 0x76000~1000 writing nocache)
2019-11-20 17:06:44.511672 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x77000~1000 buffer(0x558e2b3cb5f0
space 0x558e2dd76f68 0x77000~1000 writing nocache)
2019-11-20 17:06:44.511674 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x78000~1000 buffer(0x558e2e11e2d0
space 0x558e2dd76f68 0x78000~1000 writing nocache)
2019-11-20 17:06:44.511675 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x79000~1000 buffer(0x558e2e0d9320
space 0x558e2dd76f68 0x79000~1000 writing nocache)
2019-11-20 17:06:44.511676 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7a000~1000 buffer(0x558e2e0026c0
space 0x558e2dd76f68 0x7a000~1000 writing nocache)
2019-11-20 17:06:44.511678 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7b000~1000 buffer(0x558e2dfb25a0
space 0x558e2dd76f68 0x7b000~1000 writing nocache)
2019-11-20 17:06:44.511679 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7c000~1000 buffer(0x558e2dfb27e0
space 0x558e2dd76f68 0x7c000~1000 writing nocache)
2019-11-20 17:06:44.511681 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7d000~1000 buffer(0x558e2e14f050
space 0x558e2dd76f68 0x7d000~1000 writing nocache)
2019-11-20 17:06:44.511682 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7e000~1000 buffer(0x558e2e11e1b0
space 0x558e2dd76f68 0x7e000~1000 writing nocache)
2019-11-20 17:06:44.511683 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7f000~1000 buffer(0x558e2dfb3830
space 0x558e2dd76f68 0x7f000~1000 writing nocache)
2019-11-20 17:06:44.511685 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map  0x80000~1000: 0x0~1000 Blob(0x558e30a7aa80
blob([0xa10000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000) SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.511688 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map                csum:
[9f31bd54,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
2019-11-20 17:06:44.511690 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map        0x0~1000 buffer(0x558e2e11f290 space
0x558e30a7a2b8 0x0~1000 writing nocache)
2019-11-20 17:06:44.511692 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2019-11-20 17:06:44.511694 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) fault_range 0x81000~1000
2019-11-20 17:06:44.511695 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _do_write_small 0x81000~1000
2019-11-20 17:06:44.511697 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small considering Blob(0x558e30a7aa80
blob([0xa10000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000) SharedBlob(0x558e30a7a2a0 sbid 0x0)) bstart
0x80000
2019-11-20 17:06:44.511700 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small  write to unused 0x1000~1000 pad 0x0 + 0x0 of
mutable Blob(0x558e30a7aa80 blob([0xa10000~10000] csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000)
SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.511704 7ff197c0a700 20 bluestore.BufferSpace(0x558e30a7a2b8 in 0x558e2b06f420) _discard 0x1000~1000
2019-11-20 17:06:44.511706 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small deferring small 0x1000 unused write via
deferred
2019-11-20 17:06:44.511709 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small     lex  old  0x80000~1000:  0x0~1000
Blob(0x558e30a7aa80     blob([0xa10000~10000]     csum+has_unused     crc32c/0x1000     unused=0xfffe)     use_tracker(0x10000     0x1000)
SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.511713 7ff197c0a700 20 bluestore.blob(0x558e30a7aa80) get_ref 0x1000~1000 Blob(0x558e30a7aa80 blob([0xa10000~10000]
csum+has_unused crc32c/0x1000 unused=0xfffe) use_tracker(0x10000 0x1000) SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.511716 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small     lex  0x81000~1000:  0x1000~1000
Blob(0x558e30a7aa80     blob([0xa10000~10000]     csum+has_unused     crc32c/0x1000     unused=0xfffc)     use_tracker(0x10000     0x2000)
SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.511720 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write txc 0x558e2deb7340 0 blobs
2019-11-20 17:06:44.511721 7ff197c0a700 30 estimate gc range(hex): [81000, 82000)
2019-11-20 17:06:44.511722 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write extending size to 0x82000
2019-11-20 17:06:44.511724 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) compress_extent_map 0x81000~1000 next shard 0xffffffff
merging 0x80000~1000: 0x0~1000 Blob(0x558e30a7aa80 blob([0xa10000~10000] csum+has_unused crc32c/0x1000 unused=0xfffc) use_tracker(0x10000
0x2000) SharedBlob(0x558e30a7a2a0 sbid 0x0)) and 0x81000~1000: 0x1000~1000 Blob(0x558e30a7aa80 blob([0xa10000~10000] csum+has_unused
crc32c/0x1000 unused=0xfffc) use_tracker(0x10000 0x2000) SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.511729 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) dirty_range 0x81000~1000
2019-11-20 17:06:44.511730 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) dirty_range mark inline shard dirty
2019-11-20 17:06:44.511732 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _write 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x81000~1000 = 0
```

```
// 576k~4k
2019-11-20 17:06:44.580563 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _setattrs 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys
2019-11-20 17:06:44.580567 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _setattrs 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys = 0
2019-11-20 17:06:44.580570 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _set_alloc_hint 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags -
2019-11-20 17:06:44.580573 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _set_alloc_hint 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags - = 0
2019-11-20 17:06:44.580576 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _write 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x90000~1000
2019-11-20 17:06:44.580578 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x90000~1000 - have 0x90000 (589824) bytes fadvise_flags 0x0
2019-11-20 17:06:44.580581 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode 0x558e2dc0a6c0
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# nid 1278 size 0x90000 (589824) expected_object_size 1048576 expected_write_size
1048576 in 0 shards, 0 spanning blobs
2019-11-20 17:06:44.580584 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode  attr _ len 292
2019-11-20 17:06:44.580585 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode  attr snapset len 35
2019-11-20 17:06:44.580586 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map  0x0~80000: 0x0~80000 Blob(0x558e2dd77650
blob([0x990000~10000,0x9a0000~10000,0x9b0000~10000,0x9c0000~10000,0x9d0000~10000,0x9e0000~10000,0x9f0000~10000,0xa00000~10000] csum
crc32c/0x1000) use_tracker(0x8*0x10000 0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:44.580595 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map                csum:
[e8ba7e74,2cab4bfd,f8e16d5e,906def58,8c267066,70148deb,74666a62,b6857eae,5eaa119a,d59b091a,1c153456,aa5d89d6,89244de0,51d7ed58,7f3b84e2,78fa831a,999ce1c,ec
2019-11-20 17:06:44.580606 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x6a000~1000 buffer(0x558e2e169170
space 0x558e2dd76f68 0x6a000~1000 writing nocache)
2019-11-20 17:06:44.580608 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x6b000~1000 buffer(0x558e2e14efc0
space 0x558e2dd76f68 0x6b000~1000 writing nocache)
```

```
2019-11-20 17:06:44.580609 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x6c000~1000 buffer(0x558e2dfb3a70
space 0x558e2dd76f68 0x6c000~1000 writing nocache)
2019-11-20 17:06:44.580611 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x6d000~1000 buffer(0x558e2dfb3710
space 0x558e2dd76f68 0x6d000~1000 writing nocache)
2019-11-20 17:06:44.580612 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x6e000~1000 buffer(0x558e2e14fb00
space 0x558e2dd76f68 0x6e000~1000 writing nocache)
2019-11-20 17:06:44.580613 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x6f000~1000 buffer(0x558e2e14e240
space 0x558e2dd76f68 0x6f000~1000 writing nocache)
2019-11-20 17:06:44.580614 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x70000~1000 buffer(0x558e2b3cb680
space 0x558e2dd76f68 0x70000~1000 writing nocache)
2019-11-20 17:06:44.580616 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x71000~1000 buffer(0x558e2e14e750
space 0x558e2dd76f68 0x71000~1000 writing nocache)
2019-11-20 17:06:44.580617 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x72000~1000 buffer(0x558e2e14e870
space 0x558e2dd76f68 0x72000~1000 writing nocache)
2019-11-20 17:06:44.580619 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x73000~1000 buffer(0x558e2dec2480
space 0x558e2dd76f68 0x73000~1000 writing nocache)
2019-11-20 17:06:44.580621 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x74000~1000 buffer(0x558e2e11e870
space 0x558e2dd76f68 0x74000~1000 writing nocache)
2019-11-20 17:06:44.580622 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x75000~1000 buffer(0x558e2dec2360
space 0x558e2dd76f68 0x75000~1000 writing nocache)
2019-11-20 17:06:44.580624 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x76000~1000 buffer(0x558e2dec2120
space 0x558e2dd76f68 0x76000~1000 writing nocache)
2019-11-20 17:06:44.580625 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x77000~1000 buffer(0x558e2b3cb5f0
space 0x558e2dd76f68 0x77000~1000 writing nocache)
2019-11-20 17:06:44.580626 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x78000~1000 buffer(0x558e2e11e2d0
space 0x558e2dd76f68 0x78000~1000 writing nocache)
2019-11-20 17:06:44.580628 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x79000~1000 buffer(0x558e2e0d9320
space 0x558e2dd76f68 0x79000~1000 writing nocache)
2019-11-20 17:06:44.580629 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7a000~1000 buffer(0x558e2e0026c0
space 0x558e2dd76f68 0x7a000~1000 writing nocache)
2019-11-20 17:06:44.580631 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7b000~1000 buffer(0x558e2dfb25a0
space 0x558e2dd76f68 0x7b000~1000 writing nocache)
2019-11-20 17:06:44.580632 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7c000~1000 buffer(0x558e2dfb27e0
space 0x558e2dd76f68 0x7c000~1000 writing nocache)
2019-11-20 17:06:44.580633 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7d000~1000 buffer(0x558e2e14f050
space 0x558e2dd76f68 0x7d000~1000 writing nocache)
2019-11-20 17:06:44.580634 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7e000~1000 buffer(0x558e2e11e1b0
space 0x558e2dd76f68 0x7e000~1000 writing nocache)
2019-11-20 17:06:44.580635 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7f000~1000 buffer(0x558e2dfb3830
space 0x558e2dd76f68 0x7f000~1000 writing nocache)
2019-11-20 17:06:44.580637 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map       0x80000~10000:   0x0~10000
Blob(0x558e30a7aa80 blob([0xa10000~10000] csum crc32c/0x1000) use_tracker(0x10000 0x10000) SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.580640 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map                        csum:
[9f31bd54,107ed3b1,a502699d,12157b19,f4ec87f0,54f38d6e,ef9fc2f5,c2ed1542,b548706f,39a01d2b,2f05030b,96e224a3,43dcda80,8d7a7eda,24af97f1,ab14bd70]
2019-11-20 17:06:44.580642 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map        0x0~1000 buffer(0x558e2e11f290 space
0x558e30a7a2b8 0x0~1000 writing nocache)
2019-11-20 17:06:44.580643 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x1000~1000 buffer(0x558e2e11f170
space 0x558e30a7a2b8 0x1000~1000 writing nocache)
2019-11-20 17:06:44.580644 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x2000~1000 buffer(0x558e2e11e900
space 0x558e30a7a2b8 0x2000~1000 writing nocache)
2019-11-20 17:06:44.580646 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x3000~1000 buffer(0x558e2e14eb40
space 0x558e30a7a2b8 0x3000~1000 writing nocache)
2019-11-20 17:06:44.580647 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x4000~1000 buffer(0x558e2dfb2750
space 0x558e30a7a2b8 0x4000~1000 writing nocache)
2019-11-20 17:06:44.580648 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x5000~1000 buffer(0x558e2dfb2510
space 0x558e30a7a2b8 0x5000~1000 writing nocache)
2019-11-20 17:06:44.580649 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x6000~1000 buffer(0x558e2e169200
space 0x558e30a7a2b8 0x6000~1000 writing nocache)
2019-11-20 17:06:44.580651 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x7000~1000 buffer(0x558e2e1685a0
space 0x558e30a7a2b8 0x7000~1000 writing nocache)
2019-11-20 17:06:44.580652 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x8000~1000 buffer(0x558e2e0d9b90
space 0x558e30a7a2b8 0x8000~1000 writing nocache)
2019-11-20 17:06:44.580653 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x9000~1000 buffer(0x558e2dfb3b00
space 0x558e30a7a2b8 0x9000~1000 writing nocache)
2019-11-20 17:06:44.580655 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0xa000~1000 buffer(0x558e2e0027e0
space 0x558e30a7a2b8 0xa000~1000 writing nocache)
2019-11-20 17:06:44.580656 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0xb000~1000 buffer(0x558e2e0d8bd0
space 0x558e30a7a2b8 0xb000~1000 writing nocache)
2019-11-20 17:06:44.580658 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0xc000~1000 buffer(0x558e2b3cae10
space 0x558e30a7a2b8 0xc000~1000 writing nocache)
2019-11-20 17:06:44.580659 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0xd000~1000 buffer(0x558e2e11f200
space 0x558e30a7a2b8 0xd000~1000 writing nocache)
2019-11-20 17:06:44.580660 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0xe000~1000 buffer(0x558e2dfb2cf0
space 0x558e30a7a2b8 0xe000~1000 writing nocache)
2019-11-20 17:06:44.580661 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0xf000~1000 buffer(0x558e2e11f830
space 0x558e30a7a2b8 0xf000~1000 writing nocache)
2019-11-20 17:06:44.580663 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2019-11-20 17:06:44.580664 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) fault_range 0x90000~1000
2019-11-20 17:06:44.580665 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _do_write_small 0x90000~1000
2019-11-20 17:06:44.580667 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small considering Blob(0x558e30a7aa80
blob([0xa10000~10000] csum crc32c/0x1000) use_tracker(0x10000 0x10000) SharedBlob(0x558e30a7a2a0 sbid 0x0)) bstart 0x80000
2019-11-20 17:06:44.580673 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _pad_zeros 0x10000~1000 chunk_size 0x1000
2019-11-20 17:06:44.580674 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _pad_zeros pad 0x0 + 0x0 on front/back, now 0x10000~1000
2019-11-20 17:06:44.580676 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small reuse blob Blob(0x558e30a7aa80
blob([0xa10000~10000,!~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,0]) SharedBlob(0x558e30a7a2a0 sbid 0x0)) (0x10000~1000)
(0x10000~1000)
2019-11-20 17:06:44.580680 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write txc 0x558e2def0b00 1 blobs
2019-11-20 17:06:44.580683 7ff197c0a700 10 stupidalloc 0x0x558e2b07c720 allocate_int want_size 0x10000 alloc_unit 0x10000 hint 0x0
2019-11-20 17:06:44.580685 7ff197c0a700 30 stupidalloc 0x0x558e2b07c720 _choose_bin len 0x10000 -> 5
2019-11-20 17:06:44.580687 7ff197c0a700 30 stupidalloc 0x0x558e2b07c720 allocate_int got 0xa20000~10000 from bin 9
2019-11-20 17:06:44.580691 7ff197c0a700 30 stupidalloc 0x0x558e2b07c720 _choose_bin len 0x5fecd0000 -> 9
2019-11-20 17:06:44.580692 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write prealloc [0xa20000~10000]
2019-11-20 17:06:44.580695 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write blob Blob(0x558e30a7aa80
blob([0xa10000~10000,0xa20000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,0]) SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.580700 7ff197c0a700 20 bluestore.blob(0x558e30a7aa80) get_ref 0x10000~1000 Blob(0x558e30a7aa80
blob([0xa10000~10000,0xa20000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,0]) SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.580703 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write lex 0x90000~1000: 0x10000~1000
Blob(0x558e30a7aa80 blob([0xa10000~10000,0xa20000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,1000])
SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.580707 7ff197c0a700 20 bluestore.BufferSpace(0x558e30a7a2b8 in 0x558e2b06f420) _discard 0x10000~1000
2019-11-20 17:06:44.580710 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write deferring small 0x1000 write via deferred
2019-11-20 17:06:44.580712 7ff197c0a700 30 estimate gc range(hex): [90000, 91000)
2019-11-20 17:06:44.580713 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write extending size to 0x91000
```

```
2019-11-20 17:06:44.580716 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) compress_extent_map 0x90000~1000 next shard 0xffffffff
merging 0x80000~10000: 0x0~10000 Blob(0x558e30a7aa80 blob([0xa10000~10000,0xa20000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000
0x[10000,1000]) SharedBlob(0x558e30a7a2a0 sbid 0x0)) and 0x90000~1000: 0x10000~1000 Blob(0x558e30a7aa80
blob([0xa10000~10000,0xa20000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,1000]) SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.580722 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) dirty_range 0x90000~1000
2019-11-20 17:06:44.580724 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) dirty_range mark inline shard dirty
2019-11-20 17:06:44.580725 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _write 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x90000~1000 = 0


// 580k~4k
2019-11-20 17:06:44.605839 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _setattrs 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys
2019-11-20 17:06:44.605845 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _setattrs 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 2 keys = 0
2019-11-20 17:06:44.605849 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _set_alloc_hint 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags -
2019-11-20 17:06:44.605853 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _set_alloc_hint 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# object_size 1048576 write_size 1048576 flags - = 0
2019-11-20 17:06:44.605857 7ff197c0a700 15 bluestore(/sandstone-data/ceph-2) _write 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x91000~1000
2019-11-20 17:06:44.605861 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x91000~1000 - have 0x91000 (593920) bytes fadvise_flags 0x0
2019-11-20 17:06:44.605864 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode 0x558e2dc0a6c0
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# nid 1278 size 0x91000 (593920) expected_object_size 1048576 expected_write_size
1048576 in 0 shards, 0 spanning blobs
2019-11-20 17:06:44.605868 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode  attr _ len 292
2019-11-20 17:06:44.605869 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_onode  attr snapset len 35
2019-11-20 17:06:44.605871 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map  0x0~80000: 0x0~80000 Blob(0x558e2dd77650
blob([0x990000~10000,0x9a0000~10000,0x9b0000~10000,0x9c0000~10000,0x9d0000~10000,0x9e0000~10000,0x9f0000~10000,0xa00000~10000] csum
crc32c/0x1000) use_tracker(0x8*0x10000 0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x558e2dd76f50 sbid 0x0))
2019-11-20 17:06:44.605887 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map csum:
[e8ba7e74,2cab4bfd,f8e16d5e,906def58,8c267066,70148deb,74666a62,b6857eae,5eaa119a,d59b091a,1c153456,aa5d89d6,89244de0,51d7ed58,7f3b84e2,78fa831a,999ce1c,ec
2019-11-20 17:06:44.605905 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x6a000~1000 buffer(0x558e2e169170
space 0x558e2dd76f68 0x6a000~1000 writing nocache)
2019-11-20 17:06:44.605909 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x6b000~1000 buffer(0x558e2e14efc0
space 0x558e2dd76f68 0x6b000~1000 writing nocache)
2019-11-20 17:06:44.605910 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x6c000~1000 buffer(0x558e2dfb3a70
space 0x558e2dd76f68 0x6c000~1000 writing nocache)
2019-11-20 17:06:44.605912 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x6d000~1000 buffer(0x558e2dfb3710
space 0x558e2dd76f68 0x6d000~1000 writing nocache)
2019-11-20 17:06:44.605914 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x6e000~1000 buffer(0x558e2e14fb00
space 0x558e2dd76f68 0x6e000~1000 writing nocache)
2019-11-20 17:06:44.605916 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x6f000~1000 buffer(0x558e2e14e240
space 0x558e2dd76f68 0x6f000~1000 writing nocache)
2019-11-20 17:06:44.605917 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x70000~1000 buffer(0x558e2b3cb680
space 0x558e2dd76f68 0x70000~1000 writing nocache)
2019-11-20 17:06:44.605919 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x71000~1000 buffer(0x558e2e14e750
space 0x558e2dd76f68 0x71000~1000 writing nocache)
2019-11-20 17:06:44.605921 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x72000~1000 buffer(0x558e2e14e870
space 0x558e2dd76f68 0x72000~1000 writing nocache)
2019-11-20 17:06:44.605922 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x73000~1000 buffer(0x558e2dec2480
space 0x558e2dd76f68 0x73000~1000 writing nocache)
2019-11-20 17:06:44.605924 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x74000~1000 buffer(0x558e2e11e870
space 0x558e2dd76f68 0x74000~1000 writing nocache)
2019-11-20 17:06:44.605925 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x75000~1000 buffer(0x558e2dec2360
space 0x558e2dd76f68 0x75000~1000 writing nocache)
2019-11-20 17:06:44.605927 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x76000~1000 buffer(0x558e2dec2120
space 0x558e2dd76f68 0x76000~1000 writing nocache)
2019-11-20 17:06:44.605928 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x77000~1000 buffer(0x558e2b3cb5f0
space 0x558e2dd76f68 0x77000~1000 writing nocache)
2019-11-20 17:06:44.605929 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x78000~1000 buffer(0x558e2e11e2d0
space 0x558e2dd76f68 0x78000~1000 writing nocache)
2019-11-20 17:06:44.605931 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x79000~1000 buffer(0x558e2e0d9320
space 0x558e2dd76f68 0x79000~1000 writing nocache)
2019-11-20 17:06:44.605932 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x7a000~1000 buffer(0x558e2e0026c0
space 0x558e2dd76f68 0x7a000~1000 writing nocache)
2019-11-20 17:06:44.605934 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x7b000~1000 buffer(0x558e2dfb25a0
space 0x558e2dd76f68 0x7b000~1000 writing nocache)
2019-11-20 17:06:44.605936 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x7c000~1000 buffer(0x558e2dfb27e0
space 0x558e2dd76f68 0x7c000~1000 writing nocache)
2019-11-20 17:06:44.605937 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x7d000~1000 buffer(0x558e2e14f050
space 0x558e2dd76f68 0x7d000~1000 writing nocache)
2019-11-20 17:06:44.605939 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x7e000~1000 buffer(0x558e2e11e1b0
space 0x558e2dd76f68 0x7e000~1000 writing nocache)
2019-11-20 17:06:44.605940 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x7f000~1000 buffer(0x558e2dfb3830
space 0x558e2dd76f68 0x7f000~1000 writing nocache)
2019-11-20 17:06:44.605942 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x80000~11000: 0x0~11000
Blob(0x558e30a7aa80 blob([0xa10000~10000,0xa20000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,1000])
SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.605947 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map csum:
[9f31bd54,107ed3b1,a502699d,12157b19,f4ec87f0,54f38d6e,ef9fc2f5,c2ed1542,b548706f,39a01d2b,2f05030b,96e224a3,43dcda80,8d7a7eda,24af97f1,ab14bd70,4be1e504,6
2019-11-20 17:06:44.605950 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x0~1000 buffer(0x558e2e11f290 space
0x558e30a7a2b8 0x0~1000 writing nocache)
2019-11-20 17:06:44.605952 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x1000~1000 buffer(0x558e2e11f170
space 0x558e30a7a2b8 0x1000~1000 writing nocache)
2019-11-20 17:06:44.605954 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x2000~1000 buffer(0x558e2e11e900
space 0x558e30a7a2b8 0x2000~1000 writing nocache)
2019-11-20 17:06:44.605955 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x3000~1000 buffer(0x558e2e14eb40
space 0x558e30a7a2b8 0x3000~1000 writing nocache)
2019-11-20 17:06:44.605957 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x4000~1000 buffer(0x558e2dfb2750
space 0x558e30a7a2b8 0x4000~1000 writing nocache)
2019-11-20 17:06:44.605958 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x5000~1000 buffer(0x558e2dfb2510
space 0x558e30a7a2b8 0x5000~1000 writing nocache)
2019-11-20 17:06:44.605960 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x6000~1000 buffer(0x558e2e169200
space 0x558e30a7a2b8 0x6000~1000 writing nocache)
2019-11-20 17:06:44.605961 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x7000~1000 buffer(0x558e2e1685a0
space 0x558e30a7a2b8 0x7000~1000 writing nocache)
2019-11-20 17:06:44.605963 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x8000~1000 buffer(0x558e2e0d9b90
space 0x558e30a7a2b8 0x8000~1000 writing nocache)
2019-11-20 17:06:44.605965 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0x9000~1000 buffer(0x558e2dfb3b00
space 0x558e30a7a2b8 0x9000~1000 writing nocache)
2019-11-20 17:06:44.605968 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map 0xa000~1000 buffer(0x558e2e0027e0
space 0x558e30a7a2b8 0xa000~1000 writing nocache)
```

```
2019-11-20 17:06:44.605971 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xb000~1000 buffer(0x558e2e0d8bd0
space 0x558e30a7a2b8 0xb000~1000 writing nocache)
2019-11-20 17:06:44.605974 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xc000~1000 buffer(0x558e2b3cae10
space 0x558e30a7a2b8 0xc000~1000 writing nocache)
2019-11-20 17:06:44.605977 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xd000~1000 buffer(0x558e2e11f200
space 0x558e30a7a2b8 0xd000~1000 writing nocache)
2019-11-20 17:06:44.605980 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xe000~1000 buffer(0x558e2dfb2cf0
space 0x558e30a7a2b8 0xe000~1000 writing nocache)
2019-11-20 17:06:44.605983 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map          0xf000~1000 buffer(0x558e2e11f830
space 0x558e30a7a2b8 0xf000~1000 writing nocache)
2019-11-20 17:06:44.605986 7ff197c0a700 30 bluestore(/sandstone-data/ceph-2) _dump_extent_map         0x10000~1000 buffer(0x558e2e0d8f30
space 0x558e30a7a2b8 0x10000~1000 writing nocache)
2019-11-20 17:06:44.605990 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2019-11-20 17:06:44.605994 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) fault_range 0x91000~1000
2019-11-20 17:06:44.605998 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _do_write_small 0x91000~1000
2019-11-20 17:06:44.606033 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small considering Blob(0x558e30a7aa80
blob([0xa10000~10000,0xa20000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,1000]) SharedBlob(0x558e30a7a2a0 sbid 0x0))
bstart 0x80000
2019-11-20 17:06:44.606042 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small reading head 0x0 and tail 0x0
2019-11-20 17:06:44.606047 7ff197c0a700 20 bluestore.BufferSpace(0x558e30a7a2b8 in 0x558e2b06f420) _discard 0x11000~1000
2019-11-20 17:06:44.606056 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small deferred write 0x11000~1000 of mutable
Blob(0x558e30a7aa80 blob([0xa10000~10000,0xa20000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,1000])
SharedBlob(0x558e30a7a2a0 sbid 0x0)) at [0xa21000~1000]
2019-11-20 17:06:44.606065 7ff197c0a700 20 bluestore.blob(0x558e30a7aa80) get_ref 0x11000~1000 Blob(0x558e30a7aa80
blob([0xa10000~10000,0xa20000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,1000]) SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.606072 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write_small lex 0x91000~1000: 0x11000~1000
Blob(0x558e30a7aa80 blob([0xa10000~10000,0xa20000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,2000])
SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.606079 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write txc 0x558e2b3858c0 0 blobs
2019-11-20 17:06:44.606082 7ff197c0a700 30 estimate gc range(hex): [91000, 92000)
2019-11-20 17:06:44.606085 7ff197c0a700 20 bluestore(/sandstone-data/ceph-2) _do_write extending size to 0x92000
2019-11-20 17:06:44.606088 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) compress_extent_map 0x91000~1000 next shard 0xffffffff
merging 0x80000~11000: 0x0~11000 Blob(0x558e30a7aa80 blob([0xa10000~10000,0xa20000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000
0x[10000,2000]) SharedBlob(0x558e30a7a2a0 sbid 0x0)) and 0x91000~1000: 0x11000~1000 Blob(0x558e30a7aa80
blob([0xa10000~10000,0xa20000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,2000]) SharedBlob(0x558e30a7a2a0 sbid 0x0))
2019-11-20 17:06:44.606100 7ff197c0a700 30 bluestore.extentmap(0x558e2dc0a7b0) dirty_range 0x91000~1000
2019-11-20 17:06:44.606102 7ff197c0a700 20 bluestore.extentmap(0x558e2dc0a7b0) dirty_range mark inline shard dirty
2019-11-20 17:06:44.606106 7ff197c0a700 10 bluestore(/sandstone-data/ceph-2) _write 2.b_head
#2:d192e602:::rbd_data.101f241b25c.0000000000000000:head# 0x91000~1000 = 0
```

2、EC场景

```
```

3、small write总结：[off,len]肯定是是在某个最小分配单元内
(1) 初次写**首先创建blob并申请物理空间**, off和len按block对齐。len <= prefer_deferred_size时defered write，反之是构造aio write。
举例：初次写[0,2k]或初次写[0~4k]都会分配一个blob和对应的物理空间，相应的[0~4k]标记为已使用，其他标记为未使用

(2) [off,len]空间已经分配且未使用，**复用blob**，则按block对齐后直接写到其**已分配未使用**的block中。len <= prefer_deferred_size时defered write，反之是构造aio write。
举例：在[0~4k]已写过的基础上，写[4k~8k]，直接复用blob写到未使用的空间上

(3) [off,len]空间已经分配但是已经写过(完全覆盖写)，**复用blob**，构造defered write(**如果是非block对齐则需要先进行RMW**)。
举例：在[0~4k]已写过的基础上，覆盖写[0~4k]，小块的覆盖写都构造defered write
注意一种情况：
初次写[64k~68k]会首先申请一片64k的物理空间，只标记[0~64k]为未使用，而[64k~128k]被标记为已使用，所以后续无论是初次写还是覆盖写[64k~128k]区间都是按照覆盖写的逻辑来处理。这一有点奇怪？？

(4) [off,len]空间没有分配，并且blob中分配的空间已使用完(unuse位图全置位)，**则复用blob**，申请新的物理空间扩充blob，重新初始化unuse位图
举例：在[0~64k]已写过的基础上，现在写[64k~68k],此时会申请新的物理空间(min_alloc_size)并向后扩充当前复用的blob；
在[64k~128k]已写过的基础上，现在写[0~4k],此时会申请新的物理空间(min_alloc_size)并向前扩充当前复用的blob

(5) [off,len]空间没有分配，但是blob中仍然有未使用的空间，此时不能复用blob(unuse位图已初始化)，重新分配一个new blob
举例：初次写[0~4k]会分配一个blob和对应的物理空间，但是物理空间[4k~60k]还未使用，此时写[64k~4k]不能复用已有的blob

4、blob复用总结
(1) blob已经compressed 或 shared时不允许复用
(2) 开启crc，但是off和len非csum_chunk_size对齐时不允许复用
(3) 写的范围超过了max_blob_size大小不允许复用
(4) blob中仍然有未使用的空间不允许复用
举例：先写[0~4k]，然后写[64k~68k]
(5) 新写范围完全覆盖blob的逻辑空间，
举例：先写[64k~68k]或者[64k~128k]或者[0~128k], 然后写[0~128k]

通过上述方式可以搜集到Bluestore在写入数据时，object的数据分配和映射过程，可以帮助理解其实现。

```
///////[8k~4k]
2020-06-24 10:24:24.294063 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_write
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x2000~1000 - have 0x0 (0) bytes fadvise_flags 0x0
```

```
2020-06-24 10:24:24.294067 7fb5ce895700 30 _dump_onode 0x559d0f129200 #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# nid 20045
size 0x0 (0) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2020-06-24 10:24:24.294071 7fb5ce895700 30 _dump_onode  attr _ len 293
2020-06-24 10:24:24.294073 7fb5ce895700 30 _dump_onode  attr snapset len 35
2020-06-24 10:24:24.294075 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2020-06-24 10:24:24.294078 7fb5ce895700 30 bluestore.extentmap(0x559d0f1292f0) fault_range 0x2000~1000
2020-06-24 10:24:24.294081 7fb5ce895700 10 bluestore(/sandstone-data/ceph-4) _do_write_small 0x2000~1000
2020-06-24 10:24:24.294083 7fb5ce895700 30 bluestore.extentmap(0x559d0f1292f0) fault_range 0x0~82000
2020-06-24 10:24:24.294086 7fb5ce895700 30 bluestore(/sandstone-data/ceph-4) _pad_zeros 0x2000~1000 chunk_size 0x1000
2020-06-24 10:24:24.294088 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _pad_zeros pad 0x0 + 0x0 on front/back, now 0x2000~1000
2020-06-24 10:24:24.294092 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x559d0eb2d8c0 1 blobs
2020-06-24    10:24:24.294094    7fb5ce895700    20    bluestore(/sandstone-data/ceph-4)    _do_alloc_write        buffered:0    compress:0
target_blob_size:524288 write_flags:0
write_item:    blob  0x559d0f4b41c8  logical_offset:8192  blob_length:65536  b_off:8192  b_off0:8192  length0:4096  mark_unused:1  new_blob:1
compressed_len:0

2020-06-24 10:24:24.294100 7fb5ce895700 10 fbmap_alloc 0x559d0ec97400 allocate 0x10000/10000,10000,0
2020-06-24 10:24:24.294104 7fb5ce895700 10 fbmap_alloc 0x559d0ec97400 allocate extent: 0x24e0000~10000/10000,10000,0
2020-06-24 10:24:24.294106 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0x24e0000~10000]
2020-06-24 10:24:24.294109 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write forcing csum_order to block_size_order 12
2020-06-24  10:24:24.294110  7fb5ce895700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write  initialize  csum  setting  for  new  blob
Blob(0x559d0f41dab0 blob([]) use_tracker(0x0 0x0) SharedBlob(0x559d0f3fa4d0 sbid 0x0)) csum_type crc32c csum_order 12 csum_length 0x10000
2020-06-24    10:24:24.294117    7fb5ce895700    20    bluestore(/sandstone-data/ceph-4)    _do_alloc_write    blob    Blob(0x559d0f41dab0
blob([0x24e0000~10000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x559d0f3fa4d0 sbid 0x0))
2020-06-24 10:24:24.294124 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write  b_off=8192 b_end=12288 blob_length 65536
2020-06-24 10:24:24.294126 7fb5ce895700 20 bluestore.blob(0x559d0f41dab0) get_ref 0x2000~1000 Blob(0x559d0f41dab0 blob([0x24e0000~10000]
csum+has_unused crc32c/0x1000 unused=0xfffb) use_tracker(0x0 0x0) SharedBlob(0x559d0f3fa4d0 sbid 0x0))
2020-06-24 10:24:24.294130 7fb5ce895700 20 bluestore.blob(0x559d0f41dab0) get_ref init 0x10000, 10000
2020-06-24  10:24:24.294132  7fb5ce895700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write     lex  0x2000~1000:  0x2000~1000
Blob(0x559d0f41dab0     blob([0x24e0000~10000]     csum+has_unused     crc32c/0x1000     unused=0xfffb)     use_tracker(0x10000     0x1000)
SharedBlob(0x559d0f3fa4d0 sbid 0x0))
2020-06-24 10:24:24.294137 7fb5ce895700 20 bluestore.BufferSpace(0x559d0f3fa4e8 in 0x559d0eb2a1c0) _discard 0x2000~1000
2020-06-24 10:24:24.294140 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write deferring small 0x1000 write via deferred
2020-06-24 10:24:24.294144 7fb5ce895700 30 estimate gc range(hex): [2000, 3000)
2020-06-24 10:24:24.294146 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_write extending size to 0x3000
2020-06-24 10:24:24.294149 7fb5ce895700 30 bluestore.extentmap(0x559d0f1292f0) dirty_range 0x2000~1000
2020-06-24 10:24:24.294150 7fb5ce895700 20 bluestore.extentmap(0x559d0f1292f0) dirty_range mark inline shard dirty
2020-06-24        10:24:24.294152        7fb5ce895700        10        bluestore(/sandstone-data/ceph-4)        _write        2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x2000~1000 = 0


//////[70k~4k]
2020-06-24           10:24:25.060005           7fb5ce895700           20           bluestore(/sandstone-data/ceph-4)           _do_write
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x11800~1000 - have 0x3000 (12288) bytes fadvise_flags 0x0
2020-06-24 10:24:25.060010 7fb5ce895700 30 _dump_onode 0x559d0f129200 #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# nid 20045
size 0x3000 (12288) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2020-06-24 10:24:25.060016 7fb5ce895700 30 _dump_onode  attr _ len 293
2020-06-24 10:24:25.060018 7fb5ce895700 30 _dump_onode  attr snapset len 35
2020-06-24  10:24:25.060020  7fb5ce895700  30  _dump_extent_map    0x2000~1000:  0x2000~1000  Blob(0x559d0f41dab0  blob([0x24e0000~10000]
csum+has_unused crc32c/0x1000 unused=0xfffb) use_tracker(0x10000 0x1000) SharedBlob(0x559d0f3fa4d0 sbid 0x0))
2020-06-24 10:24:25.060043 7fb5ce895700 30 _dump_extent_map      csum: [0,0,1a7a4a9c,0,0,0,0,0,0,0,0,0,0,0,0,0]
2020-06-24 10:24:25.060048 7fb5ce895700 30 _dump_extent_map          0x2000~1000 buffer(0x559d0ec6cf30 space 0x559d0f3fa4e8 0x2000~1000
writing nocache)
2020-06-24 10:24:25.060054 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2020-06-24 10:24:25.060058 7fb5ce895700 30 bluestore.extentmap(0x559d0f1292f0) fault_range 0x11800~1000
2020-06-24 10:24:25.060063 7fb5ce895700 10 bluestore(/sandstone-data/ceph-4) _do_write_small 0x11800~1000
2020-06-24 10:24:25.060067 7fb5ce895700 30 bluestore.extentmap(0x559d0f1292f0) fault_range 0x0~91800
2020-06-24   10:24:25.060071   7fb5ce895700   20   bluestore(/sandstone-data/ceph-4)   _do_write_small   considering   Blob(0x559d0f41dab0
blob([0x24e0000~10000]  csum+has_unused  crc32c/0x1000  unused=0xfffb)  use_tracker(0x10000  0x1000)  SharedBlob(0x559d0f3fa4d0  sbid  0x0))
bstart 0x0
2020-06-24 10:24:25.060079 7fb5ce895700 30 bluestore(/sandstone-data/ceph-4) _pad_zeros 0x1800~1000 chunk_size 0x1000
2020-06-24 10:24:25.060089 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _pad_zeros pad 0x800 + 0x800 on front/back, now 0x1000~2000
2020-06-24 10:24:25.060098 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x559d0fa44000 1 blobs
2020-06-24    10:24:25.060101    7fb5ce895700    20    bluestore(/sandstone-data/ceph-4)    _do_alloc_write        buffered:0    compress:0
target_blob_size:524288 write_flags:0
write_item:    blob  0x559d0ec99348  logical_offset:71680  blob_length:65536  b_off:4096  b_off0:6144  length0:4096  mark_unused:1  new_blob:1
compressed_len:0

2020-06-24 10:24:25.060107 7fb5ce895700 10 fbmap_alloc 0x559d0ec97400 allocate 0x10000/10000,10000,0
2020-06-24 10:24:25.060113 7fb5ce895700 10 fbmap_alloc 0x559d0ec97400 allocate extent: 0x24f0000~10000/10000,10000,0
2020-06-24 10:24:25.060116 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0x24f0000~10000]
2020-06-24 10:24:25.060119 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write forcing csum_order to block_size_order 12
2020-06-24 10:24:25.060121 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write forcing blob_offset to 0x11000
2020-06-24  10:24:25.060139  7fb5ce895700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write  initialize  csum  setting  for  new  blob
Blob(0x559d0f46ccb0 blob([]) use_tracker(0x0 0x0) SharedBlob(0x559d0f354000 sbid 0x0)) csum_type crc32c csum_order 12 csum_length 0x20000
2020-06-24    10:24:25.060150    7fb5ce895700    20    bluestore(/sandstone-data/ceph-4)    _do_alloc_write    blob    Blob(0x559d0f46ccb0
blob([!~10000,0x24f0000~10000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x559d0f354000 sbid 0x0))
2020-06-24 10:24:25.060161 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write  b_off=69632 b_end=77824 blob_length 65536
2020-06-24     10:24:25.060164     7fb5ce895700     20     bluestore.blob(0x559d0f46ccb0)     get_ref     0x11800~1000     Blob(0x559d0f46ccb0
blob([!~10000,0x24f0000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x0 0x0) SharedBlob(0x559d0f354000 sbid 0x0))
2020-06-24 10:24:25.060171 7fb5ce895700 20 bluestore.blob(0x559d0f46ccb0) get_ref init 0x20000, 10000
2020-06-24  10:24:25.060175  7fb5ce895700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write     lex  0x11800~1000:  0x11800~1000
Blob(0x559d0f46ccb0  blob([!~10000,0x24f0000~10000]  csum+has_unused  crc32c/0x1000  unused=0xff)  use_tracker(0x2*0x10000  0x[0,1000])
SharedBlob(0x559d0f354000 sbid 0x0))
2020-06-24 10:24:25.060182 7fb5ce895700 20 bluestore.BufferSpace(0x559d0f354018 in 0x559d0eb2a1c0) _discard 0x11000~2000
2020-06-24 10:24:25.060188 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write deferring small 0x2000 write via deferred
2020-06-24 10:24:25.060193 7fb5ce895700 30 estimate gc range(hex): [11800, 12800)
2020-06-24 10:24:25.060195 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_write extending size to 0x12800
2020-06-24 10:24:25.060199 7fb5ce895700 30 bluestore.extentmap(0x559d0f1292f0) dirty_range 0x11800~1000
2020-06-24 10:24:25.060201 7fb5ce895700 20 bluestore.extentmap(0x559d0f1292f0) dirty_range mark inline shard dirty
2020-06-24        10:24:25.060204        7fb5ce895700        10        bluestore(/sandstone-data/ceph-4)        _write        2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x11800~1000 = 0


//////////[0~64k]
2020-06-24           10:24:25.884748           7fb5ce895700           20           bluestore(/sandstone-data/ceph-4)           _do_write
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x0~10000 - have 0x12800 (75776) bytes fadvise_flags 0x0
2020-06-24 10:24:25.884752 7fb5ce895700 30 _dump_onode 0x559d0f129200 #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# nid 20045
size 0x12800 (75776) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2020-06-24 10:24:25.884756 7fb5ce895700 30 _dump_onode  attr _ len 293
2020-06-24 10:24:25.884767 7fb5ce895700 30 _dump_onode  attr snapset len 35
2020-06-24  10:24:25.884768  7fb5ce895700  30  _dump_extent_map    0x2000~1000:  0x2000~1000  Blob(0x559d0f41dab0  blob([0x24e0000~10000]
csum+has_unused crc32c/0x1000 unused=0xfffb) use_tracker(0x10000 0x1000) SharedBlob(0x559d0f3fa4d0 sbid 0x0))
```

```
2020-06-24 10:24:25.884780 7fb5ce895700 30 _dump_extent_map      csum: [0,0,1a7a4a9c,0,0,0,0,0,0,0,0,0,0,0,0,0]
2020-06-24 10:24:25.884782 7fb5ce895700 30 _dump_extent_map          0x2000~1000 buffer(0x559d0ec6cf30 space 0x559d0f3fa4e8 0x2000~1000
writing nocache)
2020-06-24 10:24:25.884785 7fb5ce895700 30 _dump_extent_map  0x11800~1000: 0x11800~1000 Blob(0x559d0f46ccb0 blob([!~10000,0x24f0000~10000]
csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,1000]) SharedBlob(0x559d0f354000 sbid 0x0))
2020-06-24       10:24:25.884791       7fb5ce895700       30       _dump_extent_map                                                    csum:
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,18d9a92f,1cde6b1,0,0,0,0,0,0,0,0,0,0,0,0]
2020-06-24 10:24:25.884794 7fb5ce895700 30 _dump_extent_map         0x11000~2000 buffer(0x559d0ec6d320 space 0x559d0f354018 0x11000~2000
writing nocache)
2020-06-24 10:24:25.884798 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2020-06-24 10:24:25.884800 7fb5ce895700 30 bluestore.extentmap(0x559d0f1292f0) fault_range 0x0~10000
2020-06-24 10:24:25.884804 7fb5ce895700 10 bluestore(/sandstone-data/ceph-4) _do_write_big 0x0~10000 target_blob_size 0x80000 compress 0
2020-06-24 10:24:25.884819 7fb5ce895700 20 bluestore.blob(0x559d0f41dab0) put_ref 0x2000~1000 Blob(0x559d0f41dab0 blob([0x24e0000~10000]
csum+has_unused crc32c/0x1000 unused=0xfffb) use_tracker(0x10000 0x1000) SharedBlob(0x559d0f3fa4d0 sbid 0x0))
2020-06-24   10:24:25.884841   7fb5ce895700   20   bluestore(/sandstone-data/ceph-4)   _do_write_big   reuse   blob   Blob(0x559d0f46ccb0
blob([!~10000,0x24f0000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,1000]) SharedBlob(0x559d0f354000
sbid 0x0)) (0x0~10000)
2020-06-24 10:24:25.884847 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x559d0fa44580 1 blobs
2020-06-24    10:24:25.884849    7fb5ce895700    20    bluestore(/sandstone-data/ceph-4)    _do_alloc_write        buffered:0    compress:0
target_blob_size:524288 write_flags:0
write_item:     blob   0x559d0f51f968   logical_offset:0   blob_length:65536   b_off:0   b_off0:0   length0:65536   mark_unused:0   new_blob:0
compressed_len:0

2020-06-24 10:24:25.884855 7fb5ce895700 10 fbmap_alloc 0x559d0ec97400 allocate 0x10000/10000,10000,0
2020-06-24 10:24:25.884860 7fb5ce895700 10 fbmap_alloc 0x559d0ec97400 allocate extent: 0x2500000~10000/10000,10000,0
2020-06-24 10:24:25.884862 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0x2500000~10000]
2020-06-24    10:24:25.884867    7fb5ce895700    20    bluestore(/sandstone-data/ceph-4)    _do_alloc_write    blob    Blob(0x559d0f46ccb0
blob([0x2500000~10000,0x24f0000~10000]     csum+has_unused     crc32c/0x1000     unused=0xff)     use_tracker(0x2*0x10000     0x[0,1000])
SharedBlob(0x559d0f354000 sbid 0x0))
2020-06-24     10:24:25.884880     7fb5ce895700     20     bluestore.blob(0x559d0f46ccb0)     get_ref     0x0~10000     Blob(0x559d0f46ccb0
blob([0x2500000~10000,0x24f0000~10000]     csum+has_unused     crc32c/0x1000     unused=0xff)     use_tracker(0x2*0x10000     0x[0,1000])
SharedBlob(0x559d0f354000 sbid 0x0))
2020-06-24 10:24:25.884885 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write  lex 0x0~10000: 0x0~10000 Blob(0x559d0f46ccb0
blob([0x2500000~10000,0x24f0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,1000]) SharedBlob(0x559d0f354000 sbid 0x0))
2020-06-24 10:24:25.884889 7fb5ce895700 20 bluestore.BufferSpace(0x559d0f354018 in 0x559d0eb2a1c0) _discard 0x0~10000
2020-06-24 10:24:25.884898 7fb5ce895700 30 estimate gc range(hex): [0, 10000)
2020-06-24   10:24:25.884900   7fb5ce895700   20   bluestore(/sandstone-data/ceph-4)   _wctx_finish   lex_old   0x2000~1000:   0x2000~1000
Blob(0x559d0f41dab0 blob([!~10000] csum+has_unused crc32c/0x1000 unused=0xfffb) use_tracker(0x10000 0x0) SharedBlob(0x559d0f3fa4d0 sbid
0x0))
2020-06-24 10:24:25.884904 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _wctx_finish  blob release [0x24e0000~10000]
2020-06-24 10:24:25.884906 7fb5ce895700 20 bluestore.blob(0x559d0f41dab0) discard_unallocated 0x0~10000
2020-06-24 10:24:25.884907 7fb5ce895700 20 bluestore.BufferSpace(0x559d0f3fa4e8 in 0x559d0eb2a1c0) _discard 0x0~10000
2020-06-24 10:24:25.884910 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _wctx_finish  release 0x24e0000~10000
2020-06-24 10:24:25.884913 7fb5ce895700 30 bluestore.extentmap(0x559d0f1292f0) dirty_range 0x0~10000
2020-06-24 10:24:25.884914 7fb5ce895700 20 bluestore.extentmap(0x559d0f1292f0) dirty_range mark inline shard dirty
2020-06-24        10:24:25.884917        7fb5ce895700        10        bluestore(/sandstone-data/ceph-4)        _write        2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x0~10000 = 0
```

5、空间释放

```
/var/lib/ceph/bin/ceph daemon osd.4 config set debug_bluestore 30 ; fio -filename=/dev/sdd -direct=1 -iodepth 1 -thread -rw=write -
ioengine=libaio -bs=4k -size=4k -offset=1088k -numjobs=1   -group_reporting -name=test ; fio -filename=/dev/sdd -direct=1 -iodepth 1 -
thread -rw=write -ioengine=libaio -bs=2k -size=2k -offset=1086k -numjobs=1   -group_reporting -name=test; sg_unmap --lba=2176 --num=8
/dev/sdd ;  /var/lib/ceph/bin/ceph daemon osd.4 config set debug_bluestore 1;
写[64k,4k]
写[62k,2k]
unmap[64k,4k]


2020-06-02        11:54:44.928164        7f6204ab8700        15        bluestore(/sandstone-data/ceph-4)        _touch        2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head#
2020-06-02 11:54:44.928168 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _assign_nid 2614
2020-06-02        11:54:44.928169        7f6204ab8700        10        bluestore(/sandstone-data/ceph-4)        _touch        2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# = 0
2020-06-02        11:54:44.928174        7f6204ab8700        15        bluestore(/sandstone-data/ceph-4)        _setattrs        2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 2 keys
2020-06-02        11:54:44.928196        7f6204ab8700        10        bluestore(/sandstone-data/ceph-4)        _setattrs        2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 2 keys = 0
2020-06-02       11:54:44.928201       7f6204ab8700       15       bluestore(/sandstone-data/ceph-4)       _set_alloc_hint       2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# object_size 1048576 write_size 1048576 flags -
2020-06-02       11:54:44.928205       7f6204ab8700       10       bluestore(/sandstone-data/ceph-4)       _set_alloc_hint       2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# object_size 1048576 write_size 1048576 flags - = 0
2020-06-02        11:54:44.928209        7f6204ab8700        15        bluestore(/sandstone-data/ceph-4)        _write        2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x10000~1000
2020-06-02           11:54:44.928212           7f6204ab8700           20           bluestore(/sandstone-data/ceph-4)           _do_write
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x10000~1000 - have 0x0 (0) bytes fadvise_flags 0x0
2020-06-02 11:54:44.928215 7f6204ab8700 30 _dump_onode 0x561bc188f8c0 #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# nid 2614
size 0x0 (0) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2020-06-02 11:54:44.928220 7f6204ab8700 30 _dump_onode  attr _ len 293
2020-06-02 11:54:44.928221 7f6204ab8700 30 _dump_onode  attr snapset len 35
2020-06-02 11:54:44.928223 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2020-06-02 11:54:44.928225 7f6204ab8700 30 bluestore.extentmap(0x561bc188f9b0) fault_range 0x10000~1000
2020-06-02 11:54:44.928234 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _do_write_small 0x10000~1000
2020-06-02 11:54:44.928236 7f6204ab8700 30 bluestore.extentmap(0x561bc188f9b0) fault_range 0x0~90000
2020-06-02 11:54:44.928239 7f6204ab8700 30 bluestore(/sandstone-data/ceph-4) _pad_zeros 0x0~1000 chunk_size 0x1000
2020-06-02 11:54:44.928241 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _pad_zeros pad 0x0 + 0x0 on front/back, now 0x0~1000
2020-06-02 11:54:44.928245 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x561bc070a2c0 1 blobs
2020-06-02 11:54:44.928249 7f6204ab8700 10 fbmap_alloc 0x561bbfb53500 allocate 0x10000/10000,10000,0
2020-06-02 11:54:44.928253 7f6204ab8700 10 fbmap_alloc 0x561bbfb53500 allocate extent: 0x1130000~10000/10000,10000,0
2020-06-02 11:54:44.928256 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0x1130000~10000]
2020-06-02 11:54:44.928258 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write forcing csum_order to block_size_order 12
2020-06-02 11:54:44.928260 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write forcing blob_offset to 0x10000
2020-06-02  11:54:44.928261  7f6204ab8700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write  initialize  csum  setting  for  new  blob
Blob(0x561bc0756700 blob([]) use_tracker(0x0 0x0) SharedBlob(0x561bc0757a40 sbid 0x0)) csum_type crc32c csum_order 12 csum_length 0x20000
2020-06-02    11:54:44.928267    7f6204ab8700    20    bluestore(/sandstone-data/ceph-4)    _do_alloc_write    blob    Blob(0x561bc0756700
blob([!~10000,0x1130000~10000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x561bc0757a40 sbid 0x0))
```

```
2020-06-02    11:54:44.928274    7f6204ab8700    20    bluestore.blob(0x561bc0756700)    get_ref    0x10000~1000    Blob(0x561bc0756700
blob([!~10000,0x1130000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x0 0x0) SharedBlob(0x561bc0757a40 sbid 0x0))
2020-06-02 11:54:44.928278 7f6204ab8700 20 bluestore.blob(0x561bc0756700) get_ref init 0x20000, 10000
2020-06-02  11:54:44.928281  7f6204ab8700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write    lex  0x10000~1000:  0x10000~1000
Blob(0x561bc0756700  blob([!~10000,0x1130000~10000]  csum+has_unused  crc32c/0x1000  unused=0xff)  use_tracker(0x2*0x10000  0x[0,1000])
SharedBlob(0x561bc0757a40 sbid 0x0))
2020-06-02 11:54:44.928286 7f6204ab8700 20 bluestore.BufferSpace(0x561bc0757a58 in 0x561bbfe5c620) _discard 0x10000~1000
2020-06-02 11:54:44.928289 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write deferring small 0x1000 write via deferred
2020-06-02 11:54:44.928293 7f6204ab8700 30 estimate gc range(hex): [10000, 11000)
2020-06-02 11:54:44.928295 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _do_write extending size to 0x11000
2020-06-02 11:54:44.928297 7f6204ab8700 30 bluestore.extentmap(0x561bc188f9b0) dirty_range 0x10000~1000
2020-06-02 11:54:44.928299 7f6204ab8700 20 bluestore.extentmap(0x561bc188f9b0) dirty_range mark inline shard dirty
2020-06-02        11:54:44.928301        7f6204ab8700        10        bluestore(/sandstone-data/ceph-4)        _write        2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x10000~1000 = 0
2020-06-02 11:54:44.928304 7f6204ab8700 30 bluestore.OnodeSpace(0x561bc20eb488 in 0x561bbfe5c620) lookup
2020-06-02  11:54:44.928306  7f6204ab8700  30  bluestore.OnodeSpace(0x561bc20eb488  in  0x561bbfe5c620)  lookup  #2:34000000::::head#  hit
0x561bc05dd200
2020-06-02 11:54:44.928310 7f6204ab8700 15 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000::::head#
2020-06-02                11:54:44.928313                7f6204ab8700                30                bluestore(/sandstone-data/ceph-4)
_omap_setkeys  0x0000000000000466'.0000000129.00000000000000000005' <- 0000000129.00000000000000000005
2020-06-02 11:54:44.928326 7f6204ab8700 30 bluestore(/sandstone-data/ceph-4) _omap_setkeys  0x0000000000000466'._fastinfo' <- _fastinfo
2020-06-02 11:54:44.928330 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000::::head# = 0
2020-06-02 11:54:44.928333 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _txc_calc_cost 0x561bc070a2c0 cost 675458 (1 ios * 670000 +
5458 bytes)
2020-06-02 11:54:44.928335 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _txc_write_nodes txc 0x561bc070a2c0 onodes 0x561bc188f8c0
shared_blobs
2020-06-02           11:54:44.928337           7f6204ab8700           20           bluestore.extentmap(0x561bc188f9b0)           update
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head#
2020-06-02 11:54:44.928343 7f6204ab8700 20 bluestore.extentmap(0x561bc188f9b0) update  inline shard 158 bytes from 1 extents
2020-06-02        11:54:44.928348        7f6204ab8700        20        bluestore(/sandstone-data/ceph-4)                    onode
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# is 543 (379 bytes onode + 2 bytes spanning blobs + 162 bytes inline extents)
2020-06-02  11:54:44.928364  7f6204ab8700  20  bluestore(/sandstone-data/ceph-4)  _txc_finalize_kv  txc  0x561bc070a2c0  allocated
0x[1130000~10000] released 0x[]
2020-06-02 11:54:44.928368 7f6204ab8700 10 freelist allocate 0x1130000~10000
2020-06-02 11:54:44.928370 7f6204ab8700 20 freelist _xor first_key 0x1000000 last_key 0x1000000
2020-06-02 11:54:44.928371 7f6204ab8700 30 freelist _xor 0x1000000: 00000000  00 00 08 00 00 00 00 00  00 00 00 00 00 00 00 00
00 |................|
2020-06-02 11:54:44.928382 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _txc_state_proc txc 0x561bc070a2c0 prepare
2020-06-02 11:54:44.928384 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _txc_finish_io 0x561bc070a2c0
2020-06-02 11:54:44.928385 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _txc_state_proc txc 0x561bc070a2c0 io_done
2020-06-02 11:54:44.928426 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _kv_sync_thread wake
2020-06-02 11:54:44.928435 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _kv_sync_thread committing 1 submitting 1 deferred done 0
stable 0
2020-06-02 11:54:44.928438 7f62122d3700 30 bluestore(/sandstone-data/ceph-4) _kv_sync_thread committing <0x561bc070a2c0>
2020-06-02 11:54:44.928440 7f62122d3700 30 bluestore(/sandstone-data/ceph-4) _kv_sync_thread submitting <0x561bc070a2c0>
2020-06-02 11:54:44.928441 7f62122d3700 30 bluestore(/sandstone-data/ceph-4) _kv_sync_thread deferred_done <>
2020-06-02 11:54:44.928442 7f62122d3700 30 bluestore(/sandstone-data/ceph-4) _kv_sync_thread deferred_stable <>
2020-06-02 11:54:44.928527 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _txc_applied_kv onode 0x561bc188f8c0 had 1
2020-06-02 11:54:44.928531 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _txc_applied_kv onode 0x561bc05dd200 had 1
2020-06-02 11:54:44.935035 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _kv_sync_thread committed 1 cleaned 0 in 0.006596 (0.000006
flush + 0.006590 kv commit)
2020-06-02 11:54:44.935061 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _kv_sync_thread releasing old bluefs 0x[]
2020-06-02 11:54:44.935065 7f62122d3700 10 fbmap_alloc 0x561bbfb53500 release done


2020-06-02        11:54:45.552609        7f6204ab8700        15        bluestore(/sandstone-data/ceph-4)        _write        2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0xf800~800
2020-06-02           11:54:45.552613           7f6204ab8700           20           bluestore(/sandstone-data/ceph-4)           _do_write
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0xf800~800 - have 0x11000 (69632) bytes fadvise_flags 0x0
2020-06-02 11:54:45.552617 7f6204ab8700 30 _dump_onode 0x561bc188f8c0 #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# nid 2614
size 0x11000 (69632) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2020-06-02 11:54:45.552621 7f6204ab8700 30 _dump_onode  attr _ len 293
2020-06-02 11:54:45.552622 7f6204ab8700 30 _dump_onode  attr snapset len 35
2020-06-02 11:54:45.552623 7f6204ab8700 30 _dump_extent_map  0x10000~1000: 0x10000~1000 Blob(0x561bc0756700 blob([!~10000,0x1130000~10000]
csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,1000]) SharedBlob(0x561bc0757a40 sbid 0x0))
2020-06-02       11:54:45.552632       7f6204ab8700       30       _dump_extent_map                                                    csum:
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,86deadaf,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
2020-06-02 11:54:45.552635 7f6204ab8700 30 _dump_extent_map         0x10000~1000 buffer(0x561bc0f6fef0 space 0x561bc0757a58 0x10000~1000
writing nocache)
2020-06-02 11:54:45.552638 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2020-06-02 11:54:45.552641 7f6204ab8700 30 bluestore.extentmap(0x561bc188f9b0) fault_range 0xf800~800
2020-06-02 11:54:45.552644 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _do_write_small 0xf800~800
2020-06-02 11:54:45.552646 7f6204ab8700 30 bluestore.extentmap(0x561bc188f9b0) fault_range 0x0~8f800
2020-06-02  11:54:45.552647  7f6204ab8700  20  bluestore(/sandstone-data/ceph-4)  _do_write_small  considering  Blob(0x561bc0756700
blob([!~10000,0x1130000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,1000]) SharedBlob(0x561bc0757a40
sbid 0x0)) bstart 0x0
2020-06-02 11:54:45.552651 7f6204ab8700 30 bluestore.extentmap(0x561bc188f9b0) fault_range 0xf000~1000
2020-06-02 11:54:45.552655 7f6204ab8700 30 bluestore(/sandstone-data/ceph-4) _pad_zeros 0xf800~800 chunk_size 0x1000
2020-06-02 11:54:45.552658 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _pad_zeros pad 0x800 + 0x0 on front/back, now 0xf000~1000
2020-06-02  11:54:45.552660  7f6204ab8700  20  bluestore(/sandstone-data/ceph-4)  _do_write_small  reuse  blob  Blob(0x561bc0756700
blob([!~10000,0x1130000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,1000]) SharedBlob(0x561bc0757a40
sbid 0x0)) (0xf000~1000) (0xf800~800)
2020-06-02 11:54:45.552666 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x561bc072b340 1 blobs
2020-06-02 11:54:45.552669 7f6204ab8700 10 fbmap_alloc 0x561bbfb53500 allocate 0x10000/10000,10000,0
2020-06-02 11:54:45.552674 7f6204ab8700 10 fbmap_alloc 0x561bbfb53500 allocate extent: 0x1140000~10000/10000,10000,0
2020-06-02 11:54:45.552675 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0x1140000~10000]
2020-06-02    11:54:45.552679    7f6204ab8700    20    bluestore(/sandstone-data/ceph-4)    _do_alloc_write    blob    Blob(0x561bc0756700
blob([0x1140000~10000,0x1130000~10000]     csum+has_unused     crc32c/0x1000     unused=0xff)     use_tracker(0x2*0x10000     0x[0,1000])
SharedBlob(0x561bc0757a40 sbid 0x0))
2020-06-02     11:54:45.552686     7f6204ab8700     20     bluestore.blob(0x561bc0756700)     get_ref     0xf800~800     Blob(0x561bc0756700
blob([0x1140000~10000,0x1130000~10000]     csum+has_unused     crc32c/0x1000     unused=0xff)     use_tracker(0x2*0x10000     0x[0,1000])
SharedBlob(0x561bc0757a40 sbid 0x0))
2020-06-02  11:54:45.552691  7f6204ab8700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write  lex  0xf800~800:  0xf800~800
Blob(0x561bc0756700  blob([0x1140000~10000,0x1130000~10000]  csum+has_unused  crc32c/0x1000  unused=0x7f)  use_tracker(0x2*0x10000
0x[800,1000]) SharedBlob(0x561bc0757a40 sbid 0x0))
2020-06-02 11:54:45.552696 7f6204ab8700 20 bluestore.BufferSpace(0x561bc0757a58 in 0x561bbfe5c620) _discard 0xf000~1000
2020-06-02 11:54:45.552699 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write deferring small 0x1000 write via deferred
2020-06-02 11:54:45.552703 7f6204ab8700 30 estimate gc range(hex): [f800, 10000)
2020-06-02 11:54:45.552706 7f6204ab8700 20 bluestore.extentmap(0x561bc188f9b0) compress_extent_map 0xf800~800 next shard 0xffffffff
merging 0xf800~800: 0xf800~800 Blob(0x561bc0756700 blob([0x1140000~10000,0x1130000~10000] csum+has_unused crc32c/0x1000 unused=0x7f)
use_tracker(0x2*0x10000 0x[800,1000]) SharedBlob(0x561bc0757a40 sbid 0x0)) and 0x10000~1000: 0x10000~1000 Blob(0x561bc0756700
blob([0x1140000~10000,0x1130000~10000]     csum+has_unused     crc32c/0x1000     unused=0x7f)     use_tracker(0x2*0x10000     0x[800,1000])
SharedBlob(0x561bc0757a40 sbid 0x0))
2020-06-02 11:54:45.552714 7f6204ab8700 30 bluestore.extentmap(0x561bc188f9b0) dirty_range 0xf800~800
```

```
2020-06-02 11:54:45.552716 7f6204ab8700 20 bluestore.extentmap(0x561bc188f9b0) dirty_range mark inline shard dirty
2020-06-02 11:54:45.552718 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _write 2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0xf800~800 = 0
2020-06-02 11:54:45.552721 7f6204ab8700 30 bluestore.OnodeSpace(0x561bc20eb488 in 0x561bbfe5c620) lookup
2020-06-02 11:54:45.552723 7f6204ab8700 30 bluestore.OnodeSpace(0x561bc20eb488 in 0x561bbfe5c620) lookup #2:34000000:::head# hit
0x561bc05dd200
2020-06-02 11:54:45.552726 7f6204ab8700 15 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000:::head#
2020-06-02 11:54:45.552742 7f6204ab8700 30 bluestore(/sandstone-data/ceph-4)
_omap_setkeys 0x0000000000000466'.0000000129.00000000000000000006' <- 0000000129.00000000000000000006
2020-06-02 11:54:45.552753 7f6204ab8700 30 bluestore(/sandstone-data/ceph-4) _omap_setkeys 0x0000000000000466'._info' <- _info
2020-06-02 11:54:45.552756 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000:::head# = 0
2020-06-02 11:54:45.552760 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _txc_calc_cost 0x561bc072b340 cost 674082 (1 ios * 670000 +
4082 bytes)
2020-06-02 11:54:45.552762 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _txc_write_nodes txc 0x561bc072b340 onodes 0x561bc188f8c0
shared_blobs
2020-06-02 11:54:45.552765 7f6204ab8700 20 bluestore.extentmap(0x561bc188f9b0) update
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head#
2020-06-02 11:54:45.552770 7f6204ab8700 20 bluestore.extentmap(0x561bc188f9b0) update inline shard 154 bytes from 1 extents
2020-06-02 11:54:45.552776 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) onode
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# is 539 (379 bytes onode + 2 bytes spanning blobs + 158 bytes inline extents)
2020-06-02 11:54:45.552793 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _txc_finalize_kv txc 0x561bc072b340 allocated
0x[1140000~10000] released 0x[]
2020-06-02 11:54:45.552797 7f6204ab8700 10 freelist allocate 0x1140000~10000
2020-06-02 11:54:45.552799 7f6204ab8700 20 freelist _xor first_key 0x1000000 last_key 0x1000000
2020-06-02 11:54:45.552800 7f6204ab8700 30 freelist _xor 0x1000000: 00000000 00 00 10 00 00 00 00 00 00 00 00 00 00 00 00
00 |................|
2020-06-02 11:54:45.552811 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _txc_state_proc txc 0x561bc072b340 prepare
2020-06-02 11:54:45.552813 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _txc_finish_io 0x561bc072b340
2020-06-02 11:54:45.552815 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _txc_state_proc txc 0x561bc072b340 io_done
2020-06-02 11:54:45.552840 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _kv_sync_thread wake
2020-06-02 11:54:45.552846 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _kv_sync_thread committing 1 submitting 1 deferred done 0
stable 0
2020-06-02 11:54:45.552849 7f62122d3700 30 bluestore(/sandstone-data/ceph-4) _kv_sync_thread committing <0x561bc072b340>
2020-06-02 11:54:45.552851 7f62122d3700 30 bluestore(/sandstone-data/ceph-4) _kv_sync_thread submitting <0x561bc072b340>
2020-06-02 11:54:45.552852 7f62122d3700 30 bluestore(/sandstone-data/ceph-4) _kv_sync_thread deferred_done <>
2020-06-02 11:54:45.552853 7f62122d3700 30 bluestore(/sandstone-data/ceph-4) _kv_sync_thread deferred_stable <>
2020-06-02 11:54:45.552966 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _txc_applied_kv onode 0x561bc188f8c0 had 1
2020-06-02 11:54:45.552971 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _txc_applied_kv onode 0x561bc05dd200 had 1
2020-06-02 11:54:45.555342 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _kv_sync_thread committed 1 cleaned 0 in 0.002493 (0.000006
flush + 0.002487 kv commit)
2020-06-02 11:54:45.555358 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _kv_sync_thread releasing old bluefs 0x[]
2020-06-02 11:54:45.555361 7f62122d3700 10 fbmap_alloc 0x561bbfb53500 release done



2020-06-02 11:54:45.642145 7f6204ab8700 15 bluestore(/sandstone-data/ceph-4) _truncate 2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x10000
2020-06-02 11:54:45.642149 7f6204ab8700 15 bluestore(/sandstone-data/ceph-4) _do_truncate 2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x10000
2020-06-02 11:54:45.642152 7f6204ab8700 30 _dump_onode 0x561bc188f8c0 #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# nid 2614
size 0x11000 (69632) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2020-06-02 11:54:45.642156 7f6204ab8700 30 _dump_onode attr _ len 293
2020-06-02 11:54:45.642158 7f6204ab8700 30 _dump_onode attr snapset len 35
2020-06-02 11:54:45.642160 7f6204ab8700 30 _dump_extent_map 0xf800~1800: 0xf800~1800 Blob(0x561bc0756700
blob([0x1140000~10000,0x1130000~10000] csum+has_unused crc32c/0x1000 unused=0x7f) use_tracker(0x2*0x10000 0x[800,1000])
SharedBlob(0x561bc0757a40 sbid 0x0))
2020-06-02 11:54:45.642169 7f6204ab8700 30 _dump_extent_map csum:
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,49a8d579,86deadaf,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
2020-06-02 11:54:45.642173 7f6204ab8700 30 _dump_extent_map 0xf000~1000 buffer(0x561bc1568a20 space 0x561bc0757a58 0xf000~1000
writing nocache)
2020-06-02 11:54:45.642175 7f6204ab8700 30 _dump_extent_map 0x10000~1000 buffer(0x561bc0f6fef0 space 0x561bc0757a58 0x10000~1000
writing nocache)
2020-06-02 11:54:45.642177 7f6204ab8700 30 bluestore.extentmap(0x561bc188f9b0) fault_range 0x10000~1000
2020-06-02 11:54:45.642180 7f6204ab8700 20 bluestore.blob(0x561bc0756700) put_ref 0x10000~1000 Blob(0x561bc0756700
blob([0x1140000~10000,0x1130000~10000] csum+has_unused crc32c/0x1000 unused=0x7f) use_tracker(0x2*0x10000 0x[800,1000])
SharedBlob(0x561bc0757a40 sbid 0x0))
2020-06-02 11:54:45.642189 7f6204ab8700 30 bluestore.extentmap(0x561bc188f9b0) dirty_range 0x10000~1000
2020-06-02 11:54:45.642191 7f6204ab8700 20 bluestore.extentmap(0x561bc188f9b0) dirty_range mark inline shard dirty
2020-06-02 11:54:45.642193 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _wctx_finish lex_old 0x10000~1000: 0x10000~1000
Blob(0x561bc0756700 blob([0x1140000~10000,!~10000] csum+has_unused crc32c/0x1000 unused=0x7f) use_tracker(0x2*0x10000 0x[800,0])
SharedBlob(0x561bc0757a40 sbid 0x0))
2020-06-02 11:54:45.642199 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _wctx_finish blob release [0x1130000~10000]
2020-06-02 11:54:45.642201 7f6204ab8700 20 bluestore.blob(0x561bc0756700) discard_unallocated 0x10000~10000
2020-06-02 11:54:45.642202 7f6204ab8700 20 bluestore.BufferSpace(0x561bc0757a58 in 0x561bbfe5c620) _discard 0x10000~10000
2020-06-02 11:54:45.642215 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _wctx_finish release 0x1130000~10000
2020-06-02 11:54:45.642218 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _truncate 2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x10000 = 0
2020-06-02 11:54:45.642228 7f6204ab8700 15 bluestore(/sandstone-data/ceph-4) _setattrs 2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 2 keys
2020-06-02 11:54:45.642234 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _setattrs 2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 2 keys = 0
2020-06-02 11:54:45.642237 7f6204ab8700 30 bluestore.OnodeSpace(0x561bc20eb488 in 0x561bbfe5c620) lookup
2020-06-02 11:54:45.642238 7f6204ab8700 30 bluestore.OnodeSpace(0x561bc20eb488 in 0x561bbfe5c620) lookup #2:34000000:::head# hit
0x561bc05dd200
2020-06-02 11:54:45.642242 7f6204ab8700 15 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000:::head#
2020-06-02 11:54:45.642245 7f6204ab8700 30 bluestore(/sandstone-data/ceph-4)
_omap_setkeys 0x0000000000000466'.0000000129.00000000000000000007' <- 0000000129.00000000000000000007
2020-06-02 11:54:45.642255 7f6204ab8700 30 bluestore(/sandstone-data/ceph-4) _omap_setkeys 0x0000000000000466'._info' <- _info
2020-06-02 11:54:45.642258 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000:::head# = 0
2020-06-02 11:54:45.642261 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _txc_calc_cost 0x561bc0e37080 cost 671958 (1 ios * 670000 +
1958 bytes)
2020-06-02 11:54:45.642264 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _txc_write_nodes txc 0x561bc0e37080 onodes 0x561bc188f8c0
shared_blobs
2020-06-02 11:54:45.642274 7f6204ab8700 20 bluestore.extentmap(0x561bc188f9b0) update
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head#
2020-06-02 11:54:45.642280 7f6204ab8700 20 bluestore.extentmap(0x561bc188f9b0) update inline shard 160 bytes from 1 extents
2020-06-02 11:54:45.642285 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) onode
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# is 545 (379 bytes onode + 2 bytes spanning blobs + 164 bytes inline extents)
2020-06-02 11:54:45.642292 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _txc_finalize_kv txc 0x561bc0e37080 allocated 0x[] released
0x[1130000~10000]
2020-06-02 11:54:45.642295 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _txc_finalize_kv release 0x1130000~10000
2020-06-02 11:54:45.642297 7f6204ab8700 10 freelist release 0x1130000~10000 // 释放空间
2020-06-02 11:54:45.642298 7f6204ab8700 20 freelist _xor first_key 0x1000000 last_key 0x1000000
```

```
2020-06-02 11:54:45.642300 7f6204ab8700 30 freelist _xor 0x1000000: 00000000 00 00 08 00 00 00 00 00 00 00 00 00 00 00 00 00 |................|
2020-06-02 11:54:45.642310 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _txc_state_proc txc 0x561bc0e37080 prepare
2020-06-02 11:54:45.642312 7f6204ab8700 20 bluestore(/sandstone-data/ceph-4) _txc_finish_io 0x561bc0e37080
2020-06-02 11:54:45.642314 7f6204ab8700 10 bluestore(/sandstone-data/ceph-4) _txc_state_proc txc 0x561bc0e37080 io_done
2020-06-02 11:54:45.642342 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _kv_sync_thread wake
2020-06-02 11:54:45.642351 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _kv_sync_thread committing 1 submitting 1 deferred done 0 stable 0
2020-06-02 11:54:45.642357 7f62122d3700 30 bluestore(/sandstone-data/ceph-4) _kv_sync_thread committing <0x561bc0e37080>
2020-06-02 11:54:45.642369 7f62122d3700 30 bluestore(/sandstone-data/ceph-4) _kv_sync_thread submitting <0x561bc0e37080>
2020-06-02 11:54:45.642370 7f62122d3700 30 bluestore(/sandstone-data/ceph-4) _kv_sync_thread deferred_done <>
2020-06-02 11:54:45.642371 7f62122d3700 30 bluestore(/sandstone-data/ceph-4) _kv_sync_thread deferred_stable <>
2020-06-02 11:54:45.642433 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _txc_applied_kv onode 0x561bc188f8c0 had 1
2020-06-02 11:54:45.642440 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _txc_applied_kv onode 0x561bc05dd200 had 1
2020-06-02 11:54:45.643372 7f62188da700 20 bluestore.MempoolThread(0x561bbfb831e0) _trim_shards cache_size: 2845415832 kv_alloc: 1218696969 kv_used: 537066 meta_alloc: 1084479241 meta_used: 497038 data_alloc: 542239620 data_used: 0
2020-06-02 11:54:45.643380 7f62188da700 30 bluestore.MempoolThread(0x561bbfb831e0) _trim_shards max_shard_onodes: 167568 max_shard_buffer: 108447924
2020-06-02 11:54:45.643382 7f62188da700 20 bluestore.2QCache(0x561bbfb2bdc0) _trim onodes 227 / 167568 buffers 0 / 108447924
2020-06-02 11:54:45.643385 7f62188da700 20 bluestore.2QCache(0x561bbfe5c380) _trim onodes 34 / 167568 buffers 0 / 108447924
2020-06-02 11:54:45.643387 7f62188da700 20 bluestore.2QCache(0x561bbfe5c460) _trim onodes 39 / 167568 buffers 0 / 108447924
2020-06-02 11:54:45.643388 7f62188da700 20 bluestore.2QCache(0x561bbfe5c540) _trim onodes 40 / 167568 buffers 0 / 108447924
2020-06-02 11:54:45.643390 7f62188da700 20 bluestore.2QCache(0x561bbfe5c620) _trim onodes 44 / 167568 buffers 0 / 108447924
2020-06-02 11:54:45.652815 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _kv_sync_thread committed 1 cleaned 0 in 0.010456 (0.000017 flush + 0.010439 kv commit)
2020-06-02 11:54:45.652839 7f62122d3700 20 bluestore(/sandstone-data/ceph-4) _kv_sync_thread releasing old bluefs 0x[]
2020-06-02 11:54:45.652842 7f62122d3700 10 fbmap_alloc 0x561bbfb53500 release done
```

BlueStore big write Log分析

big write流程如果是未开启压缩，则可以尝试寻找可复用的blob。如果开启压缩则每次big write都是构造新的blob。默认是不开启，通过参数bluestore_compression_mode配置

```
// 初次写0~128k
2019-11-27 12:24:34.640865 7f5e0c649700 15 bluestore(/sandstone-data/ceph-4) _touch 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head#
2019-11-27 12:24:34.640868 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _assign_nid 1547
2019-11-27 12:24:34.640870 7f5e0c649700 10 bluestore(/sandstone-data/ceph-4) _touch 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# = 0
2019-11-27 12:24:34.640875 7f5e0c649700 15 bluestore(/sandstone-data/ceph-4) _setattrs 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 2 keys
2019-11-27 12:24:34.640880 7f5e0c649700 10 bluestore(/sandstone-data/ceph-4) _setattrs 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 2 keys = 0
2019-11-27 12:24:34.640883 7f5e0c649700 15 bluestore(/sandstone-data/ceph-4) _set_alloc_hint 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# object_size 1048576 write_size 1048576 flags -
2019-11-27 12:24:34.640887 7f5e0c649700 10 bluestore(/sandstone-data/ceph-4) _set_alloc_hint 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# object_size 1048576 write_size 1048576 flags - = 0
2019-11-27 12:24:34.640891 7f5e0c649700 15 bluestore(/sandstone-data/ceph-4) _write 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 0x0~20000
2019-11-27 12:24:34.640894 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_write #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 0x0~20000 - have 0x0 (0) bytes fadvise_flags 0x0
2019-11-27 12:24:34.640897 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_onode 0x559f75180240 #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# nid 1547 size 0x0 (0) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2019-11-27 12:24:34.640902 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_onode  attr _ len 292
2019-11-27 12:24:34.640903 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_onode  attr snapset len 35
2019-11-27 12:24:34.640905 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size 0x80000 compress=0 buffered=0
2019-11-27 12:24:34.640907 7f5e0c649700 30 bluestore.extentmap(0x559f75180330) fault_range 0x0~20000
2019-11-27 12:24:34.640914 7f5e0c649700 10 bluestore(/sandstone-data/ceph-4) _do_write_big 0x0~20000 target_blob_size 0x80000 compress 0
2019-11-27 12:24:34.640922 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x559f72e81b80 1 blobs
2019-11-27 12:24:34.640926 7f5e0c649700 10 stupidalloc 0x0x559f72b1a720 allocate_int want_size 0x20000 alloc_unit 0x10000 hint 0x0
2019-11-27 12:24:34.640930 7f5e0c649700 30 stupidalloc 0x0x559f72b1a720 _choose_bin len 0x20000 -> 6
2019-11-27 12:24:34.640935 7f5e0c649700 30 stupidalloc 0x0x559f72b1a720 allocate_int got 0x80200000~20000 from bin 9
2019-11-27 12:24:34.640939 7f5e0c649700 30 stupidalloc 0x0x559f72b1a720 _choose_bin len 0x57f4e0000 -> 9

// 分配空间，创建lextent[0x0~20000], logical_offset=0x0, blob_offset=0x0, length=0x20000.
2019-11-27 12:24:34.640940 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0x80200000~20000]
2019-11-27 12:24:34.640943 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write initialize csum setting for new blob Blob(0x559f75a21730 blob([]) use_tracker(0x0 0x0) SharedBlob(0x559f75a20d20 sbid 0x0)) csum_type crc32c csum_order 12 csum_length 0x20000
2019-11-27 12:24:34.640949 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write blob Blob(0x559f75a21730 blob([0x80200000~20000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 12:24:34.640981 7f5e0c649700 20 bluestore.blob(0x559f75a21730) get_ref 0x0~20000 Blob(0x559f75a21730 blob([0x80200000~20000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 12:24:34.640985 7f5e0c649700 20 bluestore.blob(0x559f75a21730) get_ref init 0x20000, 10000
2019-11-27 12:24:34.640987 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write  lex 0x0~20000: 0x0~20000 Blob(0x559f75a21730 blob([0x80200000~20000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))

2019-11-27 12:24:34.640992 7f5e0c649700 20 bluestore.BufferSpace(0x559f75a20d38 in 0x559f72751dc0) _discard 0x0~20000
2019-11-27 12:24:34.641014 7f5e0c649700 30 estimate gc range(hex): [0, 20000)
2019-11-27 12:24:34.641018 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_write extending size to 0x20000
2019-11-27 12:24:34.641020 7f5e0c649700 30 bluestore.extentmap(0x559f75180330) dirty_range 0x0~20000
2019-11-27 12:24:34.641022 7f5e0c649700 20 bluestore.extentmap(0x559f75180330) dirty_range mark inline shard dirty
2019-11-27 12:24:34.641024 7f5e0c649700 10 bluestore(/sandstone-data/ceph-4) _write 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 0x0~20000 = 0


// 初次写128k~128k
2019-11-27 12:24:34.675440 7f5e0c649700 15 bluestore(/sandstone-data/ceph-4) _setattrs 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 2 keys
2019-11-27 12:24:34.675446 7f5e0c649700 10 bluestore(/sandstone-data/ceph-4) _setattrs 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 2 keys = 0
2019-11-27 12:24:34.675450 7f5e0c649700 15 bluestore(/sandstone-data/ceph-4) _set_alloc_hint 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# object_size 1048576 write_size 1048576 flags -
```

```
2019-11-27 12:24:34.675453 7f5e0c649700 10 bluestore(/sandstone-data/ceph-4) _set_alloc_hint 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# object_size 1048576 write_size 1048576 flags - = 0
2019-11-27 12:24:34.675457 7f5e0c649700 15 bluestore(/sandstone-data/ceph-4) _write 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 0x20000~20000
2019-11-27 12:24:34.675460 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_write #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 0x20000~20000 - have 0x20000 (131072) bytes fadvise_flags 0x0
2019-11-27 12:24:34.675464 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_onode 0x559f75180240 #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# nid 1547 size 0x20000 (131072) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2019-11-27 12:24:34.675468 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_onode  attr _ len 292
2019-11-27 12:24:34.675470 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_onode  attr snapset len 35
2019-11-27 12:24:34.675471 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_extent_map  0x0~20000: 0x0~20000 Blob(0x559f75a21730 blob([0x80200000~20000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 12:24:34.675480 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_extent_map csum: [6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6
2019-11-27 12:24:34.675487 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size 0x80000 compress=0 buffered=0
2019-11-27 12:24:34.675489 7f5e0c649700 30 bluestore.extentmap(0x559f75180330) fault_range 0x20000~20000
2019-11-27 12:24:34.675491 7f5e0c649700 10 bluestore(/sandstone-data/ceph-4) _do_write_big 0x20000~20000 target_blob_size 0x80000 compress 0

// 复用blob[0x559f75a21730]
2019-11-27 12:24:34.675497 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_write_big reuse blob Blob(0x559f75a21730 blob([0x80200000~20000,!~20000] csum crc32c/0x1000) use_tracker(0x4*0x10000 0x[10000,10000,0,0]) SharedBlob(0x559f75a20d20 sbid 0x0)) (0x20000~20000)
2019-11-27 12:24:34.675502 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x559f75b0a2c0 1 blobs
2019-11-27 12:24:34.675505 7f5e0c649700 10 stupidalloc 0x0x559f72b1a720 allocate_int want_size 0x20000 alloc_unit 0x10000 hint 0x0
2019-11-27 12:24:34.675507 7f5e0c649700 30 stupidalloc 0x0x559f72b1a720 _choose_bin len 0x20000 -> 6
2019-11-27 12:24:34.675510 7f5e0c649700 30 stupidalloc 0x0x559f72b1a720 allocate_int got 0x80220000~20000 from bin 9
2019-11-27 12:24:34.675514 7f5e0c649700 30 stupidalloc 0x0x559f72b1a720 _choose_bin len 0x57f4c0000 -> 9

// 创建lextent[0x0~20000], logical_offset=0x20000, blob_offset=0x20000, length=0x20000. 可以与前面的extent合并为lextent[0x0~40000]
2019-11-27 12:24:34.675515 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0x80220000~20000]
2019-11-27 12:24:34.675524 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write blob Blob(0x559f75a21730 blob([0x80200000~20000,0x80220000~20000] csum crc32c/0x1000) use_tracker(0x4*0x10000 0x[10000,10000,0,0]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 12:24:34.675538 7f5e0c649700 20 bluestore.blob(0x559f75a21730) get_ref 0x20000~20000 Blob(0x559f75a21730 blob([0x80200000~20000,0x80220000~20000] csum crc32c/0x1000) use_tracker(0x4*0x10000 0x[10000,10000,0,0]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 12:24:34.675547 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write  lex 0x20000~20000: 0x20000~20000 Blob(0x559f75a21730 blob([0x80200000~20000,0x80220000~20000] csum crc32c/0x1000) use_tracker(0x4*0x10000 0x[10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))

2019-11-27 12:24:34.675552 7f5e0c649700 20 bluestore.BufferSpace(0x559f75a20d38 in 0x559f72751dc0) _discard 0x20000~20000
2019-11-27 12:24:34.675558 7f5e0c649700 30 estimate gc range(hex): [20000, 40000)
2019-11-27 12:24:34.675560 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_write extending size to 0x40000
2019-11-27 12:24:34.675562 7f5e0c649700 20 bluestore.extentmap(0x559f75180330) compress_extent_map 0x20000~20000 next shard 0xffffffff merging 0x0~20000: 0x0~20000 Blob(0x559f75a21730 blob([0x80200000~20000,0x80220000~20000] csum crc32c/0x1000) use_tracker(0x4*0x10000 0x[10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0)) and 0x20000~20000: 0x20000~20000 Blob(0x559f75a21730 blob([0x80200000~20000,0x80220000~20000] csum crc32c/0x1000) use_tracker(0x4*0x10000 0x[10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 12:24:34.675571 7f5e0c649700 30 bluestore.extentmap(0x559f75180330) dirty_range 0x20000~20000
2019-11-27 12:24:34.675572 7f5e0c649700 20 bluestore.extentmap(0x559f75180330) dirty_range mark inline shard dirty
2019-11-27 12:24:34.675574 7f5e0c649700 10 bluestore(/sandstone-data/ceph-4) _write 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 0x20000~20000 = 0



// 覆盖写0~128k
2019-11-27 12:24:38.108077 7f5e0c649700 15 bluestore(/sandstone-data/ceph-4) _setattrs 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 2 keys
2019-11-27 12:24:38.108083 7f5e0c649700 10 bluestore(/sandstone-data/ceph-4) _setattrs 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 2 keys = 0
2019-11-27 12:24:38.108086 7f5e0c649700 15 bluestore(/sandstone-data/ceph-4) _set_alloc_hint 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# object_size 1048576 write_size 1048576 flags -
2019-11-27 12:24:38.108090 7f5e0c649700 10 bluestore(/sandstone-data/ceph-4) _set_alloc_hint 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# object_size 1048576 write_size 1048576 flags - = 0
2019-11-27 12:24:38.108094 7f5e0c649700 15 bluestore(/sandstone-data/ceph-4) _write 2.0_head #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 0x0~20000
2019-11-27 12:24:38.108097 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_write #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 0x0~20000 - have 0x100000 (1048576) bytes fadvise_flags 0x0
2019-11-27 12:24:38.108100 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_onode 0x559f75180240 #2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# nid 1547 size 0x100000 (1048576) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2019-11-27 12:24:38.108115 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_onode  attr _ len 292
2019-11-27 12:24:38.108117 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_onode  attr snapset len 35
2019-11-27 12:24:38.108118 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_extent_map  0x0~80000: 0x0~80000 Blob(0x559f75a21730 blob([0x80200000~20000,0x80220000~20000,0x80240000~20000,0x80260000~20000] csum crc32c/0x1000) use_tracker(0x8*0x10000 0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 12:24:38.108138 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_extent_map csum: [6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6
2019-11-27 12:24:38.108149 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_extent_map      0x0~1000 buffer(0x559f75a361b0 space 0x559f75a20d38 0x0~1000 clean)
2019-11-27 12:24:38.108151 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_extent_map  0x80000~80000: 0x0~80000 Blob(0x559f756a60e0 blob([0x80280000~20000,0x802a0000~20000,0x802c0000~20000,0x802e0000~20000] csum crc32c/0x1000) use_tracker(0x8*0x10000 0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f756a79d0 sbid 0x0))
2019-11-27 12:24:38.108157 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_extent_map csum: [6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6
2019-11-27 12:24:38.108168 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size 0x80000 compress=0 buffered=0
2019-11-27 12:24:38.108170 7f5e0c649700 30 bluestore.extentmap(0x559f75180330) fault_range 0x0~20000
2019-11-27 12:24:38.108173 7f5e0c649700 10 bluestore(/sandstone-data/ceph-4) _do_write_big 0x0~20000 target_blob_size 0x80000 compress 0

// [0x0~20000]被punch hole
2019-11-27 12:24:38.108176 7f5e0c649700 20 bluestore.blob(0x559f75a21730) put_ref 0x0~20000 Blob(0x559f75a21730 blob([0x80200000~20000,0x80220000~20000,0x80240000~20000,0x80260000~20000] csum crc32c/0x1000) use_tracker(0x8*0x10000 0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 12:24:38.108202 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_write_big reuse blob Blob(0x559f75a21730 blob([!~20000,0x80220000~20000,0x80240000~20000,0x80260000~20000] csum crc32c/0x1000) use_tracker(0x8*0x10000 0x[0,0,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0)) (0x0~20000)
2019-11-27 12:24:38.108209 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x559f72e80840 1 blobs
2019-11-27 12:24:38.108213 7f5e0c649700 10 stupidalloc 0x0x559f72b1a720 allocate_int want_size 0x20000 alloc_unit 0x10000 hint 0x0
2019-11-27 12:24:38.108215 7f5e0c649700 30 stupidalloc 0x0x559f72b1a720 _choose_bin len 0x20000 -> 6
2019-11-27 12:24:38.108218 7f5e0c649700 30 stupidalloc 0x0x559f72b1a720 allocate_int got 0x80300000~20000 from bin 9
2019-11-27 12:24:38.108222 7f5e0c649700 30 stupidalloc 0x0x559f72b1a720 _choose_bin len 0x57f3e0000 -> 9
```

```
// 分配新的空间[0x80300000~20000]
2019-11-27 12:24:38.108223 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0x80300000~20000]
2019-11-27    12:24:38.108226    7f5e0c649700    20    bluestore(/sandstone-data/ceph-4)    _do_alloc_write    blob    Blob(0x559f75a21730
blob([0x80300000~20000,0x80220000~20000,0x80240000~20000,0x80260000~20000]        csum        crc32c/0x1000)        use_tracker(0x8*0x10000
0x[0,0,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27     12:24:38.108242     7f5e0c649700     20     bluestore.blob(0x559f75a21730)     get_ref     0x0~20000     Blob(0x559f75a21730
blob([0x80300000~20000,0x80220000~20000,0x80240000~20000,0x80260000~20000]        csum        crc32c/0x1000)        use_tracker(0x8*0x10000
0x[0,0,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 12:24:38.108248 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write  lex 0x0~20000: 0x0~20000 Blob(0x559f75a21730
blob([0x80300000~20000,0x80220000~20000,0x80240000~20000,0x80260000~20000]        csum        crc32c/0x1000)        use_tracker(0x8*0x10000
0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 12:24:38.108252 7f5e0c649700 20 bluestore.BufferSpace(0x559f75a20d38 in 0x559f72751dc0) _discard 0x0~20000
2019-11-27  12:24:38.108254  7f5e0c649700  20  bluestore.2QCache(0x559f72751dc0)  _rm_buffer  buffer(0x559f75a361b0  space  0x559f75a20d38
0x0~1000 clean)
2019-11-27 12:24:38.108263 7f5e0c649700 30 estimate gc range(hex): [0, 20000)
2019-11-27 12:24:38.108264 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _wctx_finish lex_old 0x0~20000: 0x0~20000 Blob(0x559f75a21730
blob([0x80300000~20000,0x80220000~20000,0x80240000~20000,0x80260000~20000]        csum        crc32c/0x1000)        use_tracker(0x8*0x10000
0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 12:24:38.108269 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _wctx_finish  blob release [0x80200000~20000]
2019-11-27 12:24:38.108270 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _wctx_finish  release 0x80200000~20000

// 合并连续的lextent
2019-11-27 12:24:38.108273 7f5e0c649700 20 bluestore.extentmap(0x559f75180330) compress_extent_map 0x0~20000 next shard 0xffffffff merging
0x0~20000:  0x0~20000  Blob(0x559f75a21730  blob([0x80300000~20000,0x80220000~20000,0x80240000~20000,0x80260000~20000]  csum  crc32c/0x1000)
use_tracker(0x8*0x10000  0x[10000,10000,10000,10000,10000,10000,10000,10000])  SharedBlob(0x559f75a20d20  sbid  0x0))  and  0x20000~60000:
0x20000~60000    Blob(0x559f75a21730    blob([0x80300000~20000,0x80220000~20000,0x80240000~20000,0x80260000~20000]    csum    crc32c/0x1000)
use_tracker(0x8*0x10000 0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 12:24:38.108280 7f5e0c649700 30 bluestore.extentmap(0x559f75180330) dirty_range 0x0~20000
2019-11-27 12:24:38.108282 7f5e0c649700 20 bluestore.extentmap(0x559f75180330) dirty_range mark inline shard dirty
2019-11-27         12:24:38.108284         7f5e0c649700         10         bluestore(/sandstone-data/ceph-4)         _write         2.0_head
#2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 0x0~20000 = 0


// 覆盖写0~64k
2019-11-27         15:17:05.703542         7f5e0c649700         15         bluestore(/sandstone-data/ceph-4)         _setattrs         2.0_head
#2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 2 keys
2019-11-27         15:17:05.703550         7f5e0c649700         10         bluestore(/sandstone-data/ceph-4)         _setattrs         2.0_head
#2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 2 keys = 0
2019-11-27        15:17:05.703556        7f5e0c649700        15        bluestore(/sandstone-data/ceph-4)        _set_alloc_hint        2.0_head
#2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# object_size 1048576 write_size 1048576 flags -
2019-11-27        15:17:05.703562        7f5e0c649700        10        bluestore(/sandstone-data/ceph-4)        _set_alloc_hint        2.0_head
#2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# object_size 1048576 write_size 1048576 flags - = 0
2019-11-27         15:17:05.703569         7f5e0c649700         15         bluestore(/sandstone-data/ceph-4)         _write         2.0_head
#2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 0x0~10000
2019-11-27           15:17:05.703574           7f5e0c649700           20           bluestore(/sandstone-data/ceph-4)           _do_write
#2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 0x0~10000 - have 0x100000 (1048576) bytes fadvise_flags 0x0
2019-11-27       15:17:05.703580       7f5e0c649700       30       bluestore(/sandstone-data/ceph-4)       _dump_onode       0x559f75180240
#2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head#    nid    1547    size    0x100000    (1048576)    expected_object_size    1048576
expected_write_size 1048576 in 0 shards, 0 spanning blobs
2019-11-27 15:17:05.703584 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_onode  attr _ len 292
2019-11-27 15:17:05.703585 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_onode  attr snapset len 35
2019-11-27 15:17:05.703587 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_extent_map   0x0~80000: 0x0~80000 Blob(0x559f75a21730
blob([0x80300000~20000,0x80320000~20000,0x80340000~20000,0x80360000~20000]        csum        crc32c/0x1000)        use_tracker(0x8*0x10000
0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27    15:17:05.703599    7f5e0c649700    30    bluestore(/sandstone-data/ceph-4)    _dump_extent_map                          csum:
[6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6
2019-11-27 15:17:05.703610 7f5e0c649700 30 bluestore(/sandstone-data/ceph-4) _dump_extent_map        0x0~1000 buffer(0x559f75e0c870 space
0x559f75a20d38 0x0~1000 clean)
2019-11-27   15:17:05.703613   7f5e0c649700   30   bluestore(/sandstone-data/ceph-4)   _dump_extent_map      0x80000~80000:   0x0~80000
Blob(0x559f756a60e0 blob([0x80380000~20000,0x803a0000~20000,0x803c0000~20000,0x803e0000~20000] csum crc32c/0x1000) use_tracker(0x8*0x10000
0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f756a79d0 sbid 0x0))
2019-11-27    15:17:05.703619    7f5e0c649700    30    bluestore(/sandstone-data/ceph-4)    _dump_extent_map                          csum:
[6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6706be76,6
2019-11-27 15:17:05.703636 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2019-11-27 15:17:05.703638 7f5e0c649700 30 bluestore.extentmap(0x559f75180330) fault_range 0x0~10000
2019-11-27 15:17:05.703644 7f5e0c649700 10 bluestore(/sandstone-data/ceph-4) _do_write_big 0x0~10000 target_blob_size 0x80000 compress 0
2019-11-27     15:17:05.703652     7f5e0c649700     20     bluestore.blob(0x559f75a21730)     put_ref     0x0~10000     Blob(0x559f75a21730
blob([0x80300000~20000,0x80320000~20000,0x80340000~20000,0x80360000~20000]        csum        crc32c/0x1000)        use_tracker(0x8*0x10000
0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27   15:17:05.703663   7f5e0c649700   20   bluestore(/sandstone-data/ceph-4)   _do_write_big   reuse   blob   Blob(0x559f75a21730
blob([!~10000,0x80310000~10000,0x80320000~20000,0x80340000~20000,0x80360000~20000]     csum     crc32c/0x1000)     use_tracker(0x8*0x10000
0x[0,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0)) (0x0~10000)
2019-11-27 15:17:05.703670 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x559f72e818c0 1 blobs
2019-11-27 15:17:05.703673 7f5e0c649700 10 stupidalloc 0x0x559f72b1a720 allocate_int want_size 0x10000 alloc_unit 0x10000 hint 0x0
2019-11-27 15:17:05.703675 7f5e0c649700 30 stupidalloc 0x0x559f72b1a720 _choose_bin len 0x10000 -> 5
2019-11-27 15:17:05.703679 7f5e0c649700 30 stupidalloc 0x0x559f72b1a720 allocate_int got 0x83000000~10000 from bin 9
2019-11-27 15:17:05.703682 7f5e0c649700 30 stupidalloc 0x0x559f72b1a720 _choose_bin len 0x57c6f0000 -> 9
2019-11-27 15:17:05.703684 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0x83000000~10000]
2019-11-27    15:17:05.703687    7f5e0c649700    20    bluestore(/sandstone-data/ceph-4)    _do_alloc_write    blob    Blob(0x559f75a21730
blob([0x83000000~10000,0x80310000~10000,0x80320000~20000,0x80340000~20000,0x80360000~20000]  csum  crc32c/0x1000)  use_tracker(0x8*0x10000
0x[0,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27     15:17:05.703699     7f5e0c649700     20     bluestore.blob(0x559f75a21730)     get_ref     0x0~10000     Blob(0x559f75a21730
blob([0x83000000~10000,0x80310000~10000,0x80320000~20000,0x80340000~20000,0x80360000~20000]  csum  crc32c/0x1000)  use_tracker(0x8*0x10000
0x[0,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 15:17:05.703705 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write  lex 0x0~10000: 0x0~10000 Blob(0x559f75a21730
blob([0x83000000~10000,0x80310000~10000,0x80320000~20000,0x80340000~20000,0x80360000~20000]  csum  crc32c/0x1000)  use_tracker(0x8*0x10000
0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 15:17:05.703710 7f5e0c649700 20 bluestore.BufferSpace(0x559f75a20d38 in 0x559f72751dc0) _discard 0x0~10000
2019-11-27  15:17:05.703712  7f5e0c649700  20  bluestore.2QCache(0x559f72751dc0)  _rm_buffer  buffer(0x559f75e0c870  space  0x559f75a20d38
0x0~1000 clean)
2019-11-27 15:17:05.703750 7f5e0c649700 30 estimate gc range(hex): [0, 10000)
2019-11-27 15:17:05.703754 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _wctx_finish lex_old 0x0~10000: 0x0~10000 Blob(0x559f75a21730
blob([0x83000000~10000,0x80310000~10000,0x80320000~20000,0x80340000~20000,0x80360000~20000]  csum  crc32c/0x1000)  use_tracker(0x8*0x10000
0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 15:17:05.703760 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _wctx_finish  blob release [0x80300000~10000]
2019-11-27 15:17:05.703762 7f5e0c649700 20 bluestore(/sandstone-data/ceph-4) _wctx_finish  release 0x80300000~10000
2019-11-27 15:17:05.703765 7f5e0c649700 20 bluestore.extentmap(0x559f75180330) compress_extent_map 0x0~10000 next shard 0xffffffff merging
0x0~10000: 0x0~10000 Blob(0x559f75a21730 blob([0x83000000~10000,0x80310000~10000,0x80320000~20000,0x80340000~20000,0x80360000~20000] csum
crc32c/0x1000)  use_tracker(0x8*0x10000  0x[10000,10000,10000,10000,10000,10000,10000,10000])  SharedBlob(0x559f75a20d20  sbid  0x0))  and
0x10000~70000:                                        0x10000~70000                                                    Blob(0x559f75a21730
blob([0x83000000~10000,0x80310000~10000,0x80320000~20000,0x80340000~20000,0x80360000~20000]  csum  crc32c/0x1000)  use_tracker(0x8*0x10000
0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x559f75a20d20 sbid 0x0))
2019-11-27 15:17:05.703774 7f5e0c649700 30 bluestore.extentmap(0x559f75180330) dirty_range 0x0~10000
```

```
2019-11-27 15:17:05.703775 7f5e0c649700 20 bluestore.extentmap(0x559f75180330) dirty_range mark inline shard dirty
2019-11-27 15:17:05.703777 7f5e0c649700 10 bluestore(/sandstone-data/ceph-4) _write 2.0_head
#2:00bfa142:::rbd_data.101f241b25c.0000000000000002:head# 0x0~10000 = 0




///////初次写448k~128k
2020-06-24 11:50:21.565869 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_write
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x70000~20000 - have 0x0 (0) bytes fadvise_flags 0x0
2020-06-24 11:50:21.565875 7fb5ce895700 30 _dump_onode 0x559d0f129200 #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# nid 20050
size 0x0 (0) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2020-06-24 11:50:21.565882 7fb5ce895700 30 _dump_onode  attr _ len 293
2020-06-24 11:50:21.565884 7fb5ce895700 30 _dump_onode  attr snapset len 35
2020-06-24 11:50:21.565888 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2020-06-24 11:50:21.565891 7fb5ce895700 30 bluestore.extentmap(0x559d0f1292f0) fault_range 0x70000~20000
2020-06-24 11:50:21.565895 7fb5ce895700 10 bluestore(/sandstone-data/ceph-4) _do_write_big 0x70000~20000 target_blob_size 0x80000 compress
0
2020-06-24 11:50:21.565904 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x559d0f24b600 1 blobs
2020-06-24 11:50:21.565907 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write buffered:0 compress:0
target_blob_size:524288 write_flags:0
write_item:  blob  0x559d0f35b428  logical_offset:458752  blob_length:131072  b_off:0  b_off0:0  length0:131072  mark_unused:0  new_blob:1
compressed_len:0

2020-06-24 11:50:21.565916 7fb5ce895700 10 fbmap_alloc 0x559d0ec97400 allocate 0x20000/10000,20000,0
2020-06-24 11:50:21.565921 7fb5ce895700 10 fbmap_alloc 0x559d0ec97400 allocate extent: 0xd90000~20000/10000,20000,0
2020-06-24 11:50:21.565923 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0xd90000~20000]
2020-06-24 11:50:21.565926 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write initialize csum setting for new blob
Blob(0x559d0f433810 blob([]) use_tracker(0x0 0x0) SharedBlob(0x559d0f018850 sbid 0x0)) csum_type crc32c csum_order 12 csum_length 0x20000
2020-06-24 11:50:21.565933 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write blob Blob(0x559d0f433810
blob([0xd90000~20000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x559d0f018850 sbid 0x0))
2020-06-24 11:50:21.565952 7fb5ce895700 20 bluestore.blob(0x559d0f433810) get_ref 0x0~20000 Blob(0x559d0f433810 blob([0xd90000~20000] csum
crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x559d0f018850 sbid 0x0))
2020-06-24 11:50:21.565956 7fb5ce895700 20 bluestore.blob(0x559d0f433810) get_ref init 0x20000, 10000
2020-06-24 11:50:21.565959 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write lex 0x70000~20000: 0x0~20000
Blob(0x559d0f433810 blob([0xd90000~20000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,10000]) SharedBlob(0x559d0f018850 sbid
0x0))
2020-06-24 11:50:21.565964 7fb5ce895700 20 bluestore.BufferSpace(0x559d0f018868 in 0x559d0eb2a1c0) _discard 0x0~20000
2020-06-24 11:50:21.565974 7fb5ce895700 30 estimate gc range(hex): [70000, 90000)
2020-06-24 11:50:21.565976 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_write extending size to 0x90000
2020-06-24 11:50:21.565978 7fb5ce895700 30 bluestore.extentmap(0x559d0f1292f0) dirty_range 0x70000~20000
2020-06-24 11:50:21.565980 7fb5ce895700 20 bluestore.extentmap(0x559d0f1292f0) dirty_range mark inline shard dirty
2020-06-24 11:50:21.565982 7fb5ce895700 10 bluestore(/sandstone-data/ceph-4) _write 2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x70000~20000 = 0
2020-06-24 11:50:21.565986 7fb5ce895700 30 bluestore.OnodeSpace(0x559d0f706148 in 0x559d0eb2a1c0) lookup
2020-06-24 11:50:21.565988 7fb5ce895700 30 bluestore.OnodeSpace(0x559d0f706148 in 0x559d0eb2a1c0) lookup #2:34000000:::head# hit
0x559d0f0ef200
2020-06-24 11:50:21.565992 7fb5ce895700 15 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000:::head#
2020-06-24 11:50:21.565996 7fb5ce895700 30 bluestore(/sandstone-data/ceph-4)
_omap_setkeys  0x0000000000000466'.0000000282.00000000000000000118' <- 0000000282.0000000000000000118
2020-06-24 11:50:21.566005 7fb5ce895700 30 bluestore(/sandstone-data/ceph-4) _omap_setkeys  0x0000000000000466'._fastinfo' <- _fastinfo
2020-06-24 11:50:21.566010 7fb5ce895700 10 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000:::head# = 0
2020-06-24 11:50:21.566015 7fb5ce895700 10 bluestore(/sandstone-data/ceph-4) _txc_calc_cost 0x559d0f24b600 cost 1472434 (2 ios * 670000 +
132434 bytes)
2020-06-24 11:50:21.566019 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _txc_write_nodes txc 0x559d0f24b600 onodes 0x559d0f129200
shared_blobs
2020-06-24 11:50:21.566022 7fb5ce895700 20 bluestore.extentmap(0x559d0f1292f0) update
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head#
2020-06-24 11:50:21.566030 7fb5ce895700 20 bluestore.extentmap(0x559d0f1292f0) update  inline shard 147 bytes from 1 extents
2020-06-24 11:50:21.566038 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) onode
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# is 533 (380 bytes onode + 2 bytes spanning blobs + 151 bytes inline extents)

////覆盖写448k~64k,首先会punch hole掉覆盖的范围
2020-06-24 11:50:22.331699 7fb5ce895700 15 bluestore(/sandstone-data/ceph-4) _write 2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x70000~10000
2020-06-24 11:50:22.331704 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_write
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x70000~10000 - have 0x90000 (589824) bytes fadvise_flags 0x0
2020-06-24 11:50:22.331710 7fb5ce895700 30 _dump_onode 0x559d0f129200 #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# nid 20050
size 0x90000 (589824) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2020-06-24 11:50:22.331716 7fb5ce895700 30 _dump_onode  attr _ len 293
2020-06-24 11:50:22.331717 7fb5ce895700 30 _dump_onode  attr snapset len 35
2020-06-24 11:50:22.331718 7fb5ce895700 30 _dump_extent_map   0x70000~20000:  0x0~20000 Blob(0x559d0f433810 blob([0xd90000~20000] csum
crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,10000]) SharedBlob(0x559d0f018850 sbid 0x0))
2020-06-24 11:50:22.331728 7fb5ce895700 30 _dump_extent_map                                                                    csum:
[c59ad9d4,5ccf7fa,b8b2b261,26989e96,9503541d,c4809830,c58b617d,49220fe4,a607b233,14d5e0de,581fddc3,b975ffdf,24f55bcc,6d8bd893,a78c24a3,8c5586e5,bf2d964b,b5
2020-06-24 11:50:22.331734 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2020-06-24 11:50:22.331737 7fb5ce895700 30 bluestore.extentmap(0x559d0f1292f0) fault_range 0x70000~10000
2020-06-24 11:50:22.331740 7fb5ce895700 10 bluestore(/sandstone-data/ceph-4) _do_write_big 0x70000~10000 target_blob_size 0x80000 compress
0
2020-06-24 11:50:22.331745 7fb5ce895700 20 bluestore.blob(0x559d0f433810) put_ref 0x0~10000 Blob(0x559d0f433810 blob([0xd90000~20000] csum
crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,10000]) SharedBlob(0x559d0f018850 sbid 0x0)) ////push invalid extent
2020-06-24 11:50:22.331756 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_write_big reuse blob Blob(0x559d0f433810
blob([!~10000,0xda0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[0,10000]) SharedBlob(0x559d0f018850 sbid 0x0)) (0x0~10000)
2020-06-24 11:50:22.331763 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x559d0eff4dc0 1 blobs
2020-06-24 11:50:22.331765 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write buffered:0 compress:0
target_blob_size:524288 write_flags:0
write_item:  blob  0x559d0f4b4468  logical_offset:458752  blob_length:65536  b_off:0  b_off0:0  length0:65536  mark_unused:0  new_blob:0
compressed_len:0

2020-06-24 11:50:22.331771 7fb5ce895700 10 fbmap_alloc 0x559d0ec97400 allocate 0x10000/10000,10000,0
2020-06-24 11:50:22.331776 7fb5ce895700 10 fbmap_alloc 0x559d0ec97400 allocate extent: 0xdd0000~10000/10000,10000,0
2020-06-24 11:50:22.331778 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0xdd0000~10000]
2020-06-24 11:50:22.331783 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write blob Blob(0x559d0f433810
blob([0xdd0000~10000,0xda0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[0,10000]) SharedBlob(0x559d0f018850 sbid 0x0))
2020-06-24 11:50:22.331797 7fb5ce895700 20 bluestore.blob(0x559d0f433810) get_ref 0x0~10000 Blob(0x559d0f433810
blob([0xdd0000~10000,0xda0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[0,10000]) SharedBlob(0x559d0f018850 sbid 0x0))
2020-06-24 11:50:22.331801 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write lex 0x70000~10000: 0x0~10000
Blob(0x559d0f433810 blob([0xdd0000~10000,0xda0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,10000])
SharedBlob(0x559d0f018850 sbid 0x0))
2020-06-24 11:50:22.331806 7fb5ce895700 20 bluestore.BufferSpace(0x559d0f018868 in 0x559d0eb2a1c0) _discard 0x0~10000
2020-06-24 11:50:22.331824 7fb5ce895700 30 estimate gc range(hex): [70000, 80000)
```

2020-06-24 11:50:22.331828 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _wctx_finish lex_old 0x70000~10000: 0x0~10000 Blob(0x559d0f433810 blob([0xdd0000~10000,0xda0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,10000]) SharedBlob(0x559d0f018850 sbid 0x0))
2020-06-24 11:50:22.331833 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _wctx_finish blob release [0xd90000~10000]
2020-06-24 11:50:22.331835 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _wctx_finish release 0xd90000~10000
2020-06-24 11:50:22.331838 7fb5ce895700 20 bluestore.extentmap(0x559d0f1292f0) compress_extent_map 0x70000~10000 next shard 0xffffffff merging 0x70000~10000: 0x0~10000 Blob(0x559d0f433810 blob([0xdd0000~10000,0xda0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,10000]) SharedBlob(0x559d0f018850 sbid 0x0)) and 0x80000~10000: 0x10000~10000 Blob(0x559d0f433810 blob([0xdd0000~10000,0xda0000~10000] csum crc32c/0x1000) use_tracker(0x2*0x10000 0x[10000,10000]) SharedBlob(0x559d0f018850 sbid 0x0))
2020-06-24 11:50:22.331847 7fb5ce895700 30 bluestore.extentmap(0x559d0f1292f0) dirty_range 0x70000~10000
2020-06-24 11:50:22.331848 7fb5ce895700 20 bluestore.extentmap(0x559d0f1292f0) dirty_range mark inline shard dirty
2020-06-24 11:50:22.331851 7fb5ce895700 10 bluestore(/sandstone-data/ceph-4) _write 2.2c_head #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x70000~10000 = 0
2020-06-24 11:50:22.331855 7fb5ce895700 30 bluestore.OnodeSpace(0x559d0f706148 in 0x559d0eb2a1c0) lookup
2020-06-24 11:50:22.331857 7fb5ce895700 30 bluestore.OnodeSpace(0x559d0f706148 in 0x559d0eb2a1c0) lookup #2:34000000:::head# hit 0x559d0f0ef200
2020-06-24 11:50:22.331860 7fb5ce895700 15 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000:::head#
2020-06-24 11:50:22.331864 7fb5ce895700 30 bluestore(/sandstone-data/ceph-4) _omap_setkeys 0x0000000000000466'.0000000282.00000000000000000119' <- 0000000282.0000000000000000119
2020-06-24 11:50:22.331886 7fb5ce895700 30 bluestore(/sandstone-data/ceph-4) _omap_setkeys 0x0000000000000466'._info' <- _info
2020-06-24 11:50:22.331896 7fb5ce895700 10 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000:::head# = 0
2020-06-24 11:50:22.331899 7fb5ce895700 10 bluestore(/sandstone-data/ceph-4) _txc_calc_cost 0x559d0eff4dc0 cost 1407570 (2 ios * 670000 + 67570 bytes)
2020-06-24 11:50:22.331902 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) _txc_write_nodes txc 0x559d0eff4dc0 onodes 0x559d0f129200 shared_blobs
2020-06-24 11:50:22.331905 7fb5ce895700 20 bluestore.extentmap(0x559d0f1292f0) update #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head#
2020-06-24 11:50:22.331911 7fb5ce895700 20 bluestore.extentmap(0x559d0f1292f0) update inline shard 151 bytes from 1 extents
2020-06-24 11:50:22.331916 7fb5ce895700 20 bluestore(/sandstone-data/ceph-4) onode #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# is 537 (380 bytes onode + 2 bytes spanning blobs + 155 bytes inline extents)

[960k, 64k]
[448k, 64k]
[512k, 128k]

//////// [960k, 64k]
2020-06-28 15:50:16.507237 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _touch 2.2c_head #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# = 0
2020-06-28 15:50:16.507266 7fc16b64f700 15 bluestore(/sandstone-data/ceph-4) _setattrs 2.2c_head #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 2 keys
2020-06-28 15:50:16.507279 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _setattrs 2.2c_head #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 2 keys = 0
2020-06-28 15:50:16.507285 7fc16b64f700 15 bluestore(/sandstone-data/ceph-4) _set_alloc_hint 2.2c_head #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# object_size 1048576 write_size 1048576 flags -
2020-06-28 15:50:16.507290 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _set_alloc_hint 2.2c_head #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# object_size 1048576 write_size 1048576 flags - = 0
2020-06-28 15:50:16.507295 7fc16b64f700 15 bluestore(/sandstone-data/ceph-4) _write 2.2c_head #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0xf0000~10000
2020-06-28 15:50:16.507300 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_write #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0xf0000~10000 - have 0x0 (0) bytes fadvise_flags 0x0
2020-06-28 15:50:16.507305 7fc16b64f700 30 _dump_onode 0x55d1b1cd9b00 #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# nid 21011 size 0x0 (0) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2020-06-28 15:50:16.507310 7fc16b64f700 30 _dump_onode attr _ len 293
2020-06-28 15:50:16.507311 7fc16b64f700 30 _dump_onode attr snapset len 35
2020-06-28 15:50:16.507315 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size 0x80000 compress=0 buffered=0
2020-06-28 15:50:16.507318 7fc16b64f700 30 bluestore.extentmap(0x55d1b1cd9bf0) fault_range 0xf0000~10000
2020-06-28 15:50:16.507322 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _do_write_big 0xf0000~10000 target_blob_size 0x80000 compress 0
2020-06-28 15:50:16.507329 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x55d1b2219600 1 blobs
2020-06-28 15:50:16.507352 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write buffered:0 compress:0 target_blob_size:524288 write_flags:0
write_item: blob 0x55d1b25735e8 logical_offset:983040 blob_length:65536 b_off:0 b_off0:0 length0:65536 mark_unused:0 new_blob:1 compressed_len:0

2020-06-28 15:50:16.507361 7fc16b64f700 10 fbmap_alloc 0x55d1b0ff1400 allocate 0x10000/10000,10000,0
2020-06-28 15:50:16.507367 7fc16b64f700 10 fbmap_alloc 0x55d1b0ff1400 allocate extent: 0xdd0000~10000/10000,10000,0
2020-06-28 15:50:16.507370 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0xdd0000~10000]
2020-06-28 15:50:16.507373 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write forcing blob_offset to 0x70000
2020-06-28 15:50:16.507375 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write initialize csum setting for new blob Blob(0x55d1b2031650 blob([]) use_tracker(0x0 0x0) SharedBlob(0x55d1b1b31b90 sbid 0x0)) csum_type crc32c csum_order 12 csum_length 0x80000
2020-06-28 15:50:16.507386 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write blob Blob(0x55d1b2031650 blob([!~70000,0xdd0000~10000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x55d1b1b31b90 sbid 0x0))
2020-06-28 15:50:16.507404 7fc16b64f700 20 bluestore.blob(0x55d1b2031650) get_ref 0x70000~10000 Blob(0x55d1b2031650 blob([!~70000,0xdd0000~10000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x55d1b1b31b90 sbid 0x0))
2020-06-28 15:50:16.507409 7fc16b64f700 20 bluestore.blob(0x55d1b2031650) get_ref init 0x80000, 10000
2020-06-28 15:50:16.507412 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write lex **0xf0000~10000: 0x70000~10000** Blob(0x55d1b2031650 blob([!~70000,0xdd0000~10000] csum crc32c/0x1000) use_tracker(0x8*0x10000 0x[0,0,0,0,0,0,0,10000]) SharedBlob(0x55d1b1b31b90 sbid 0x0))
2020-06-28 15:50:16.507434 7fc16b64f700 20 bluestore.BufferSpace(0x55d1b1b31ba8 in 0x55d1b13a01c0) _discard 0x70000~10000
2020-06-28 15:50:16.507446 7fc16b64f700 30 estimate gc range(hex): [f0000, 100000)
2020-06-28 15:50:16.507449 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_write extending size to 0x100000
2020-06-28 15:50:16.507452 7fc16b64f700 30 bluestore.extentmap(0x55d1b1cd9bf0) dirty_range 0xf0000~10000
2020-06-28 15:50:16.507454 7fc16b64f700 20 bluestore.extentmap(0x55d1b1cd9bf0) dirty_range mark inline shard dirty
2020-06-28 15:50:16.507457 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _write 2.2c_head #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0xf0000~10000 = 0
2020-06-28 15:50:16.507463 7fc16b64f700 30 bluestore.OnodeSpace(0x55d1b15272c8 in 0x55d1b13a01c0) lookup
2020-06-28 15:50:16.507465 7fc16b64f700 30 bluestore.OnodeSpace(0x55d1b15272c8 in 0x55d1b13a01c0) lookup #2:34000000:::head# hit 0x55d1b18bb440
2020-06-28 15:50:16.507482 7fc16b64f700 15 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000:::head#
2020-06-28 15:50:16.507488 7fc16b64f700 30 bluestore(/sandstone-data/ceph-4) _omap_setkeys 0x0000000000000466'.0000000288.00000000000000000145' <- 0000000288.0000000000000000145
2020-06-28 15:50:16.507502 7fc16b64f700 30 bluestore(/sandstone-data/ceph-4) _omap_setkeys 0x0000000000000466'._fastinfo' <- _fastinfo
2020-06-28 15:50:16.507538 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000:::head# = 0
2020-06-28 15:50:16.507546 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _txc_calc_cost 0x55d1b2219600 cost 1406898 (2 ios * 670000 + 66898 bytes)
2020-06-28 15:50:16.507549 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _txc_write_nodes txc 0x55d1b2219600 onodes 0x55d1b1cd9b00 shared_blobs

```
2020-06-28 15:50:16.507552 7fc16b64f700 20 bluestore.extentmap(0x55d1b1cd9bf0) update
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head#
2020-06-28 15:50:16.507582 7fc16b64f700 20 bluestore.extentmap(0x55d1b1cd9bf0) update  inline shard 543 bytes from 1 extents
2020-06-28 15:50:16.507589 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) onode
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# is 929 (380 bytes onode + 2 bytes spanning blobs + 547 bytes inline extents)


//// [448k, 64k]
2020-06-28 15:50:17.263329 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_write
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x70000~10000 - have 0x100000 (1048576) bytes fadvise_flags 0x0
2020-06-28 15:50:17.263338 7fc16b64f700 30 _dump_onode 0x55d1b1cd9b00 #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# nid 21011
size 0x100000 (1048576) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2020-06-28 15:50:17.263343 7fc16b64f700 30 _dump_onode  attr _ len 293
2020-06-28 15:50:17.263344 7fc16b64f700 30 _dump_onode  attr snapset len 35
2020-06-28 15:50:17.263345 7fc16b64f700 30 _dump_extent_map 0xf0000~10000: 0x70000~10000 Blob(0x55d1b2031650
blob([!~70000,0xdd0000~10000] csum crc32c/0x1000) use_tracker(0x8*0x10000 0x[0,0,0,0,0,0,0,10000]) SharedBlob(0x55d1b1b31b90 sbid 0x0))
2020-06-28 15:50:17.263361 7fc16b64f700 30 _dump_extent_map csum:
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
2020-06-28 15:50:17.263379 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2020-06-28 15:50:17.263383 7fc16b64f700 30 bluestore.extentmap(0x55d1b1cd9bf0) fault_range 0x70000~10000
2020-06-28 15:50:17.263386 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _do_write_big 0x70000~10000 target_blob_size 0x80000 compress
0
2020-06-28 15:50:17.263392 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x55d1b22182c0 1 blobs
2020-06-28 15:50:17.263393 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write buffered:0 compress:0
target_blob_size:524288 write_flags:0
write_item:  blob 0x55d1b0fc90a8 logical_offset:458752 blob_length:65536 b_off:0 b_off0:0 length0:65536 mark_unused:0 new_blob:1
compressed_len:0

2020-06-28 15:50:17.263411 7fc16b64f700 10 fbmap_alloc 0x55d1b0ff1400 allocate 0x10000/10000,10000,0
2020-06-28 15:50:17.263424 7fc16b64f700 10 fbmap_alloc 0x55d1b0ff1400 allocate extent: 0xe00000~10000/10000,10000,0
2020-06-28 15:50:17.263439 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0xe00000~10000]
2020-06-28 15:50:17.263443 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write forcing blob_offset to 0x70000
2020-06-28 15:50:17.263444 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write initialize csum setting for new blob
Blob(0x55d1b203ed20 blob([]) use_tracker(0x0 0x0) SharedBlob(0x55d1b1c432d0 sbid 0x0)) csum_type crc32c csum_order 12 csum_length 0x80000
2020-06-28 15:50:17.263451 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write blob Blob(0x55d1b203ed20
blob([!~70000,0xe00000~10000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x55d1b1c432d0 sbid 0x0))
2020-06-28 15:50:17.263464 7fc16b64f700 20 bluestore.blob(0x55d1b203ed20) get_ref 0x70000~10000 Blob(0x55d1b203ed20
blob([!~70000,0xe00000~10000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x55d1b1c432d0 sbid 0x0))
2020-06-28 15:50:17.263483 7fc16b64f700 20 bluestore.blob(0x55d1b203ed20) get_ref init 0x80000, 10000
2020-06-28 15:50:17.263487 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write lex 0x70000~10000: 0x70000~10000
Blob(0x55d1b203ed20 blob([!~70000,0xe00000~10000] csum crc32c/0x1000) use_tracker(0x8*0x10000 0x[0,0,0,0,0,0,0,10000])
SharedBlob(0x55d1b1c432d0 sbid 0x0))
2020-06-28 15:50:17.263492 7fc16b64f700 20 bluestore.BufferSpace(0x55d1b1c432e8 in 0x55d1b13a01c0) _discard 0x70000~10000
2020-06-28 15:50:17.263502 7fc16b64f700 30 estimate gc range(hex): [70000, 80000)
2020-06-28 15:50:17.263506 7fc16b64f700 30 bluestore.extentmap(0x55d1b1cd9bf0) dirty_range 0x70000~10000
2020-06-28 15:50:17.263507 7fc16b64f700 20 bluestore.extentmap(0x55d1b1cd9bf0) dirty_range mark inline shard dirty
2020-06-28 15:50:17.263512 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _write 2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x70000~10000 = 0
2020-06-28 15:50:17.263517 7fc16b64f700 30 bluestore.OnodeSpace(0x55d1b15272c8 in 0x55d1b13a01c0) lookup
2020-06-28 15:50:17.263518 7fc16b64f700 30 bluestore.OnodeSpace(0x55d1b15272c8 in 0x55d1b13a01c0) lookup #2:34000000::::head# hit
0x55d1b18bb440
2020-06-28 15:50:17.263535 7fc16b64f700 15 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000::::head#
2020-06-28 15:50:17.263539 7fc16b64f700 30 bluestore(/sandstone-data/ceph-4)
_omap_setkeys  0x0000000000000466'.0000000288.00000000000000000146' <- 0000000288.00000000000000000146
2020-06-28 15:50:17.263549 7fc16b64f700 30 bluestore(/sandstone-data/ceph-4) _omap_setkeys  0x0000000000000466'._info' <- _info
2020-06-28 15:50:17.263553 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000::::head# = 0
2020-06-28 15:50:17.263557 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _txc_calc_cost 0x55d1b22182c0 cost 1407570 (2 ios * 670000 +
67570 bytes)
2020-06-28 15:50:17.263559 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _txc_write_nodes txc 0x55d1b22182c0 onodes 0x55d1b1cd9b00
shared_blobs
2020-06-28 15:50:17.263562 7fc16b64f700 20 bluestore.extentmap(0x55d1b1cd9bf0) update
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head#
2020-06-28 15:50:17.263569 7fc16b64f700 20 bluestore.extentmap(0x55d1b1cd9bf0) update  inline shard 1083 bytes from 2 extents
2020-06-28 15:50:17.263577 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) onode
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# is 1469 (380 bytes onode + 2 bytes spanning blobs + 1087 bytes inline extents)


////// [512k,  128k]
2020-06-28 15:50:17.997019 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_write
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x80000~20000 - have 0x100000 (1048576) bytes fadvise_flags 0x0
2020-06-28 15:50:17.997024 7fc16b64f700 30 _dump_onode 0x55d1b1cd9b00 #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# nid 21011
size 0x100000 (1048576) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2020-06-28 15:50:17.997029 7fc16b64f700 30 _dump_onode  attr _ len 293
2020-06-28 15:50:17.997030 7fc16b64f700 30 _dump_onode  attr snapset len 35
2020-06-28 15:50:17.997032 7fc16b64f700 30 _dump_extent_map 0x70000~10000: 0x70000~10000 Blob(0x55d1b203ed20
blob([!~70000,0xe00000~10000] csum crc32c/0x1000) use_tracker(0x8*0x10000 0x[0,0,0,0,0,0,0,10000]) SharedBlob(0x55d1b1c432d0 sbid 0x0))
2020-06-28 15:50:17.997047 7fc16b64f700 30 _dump_extent_map csum:
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
2020-06-28 15:50:17.997061 7fc16b64f700 30 _dump_extent_map 0xf0000~10000: 0x70000~10000 Blob(0x55d1b2031650
blob([!~70000,0xdd0000~10000] csum crc32c/0x1000) use_tracker(0x8*0x10000 0x[0,0,0,0,0,0,0,10000]) SharedBlob(0x55d1b1b31b90 sbid 0x0))
2020-06-28 15:50:17.997068 7fc16b64f700 30 _dump_extent_map csum:
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
2020-06-28 15:50:17.997083 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2020-06-28 15:50:17.997087 7fc16b64f700 30 bluestore.extentmap(0x55d1b1cd9bf0) fault_range 0x80000~20000
2020-06-28 15:50:17.997090 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _do_write_big 0x80000~20000 target_blob_size 0x80000 compress
0
2020-06-28 15:50:17.997094 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_write_big reuse blob Blob(0x55d1b2031650
blob([!~70000,0xdd0000~10000] csum crc32c/0x1000) use_tracker(0x8*0x10000 0x[0,0,0,0,0,0,0,10000]) SharedBlob(0x55d1b1b31b90 sbid 0x0))
(0x0~20000)
2020-06-28 15:50:17.997101 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x55d1b22198c0 1 blobs
2020-06-28 15:50:17.997103 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write buffered:0 compress:0
target_blob_size:524288 write_flags:0
write_item:  blob 0x55d1b2051508 logical_offset:524288 blob_length:131072 b_off:0 b_off0:0 length0:131072 mark_unused:0 new_blob:0
compressed_len:0

2020-06-28 15:50:17.997110 7fc16b64f700 10 fbmap_alloc 0x55d1b0ff1400 allocate 0x20000/10000,20000,0
2020-06-28 15:50:17.997116 7fc16b64f700 10 fbmap_alloc 0x55d1b0ff1400 allocate extent: 0x2440000~20000/10000,20000,0
2020-06-28 15:50:17.997119 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0x2440000~20000]
2020-06-28 15:50:17.997124 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write blob Blob(0x55d1b2031650
blob([0x2440000~20000,!~50000,0xdd0000~10000] csum crc32c/0x1000) use_tracker(0x8*0x10000 0x[0,0,0,0,0,0,0,10000])
SharedBlob(0x55d1b1b31b90 sbid 0x0))
2020-06-28 15:50:17.997147 7fc16b64f700 20 bluestore.blob(0x55d1b2031650) get_ref 0x0~20000 Blob(0x55d1b2031650
blob([0x2440000~20000,!~50000,0xdd0000~10000] csum crc32c/0x1000) use_tracker(0x8*0x10000 0x[0,0,0,0,0,0,0,10000])
```

```
SharedBlob(0x55d1b1b31b90 sbid 0x0))
2020-06-28  15:50:17.997155  7fc16b64f700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write  lex  0x80000~20000:  0x0~20000
Blob(0x55d1b2031650  blob([0x2440000~20000,!~50000,0xdd0000~10000]  csum  crc32c/0x1000)  use_tracker(0x8*0x10000
0x[10000,10000,0,0,0,0,0,10000]) SharedBlob(0x55d1b1b31b90 sbid 0x0))
2020-06-28 15:50:17.997161 7fc16b64f700 20 bluestore.BufferSpace(0x55d1b1b31ba8 in 0x55d1b13a01c0) _discard 0x0~20000
2020-06-28 15:50:17.997173 7fc16b64f700 30 estimate gc range(hex): [80000, a0000)
2020-06-28 15:50:17.997177 7fc16b64f700 30 bluestore.extentmap(0x55d1b1cd9bf0) dirty_range 0x80000~20000
2020-06-28 15:50:17.997179 7fc16b64f700 20 bluestore.extentmap(0x55d1b1cd9bf0) dirty_range mark inline shard dirty
2020-06-28  15:50:17.997182  7fc16b64f700  10  bluestore(/sandstone-data/ceph-4)  _write  2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x80000~20000 = 0
2020-06-28 15:50:17.997187 7fc16b64f700 30 bluestore.OnodeSpace(0x55d1b15272c8 in 0x55d1b13a01c0) lookup
2020-06-28  15:50:17.997192  7fc16b64f700  30  bluestore.OnodeSpace(0x55d1b15272c8  in  0x55d1b13a01c0)  lookup  #2:34000000:::head#  hit
0x55d1b18bb440
2020-06-28 15:50:17.997197 7fc16b64f700 15 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000:::head#
2020-06-28  15:50:17.997202  7fc16b64f700  30  bluestore(/sandstone-data/ceph-4)
_omap_setkeys  0x0000000000000466'.0000000288.00000000000000000147' <- 0000000288.0000000000000000147
2020-06-28 15:50:17.997218 7fc16b64f700 30 bluestore(/sandstone-data/ceph-4) _omap_setkeys  0x0000000000000466'._info' <- _info
2020-06-28 15:50:17.997224 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000:::head# = 0
2020-06-28 15:50:17.997230 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _txc_calc_cost 0x55d1b22198c0 cost 1473106 (2 ios * 670000 +
133106 bytes)
2020-06-28 15:50:17.997233 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _txc_write_nodes txc 0x55d1b22198c0 onodes 0x55d1b1cd9b00
shared_blobs
2020-06-28  15:50:17.997238  7fc16b64f700  20  bluestore.extentmap(0x55d1b1cd9bf0)  update
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head#
2020-06-28 15:50:17.997246 7fc16b64f700 20 bluestore.extentmap(0x55d1b1cd9bf0) update  inline shard 1093 bytes from 3 extents
2020-06-28  15:50:17.997253  7fc16b64f700  20  bluestore(/sandstone-data/ceph-4)  onode
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# is 1479 (380 bytes onode + 2 bytes spanning blobs + 1097 bytes inline extents)
```

```
/////[448k, 576k],生成0x70000~80000:0x0~80000和0xf0000~10000:0x70000~10000两个extent
2020-06-29  15:44:51.237176  7fc16b64f700  20  bluestore(/sandstone-data/ceph-4)  _do_write
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x70000~90000 - have 0x0 (0) bytes fadvise_flags 0x0
2020-06-29 15:44:51.237179 7fc16b64f700 30 _dump_onode 0x55d1b1cd9b00 #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# nid 21140
size 0x0 (0) expected_object_size 1048576 expected_write_size 1048576 in 0 shards, 0 spanning blobs
2020-06-29 15:44:51.237184 7fc16b64f700 30 _dump_onode  attr _ len 293
2020-06-29 15:44:51.237185 7fc16b64f700 30 _dump_onode  attr snapset len 35
2020-06-29 15:44:51.237188 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2020-06-29 15:44:51.237191 7fc16b64f700 30 bluestore.extentmap(0x55d1b1cd9bf0) fault_range 0x70000~90000
2020-06-29 15:44:51.237194 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _do_write_big 0x70000~90000 target_blob_size 0x80000 compress
0
2020-06-29 15:44:51.237207 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x55d1b2d64840 2 blobs
2020-06-29  15:44:51.237210  7fc16b64f700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write  buffered:0  compress:0
target_blob_size:524288 write_flags:0
write_item:  blob  0x55d1b2b36708  logical_offset:458752  blob_length:524288  b_off:0  b_off0:0  length0:524288  mark_unused:0  new_blob:1
compressed_len:0
             blob  0x55d1b2b367e8  logical_offset:983040  blob_length:65536  b_off:0  b_off0:0  length0:65536  mark_unused:0  new_blob:1
compressed_len:0

2020-06-29 15:44:51.237261 7fc16b64f700 10 fbmap_alloc 0x55d1b0ff1400 allocate 0x90000/10000,90000,0
2020-06-29 15:44:51.237280 7fc16b64f700 10 fbmap_alloc 0x55d1b0ff1400 allocate extent: 0xdd0000~10000/10000,90000,0
2020-06-29 15:44:51.237282 7fc16b64f700 10 fbmap_alloc 0x55d1b0ff1400 allocate extent: 0xe00000~10000/10000,90000,0
2020-06-29 15:44:51.237283 7fc16b64f700 10 fbmap_alloc 0x55d1b0ff1400 allocate extent: 0x2440000~20000/10000,90000,0
2020-06-29 15:44:51.237286 7fc16b64f700 10 fbmap_alloc 0x55d1b0ff1400 allocate extent: 0x28e0000~50000/10000,90000,0
2020-06-29  15:44:51.237288  7fc16b64f700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write  prealloc
[0xdd0000~10000,0xe00000~10000,0x2440000~20000,0x28e0000~50000]
2020-06-29  15:44:51.237292  7fc16b64f700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write  initialize  csum  setting  for  new  blob
Blob(0x55d1b203ed20 blob([]) use_tracker(0x0 0x0) SharedBlob(0x55d1b1c432d0 sbid 0x0)) csum_type crc32c csum_order 12 csum_length 0x80000
2020-06-29  15:44:51.237310  7fc16b64f700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write  blob  Blob(0x55d1b203ed20
blob([0xdd0000~10000,0xe00000~10000,0x2440000~20000,0x28e0000~40000]  csum  crc32c/0x1000)  use_tracker(0x0  0x0)  SharedBlob(0x55d1b1c432d0
sbid 0x0))
2020-06-29  15:44:51.237424  7fc16b64f700  20  bluestore.blob(0x55d1b203ed20)  get_ref  0x0~80000  Blob(0x55d1b203ed20
blob([0xdd0000~10000,0xe00000~10000,0x2440000~20000,0x28e0000~40000]  csum  crc32c/0x1000)  use_tracker(0x0  0x0)  SharedBlob(0x55d1b1c432d0
sbid 0x0))
2020-06-29 15:44:51.237433 7fc16b64f700 20 bluestore.blob(0x55d1b203ed20) get_ref init 0x80000, 10000
2020-06-29  15:44:51.237437  7fc16b64f700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write  lex  0x70000~80000:  0x0~80000
Blob(0x55d1b203ed20  blob([0xdd0000~10000,0xe00000~10000,0x2440000~20000,0x28e0000~40000]  csum  crc32c/0x1000)  use_tracker(0x8*0x10000
0x[10000,10000,10000,10000,10000,10000,10000,10000]) SharedBlob(0x55d1b1c432d0 sbid 0x0))
2020-06-29 15:44:51.237454 7fc16b64f700 20 bluestore.BufferSpace(0x55d1b1c432e8 in 0x55d1b13a01c0) _discard 0x0~80000
2020-06-29 15:44:51.237467 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write forcing blob_offset to 0x70000
2020-06-29  15:44:51.237468  7fc16b64f700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write  initialize  csum  setting  for  new  blob
Blob(0x55d1b2675880 blob([]) use_tracker(0x0 0x0) SharedBlob(0x55d1b23b8d90 sbid 0x0)) csum_type crc32c csum_order 12 csum_length 0x80000
2020-06-29  15:44:51.239214  7fc16b64f700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write  blob  Blob(0x55d1b2675880
blob([!~70000,0x2920000~10000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x55d1b23b8d90 sbid 0x0))
2020-06-29  15:44:51.239261  7fc16b64f700  20  bluestore.blob(0x55d1b2675880)  get_ref  0x70000~10000  Blob(0x55d1b2675880
blob([!~70000,0x2920000~10000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x55d1b23b8d90 sbid 0x0))
2020-06-29 15:44:51.239268 7fc16b64f700 20 bluestore.blob(0x55d1b2675880) get_ref init 0x80000, 10000
2020-06-29  15:44:51.239272  7fc16b64f700  20  bluestore(/sandstone-data/ceph-4)  _do_alloc_write  lex  0xf0000~10000:  0x70000~10000
Blob(0x55d1b2675880  blob([!~70000,0x2920000~10000]  csum  crc32c/0x1000)  use_tracker(0x8*0x10000  0x[0,0,0,0,0,0,0,10000])
SharedBlob(0x55d1b23b8d90 sbid 0x0))
2020-06-29 15:44:51.239277 7fc16b64f700 20 bluestore.BufferSpace(0x55d1b23b8da8 in 0x55d1b13a01c0) _discard 0x70000~10000
2020-06-29 15:44:51.239287 7fc16b64f700 30 estimate gc range(hex): [70000, 100000)
2020-06-29 15:44:51.239299 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _do_write extending size to 0x100000
2020-06-29 15:44:51.239303 7fc16b64f700 30 bluestore.extentmap(0x55d1b1cd9bf0) dirty_range 0x70000~90000
2020-06-29 15:44:51.239305 7fc16b64f700 20 bluestore.extentmap(0x55d1b1cd9bf0) dirty_range mark inline shard dirty
2020-06-29  15:44:51.239308  7fc16b64f700  10  bluestore(/sandstone-data/ceph-4)  _write  2.2c_head
#2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# 0x70000~90000 = 0
2020-06-29 15:44:51.239328 7fc16b64f700 30 bluestore.OnodeSpace(0x55d1b15272c8 in 0x55d1b13a01c0) lookup
2020-06-29  15:44:51.239331  7fc16b64f700  30  bluestore.OnodeSpace(0x55d1b15272c8  in  0x55d1b13a01c0)  lookup  #2:34000000:::head#  hit
0x55d1b18bb440
2020-06-29 15:44:51.239336 7fc16b64f700 15 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000:::head#
2020-06-29  15:44:51.239341  7fc16b64f700  30  bluestore(/sandstone-data/ceph-4)
_omap_setkeys  0x0000000000000466'.0000000318.00000000000000000150' <- 0000000318.0000000000000000150
2020-06-29 15:44:51.239357 7fc16b64f700 30 bluestore(/sandstone-data/ceph-4) _omap_setkeys  0x0000000000000466'._fastinfo' <- _fastinfo
2020-06-29 15:44:51.239379 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.2c_head #2:34000000:::head# = 0
```

```
2020-06-29 15:44:51.239388 7fc16b64f700 10 bluestore(/sandstone-data/ceph-4) _txc_calc_cost 0x55d1b2d64840 cost 4611186 (6 ios * 670000 + 591186 bytes)
2020-06-29 15:44:51.239392 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) _txc_write_nodes txc 0x55d1b2d64840 onodes 0x55d1b1cd9b00 shared_blobs
2020-06-29 15:44:51.239394 7fc16b64f700 20 bluestore.extentmap(0x55d1b1cd9bf0) update #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head#
2020-06-29 15:44:51.239403 7fc16b64f700 20 bluestore.extentmap(0x55d1b1cd9bf0) update  inline shard 1086 bytes from 2 extents
2020-06-29 15:44:51.239409 7fc16b64f700 20 bluestore(/sandstone-data/ceph-4) onode #2:35b36518:::rbd_data.10557d0a41f0.0000000000000001:head# is 1472 (380 bytes onode + 2 bytes spanning blobs + 1090 bytes inline extents)
```

BlueStore GC

压缩模式由bluestore_compression_mode参数设置，默认是none，即是不压缩。开启压缩才会走GC流程。

```
BlueStore::GarbageCollector::estimate()


BlueStore::_do_gc()
```

BlueStore快照实现

```
[root@node3]# /var/lib/ceph/bin/ceph-kvstore-tool rocksdb db/ list X
X    %00%00%00%00%00%01%40%01
X    %00%00%00%00%00%01%40%02
X    %00%00%00%00%00%01%40%03
X    %00%00%00%00%00%01%40%04
[root@node3]# /var/lib/ceph/bin/ceph-kvstore-tool rocksdb db/ get X %00%00%00%00%00%01%40%04 out X.bin
[root@node3 osd_4_db]# /root/ceph-dencoder import X.bin  type bluestore_extent_ref_map_t skip 6 decode dump_json
{
    "ref_map": [
        {
            "offset": 24051712, // 0x16f0000, 编码用3Bytes
            "length": 65536,    // 编码用1Byte
            "refs": 2           // 有两个对象在引用该pextent, 编码用1Byte
        }
    ]
}
```

extent map reshard

bluestone元数据的组织结构

```
bluestore_extent_map_inline_shard_prealloc_size = 256 //
bluestore_extent_map_shard_max_size = 1200  // ExtentMap inline编码的最大长度, 超过后则进行reshard; shard分片的最大长度
bluestore_extent_map_shard_min_size = 150 // shard分片的最小长度, 小于该长度则尝试向前后合并
bluestore_extent_map_shard_target_size = 500 //
bluestore_extent_map_shard_target_size_slop = 0.2 //


bluestore对象的元数据包含对象大小, 对象在store层唯一标识, attr属性, 对象的extent map信息, 对象数据的crc, 对象的omap属性等. 元数据的持久化分为两大类, 以'O'为前缀的kv和以'M'为前缀的kv, 其中'M'前缀的kv保存对象的omap属性, 除此外的元数据都是保存在'O'前缀的kv中.
根据对象的元数据的大小持久化方式分为两种：inline模式和reshard模式.

inline模式指ExtentMap inline编码, 对象元数据仅仅存在一个'O'前缀的k/v中, 编码格式:
bluestore_onode_t + {struct_v(1B) + spanning_blob_map_num(1B)} + extent_map_inline_bl

(1) 编码bluestore_onode_t结构
denc(o->onode, bound), 即是:
DENC(bluestore_onode_t, v, p) {
    DENC_START(1, 1, p); // struct_v(1B) + struct_compat(1B) + struct_len(4B)
    denc_varint(v.nid, p); // uint64_t类型, 可变1~10B
    denc_varint(v.size, p); // uint64_t类型, 可变1~10B
      denc(v.attrs, p); // map 类 型 , {attrs_num(4B) + {key_size(4B) + key(1B:'_') + value_size(4B) + value(295B) + key_size(4B) + key(7B:'snapset') + value_size(4B) + value(35B)}}, 这部分占用空间最多的!!!
    denc(v.flags, p); // uint8_t类型, 1B
    denc(v.extent_map_shards, p); // vector类型, {extent_map_shards_num(4B) + {shard_offset(0) + shard_bytes(0)}}, inline模式下数组为空
    denc_varint(v.expected_object_size, p); // uint32_t类型, 可变1~5B
    denc_varint(v.expected_write_size, p); // uint32_t类型, 可变1~5B
    denc_varint(v.alloc_hint_flags, p); // uint32_t类型, 可变1~5B
    DENC_FINISH(p);
  }
(2) 编码ExtentMap::spanning_blob_map
o->extent_map.bound_encode_spanning_blobs(bound) // struct_v(1B) + spanning_blob_map_size(4B), inline模式下map为空所以这固定是2B
```

```
(3) extent_map_shards为empty, 则表示inline模式
denc(o->extent_map.inline_bl, bound);
其中inline_bl的内存包含有:
// 版本和extents数量
struct_v       // 1Byte
extents数量    // 可变长度1Byte ~ 5Bytes

// 每个Extent
blobid         // 可变长度1Byte ~ 5Bytes
logical_offset // 可变长度1Byte ~ 5Bytes
length         // 可变长度1Byte ~ 5Bytes
blob_offset    // 可变长度1Byte ~ 5Bytes


// Extent对应的bluestore_blob_t
extents     // vector类型, blob对应的pextent, 编码格式如{pextents_num(1B) + {poffset(4B) + plength(2B)}...}
flags       // uint32_t类型, 可变长度1Byte ~ 5Bytes
logical_length // uint32_t类型, 可变长度1Byte ~ 5Bytes, 开启压缩才会编码
compressed_length // uint32_t类型, 可变长度1Byte ~ 5Bytes, 开启压缩才会编码
csum_type         // uint8_t类型
csum_chunk_order  // uint8_t类型
csum_data.length() // uint32_t类型, csum_data的长度
csum_data          // csum_data buffer, 这部分占用空间最多的!!!
unused             // uint16_t类型

// Extent对应的sbid(shared时)
sbid // uint64_t类型, 可变1~10B

// Extent对应的bluestore_blob_use_tracker_t(reshard时)
au_size == 0
au_size // uint32_t类型, 1Byte

au_size != 0 && num_au == 0
au_size // uint32_t类型, 1Byte
num_au // uint32_t类型, 1Byte
total_bytes // uint32_t类型, 1Byte~5Bytes

au_size != 0 && num_au != 0
au_size // uint32_t类型, 1Byte~5Bytes
num_au // uint32_t类型, 1Byte~5Bytes
elem_size // uint32_t类型, 1Byte~5Bytes
bytes_per_au // uint32_t类型, 1Byte~5Bytes, 总有elem_size个

总结:
bluestore在编码extent map时做了很多节省空间的优化: (1)所有字段都采用变长编码方式编码; (2)对不同特点的extent优化以节省logical_offset,length,blob_offset字段的编码空间; (3)引用相同blob的不同extent只有在编码第一个
extent时会编码blob信息,避免blob重复编码;
BLOBID_FLAG_CONTIGUOUS: this extent starts at end of previous
BLOBID_FLAG_ZEROOFFSET: blob_offset is 0
BLOBID_FLAG_SAMELENGTH: length matches previous extent



reshard模式指ExtentMap分多个shard, 每个shard的extents对应一个独立的k/v保存, 独立的k/v第一个key中可能也编码了某些extent对应的bluestore_blob_t, 编码格式:
blobid(1B) + {logical_offset + blob_offset + length} + {bluestore_blob_t + ...}
reshard模式下Onode元数据的编码格式为:
bluestore_onode_t + {struct_v(1B) + spanning_blob_map_num + {bluestore_blob_t + ...}} + extent_map_inline_bl(0)
举例:
O                                                                                                                %7f%80%00%00%00%00%00%02%e4.%9d-
%21rbd_data.10557d0a41f0.0000000000000000%21%3d%ff%ff%ff%ff%ff%ff%ff%fe%ff%ff%ff%ff%ff%ff%ff%ffo // onode
O                                                                                                                %7f%80%00%00%00%00%00%02%e4.%9d-
%21rbd_data.10557d0a41f0.0000000000000000%21%3d%ff%ff%ff%ff%ff%ff%ff%fe%ff%ff%ff%ff%ff%ff%ff%ffo%00%00%00%00x // shard offset: 0x0
O                                                                                                                %7f%80%00%00%00%00%00%02%e4.%9d-
%21rbd_data.10557d0a41f0.0000000000000000%21%3d%ff%ff%ff%ff%ff%ff%ff%fe%ff%ff%ff%ff%ff%ff%ff%ffo%00%02%08%00x // shard offset: 0x20800
O                                                                                                                %7f%80%00%00%00%00%00%02%e4.%9d-
%21rbd_data.10557d0a41f0.0000000000000000%21%3d%ff%ff%ff%ff%ff%ff%ff%fe%ff%ff%ff%ff%ff%ff%ff%ffo%00%09%e0%00x // shard offset: 0x9e000


从上面分析可知extent对应的bluestore_blob_t可能存在多个k/v中保存
```

日志分析:

```
// 第一次分shard
2020-07-28 17:41:15.819061 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) dirty_range 0xaf000~1000
2020-07-28 17:41:15.819062 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) dirty_range mark inline shard dirty
2020-07-28 17:41:15.819064 7f351d5fc700 10 bluestore(/sandstone-data/ceph-4) _write 2.27_head
#2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# 0xaf000~1000 = 0
2020-07-28 17:41:15.819069 7f351d5fc700 10 bluestore(/sandstone-data/ceph-4) _txc_calc_cost 0x55b5f8a65b80 cost 676200 (1 ios * 670000 +
6200 bytes)
2020-07-28 17:41:15.819072 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _txc_write_nodes txc 0x55b5f8a65b80 onodes 0x55b5f8abcb40
shared_blobs
2020-07-28 17:41:15.819076 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) update
#2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head#
2020-07-28 17:41:15.819089 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) update inline shard 1339 bytes from 8 extents // 超过
bluestore_extent_map_shard_max_size=1200Bytes, 需要reshard
2020-07-28 17:41:15.819108 7f351d5fc700 10 bluestore.extentmap(0x55b5f8abcc30) reshard 0x[0,ffffffff) of 0 shards on
#2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# // 整对象reshard
2020-07-28 17:41:15.819120 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) fault_range 0x0~ffffffff
2020-07-28 17:41:15.819122 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard extent_avg 167, target 500, slop 100
// extent_avg=1339 / 8 = 167, slop = target * 0.2 = 100
2020-07-28 17:41:15.819130 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x1800~1000: 0x1800~1000 Blob(0x55b5f8969180
blob([0x2ec0000~10000] csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55b5f22a2150 sbid 0x0))
```

```
2020-07-28 17:41:15.819139 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0xf800~800: 0xf800~800 Blob(0x55b5f8969180 blob([0x2ec0000~10000] csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55b5f22a2150 sbid 0x0))
2020-07-28 17:41:15.819142 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x10000~800: 0x10000~800 Blob(0x55b5f6a88c40 blob([!~10000,0x2ed0000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,800]) SharedBlob(0x55b5f8a1ec40 sbid 0x0))
2020-07-28 17:41:15.819145 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x20800~1000: 0x20800~1000 Blob(0x55b5f8968cb0 blob([!~20000,0x32a0000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55b5f8968d90 sbid 0x0))
2020-07-28 17:41:15.819149 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard new shard 0x0
2020-07-28 17:41:15.819151 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard new shard 0x20800
2020-07-28 17:41:15.819152 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x30000~30000: 0x30000~30000 Blob(0x55b5f89688c0 blob([!~30000,0x32b0000~30000] csum crc32c/0x1000) use_tracker(0x6*0x10000 0x[0,0,0,10000,10000,10000]) SharedBlob(0x55b5f8968e70 sbid 0x0))
2020-07-28 17:41:15.819155 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x8c800~1000: 0xc800~1000 Blob(0x55b5f8968ee0 blob([0x32e0000~10000] csum+has_unused crc32c/0x1000 unused=0xcfff) use_tracker(0x10000 0x1000) SharedBlob(0x55b5f8968f50 sbid 0x0))
2020-07-28 17:41:15.819158 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x9e000~2000: 0x1e000~2000 Blob(0x55b5f884d7a0 blob([!~10000,0x32f0000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x2*0x10000 0x[0,2000]) SharedBlob(0x55b5f8968af0 sbid 0x0))
2020-07-28 17:41:15.819161 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard new shard 0x9e000
2020-07-28 17:41:15.819162 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0xaf000~1000: 0x2f000~1000 Blob(0x55b5f884cd20 blob([!~20000,0x3300000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55b5f8968070 sbid 0x0))
2020-07-28 17:41:15.819166 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard new [0x0(0x0 bytes),0x20800(0x0 bytes),0x9e000(0x0 bytes)] // 分成3个shard
2020-07-28 17:41:15.819169 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard old []
2020-07-28 17:41:15.819171 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard fin [0x0(0x0 bytes),0x20800(0x0 bytes),0x9e000(0x0 bytes)]
2020-07-28 17:41:15.819173 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard checking spanning blobs 0x[0,ffffffff)
2020-07-28 17:41:15.819175 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x1800~1000: 0x1800~1000 Blob(0x55b5f8969180 blob([0x2ec0000~10000] csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55b5f22a2150 sbid 0x0))
2020-07-28 17:41:15.819178 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard shard_start: 0 shard_end: 133120 extent 0x1800~1000: 0x1800~1000 Blob(0x55b5f8969180 blob([0x2ec0000~10000] csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55b5f22a2150 sbid 0x0))
2020-07-28 17:41:15.819182 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0xf800~800: 0xf800~800 Blob(0x55b5f8969180 blob([0x2ec0000~10000] csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55b5f22a2150 sbid 0x0))
2020-07-28 17:41:15.819185 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard shard_start: 0 shard_end: 133120 extent 0xf800~800: 0xf800~800 Blob(0x55b5f8969180 blob([0x2ec0000~10000] csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55b5f22a2150 sbid 0x0))
2020-07-28 17:41:15.819190 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x10000~800: 0x10000~800 Blob(0x55b5f6a88c40 blob([!~10000,0x2ed0000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,800]) SharedBlob(0x55b5f8a1ec40 sbid 0x0))
2020-07-28 17:41:15.819193 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard shard_start: 0 shard_end: 133120 extent 0x10000~800: 0x10000~800 Blob(0x55b5f6a88c40 blob([!~10000,0x2ed0000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,800]) SharedBlob(0x55b5f8a1ec40 sbid 0x0))
2020-07-28 17:41:15.819197 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x20800~1000: 0x20800~1000 Blob(0x55b5f8968cb0 blob([!~20000,0x32a0000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55b5f8968d90 sbid 0x0))
2020-07-28 17:41:15.819200 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) reshard shard 0x20800 to 0x9e000
2020-07-28 17:41:15.819201 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard cann't split Blob(0x55b5f8968cb0 blob([!~20000,0x32a0000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55b5f8968d90 sbid 0x0))
2020-07-28 17:41:15.819205 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard adding spanning Blob(0x55b5f8968cb0 spanning 0 blob([!~20000,0x32a0000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55b5f8968d90 sbid 0x0))
2020-07-28 17:41:15.819222 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x30000~30000: 0x30000~30000 Blob(0x55b5f89688c0 blob([!~30000,0x32b0000~30000] csum crc32c/0x1000) use_tracker(0x6*0x10000 0x[0,0,0,10000,10000,10000]) SharedBlob(0x55b5f8968e70 sbid 0x0))
2020-07-28 17:41:15.819226 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard cann't split shard_offset:133120 blob_offset:133120 Blob(0x55b5f89688c0 blob([!~30000,0x32b0000~30000] csum crc32c/0x1000) use_tracker(0x6*0x10000 0x[0,0,0,10000,10000,10000]) SharedBlob(0x55b5f8968e70 sbid 0x0))
2020-07-28 17:41:15.819230 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard adding spanning Blob(0x55b5f89688c0 spanning 1 blob([!~30000,0x32b0000~30000] csum crc32c/0x1000) use_tracker(0x6*0x10000 0x[0,0,0,10000,10000,10000]) SharedBlob(0x55b5f8968e70 sbid 0x0))
2020-07-28 17:41:15.819233 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x8c800~1000: 0xc800~1000 Blob(0x55b5f8968ee0 blob([0x32e0000~10000] csum+has_unused crc32c/0x1000 unused=0xcfff) use_tracker(0x10000 0x1000) SharedBlob(0x55b5f8968f50 sbid 0x0))
2020-07-28 17:41:15.819237 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard shard_start: 133120 shard_end: 647168 extent 0x8c800~1000: 0xc800~1000 Blob(0x55b5f8968ee0 blob([0x32e0000~10000] csum+has_unused crc32c/0x1000 unused=0xcfff) use_tracker(0x10000 0x1000) SharedBlob(0x55b5f8968f50 sbid 0x0))
2020-07-28 17:41:15.819244 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x9e000~2000: 0x1e000~2000 Blob(0x55b5f884d7a0 blob([!~10000,0x32f0000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x2*0x10000 0x[0,2000]) SharedBlob(0x55b5f8968af0 sbid 0x0))
2020-07-28 17:41:15.819247 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) reshard shard 0x9e000 to 0xffffffff
2020-07-28 17:41:15.819248 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard cann't split Blob(0x55b5f884d7a0 blob([!~10000,0x32f0000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x2*0x10000 0x[0,2000]) SharedBlob(0x55b5f8968af0 sbid 0x0))
2020-07-28 17:41:15.819251 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard adding spanning Blob(0x55b5f884d7a0 spanning 2 blob([!~10000,0x32f0000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x2*0x10000 0x[0,2000]) SharedBlob(0x55b5f8968af0 sbid 0x0))
2020-07-28 17:41:15.819255 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0xaf000~1000: 0x2f000~1000 Blob(0x55b5f884cd20 blob([!~20000,0x3300000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55b5f8968070 sbid 0x0))
2020-07-28 17:41:15.819258 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard cann't split Blob(0x55b5f884cd20 blob([!~20000,0x3300000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55b5f8968070 sbid 0x0))
2020-07-28 17:41:15.819261 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard adding spanning Blob(0x55b5f884cd20 spanning 3 blob([!~20000,0x3300000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55b5f8968070 sbid 0x0))
2020-07-28 17:41:15.819265 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) update #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# force
2020-07-28 17:41:15.819270 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) update shard 0x0 is 241 bytes (was 0) from 3 extents
2020-07-28 17:41:15.819273 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) update shard 0x20800 is 97 bytes (was 0) from 3 extents
2020-07-28 17:41:15.819278 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) update shard 0x9e000 is 12 bytes (was 0) from 2 extents
2020-07-28 17:41:15.819288 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) onode #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# is 1430 (391 bytes onode + 1039 bytes spanning blobs + 0 bytes inline extents)
……
// 分shard后第一次写
2020-07-28 21:52:20.109639 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _choose_write_options prefer csum_order 12 target_blob_size 0x80000 compress=0 buffered=0
2020-07-28 21:52:20.109643 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) fault_range 0xb0000~1000
2020-07-28 21:52:20.109648 7f351d5fc700 10 bluestore(/sandstone-data/ceph-4) _do_write_small 0xb0000~1000
2020-07-28 21:52:20.109652 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) fault_range 0x30000~100000
2020-07-28 21:52:20.109657 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _do_write_small considering Blob(0x55b5f884cd20 spanning 3 blob([!~20000,0x3300000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55b5f8968070 sbid 0x0)) bstart 0x80000
```

```
2020-07-28 21:52:20.109664 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _do_write_small considering Blob(0x55b5f884d7a0 spanning 2 blob([!~10000,0x32f0000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x2*0x10000 0x[0,2000]) SharedBlob(0x55b5f8968af0 sbid 0x0)) bstart 0x80000
2020-07-28 21:52:20.109669 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _do_write_small considering Blob(0x55b5f8968ee0 blob([0x32e0000~10000] csum+has_unused crc32c/0x1000 unused=0xcfff) use_tracker(0x10000 0x1000) SharedBlob(0x55b5f8968f50 sbid 0x0)) bstart 0x80000
2020-07-28 21:52:20.109674 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _do_write_small considering Blob(0x55b5f89688c0 spanning 1 blob([!~30000,0x32b0000~30000] csum crc32c/0x1000) use_tracker(0x6*0x10000 0x[0,0,0,10000,10000,10000]) SharedBlob(0x55b5f8968e70 sbid 0x0)) bstart 0x0
2020-07-28 21:52:20.109681 7f351d5fc700 30 bluestore(/sandstone-data/ceph-4) _pad_zeros 0x0~1000 chunk_size 0x1000
2020-07-28 21:52:20.109684 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _pad_zeros pad 0x0 + 0x0 on front/back, now 0x0~1000
2020-07-28 21:52:20.109689 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x55b5f204d8c0 1 blobs
2020-07-28 21:52:20.109692 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write buffered:0 compress:0 target_blob_size:524288 write_flags:524288
write_item: blob 0x55b5f93b1a48 logical_offset:720896 blob_length:65536 b_off:0 b_off0:0 length0:4096 mark_unused:1 new_blob:1 compressed_len:0
2020-07-28 21:52:20.109700 7f351d5fc700 10 fbmap_alloc 0x55b5f1aa9400 allocate 0x10000/10000,10000,0
2020-07-28 21:52:20.109705 7f351d5fc700 10 fbmap_alloc 0x55b5f1aa9400 allocate extent: 0x2f80000~10000/10000,10000,0
2020-07-28 21:52:20.109708 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0x2f80000~10000]
2020-07-28 21:52:20.109711 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write forcing csum_order to block_size_order 12
2020-07-28 21:52:20.109713 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write forcing blob_offset to 0x30000
2020-07-28 21:52:20.109714 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write initialize csum setting for new blob Blob(0x55b5f22a3e30 blob([]) use_tracker(0x0 0x0) SharedBlob(0x55b5f22a3ea0 sbid 0x0)) csum_type crc32c csum_order 12 csum_length 0x40000
2020-07-28 21:52:20.109727 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write blob Blob(0x55b5f22a3e30 blob([!~30000,0x2f80000~10000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x55b5f22a3ea0 sbid 0x0))
2020-07-28 21:52:20.109744 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write b_off=196608 b_end=200704 blob_length 65536
2020-07-28 21:52:20.109748 7f351d5fc700 20 bluestore.blob(0x55b5f22a3e30) get_ref 0x30000~1000 Blob(0x55b5f22a3e30 blob([!~30000,0x2f80000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x0 0x0) SharedBlob(0x55b5f22a3ea0 sbid 0x0))
2020-07-28 21:52:20.109754 7f351d5fc700 20 bluestore.blob(0x55b5f22a3e30) get_ref init 0x40000, 10000
2020-07-28 21:52:20.109758 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write lex 0xb0000~1000: 0x30000~1000 Blob(0x55b5f22a3e30 blob([!~30000,0x2f80000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x4*0x10000 0x[0,0,0,1000]) SharedBlob(0x55b5f22a3ea0 sbid 0x0))
2020-07-28 21:52:20.109765 7f351d5fc700 20 bluestore.BufferSpace(0x55b5f22a3eb8 in 0x55b5f1db4620) _discard 0x30000~1000
2020-07-28 21:52:20.109771 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write deferring small 0x1000 write via deferred
2020-07-28 21:52:20.109777 7f351d5fc700 30 estimate gc range(hex): [b0000, b1000)
2020-07-28 21:52:20.109785 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _do_write extending size to 0xb1000
2020-07-28 21:52:20.109789 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) dirty_range 0xb0000~1000
2020-07-28 21:52:20.109792 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) dirty_range mark shard 0x9e000 dirty
2020-07-28 21:52:20.109795 7f351d5fc700 10 bluestore(/sandstone-data/ceph-4) _write 2.27_head #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# 0xb0000~1000 = 0
2020-07-28 21:52:20.109814 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _txc_write_nodes txc 0x55b5f204d8c0 onodes 0x55b5f8abcb40 shared_blobs
2020-07-28 21:52:20.109817 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) update #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head#
2020-07-28 21:52:20.109823 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) encode_some 0x9e000~fff61fff hit new spanning blob 0xb0000~1000: 0x30000~1000 Blob(0x55b5f22a3e30 blob([!~30000,0x2f80000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x4*0x10000 0x[0,0,0,1000]) SharedBlob(0x55b5f22a3ea0 sbid 0x0)) //// 新分配的blob
2020-07-28 21:52:20.109832 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) update shard 0x9e000 is 0 bytes (was 12) from 2 extents
2020-07-28 21:52:20.109836 7f351d5fc700 10 bluestore.extentmap(0x55b5f8abcc30) reshard 0x[0,c0000) of 3 shards on #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head#
2020-07-28 21:52:20.109840 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard spanning blob 0 Blob(0x55b5f8968cb0 spanning 0 blob([!~20000,0x32a0000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55b5f8968d90 sbid 0x0))
2020-07-28 21:52:20.109850 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard spanning blob 1 Blob(0x55b5f89688c0 spanning 1 blob([!~30000,0x32b0000~30000] csum crc32c/0x1000) use_tracker(0x6*0x10000 0x[0,0,0,10000,10000,10000]) SharedBlob(0x55b5f8968e70 sbid 0x0))
2020-07-28 21:52:20.109856 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard spanning blob 2 Blob(0x55b5f884d7a0 spanning 2 blob([!~10000,0x32f0000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x2*0x10000 0x[0,2000]) SharedBlob(0x55b5f8968af0 sbid 0x0))
2020-07-28 21:52:20.109861 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard spanning blob 3 Blob(0x55b5f884cd20 spanning 3 blob([!~20000,0x3300000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55b5f8968070 sbid 0x0))
2020-07-28 21:52:20.109867 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard shards [0,3) over 0x[0,ffffffff)
2020-07-28 21:52:20.109870 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) fault_range 0x0~ffffffff
2020-07-28 21:52:20.109877 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard extent_avg 43, target 500, slop 100
2020-07-28 21:52:20.109880 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x1800~1000: 0x1800~1000 Blob(0x55b5f8969180 blob([0x2ec0000~10000] csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55b5f22a2150 sbid 0x0))
2020-07-28 21:52:20.109886 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0xf800~800: 0xf800~800 Blob(0x55b5f8969180 blob([0x2ec0000~10000] csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55b5f22a2150 sbid 0x0))
2020-07-28 21:52:20.109891 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x10000~800: 0x10000~800 Blob(0x55b5f6a88c40 blob([!~10000,0x2ed0000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,800]) SharedBlob(0x55b5f8a1ec40 sbid 0x0))
2020-07-28 21:52:20.109897 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x20800~1000: 0x20800~1000 Blob(0x55b5f8968cb0 spanning 0 blob([!~20000,0x32a0000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55b5f8968d90 sbid 0x0))
2020-07-28 21:52:20.109903 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x30000~30000: 0x30000~30000 Blob(0x55b5f89688c0 spanning 1 blob([!~30000,0x32b0000~30000] csum crc32c/0x1000) use_tracker(0x6*0x10000 0x[0,0,0,10000,10000,10000]) SharedBlob(0x55b5f8968e70 sbid 0x0))
2020-07-28 21:52:20.109909 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x8c800~1000: 0xc800~1000 Blob(0x55b5f8968ee0 blob([0x32e0000~10000] csum+has_unused crc32c/0x1000 unused=0xcfff) use_tracker(0x10000 0x1000) SharedBlob(0x55b5f8968f50 sbid 0x0))
2020-07-28 21:52:20.109914 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0x9e000~2000: 0x1e000~2000 Blob(0x55b5f884d7a0 spanning 2 blob([!~10000,0x32f0000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x2*0x10000 0x[0,2000]) SharedBlob(0x55b5f8968af0 sbid 0x0))
2020-07-28 21:52:20.109921 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0xaf000~1000: 0x2f000~1000 Blob(0x55b5f884cd20 spanning 3 blob([!~20000,0x3300000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55b5f8968070 sbid 0x0))
2020-07-28 21:52:20.109927 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) extent 0xb0000~1000: 0x30000~1000 Blob(0x55b5f22a3e30 blob([!~30000,0x2f80000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x4*0x10000 0x[0,0,0,1000]) SharedBlob(0x55b5f22a3ea0 sbid 0x0))
2020-07-28 21:52:20.109933 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard new []
2020-07-28 21:52:20.109935 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard old [0x0(0xf1 bytes),0x20800(0x61 bytes),0x9e000(0xc bytes)]
2020-07-28 21:52:20.109939 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) reshard fin []
2020-07-28 21:52:20.109941 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) reshard un-spanning Blob(0x55b5f8968cb0 blob([!~20000,0x32a0000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55b5f8968d90 sbid 0x0))
2020-07-28 21:52:20.109947 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) reshard un-spanning Blob(0x55b5f89688c0 blob([!~30000,0x32b0000~30000] csum crc32c/0x1000) use_tracker(0x6*0x10000 0x[0,0,0,10000,10000,10000]) SharedBlob(0x55b5f8968e70 sbid 0x0))
2020-07-28 21:52:20.109953 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) reshard un-spanning Blob(0x55b5f884d7a0 blob([!~10000,0x32f0000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x2*0x10000 0x[0,2000]) SharedBlob(0x55b5f8968af0 sbid 0x0))
2020-07-28 21:52:20.109958 7f351d5fc700 30 bluestore.extentmap(0x55b5f8abcc30) reshard un-spanning Blob(0x55b5f884cd20 blob([!~20000,0x3300000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55b5f8968070 sbid 0x0))
```

2020-07-28 21:52:20.109970 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) update #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# force
2020-07-28 21:52:20.110016 7f351d5fc700 20 bluestore.extentmap(0x55b5f8abcc30) update inline shard 1623 bytes from 9 extents
2020-07-28 21:52:20.110025 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) onode #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# is 2009 (380 bytes onode + 2 bytes spanning blobs + 1627 bytes inline extents)
2020-07-28 21:52:20.110102 7f351d5fc700 20 bluestore(/sandstone-data/ceph-4) _txc_finalize_kv txc 0x55b5f204d8c0 allocated 0x[2f80000~10000] released 0x[]

2020-06-01 11:11:23.959765 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _txc_write_nodes txc 0x55973bf242c0 onodes 0x55973c434d80 shared_blobs
2020-06-01 11:11:23.959768 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) update #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head#
2020-06-01 11:11:23.959783 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) update inline shard 1546 bytes from 9 extents
2020-06-01 11:11:23.959809 7f2ad2390700 10 bluestore.extentmap(0x55973c434e70) reshard 0x[0,ffffffff) of 0 shards on #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head#
2020-06-01 11:11:23.959818 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) fault_range 0x0~ffffffff
2020-06-01 11:11:23.959820 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) reshard extent_avg 171, target 500, slop 100
2020-06-01 11:11:23.959829 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) extent 0x1800~1000: 0x1800~1000 Blob(0x55973c3a7ea0 blob([0xd20000~10000] csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55973c6cfe30 sbid 0x0))
2020-06-01 11:11:23.959839 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) extent 0xf800~800: 0xf800~800 Blob(0x55973c3a7ea0 blob([0xd20000~10000] csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55973c6cfe30 sbid 0x0))
2020-06-01 11:11:23.959846 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) extent 0x10000~800: 0x10000~800 Blob(0x55973c6cff10 blob([!~10000,0xd30000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,800]) SharedBlob(0x55973c6cff80 sbid 0x0))
2020-06-01 11:11:23.959888 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) extent 0x20800~1000: 0x20800~1000 Blob(0x55973c728850 blob([!~20000,0xd80000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55973c7288c0 sbid 0x0))
2020-06-01 11:11:23.959894 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) reshard new shard 0x0
2020-06-01 11:11:23.959896 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) reshard new shard 0x20800
2020-06-01 11:11:23.959897 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) extent 0x30000~30000: 0x30000~30000 Blob(0x55973c729030 blob([!~30000,0xd90000~10000,0xda0000~20000] csum crc32c/0x1000) use_tracker(0x6*0x10000 0x[0,0,0,10000,10000,10000]) SharedBlob(0x55973c7290a0 sbid 0x0))
2020-06-01 11:11:23.959912 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) extent 0x8c800~1000: 0xc800~1000 Blob(0x55973c728d90 blob([0xdd0000~10000,0xdc0000~10000] csum+has_unused crc32c/0x1000 unused=0xbf) use_tracker(0x2*0x10000 0x[1000,2000]) SharedBlob(0x55973c6c7c70 sbid 0x0))
2020-06-01 11:11:23.959917 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) extent 0x9e000~2000: 0x1e000~2000 Blob(0x55973c728d90 blob([0xdd0000~10000,0xdc0000~10000] csum+has_unused crc32c/0x1000 unused=0xbf) use_tracker(0x2*0x10000 0x[1000,2000]) SharedBlob(0x55973c6c7c70 sbid 0x0))
2020-06-01 11:11:23.959922 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) reshard new shard 0x9e000
2020-06-01 11:11:23.959923 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) extent 0xaf000~1000: 0x2f000~1000 Blob(0x55973c2564d0 blob([!~20000,0xf40000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55973cd35b90 sbid 0x0))
2020-06-01 11:11:23.959927 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) extent 0xb0000~1000: 0x30000~1000 Blob(0x55973cda72d0 blob([!~30000,0xf50000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x4*0x10000 0x[0,0,0,1000]) SharedBlob(0x55973cda65b0 sbid 0x0))
2020-06-01 11:11:23.959959 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) reshard new [0x0(0x0 bytes),0x20800(0x0 bytes),0x9e000(0x0 bytes)]
2020-06-01 11:11:23.959964 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) reshard old []
2020-06-01 11:11:23.959968 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) reshard fin [0x0(0x0 bytes),0x20800(0x0 bytes),0x9e000(0x0 bytes)]
2020-06-01 11:11:23.959974 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) reshard checking spanning blobs 0x[0,ffffffff)
2020-06-01 11:11:23.959977 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) extent 0x1800~1000: 0x1800~1000 Blob(0x55973c3a7ea0 blob([0xd20000~10000] csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55973c6cfe30 sbid 0x0))
2020-06-01 11:11:23.959985 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) extent 0xf800~800: 0xf800~800 Blob(0x55973c3a7ea0 blob([0xd20000~10000] csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55973c6cfe30 sbid 0x0))

```
2020-06-01 11:11:23.959991 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70)  extent 0x10000~800: 0x10000~800 Blob(0x55973c6cff10 blob([!~10000,0xd30000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,800]) SharedBlob(0x55973c6cff80 sbid 0x0))
2020-06-01 11:11:23.959999 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70)  extent 0x20800~1000: 0x20800~1000 Blob(0x55973c728850 blob([!~20000,0xd80000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55973c7288c0 sbid 0x0))
2020-06-01 11:11:23.960006 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) reshard  shard 0x20800 to 0x9e000
2020-06-01 11:11:23.960010 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) reshard    adding spanning Blob(0x55973c728850 spanning 0 blob([!~20000,0xd80000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55973c7288c0 sbid 0x0))
2020-06-01 11:11:23.960073 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70)  extent 0x30000~30000: 0x30000~30000 Blob(0x55973c729030 blob([!~30000,0xd90000~10000,0xda0000~20000] csum crc32c/0x1000) use_tracker(0x6*0x10000 0x[0,0,0,10000,10000,10000]) SharedBlob(0x55973c7290a0 sbid 0x0))
2020-06-01 11:11:23.960083 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) reshard    adding spanning Blob(0x55973c729030 spanning 1 blob([!~30000,0xd90000~10000,0xda0000~20000] csum crc32c/0x1000) use_tracker(0x6*0x10000 0x[0,0,0,10000,10000,10000]) SharedBlob(0x55973c7290a0 sbid 0x0))
2020-06-01 11:11:23.960091 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70)  extent 0x8c800~1000: 0xc800~1000 Blob(0x55973c728d90 blob([0xdd0000~10000,0xdc0000~10000] csum+has_unused crc32c/0x1000 unused=0xbf) use_tracker(0x2*0x10000 0x[1000,2000]) SharedBlob(0x55973c6c7c70 sbid 0x0))
2020-06-01 11:11:23.960100 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) reshard    adding spanning Blob(0x55973c728d90 spanning 2 blob([0xdd0000~10000,0xdc0000~10000] csum+has_unused crc32c/0x1000 unused=0xbf) use_tracker(0x2*0x10000 0x[1000,2000]) SharedBlob(0x55973c6c7c70 sbid 0x0))
2020-06-01 11:11:23.960108 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70)  extent 0x9e000~2000: 0x1e000~2000 Blob(0x55973c728d90 spanning 2 blob([0xdd0000~10000,0xdc0000~10000] csum+has_unused crc32c/0x1000 unused=0xbf) use_tracker(0x2*0x10000 0x[1000,2000]) SharedBlob(0x55973c6c7c70 sbid 0x0))
2020-06-01 11:11:23.960116 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) reshard  shard 0x9e000 to 0xffffffff
2020-06-01 11:11:23.960118 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70)  extent 0xaf000~1000: 0x2f000~1000 Blob(0x55973c2564d0 blob([!~20000,0xf40000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55973cd35b90 sbid 0x0))
2020-06-01 11:11:23.960126 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) reshard    adding spanning Blob(0x55973c2564d0 spanning 3 blob([!~20000,0xf40000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55973cd35b90 sbid 0x0))
2020-06-01 11:11:23.960134 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70)  extent 0xb0000~1000: 0x30000~1000 Blob(0x55973cda72d0 blob([!~30000,0xf50000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x4*0x10000 0x[0,0,0,1000]) SharedBlob(0x55973cda65b0 sbid 0x0))
2020-06-01 11:11:23.960142 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) reshard    adding spanning Blob(0x55973cda72d0 spanning 4 blob([!~30000,0xf50000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x4*0x10000 0x[0,0,0,1000]) SharedBlob(0x55973cda65b0 sbid 0x0))

2020-06-01 11:11:23.960150 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) update #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# force
2020-06-01 11:11:23.960160 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) update  shard 0x0 is 241 bytes (was 0) from 3 extents  // inline编码ExtentMap, bluestore_blob_t
2020-06-01 11:11:23.960165 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) update  shard 0x20800 is 21 bytes (was 0) from 3 extents // 编码ExtentMap, struct_v(1B) + extent_num(1B) + {blobid(1B) + 2B + 2B + 1B} + {blobid(1B) + 2B + 2B + 2B} + {blobid(1B) + 2B + 2B + 1B}
2020-06-01 11:11:23.960170 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) update  shard 0x9e000 is 15 bytes (was 0) from 3 extents // 编码ExtentMap, struct_v(1B) + extent_num(1B) + {blobid(1B) + 2B + 1B + 1B} + {blobid(1B) + 1B + 2B + 1B} + {blobid(1B) + 0B + 2B + 0B}
2020-06-01 11:11:23.960190 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) onode #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# is 1721 (391 bytes onode + 1330 bytes spanning blobs + 0 bytes inline extents)
2020-06-01 11:11:23.960251 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _txc_finalize_kv txc 0x55973bf242c0 allocated 0x[f40000~20000] released 0x[]





// 分shard后第一次写
2020-06-01 11:11:24.767805 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _do_write #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# 0xb1000~f000 - have 0xb1000 (724992) bytes fadvise_flags 0x0
2020-06-01 11:11:24.767809 7f2ad2390700 30 _dump_onode 0x55973c434d80 #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# nid 17419 size 0xb1000 (724992) expected_object_size 1048576 expected_write_size 1048576 in 3 shards, 5 spanning blobs
2020-06-01 11:11:24.767813 7f2ad2390700 30 _dump_onode  attr _ len 293
2020-06-01 11:11:24.767815 7f2ad2390700 30 _dump_onode  attr snapset len 35
2020-06-01 11:11:24.767816 7f2ad2390700 30 _dump_extent_map  shard 0x0(0xf1 bytes) (loaded)
2020-06-01 11:11:24.767818 7f2ad2390700 30 _dump_extent_map  shard 0x20800(0x15 bytes) (loaded)
2020-06-01 11:11:24.767819 7f2ad2390700 30 _dump_extent_map  shard 0x9e000(0xf bytes) (loaded)
2020-06-01 11:11:24.767821 7f2ad2390700 30 _dump_extent_map   0x1800~1000: 0x1800~1000 Blob(0x55973c3a7ea0 blob([0xd20000~10000] csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55973c6cfe30 sbid 0x0))
2020-06-01 11:11:24.767829 7f2ad2390700 30 _dump_extent_map      csum: [0,37109b85,ddbb1646,0,0,0,0,0,0,0,0,0,0,0,0,79ae07b2]
2020-06-01 11:11:24.767832 7f2ad2390700 30 _dump_extent_map        0x1000~2000 buffer(0x55973bf74360 space 0x55973c6cfe48 0x1000~2000 writing nocache)
2020-06-01 11:11:24.767835 7f2ad2390700 30 _dump_extent_map        0xf000~1000 buffer(0x55973c2d9170 space 0x55973c6cfe48 0xf000~1000 writing nocache)
2020-06-01 11:11:24.767836 7f2ad2390700 30 _dump_extent_map   0xf800~800: 0xf800~800 Blob(0x55973c3a7ea0 blob([0xd20000~10000] csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55973c6cfe30 sbid 0x0))
2020-06-01 11:11:24.767840 7f2ad2390700 30 _dump_extent_map      csum: [0,37109b85,ddbb1646,0,0,0,0,0,0,0,0,0,0,0,0,79ae07b2]
2020-06-01 11:11:24.767842 7f2ad2390700 30 _dump_extent_map        0x1000~2000 buffer(0x55973bf74360 space 0x55973c6cfe48 0x1000~2000 writing nocache)
2020-06-01 11:11:24.767844 7f2ad2390700 30 _dump_extent_map        0xf000~1000 buffer(0x55973c2d9170 space 0x55973c6cfe48 0xf000~1000 writing nocache)
2020-06-01 11:11:24.767845 7f2ad2390700 30 _dump_extent_map  0x10000~800: 0x10000~800 Blob(0x55973c6cff10 blob([!~10000,0xd30000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,800]) SharedBlob(0x55973c6cff80 sbid 0x0))
2020-06-01 11:11:24.767851 7f2ad2390700 30 _dump_extent_map csum: [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,d93c2f8b,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
2020-06-01 11:11:24.767871 7f2ad2390700 30 _dump_extent_map        0x10000~1000 buffer(0x55973c2d9290 space 0x55973c6cff98 0x10000~1000 writing nocache)
2020-06-01 11:11:24.767873 7f2ad2390700 30 _dump_extent_map   0x20800~1000: 0x20800~1000 Blob(0x55973c728850 spanning 0 blob([!~20000,0xd80000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55973c7288c0 sbid 0x0))
2020-06-01 11:11:24.767880 7f2ad2390700 30 _dump_extent_map csum: [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,fdf59e9b,35342794,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
2020-06-01 11:11:24.767884 7f2ad2390700 30 _dump_extent_map        0x20000~2000 buffer(0x55973c2d9440 space 0x55973c7288d8 0x20000~2000 writing nocache)
2020-06-01 11:11:24.767886 7f2ad2390700 30 _dump_extent_map   0x30000~30000: 0x30000~30000 Blob(0x55973c729030 spanning 1 blob([!~30000,0xd90000~10000,0xda0000~20000] csum crc32c/0x1000) use_tracker(0x6*0x10000 0x[0,0,0,10000,10000,10000]) SharedBlob(0x55973c7290a0 sbid 0x0))
2020-06-01 11:11:24.767893 7f2ad2390700 30 _dump_extent_map csum: [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,77dd1e7,c72bffc9,7a55ba52,e47f96a5,57e45c2e,6679003,76c694
2020-06-01 11:11:24.767936 7f2ad2390700 30 _dump_extent_map        0x30000~10000 buffer(0x55973c2d94d0 space 0x55973c7290b8 0x30000~10000 clean)
2020-06-01 11:11:24.767939 7f2ad2390700 30 _dump_extent_map        0x40000~20000 buffer(0x55973c2d98c0 space 0x55973c7290b8 0x40000~20000 clean)
2020-06-01 11:11:24.767941 7f2ad2390700 30 _dump_extent_map   0x8c800~1000: 0xc800~1000 Blob(0x55973c728d90 spanning 2 blob([0xdd0000~10000,0xdc0000~10000] csum+has_unused crc32c/0x1000 unused=0xbf) use_tracker(0x2*0x10000 0x[1000,2000])
```

```
SharedBlob(0x55973c6c7c70 sbid 0x0))
2020-06-01 11:11:24.767948 7f2ad2390700 30 _dump_extent_map csum:
[0,0,0,0,0,0,0,0,0,0,0,0,9256d1e3,159f7524,0,0,d4e0c7e1,33c48c65,af9bea44,5b305dd7,2dfcd4be,1d9ad8bb,a598be3d,96738999,eeea7e5b,bfdeecc9,2982fe52,d6c441e8,
2020-06-01 11:11:24.767951 7f2ad2390700 30 _dump_extent_map 0xc000~2000 buffer(0x55973bf741b0 space 0x55973c6c7c88 0xc000~2000
writing nocache)
2020-06-01 11:11:24.767953 7f2ad2390700 30 _dump_extent_map 0x10000~2000 buffer(0x55973c2d9710 space 0x55973c6c7c88 0x10000~2000
writing nocache)
2020-06-01 11:11:24.767968 7f2ad2390700 30 _dump_extent_map 0x12000~e000 buffer(0x55973ced01b0 space 0x55973c6c7c88 0x12000~e000
writing nocache)
2020-06-01 11:11:24.767970 7f2ad2390700 30 _dump_extent_map 0x9e000~2000: 0x1e000~2000 Blob(0x55973c728d90 spanning 2
blob([0xdd0000~10000,0xdc0000~10000] csum+has_unused crc32c/0x1000 unused=0xbf) use_tracker(0x2*0x10000 0x[1000,2000])
SharedBlob(0x55973c6c7c70 sbid 0x0))
2020-06-01 11:11:24.767975 7f2ad2390700 30 _dump_extent_map csum:
[0,0,0,0,0,0,0,0,0,0,0,0,9256d1e3,159f7524,0,0,d4e0c7e1,33c48c65,af9bea44,5b305dd7,2dfcd4be,1d9ad8bb,a598be3d,96738999,eeea7e5b,bfdeecc9,2982fe52,d6c441e8,
2020-06-01 11:11:24.767979 7f2ad2390700 30 _dump_extent_map 0xc000~2000 buffer(0x55973bf741b0 space 0x55973c6c7c88 0xc000~2000
writing nocache)
2020-06-01 11:11:24.767980 7f2ad2390700 30 _dump_extent_map 0x10000~2000 buffer(0x55973c2d9710 space 0x55973c6c7c88 0x10000~2000
writing nocache)
2020-06-01 11:11:24.767981 7f2ad2390700 30 _dump_extent_map 0x12000~e000 buffer(0x55973ced01b0 space 0x55973c6c7c88 0x12000~e000
writing nocache)
2020-06-01 11:11:24.767983 7f2ad2390700 30 _dump_extent_map 0xaf000~1000: 0x2f000~1000 Blob(0x55973c2564d0 spanning 3
blob([!~20000,0xf40000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55973cd35b90
sbid 0x0))
2020-06-01 11:11:24.767988 7f2ad2390700 30 _dump_extent_map csum:
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,d928d45]
2020-06-01 11:11:24.767992 7f2ad2390700 30 _dump_extent_map 0x2f000~1000 buffer(0x55973bf74480 space 0x55973cd35ba8 0x2f000~1000
writing nocache)
2020-06-01 11:11:24.767994 7f2ad2390700 30 _dump_extent_map 0xb0000~1000: 0x30000~1000 Blob(0x55973cda72d0 spanning 4
blob([!~30000,0xf50000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x4*0x10000 0x[0,0,0,1000])
SharedBlob(0x55973cda65b0 sbid 0x0))
2020-06-01 11:11:24.767999 7f2ad2390700 30 _dump_extent_map csum:
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,6ca064cd,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
2020-06-01 11:11:24.768005 7f2ad2390700 30 _dump_extent_map 0x30000~1000 buffer(0x55973c2d9680 space 0x55973cda65c8 0x30000~1000
writing nocache)
2020-06-01 11:11:24.768008 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _choose_write_options prefer csum_order 12 target_blob_size
0x80000 compress=0 buffered=0
2020-06-01 11:11:24.768011 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) fault_range 0xb1000~f000
2020-06-01 11:11:24.768014 7f2ad2390700 10 bluestore(/sandstone-data/ceph-2) _do_write_small 0xb1000~f000
2020-06-01 11:11:24.768016 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) fault_range 0x31000~100000
2020-06-01 11:11:24.768019 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _do_write_small considering Blob(0x55973cda72d0 spanning 4
blob([!~30000,0xf50000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x4*0x10000 0x[0,0,0,1000])
SharedBlob(0x55973cda65b0 sbid 0x0)) bstart 0x80000
2020-06-01 11:11:24.768025 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _do_write_small  reading head 0x0 and tail 0x0
2020-06-01 11:11:24.768029 7f2ad2390700 20 bluestore.BufferSpace(0x55973cda65c8 in 0x55973bd54380) _discard 0x31000~f000
2020-06-01 11:11:24.768047 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _do_write_small  deferred write 0x31000~f000 of mutable
Blob(0x55973cda72d0 spanning 4 blob([!~30000,0xf50000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x4*0x10000
0x[0,0,0,1000]) SharedBlob(0x55973cda65b0 sbid 0x0)) at [0xf51000~f000]
2020-06-01 11:11:24.768054 7f2ad2390700 20 bluestore.blob(0x55973cda72d0) get_ref 0x31000~f000 Blob(0x55973cda72d0 spanning 4
blob([!~30000,0xf50000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x4*0x10000 0x[0,0,0,1000])
SharedBlob(0x55973cda65b0 sbid 0x0))
2020-06-01 11:11:24.768061 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _do_write_small  lex 0xb1000~f000: 0x31000~f000
Blob(0x55973cda72d0 spanning 4 blob([!~30000,0xf50000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x4*0x10000
0x[0,0,0,10000]) SharedBlob(0x55973cda65b0 sbid 0x0))
2020-06-01 11:11:24.768067 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write txc 0x55973bfbcdc0 0 blobs
2020-06-01 11:11:24.768069 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write buffered:0 compress:0 target_blob_size:524288
csum_order:12
2020-06-01 11:11:24.768072 7f2ad2390700 30 estimate gc range(hex): [b1000, c0000)
2020-06-01 11:11:24.768074 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _do_write extending size to 0xc0000
2020-06-01 11:11:24.768077 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) compress_extent_map 0xb1000~f000 next shard 0xffffffff
merging 0xb0000~1000: 0x30000~1000 Blob(0x55973cda72d0 spanning 4 blob([!~30000,0xf50000~10000] csum+has_unused crc32c/0x1000
unused=0xfff) use_tracker(0x4*0x10000 0x[0,0,0,10000]) SharedBlob(0x55973cda65b0 sbid 0x0)) and 0xb1000~f000: 0x31000~f000
Blob(0x55973cda72d0 spanning 4 blob([!~30000,0xf50000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x4*0x10000
0x[0,0,0,10000]) SharedBlob(0x55973cda65b0 sbid 0x0))
2020-06-01 11:11:24.768086 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) dirty_range 0xb1000~f000
2020-06-01 11:11:24.768088 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) dirty_range mark shard 0x9e000 dirty // 标记为dirty，需要更
新该key
2020-06-01 11:11:24.768090 7f2ad2390700 10 bluestore(/sandstone-data/ceph-2) _write 2.27_head
#2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# 0xb1000~f000 = 0
2020-06-01 11:11:24.768095 7f2ad2390700 30 bluestore.OnodeSpace(0x55973df04f48 in 0x55973bd54380) lookup
2020-06-01 11:11:24.768097 7f2ad2390700 30 bluestore.OnodeSpace(0x55973df04f48 in 0x55973bd54380) lookup #2:e4000000::::head# hit
0x55973c4698c0
2020-06-01 11:11:24.768101 7f2ad2390700 15 bluestore(/sandstone-data/ceph-2) _omap_setkeys 2.27_head #2:e4000000::::head#
2020-06-01 11:11:24.768105 7f2ad2390700 30 bluestore(/sandstone-data/ceph-2)
_omap_setkeys  0x000000000000045f'.0000000117.00000000000000000056' <- 0000000117.00000000000000000056
2020-06-01 11:11:24.768116 7f2ad2390700 30 bluestore(/sandstone-data/ceph-2) _omap_setkeys  0x000000000000045f'._info' <- _info
2020-06-01 11:11:24.768121 7f2ad2390700 10 bluestore(/sandstone-data/ceph-2) _omap_setkeys 2.27_head #2:e4000000::::head# = 0
2020-06-01 11:11:24.768125 7f2ad2390700 10 bluestore(/sandstone-data/ceph-2) _txc_calc_cost 0x55973bfbcdc0 cost 733474 (1 ios * 670000 +
63474 bytes)
2020-06-01 11:11:24.768127 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _txc_write_nodes txc 0x55973bfbcdc0 onodes 0x55973c434d80
shared_blobs
2020-06-01 11:11:24.768130 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) update
#2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head#
2020-06-01 11:11:24.768136 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) update  shard 0x9e000 is 16 bytes (was 15) from 3 extents
// 更新shard 0x9e000
2020-06-01 11:11:24.768147 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) onode
#2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# is 1722 (391 bytes onode + 1331 bytes spanning blobs + 0 bytes inline extents)


//
2020-06-01 11:11:24.872779 7f2ad2390700 15 bluestore(/sandstone-data/ceph-2) _truncate 2.27_head
#2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# 0xb0000
2020-06-01 11:11:24.872784 7f2ad2390700 15 bluestore(/sandstone-data/ceph-2) _do_truncate 2.27_head
#2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# 0xb0000
2020-06-01 11:11:24.872788 7f2ad2390700 30 _dump_onode 0x55973c434d80 #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# nid 17419
size 0xc0000 (786432) expected_object_size 1048576 expected_write_size 1048576 in 3 shards, 5 spanning blobs
2020-06-01 11:11:24.872794 7f2ad2390700 30 _dump_onode  attr _ len 293
2020-06-01 11:11:24.872796 7f2ad2390700 30 _dump_onode  attr snapset len 35
2020-06-01 11:11:24.872797 7f2ad2390700 30 _dump_extent_map  shard 0x0(0xf1 bytes) (loaded)
2020-06-01 11:11:24.872799 7f2ad2390700 30 _dump_extent_map  shard 0x20800(0x15 bytes) (loaded)
2020-06-01 11:11:24.872800 7f2ad2390700 30 _dump_extent_map  shard 0x9e000(0x10 bytes) (loaded)
2020-06-01 11:11:24.872801 7f2ad2390700 30 _dump_extent_map 0x1800~1000: 0x1800~1000 Blob(0x55973c3a7ea0 blob([0xd20000~10000]
csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55973c6cfe30 sbid 0x0))
2020-06-01 11:11:24.872811 7f2ad2390700 30 _dump_extent_map  csum: [0,37109b85,ddbb1646,0,0,0,0,0,0,0,0,0,0,0,0,79ae07b2]
2020-06-01 11:11:24.872814 7f2ad2390700 30 _dump_extent_map 0x1000~2000 buffer(0x55973bf74360 space 0x55973c6cfe48 0x1000~2000
writing nocache)
```

```
2020-06-01 11:11:24.872816 7f2ad2390700 30 _dump_extent_map          0xf000~1000 buffer(0x55973c2d9170 space 0x55973c6cfe48 0xf000~1000
writing nocache)
2020-06-01  11:11:24.872818  7f2ad2390700  30  _dump_extent_map     0xf800~800:  0xf800~800  Blob(0x55973c3a7ea0  blob([0xd20000~10000]
csum+has_unused crc32c/0x1000 unused=0x7ff9) use_tracker(0x10000 0x1800) SharedBlob(0x55973c6cfe30 sbid 0x0))
2020-06-01 11:11:24.872822 7f2ad2390700 30 _dump_extent_map      csum: [0,37109b85,ddbb1646,0,0,0,0,0,0,0,0,0,0,0,0,0,79ae07b2]
2020-06-01 11:11:24.872824 7f2ad2390700 30 _dump_extent_map          0x1000~2000 buffer(0x55973bf74360 space 0x55973c6cfe48 0x1000~2000
writing nocache)
2020-06-01 11:11:24.872825 7f2ad2390700 30 _dump_extent_map          0xf000~1000 buffer(0x55973c2d9170 space 0x55973c6cfe48 0xf000~1000
writing nocache)
2020-06-01 11:11:24.872827 7f2ad2390700 30 _dump_extent_map  0x10000~800: 0x10000~800 Blob(0x55973c6cff10 blob([!~10000,0xd30000~10000]
csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,800]) SharedBlob(0x55973c6cff80 sbid 0x0))
2020-06-01        11:11:24.872834        7f2ad2390700        30        _dump_extent_map                                                  csum:
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,d93c2f8b,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
2020-06-01 11:11:24.872838 7f2ad2390700 30 _dump_extent_map         0x10000~1000 buffer(0x55973c2d9290 space 0x55973c6cff98 0x10000~1000
writing nocache)
2020-06-01  11:11:24.872841  7f2ad2390700  30  _dump_extent_map     0x20800~1000:  0x20800~1000  Blob(0x55973c728850  spanning  0
blob([!~20000,0xd80000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55973c7288c0
sbid 0x0))
2020-06-01        11:11:24.872851        7f2ad2390700        30        _dump_extent_map                                                  csum:
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,fdf59e9b,35342794,0,0,0,0,0,0,0,0,0,0,0,0]
2020-06-01 11:11:24.872879 7f2ad2390700 30 _dump_extent_map         0x20000~2000 buffer(0x55973c2d9440 space 0x55973c7288d8 0x20000~2000
writing nocache)
2020-06-01  11:11:24.872882  7f2ad2390700  30  _dump_extent_map     0x30000~30000:  0x30000~30000  Blob(0x55973c729030  spanning  1
blob([!~30000,0xd90000~10000,0xda0000~20000]      csum      crc32c/0x1000)      use_tracker(0x6*0x10000      0x[0,0,0,10000,10000,10000])
SharedBlob(0x55973c7290a0 sbid 0x0))
2020-06-01        11:11:24.872891        7f2ad2390700        30        _dump_extent_map                                                  csum:
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,77dd1e7,c72bffc9,7a55ba52,e47f96a5,57e45c2e,6679003,76c694
2020-06-01 11:11:24.872901 7f2ad2390700 30 _dump_extent_map        0x30000~10000 buffer(0x55973c2d94d0 space 0x55973c7290b8 0x30000~10000
clean)
2020-06-01 11:11:24.872903 7f2ad2390700 30 _dump_extent_map        0x40000~20000 buffer(0x55973c2d98c0 space 0x55973c7290b8 0x40000~20000
clean)
2020-06-01  11:11:24.872904  7f2ad2390700  30  _dump_extent_map     0x8c800~1000:  0xc800~1000  Blob(0x55973c728d90  spanning  2
blob([0xdd0000~10000,0xdc0000~10000]     csum+has_unused     crc32c/0x1000     unused=0xbf)     use_tracker(0x2*0x10000     0x[1000,2000])
SharedBlob(0x55973c6c7c70 sbid 0x0))
2020-06-01        11:11:24.872909        7f2ad2390700        30        _dump_extent_map                                                  csum:
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,9256d1e3,159f7524,0,0,d4e0c7e1,33c48c65,af9bea44,5b305dd7,2dfcd4be,1d9ad8bb,a598be3d,96738999,eeea7e5b,bfdeecc9,2982fe52,d6c441e8,
2020-06-01 11:11:24.872913 7f2ad2390700 30 _dump_extent_map          0xc000~2000 buffer(0x55973bf741b0 space 0x55973c6c7c88 0xc000~2000
writing nocache)
2020-06-01 11:11:24.872915 7f2ad2390700 30 _dump_extent_map         0x10000~2000 buffer(0x55973c2d9710 space 0x55973c6c7c88 0x10000~2000
writing nocache)
2020-06-01 11:11:24.872916 7f2ad2390700 30 _dump_extent_map         0x12000~e000 buffer(0x55973ced01b0 space 0x55973c6c7c88 0x12000~e000
writing nocache)
2020-06-01  11:11:24.872918  7f2ad2390700  30  _dump_extent_map     0x9e000~2000:  0x1e000~2000  Blob(0x55973c728d90  spanning  2
blob([0xdd0000~10000,0xdc0000~10000]     csum+has_unused     crc32c/0x1000     unused=0xbf)     use_tracker(0x2*0x10000     0x[1000,2000])
SharedBlob(0x55973c6c7c70 sbid 0x0))
2020-06-01        11:11:24.872922        7f2ad2390700        30        _dump_extent_map                                                  csum:
[0,0,0,0,0,0,0,0,0,0,0,0,0,9256d1e3,159f7524,0,0,d4e0c7e1,33c48c65,af9bea44,5b305dd7,2dfcd4be,1d9ad8bb,a598be3d,96738999,eeea7e5b,bfdeecc9,2982fe52,d6c441e8,
2020-06-01 11:11:24.872926 7f2ad2390700 30 _dump_extent_map          0xc000~2000 buffer(0x55973bf741b0 space 0x55973c6c7c88 0xc000~2000
writing nocache)
2020-06-01 11:11:24.872928 7f2ad2390700 30 _dump_extent_map         0x10000~2000 buffer(0x55973c2d9710 space 0x55973c6c7c88 0x10000~2000
writing nocache)
2020-06-01 11:11:24.872929 7f2ad2390700 30 _dump_extent_map         0x12000~e000 buffer(0x55973ced01b0 space 0x55973c6c7c88 0x12000~e000
writing nocache)
2020-06-01  11:11:24.872930  7f2ad2390700  30  _dump_extent_map     0xaf000~1000:  0x2f000~1000  Blob(0x55973c2564d0  spanning  3
blob([!~20000,0xf40000~10000] csum+has_unused crc32c/0x1000 unused=0x7fff) use_tracker(0x3*0x10000 0x[0,0,1000]) SharedBlob(0x55973cd35b90
sbid 0x0))
2020-06-01        11:11:24.872935        7f2ad2390700        30        _dump_extent_map                                                  csum:
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,d928d45]
2020-06-01 11:11:24.872939 7f2ad2390700 30 _dump_extent_map         0x2f000~1000 buffer(0x55973bf74480 space 0x55973cd35ba8 0x2f000~1000
writing nocache)
2020-06-01  11:11:24.872941  7f2ad2390700  30  _dump_extent_map     0xb0000~10000:  0x30000~10000  Blob(0x55973cda72d0  spanning  4
blob([!~30000,0xf50000~10000]      csum+has_unused      crc32c/0x1000      unused=0xfff)      use_tracker(0x4*0x10000      0x[0,0,0,10000])
SharedBlob(0x55973cda65b0 sbid 0x0))
2020-06-01        11:11:24.872946        7f2ad2390700        30        _dump_extent_map                                                  csum:
[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,6ca064cd,33ccc7fa,c3b50c70,245a7826,33ffd5a0,17fc86bb,148b
2020-06-01 11:11:24.872952 7f2ad2390700 30 _dump_extent_map         0x30000~1000 buffer(0x55973c2d9680 space 0x55973cda65c8 0x30000~1000
writing nocache)
2020-06-01 11:11:24.872953 7f2ad2390700 30 _dump_extent_map         0x31000~f000 buffer(0x55973bf753b0 space 0x55973cda65c8 0x31000~f000
writing nocache)
2020-06-01 11:11:24.872955 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) fault_range 0xb0000~10000
2020-06-01  11:11:24.872961  7f2ad2390700  20  bluestore.blob(0x55973cda72d0)  put_ref  0x30000~10000  Blob(0x55973cda72d0  spanning  4
blob([!~30000,0xf50000~10000]      csum+has_unused      crc32c/0x1000      unused=0xfff)      use_tracker(0x4*0x10000      0x[0,0,0,10000])
SharedBlob(0x55973cda65b0 sbid 0x0))
2020-06-01 11:11:24.872982 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) dirty_range 0xb0000~10000
2020-06-01 11:11:24.872984 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) dirty_range mark shard 0x9e000 dirty
2020-06-01  11:11:24.872986  7f2ad2390700  20  bluestore(/sandstone-data/ceph-2)  _wctx_finish  lex_old  0xb0000~10000:  0x30000~10000
Blob(0x55973cda72d0  spanning  4  blob([!~40000]  csum+has_unused  crc32c/0x1000  unused=0xfff)  use_tracker(0x4*0x10000  0x[0,0,0,0])
SharedBlob(0x55973cda65b0 sbid 0x0))
2020-06-01 11:11:24.872992 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _wctx_finish  blob release [0xf50000~10000]
2020-06-01 11:11:24.872995 7f2ad2390700 20 bluestore.blob(0x55973cda72d0) discard_unallocated 0x0~40000
2020-06-01 11:11:24.872997 7f2ad2390700 20 bluestore.BufferSpace(0x55973cda65c8 in 0x55973bd54380) _discard 0x0~40000
2020-06-01 11:11:24.873001 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _wctx_finish  release 0xf50000~10000
2020-06-01  11:11:24.873003  7f2ad2390700  20  bluestore(/sandstone-data/ceph-2)  _wctx_finish  spanning_blob_map  removing  empty
Blob(0x55973cda72d0  spanning  4  blob([!~40000]  csum+has_unused  crc32c/0x1000  unused=0xfff)  use_tracker(0x4*0x10000  0x[0,0,0,0])
SharedBlob(0x55973cda65b0 sbid 0x0))
2020-06-01        11:11:24.873014        7f2ad2390700        10        bluestore(/sandstone-data/ceph-2)        _truncate        2.27_head
#2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# 0xb0000 = 0
2020-06-01        11:11:24.873024        7f2ad2390700        15        bluestore(/sandstone-data/ceph-2)        _setattrs        2.27_head
#2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# 2 keys
2020-06-01        11:11:24.873031        7f2ad2390700        10        bluestore(/sandstone-data/ceph-2)        _setattrs        2.27_head
#2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# 2 keys = 0
2020-06-01 11:11:24.873035 7f2ad2390700 30 bluestore.OnodeSpace(0x55973df04f48 in 0x55973bd54380) lookup
2020-06-01  11:11:24.873037  7f2ad2390700  30  bluestore.OnodeSpace(0x55973df04f48  in  0x55973bd54380)  lookup  #2:e4000000:::head#  hit
0x55973c4698c0
2020-06-01 11:11:24.873042 7f2ad2390700 15 bluestore(/sandstone-data/ceph-2) _omap_setkeys 2.27_head #2:e4000000:::head#
2020-06-01                11:11:24.873045                7f2ad2390700                30                bluestore(/sandstone-data/ceph-2)
_omap_setkeys  0x000000000000045f'.0000000117.00000000000000000057' <- 0000000117.0000000000000000057
2020-06-01 11:11:24.873056 7f2ad2390700 30 bluestore(/sandstone-data/ceph-2) _omap_setkeys  0x000000000000045f'._info' <- _info
2020-06-01 11:11:24.873061 7f2ad2390700 10 bluestore(/sandstone-data/ceph-2) _omap_setkeys 2.27_head #2:e4000000:::head# = 0
2020-06-01 11:11:24.873065 7f2ad2390700 10 bluestore(/sandstone-data/ceph-2) _txc_calc_cost 0x55973cad3600 cost 671958 (1 ios * 670000 +
1958 bytes)
2020-06-01 11:11:24.873068 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _txc_write_nodes txc 0x55973cad3600 onodes 0x55973c434d80
shared_blobs
2020-06-01           11:11:24.873070           7f2ad2390700           20           bluestore.extentmap(0x55973c434e70)           update
#2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head#
```

2020-06-01 11:11:24.873076 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) update shard 0x9e000 is 12 bytes (was 16) from 2 extents // 更新shard 0x9e000
2020-06-01 11:11:24.873087 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) onode #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# is 1430 (391 bytes onode + 1039 bytes spanning blobs + 0 bytes inline extents)

2020-06-01 14:45:56.355940 7f2ad2390700 10 bluestore(/sandstone-data/ceph-2) _do_write_small 0x1c000~2000
2020-06-01 14:45:56.355942 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) fault_range 0x0~9c000
2020-06-01 14:45:56.355945 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _do_write_small considering Blob(0x55973c6cff10 blob([!~10000,0xd30000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,800]) SharedBlob(0x55973c6cff80 sbid 0x0)) bstart 0x0
2020-06-01 14:45:56.355950 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _do_write_small reading head 0x0 and tail 0x0
2020-06-01 14:45:56.355953 7f2ad2390700 20 bluestore.BufferSpace(0x55973c6cff98 in 0x55973bd54380) _discard 0x1c000~2000
2020-06-01 14:45:56.355962 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _do_write_small deferred write 0x1c000~2000 of mutable Blob(0x55973c6cff10 blob([!~10000,0xd30000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,800]) SharedBlob(0x55973c6cff80 sbid 0x0)) at [0xd3c000~2000]
2020-06-01 14:45:56.355967 7f2ad2390700 20 bluestore.blob(0x55973c6cff10) get_ref 0x1c000~2000 Blob(0x55973c6cff10 blob([!~10000,0xd30000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,800]) SharedBlob(0x55973c6cff80 sbid 0x0))
2020-06-01 14:45:56.355973 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _do_write_small lex 0x1c000~2000: 0x1c000~2000 Blob(0x55973c6cff10 blob([!~10000,0xd30000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,2800]) SharedBlob(0x55973c6cff80 sbid 0x0))
2020-06-01 14:45:56.355978 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write txc 0x55973cb97080 0 blobs
2020-06-01 14:45:56.355979 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _do_alloc_write buffered:0 compress:0 target_blob_size:524288 csum_order:12
2020-06-01 14:45:56.355982 7f2ad2390700 30 estimate gc range(hex): [1c000, 1e000)
2020-06-01 14:45:56.355985 7f2ad2390700 30 bluestore.extentmap(0x55973c434e70) dirty_range 0x1c000~2000
2020-06-01 14:45:56.355986 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) dirty_range mark shard 0x0 dirty //
2020-06-01 14:45:56.355988 7f2ad2390700 10 bluestore(/sandstone-data/ceph-2) _write 2.27_head #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# 0x1c000~2000 = 0
2020-06-01 14:45:56.355992 7f2ad2390700 30 bluestore.OnodeSpace(0x55973df04f48 in 0x55973bd54380) lookup
2020-06-01 14:45:56.355994 7f2ad2390700 30 bluestore.OnodeSpace(0x55973df04f48 in 0x55973bd54380) lookup #2:e4000000:::head# hit 0x55973c4698c0
2020-06-01 14:45:56.355998 7f2ad2390700 15 bluestore(/sandstone-data/ceph-2) _omap_setkeys 2.27_head #2:e4000000:::head#
2020-06-01 14:45:56.356001 7f2ad2390700 30 bluestore(/sandstone-data/ceph-2) _omap_setkeys 0x000000000000045f'.0000000125.00000000000000000058' <- 0000000125.00000000000000000058
2020-06-01 14:45:56.356011 7f2ad2390700 30 bluestore(/sandstone-data/ceph-2) _omap_setkeys 0x000000000000045f'._epoch' <- _epoch
2020-06-01 14:45:56.356015 7f2ad2390700 30 bluestore(/sandstone-data/ceph-2) _omap_setkeys 0x000000000000045f'._info' <- _info
2020-06-01 14:45:56.356017 7f2ad2390700 10 bluestore(/sandstone-data/ceph-2) _omap_setkeys 2.27_head #2:e4000000:::head# = 0
2020-06-01 14:45:56.356021 7f2ad2390700 10 bluestore(/sandstone-data/ceph-2) _txc_calc_cost 0x55973cb97080 cost 680244 (1 ios * 670000 + 10244 bytes)
2020-06-01 14:45:56.356023 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) _txc_write_nodes txc 0x55973cb97080 onodes 0x55973c434d80 shared_blobs
2020-06-01 14:45:56.356025 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) update #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head#
2020-06-01 14:45:56.356034 7f2ad2390700 20 bluestore.extentmap(0x55973c434e70) update shard 0x0 is 246 bytes (was 241) from 4 extents //
2020-06-01 14:45:56.356045 7f2ad2390700 20 bluestore(/sandstone-data/ceph-2) onode #2:e42e9d2d:::rbd_data.10557d0a41f0.0000000000000000:head# is 1430 (391 bytes onode + 1039 bytes spanning blobs + 0 bytes inline extents)

```
//////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
2020-06-19 16:04:26.095350 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) update
#2:ddb85c4a:::rbd_data.4e60666b8b4567.0000000000000000:head#
2020-06-19 16:04:26.095357 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) update  inline shard 1218 bytes from 5 extents
2020-06-19 16:04:26.095368 7f175f7ec700 10 bluestore.extentmap(0x563a819790b0) reshard 0x[0,ffffffff) of 0 shards on
#2:ddb85c4a:::rbd_data.4e60666b8b4567.0000000000000000:head#
2020-06-19 16:04:26.095381 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) fault_range 0x0~ffffffff
2020-06-19 16:04:26.095383 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard  extent_avg 243, target 500, slop 100
2020-06-19 16:04:26.095385 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0)     extent  0x0~f000:  0x0~f000  Blob(0x563a81a0c8c0
blob([0x1860000~10000] csum+has_unused crc32c/0x1000 unused=0x8000) use_tracker(0x10000 0xf000) SharedBlob(0x563a81a0c930 sbid 0x0))
2020-06-19 16:04:26.095390 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0)   extent 0x10000~f000: 0x10000~f000 Blob(0x563a81a0d490
blob([!~10000,0x1870000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,f000]) SharedBlob(0x563a81a0d500
sbid 0x0))
2020-06-19 16:04:26.095401 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0)   extent 0x20000~f000: 0x20000~f000 Blob(0x563a81aca070
blob([!~20000,0x1880000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,f000]) SharedBlob(0x563a81a0cd90
sbid 0x0))
2020-06-19 16:04:26.095406 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard  new shard 0x0
2020-06-19 16:04:26.095407 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard  new shard 0x20000
2020-06-19 16:04:26.095412 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0)   extent 0x40000~f000: 0x40000~f000 Blob(0x563a81aca4d0
blob([!~40000,0x1890000~10000]     csum+has_unused     crc32c/0x1000     unused=0xfff)     use_tracker(0x5*0x10000     0x[0,0,0,0,f000])
SharedBlob(0x563a81aca2a0 sbid 0x0))
2020-06-19 16:04:26.095416 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0)   extent 0x50000~f000: 0x50000~f000 Blob(0x563a81acaf50
blob([!~50000,0x18a0000~10000]     csum+has_unused     crc32c/0x1000     unused=0x1fff)     use_tracker(0x6*0x10000     0x[0,0,0,0,0,f000])
SharedBlob(0x563a81acaee0 sbid 0x0))
2020-06-19 16:04:26.095420 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard  new shard 0x50000
2020-06-19 16:04:26.095421 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard  new [0x0(0x0 bytes),0x20000(0x0 bytes),0x50000(0x0
bytes)]
2020-06-19 16:04:26.095424 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard  old []
2020-06-19 16:04:26.095425 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard  fin [0x0(0x0 bytes),0x20000(0x0 bytes),0x50000(0x0
bytes)]
2020-06-19 16:04:26.095428 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard checking spanning blobs 0x[0,ffffffff)
2020-06-19 16:04:26.095429 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0)     extent  0x0~f000:  0x0~f000  Blob(0x563a81a0c8c0
blob([0x1860000~10000] csum+has_unused crc32c/0x1000 unused=0x8000) use_tracker(0x10000 0xf000) SharedBlob(0x563a81a0c930 sbid 0x0))
2020-06-19 16:04:26.095432 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0)   extent 0x10000~f000: 0x10000~f000 Blob(0x563a81a0d490
blob([!~10000,0x1870000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,f000]) SharedBlob(0x563a81a0d500
sbid 0x0))
2020-06-19 16:04:26.095436 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0)   extent 0x20000~f000: 0x20000~f000 Blob(0x563a81aca070
blob([!~20000,0x1880000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,f000]) SharedBlob(0x563a81a0cd90
sbid 0x0))
2020-06-19 16:04:26.095439 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) reshard  shard 0x20000 to 0x50000
2020-06-19 16:04:26.095440 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard cann't split Blob(0x563a81aca070
blob([!~20000,0x1880000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,f000]) SharedBlob(0x563a81a0cd90
sbid 0x0))
2020-06-19 16:04:26.095445 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard     adding spanning Blob(0x563a81aca070 spanning 0
blob([!~20000,0x1880000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,f000]) SharedBlob(0x563a81a0cd90
sbid 0x0))
2020-06-19 16:04:26.095463 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0)   extent 0x40000~f000: 0x40000~f000 Blob(0x563a81aca4d0
blob([!~40000,0x1890000~10000]     csum+has_unused     crc32c/0x1000     unused=0xfff)     use_tracker(0x5*0x10000     0x[0,0,0,0,f000])
SharedBlob(0x563a81aca2a0 sbid 0x0))
2020-06-19 16:04:26.095467 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard cann't split Blob(0x563a81aca4d0
blob([!~40000,0x1890000~10000]     csum+has_unused     crc32c/0x1000     unused=0xfff)     use_tracker(0x5*0x10000     0x[0,0,0,0,f000])
SharedBlob(0x563a81aca2a0 sbid 0x0))
2020-06-19 16:04:26.095470 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard     adding spanning Blob(0x563a81aca4d0 spanning 1
blob([!~40000,0x1890000~10000]     csum+has_unused     crc32c/0x1000     unused=0xfff)     use_tracker(0x5*0x10000     0x[0,0,0,0,f000])
SharedBlob(0x563a81aca2a0 sbid 0x0))
2020-06-19 16:04:26.095473 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0)   extent 0x50000~f000: 0x50000~f000 Blob(0x563a81acaf50
blob([!~50000,0x18a0000~10000]     csum+has_unused     crc32c/0x1000     unused=0x1fff)     use_tracker(0x6*0x10000     0x[0,0,0,0,0,f000])
SharedBlob(0x563a81acaee0 sbid 0x0))
2020-06-19 16:04:26.095476 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) reshard  shard 0x50000 to 0xffffffff
2020-06-19 16:04:26.095478 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard cann't split Blob(0x563a81acaf50
blob([!~50000,0x18a0000~10000]     csum+has_unused     crc32c/0x1000     unused=0x1fff)     use_tracker(0x6*0x10000     0x[0,0,0,0,0,f000])
SharedBlob(0x563a81acaee0 sbid 0x0))
2020-06-19 16:04:26.095482 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard     adding spanning Blob(0x563a81acaf50 spanning 2
blob([!~50000,0x18a0000~10000]     csum+has_unused     crc32c/0x1000     unused=0x1fff)     use_tracker(0x6*0x10000     0x[0,0,0,0,0,f000])
SharedBlob(0x563a81acaee0 sbid 0x0))
2020-06-19 16:04:26.095496 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) update
#2:ddb85c4a:::rbd_data.4e60666b8b4567.0000000000000000:head# force
2020-06-19 16:04:26.095506 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) update  shard 0x0 is 235 bytes (was 0) from 2 extents
2020-06-19 16:04:26.095510 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) update  shard 0x20000 is 12 bytes (was 0) from 2 extents
2020-06-19 16:04:26.095512 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) update  shard 0x50000 is 8 bytes (was 0) from 1 extents
2020-06-19 16:04:26.095523 7f175f7ec700 20 bluestore(/sandstone-data/ceph-4) onode
#2:ddb85c4a:::rbd_data.4e60666b8b4567.0000000000000000:head# is 1400 (392 bytes onode + 1008 bytes spanning blobs + 0 bytes inline
extents)



2020-06-19 16:04:26.659448 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) fault_range 0x60000~10000
2020-06-19 16:04:26.659450 7f175f7ec700 10 bluestore(/sandstone-data/ceph-4) _do_write_big 0x60000~10000 target_blob_size 0x80000 compress
0
2020-06-19 16:04:26.659463 7f175f7ec700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write txc 0x563a81848b00 1 blobs
2020-06-19 16:04:26.659467 7f175f7ec700 10 fbmap_alloc 0x563a8096f700 allocate 0x10000/10000,10000,0
2020-06-19 16:04:26.659471 7f175f7ec700 10 fbmap_alloc 0x563a8096f700 allocate extent: 0x18b0000~10000/10000,10000,0
2020-06-19 16:04:26.659473 7f175f7ec700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write prealloc [0x18b0000~10000]
2020-06-19 16:04:26.659475 7f175f7ec700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write forcing blob_offset to 0x60000
2020-06-19 16:04:26.659476 7f175f7ec700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write initialize csum setting for new blob
Blob(0x563a81acb810 blob([]) use_tracker(0x0 0x0) SharedBlob(0x563a81acb880 sbid 0x0)) csum_type crc32c csum_order 12 csum_length 0x70000
2020-06-19 16:04:26.659481 7f175f7ec700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write blob Blob(0x563a81acb810
blob([!~60000,0x18b0000~10000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x563a81acb880 sbid 0x0))
2020-06-19 16:04:26.659493 7f175f7ec700 20 bluestore.blob(0x563a81acb810) get_ref 0x60000~10000 Blob(0x563a81acb810
blob([!~60000,0x18b0000~10000] csum crc32c/0x1000) use_tracker(0x0 0x0) SharedBlob(0x563a81acb880 sbid 0x0))
2020-06-19 16:04:26.659501 7f175f7ec700 20 bluestore.blob(0x563a81acb810) get_ref init 0x70000, 10000
2020-06-19 16:04:26.659504 7f175f7ec700 20 bluestore(/sandstone-data/ceph-4) _do_alloc_write lex 0x60000~10000: 0x60000~10000
Blob(0x563a81acb810 blob([!~60000,0x18b0000~10000] csum crc32c/0x1000) use_tracker(0x7*0x10000 0x[0,0,0,0,0,0,10000])
SharedBlob(0x563a81acb880 sbid 0x0))
2020-06-19 16:04:26.659509 7f175f7ec700 20 bluestore.BufferSpace(0x563a81acb898 in 0x563a80d1e380) _discard 0x60000~10000
2020-06-19 16:04:26.659517 7f175f7ec700 30 estimate gc range(hex): [60000, 70000)
2020-06-19 16:04:26.659522 7f175f7ec700 20 bluestore(/sandstone-data/ceph-4) _do_write extending size to 0x70000
2020-06-19 16:04:26.659524 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) dirty_range 0x60000~10000
2020-06-19 16:04:26.659525 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) dirty_range mark shard 0x50000 dirty
```

```
2020-06-19 16:04:26.659527 7f175f7ec700 10 bluestore(/sandstone-data/ceph-4) _write 2.3b_head #2:ddb85c4a:::rbd_data.4e60666b8b4567.0000000000000000:head# 0x60000~10000 = 0
2020-06-19 16:04:26.659531 7f175f7ec700 30 bluestore.OnodeSpace(0x563a82f02bc8 in 0x563a80d1e380) lookup
2020-06-19 16:04:26.659533 7f175f7ec700 30 bluestore.OnodeSpace(0x563a82f02bc8 in 0x563a80d1e380) lookup #2:dc000000::::head# hit 0x563a81979200
2020-06-19 16:04:26.659537 7f175f7ec700 15 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.3b_head #2:dc000000::::head#
2020-06-19 16:04:26.659540 7f175f7ec700 30 bluestore(/sandstone-data/ceph-4) _omap_setkeys 0x000000000000046b'.0000000245.00000000000000000212' <- 0000000245.00000000000000000212
2020-06-19 16:04:26.659549 7f175f7ec700 30 bluestore(/sandstone-data/ceph-4) _omap_setkeys 0x000000000000046b'._info' <- _info
2020-06-19 16:04:26.659552 7f175f7ec700 10 bluestore(/sandstone-data/ceph-4) _omap_setkeys 2.3b_head #2:dc000000::::head# = 0
2020-06-19 16:04:26.659555 7f175f7ec700 10 bluestore(/sandstone-data/ceph-4) _txc_calc_cost 0x563a81848b00 cost 1407576 (2 ios * 670000 + 67576 bytes)
2020-06-19 16:04:26.659557 7f175f7ec700 20 bluestore(/sandstone-data/ceph-4) _txc_write_nodes txc 0x563a81848b00 onodes 0x563a81978fc0 shared_blobs
2020-06-19 16:04:26.659559 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) update #2:ddb85c4a:::rbd_data.4e60666b8b4567.0000000000000000:head#
2020-06-19 16:04:26.659562 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) encode_some 0x50000~fffaffff hit new spanning blob 0x60000~10000: 0x60000~10000 Blob(0x563a81acb810 blob([!~60000,0x18b0000~10000] csum crc32c/0x1000) use_tracker(0x7*0x10000 0x[0,0,0,0,0,0,10000]) SharedBlob(0x563a81acb880 sbid 0x0))  ///////////// 本次io生成一个新的extent 0x60000~10000: 0x60000~10000
2020-06-19 16:04:26.659568 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) update  shard 0x50000 is 0 bytes (was 8) from 1 extents
2020-06-19 16:04:26.659570 7f175f7ec700 10 bluestore.extentmap(0x563a819790b0) reshard 0x[0,70000) of 3 shards on #2:ddb85c4a:::rbd_data.4e60666b8b4567.0000000000000000:head#
2020-06-19 16:04:26.659573 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard   spanning blob 0 Blob(0x563a81aca070 spanning 0 blob([!~20000,0x1880000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,f000]) SharedBlob(0x563a81a0cd90 sbid 0x0))
2020-06-19 16:04:26.659578 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard   spanning blob 1 Blob(0x563a81aca4d0 spanning 1 blob([!~40000,0x1890000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x5*0x10000 0x[0,0,0,0,f000]) SharedBlob(0x563a81aca2a0 sbid 0x0))
2020-06-19 16:04:26.659582 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard   spanning blob 2 Blob(0x563a81acaf50 spanning 2 blob([!~50000,0x18a0000~10000] csum+has_unused crc32c/0x1000 unused=0x1fff) use_tracker(0x6*0x10000 0x[0,0,0,0,0,f000]) SharedBlob(0x563a81acaee0 sbid 0x0))
2020-06-19 16:04:26.659585 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard   shards [0,3) over 0x[0,ffffffff)
2020-06-19 16:04:26.659587 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) fault_range 0x0~ffffffff
2020-06-19 16:04:26.659593 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard  extent_avg 51, target 500, slop 100
2020-06-19 16:04:26.659595 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) extent 0x0~f000: 0x0~f000 Blob(0x563a81a0c8c0 blob([0x1860000~10000] csum+has_unused crc32c/0x1000 unused=0x8000) use_tracker(0x10000 0xf000) SharedBlob(0x563a81a0c930 sbid 0x0))
2020-06-19 16:04:26.659602 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) extent 0x10000~f000: 0x10000~f000 Blob(0x563a81a0d490 blob([!~10000,0x1870000~10000] csum+has_unused crc32c/0x1000 unused=0xff) use_tracker(0x2*0x10000 0x[0,f000]) SharedBlob(0x563a81a0d500 sbid 0x0))
2020-06-19 16:04:26.659607 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) extent 0x20000~f000: 0x20000~f000 Blob(0x563a81aca070 spanning 0 blob([!~20000,0x1880000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,f000]) SharedBlob(0x563a81a0cd90 sbid 0x0))
2020-06-19 16:04:26.659610 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) extent 0x40000~f000: 0x40000~f000 Blob(0x563a81aca4d0 spanning 1 blob([!~40000,0x1890000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x5*0x10000 0x[0,0,0,0,f000]) SharedBlob(0x563a81aca2a0 sbid 0x0))
2020-06-19 16:04:26.659613 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) extent 0x50000~f000: 0x50000~f000 Blob(0x563a81acaf50 spanning 2 blob([!~50000,0x18a0000~10000] csum+has_unused crc32c/0x1000 unused=0x1fff) use_tracker(0x6*0x10000 0x[0,0,0,0,0,f000]) SharedBlob(0x563a81acaee0 sbid 0x0))
2020-06-19 16:04:26.659617 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) extent 0x60000~10000: 0x60000~10000 Blob(0x563a81acb810 blob([!~60000,0x18b0000~10000] csum crc32c/0x1000) use_tracker(0x7*0x10000 0x[0,0,0,0,0,0,10000]) SharedBlob(0x563a81acb880 sbid 0x0))
2020-06-19 16:04:26.659620 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard  new []
2020-06-19 16:04:26.659621 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard old [0x0(0xeb bytes),0x20000(0xc bytes),0x50000(0x8 bytes)]
2020-06-19 16:04:26.659624 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) reshard  fin []
2020-06-19 16:04:26.659625 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) reshard un-spanning Blob(0x563a81aca070 blob([!~20000,0x1880000~10000] csum+has_unused crc32c/0x1000 unused=0x3ff) use_tracker(0x3*0x10000 0x[0,0,f000]) SharedBlob(0x563a81a0cd90 sbid 0x0))
2020-06-19 16:04:26.659629 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) reshard un-spanning Blob(0x563a81aca4d0 blob([!~40000,0x1890000~10000] csum+has_unused crc32c/0x1000 unused=0xfff) use_tracker(0x5*0x10000 0x[0,0,0,0,f000]) SharedBlob(0x563a81aca2a0 sbid 0x0))
2020-06-19 16:04:26.659633 7f175f7ec700 30 bluestore.extentmap(0x563a819790b0) reshard un-spanning Blob(0x563a81acaf50 blob([!~50000,0x18a0000~10000] csum+has_unused crc32c/0x1000 unused=0x1fff) use_tracker(0x6*0x10000 0x[0,0,0,0,0,f000]) SharedBlob(0x563a81acaee0 sbid 0x0))
2020-06-19 16:04:26.659637 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) update #2:ddb85c4a:::rbd_data.4e60666b8b4567.0000000000000000:head# force
2020-06-19 16:04:26.659642 7f175f7ec700 20 bluestore.extentmap(0x563a819790b0) update  inline shard 1694 bytes from 6 extents
2020-06-19 16:04:26.659648 7f175f7ec700 20 bluestore(/sandstone-data/ceph-4) onode #2:ddb85c4a:::rbd_data.4e60666b8b4567.0000000000000000:head# is 2081 (381 bytes onode + 2 bytes spanning blobs + 1698 bytes inline extents)

\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\
```

状态机

保序

bluestore/agilestore，每个写请求，会对应至少2个aio write（一个store发起的data的aio，一个rocksdb发起的写wal的aio），2个bdev sync（两个都是kv_sync_thread发起的，一个是为了使data io persistent的bdev sync，一个是kv_sync_thread调用db->Write(sync=true)从而触发bluefs去sync脏的wal aio，这个时候该线程会在std::condition_variable上等待未完成的aio，可能会导致cs）

SHA-1: 1a39872b9e64801c8878703dc5d7b6799cceb74a
* **os/bluestore: try to split blobs instead of spanning them**

```
avg-cpu:  %user   %nice %system %iowait  %steal   %idle
          24.31    0.00   10.58   50.15    0.00   14.96

Device:         rrqm/s   wrqm/s     r/s     w/s    rMB/s    wMB/s avgrq-sz avgqu-sz   await r_await w_await  svctm  %util
nvme0n1           0.00 35645.00 21905.00 6258.00   607.25   721.62    96.64   160.80    4.27    0.40   17.82   0.04 100.10
nvme1n1           0.00  9473.00 29647.00 5082.00   566.01   611.68    69.45   168.87    4.86    0.28   31.58   0.03 100.10
```