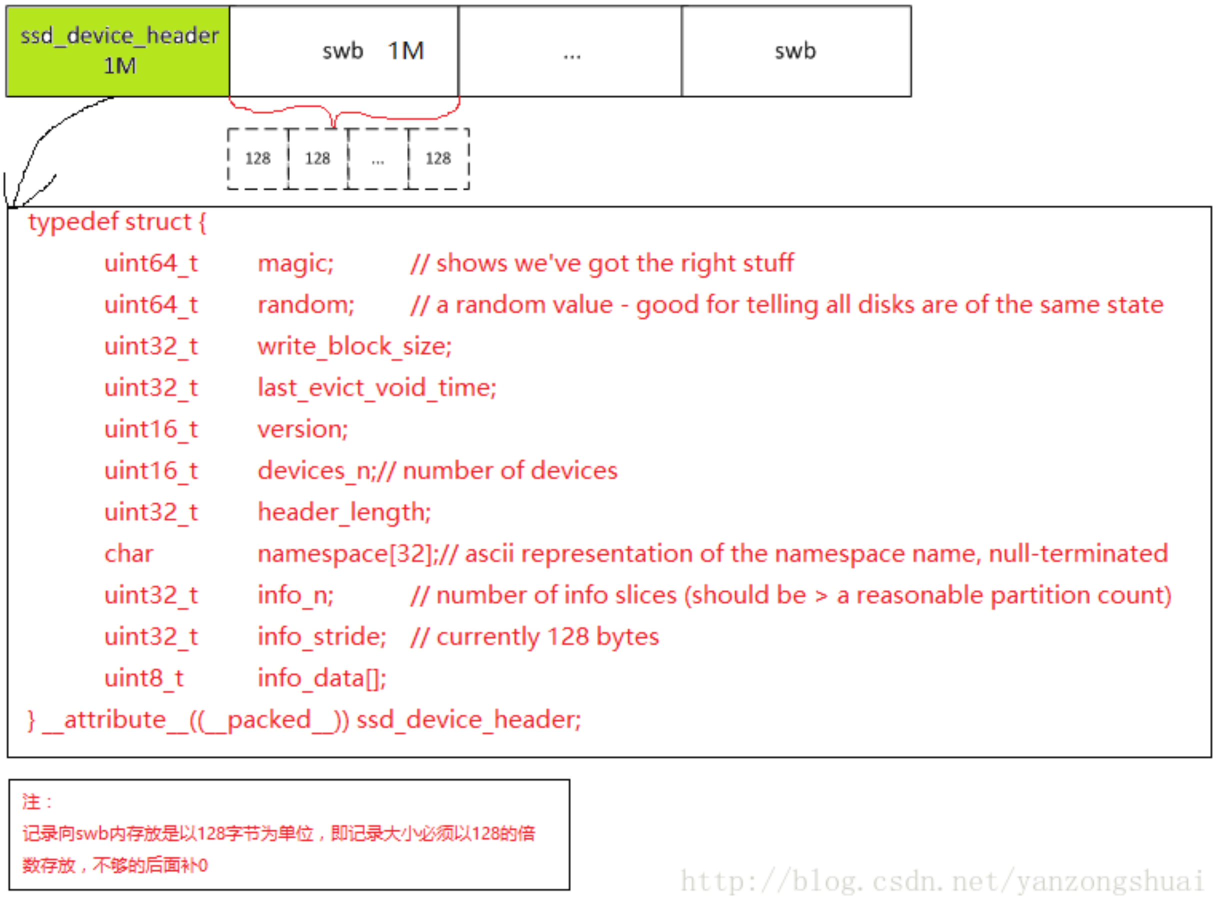
Aerospike SSD模式时，磁盘数据存储格式

1、磁盘数据格式



```
typedef struct {
        uint64_t        magic;          // shows we've got the right stuff
        uint64_t        random;         // a random value - good for telling all disks are of the same state
        uint32_t        write_block_size;
        uint32_t        last_evict_void_time;
        uint16_t        version;
        uint16_t        devices_n;// number of devices
        uint32_t        header_length;
        char            namespace[32];// ascii representation of the namespace name, null-terminated
        uint32_t        info_n;         // number of info slices (should be > a reasonable partition count)
        uint32_t        info_stride;    // currently 128 bytes
        uint8_t         info_data[];
} __attribute__((__packed__)) ssd_device_header;
```

注：
记录向swb内存放是以128字节为单位，即记录大小必须以128的倍数存放，不够的后面补0

2、代码分析

```

//磁盘头初始化函数
ssd_device_header *
ssd_init_header(as_namespace *ns)
{    //header的大小是1M
    ssd_device_header *h = cf_valloc(SSD_DEFAULT_HEADER_LENGTH);

    if (! h) {
        return 0;
    }

    memset(h, 0, SSD_DEFAULT_HEADER_LENGTH);

    h->magic = SSD_HEADER_MAGIC;
    h->random = 0;
    h->write_block_size = ns->storage_write_block_size;
    h->last_evict_void_time = 0;
    h->version = SSD_VERSION;
    h->devices_n = 0;
    h->header_length = SSD_DEFAULT_HEADER_LENGTH;
    memset(h->namespace, 0, sizeof(h->namespace));
    strcpy(h->namespace, ns->name);
    h->info_n = AS_PARTITIONS;
    h->info_stride = SSD_HEADER_INFO_STRIDE;

    return h;
}
```

3、SSD模式下，刷盘是随机的

```
//当current_swb写满时，从ssd->swb_free_q队列获取一个空闲的swb
ssd_write_bins->swb = swb_get(ssd)->cf_queue_pop(ssd->swb_free_q, &swb, CF_QUEUE_NOWAIT)
/*
1、而ssd->swb_free_q链表里的swb并不是按磁盘从头到尾的顺序排列的
2、后台线程从脏队列拿出一个刷完后放到swb_free_q队列里
*/
ssd_write_worker->cf_queue_pop(ssd->swb_write_q, &swb, 100)->
ssd_flush_swb(ssd, swb)->ssd_post_write->swb_dereference_and_release->
swb_release->cf_queue_push(swb->ssd->swb_free_q, &swb)


//swb和磁盘的关系是1M1M对应的
ssd_flush_swb->off_t write_offset = (off_t)WBLOCK_ID_TO_BYTES(ssd, swb->wblock_id);
            ->lseek(fd, write_offset, SEEK_SET)
            ->write(fd, swb->buf, ssd->write_block_size)
static inline uint64_t WBLOCK_ID_TO_BYTES(drv_ssd *ssd, uint32_t wblock_id) {
    return (uint64_t)wblock_id * (uint64_t)ssd->write_block_size;
}
```

swb不按照磁盘从小到大进行取，刷写时磁盘可能跳来跳去，即刷写时随机写。对于普通硬盘来说性能是不容乐观的。所以Aerospike官方对于SSD模式也推荐使用SSD盘进行存储数据。