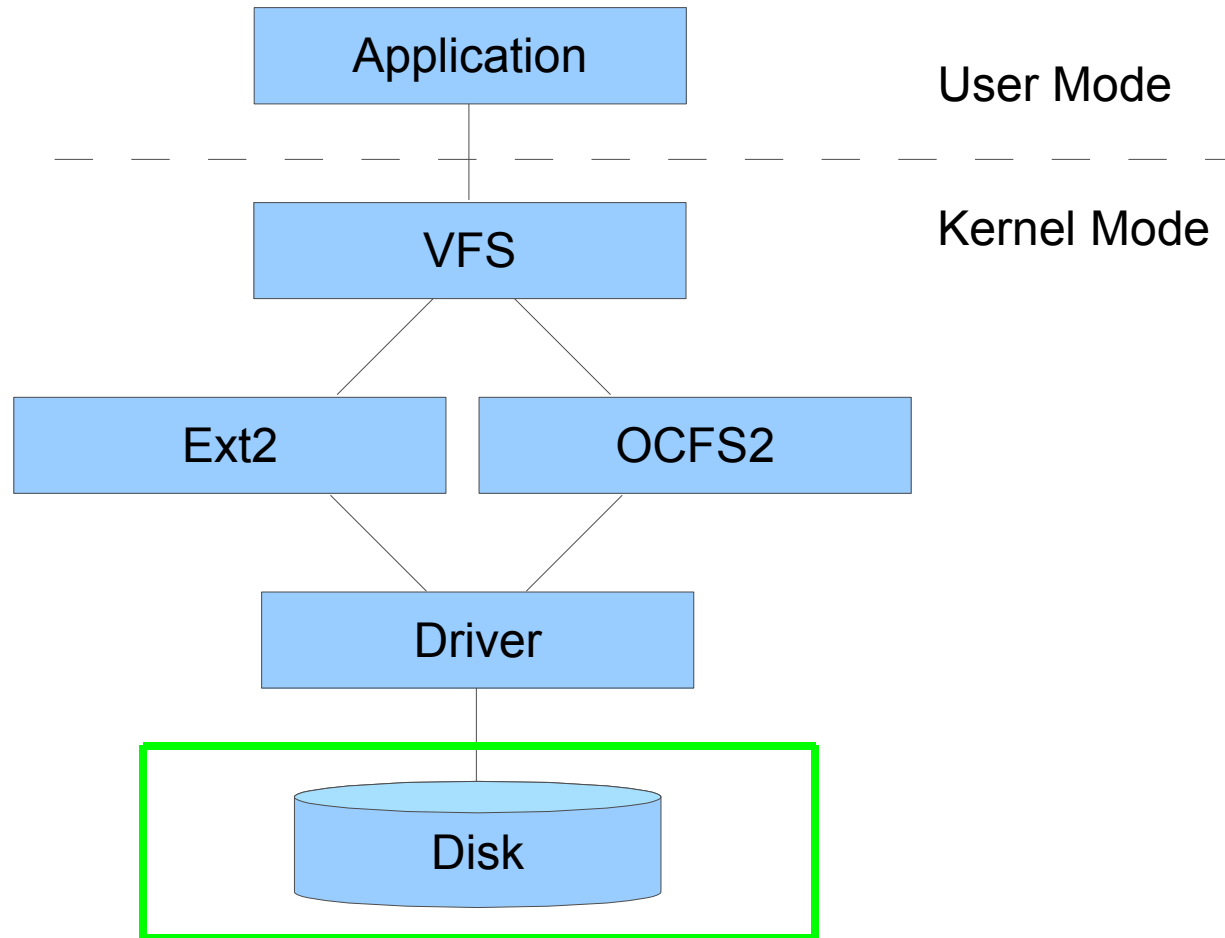# Disk layout of ext2 and ocfs2
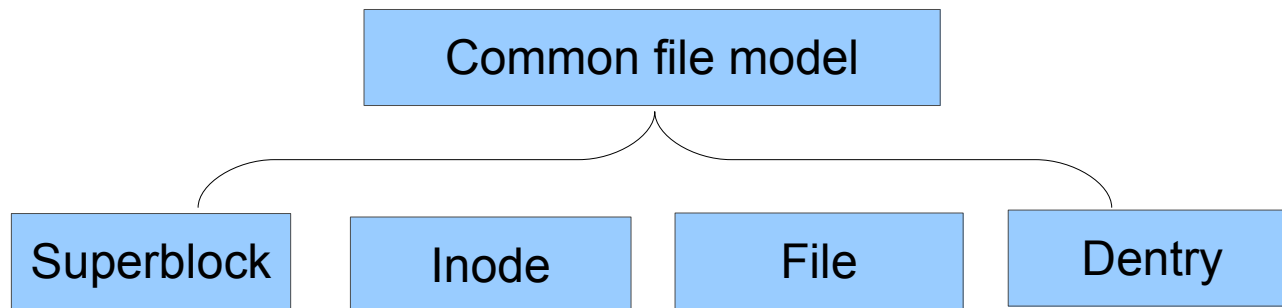
ETSD

tiger yang

2006-8-4

Linux file system

ORACLE

Virtual file system

➢ Providing a common interface to various file systems
➢ System call: mount, umount, mkdir, symlink, chown, stat, chmod, open, close, select, read, write, truncate...
➢ Common file model:

```
                    ┌──────────────────────┐
                    │  Common file model   │
                    └──────────────────────┘
          ┌──────────────┬─────────┴──────┬──────────────┐
   ┌────────────┐  ┌──────────┐      ┌──────────┐   ┌──────────┐
   │ Superblock │  │  Inode   │      │   File   │   │  Dentry  │
   └────────────┘  └──────────┘      └──────────┘   └──────────┘
```

Data structure of Virtual file system

- in /include/linux/fs.h
- struct super_block and struct super_operations
- struct inode and struct inode_operations
- struct dentry and struct dentry_operations
- struct file and struct file_operations

ORACLE

Ext2 file system

- Since 1994
- Choose the block size (1024,2048,4096 bytes)
- Partition disk blocks into group (lower average disk seek time)
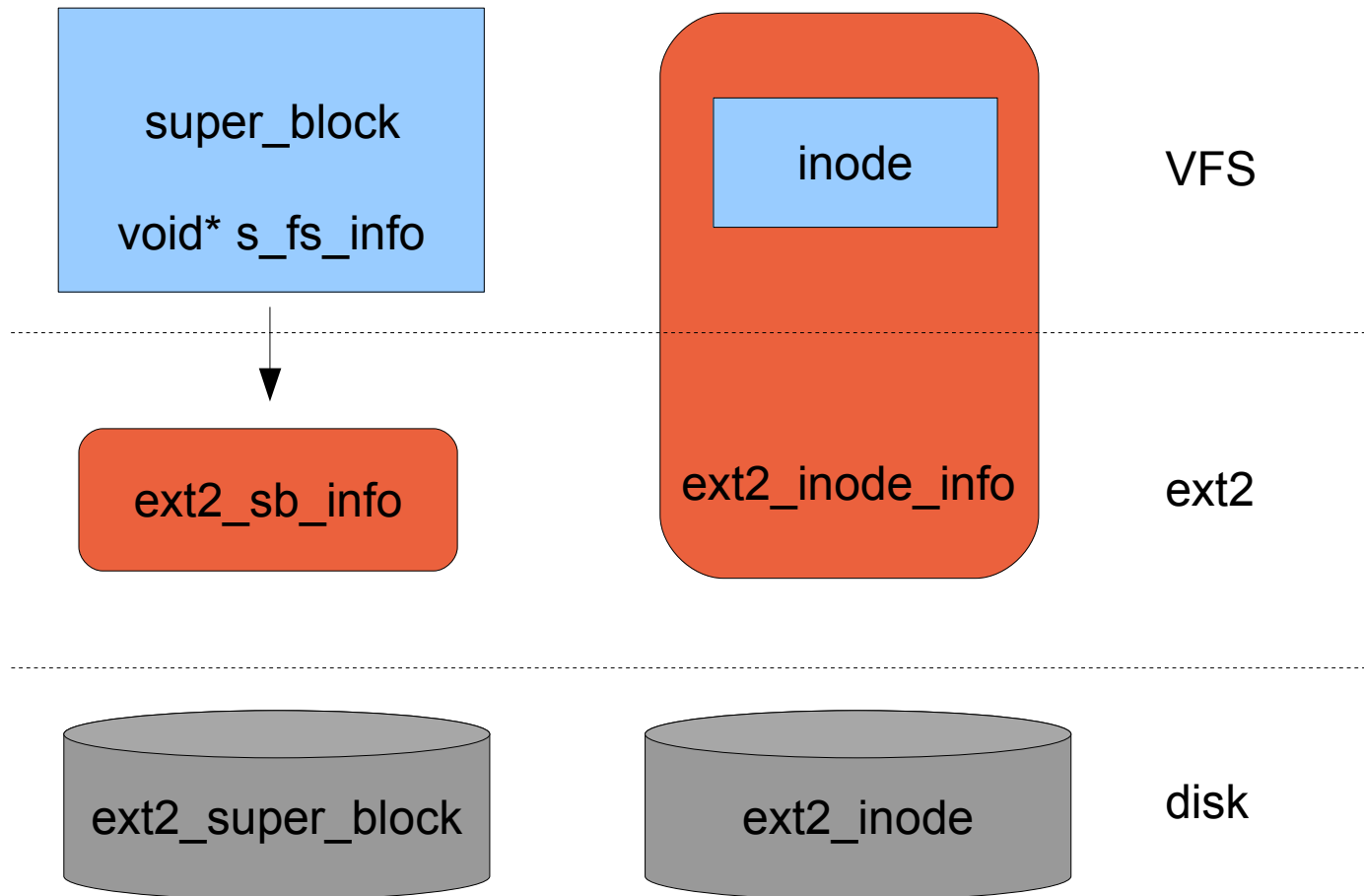- Preallocate disk data block to regular files
- Fast symbolic links
- journaling

ORACLE

Ext2 memory data structures

- in /fs/ext2/ext2.h
- struct ext2_sb_info
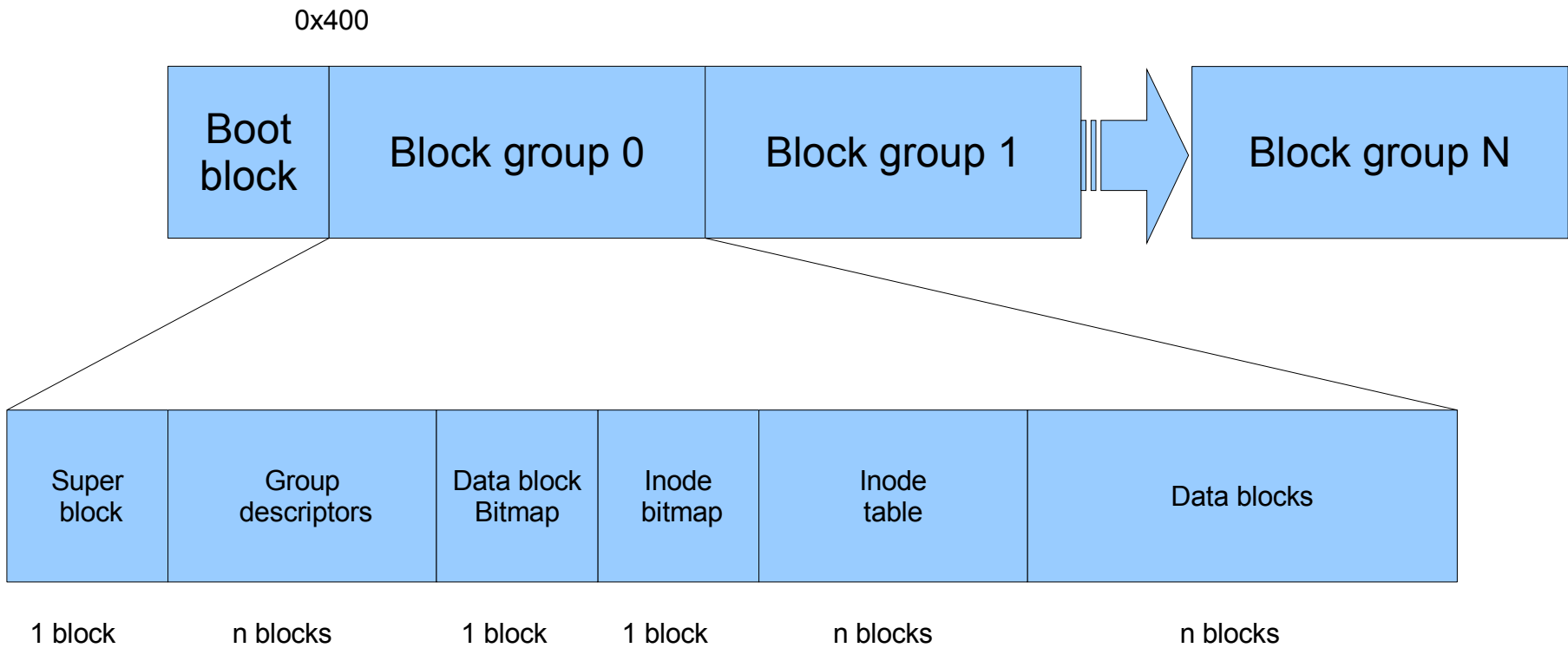- struct ext2_inode_info
- struct ext2_group_desc

Ext2 disk data structures

- in /include/linux/ext2_fs.h
- struct ext2_super_block
- struct ext2_inode
- struct ext2_group_desc

Data structures

Ext2 disk layout

Ext2 super block in disk

```
0000:03f0   00 00 00 00   00 00 00 00   00 00 00 00   00 00 00 00
0000:0400   00 0a 00 00   00 28 00 00   00 02 00 00   5b 26 00 00
0000:0410   f5 09 00 00   01 00 00 00   00 00 00 00   00 00 00 00
0000:0420   00 20 00 00   00 20 00 00   00 05 00 00   00 00 00 00
0000:0430   13 d4 c6 44   00 00 1c 00   53 ef 01 00   01 00 00 00
0000:0440   13 d4 c6 44   00 4e ed 00   00 00 00 00   01 00 00 00
0000:0450   00 00 00 00   0b 00 00 00   80 00 00 00   10 00 00 00
0000:0460   02 00 00 00   01 00 00 00   72 e8 15 43   b5 f1 45 bb
0000:0470   96 69 9c b3   0f c2 fa 6b   00 00 00 00   00 00 00 00
0000:0480   00 00 00 00   00 00 00 00   00 00 00 00   00 00 00 00
0000:0490   00 00 00 00   00 00 00 00   00 00 00 00   00 00 00 00
0000:04a0   00 00 00 00   00 00 00 00   00 00 00 00   00 00 00 00
0000:04b0   00 00 00 00   00 00 00 00   00 00 00 00   00 00 00 00
0000:04c0   00 00 00 00   00 00 00 00   00 00 00 00   00 00 27 00
0000:04d0   00 00 00 00   00 00 00 00   00 00 00 00   00 00 00 00
0000:04e0   00 00 00 00   00 00 00 00   00 00 00 00   bf 89 9e f9
0000:04f0   98 ba 42 b3   88 b3 d6 de   59 e5 92 d9   02 00 00 00
0000:0500   00 00 00 00   00 00 00 00   13 d4 c6 44   00 00 00 00
0000:0510   00 00 00 00   00 00 00 00   00 00 00 00   00 00 00 00
0000:0520   00 00 00 00   00 00 00 00   00 00 00 00   00 00 00 00
0000:0530   00 00 00 00   00 00 00 00   00 00 00 00   00 00 00 00
```

```
struct ext2_super_block {
00    __le32 s_inodes_count;            /* Inodes count */
      __le32 s_blocks_count;            /* Blocks count */
      __le32 s_r_blocks_count;          /* Reserved blocks count */
      __le32 s_free_blocks_count;       /* Free blocks count */
10    __le32 s_free_inodes_count;       /* Free inodes count */
      __le32 s_first_data_block;        /* First Data Block */
      __le32 s_log_block_size;          /* Block size */
      __le32 s_log_frag_size;           /* Fragment size */
20    __le32 s_blocks_per_group;        /* # Blocks per group */
      __le32 s_frags_per_group;         /* # Fragments per group */
      __le32 s_inodes_per_group;        /* # Inodes per group */
      __le32 s_mtime;                   /* Mount time */
30    __le32 s_wtime;                   /* Write time */
      __le16 s_mnt_count;               /* Mount count */
      __le16 s_max_mnt_count;           /* Maximal mount count */
      __le16 s_magic;                   /* Magic signature */
      __le16 s_state;                   /* File system state */
      __le16 s_errors;                  /* Behaviour when detecting errors */
      __le16 s_minor_rev_level;         /* minor revision level */
40    __le32 s_lastcheck;               /* time of last check */
      __le32 s_checkinterval;           /* max. time between checks */
      __le32 s_creator_os;              /* OS */
      __le32 s_rev_level;               /* Revision level */
50    __le16 s_def_resuid;              /* Default uid for reserved blocks */
      __le16 s_def_resgid;              /* Default gid for reserved blocks */
      __le32 s_first_ino;               /* First non-reserved inode */
      __le16  s_inode_size;             /* size of inode structure */
      __le16 s_block_group_nr;          /* block group # of this superblock */
      __le32 s_feature_compat;          /* compatible feature set */
60    __le32 s_feature_incompat;        /* incompatible feature set */
      __le32 s_feature_ro_compat;       /* readonly-compatible feature set */
68    __u8    s_uuid[16];               /* 128-bit uuid for volume */
78    char   s_volume_name[16];         /* volume name */
88    char   s_last_mounted[64];        /* directory where last mounted */
c8    __le32 s_algorithm_usage_bitmap;  /* For compression */
      __u8    s_prealloc_blocks;        /* Nr of blocks to try to preallocate*/
      __u8    s_prealloc_dir_blocks;    /* Nr to preallocate for dirs */
      __u16  s_padding1;
d0    __u8    s_journal_uuid[16];       /* uuid of journal superblock */
e0    __u32  s_journal_inum;            /* inode number of journal file */
      __u32  s_journal_dev;             /* device number of journal file */
      __u32  s_last_orphan;             /* start of list of inodes to delete */
ec    __u32  s_hash_seed[4];            /* HTREE hash seed */
fc    __u8    s_def_hash_version;       /* Default hash version to use */
      __u8    s_reserved_char_pad;
      __u16  s_reserved_word_pad;
100   __le32 s_default_mount_opts;
      __le32 s_first_meta_bg;           /* First metablock block group */
      __u32  s_reserved[190];           /* Padding to the end of the block */
}
```

Ext2 group descriptors in disk

0000:07f0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:0800	2a	00	00	00	2b	00	00	00	2c	00	00	00	27	1f	f5	04	* . . . + . .
0000:0810	02	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:0820	2a	20	00	00	2b	20	00	00	2c	20	00	00	34	07	00	05	* . . + .
0000:0830	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:0840	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .

```
struct ext2_group_desc
{
        __le32  bg_block_bitmap;            /* Blocks bitmap block */
        __le32  bg_inode_bitmap;            /* Inodes bitmap block */
        __le32  bg_inode_table;          /* Inodes table block */
        __le16  bg_free_blocks_count;   /* Free blocks count */
        __le16  bg_free_inodes_count;   /* Free inodes count */
        __le16  bg_used_dirs_count;     /* Directories count */
        __le16  bg_pad;
        __le32  bg_reserved[3];
};
```

0x2a*0x400=0xA800

0x2b*0x400=0xAC00

0x2c*0x400=0xB000

ORACLE

Ext2 block bitmap in disk

0000:a7f0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:a800	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0000:a810	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	03	00	00	00	00	. . . . .
0000:a820	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:a830	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .

Block group0

0080:a7f0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0080:a800	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0080:a810	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0080:a820	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0080:a830	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0080:a840	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0080:a850	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0080:a860	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0080:a870	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0080:a880	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0080:a890	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0080:a8a0	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0080:a8b0	ff	ff	ff	ff	ff	ff	ff	03	00	00	00	00	00	00	00	00	
0080:a8c0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0080:a8d0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0080:a8e0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0080:a8f0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	80	. . . .
0080:a900	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0080:a910	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	

Block group1

ORACLE

Ext2 inode bitmap in disk

Ext2 inode table in disk

0000:aff0	f	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0000:b000	00	00	00	00	00	00	00	00	13	d4	c6	44	13	d4	c6	44	. . . .
0000:b010	13	d4	c6	44	00	00	00	00	00	00	00	00	00	00	00	00	. D . . .
0000:b020	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:b030	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:b040	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:b050	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:b060	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:b070	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:b080	ed	41	f4	01	00	04	00	00	42	e0	c6	44	3c	e0	c6	44	A . . .
0000:b090	3c	e0	c6	44	00	00	00	00	f5	01	04	00	02	00	00	00	< D . . .
0000:b0a0	00	00	00	00	00	00	00	00	cc	00	00	00	00	00	00	00	. . . .
0000:b0b0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:b0c0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:b0d0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:b0e0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:b0f0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:b100	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:b110	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:b120	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .

Inode of root

ORACLE

Ext2 inode of root

```
0000:b070   00 00 00 00  00 00 00 00  00 00 00 00  00 00 00 00   ........
0000:b080   ed 41 f4 01  00 04 00 00  cc 12 c7 44  c4 12 c7 44   A......
0000:b090   c4 12 c7 44  00 00 00 00  f5 01 04 00  02 00 00 00   . D...
0000:b0a0   00 00 00 00  00 00 00 00  cc 00 00 00  00 00 00 00   ........
0000:b0b0   00 00 00 00  00 00 00 00  00 00 00 00  00 00 00 00   ........
0000:b0c0   00 00 00 00  00 00 00 00  00 00 00 00  00 00 00 00   ........
0000:b0d0   00 00 00 00  00 00 00 00  00 00 00 00  00 00 00 00   ........
0000:b0e0   00 00 00 00  00 00 00 00  00 00 00 00  00 00 00 00   ........
```

```
struct ext2_inode {
        __le16  i_mode;          /* File mode */
        __le16  i_uid;           /* Low 16 bits of Owner Uid */
        __le32  i_size;          /* Size in bytes */
        __le32  i_atime;         /* Access time */
        __le32  i_ctime;         /* Creation time */
        __le32  i_mtime;         /* Modification time */
        __le32  i_dtime;         /* Deletion Time */
        __le16  i_gid;           /* Low 16 bits of Group Id */
        __le16  i_links_count;   /* Links count */
        __le32  i_blocks;        /* Blocks count */
        __le32  i_flags;         /* File flags */
        union { __le32 } osd1;                  /* OS dependent 1 */
        __le32  i_block[15];/* Pointers to blocks */          0xcc*0x400=0x33000
        __le32  i_generation;    /* File version (for NFS) */
        __le32  i_file_acl;      /* File ACL */
        __le32  i_dir_acl;       /* Directory ACL */
        __le32  i_faddr;         /* Fragment address */
        union {} osd2;                          /* OS dependent 2 */
};
```

Ext2 root directory in disk (data)

```
0003:2ff0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0003:3000   02 00 00 00 0c 00 01 02 2e 00 00 00 02 00 00 00   .............
0003:3010   0c 00 02 02 2e 2e 00 00 0b 00 00 00 14 00 0a 02   ................
0003:3020   6c 6f 73 74 2b 66 6f 75 6e 64 00 00 01 05 00 00   lost+found......
0003:3030   0c 00 03 02 61 62 63 00 11 00 00 00 2c 00 22 01   ....abc.....,.".
0003:3040   6f 63 66 73 32 2d 32 2e 36 2e 39 2d 32 32 2e 45   ocfs2-2.6.9-22.E
0003:3050   4c 2d 31 2e 32 2e 31 2d 31 2e 69 36 38 36 2e 72   L-1.2.1-1.i686.r
0003:3060   70 6d 00 00 12 00 00 00 14 00 09 07 6f 63 66 73   pm..........ocfs
0003:3070   32 2e 72 70 6d 00 00 00 13 00 00 00 88 03 07 07   2.rpm...........
0003:3080   61 62 63 64 65 66 67 00 00 00 00 00 00 00 00 00   abcdefg.........
0003:3090   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
```

```c
struct ext2_dir_entry_2 {
    __le32  inode;                        /* Inode number */
    __le16  rec_len;                      /* Directory entry length */
    __u8    name_len;                     /* Name length */
    __u8    file_type;
    char    name[EXT2_NAME_LEN];   /* File name */
};
```

```c
enum {
    EXT2_FT_UNKNOWN,
    EXT2_FT_REG_FILE,
    EXT2_FT_DIR,
    EXT2_FT_CHRDEV,
    EXT2_FT_BLKDEV,
    EXT2_FT_FIFO,
    EXT2_FT_SOCK,
    EXT2_FT_SYMLINK,
    EXT2_FT_MAX
};
```
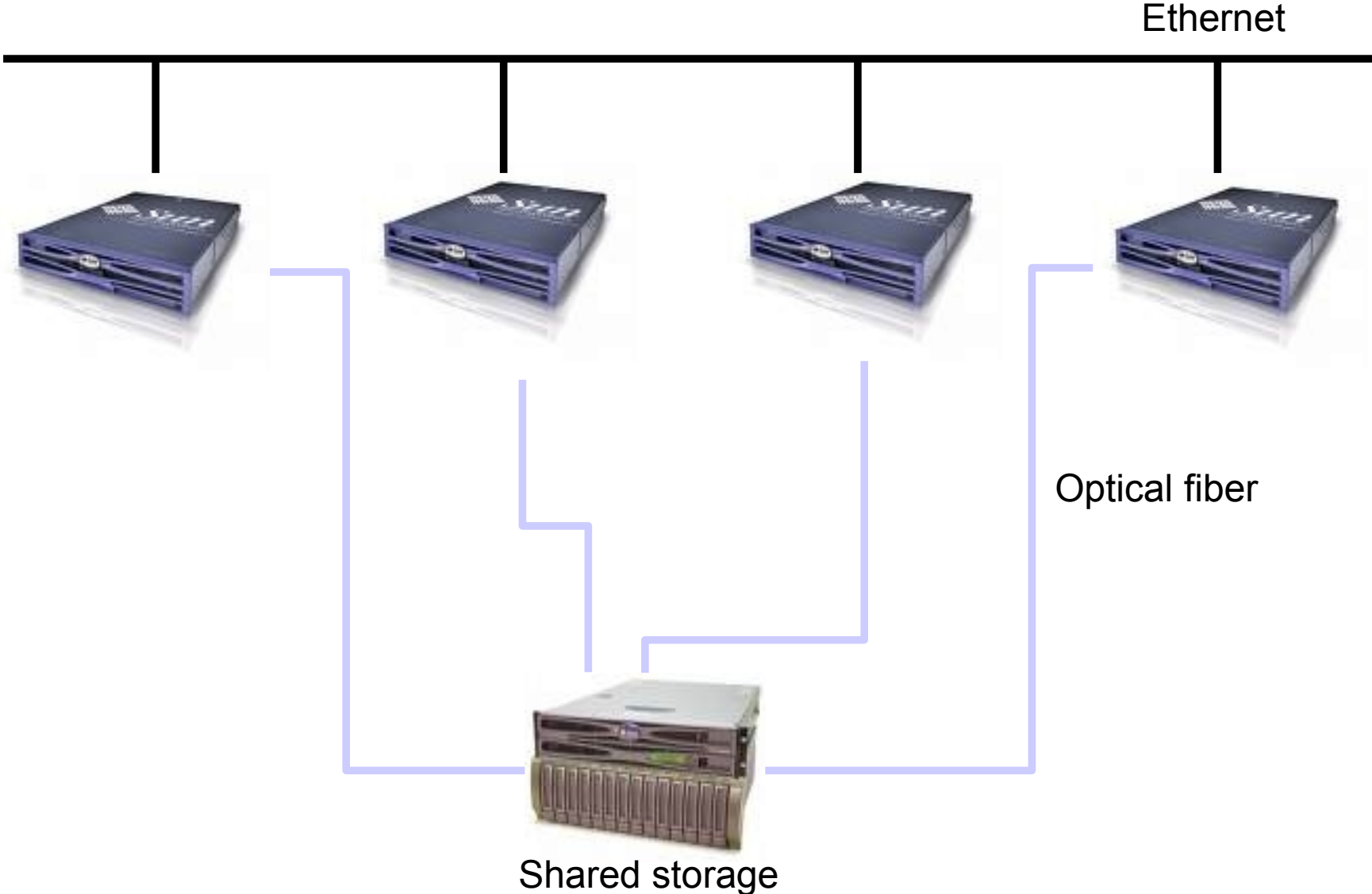
ORACLE

Ext2 inode of symbolic link

0000:b870	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
0000:b880	ff	a1	00	00	22	00	00	00	cc	12	c7	44	66	12	c7	44	.." .... . Df . D
0000:b890	66	12	c7	44	00	00	00	00	00	00	01	00	00	00	00	00	f . D............
0000:b8a0	00	00	00	00	00	00	00	00	6f	63	66	73	32	2d	32	2e	........ocfs2-2.
0000:b8b0	36	2e	39	2d	32	32	2e	45	4c	2d	31	2e	32	2e	31	2d	6.9-22.EL-1.2.1-
0000:b8c0	31	2e	69	36	38	36	2e	72	70	6d	00	00	00	00	00	00	1.i686.rpm......
0000:b8d0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
0000:b8e0	00	00	00	00	bc	f3	e5	b9	00	00	00	00	00	00	00	00	....
0000:b8f0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....

```
struct ext2_inode {
        __le16  i_mode;      /* File mode */
        __le16  i_uid;       /* Low 16 bits of Owner Uid */
        __le32  i_size;      /* Size in bytes */
..............
        __le32  i_block[15];/* Pointers to blocks */
...............
};
```
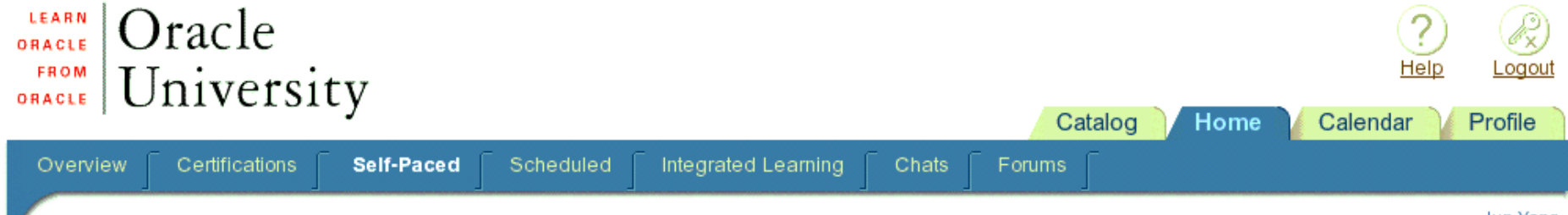
ORACLE

Oracle cluster file system 2

Ethernet

Optical fiber

Shared storage

Introduction to OCFS2

➢ http://ilearning.oracle.com/ilearn/en/learner/jsp/offering_details_home.jsp?classid=183125358

Goals for OCFS2

➤ Become a high performance general purpose Cluster FS

➤ ● Meta data caching

➤ ● Meta data journaling

➤ ● Cross node file data consistency

➤ ● Easy to administer, including operation as shared root fs

➤ ● Integrate well with the Linux Kernel (2.6.16)

➤ ● Cluster aware tools

➤ ● Multiple cluster size (4K ~ 1M)

➤ ● Multiple block sizes (512,1024,2048,4096 bytes)

ORACLE

ocfs2 memory data structures

➢ in fs/ocfs2/ocfs2.h
➢ struct ocfs2_super
➢ in fs/ocfs2/inode.h
➢ struct ocfs2_inode_info

ORACLE

ocfs2 disk data structures

- in fs/ocfs2/ocfs2_fs.h
- struct ocfs2_super_block
- struct ocfs2_dinode
- struct ocfs2_group_desc

ORACLE

Data structures

super_block

void* s_fs_info

inode

VFS

ocfs2_super

ocfs2_inode_info

ocfs2

ocfs2_super_block

ocfs2_dinode
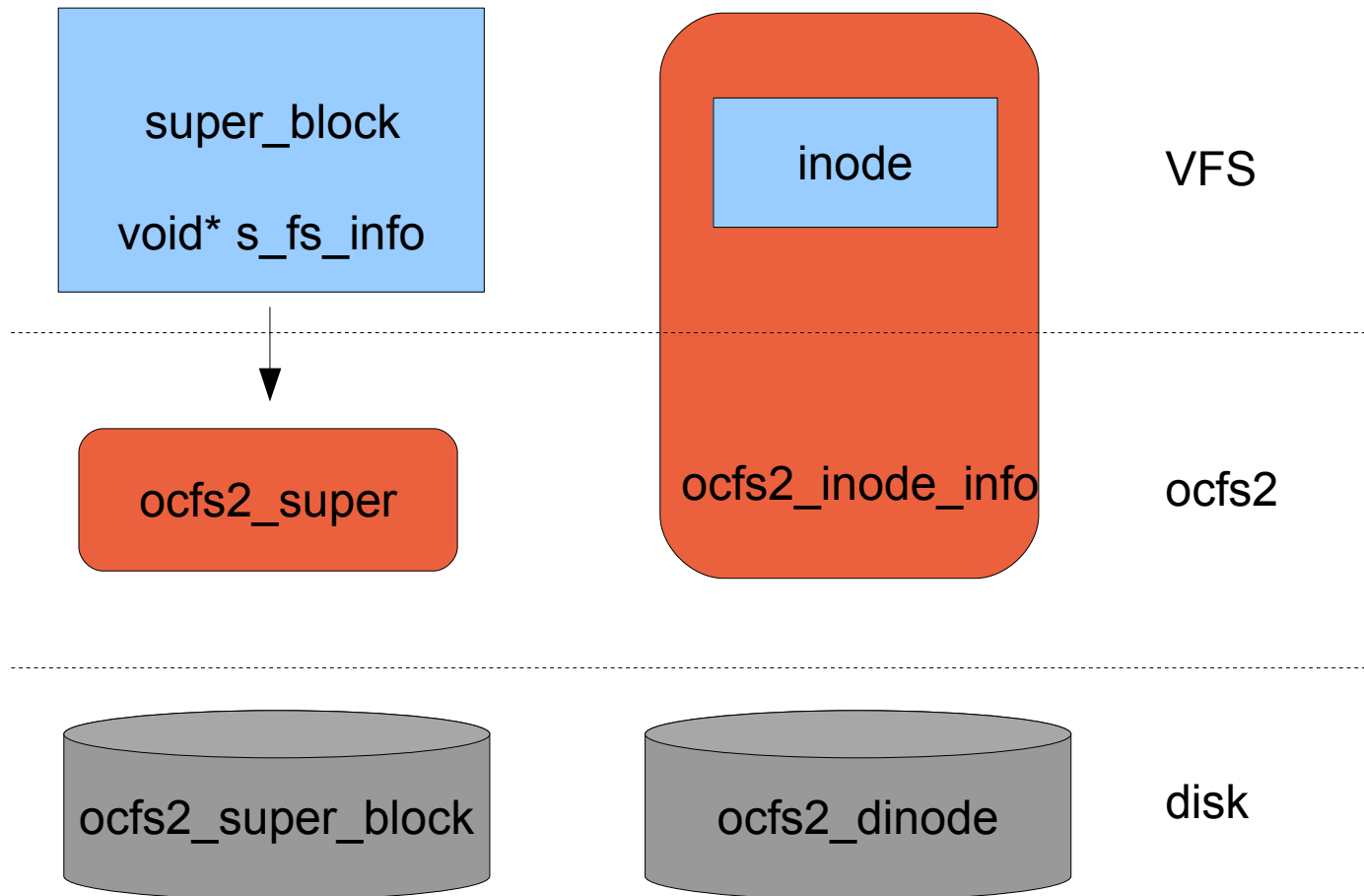
disk

ocfs2 disk layout

Block #2

		cluster group 0	cluster group 1	⇒	cluster group N
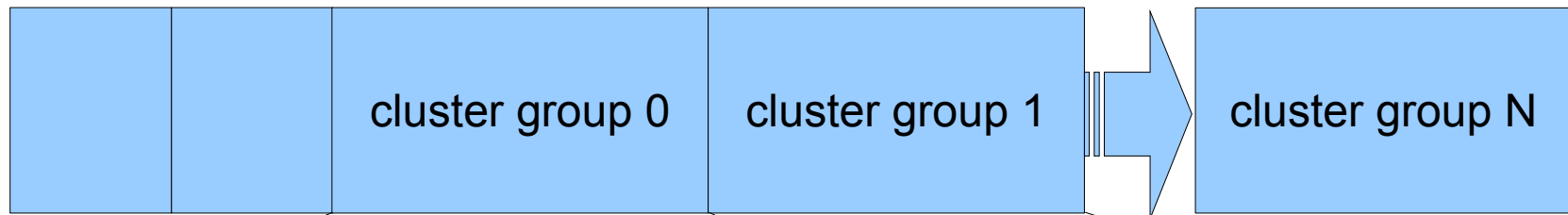
Super block	global_bitmap	global_inode_alloc	System file's data	Data clusters	global_bitmap	Data clusters
1 block	1 cluster	n clusters	n clusters	n clusters	1 cluster	n clusters

ORACLE

Ocfs2 super block in disk


```
0000:07f0   02 02 02 02 02 02 02 02 02 02 02 02 02 02 02 02   .. .....
0000:0800   4f 43 46 53 56 32 00 00 42 69 6f b5 ff ff ff ff   O C F S V 2 .
0000:0810   00 00 00 00 00 28 00 00 00 00 00 00 00 00 00 00   . . . . . (
0000:0820   00 00 00 00 00 00 00 00 00 00 00 00 31 00 00 00   . . . . . . . .
0000:0830   00 00 00 00 00 00 00 00 8d 7f c8 44 00 00 00 00   . . . . . .
0000:0840   8d 7f c8 44 00 00 00 00 00 00 00 00 00 00 00 00   .. D ...
0000:0850   02 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   . . . . . . . .
0000:0860   42 69 6f b5 00 00 00 00 00 00 00 00 00 00 00 00   Bi o . . .
0000:0870   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   . . . . . . . .
0000:0880   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   . . . . . . . .
0000:0890   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   . . . . . . . .
0000:08a0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   . . . . . . . .
0000:08b0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   . . . . . . . .
0000:08c0   00 00 5a 00 00 00 14 00 00 00 00 00 00 00 00 00   . Z . . .
0000:08d0   8d 7f c8 44 00 00 00 00 00 00 00 00 00 00 00 00   . D . . .
0000:08e0   00 00 00 00 00 00 00 00 09 00 00 00 00 00 00 00   . . . . . . . .
0000:08f0   0a 00 00 00 00 00 00 00 0a 00 00 00 0c 00 00 00   . . . . . . . .
0000:0900   02 00 00 00 00 00 00 00 04 00 00 00 00 00 00 00   . . . . . . . .
0000:0910   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   . . . . . . . .
0000:0920   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   . . . . . . . .
0000:0930   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   . . . . . . . .
0000:0940   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   . . . . . . . .
0000:0950   0f 5f 8b 84 ea 5e 45 ed 89 db 45 42 2b bd b8 8d   . _ . . ^E
0000:0960   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   . . . . . . . .
0000:0970   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   . . . . . . . .
```

```
struct ocfs2_super_block {
/*00*/   __le16 s_major_rev_level;
         __le16 s_minor_rev_level;
         __le16 s_mnt_count;
         __le16 s_max_mnt_count;
         __le16 s_state;
         __le16 s_errors;
         __le32 s_checkinterval;
/*10*/   __le64 s_lastcheck;
         __le32 s_creator_os;
         __le32 s_feature_compat;
/*20*/   __le32 s_feature_incompat;
         __le32 s_feature_ro_compat;
         __le64 s_root_blkno;
/*30*/   __le64 s_system_dir_blkno;
         __le32 s_blocksize_bits;
         __le32 s_clustersize_bits;
/*40*/   __le16 s_max_slots;
         __le16 s_reserved1;
         __le32 s_reserved2;
         __le64 s_first_cluster_group;
/*50*/   __u8  s_label[64];
/*90*/   __u8  s_uuid[16];
/*A0*/
};
```

Address of root inode: 0x09*0x400=0x2400

Address of system_dir inode:0x0a*0x400=0x2800

ORACLE

Global bitmap in disk

0000:0ff0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:1000	47	52	4f	55	50	30	31	00	c0	03	00	1e	85	13	00	00	G R O U P 0 1 .
0000:1010	42	69	6f	b5	00	00	00	00	00	00	00	00	00	00	00	00	B i o . . . . . .
0000:1020	0f	00	00	00	00	00	00	00	04	00	00	00	00	00	00	00	. . . .
0000:1030	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	. . . .
0000:1040	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0000:1050	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0000:1060	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0000:1070	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0000:1080	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0000:1090	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	
0000:10a0	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	ff	

```
struct ocfs2_group_desc
{
/*00*/  __u8   bg_signature[8];        /* Signature for validation */
        __le16   bg_size;              /* Size of included bitmap in bytes. */
        __le16   bg_bits;              /* Bits represented by this group. */
        __le16  bg_free_bits_count;    /* Free bits count */
        __le16   bg_chain;             /* What chain I am in. */
/*10*/  __le32   bg_generation;
        __le32  bg_reserved1;
        __le64   bg_next_group;        /* Next group in my list, in blocks */
/*20*/  __le64   bg_parent_dinode;     /* dinode which owns me, in blocks */
        __le64   bg_blkno;             /* Offset on disk, in blocks */
/*30*/  __le64   bg_reserved2[2];
/*40*/  __u8   bg_bitmap[0];
};
```

ORACLE

global_inode_alloc in disk

group_desc &inode_bitmap
root
system_dir
bad_blocks
global_inode_alloc
slot_map
heartbeat
global_bitmap
orphan_dir:0000
orphan-dir:0001
extent_alloc:0000
extent_alloc:0001
inode_alloc:0000
inode_alloc:0001
journal:0000
journal:0001
local_alloc:0000
local_alloc:0001
truncate_log:0000
truncate_log:0001

1 block

global_inode_alloc

ORACLE

Inode of root

```
0000:23f0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000:2400   49 4e 4f 44 45 30 31 00 42 69 6f b5 ff ff 01 00   INODE01.Bio  ..
0000:2410   00 00 00 00 01 00 00 00 00 00 00 00 00 00 00 00   ................
0000:2420   00 04 00 00 00 00 00 00 ed 41 03 00 11 00 00 00   .........A......
0000:2430   8d 7f c8 44 00 00 00 00 8d 7f c8 44 00 00 00 00   ...D.......D....
0000:2440   8d 7f c8 44 00 00 00 00 00 00 00 00 00 00 00 00   ...D............
0000:2450   09 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000:2460   42 69 6f b5 00 00 00 00 00 00 00 00 00 00 00 00   Bio.............
0000:2470   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000:2480   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000:2490   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000:24a0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000:24b0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000:24c0   00 00 33 00 01 00 00 00 00 00 00 00 00 00 00 00   ..3.............
0000:24d0   00 00 00 00 01 00 00 00 d4 00 00 00 00 00 00 00   ................
0000:24e0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
```

0xD4 * 0x400=0x35000

```
0003:4ff0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0003:5000   09 00 00 00 00 00 00 00 10 00 01 02 2e 00 00 00   ................
0003:5010   09 00 00 00 00 00 00 00 10 00 02 02 2e 2e 00 00   ................
0003:5020   e9 21 00 00 00 00 00 00 e0 03 0a 02 6c 6f 73 74   .!..........lost
0003:5030   2b 66 6f 75 6e 64 00 00 00 00 00 00 00 00 00 00   +found..........
0003:5040   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0003:5050   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0003:5060   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0003:5070   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0003:5080   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
```

0x21e9 * 0x400=0x87A4000

Inode of system_dir


```
0000:27f0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ........
0000:2800   49 4e 4f 44 45 30 31 00 42 69 6f b5 ff ff 02 00   INODE01.Bio....
0000:2810   00 00 00 00 01 00 00 00 00 00 00 00 00 00 00 00   ........
0000:2820   00 04 00 00 00 00 00 00 ed 41 04 00 11 00 00 00   ........
0000:2830   8d 7f c8 44 00 00 00 00 8d 7f c8 44 00 00 00 00   ..D.....D...
0000:2840   8d 7f c8 44 00 00 00 00 00 00 00 00 00 00 00 00   ..D.....
0000:2850   0a 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ........
0000:2860   42 69 6f b5 00 00 00 00 00 00 00 00 00 00 00 00   Bio....
0000:2870   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ........
0000:2880   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ........
0000:2890   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ........
0000:28a0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ........
0000:28b0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ........
0000:28c0   00 00 33 00 01 00 00 00 00 00 00 00 00 00 00 00   ..3....
0000:28d0   00 00 00 00 01 00 00 00 d8 00 00 00 00 00 00 00   ........
0000:28e0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ........
0000:28f0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ........
0000:2900   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ........
```

struct ocfs2_extent_list {
/*00*/ __le16 l_tree_depth;
        __le16 l_count;
        __le16 l_next_free_rec;
        __le16 l_reserved1;
        __le64 l_reserved2;
/*10*/ struct ocfs2_extent_rec l_recs[0];
};

struct ocfs2_extent_rec {
/*00*/ __le32 e_cpos;
        __le32 e_clusters;
        __le64 e_blkno;   /* Physical disk offset, in blocks */
};

0xD8 * 0x400=0x36000

struct ocfs2_dinode {
/*00*/ __u8 i_signature[8];
        __le32 i_generation;
        __le16 i_suballoc_slot;
        __le16 i_suballoc_bit;
/*10*/ __le32 i_reserved0;
        __le32 i_clusters;
        __le32 i_uid;
        __le32 i_gid;
/*20*/ __le64 i_size;
        __le16 i_mode;
        __le16 i_links_count;
        __le32 i_flags;
/*30*/ __le64 i_atime;
        __le64 i_ctime;
/*40*/ __le64 i_mtime;
        __le64 i_dtime;
/*50*/ __le64 i_blkno;
        __le64 i_last_eb_blk;
/*60*/ __le32 i_fs_generation;
        __le32 i_atime_nsec;
        __le32 i_ctime_nsec;
        __le32 i_mtime_nsec;
/*70*/ __le64 i_reserved1[9];
/*B8*/ union {
        } id1;
/*C0*/ union {
        struct ocfs2_super_block   i_super;
        struct ocfs2_local_alloc   i_lab;
        struct ocfs2_chain_list   i_chain;
        struct ocfs2_extent_list   i_list;
        struct ocfs2_truncate_log   i_dealloc;
        __u8   i_symlink[0];
        } id2;
};

ORACLE

Extent allocation

Block Allocation

Extent Allocation

system_dir in disk

0003:5ff0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	....
0003:6000	0a	00	00	00	00	00	00	00	10	00	01	02	2e	00	00	00	....
0003:6010	0a	00	00	00	00	00	00	00	10	00	02	02	2e	2e	00	00	....
0003:6020	0b	00	00	00	00	00	00	00	18	00	0a	01	62	61	64	5f	.... bad_
0003:6030	62	6c	6f	63	6b	73	00	00	0c	00	00	00	00	00	00	00	blocks..........
0003:6040	20	00	12	01	67	6c	6f	62	61	6c	5f	69	6e	6f	64	65	...global_inode
0003:6050	5f	61	6c	6c	6f	63	00	00	0d	00	00	00	00	00	00	00	_alloc..........
0003:6060	14	00	08	01	73	6c	6f	74	5f	6d	61	70	0e	00	00	00	....slot_map....
0003:6070	00	00	00	00	18	00	09	01	68	65	61	72	74	62	65	61	........heartbea
0003:6080	74	00	00	00	0f	00	00	00	00	00	00	00	1c	00	0d	01	t...............
0003:6090	67	6c	6f	62	61	6c	5f	62	69	74	6d	61	70	00	00	00	global_bitmap...
0003:60a0	10	00	00	00	00	00	00	00	1c	00	0f	02	6f	72	70	68	............orph
0003:60b0	61	6e	5f	64	69	72	3a	30	30	30	30	00	11	00	00	00	an_dir:0000.....
0003:60c0	00	00	00	00	1c	00	0f	02	6f	72	70	68	61	6e	5f	64	........orphan_d
0003:60d0	69	72	3a	30	30	30	31	00	12	00	00	00	00	00	00	00	ir:0001.........
0003:60e0	20	00	11	01	65	78	74	65	6e	74	5f	61	6c	6c	6f	63	...extent_alloc
0003:60f0	3a	30	30	30	30	00	00	00	13	00	00	00	00	00	00	00	:0000..........
0003:6100	20	00	11	01	65	78	74	65	6e	74	5f	61	6c	6c	6f	63	...extent_alloc
0003:6110	3a	30	30	30	31	00	00	00	14	00	00	00	00	00	00	00	:0001..........
0003:6120	1c	00	10	01	69	6e	6f	64	65	5f	61	6c	6c	6f	63	3a	....inode_alloc:
0003:6130	30	30	30	30	15	00	00	00	00	00	00	00	1c	00	10	01	0000............
0003:6140	69	6e	6f	64	65	5f	61	6c	6c	6f	63	3a	30	30	30	31	inode_alloc:0001
0003:6150	16	00	00	00	00	00	00	00	18	00	0c	01	6a	6f	75	72	............jour
0003:6160	6e	61	6c	3a	30	30	30	30	17	00	00	00	00	00	00	00	nal:0000.........
0003:6170	18	00	0c	01	6a	6f	75	72	6e	61	6c	3a	30	30	30	31	....journal:0001
0003:6180	18	00	00	00	00	00	00	00	1c	00	10	01	6c	6f	63	61	............loca
0003:6190	6c	5f	61	6c	6c	6f	63	3a	30	30	30	30	19	00	00	00	l_alloc:0000....
0003:61a0	00	00	00	00	1c	00	10	01	6c	6f	63	61	6c	5f	61	6c	........local_al
0003:61b0	6c	6f	63	3a	30	30	30	31	1a	00	00	00	00	00	00	00	loc:0001.........
0003:61c0	20	00	11	01	74	72	75	6e	63	61	74	65	5f	6c	6f	67	...truncate_log
0003:61d0	3a	30	30	30	30	00	00	00	1b	00	00	00	00	00	00	00	:0000..........
0003:61e0	28	02	11	01	74	72	75	6e	63	61	74	65	5f	6c	6f	67	(...truncate_log
0003:61f0	3a	30	30	30	31	00	00	00	00	00	00	00	00	00	00	00	:0001..........
0003:6200	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	

ORACLE

inode_alloc:0000 in disk

```
0000:4ff0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000:5000   49 4e 4f 44 45 30 31 00 42 69 6f b5 ff ff 0c 00   INODE01.Bio ...
0000:5010   00 00 00 00 00 02 00 00 00 00 00 00 00 00 00 00   ................
0000:5020   00 00 20 00 00 00 00 00 a4 81 01 00 91 04 00 00   ................
0000:5030   8d 7f c8 44 00 00 00 00 8d 7f c8 44 00 00 00 00   .. D..... D....
0000:5040   8d 7f c8 44 00 00 00 00 00 00 00 00 00 00 00 00   .. D............
0000:5050   14 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000:5060   42 69 6f b5 00 00 00 00 00 00 00 00 00 00 00 00   Bio ............
0000:5070   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000:5080   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000:5090   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000:50a0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000:50b0   00 00 00 00 00 00 00 00 02 00 00 00 00 08 00 00   ................
0000:50c0   00 02 04 00 33 00 01 00 00 00 00 00 00 00 00 00   ....3...........
0000:50d0   fe 07 00 00 00 08 00 00 e8 21 00 00 00 00 00 00   .........!......
0000:50e0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
```

0x21e8 * 0x400=0x87A0000

```
0087:9ff0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0087:a000   47 52 4f 55 50 30 31 00 c0 03 00 08 fe 07 00 00   GROUP01. .......
0087:a010   42 69 6f b5 00 00 00 00 00 00 00 00 00 00 00 00   Bio ............
0087:a020   14 00 00 00 00 00 00 00 e8 21 00 00 00 00 00 00   .........!......
0087:a030   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0087:a040   03 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0087:a050   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0087:a060   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0087:a070   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
```

```
0087:a3f0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0087:a400   49 4e 4f 44 45 30 31 00 42 69 6f b5 00 00 01 00   INODE01.Bio ....
0087:a410   00 00 00 00 01 00 00 00 00 00 00 00 00 00 00 00   ................
0087:a420   00 04 00 00 00 00 00 00 ed 41 02 00 01 00 00 00   .........A......
0087:a430   8d 7f c8 44 00 00 00 00 8d 7f c8 44 00 00 00 00   .. D..... D....
0087:a440   8d 7f c8 44 00 00 00 00 00 00 00 00 00 00 00 00   .. D............
0087:a450   e9 21 00 00 00 00 00 00 00 00 00 00 00 00 00 00   .!..............
0087:a460   42 69 6f b5 00 00 00 00 00 00 00 00 00 00 00 00   Bio ............
0087:a470   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0087:a480   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0087:a490   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0087:a4a0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0087:a4b0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0087:a4c0   00 00 33 00 01 00 00 00 00 00 00 00 00 00 00 00   ..3.............
0087:a4d0   00 00 00 00 01 00 00 00 e8 29 00 00 00 00 00 00   .........)......
0087:a4e0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
```

debugfs.ocfs2

1. stats -h   -view the super block
2. ls -l /   -list root directory
3. ls -l //   -list system_dir directory
4. stat //global_inode_alloc   -view inode

ORACLE

Q & A

ORACLE®

Thanks !

ORACLE®