Object Based Storage Devices Presentation

- Motivation behind this proposal

- Architectural model overview

- Relevant topics

  - Objects

  - Groups

  - Sessions

- Commands

- Attributes

Seagate.

Information the way you want it.

OBSD: Where are we, What do we want

Today, nothing etched in silicon

- Everything candidate to change

Research has proven concept

- CMU, Seagate, IBM all have done something
- Reasonable level of confidence in proposal

Flocks of Chickens & Dozens of Eggs

- To progress need broad participation
    - HBA vendors
    - File system suppliers
    - Storage vendors
    - Interconnect providers
- They need to have confidence, reference spec

Need to collaborate with SNIA

- File systems, other software contributions

Seagate

Information the way you want it.

OBSD Goal: Exploit Storage Intelligence

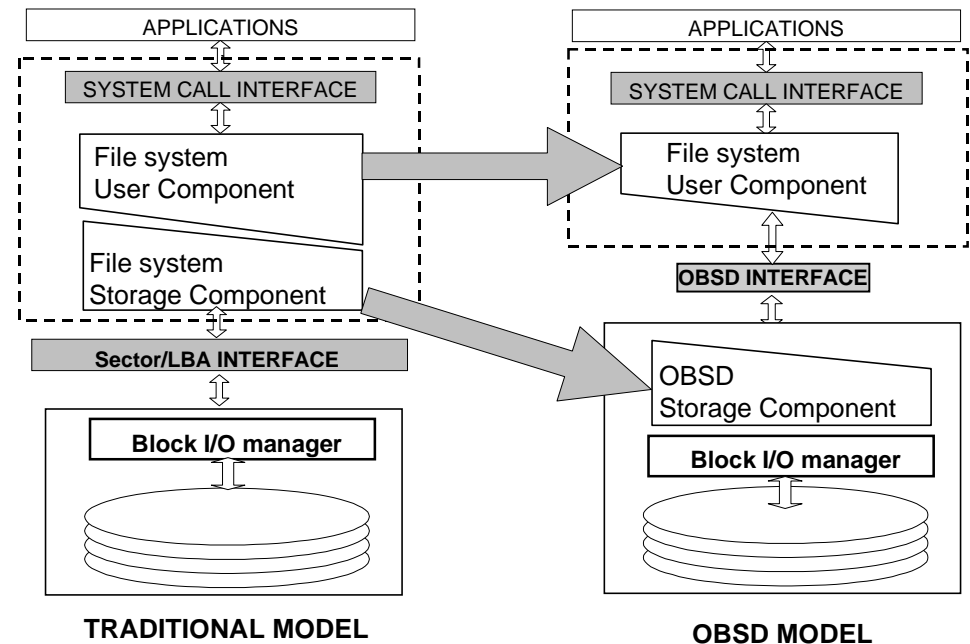OBSD: A new boundary subdividing File System

- Completes device abstraction
- Applicable to Disc, Tape, Array Subsystem, Library, etc

File system - user component

- Naming
- Hierarchy management
- File access control

Objects - storage component

- Space Management
- Infrastructure security
- Foundation for automated management

TRADITIONAL MODEL

APPLICATIONS
SYSTEM CALL INTERFACE
File system User Component
File system Storage Component
Sector/LBA INTERFACE
Block I/O manager

OBSD MODEL

APPLICATIONS
SYSTEM CALL INTERFACE
File system User Component
OBSD INTERFACE
OBSD Storage Component
Block I/O manager

Seagate.
Information the way you want it.

OBSD: Constituents

SAN

- The interconnect or network
- No presumption of which

Requester

- Client or server
- Issuer of OBSD commands

OBSD

- Storage devices - disc tape, array, library, etc

Policy/Storage Manager

- Manages security
- Access to OBSD after permission from Manager
- Not essential, but function must be somewhere for security

Policy/Storage Manager

SAN

Requesters

Object Based Storage

Seagate.
Information the way you want it.
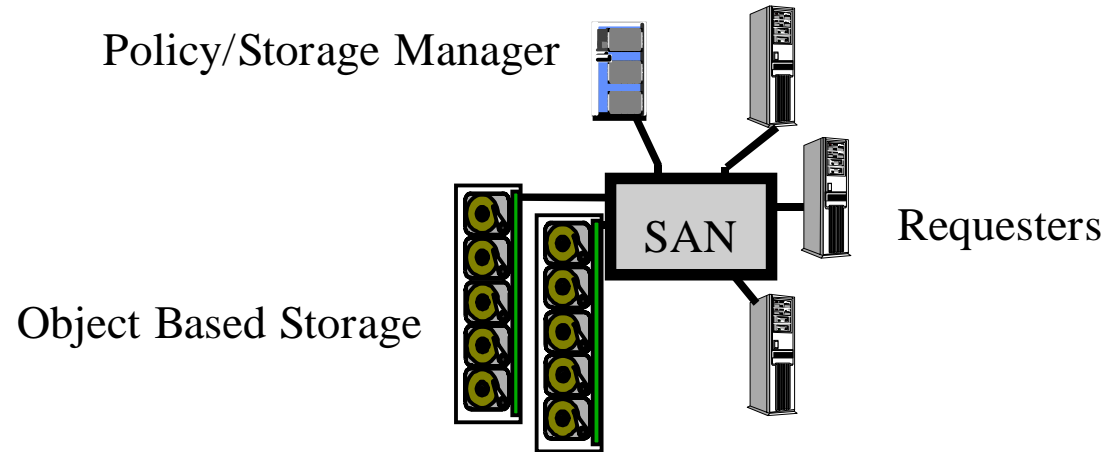
OBSD: Objects

Object

- Sequence of bytes associated with unique identifier
- Located & allocated by storage device
- Only addressing is relative to start of Object

Accessing an Object

- OBSD -
    - Addressed via transport addressing mechanism
    - ** Import Action - OBSD name is a parameter
- Object Group ID
    - 32-bit unsigned integer
    - Identifies subdivision of OBSD
- Object ID
    - 64-bit unsigned integer
    - Some argue for 128-bit
        - Global ID
        - Some OS use Global ID as separate identifier
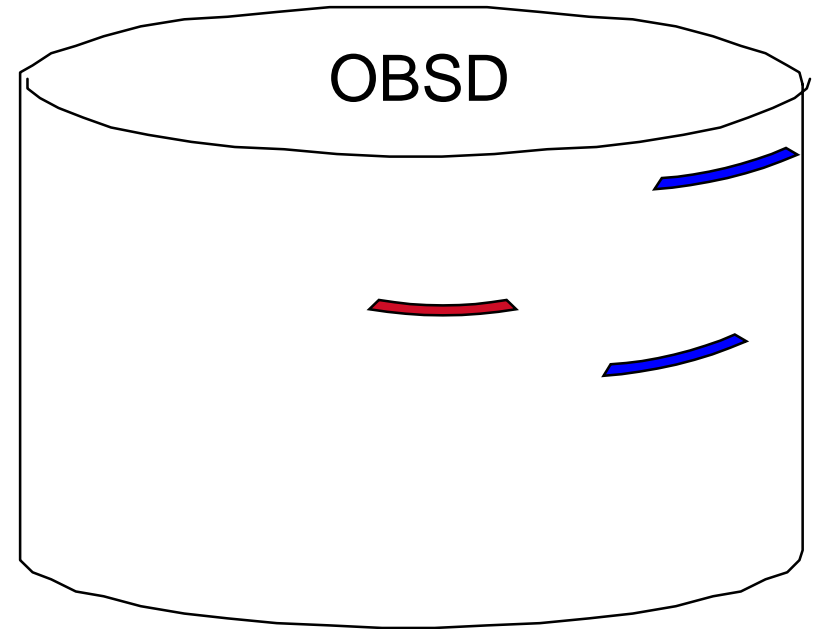
OBSD

Red Object: ID = 627
Blue Object: ID = 54

Seagate.
Information the way you want it.

OBSD: How the Model Works:

Create an OBSD

- Format OBSD - defines OBSD structure on device
- Create Object Group - defines a set in which to create objects

Create an Object

- Create - get Object ID "x" back
- Open "x"
- Write"x",starting byte,length
- Write "x",starting byte,length
- Close "x"

Accessing data

- Get Attribute (Group Control Object) = Root Object ID = "x" = root directory?
- Open Object "x"
- Read Object "x" , starting byte, length
- Read Object "x" , starting byte, length
- Close Object "x"

Seagate.
Information the way you want it.

OBSD: What about Groups?

Group

- Subset of Objects on OBSD
- Any Object in only 1 Group
- Has capacity quota
- Not Partition

Facilitates Object management

- Multiple OS's consuming of space concurrently
- Objects collected by management goal

Attributes, not Group, guide location

- co-locate with others in Group or near any Object
- High data rate - outer zone(s)
- Contiguity

Would really prefer LUN's, I think
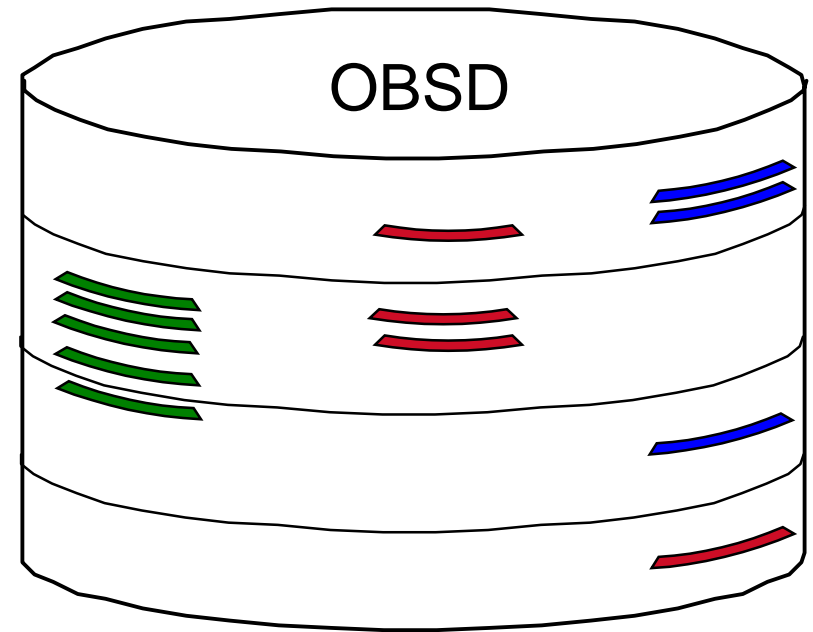
OBSD

Green Group: Co-located
Red Group: 1 Object in Outer zone
Blue Group: 2 Objects in Outer zone

Seagate.
Information the way you want it.

OBSD: What about Sessions?

Session

- I/O with QoS criteria
- Criteria set by attributes
    - Set Attribute action
    - Open
    - Create
- Session ID returned by OBSD
- Session ID used by Requester to mark I/O
    - OBSD
- OBSD rejects OPEN if impossible to satisfy

OPEN Session 1 on Object n
     Read n
     Read n
     Write n
     Read n
OPEN Object n
     Read n
     Read n
     Read n
     Read n
     Read n
CLOSE Session 1 Object n
     Read n
Close Object n

Seagate.
Information the way you want it.

OBSD: What about Priority?

Quantify relative importance of I/O

- 4 bits
- Possible interpretation
    - 13 - 16 reserved
    - 12 - Head of I/O queue
    - 9 - 11 reserved
    - 8 - Normal priority
    - 5 - 7 reserved
    - 4 - Low priority
    - 2 - 3 reserved
    - 1 - Do only after idle for xx seconds
- Ranges allow OS to use several levels of priority

Seagate.
Information the way you want it.

OBSD Actions: Format OBSD

Create OBSD structure on device

- Set up device control object

Options

Transfer

Response

Bit Byte	7	6	5	4	3	2	1	0
8	(MSB)			FORMAT OBSD ACTION CODE				
9								(LSB)
10				OPTION BYTE 1				
11				OPTION BYTE 2				
24	(MSB)							
25				Reserved				
26								
27								(LSB)
28	(MSB)							
29								
30				LENGTH				
31				(OBSD_CAPACITY)				
32								
23								
34								
35								(LSB)

● Figure 1. Format OBSD

Seagate.

Information the way you want it™

OBSD Actions: Create Object

Parameters

- Allocate Object ID
- Length -Preallocate bytes
- Attribute mask

Options

- ATTR: Attributes included
- SESS: Set up a session
- CMPL: Object data with Create

Transfer: attributes, data

Response

- Object ID
- Capacity remaining in Group
- Session ID

Bit Byte	7	6	5	4	3	2	1	0
8	(MSB)			CREATE ACTION CODE				
9								(LSB)
10				OPTION BYTE 1				
11				OPTION BYTE 2				
12	(MSB)							
13				OBJECT GROUP ID				
14								
15								(LSB)
16	(MSB)							
17								
18								
19				STARTING BYTE				
20				ADDRESS				
21								
22								
23								(LSB)
24	(MSB)							
25								
26								
27				LENGTH				
28				(optional)				
29								
30								
31								(LSB)
32	(MSB)							
33								
34				ATTRIBUTE				
35				MASK				
36				(optional)				
37								
38								
39								(LSB)

Figure 1 CREATE Object Action

Seagate.
Information the way you want it.

OBSD Actions: Open

Parameters

- Length -Preallocate bytes
- Attribute mask

Options

- ATTR: Attributes included
- SESS: Set up a session
- SEQ: Sequential access
- WRNLY: Write only
- RDNLY: Read only

Transfer: attributes

Response

- Session ID
- Object Logical Length
- Group remaining capacity

Bit Byte	7	6	5	4	3	2	1	0
8	(MSB)			OPEN ACTION CODE				
9								(LSB)
10				OPTION BYTE 1				
11				OPTION BYTE 2				
12	(MSB)							
13								
14				OBJECT GROUP ID				
15								(LSB)
16	(MSB)							
17				OBJECT ID				
18								
19								
20								
21								
22								
23								(LSB)
24	(MSB)							
25								
26								
27				LENGTH				
28								
29								
30								
31								(LSB)
32	(MSB)							
33								
34								
35				ATTRIBUTE MASK				
36				(optional)				
37								
38								
39								(LSB)
40	(MSB)			SESSION ID				
41				(optional)				
42								
43								(LSB)

Figure 1 OPEN Object Action

OBSD Actions: Read

Parameters

- Starting byte
- Length of transfer
- Session ID

Options

- Priority

Transfer: requested data

Response

- Length of returned data

Bit Byte	7	6	5	4	3	2	1	0
8	(MSB)			READ ACTION CODE				
9								(LSB)
10				OPTION BYTE 1				
11				OPTION BYTE 2				
12	(MSB)							
13								
14				OBJECT GROUP ID				
15								(LSB)
16	(MSB)							
17								
18								
19				OBJECT ID				
20								
21								
22								
23								(LSB)
24	(MSB)							
25								
26								
27				STARTING BYTE				
28				ADDRESS				
29								
30								
31								(LSB)
32	(MSB)							
33								
34								
35				TRANSFER LENGTH				
36								
37								
38								
39								
40	(MSB)							
41				SESSION ID				
42				(optional)				
43								(LSB)

Figure 1 Read Action

OBSD Actions: Write

Parameters

- Starting byte
- Length of transfer
- Session ID

Options

- Priority

Transfer: data

Response

Bit Byte	7	6	5	4	3	2	1	0
8	(MSB)			WRITE ACTION CODE				
9								(LSB)
10				OPTION BYTE 1				
11				OPTION BYTE 2				
12	(MSB)							
13								
14				OBJECT GROUP ID				
15								(LSB)
16	(MSB)							
17								
18								
19								
20				Object ID				
21								
22								
23								(LSB)
24	(MSB)							
25								
26								
27				STARTING BYTE				
28				ADDRESS				
29								
30								
31								(LSB)
32	(MSB)							
33								
34								
35				LENGTH				
36								
37								
38								
39								(LSB)
40	(MSB)							
41				SESSION ID				
42				(optional)				
43								(LSB)

Figure 1 WRITE Object Action

Seagate.
Information the way you want it.

OBSD Actions: Append (could be Write option)

OBSD does concurrency control

Parameters

- Length of transfer

Options

- Priority

Transfer: data

Response

- Starting byte address of transfer
- Group remaining capacity

Example: Log file

Bit Byte	7	6	5	4	3	2	1	0
8	(MSB)			APPEND ACTION CODE				
9								(LSB)
10				OPTION BYTE 1				
11				OPTION BYTE 2				
12	(MSB)							
13								
14				OBJECT GROUP ID				
15								(LSB)
16	(MSB)							
17								
18								
19								
20								
21								
22				OBJECT ID				
23								(LSB)
24	(MSB)							
25								
26								
27				TRANSFER LENGTH				
28								
29								
30								
31								(LSB)
32	(MSB)							
23				SESSION ID				
34				(optional)				
35								(LSB)

- Figure 1 APPEND

| System 1 | System 2 | System 1 | System 1 | System 2 | System 3 | ... |

Seagate

Information the way you want it.

OBSD Actions: Flush (Synchronize)

Parameters

- none - flush OBSD
- Group ID - flush group
- Object ID - flush object

Options

Transfer

Response

Bit Byte	7	6	5	4	3	2	1	0
8	(MSB)			FLUSH OBJECT ACTION CODE				
9								(LSB)
10				OPTION BYTE 1				
11				OPTION BYTE 2				
12	(MSB)							
13								
14				OBJECT GROUP ID				
15								(LSB)
16	(MSB)							
17								
18								
19				OBJECT ID				
20								
21								
22								
23								(LSB)

● Figure 1 FLUSH Object Operation

Seagate.

Information the way you want it.

OBSD Actions: Close

Identifies Object or session as no longer in use

Parameters

- Session ID

Options

Transfer

Response

Bit Byte	7	6	5	4	3	2	1	0
8	(MSB)			CLOSE ACTION CODE				
9								(LSB)
10				OPTION BYTE 1				
11				OPTION BYTE 2				
12	(MSB)							
13								
14				OBJECT GROUP ID				
15								(LSB)
16	(MSB)							
17				OBJECT ID				
18								
19								
20								
21								
22								
23	(MSB)							
24								
25				SESSION ID				
26								
27								(LSB)

Figure 1 CLOSE Object Action

Seagate.
Information the way you want it.™

OBSD Actions: Flush (Synchronize)

Parameters

- none
- Group ID - flush group
- Object ID - flush object

Options

Transfer

Response

Bit Byte	7	6	5	4	3	2	1	0
8	(MSB)		FLUSH OBJECT ACTION CODE					
9								(LSB)
10			OPTION BYTE 1					
11			OPTION BYTE 2					
12	(MSB)							
13								
14			OBJECT GROUP ID					
15								(LSB)
16	(MSB)							
17								
18								
19			OBJECT ID					
20								
21								
22								
23								(LSB)

● Figure 1 FLUSH Object Operation

Seagate.

Information the way you want it.

OBSD Actions: Remove

Parameters

Options

- DESTR - obliterate contents

Transfer

Bit Byte	7	6	5	4	3	2	1	0
8	(MSB)			REMOVE ACTION CODE				
9								(LSB)
10				OPTION BYTE 1				
11				OPTION BYTE 2				
12	(MSB)							
13								
14				OBJECT GROUP ID				
15								(LSB)
16	(MSB)							
17								
18								
19								
20								
21								
22				OBJECT ID				
23								(LSB)

• Figure 1 REMOVE Object Action

Response

- Group remaining capacity

Seagate
Information the way you want it.

OBSD Actions: Create Object Group

Parameters

- Capacity quota

Options

Transfer

Response

Bit Byte	7	6	5	4	3	2	1	0
8	(MSB)		CREATE OBJECT GROUP ACTION CODE					
9								(LSB)
10			OPTION BYTE 1					
11			OPTION BYTE 2					
12	(MSB)							
13								
14								
15								
16			CAPACITY QUOTA					
17								
18								
19								(LSB)

• Figure 1 CREATE OBJECT GROUP Action

Seagate

Information the way you want it.™

OBSD Actions: Remove Object Group

Parameters

Options

Transfer

Response

- Group remaining capacity

Bit Byte	7	6	5	4	3	2	1	0
8	(MSB)		REMOVE OBJECT GROUP ACTION CODE					
9								(LSB)
10				OPTION BYTE 1				
11				OPTION BYTE 2				
12	(MSB)							
13								
14				OBJECT GROUP ID				
15								(LSB)

● Figure 1 REMOVE OBJECT GROUP Action

Seagate

Information the way you want it.

OBSD Actions: Import Object

Parameters

- OBSD source: {OBSD,Group ID, Object ID}
- Attributes

Options

- Priority

Transfer

- Attributes (from Requester)
- Data (from Source OBSD)

Response

- Object ID
- Group remaining capacity

Bit Byte	7	6	5	4	3	2	1	0
8	(MSB)			IMPORT ACTION CODE				
9								(LSB)
10				OPTION BYTE 1				
11				OPTION BYTE 2				
12	(MSB)							
13				DESTINATION				
14				OBJECT GROUP ID				
15								(LSB)
16	(MSB)							
17								
18								
19				SOURCE OBJECT ID				
20								
21								
22								
23								(LSB)
24	(MSB)							
25								
26								
27								
28								
29								
30								
31				SOURCE OBSD				
32								
33								
34								
35								
36								
37								
38								
39								(LSB)
40	(MSB)							
41								
42								
43				ATTRIBUTE MASK				
44								
45								
46								
47								(LSB)
48	(MSB)							
49				SOURCE OBJECT GROUP ID				
50								
51								(LSB)

• Figure 1 Import Object Action

Seagate.
Information the way you want it.

OBSD Actions: Get Attributes

Parameters

- Session ID
- Attribute mask

Options

Transfer

- Attribute mask (from Requester)
- Attributes (from OBSD)

Response

Bit Byte	7	6	5	4	3	2	1	0
8	(MSB)			GET ATTRIBUTE ACTION CODE				
9								(LSB)
10				OPTION BYTE 1				
11				OPTION BYTE 2				
12	(MSB)							
13								
14				OBJECT GROUP ID				
15								(LSB)
16	(MSB)							
17								
18								
19				OBJECT ID				
20								
21								
22								
23								(LSB)
24	(MSB)							
25								
26								
27				ATTRIBUTE MASK				
28								
29								
30								
31								(LSB)
32	(MSB)							
33				SESSION ID				
34				(optional)				
35								(LSB)

- Figure 1 GET ATTRIBUTE Action

Seagate.
Information the way you want it.

OBSD Actions: Set Attributes

Parameters

- Session ID
- Attribute mask

Options

Transfer

- Attribute mask (from Requester)
- Attributes (from Requester)

Response

Bit Byte	7	6	5	4	3	2	1	0
8	(MSB)			SET ATTRIBUTE ACTION CODE				
9								(LSB)
10				OPTION BYTE 1				
11				OPTION BYTE 2				
12	(MSB)							
13								
14				OBJECT GROUP ID				
15								(LSB)
16	(MSB)							
17								
18								
19				OBJECT ID				
20								
21								
22								
23								(LSB)
24	(MSB)							
25				SESSION ID				
26				(optional)				
27								(LSB)
28	(MSB)							
29								
30								
31								
32				ATTRIBUTE MASK				
33								
34								
35								(LSB)

● Figure 1 SET ATTRIBUTE Action

Seagate

Information the way you want it.™

OBSD : Attributes - Object

Type	Name	Set by	Length	Semantics
Clustering	NEARBY_OBJECT	Set Attribute	16	Locate this Object near another
Depending Object	DEPENDING_OBJECT	Set Attribute	16	This Object is dependent on the named obje
Cloning	COPIED_OBJECT	OBSD	8	Object was created Copy Object
Size	OBJECT_LOGICAL_LENGTH OBJECT_SIZE	OBSD, Set Attr. OBSD, Set Attr.	8 8	Largest offset written Number of Bytes Allocated for Object
Access control	ACCESS CONTROL STATE	Set Attr	2 2	Access version Reserved
Time	CREATED_TIME DATA_MODIFIED_TIME DATA_ACCESSED_TIME ATTRIBUTE_MODIFIED_TIME EXPIRATION_TIME_STAMP	OBSD, CREATE OBSD, CLOSE OBSD, OPEN OBSD, Set Attr. CREATE, Set Attr.	8 8 8 8 8	Timestamp of object creation Timestamp of last object data modification Timestamp of last data access Timestamp of last attribute modification Timestamp after which object is not required
Miscellaneous	OBJECT_ATTRIBUTES	OBSD, Set Attr.	8	Bits of Object properties for self-mgmt 00: INDELIBILITY 01: TRANSCIENT_OBJECT
File System	FILE_SYSTEM_ID FS-SPECIFIC	Set Attribute Set Attribute	2 256	Identification of the OS creating the object 256 bytes uninterpreted by OBSD

Seagate.
Information the way you want it.

OBSD : Attributes - Session

Time to initial access (TIA) – The average time delay (in milliseconds) that can be tolerated until the first byte of data from the object is delivered.

Sustained access rate (SAR) – The on-going average data rate (in bytes per second) that data needs to be read from or written to the object.
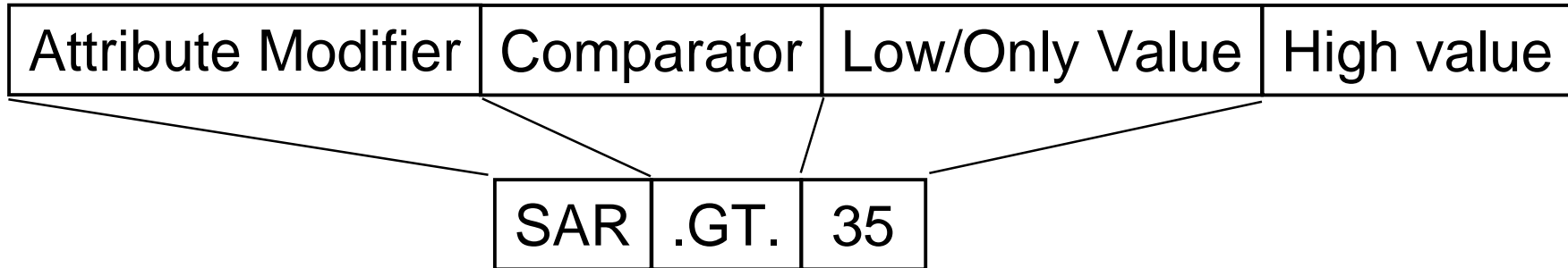
Frequency of access (IOR) – The average number of requests per second that can be expected to read data from or write data into the object.

Bandwidth of access (BOA) – The average request size of data to be read from or written into the object.

Read vs. Write bias – indicates the specification is for read requests or for write requests

Direct vs. Sequential access – indicates the specification is for direct access requests or for sequential requests.

Seagate.
Information the way you want it.

OBSD : Specifying Session Attributes

Attribute Modifier	Comparator	Low/Only Value	High value

SAR	.GT.	35

- **Attribute Identifier -** indicates the particular attribute being set; the identifier enables discriminating among static, dynamic, and extended attributes

- **Comparator** - is one of 'value', 'less than', 'greater than', or 'inclusive'. *value* indicates a specific desired amount (specified by **Low/Only Value** below) is given. *less than* and *greater than* indicate that the **Low/Only Value** is to be viewed as the <u>maximum</u> or <u>minimum</u> values (respectively) for the attribute. *inclusive* indicates that the attribute is to lie within the range specified by **Low/Only Value** and **High Value** (see below). (Needs encoding)

- **Low/Only Value –** an integer representing the bottom of a range, if one is indicated by the comparator; otherwise, it is the unique value

- **High Value –** an integer representing the top of a range, if one is indicated by the comparator; it is absent, otherwise.

  - It is anticipated that Low/Only Value and High Value may be other than integers (e.g. character strings) in the future.

Seagate.
Information the way you want it.

OBSD : Communicating Attributes

Attributes = mode pages for Objects, Groups, OBSD

- Usually bytes, not bits, in length
- Significant variation in length: 1 bit - 256 bytes

Need your help

- Ability to change any attribute independently
- Ability to transfer (approx.) only what is needed
- Predictable length required
- Extensible method

Is this a good way to do it?

- Attributes grouped by table
- 64-bit mask defined in Actions
- 1-bit per table
- Table starts with its bit mask
- Fixed length tables

| 00100......00000 | 64-bit Attribute mask |

| 0000 0101 | Attribute Table 2 - mask |

- Attribute Table 2 - Field 1
- Attribute Table 2 - Field 2
- Field 3
- Field 4
- Attribute Table 2 - Field 5
- Field 6
- Field 7
- Field 8

Seagate.
Information the way you want it.