To FUSE or Not to FUSE: Performance of User-Space File Systems

Bharath Kumar Reddy Vangoor, Vasily Tarasov,[×] and Erez Zadok.

Stony Brook University, [×]IBM Research – Almaden

http://filesystems.org/fuse/

Stony Brook University IBM Research

Introduction and Motivation

Stony Brook University IBM Research

Introduction and Motivation

- What is FUSE?
- Why FUSE?
- Performance vs. ease of development
- No previous study of FUSE
- Little documentation for FUSE
- File Systems (Industry & Academia)
  - Using FUSE: LTFS, PLFS, S3QL, SSHFS, …
  - User-level, **not** using FUSE: GPFS, HDFS, CASL

Stony Brook University IBM Research

FUSE High Level Architecture

Stony Brook University IBM Research

FUSE High Level Architecture

User

Kernel

Stony Brook University IBM Research

FUSE High Level Architecture

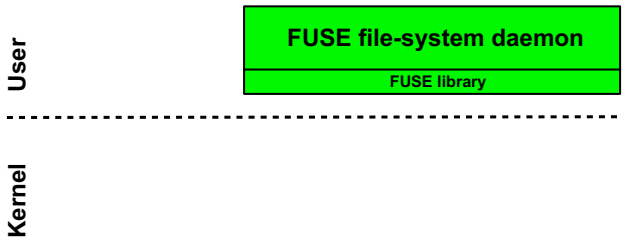
User

FUSE library

Kernel

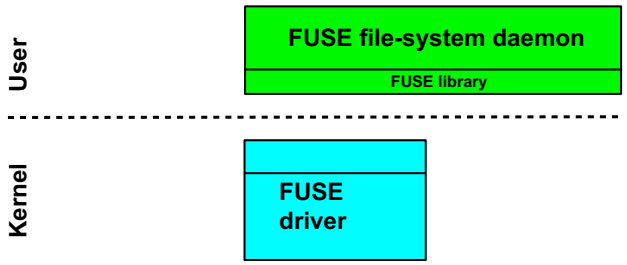
Stony Brook University IBM Research

FUSE High Level Architecture



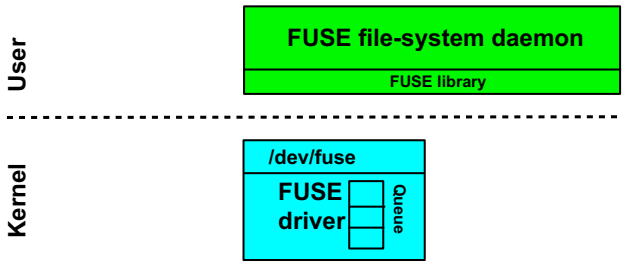
Stony Brook University IBM Research

FUSE High Level Architecture



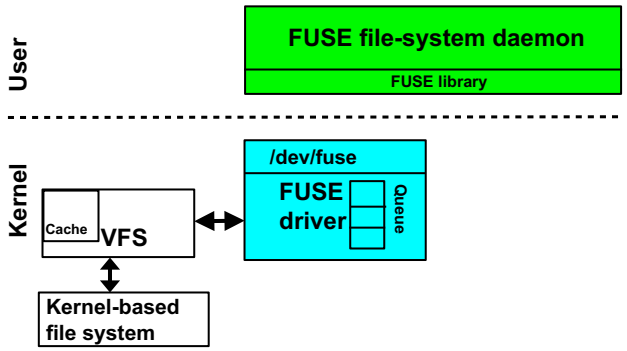
Stony Brook University IBM Research

FUSE High Level Architecture



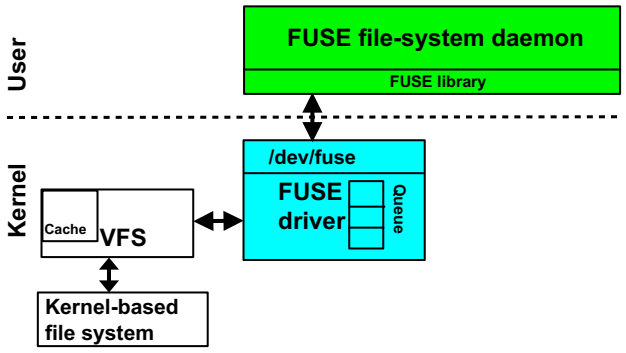
Stony Brook University IBM Research

FUSE High Level Architecture



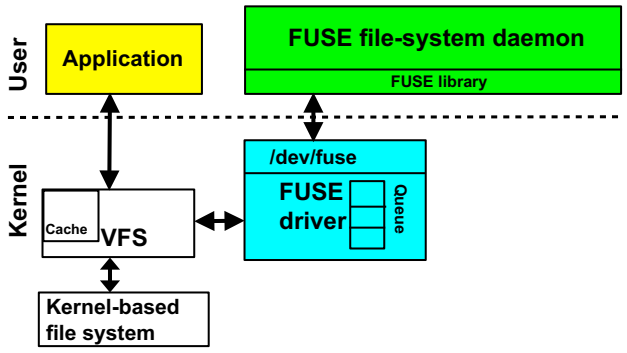
Stony Brook University IBM Research

FUSE High Level Architecture



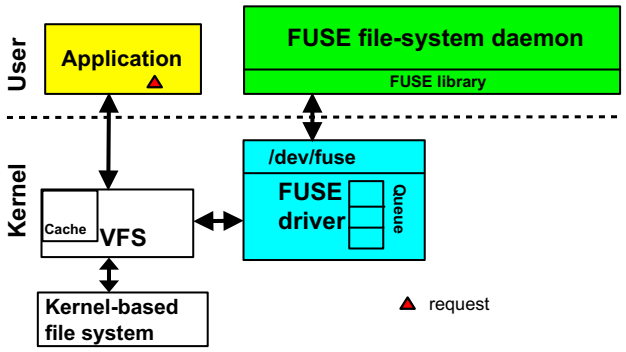
Stony Brook University IBM Research

FUSE High Level Architecture

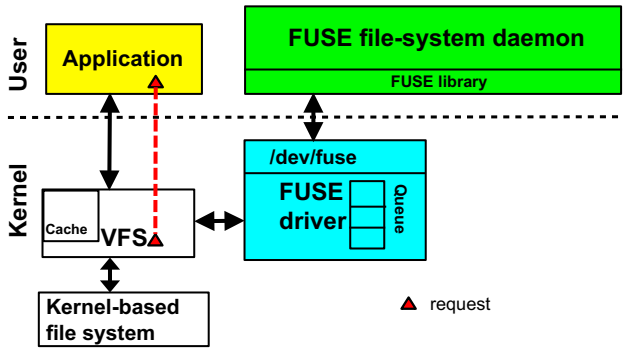


Stony Brook University IBM Research

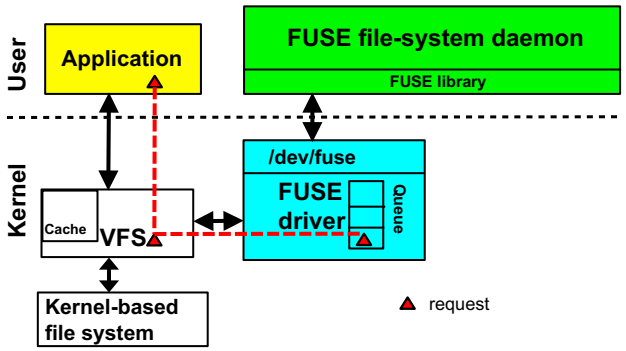
FUSE High Level Architecture



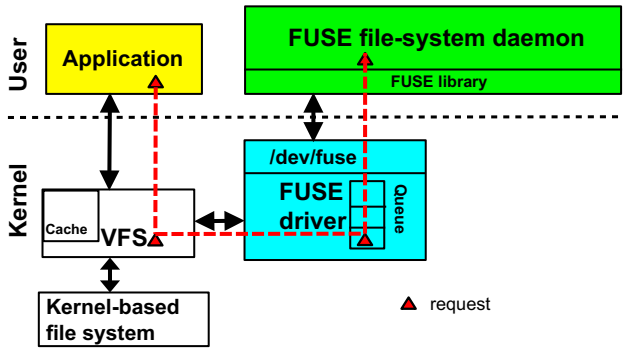
FUSE High Level Architecture



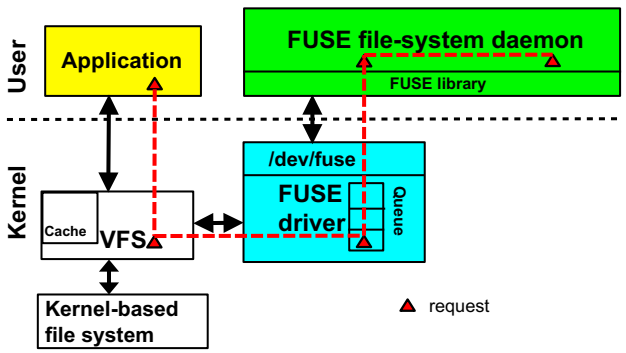
FUSE High Level Architecture



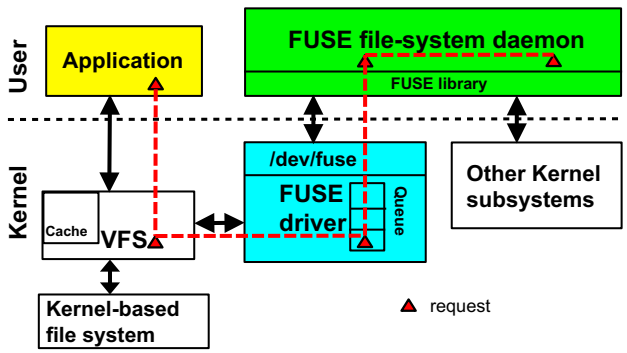
FUSE High Level Architecture



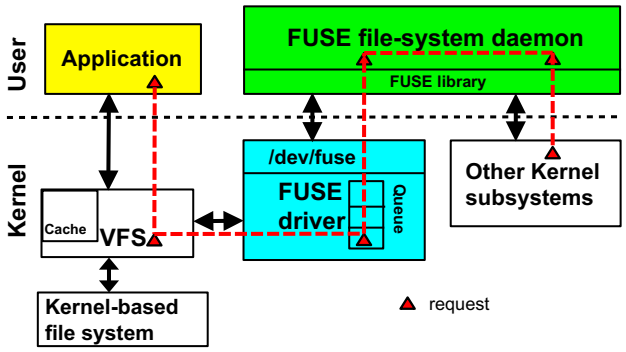
FUSE High Level Architecture



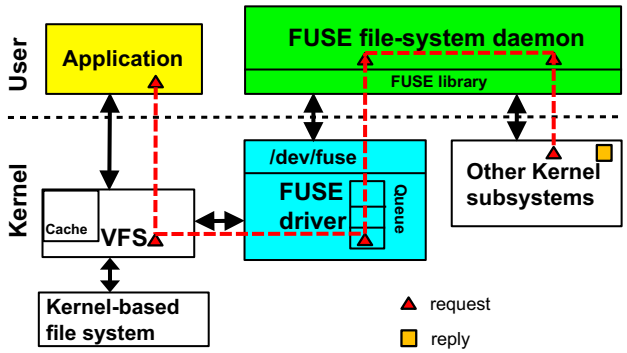
FUSE High Level Architecture



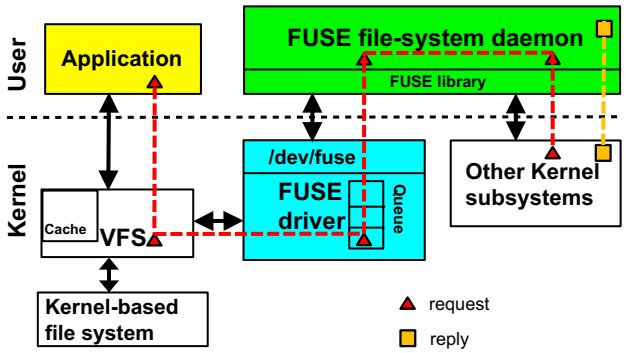
FUSE High Level Architecture



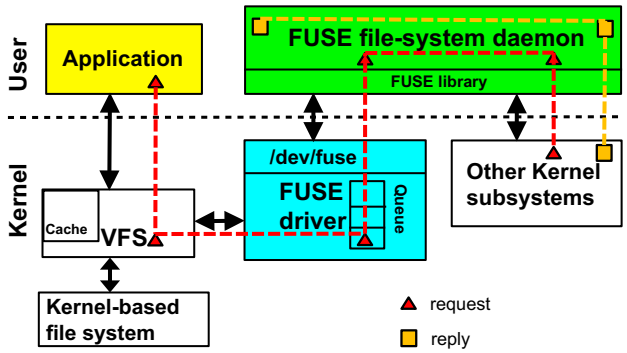
FUSE High Level Architecture



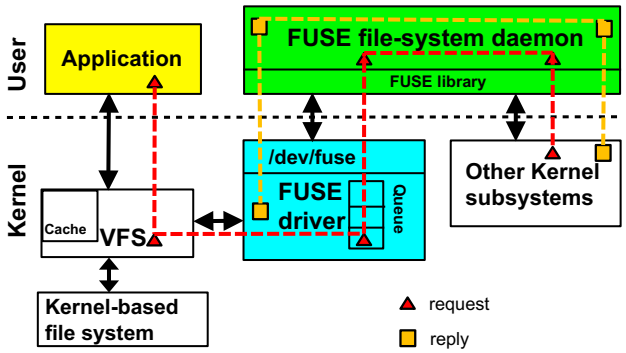
FUSE High Level Architecture



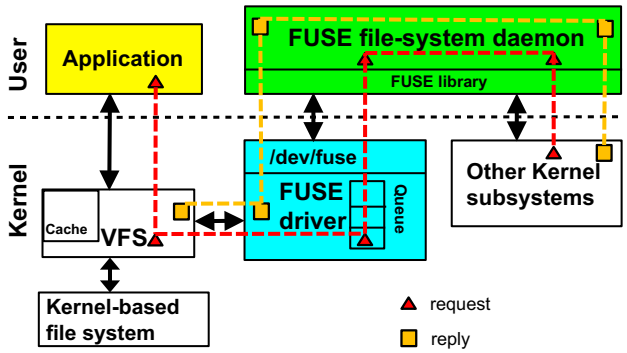
FUSE High Level Architecture



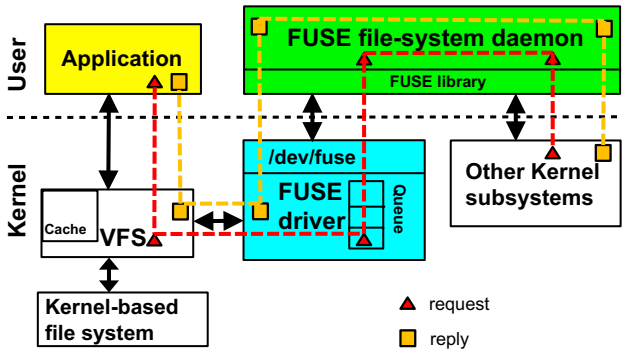
FUSE High Level Architecture



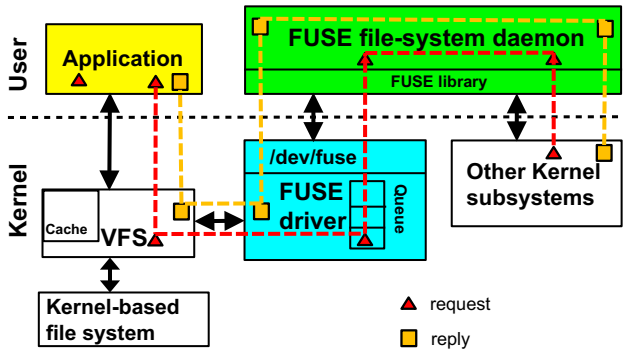
FUSE High Level Architecture



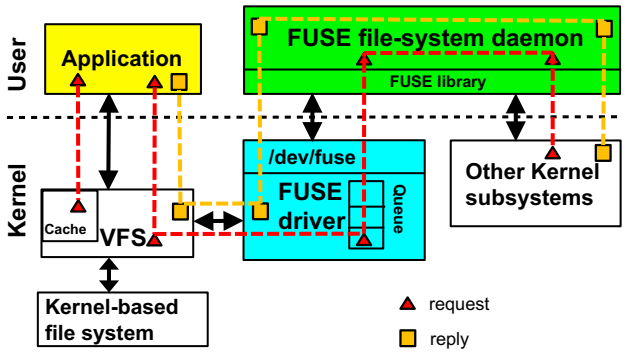
FUSE High Level Architecture



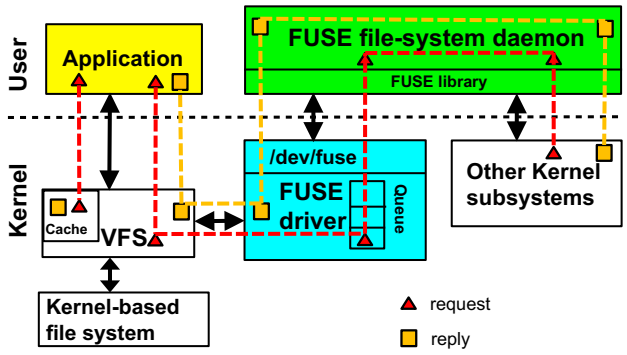
FUSE High Level Architecture



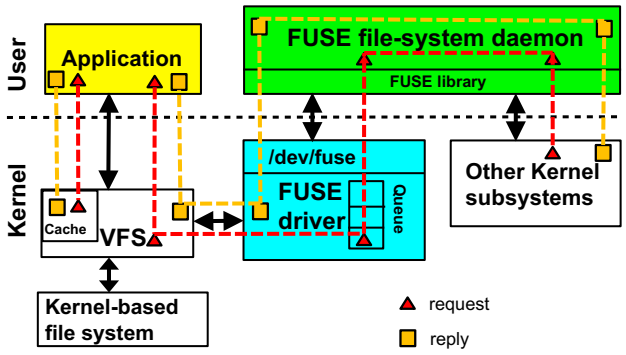
FUSE High Level Architecture



FUSE High Level Architecture



FUSE High Level Architecture



FUSE Request Types

FUSE Request Types

Group	Request Types
Special (3)	INIT, DESTROY, INTERRUPT
Metadata (14)	**LOOKUP**, **FORGET**, BATCH_FORGET, CREATE, UNLINK, LINK, RENAME, RENAME2, **OPEN**, **RELEASE**, STATFS, FSYNC, FLUSH, ACCESS
Data (2)	**READ**, **WRITE**
Attributes (2)	GETATTR, SETATTR
Extended Attributes (4)	SETXATTR, GETXATTR, LISTXATTR, REMOVEXATTR
Symlinks (2)	SYMLINK, READLINK
Directory (7)	MKDIR, RMDIR, OPENDIR, RELEASEDIR, READDIR, READDIRPLUS, FSYNCDIR
Locking (3)	GETLK, SETLK, SETLKW
Misc (6)	BMAP, FALLOCATE, MKNOD, IOCTL, POLL, NOTIFY_REPLY

FUSE Request Types

Group	Request Types
Special (3)	INIT, DESTROY, INTERRUPT
Metadata (14)	**LOOKUP** **FORGET**, BATCH_FORGET, CREATE, UNLINK, LINK, RENAME, RENAME2, **OPEN**, **RELEASE**, STATFS, FSYNC, FLUSH, ACCESS
Data (2)	**READ**, **WRITE**
Attributes (2)	GETATTR, SETATTR
Extended Attributes (4)	SETXATTR, GETXATTR, LISTXATTR, REMOVEXATTR
Symlinks (2)	SYMLINK, READLINK
Directory (7)	MKDIR, RMDIR, OPENDIR, RELEASEDIR, READDIR, READDIRPLUS, FSYNCDIR
Locking (3)	GETLK, SETLK, SETLKW
Misc (6)	BMAP, FALLOCATE, MKNOD, IOCTL, POLL, NOTIFY_REPLY

FUSE Request Types

Group	Request Types
Special (3)	INIT, DESTROY, INTERRUPT
Metadata (14)	**LOOKUP** **FORGET**, BATCH_FORGET, CREATE, UNLINK, LINK, RENAME, RENAME2, **OPEN**, **RELEASE**, STATFS, FSYNC, FLUSH, ACCESS
Data (2)	**READ**, **WRITE**
Attributes (2)	GETATTR, SETATTR
Extended Attributes (4)	SETXATTR, GETXATTR, LISTXATTR, REMOVEXATTR
Symlinks (2)	SYMLINK, READLINK
Directory (7)	MKDIR, RMDIR, OPENDIR, RELEASEDIR, READDIR, READDIRPLUS, FSYNCDIR
Locking (3)	GETLK, SETLK, SETLKW
Misc (6)	BMAP, FALLOCATE, MKNOD, IOCTL, POLL, NOTIFY_REPLY

FUSE Request Types

Group	Request Types
Special (3)	INIT, DESTROY, INTERRUPT
Metadata (14)	**LOOKUP** **FORGET**, BATCH_FORGET, CREATE, UNLINK, LINK, RENAME, RENAME2, **OPEN** **RELEASE**, STATFS, FSYNC, FLUSH, ACCESS
Data (2)	**READ**, **WRITE**
Attributes (2)	GETATTR, SETATTR
Extended Attributes (4)	SETXATTR, GETXATTR, LISTXATTR, REMOVEXATTR
Symlinks (2)	SYMLINK, READLINK
Directory (7)	MKDIR, RMDIR, OPENDIR, RELEASEDIR, READDIR, READDIRPLUS, FSYNCDIR
Locking (3)	GETLK, SETLK, SETLKW
Misc (6)	BMAP, FALLOCATE, MKNOD, IOCTL, POLL, NOTIFY_REPLY

FUSE Request Types

Group	Request Types
Special (3)	INIT, DESTROY, INTERRUPT
Metadata (14)	**LOOKUP** **FORGET**, BATCH_FORGET, CREATE, UNLINK, LINK, RENAME, RENAME2, **OPEN** **RELEASE**, STATFS, FSYNC, FLUSH, ACCESS
Data (2)	**READ**, **WRITE**
Attributes (2)	GETATTR, SETATTR
Extended Attributes (4)	SETXATTR, GETXATTR, LISTXATTR, REMOVEXATTR
Symlinks (2)	SYMLINK, READLINK
Directory (7)	MKDIR, RMDIR, OPENDIR, RELEASEDIR, READDIR, READDIRPLUS, FSYNCDIR
Locking (3)	GETLK, SETLK, SETLKW
Misc (6)	BMAP, FALLOCATE, MKNOD, IOCTL, POLL, NOTIFY_REPLY

FUSE Request Types

Group	Request Types
Special (3)	INIT, DESTROY, INTERRUPT
Metadata (14)	**LOOKUP** **FORGET**, BATCH_FORGET, CREATE, UNLINK, LINK, RENAME, RENAME2, **OPEN** **RELEASE**, STATFS, FSYNC, FLUSH, ACCESS
Data (2)	**READ**, **WRITE**
Attributes (2)	GETATTR, SETATTR
Extended Attributes (4)	SETXATTR, GETXATTR, LISTXATTR, REMOVEXATTR
Symlinks (2)	SYMLINK, READLINK
Directory (7)	MKDIR, RMDIR, OPENDIR, RELEASEDIR, READDIR, READDIRPLUS, FSYNCDIR
Locking (3)	GETLK, SETLK, SETLKW
Misc (6)	BMAP, FALLOCATE, MKNOD, IOCTL, POLL, NOTIFY_REPLY

FUSE Request Types

Group	Request Types
Special (3)	INIT, DESTROY, INTERRUPT
Metadata (14)	**LOOKUP**, **FORGET**, BATCH_FORGET, CREATE, UNLINK, LINK, RENAME, RENAME2, **OPEN**, **RELEASE**, STATFS, FSYNC, FLUSH, ACCESS
Data (2)	**READ**, **WRITE**
Attributes (2)	GETATTR, SETATTR
Extended Attributes (4)	SETXATTR, GETXATTR, LISTXATTR, REMOVEXATTR
Symlinks (2)	SYMLINK, READLINK
Directory (7)	MKDIR, RMDIR, OPENDIR, RELEASEDIR, READDIR, READDIRPLUS, FSYNCDIR
Locking (3)	GETLK, SETLK, SETLKW
Misc (6)	BMAP, FALLOCATE, MKNOD, IOCTL, POLL, NOTIFY_REPLY

FUSE Queues

FUSE Queues

USER

KERNEL

FUSE Queues

USER

FUSE Daemon

KERNEL

FUSE Queues

USER

FUSE Daemon

KERNEL

FUSE Kernel Driver

FUSE Queues

USER

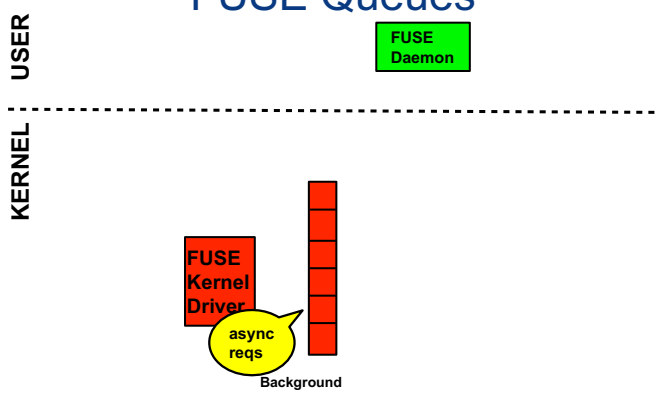
FUSE Daemon

KERNEL

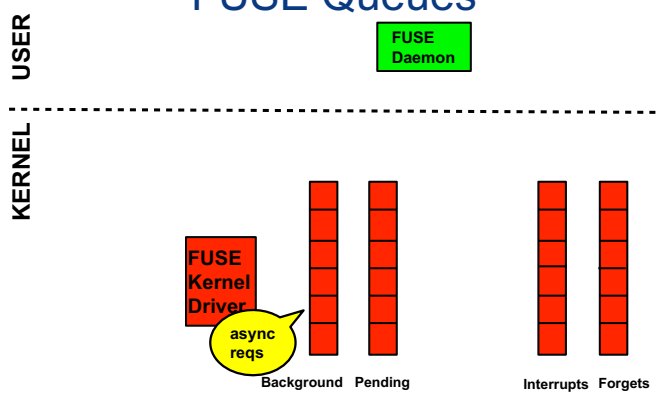
FUSE Kernel Driver

Background

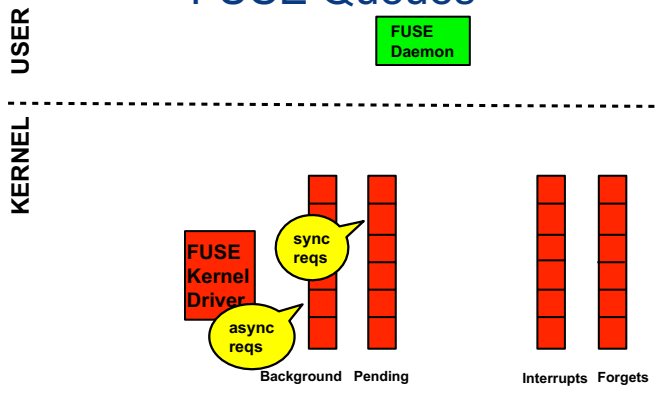
FUSE Queues



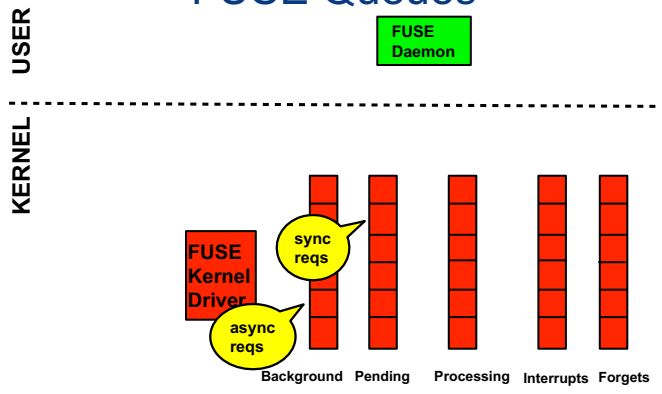
FUSE Queues



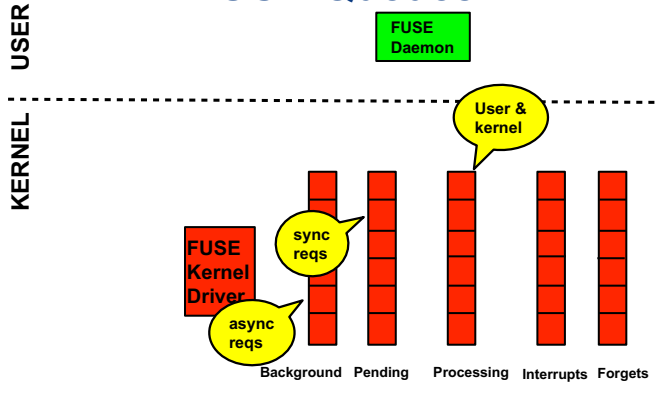
FUSE Queues



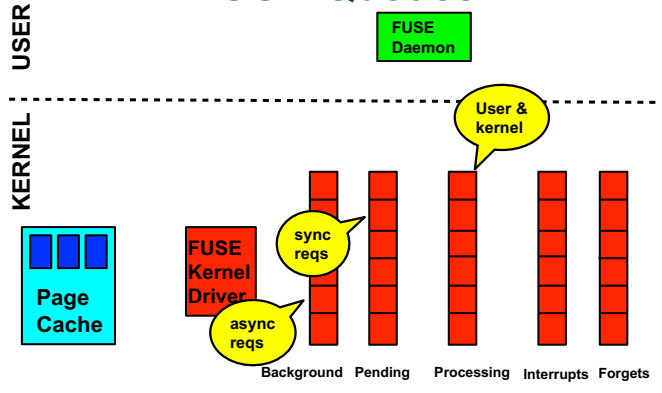
FUSE Queues



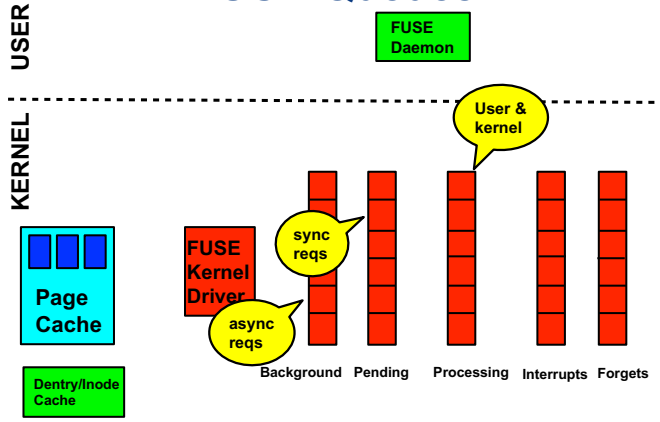
FUSE Queues



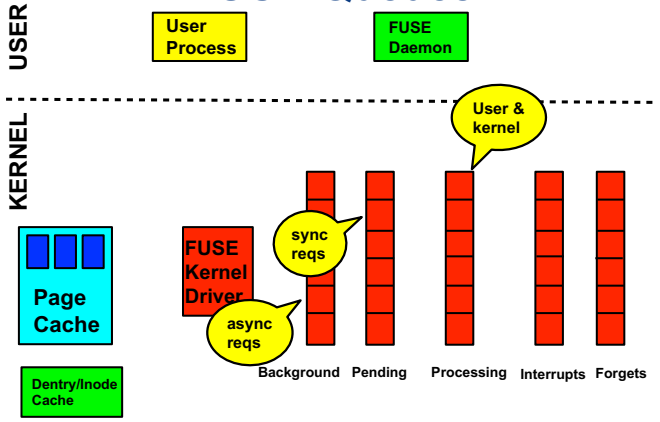
FUSE Queues



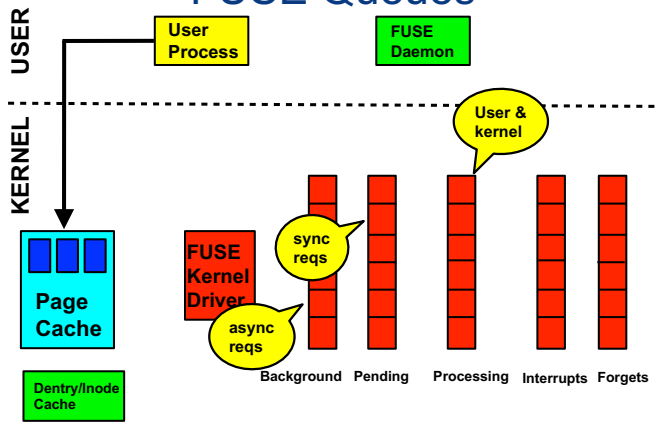
FUSE Queues



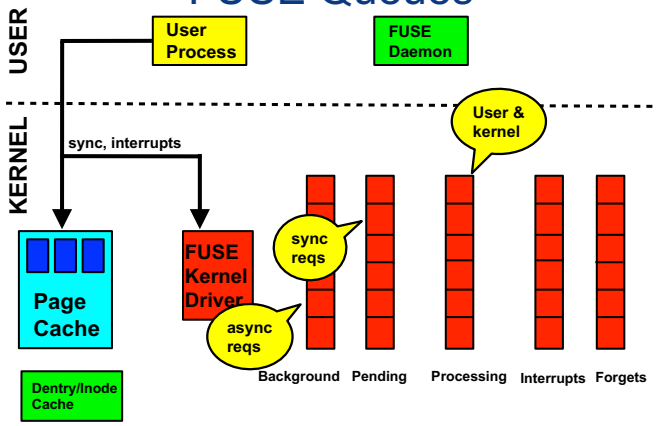
FUSE Queues



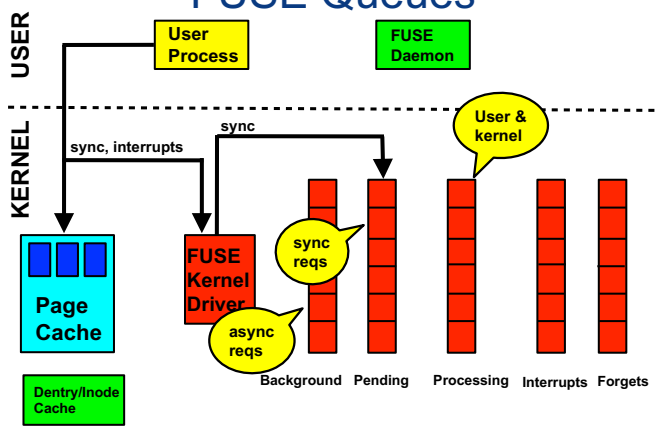
FUSE Queues



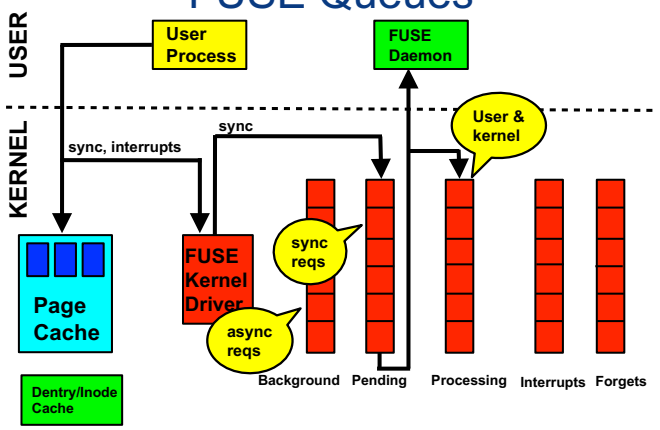
FUSE Queues



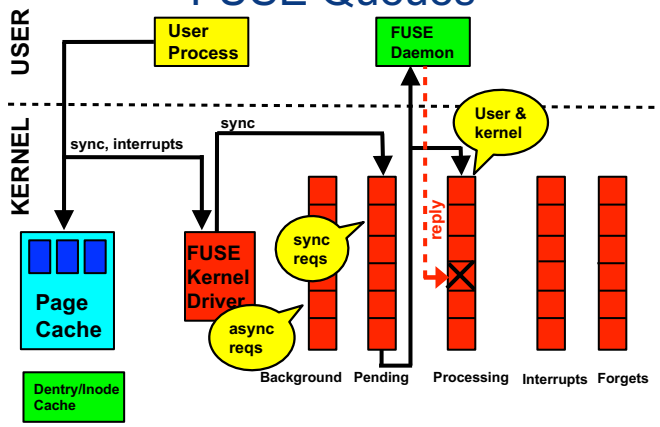
FUSE Queues



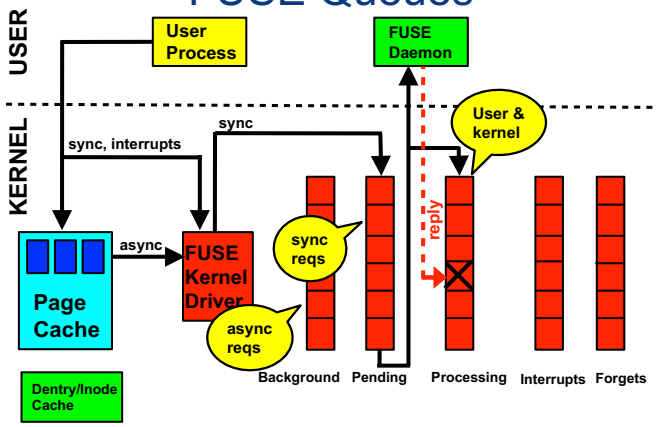
FUSE Queues



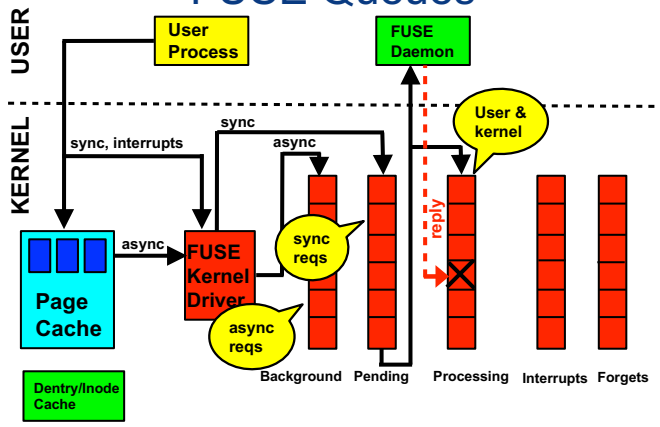
FUSE Queues



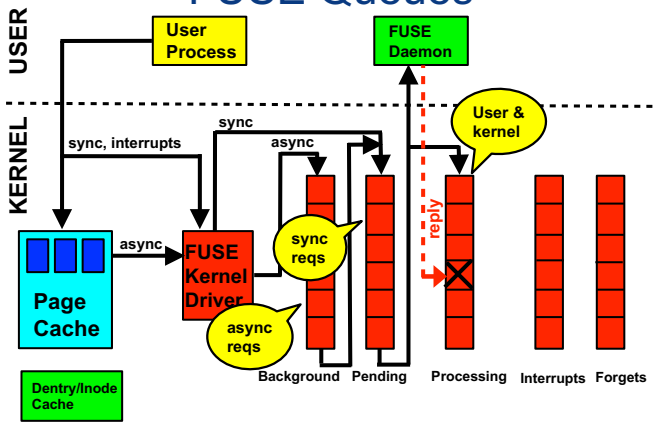
FUSE Queues



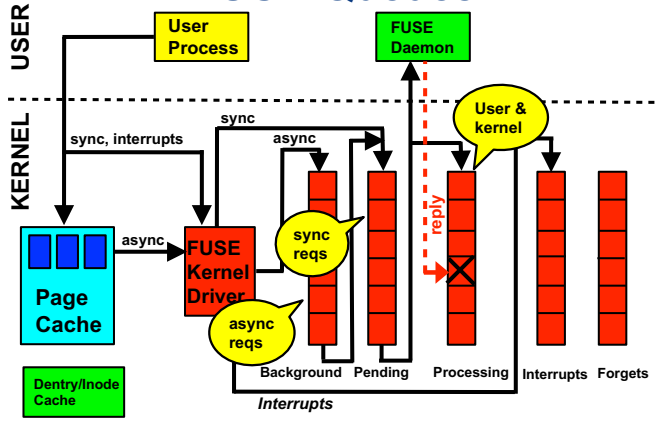
FUSE Queues



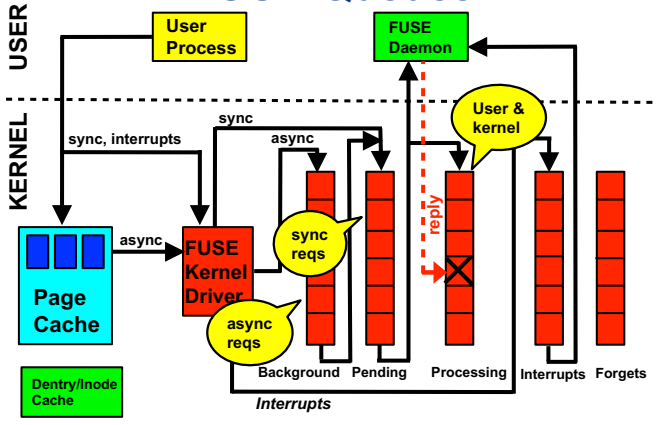
FUSE Queues



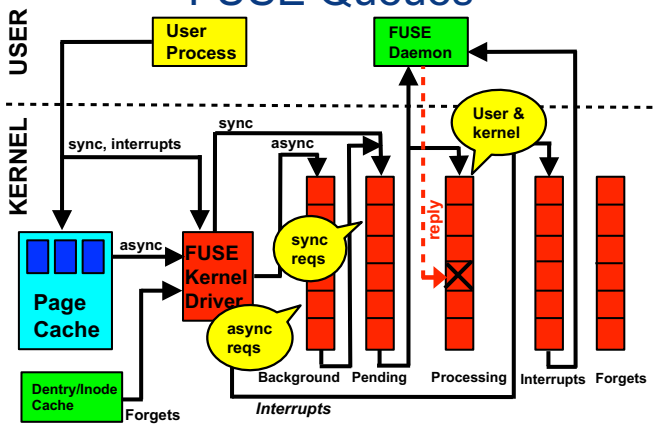
FUSE Queues



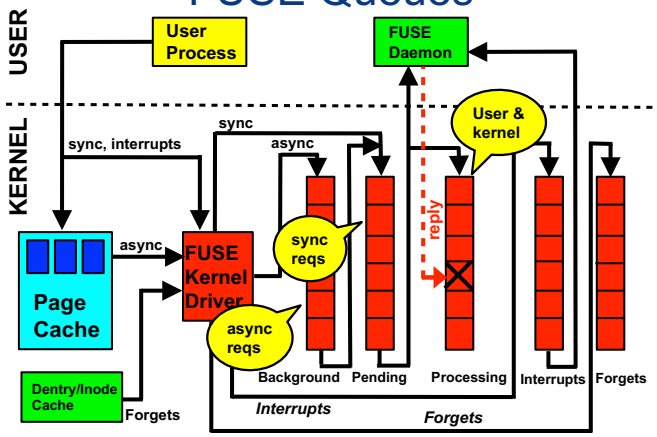
FUSE Queues



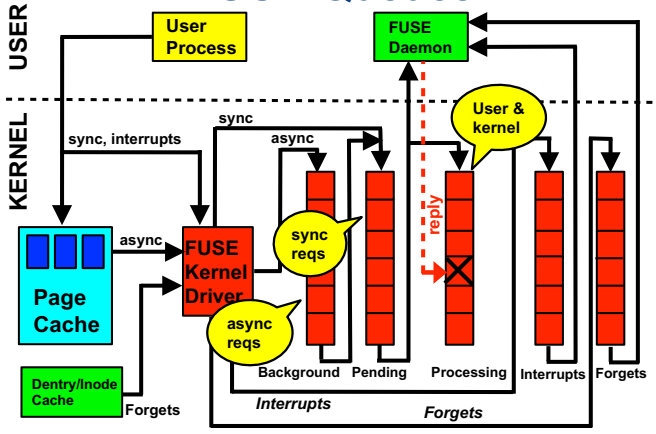
FUSE Queues



FUSE Queues



FUSE Queues



FUSE Optimizations

- Write-back cache
- Extended request size
- Zero copy (using splice)
- Multi-threaded daemon

FUSE Optimizations

☞ Write-back cache
- Extended request size
- Zero copy (using splice)
- Multi-threaded daemon

FUSE Optimizations

- Write-back cache

☞ Extended request size
- Zero copy (using splice)
- Multi-threaded daemon

FUSE Optimizations

- Write-back cache
- Extended request size
- Zero copy (using splice)
- Multi-threaded daemon

FUSE Optimizations

- Write-back cache
- Extended request size
- Zero copy (using splice)
- Multi-threaded daemon

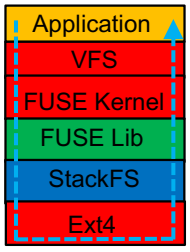
Methodology & Instrumentation

Methodology & Instrumentation

- StackFS
  - Pass-through F/S
- Performance Statistics
  - Kernel statistics
  - Library statistics
  - Tracking timing in queues
- Traces at each stage of request processing

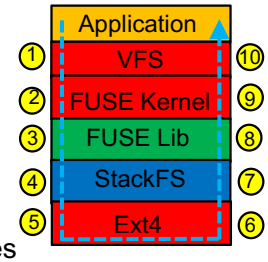
Methodology & Instrumentation

- StackFS
  - Pass-through F/S
- Performance Statistics
  - Kernel statistics
  - Library statistics
  - Tracking timing in queues
- Traces at each stage of request processing



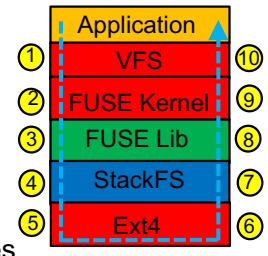
Methodology & Instrumentation

- StackFS
  - Pass-through F/S
- Performance Statistics
  - Kernel statistics
  - Library statistics
  - Tracking timing in queues
- Traces at each stage of request processing



Methodology & Instrumentation

- StackFS
  - Pass-through F/S
- Performance Statistics
  - Kernel statistics
  - Library statistics
  - Tracking timing in queues
- Traces at each stage of request processing



http://filesystems.org/fuse/

Experimental Setup and Workloads

Experimental Setup and Workloads

- Three servers
  - 4 Core 2.40 GHz CPU, 4GB RAM
- Storage
  - HDD and SSD
- Software
  - Kernel: 4.1.13, FUSE lib: commit #386b1b
- Many diverse workloads
  - Random/Sequential: reads/writes
  - Metadata and Macro workloads

Evaluation

Evaluation

- Throughput (ops/sec)
- Four classes
  - Green (< 5%)
  - Yellow (5–25%)
  - Orange (25–50%)
  - Red (> 50%)

Evaluation

- Throughput (ops/sec)
- Four classes
  - Green (< 5%)
  - Yellow (5–25%)
  - Orange (25–50%)
  - Red (> 50%)

Workload	I/O Size	HDD Base	HDD Optimized	SSD Base	SSD Optimized
Random Read 32 threads 1 file	4 KB	-60.4 %	-23.2 %	-82.5 %	-27.6 %
	32 KB	-56.2 %	-17.2 %	-55.7 %	-1.9 %
	128 KB	-34.4 %	-11.4 %	-29.1 %	-2.6 %
	1 MB	-37.0 %	-15.0 %	-12.2 %	-1.9 %

Evaluation

- Throughput (ops/sec)
- Four classes
  - Green (< 5%)
  - Yellow (5–25%)
  - Orange (25–50%)
  - Red (> 50%)

Base configuration of FUSE vs. Ext4

Workload	I/O Size	HDD Base	HDD Optimized	SSD Base	SSD Optimized
Random Read 32 threads 1 file	4 KB	-60.4 %	-23.2 %	-82.5 %	-27.6 %
	32 KB	-56.2 %	-17.2 %	-55.7 %	-1.9 %
	128 KB	-34.4 %	-11.4 %	-29.1 %	-2.6 %
	1 MB	-37.0 %	-15.0 %	-12.2 %	-1.9 %

Evaluation

- Throughput (ops/sec)
- Four classes
  - Green (< 5%)
  - Yellow (5–25%)
  - Orange (25–50%)
  - Red (> 50%)

Optimized configuration of FUSE vs. Ext4

Workload	I/O Size	HDD Base	HDD Optimized	SSD Base	SSD Optimized
Random Read 32 threads 1 file	4 KB	-60.4 %	-23.2 %	-82.5 %	-27.6 %
	32 KB	-56.2 %	-17.2 %	-55.7 %	-1.9 %
	128 KB	-34.4 %	-11.4 %	-29.1 %	-2.6 %
	1 MB	-37.0 %	-15.0 %	-12.2 %	-1.9 %

Performance Overview (1)

Performance Overview (1)

- ***Observation 1***: Relative difference varied from -83.0% to +6.2%.

Performance Overview (1)

- ***Observation 1***: Relative difference varied from -83.0% to +6.2%.

Workload	I/O Size	HDD Base	SSD Optimized
Files create 1 thread	4 KB		-83.3 %
Web server		+6.2 %	

Performance Overview (2)

Performance Overview (2)

- ***Observation 2***: FUSE optimizations improve performance. (45%)
- ***Observation 3***: But there are some exceptions.

Performance Overview (2)

- ***Observation 2***: FUSE optimizations improve performance. (45%)

Workload	HDD Base	HDD Optimized
Web server	-51.8 %	+6.2 %

- ***Observation 3***: But there are some exceptions.

Performance Overview (2)

- ***Observation 2***: FUSE optimizations improve performance. (45%)

Workload	HDD Base	HDD Optimized
Web server	-51.8 %	+6.2 %

- ***Observation 3***: But there are some exceptions.

Workload	SDD Base	SDD Optimized
Files read 1 thread	-25.0 %	-60.3 %

Performance Overview (3)

Performance Overview (3)

- ***Observation 4***: File creates performance degradation <u>increased</u> after optimizations.

Performance Overview (3)

- ***Observation 4***: File creates performance degradation <u>increased</u> after optimizations.

Workload	HDD Base	HDD Optimized	SSD Base	SSD Optimized
Files create 1 thread	-57.0 %	-81.0 %	-62.2 %	-83.3 %
Files create 32 threads	-50.2 %	-54.9 %	-57.6 %	-62.6 %

Read Workloads

Read Workloads

#	Workload	I/O Size (KB)	HDD Results		SSD Results	
			Base (% Diff)	Optimized (% Diff)	Base (% Diff)	Optimized (% Diff)
1	seq-rd-1th-1f	4	- 2.5	+ 1.7	- 0.5	- 0.9
2		32	- 0.2	- 2.2	+ 0.8	+ 0.3
3		128	- 0.9	- 2.1	+ 0.4	+ 1.7
4		1024	- 0.9	- 2.2	+ 0.2	- 0.3
5	seq-rd-32th-32f	4	- 36.9	- 26.9	- 0.1	- 0.2
6		32	- 41.5	- 30.3	- 0.1	- 1.8
7		128	- 41.3	- 29.8	- 0.1	- 0.2
8		1024	- 41.0	- 28.3	- 0.0	- 2.1
9	seq-rd-32th-1f	4	- 2.4	- 3.0	+ 0.0	+ 2.1
10		32	- 2.4	- 4.1	+ 0.7	+ 2.2
11		128	- 2.6	- 4.4	+ 1.5	+ 2.0
12		1024	- 2.5	- 4.0	- 0.1	- 0.4
13	rnd-rd-1th-1f	4	- 10.0	- 10.0	- 32.1	- 39.8
14		32	- 7.4	- 7.5	- 18.8	- 25.2
15		128	- 7.4	- 5.5	- 14.7	- 12.4
16		1024	- 9.0	- 3.1	- 15.3	- 1.5
17	rnd-rd-32th-1f	4	- 60.4	- 23.2	- 82.5	- 27.6
18		32	- 56.2	- 17.2	- 55.7	- 1.9
19		128	- 34.4	- 11.4	- 29.1	- 2.6
20		1024	- 37.0	- 15.0	- 12.2	- 1.9

Read Workloads

#	Workload	I/O Size (KB)	HDD Results		SSD Results	
			Base (% Diff)	Optimized (% Diff)	Base (% Diff)	Optimized (% Diff)
1	seq-rd-1th-1f	4	- 2.5	+ 1.7	- 0.5	- 0.9
2		32	- 0.2	- 2.2	+ 0.8	+ 0.3
3		128	- 0.9	- 2.1	+ 0.4	+ 1.7
4		1024	- 0.9	- 2.2	+ 0.2	- 0.3
5	seq-rd-32th-32f	4	- 36.9	- 26.9	- 0.1	- 0.2
6		32	- 41.5	- 30.3	- 0.1	- 1.8
7		128	- 41.3	- 29.8	- 0.1	- 0.2
8		1024	- 41.0	- 28.3	- 0.0	- 2.1
9	seq-rd-32th-1f	4	- 2.4	- 3.0	+ 0.0	+ 2.1
10		32	- 2.4	- 4.1	+ 0.7	+ 2.2
11		128	- 2.6	- 4.4	+ 1.5	+ 2.0
12		1024	- 2.5	- 4.0	- 0.1	- 0.4
13	rnd-rd-1th-1f	4	- 10.0	- 10.0	- 32.1	- 39.8
14		32	- 7.4	- 7.5	- 18.8	- 25.2
15		128	- 7.4	- 5.5	- 14.7	- 12.4
16		1024	- 9.0	- 3.1	- 15.3	- 1.5
17	rnd-rd-32th-1f	4	- 60.4	- 23.2	- 82.5	- 27.6
18		32	- 56.2	- 17.2	- 55.7	- 1.9
19		128	- 34.4	- 11.4	- 29.1	- 2.6
20		1024	- 37.0	- 15.0	- 12.2	- 1.9

Read-ahead helped

Read Workloads

#	Workload	I/O Size (KB)	HDD Results		SSD Results	
			Base (% Diff)	Optimized (% Diff)	Base (% Diff)	Optimized (% Diff)
1	seq-rd-1th-1f	4	- 2.5	+ 1.7	- 0.5	- 0.9
2		32	- 0.2	- 2.2	+ 0.8	+ 0.3
3		128	- 0.9	- 2.1	+ 0.4	+ 1.7
4		1024	- 0.9	- 2.2	+ 0.2	- 0.3
5	seq-rd-32th-32f	4	- 36.9	- 26.9	- 0.1	- 0.2
6		32	- 41.5	- 30.3	- 0.1	- 1.8
7		128	- 41.3	- 29.8	- 0.1	- 0.2
8		1024	- 41.0	- 28.3	- 0.0	- 2.1
9	seq-rd-32th-1f	4	- 2.4	- 3.0	+ 0.0	+ 2.1
10		32	- 2.4	- 4.1	+ 0.7	+ 2.2
11		128	- 2.6	- 4.4	+ 1.5	+ 2.0
12		1024	- 2.5	- 4.0	- 0.1	- 0.4
13	rnd-rd-1th-1f	4	- 10.0	- 10.0	- 32.1	- 39.8
14		32	- 7.4	- 7.5	- 18.8	- 25.2
15		128	- 7.4	- 5.5	- 14.7	- 12.4
16		1024	- 9.0	- 3.1	- 15.3	- 1.5
17	rnd-rd-32th-1f	4	- 60.4	- 23.2	- 82.5	- 27.6
18		32	- 56.2	- 17.2	- 55.7	- 1.9
19		128	- 34.4	- 11.4	- 29.1	- 2.6
20		1024	- 37.0	- 15.0	- 12.2	- 1.9

Read-ahead helped

Background queue limit

Read Workloads

#	Workload	I/O Size (KB)	HDD Results		SSD Results	
			Base (% Diff)	Optimized (% Diff)	Base (% Diff)	Optimized (% Diff)
1	seq-rd-1th-1f	4	- 2.5	+ 1.7	- 0.5	- 0.9
2		32	- 0.2	- 2.2	+ 0.8	+ 0.3
3		128	- 0.9	- 2.1	+ 0.4	+ 1.7
4		1024	- 0.9	- 2.2	+ 0.2	- 0.3
5	seq-rd-32th-32f	4	- 36.9	- 26.9	- 0.1	- 0.2
6		32	- 41.5	- 30.3	- 0.1	- 1.8
7		128	- 41.3	- 29.8	- 0.1	- 0.2
8		1024	- 41.0	- 28.3	- 0.0	- 2.1
9	seq-rd-32th-1f	4	- 2.4	- 3.0	+ 0.0	+ 2.1
10		32	- 2.4	- 4.1	+ 0.7	+ 2.2
11		128	- 2.6	- 4.4	+ 1.5	+ 2.0
12		1024	- 2.5	- 4.0	- 0.1	- 0.4
13	rnd-rd-1th-1f	4	- 10.0	- 10.0	- 32.1	- 39.8
14		32	- 7.4	- 7.5	- 18.8	- 25.2
15		128	- 7.4	- 5.5	- 14.7	- 12.4
16		1024	- 9.0	- 3.1	- 15.3	- 1.5
17	rnd-rd-32th-1f	4	- 60.4	- 23.2	- 82.5	- 27.6
18		32	- 56.2	- 17.2	- 55.7	- 1.9
19		128	- 34.4	- 11.4	- 29.1	- 2.6
20		1024	- 37.0	- 15.0	- 12.2	- 1.9

Read-ahead helped

Read-ahead doesn't help

Write Workloads

Write Workloads

#	Workload	I/O Size (KB)	HDD Results		SSD Results	
			Base (% Diff)	Optimized (% Diff)	Base (% Diff)	Optimized (% Diff)
21	seq-wr-1th-1f	4	- 26.2	- 0.1	- 9.0	+ 0.1
22		32	- 17.8	- 0.2	- 2.5	+ 0.1
23		128	- 16.6	- 0.2	- 2.1	+ 0.1
24		1024	- 17.7	- 0.3	- 2.3	- 0.1
25	seq-wr-32th-32f	4	- 2.5	+ 0.1	+ 0.1	+ 0.2
26		32	- 2.7	+ 0.0	+ 0.1	+ 0.1
27		128	- 2.6	- 0.0	- 0.0	+ 0.2
28		1024	- 2.4	- 0.2	- 0.1	+ 0.2
29	rnd-wr-1th-1f	4	- 0.7	- 1.3	+ 0.9	- 27.0
30		32	- 0.1	- 1.3	- 2.2	- 13.0
31		128	- 0.1	- 1.3	- 1.7	- 0.7
32		1024	- 0.01	- 0.8	- 0.0	- 0.3
33	rnd-wr-32th-1f	4	- 0.9	- 1.8	- 0.7	- 26.6
34		32	+ 0.1	- 0.7	- 2.2	- 13.0
35		128	+ 0.3	- 1.1	- 0.1	+ 0.0
36		1024	+ 0.1	- 0.3	+ 0.9	- 0.3

Write Workloads

#	Workload	I/O Size (KB)	HDD Results		SSD Results	
			Base (% Diff)	Optimized (% Diff)	Base (% Diff)	Optimized (% Diff)
21	seq-wr-1th-1f	4	- 26.2	- 0.1	- 9.0	+ 0.1
22		32	- 17.8	- 0.2	- 2.5	+ 0.1
23		128	- 16.6	- 0.2	- 2.1	+ 0.1
24		1024	- 17.7	- 0.3	- 2.3	- 0.1
25	seq-wr-32th-32f	4	- 2.5	+ 0.1	+ 0.1	+ 0.2
26		32	- 2.7	+ 0.0	+ 0.1	+ 0.1
27		128	- 2.6	- 0.0	- 0.0	+ 0.2
28		1024	- 2.4	- 0.2	- 0.1	+ 0.2
29	rnd-wr-1th-1f	4	- 0.7	- 1.3	+ 0.9	- 27.0
30		32	- 0.1	- 1.3	- 2.2	- 13.0
31		128	- 0.1	- 1.3	- 1.7	- 0.7
32		1024	- 0.01	- 0.8	- 0.0	- 0.3
33	rnd-wr-32th-1f	4	- 0.9	- 1.8	- 0.7	- 26.6
34		32	+ 0.1	- 0.7	- 2.2	- 13.0
35		128	+ 0.3	- 1.1	- 0.1	+ 0.0
36		1024	+ 0.1	- 0.3	+ 0.9	- 0.3

Observation 5:
At least one configuration is in green

Write Workloads

#	Workload	I/O Size (KB)	HDD Results		SSD Results	
			Base (% Diff)	Optimized (% Diff)	Base (% Diff)	Optimized (% Diff)
21	seq-wr-1th-1f	4	- 26.2	- 0.1	- 9.0	+ 0.1
22		32	- 17.8	- 0.2	- 2.5	+ 0.1
23		128	- 16.6	- 0.2	- 2.1	+ 0.1
24		1024	- 17.7	- 0.3	- 2.3	- 0.1
25	seq-wr-32th-32f	4	- 2.5	+ 0.1	+ 0.1	+ 0.2
26		32	- 2.7	+ 0.0	+ 0.1	+ 0.1
27		128	- 2.6	- 0.0	- 0.0	+ 0.2
28		1024	- 2.4	- 0.2	- 0.1	+ 0.2
29	rnd-wr-1th-1f	4	- 0.7	- 1.3	+ 0.9	- 27.0
30		32	- 0.1	- 1.3	- 2.2	- 13.0
31		128	- 0.1	- 1.3	- 1.7	- 0.7
32		1024	- 0.01	- 0.8	- 0.0	- 0.3
33	rnd-wr-32th-1f	4	- 0.9	- 1.8	- 0.7	- 26.6
34		32	+ 0.1	- 0.7	- 2.2	- 13.0
35		128	+ 0.3	- 1.1	- 0.1	+ 0.0
36		1024	+ 0.1	- 0.3	+ 0.9	- 0.3

Big requests split into 4KBs

Write Workloads

#	Workload	I/O Size (KB)	HDD Results		SSD Results	
			Base (% Diff)	Optimized (% Diff)	Base (% Diff)	Optimized (% Diff)
21	seq-wr-1th-1f	4	- 26.2	- 0.1	- 9.0	+ 0.1
22		32	- 17.8	- 0.2	- 2.5	+ 0.1
23		128	- 16.6	- 0.2	- 2.1	+ 0.1
24		1024	- 17.7	- 0.3	- 2.3	- 0.1
25	seq-wr-32th-32f	4	- 2.5	+ 0.1	+ 0.1	+ 0.2
26		32	- 2.7	+ 0.0	+ 0.1	+ 0.1
27		128	- 2.6	- 0.0	- 0.0	+ 0.2
28		1024	- 2.4	- 0.2	- 0.1	+ 0.2
29	rnd-wr-1th-1f	4	- 0.7	- 1.3	+ 0.9	- 27.0
30		32	- 0.1	- 1.3	- 2.2	- 13.0
31		128	- 0.1	- 1.3	- 1.7	- 0.7
32		1024	- 0.01	- 0.8	- 0.0	- 0.3
33	rnd-wr-32th-1f	4	- 0.9	- 1.8	- 0.7	- 26.6
34		32	+ 0.1	- 0.7	- 2.2	- 13.0
35		128	+ 0.3	- 1.1	- 0.1	+ 0.0
36		1024	+ 0.1	- 0.3	+ 0.9	- 0.3

Big requests

No batching, asynchronous

Metadata and Macro Workloads

Metadata and Macro Workloads

#	Workload	I/O Size (KB)	HDD Results		SSD Results	
			Base (% Diff)	Optimized (% Diff)	Base (% Diff)	Optimized (% Diff)
37	files-cr-1th	4	- 57.0	- 81.0	- 62.2	- 83.3
38	files-cr-32th	4	- 50.2	- 54.9	- 57.6	- 62.6
39	files-rd-1th	4	+ 0.0	- 10.6	- 25.0	- 60.3
40	files-rd-32th	4	- 50.5	- 4.5	- 74.1	- 33.0
41	files-del-1th	-	- 4.0	- 10.2	- 31.6	- 60.7
42	files-del-32th	-	- 2.8	- 6.9	- 42.9	- 52.6
43	file-server	-	- 26.3	- 1.4	- 41.2	- 1.5
44	mail-server	-	- 45.0	- 4.6	- 70.5	- 32.5
45	web-server	-	- 51.8	+ 6.2	- 72.9	- 17.3

Metadata and Macro Workloads

#	Workload	I/O Size (KB)	HDD Results		SSD Results	
			Base (% Diff)	Optimized (% Diff)	Base (% Diff)	Optimized (% Diff)
37	files-cr-1th	4	- 57.0	- 81.0	- 62.2	- 83.3
38	files-cr-32th	4	- 50.2	- 54.9	- 57.6	- 62.6
39	files-rd-1th	4	+ 0.0	- 10.6	- 25.0	- 60.3
40	files-rd-32th	4	- 50.5	- 4.5	- 74.1	- 33.0
41	files-del-1th	-	- 4.0	- 10.2	- 31.6	- 60.7
42	files-del-32th	-	- 2.8	- 6.9	- 42.9	- 52.6
43	file-server	-	- 26.3	- 1.4	- 41.2	- 1.5
44	mail-server	-	- 45.0		- 70.5	- 32.5
45	web-server	-				- 17.3

Many small requests, Optimizations didn't help

Metadata and Macro Workloads

#	Workload	I/O Size (KB)	HDD Results		SSD Results	
			Base (% Diff)	Optimized (% Diff)	Base (% Diff)	Optimized (% Diff)
37	files-cr-1th	4	- 57.0	- 81.0	- 62.2	- 83.3
38	files-cr-32th	4	- 50.2	- 54.9	- 57.6	- 62.6
39	files-rd-1th	4	+ 0.0	- 10.6	- 25.0	- 60.3
40	files-rd-32th	4	- 50.5	- 4.5	- 74.1	- 33.0
41	files-del-1th	-	- 4.0	- 10.2	- 31.6	- 60.7
42	files-del-32th	-	- 2.8	- 6.9	- 42.9	- 52.6
43	file-server	-	- 26.3	- 1.4	- 41.2	- 1.5
44	mail-server	-	- 45.0		- 70.5	- 32.5
45	web-server	-				- 17.3

Many small requests. Optimizations didn't help

Optimizations help

Evaluation Summary

- Read workloads
  - Sequential workloads are in green zone
  - Optimizations helped random workloads
- Write workloads
  - In green zone
- Metadata workloads
  - Poor performance
- Macro workloads
  - Optimizations helped a lot

Evaluation Summary

- Read workloads
  - Sequential workloads are in green zone
  - Optimizations helped random workloads
- Write workloads
  - In green zone
- Metadata workloads
  - Poor performance
- Macro workloads
  - Optimizations helped a lot

More results and analysis in paper

Conclusions

Conclusions

- Detailed description of FUSE internals
- Broad performance characterization
- In-depth performance analysis

Future Directions

Future Directions

- Shared memory
- New VFS operations
- Compounding requests
  - See vNFS talk on Thursday

To FUSE or Not to FUSE: Performance of User-Space File Systems

Q&A

Vangoor Bharath Kumar Reddy, Vasily Tarasov, and Erez Zadok

http://filesystems.org/fuse/

Stony Brook University IBM Research