

CS244b - Aurora

Learning goals

- chain replication
- write-ahead logs
- quorum systems
- rethinking abstraction layers

3-tier architecture
EC2
Public ✓

application ✓

Storage
Shard data

(dynamo)

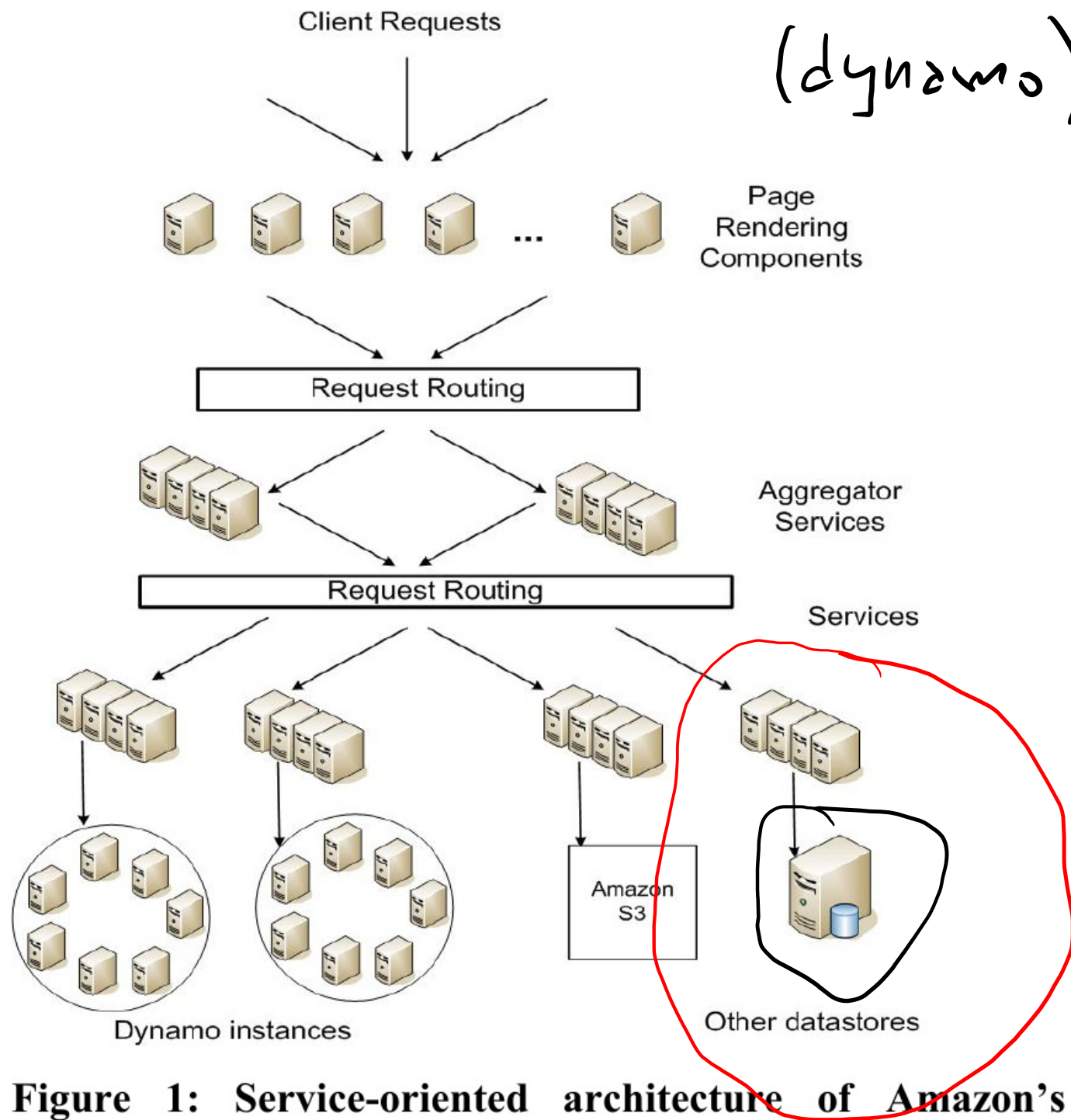
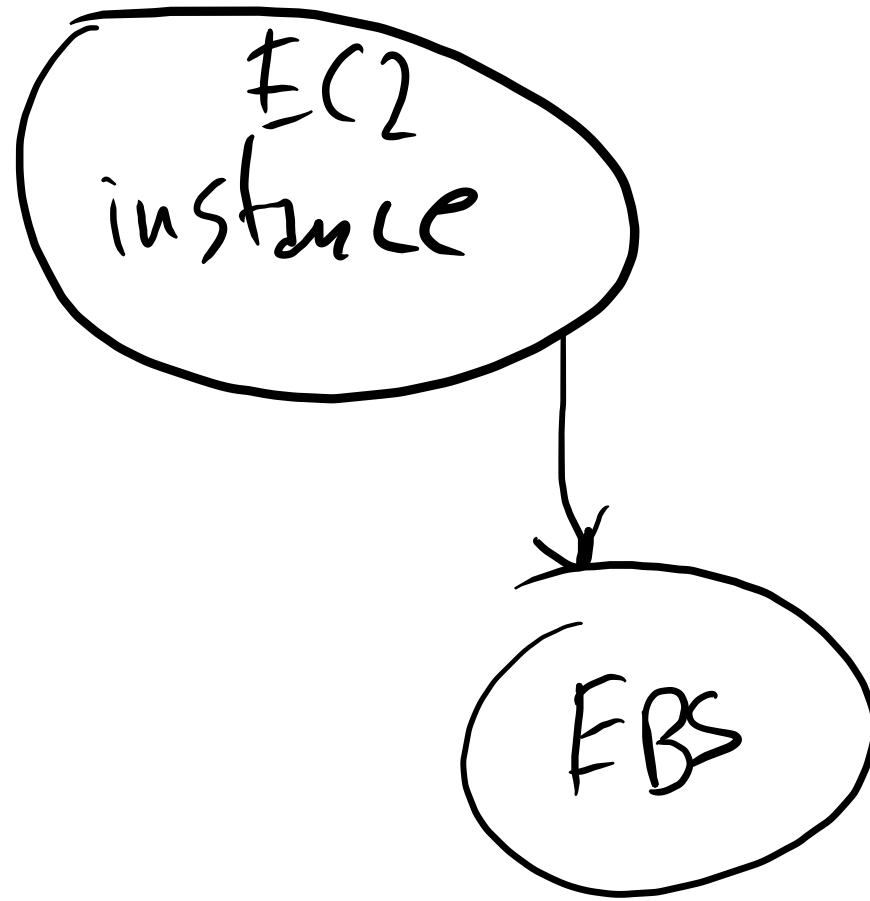
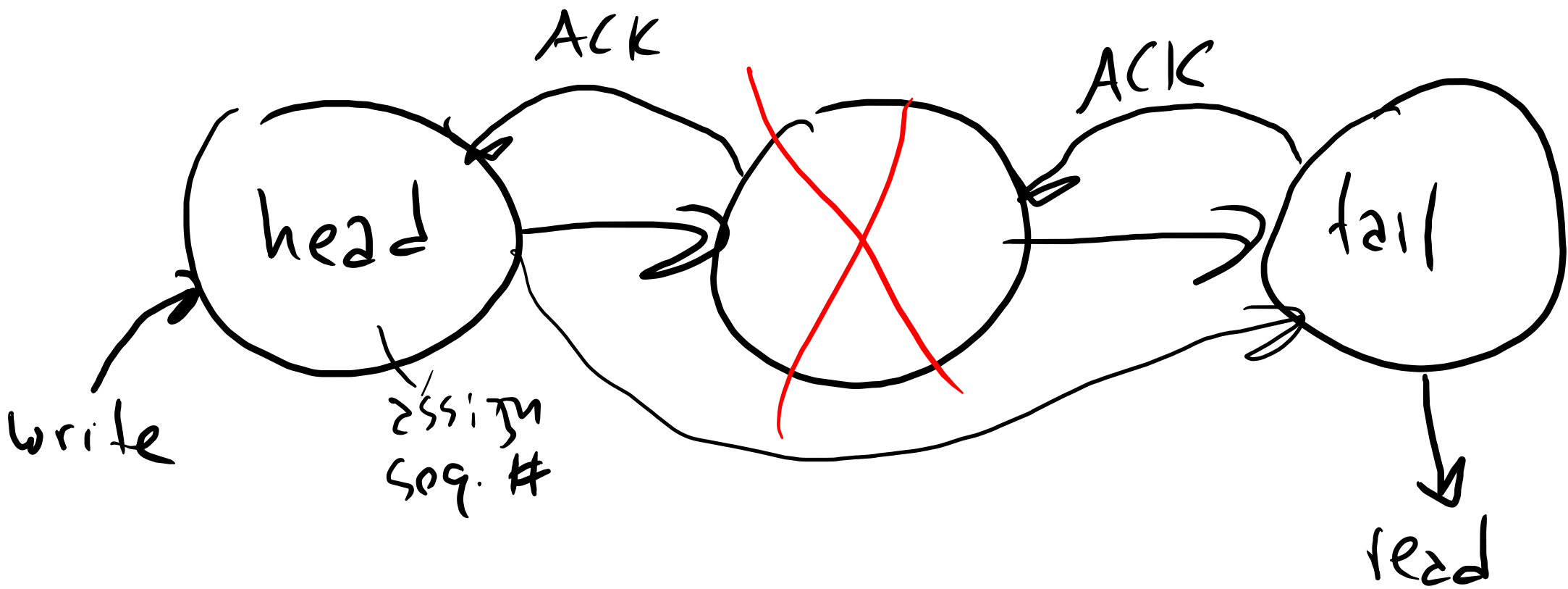


Figure 1: Service-oriented architecture of Amazon's platform

EBS



2x replication



MySQL on EC2 w. EBS

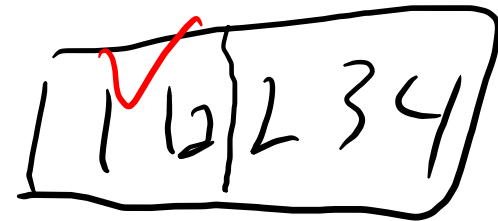
Bad for latency (median, tail)

Bad for net. ~~B~~/w

InnoDB xaction commit



- WAL write record here



- Update data on disk

- Double-write pages

- Statement log, FRM

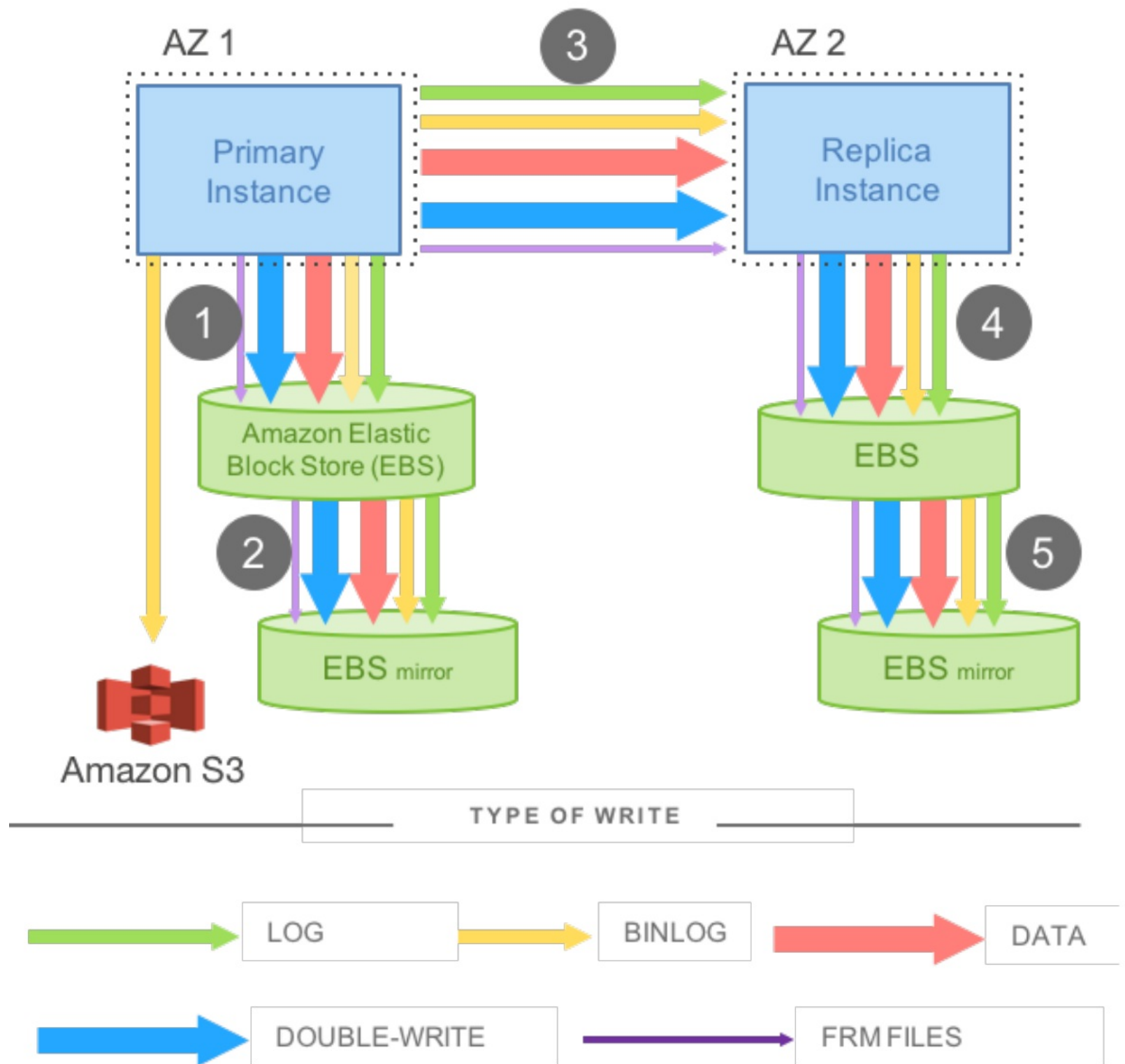
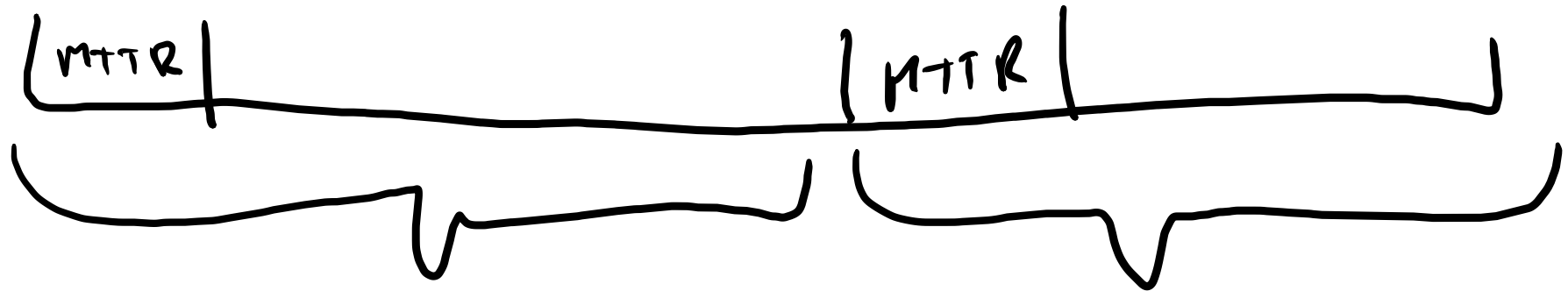


Figure 2: Network IO in mirrored MySQL

AZ+1

Survive AZ failure and failure of
Server not in AZ

MTTF, MTTR



Small segments (LOG₂B)^{MTTF}

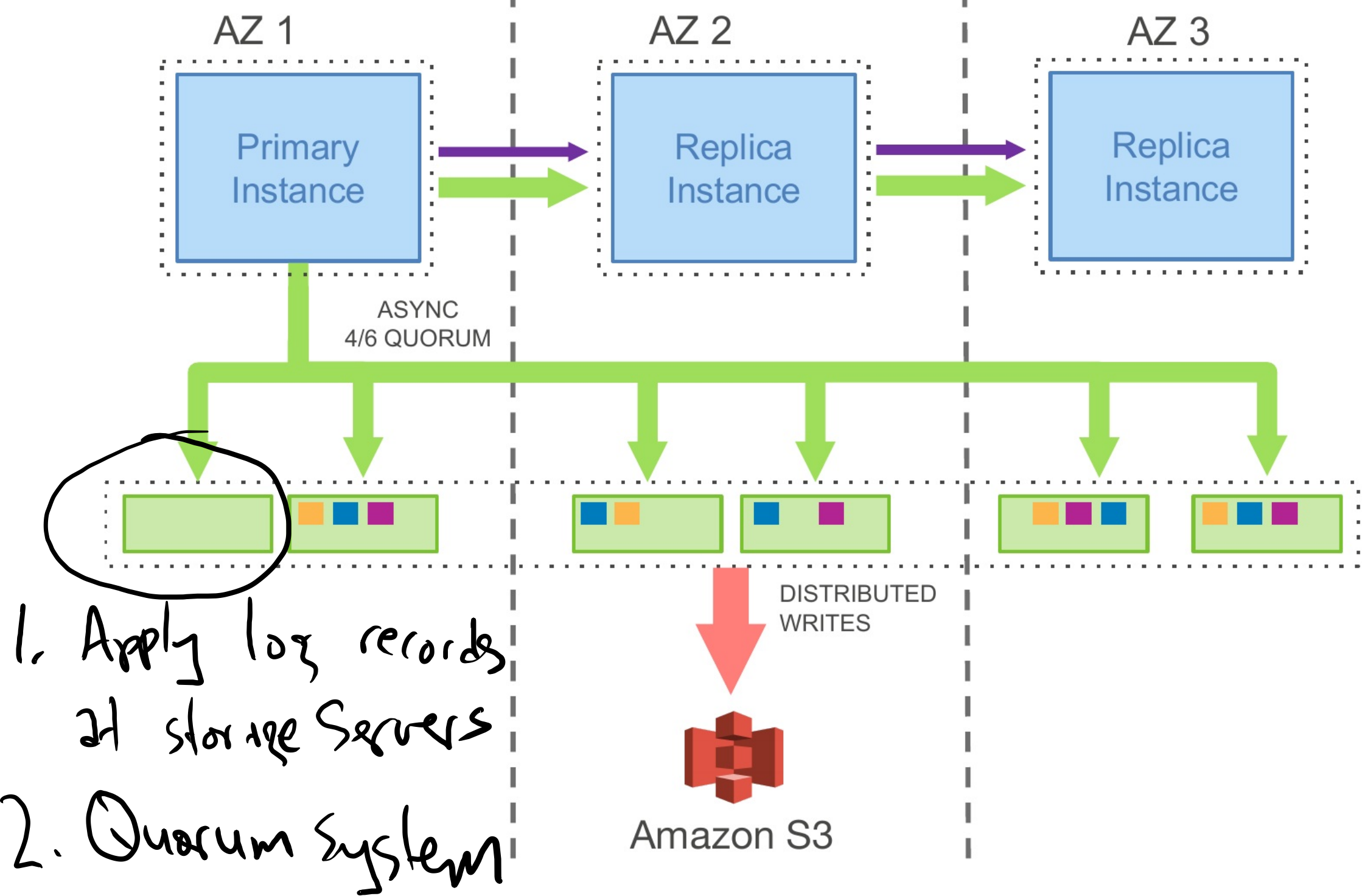


Figure 3: Network IO in Amazon Aurora

$$N=6, W=4, R=3$$

ACRONYMS

PG - protection groups

LSN - log sequence number

VDL - volume durable LSN

LAL - log allocation limit (LOM)

CPL - consistency point LSN

MTR - mini transaction

SCL - segment complete LSN (per-server)

VCL - Volume complete list

MySQL + Aurora transaction commit

- Primary create log records,
assign 2 LSN, $VDL \leq LSN \leq UDL \rightarrow LK$
- Create commit record, flag as CPL
- Send commit record, wait for $VDL \geq LSN$
Get W Acks from each PF
use backlinks to determine SCL

Storage API (RPCs)

- write log record
- read data page (as of LSN)
- truncate

Stored at Storage Server

Log, Data pages, per-page log records

Reads

Read want $SCL \geq VCL$

Wait for 1 server

evict page on storage server?

pick page w.

$LSN > VDL$

DB
Crash

- After DB instance reboot wait for R

from each PF, compute VDL

- Tell servers to truncate $(VDL, \dots, VDL+L)$

- Undo in flight transactions

Storage server crash

Read-only

Send log records to RO instances

Apply $LSN \leq UDL$

only complete MRs

Table 1: Network IOs for Aurora vs MySQL

Configuration	Transactions	IOs/Transaction
Mirrored MySQL	780,000	7.4
Aurora with Replicas	27,378,000	0.95

Table 3: SysBench OLTP (writes/sec)

Connections	Amazon Aurora	MySQL
50	40,000	10,000
500	71,000	21,000
5,000	110,000	13,000

Table 5: Percona TPC-C Variant (tpmC)

Connections/Size/ Warehouses	Amazon Aurora	MySQL 5.6	MySQL 5.7
500/10GB/100	73,955	6,093	25,289
5000/10GB/100	42,181	1,671	2,592
500/100GB/1000	70,663	3,231	11,868
5000/100GB/1000	30,221	5,575	13,005