

# ХРАНЕНИЕ ДАННЫХ

---

# ТРАНЗАКЦИИ

- Атомарность (atomicity) – гарантирует, что никакая транзакция не будет зафиксирована в системе частично. Будут либо выполнены все операции внутри транзакции, либо не выполнено ни одной.
- Согласованность (consistency) – каждая успешная транзакция по определению фиксирует только допустимые результаты.
- Изолированность (isolation) – гарантирует что никакой поток данных не может читать данные из еще не завершенной транзакции.
- Долговечность (durability) – означает что если транзакция завершена, то все данные записаны на диск.



# LOST UPDATE

Транзакция 1	Транзакция
UPDATE table1 SET f2 = f2 + 20 WHERE f1 = 1;	UPDATE table1 SET f2 = f2 + 5 WHERE f1 = 1;

# DIRTY READ

Транзакция 1	Транзакция
UPDATE table1 SET f2 = f2 + 1 WHERE f1 = 1;	
	SELECT f2 FROM table1 WHERE f1 = 1;
ROLLBACK;	



# NON REPEATABLE READ

Транзакция 1	Транзакция
	SELECT f2 FROM table1 WHERE f1 = 1;
UPDATE table1 SET f2 = f2 + 1 WHERE f1 = 1;	
COMMIT;	
	SELECT f2 FROM table1 WHERE f1 = 1;

# PHANTOM READ

Транзакция 1	Транзакция
	SELECT SUM(f2) FROM table1;
INSERT INTO tbl1 (f1,f2) VALUES (15,20);	
COMMIT;	
	SELECT SUM(f2) FROM table1;



# УРОВНИ ИЗОЛЯЦИИ ТРАНЗАКЦИЙ

Isolation level	Lost updates	Dirty reads	Non-repeatable reads	Phantoms
Read Uncommitted	don't occur	may occur	may occur	may occur
Read Committed	don't occur	don't occur	may occur	may occur
Repeatable Read	don't occur	don't occur	don't occur	may occur
Serializable	don't occur	don't occur	don't occur	don't occur

# УРОВНИ ИЗОЛЯЦИИ ТРАНЗАКЦИЙ

- Read Uncommitted – отсутствие потерянных обновлений.
- Read Committed – чтение зафиксированных данных, решение проблемы Dirty Read.
- Repeatable Read – читающая транзакция не видит изменения данных, которые были ей ранее прочитаны.
- Serializable – транзакции полностью изолируются друг от друга.



# БЛОКИРОВКИ

Блокировка – отметка о захвате объекта транзакцией в ограниченный или исключительный доступ с целью предотвращения коллизий и поддержания целостности данных.

- Pessimistic – блокируются данные в разделяемом режиме, из-за чего другие транзакции, пытающиеся изменить эти данные, приостанавливаются.
- Optimistic – при каждом изменении строки СУБД создаёт новую версию этой строки, с которой продолжает работать изменившая данные транзакция, в то время как любой другой «читающей» транзакции возвращается последняя зафиксированная версия.

# WRITE AHEAD LOG

Журнал предзаписи (WAL) — это стандартный метод обеспечения целостности данных. Основная идея WAL состоит в том, что изменения в файлах с данными (где находятся таблицы и индексы) должны записываться только после того, как эти изменения были занесены в журнал, т. е. после того как записи журнала, описывающие данные изменения, будут сохранены на постоянное устройство хранения.

Начальное состояние:  $A = 10$ ,  $B = 20$ .

Транзакция:  $A += 20$ ,  $B += 20$ .

1:  $\langle \text{start } T \rangle \rightarrow \langle T, \text{LSN}:0, \text{begin} \rangle$

2:  $\text{read } A = 10, A += 20 \rightarrow \langle T, \text{PrevLSN}:0, \text{LSN}:100, \text{update}, 10, 30 \rangle$

3:  $\text{read } B = 20, B += 20 \rightarrow \langle T, \text{PrevLSN}:100, \text{LSN}:200, \text{update}, 20, 40 \rangle$

4:  $\langle \text{commit } T \rangle \rightarrow \langle T, \text{PrevLSN}:200, 300, \text{commit} \rangle$

5:  $\langle \text{flush log} \rangle$

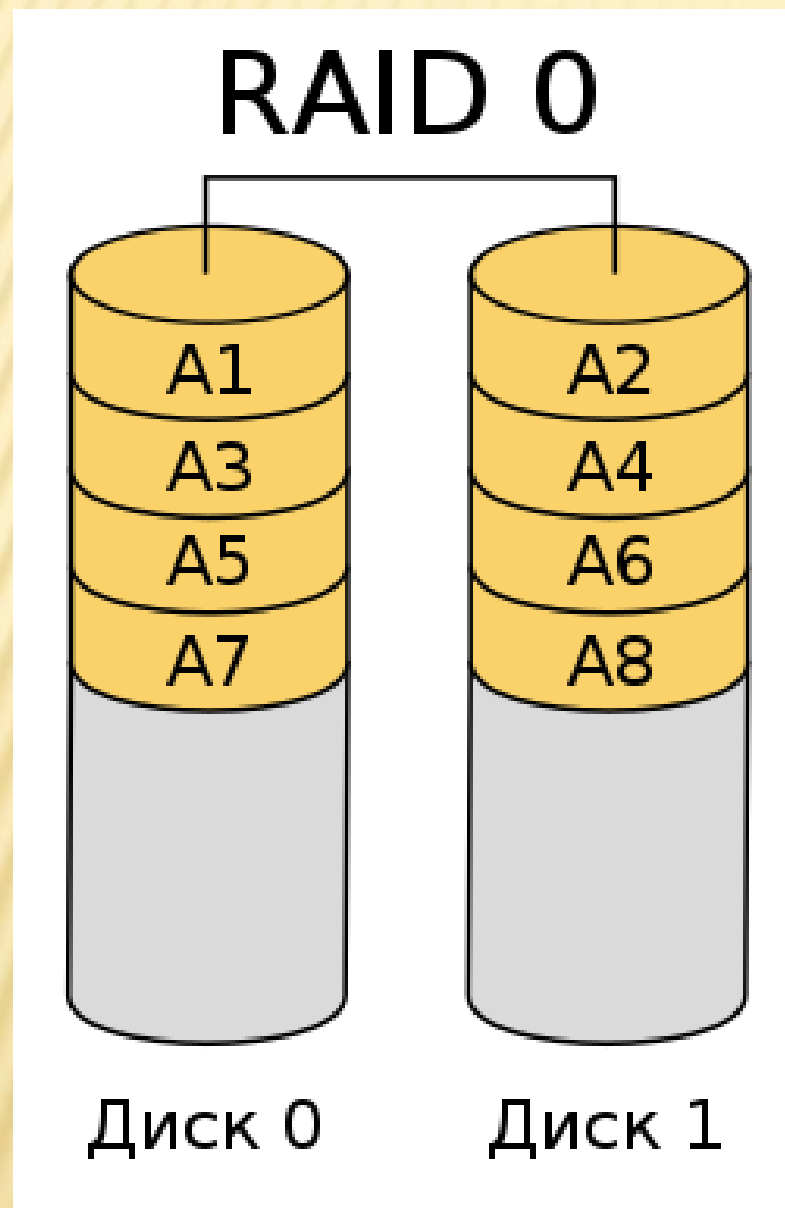
6: output A

7: output B

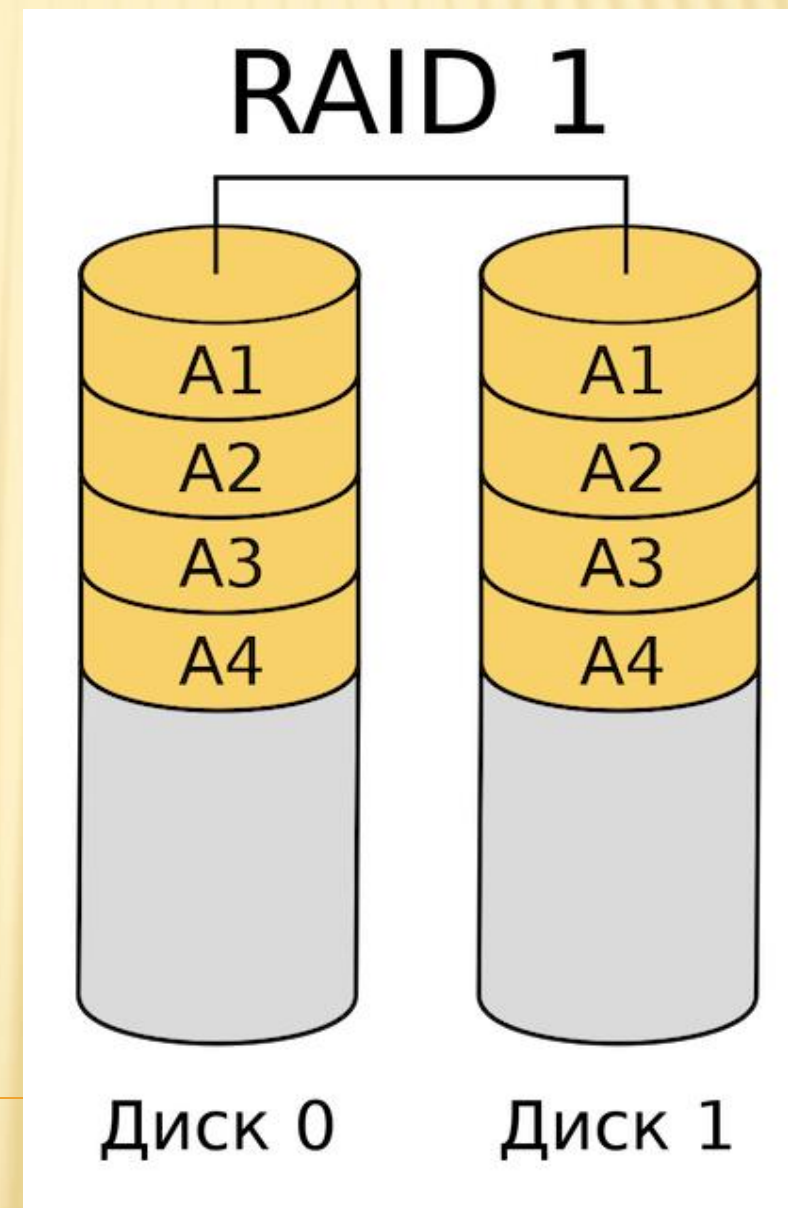


# RAID

## RAID 0 (Stripe)

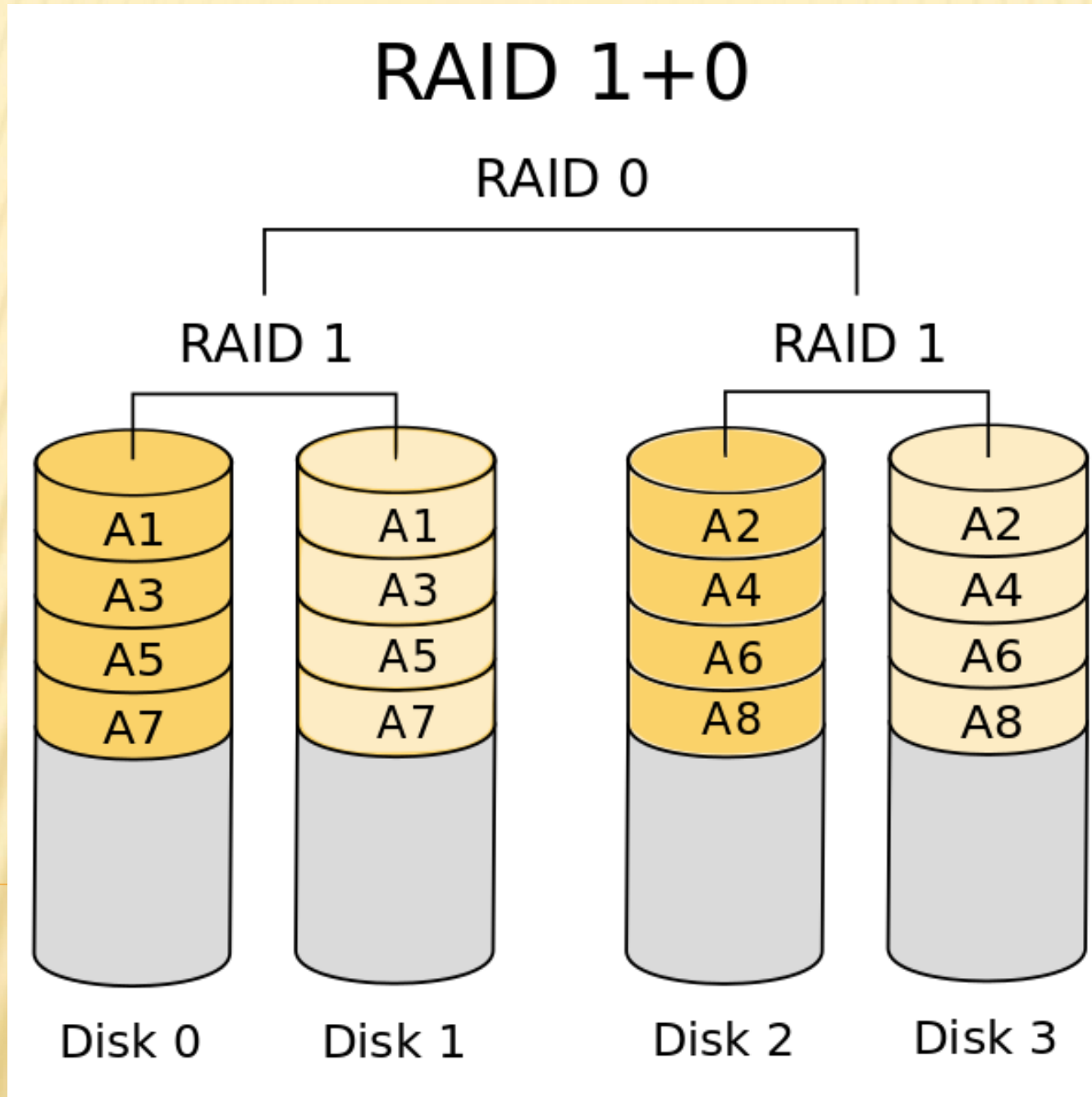


## RAID 1 (Mirror)



# RAID

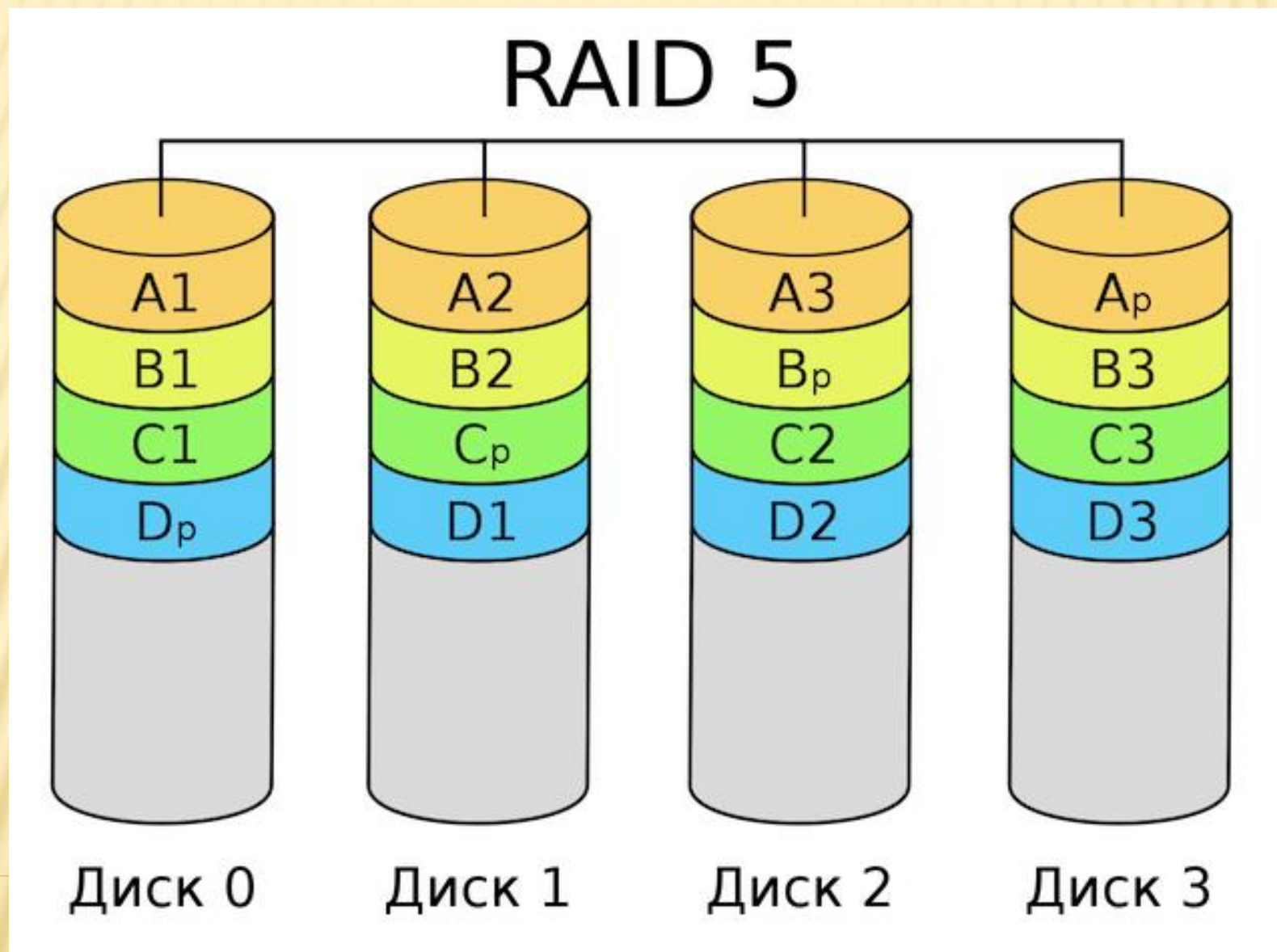
## RAID 10 (RAID 1+0)





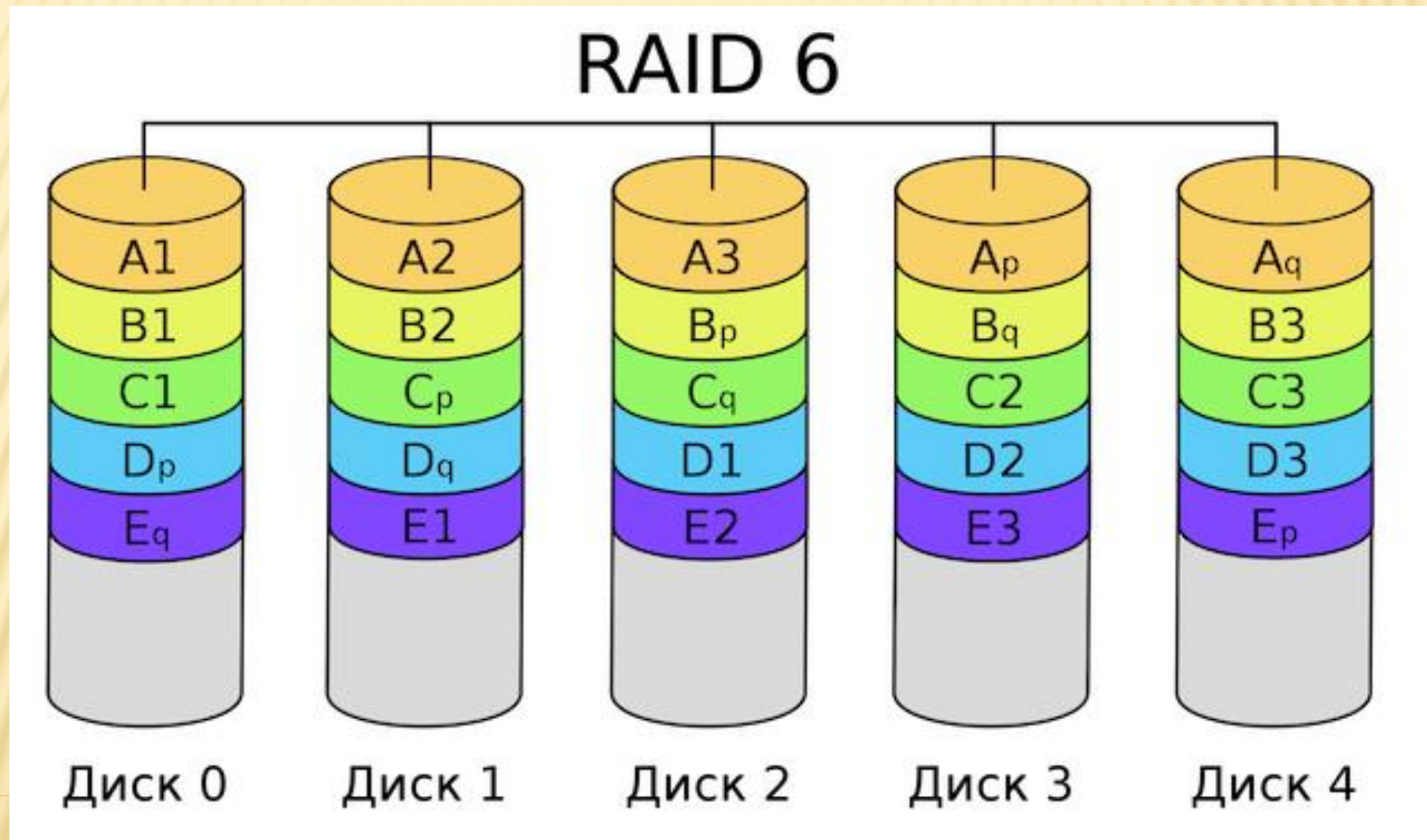
# RAID

## RAID 5 (Parity)



# RAID

## RAID 6 (Double parity)





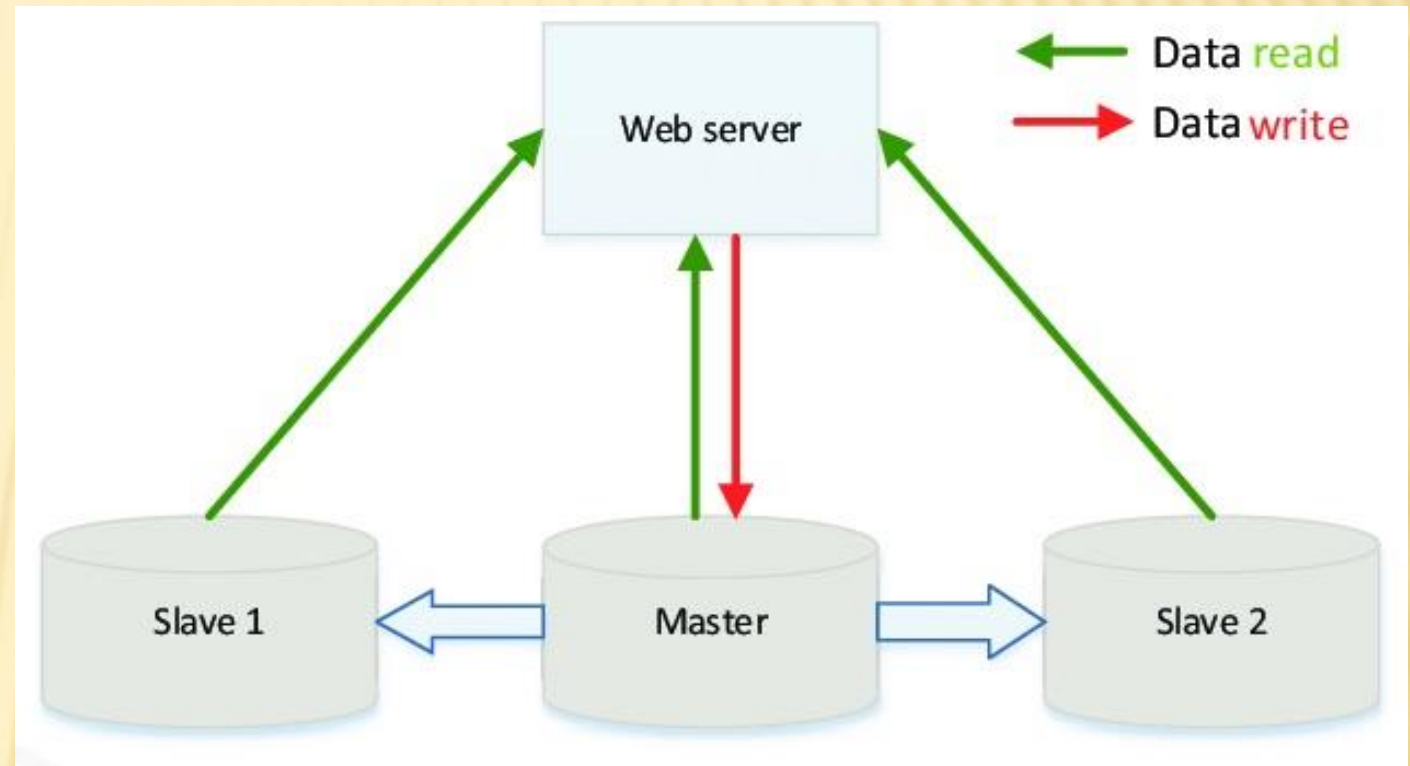
# РЕПЛИКАЦИЯ ДАННЫХ

По внутреннему устройству:

- Разделяемые диски.
- Трансляция WAL.
- Репликация запросов.

По типу репликации:

- Master-slave.
- Каскадная репликация:
  - Синхронная и Асинхронная.
  - Hot Standby.
- Master-Master.



# ШАРДИРОВАНИЕ

Original Table

CUSTOMER ID	FIRST NAME	LAST NAME	FAVORITE COLOR
1	TAEKO	OHNUKI	BLUE
2	O.V.	WRIGHT	GREEN
3	SELDA	BAĞCAN	PURPLE
4	JIM	PEPPER	AUBERGINE

Vertical Partitions

VP1

CUSTOMER ID	FIRST NAME	LAST NAME
1	TAEKO	OHNUKI
2	O.V.	WRIGHT
3	SELDA	BAĞCAN
4	JIM	PEPPER

VP2

CUSTOMER ID	FAVORITE COLOR
1	BLUE
2	GREEN
3	PURPLE
4	AUBERGINE

Horizontal Partitions

HP1

CUSTOMER ID	FIRST NAME	LAST NAME	FAVORITE COLOR
1	TAEKO	OHNUKI	BLUE
2	O.V.	WRIGHT	GREEN

HP2

CUSTOMER ID	FIRST NAME	LAST NAME	FAVORITE COLOR
3	SELDA	BAĞCAN	PURPLE
4	JIM	PEPPER	AUBERGINE



# ШАРДИРОВАНИЕ

Способ выбора ключа:

- Hash-функция.
- Табличная функция.
- Консистентное хэширование – каждый сервер отвечает за некий диапазон ключей, соответственно, при добавлении нового сервера, он забирает те диапазоны, которые находятся перед ним и после него, т.е. он частично делит диапазон. Не требуется перетасовки всего шарда.

# ШАРДИРОВАНИЕ

Доступ к данным:

- Умный клиент.
- Proxy.
- Координатор.
- Random Routing.



# РЕШАРДИНГ

Способы выполнения решардинга:

- Update is a move.
- Data expiration.
- Новые данные на новые сервера.