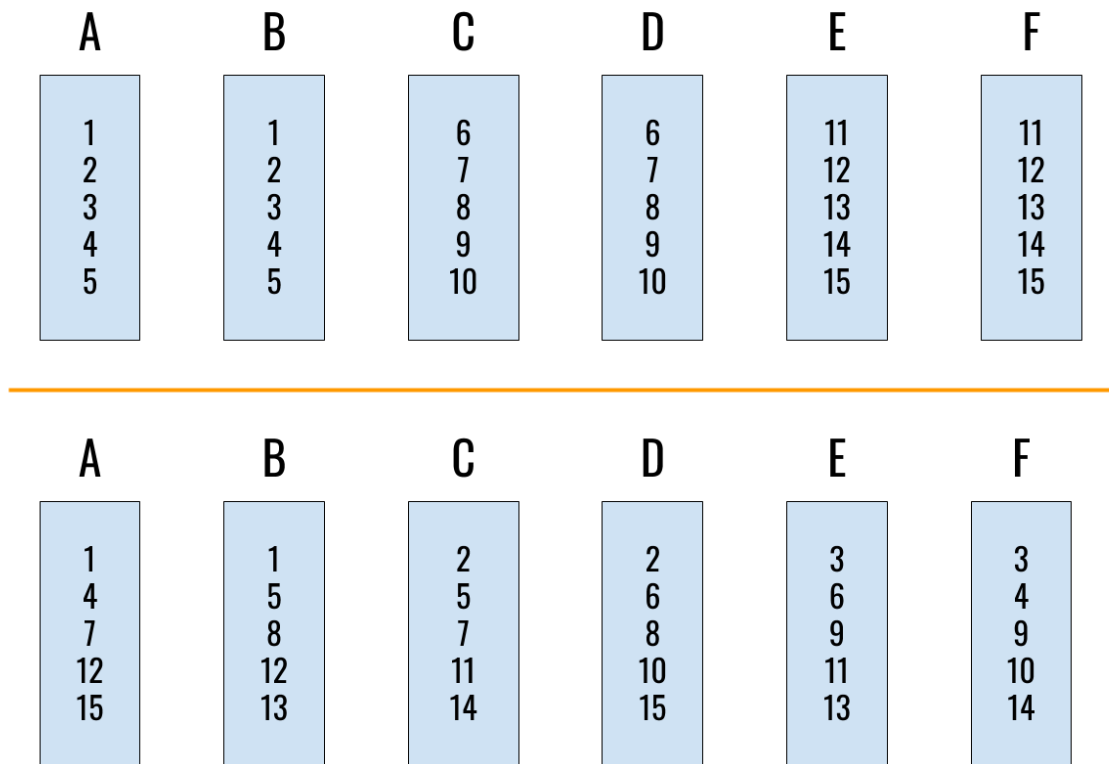


## Задание 2: Симулятор катастроф

На картинке изображены два варианта того как могут храниться данные в распределенной базе данных:



На схеме, буквами A, B, C, D, E, F обозначены сервера. Числами 1-15 обозначены фрагменты данных (чаще всего их называют shard). Где одинаковые числа означают одинаковый фрагмент данных хранится на этих серверах (такие копии называют репликами).

Предположим, что узел A вышел из строя первым. И посчитаем вероятность потери данных (когда ни на одном из оставшихся узлов не осталось реплики данных):

1. В первом случае выход из строя B означает потерю фрагментов 1-5, тогда как выход из строя любого из C, D, E, F не приводит к потере данных. Т.е. вероятность потери данных  $\frac{1}{5} = 0.2$
2. Во втором случае, при падении B мы теряем фрагмент 1, при падении C -> 7, также D -> 15, F -> 4. Т.е. вероятность потери данных  $\frac{4}{5} = 0.8$

## Задача

Нужно создать симулятор выхода из строя серверов:

```
./simulate.py -n 10 --random
```

Симулятор создает 10 виртуальных серверов. Размещает в них 100 фрагментов данных случайным образом по две реплики каждого фрагмента (важно: не размещать 2 реплики на одном и том же сервере). Проверяет что происходит при падении каждых двух серверов. И отображает вероятность потери хотя бы одного фрагмента данных<sup>1</sup>:

```
Killing 2 arbitrary servers results in data loss in 30% cases
```

И второй режим:

```
> ./simulate.py -n 6 --mirror  
Killing 2 arbitrary servers results in data loss in 20% cases
```

Который размещает данные как на верхней части схемы: каждый узел имеет еще одну полную копию всех фрагментов.

**Важно:** обе симуляции должны отличаться только размещением фрагментов данных, остальная часть кода должна быть одинаковая<sup>2</sup>.

---

<sup>1</sup> Цифра в выводе не верная, не стоит на нее ориентироваться

<sup>2</sup> Да мы знаем, что в случае --mirror можно аналитически рассчитать вероятность, но задание состоит не в этом.