

# Практическое задание: MapReduce

---

## Подготовка:

- ① Создать S3 **бакет** в Object Storage ([документация](#))
  - ② Скопировать данные желтого такси за 2020 год в созданный s3 бакет.
- 

## Задание:

Написать map-reduce приложение, использующее скопированные на Object storage данные и вычисляющее отчет на каждый месяц 2020 года вида:

---

Month	Payment type	Tips average amount
2020-01	Credit card	999.99

---

## Требования к отчету:

- ① Количество файлов — 1
- ② Формат — csv
- ③ Сортировка — не требуется
- ④ Month считаем по полю trip\_pickup\_datetime
- ⑤ Payment type считаем по полю payment\_type
- ⑥ Tips average amount вычисляем по полю tip\_amount
- ⑦ Обратите внимание, в датасете будут присутствовать неверные данные (не тот год, пустой тип оплаты) - эти данные нужно будет отсеять.
- ⑧ Маппинг типов оплат:

mapping = { 1: 'Credit card', 2: 'Cash', 3: 'No charge', 4: 'Dispute', 5: 'Unknown', 6: 'Voided trip' }