

---

## Foundations of Data Science

### Exercise sheet 9

### Exercise 3

Let  $n_1$  be number of entries with the value 1 in the matrix. Then the space needed for the first approach is  $n * n * 1$  bit. The second approach needs  $n_1 * \lceil \log_2 n \rceil$  bit.

$$n_1 * \lceil \log_2 n \rceil < n * n \Leftrightarrow \frac{n_1}{n * n} < \frac{1}{\lceil \log_2 n \rceil}$$

This means, that the fraction of 1s has to be lower than  $\frac{1}{\lceil \log_2 n \rceil}$  for the second approach to save space.

### Exercise 4

a)

```
function MAP(value)
    max=INTEGER.MIN_VALUE
    for all v in value do
        if (v>max)
            max=v
    emit(1,max)

function REDUCE(key, values)
    max=INTEGER.MIN_VALUE
    for all v in values do
        if (v>max)
            max=v
    emit(max)
```

b)

```
function MAP(value)
    for all v in value do
        emit(1,v)

function REDUCE(key, values)
    sum=0
    n=0
    for all v in values do
        sum=sum+v
        n=n+1
    a=sum/n
    emit(1,a)
```

c)

```
function MAP(value)
    for all v in value do
        emit(v,1)

function REDUCE(key, values)
    emit(1,key)
```

d)

```
function MAP(value)
    for all v in value do
        emit(v,1)
```

```
function REDUCE(keys, values)
    cnt=0
    for all k in values do
        cnt=cnt+1
    emit(1,cnt)
```