n data Sources $(D_1 \dots D_n)$

groups $\{g_1 \dots g_k\}$

Count req. $\{c_1 \dots c_k\}$

$\forall D_i = \{P_{i1}, P_{i2} \dots P_{ik}\}$

$\hookrightarrow$ Cost σ_i for a query $\leftarrow$ query(D_i) returns a Sample from D_i

goal: Collect the Target dataset with min Cost

only 2 groups g_1, g_2, $\{D_1 \dots D_n\}$

$$f(c_1, c_2) = \begin{cases} \emptyset & c_1, c_2 = 0 \\ \min\limits_{\forall D_i} \left(P_{i1} f(c_1 - 1, c_2) + P_{i2} f(c_1, c_2 - 1) + \sigma_i \right) & \end{cases}$$

$$f(C) = \min_{\forall D_i} \left[\sum_{\forall g_j, c_j \geq 1} P_{ij} f(c_1 \dots c_j - 1 \dots) + \sum_{c_j = 0} P_{ij} f(C) + \sigma_i \right]$$

If there are only Two groups. (Equal Cost)

$$D_1^* = \min_{D_i} \frac{P_{i1}}{\sigma_i} \quad , \quad D_2^* = \min_{D_i} \frac{P_{i2}}{\sigma_i}$$

$$P_1^* \leqslant P_2^*$$

the optimal Strategy is to select D_1^*

- Multi-arm Bandit Problem

  - UCB : Upper Confidence Bound
  - Thompson Sampling