Online supplemental materials accompanying the paper

Suppose a set of multiple top-k queries are \( Q = \{q_1, q_2, ..., q_x\} \) with the corresponding sets \( F = \{f_1, f_2, ..., f_x\} \) and \( K = \{k_1, k_2, ..., k_x\} \). The largest element in \( F \) is \( f_{\text{max}} \), and the largest element in \( K \) is \( k_{\text{max}} \). Assume that the groups after combination are \( G = \{G_1, ..., G_g\} \). We change the way a little to show an execution plan. Suppose there is a valid execution plan \( EP = \{(f_1, k_1'), (f_2, k_2'), ..., (f_u, k_u')\} \) where \( f_i' = 2^{i-1} \cdot f_i \), \( f_i' \leq f_{\text{max}} \), and \( k_i' = k_{\text{max}} \). Moreover, we have \( k_i' \leq k_{\text{max}} \) where \( i, j \leq u \) and \( i < j \). That is, there are some queries executed according to frequency \( f_i' \) with the value of \( k_i' \) where \( i \leq u \). Given this execution plan \( EP_i \) if we want to execute \( G_i \) effectively, the time when \( G_i \) is first executed is \( f_i'^{1} = 2^{i-1} \cdot f_i' \) where \( 2^{i-1} \cdot f_i' \leq f_i < 2^i \cdot f_i' \), and \( k_{\text{max}} \leq k_i' \). Let the cost-per-unit time with \( EP \) be \( m' \), and let \( m_{\text{opt}} \) be the cost-per-unit time of the DP approach. Obviously, \( m_{\text{opt}} \leq m' \). Let \( m' \) be the cost-per-unit time of no sharing approach.

First, we give an upper bound of \( m' \) in the following lemma.

Then, we prove Theorem 7.1.

**Lemma 11.1**: \( m' \leq 2 \cdot k_{\text{max}} / f_i' \)

**Proof**: For each element \( (f_i', k_i') \) in \( EP \), the cost of executing the query with the \( k_i' \) is \( k_i' \). In a cycle of \( f_i' \), the total cost of executing this query is \( (f_i'/f_i') \cdot k_i' \). For all the elements in \( EP \), the total cost is \( \sum_{i=1}^{u} \left(f_i'/f_i' \right) \cdot k_i' \).

Thus, we have

\[
m' = \sum_{i=1}^{u} \left(f_i'/f_i' \right) \cdot k_i' / f_i'
\]

As \( f_i' = 2^{i-1} \cdot f_i' \), we have

\[
m' = \sum_{i=1}^{u} k_i' / (2^{i-1} \cdot f_i')
\]

Since \( k_i' \leq k_{\text{max}} \) and \( \sum_{i=1}^{u} 1/2^{i-1} \leq 2 \), we get

\[
m' \leq 2 \cdot k_{\text{max}} / f_i'
\]

**Proof of Theorem 4.1**

**Proof**: Obviously,

\[
E(m / m_{\text{opt}} | x, F, K) \geq E(m / m' | x, F, K)
\]

According to conditional probability theory, we have

\[
E(m / m' | x, F, K) = \Sigma p(f_i', k_i' | x, F, K) \cdot E(m / m' | f_i', k_i', x, F, K)
\]

Replacing \( m' \) with the inequality in Lemma 11.1, we have

\[
E(m / m' | k_1', k_2', ..., k_u', f_i', k_i', x, F, K) \geq E(m | k_1', k_2', ..., k_u', f_i', f_i', x, F, K) / (2 \cdot k_{\text{max}} / f_i')
\]

As each query in \( Q \) is independent, inequality (15) can be reduced to

\[
E(m / m' | k_1', k_2', ..., k_u', f_i', f_i', x, F, K) \geq x \cdot E(m | k_1', k_2', ..., k_u', f_i', f_i', x, F, K) / (2 \cdot k_{\text{max}} / f_i')
\]

where \( m_{\text{opt}} \) is the cost-per-unit time of one query executed with no sharing according to its frequency upper bound.

Let \( A_i = \{(f, k)|2^{i-1} \cdot f_i' \leq f < 2^i \cdot f_i', 1 \leq k \leq k_i'\} \) where \( (f, k) \) is a point in the area \( A_i \), \( A_i \cap A_j = \emptyset \) and \( 1 \leq i \leq u-1 \). Suppose \( A_u = \{(f, k)|2^{u-1} \cdot f_i' \leq f < f_{\text{max}}, 1 \leq k \leq k_i'\} \). \( (f, k) \) can be considered as one query. \( A_i \) is shown in Figure 18. Suppose \( A = \cup_{i=1}^{u} A_i \) and \( |A| \) is the area of \( A \) which is the dashed area in Figure 18.

Given a query \( (f, k) \) in the area \( A \), the cost-per-unit time is \( k/f \). Suppose the probability of \( (f, k) \) existing in \( A \) is \( p(f, k) \), we know that

\[
E(m_0 | k_1', k_2', ..., k_u', f_i', f_i', x, F, K) \geq \Sigma_{(f, k) \in A} p(f, k) / k / f
\]

Obviously, \( p(f, k) = 1/|A| \). Thus, inequality (17) can be reduced to

\[
= 1/|A| \cdot \Sigma_{(f, k) \in A} k / f
\]

Since \( A > \cup_{i=1}^{u} A_i \), we can reduce inequality (18) with inequality (19)

\[
\geq 1/|A| \cdot \Sigma_{i=1}^{u-1} (k_i'/2) / f_i' \leq k_{\text{max}} / f_i'
\]

As \( 2^{i-1} \cdot f_i' \leq f < 2^i \cdot f_i' \) and \( 1 \leq k \leq k_i' \), inequality (19) can be reduced to

\[
\geq (1/|A|) \cdot \Sigma_{i=1}^{u-1} (k_i'/2) / f_i' \leq k_{\text{max}} / f_i'
\]

According to the calculus theory, we have \( \Sigma_{i=1}^{u-1} (k_i'/2) / f_i' \leq \ln (2^i / f_i' / (2^{i-1} \cdot f_i')) \). Thus, inequality (20) can be reduced to inequality (21)

\[
\geq (1/|A|) \cdot \Sigma_{i=1}^{u-1} (k_i'/2) / f_i' \leq \ln (2^i / f_i' / (2^{i-1} \cdot f_i'))\]

\[
= \Sigma_{i=1}^{u-1} (k_i'/2) / (2 \cdot \log_2 e \cdot |A|)
\]

According to the assumption, \( k_i' \geq k_i' \) and \( |A| \leq k_i' \cdot f_{\text{max}} \). Thus, equality (21) can be reduced to

\[
\geq (u - 1) \cdot (k_i'/2) / (2 \cdot \log_2 e \cdot k_i' \cdot f_{\text{max}})
\]

Suppose the query with the frequency upper bound \( f_i' \) has a \( k_i' \) value. To satisfy this frequency upper bound, we know that \( k_i' \geq k_i' \). Thus, we can reduce inequality (22) to inequality (23).

\[
E(m_0 | k_1', k_2', ..., k_u', f_i', f_i', x, F, K) \geq (u - 1) \cdot (k_i'/2) / (2 \cdot \log_2 e \cdot k_i' \cdot f_{\text{max}})
\]

As \( f_i' = 2^{i-1} \cdot f_i' \) and \( 2^{u-1} \cdot f_i' \leq f_{\text{max}} < 2^u \cdot f_i' \),

\[
u \geq \log_2 f_{\text{max}} / f_i'
\]
Based on inequalities (23) and (24), and $k_u = k_{\text{max}}$,

$$E(m_0|k_1', k_2', ..., k_u', f_1', f_k', x, F, K)$$

$$\geq (\log_2 f_{\text{max}} / f_1' - 1) * ((k_1')^2) / (2 \times \log_2 e \times k_{\text{max}} \times f_{\text{max}})$$

$$= (\log_2 f_{\text{max}} / f_1' - 1) * ((k_1')^2) / (2 \times \log_2 e \times k_{\text{max}} \times f_{\text{max}})$$

(25)

According to inequalities (13), (14), (16), and (25), we get

$$E(m/m_u^0|x, F, K)$$

$$\geq \sum p(f_1', k_1'|x, F, K) \times \sum p(k_1', k_2', ..., k_u'|f_1', k_1'|x, F, K) \times$$

$$x \times (\log_2 f_{\text{max}} / f_1' - 1) * ((k_1')^2) / (2 \times \log_2 e \times k_{\text{max}} \times f_{\text{max}}) \times (2 \times k_{\text{max}} / f_1')$$

(26)

For the above formula has nothing to do with $k_1', k_2', ..., k_u'$, we have

$$\sum p(k_1', k_2', ..., k_u'|f_1', k_1'|x, F, K) = 1.$$  

Thus, inequality (26) can be reduced to

$$= \sum p(f_1', k_1'|x, F, K) \times x \times (\log_2 f_{\text{max}} / f_1' - 1) * (k_1')^2 / (4 \times \log_2 e \times k_{\text{max}} \times f_{\text{max}} / f_1')$$

(27)

As $f_1'$ and $k_1'$ are independent, we can reduce inequality (27) to equality (28).

$$= \sum_{f_1'=1}^{f_{\text{max}}} p(f_1'|x, F, K) \times \sum_{k_1'=1}^{k_{\text{max}}} p(k_1'|x, F, K) \times$$

$$x \times (\log_2 f_{\text{max}} / f_1' - 1) * (k_1')^2 / (4 \times \log_2 e \times k_{\text{max}} \times f_{\text{max}} / f_1')$$

(28)

Since $p(k_1'|x, F, K) = 1/k_{\text{max}}$, equality (28) can be reduced to

$$= \sum_{f_1'=1}^{f_{\text{max}}} p(f_1'|x, F, K) \times x \times (\log_2 f_{\text{max}} / f_1' - 1) *$$

$$\sum_{k_1'=1}^{k_{\text{max}}} (k_1')^2 / (4 \times \log_2 e \times k_{\text{max}} \times f_{\text{max}} / f_1')$$

$$\geq \sum_{f_1'=1}^{f_{\text{max}}} p(f_1'|x, F, K) \times x \times (\log_2 f_{\text{max}} / f_1' - 1) / (12 \times \log_2 e \times f_{\text{max}} / f_1')$$

(29)

We shrink the range of $f_1'$. We handle it in two cases. When $f_{\text{max}} / 2x \geq 1$, we reduce inequality (29) to

$$\geq \sum_{f_1'=f_{\text{max}}/2x}^{f_{\text{max}}} p(f_1'|x, F, K) \times x \times (\log_2 f_{\text{max}} / f_1' - 1) / (12 \times \log_2 e \times f_{\text{max}} / f_1')$$

$$= (p(f_1' \geq f_{\text{max}} / 2x) - p(f_1' \geq f_{\text{max}} / x)) \times (\log_2 x - 1) / (24 \times \log_2 e)$$

$$= ((1/e)^{1/2} - 1/e) \times (\log_2 x - 1) / (24 \times \log_2 e)$$

$$= \Omega(\log_2 x)$$

When $f_{\text{max}} / 2x < 1$, as $f_1' \geq 1$, we have $f_{\text{max}} / f_1' < 2x$. Thus,

$$(\log_2 f_{\text{max}} / f_1' - 1) / (12 \times \log_2 e \times f_{\text{max}} / f_1')$$

$$> (\log_2 2x - 1) / (12 \times \log_2 e \times 2x)$$

Then, inequality (29) can be reduced to

$$\geq \sum_{f_1'=1}^{f_{\text{max}}/x} p(f_1'|x, F, K) \times x \times (\log_2 2x - 1) / (12 \times \log_2 e \times 2x)$$

$$= (p(f_1' \geq f_{\text{max}} / 2x) - p(f_1' \geq f_{\text{max}} / x)) \times (\log_2 2x - 1) / (24 \times \log_2 e)$$

$$= ((1/e)^{1/2} - 1/e) \times (\log_2 2x - 1) / (24 \times \log_2 e)$$

$$= \Omega(\log_2 x)$$

Fig. 18. The area of $A_i$ and $A$ where $1 \leq i \leq u$.
The analysis shows that the cost-per-unit time of no sharing is $\Omega(\log_2 x)$ times that of DP in average, where $x$ is the original number of queries in the system.

**Proof of Lemma 5.1**

Proof: Let $t_{i-1}^1$ be the first time to execute $G_{i-1}$ in the execution with $m_{i-1}$ where $t_{i-1}^1 \leq f_i$. As $f_{i-1} < f_i$ and $t_{i-1}^1 \leq f_{i-1}$, we have $t_{i-1}^1 < f_{i-1}$. Then, $f_i/t_{i-1}^1 \geq 1$ (Note that “$/$” in the formula is integer division operator). For $t_i^1 = (f_i/t_{i-1}^1)^* t_{i-1}^1$, thus, $t_i^1 \geq t_{i-1}^1$. We know that $f_i - (f_i/t_{i-1}^1)^* t_{i-1}^1 \leq t_{i-1}^1$, that is $f_i - f_i^1 \leq t_{i-1}^1$. We have $t_i^1 \geq f_i - t_{i-1}^1$.

As $t_i^1 \geq t_{i-1}^1$ and $t_i^1 \geq f_i - t_{i-1}$, we have $2t_i^1 \geq t_{i-1}^1 + f_i - t_{i-1} = f_i$. Thus, $t_i^1 \geq f_i/2$.

**Proof of Theorem 5.1**

Proof: According to Equation (11), we know that $G\text{Cost}(t_i^1) = G\text{Cost}(t_i^1 - 1) + k_i^{\text{max}}$. Thus, we have

$$m_i = (G\text{Cost}(t_i^1 - 1) + k_i^{\text{max}})/t_i^1 \quad (30)$$

The total cost in $[0, t_{i-1}^1]$ is $G\text{Cost}(t_{i-1}^1) = m_{i-1} \cdot t_{i-1}^1$.

The total cost in $[0, t_i^1 - 1]$ is:

$$G\text{Cost}(t_i^1 - 1) = t_i^1/t_{i-1}^1 \ast G\text{Cost}(t_{i-1}^1) - k_i^{\text{max}}$$

$$= t_i^1/t_{i-1}^1 \ast m_{i-1} \cdot t_{i-1}^1 - k_i^{\text{max}}$$

$$= t_i^1 \cdot m_{i-1} - k_i^{\text{max}}$$

Replacing $G\text{Cost}(t_{i-1}^1)$ with $t_i^1 \ast m_{i-1} - k_i^{\text{max}}$ in Equation (30),

$$m_i = (t_i^1 \ast m_{i-1} + k_i^{\text{max}} - k_i^{\text{max}})/t_i^1 = m_{i-1} + (k_i^{\text{max}} - k_i^{\text{max}})/t_i^1 \quad (31)$$

From Lemma 8.1, we have $t_i^1 \geq f_i/2$. Thus, we get

$$m_i \leq m_{i-1} + 2(k_i^{\text{max}} - k_i^{\text{max}})/f_i \quad (32)$$

**Proof of Theorem 5.2**

Proof: If we prove that $\forall t, 1 \leq t \leq f_i$, 

$$m_i \leq 2 \ast (D\text{Cost}(i, t - 1) + k_i^{\text{max}})/t \quad (33)$$

then the theorem is proven. We use induction on $i$ where $1 \leq i \leq g$. The basis for the induction, when $i = 1$, is to verify that inequality (31) holds. As discussed in Section 8, $m_1 = k_1^{\text{max}}/f_1$. We also know that $\forall t, 1 \leq t \leq f_1$, $D\text{Cost}(1, t - 1) = 0$. Thus,

$$(D\text{Cost}(1, t - 1) + k_1^{\text{max}})/t = k_1^{\text{max}}/t \geq k_1^{\text{max}}/f_1 = m_i$$

That is, $\forall t, 1 \leq t \leq f_1$, we know

$$m_1 \leq (D\text{Cost}(1, t - 1) + k_1^{\text{max}})/t \leq 2s(D\text{Cost}(1, t - 1) + k_1^{\text{max}})/t$$

For the induction step, suppose $\forall t, 1 \leq t \leq f_{i-1}$, we have $m_{i-1} \leq 2 \ast (D\text{Cost}(i - 1, t - 1) + k_{i-1}^{\text{max}})/t$. That is, $\forall t, 1 \leq t \leq f_{i-1}$.

$$D\text{Cost}(i - 1, t - 1) + k_{i-1}^{\text{max}} \geq m_{i-1} \ast t/2 \quad (32)$$

As discussed earlier in Section 8, when $1 \leq t \leq f_i$, we have

$$D\text{Cost}(i, t - 1) = D\text{Cost}(i - 1, t - 1) \quad (33)$$

If we want to prove $\forall t, 1 \leq t \leq f_i$, 

$$m_i \leq 2 \ast (D\text{Cost}(i, t - 1) + k_i^{\text{max}})/t \quad (34)$$

we should prove

$$m_i \leq 2 \ast (D\text{Cost}(i - 1, t - 1) + k_{i-1}^{\text{max}})/t \quad (35)$$

according to equation (33) and inequality (34). From the definition of $D\text{Cost}(i - 1, t - 1)$, we know

$$D\text{Cost}(i - 1, t - 1) + k_i^{\text{max}} = \min(\sum D\text{Cost}(i - 1, t_{j-1}) + k_{i-1}^{\text{max}}) \quad (36)$$

where $\forall t, 1 \leq t \leq f_i, \forall t, 1 \leq t_j \leq f_{i-1}$ and $\sum t_j = t$. Replacing $D\text{Cost}(i - 1, t - 1)$ in inequality (35) with the formula in equation (36), our goal is to prove

$$m_i \leq 2s(\sum D\text{Cost}(i - 1, t_{j-1}) + k_{i-1}^{\text{max}} - k_{i-1}^{\text{max}} + k_{i-1}^{\text{max}})/t \quad (37)$$

According to inequality (32), we have

$$m_{i-1} \ast t_j/2 \leq D\text{Cost}(i - 1, t_{j-1} + k_{i-1}^{\text{max}} \quad (38)$$

for each $1 \leq t_j \leq f_{i-1}$. Thus, if we want to prove inequality (37), we should prove

$$m_i \leq 2s(\sum m_{i-1} \ast t_j/2 - k_{i-1}^{\text{max}} + k_{i}^{\text{max}})/t \quad (39)$$

As $1 \leq t \leq f_i$, we should prove

$$m_i \leq m_{i-1} + 2 \ast (k_i^{\text{max}} - k_{i-1}^{\text{max}})/f_i \quad (40)$$

to achieve our original goal (34). From Theorem 8.1, Inequality (40) is proven. Then, our original goal (34) is achieved. That is, $\forall t, 1 \leq t \leq f_i$,

$$m_i \leq 2 \ast (D\text{Cost}(i, t - 1) + k_i^{\text{max}})/t \quad (34)$$

The results on real data sets for multiple streams are shown in Figure 19, Figure 20, Figure 21 and Figure 22. The trends are the same with those of synthetic data set for multiple streams.
Fig. 19. The performance of DP, GA, INCO, IN and NS with different ranges of $N$

Fig. 20. The performance of DP, GA, INCO, IN and NS with different ranges of $w$

Fig. 21. The performance of DP, GA, INCO, IN and NS with different ranges of $x$

Fig. 22. The performance of DP, GA, INCO, IN and NS with different ranges of $k$