Appendix

Proposition 1. *A time series T stored in multiple files and segments can be obtained by*

$$T = \bigcup_g S_g = \bigcup_g \left(\biguplus_h S_g^{(h)} \right).$$

Proposition 2 (Window statistics)**.** *For a window with segments, $W = S_1 \cup \cdots \cup S_\omega$, its statistics on buckets can be aggregated as*

$$|B[u]| = \sum_{g=1}^{\omega} |B_g[u]|,$$

where $|B_g[u]|$ is the size of the u-th bucket in segment S_g.

Proposition 3. *For a point p in the bucket $B[u]$, if $\lceil r/\gamma \rceil - r/\gamma < \frac{1}{2}$, then the count of its r-neighbors has an upper bound*

$$|N(p,r)| \le \sum_{i=u-\lambda}^{u+\lambda} |B[i]|,$$

where $\lambda = \lceil r/\gamma \rceil$.

Proof. For a point p in bucket $B[u]$, its r-neighbors must locate in $[p.v - r, p.v + r]$. Referring to Section IV-A, the points in bucket $B[u]$ must fall into interval $[u\gamma, (u+1)\gamma)$, i.e., $u\gamma \le p.v < (u+1)\gamma$. Since $(\lambda - \frac{1}{2})\gamma < r \le \lambda\gamma$, the following inequality can be derived.

$$(u-\lambda)\gamma \le p.v - r < (u - \lambda + \frac{3}{2})\gamma$$
$$(u + \lambda - \frac{1}{2})\gamma < p.v + r < (u + \lambda + 1)\gamma$$

That is, $N(p,r) \subseteq [(u-\lambda)\gamma, (u+\lambda+1)\gamma)$, which leads to the upper bound $|N(p,r)| \le \sum_{i=u-\lambda}^{u+\lambda} |B[i]|$. ☐

Proposition 4. *For a point p in the bucket $B[u]$, if $\lceil r/\gamma \rceil - r/\gamma \ge \frac{1}{2}$, then the count of its r-neighbors has an upper bound*

$$|N(p,r)| \le \max \left(\sum_{i=u-\lambda}^{u+\lambda-1} |B[i]|, \sum_{i=u-\lambda+1}^{u+\lambda} |B[i]| \right),$$

where $\lambda = \lceil r/\gamma \rceil$.

Proof. Similar to the proof of Proposition 3, we also have $u\gamma \le p.v < (u+1)\gamma$. Since $(\lambda - 1)\gamma < r \le (\lambda - \frac{1}{2})\gamma$, the following inequality can be derived.

$$(u - \lambda + \frac{1}{2})\gamma \le p.v - r < (u - \lambda + 2)\gamma$$

Given $(2\lambda - 2)\gamma < 2r \le (2\lambda - 1)\gamma$, $N(p,r)$ at most overlaps with 2λ buckets. Thus, for the case of $p.x - r \in B[u-\lambda]$, it has $|N(p,r)| \le \sum_{i=u-\lambda}^{u+\lambda-1} |B[i]|$. For the case of $p.x - r \in B[u-\lambda+1]$, it has $|N(p,r)| \le \sum_{i=u-\lambda+1}^{u+\lambda} |B[i]|$. Combining the above cases, the upper bound can be obtained as the maximum, i.e.,

$$\max \left(\sum_{i=u-\lambda}^{u+\lambda-1} |B[i]|, \sum_{i=u-\lambda+1}^{u+\lambda} |B[i]| \right).$$

☐

Proposition 5. *For a point p in the bucket $B[u]$, if $r \ge \gamma$, then the count of its r-neighbors has a lower bound*

$$|N(p,r)| \ge \sum_{i=u-\ell+1}^{u+\ell-1} |B[i]|,$$

where $\ell = \lfloor r/\gamma \rfloor$.

Proof. For a point p in bucket $B[u]$, its r-neighbors must locate in $[p.v - r, p.v + r]$. Referring to Section IV-A, the points in bucket $B[u]$ must fall into interval $[u\gamma, (u+1)\gamma)$, i.e., $u\gamma \le p.v < (u+1)\gamma$. Since $\ell\gamma \le r < (\ell+1)\gamma$, the following inequality can be derived.

$$(u - \ell - 1)\gamma < p.v - r < (u - \ell + 1)\gamma$$
$$(u + \ell)\gamma \le p.v + r < (u + \ell + 2)\gamma$$

It follows $[(u-\ell+1)\gamma, (u+\ell)\gamma) \subseteq N(p,r)$, leading to the lower bound $|N(p,r)| \ge \sum_{i=u-\ell+1}^{u+\ell-1} |B[i]|$.

☐

Proposition 6. *For a point p in the bucket $B[u]$, the count of its r-neighbors can be computed by loading data points in at most 4 additional buckets, having*

$$|N(p,r)| = \sum_{i=u-\ell+1}^{u+\ell-1} |B[i]| + \left| \left\{ p' \in \bigcup_{i \in \{u \pm \lambda, u \pm \ell\}} B[i] \middle| dist(p,p') \le r \right\} \right|,$$

where $\ell = \lfloor r/\gamma \rfloor \ge 1$ and $\lambda = \lceil r/\gamma \rceil$.

Proof. Referring to Propositions 3 and 5, the r-neighbors of point p in bucket $B[u]$ meet the following inequality.

$$[(u-\ell+1)\gamma, (u+\ell)\gamma) \subseteq N(p,r) \subseteq [(u-\lambda)\gamma, (u+\lambda+1)\gamma)$$

By placing the value intervals with the union sets of buckets, we can further derive

$$\bigcup_{i=u-\ell+1}^{u+\ell-1} B[i] \subseteq N(p,r) \subseteq \bigcup_{i=u-\lambda}^{u+\lambda} B[i].$$

With the lower bound obtained, we only need to further check the buckets in $\bigcup_{i=u-\lambda}^{u+\lambda} B[i]$ but not in $\bigcup_{i=u-\ell+1}^{u+\ell-1} B[i]$, i.e., up to 4 additional buckets $B[u-\lambda] \cup B[u-\ell] \cup B[u+\ell] \cup B[u+\lambda]$. ☐

Proposition 7 (File Statistics)**.** *For a number of files with time ranges overlapping with window W, $\{F_1, \ldots, F_\rho\}$, its statistics on buckets can be bounded by*

$$|\check{B}[u]| = |B^{(\rho)}[u]| \le |B[u]| \le |\hat{B}[u]| = \sum_{h=1}^{\rho} |B^{(h)}[u]|,$$

where $|B^{(h)}[u]|$ is the size of the u-th bucket for window W in file F_h, $|\check{B}[u]|$ and $|\hat{B}[u]|$ denote the lower and upper bounds of bucket size $|B[u]|$ for window W, respectively.

Proof. If all the points in $F_1, \ldots, F_\rho$ do not overwrite each other, then the bucket sizes can be simply aggregated when merging, i.e., the maximum size $\sum_{h=1}^{\rho} |B^{(h)}[u]|$. On the other hand, if all the points in $F_1, \ldots, F_{\rho-1}$ are overwritten by points in F_ρ, then the bucket size $|B[u]|$ should be the same as $|B^{(\rho)}[u]|$, the minimum size. ☐

Proposition 8. *For a point p in the bucket $B[u]$, if $\lceil r/\gamma \rceil - r/\gamma < \frac{1}{2}$, then the count of its r-neighbors has an upper bound*

$$|N(p,r)| \leq \sum_{i=u-\lambda}^{u+\lambda} |B[i]| \leq \sum_{i=u-\lambda}^{u+\lambda} \sum_{h=1}^{\rho} |B^{(h)}[i]|$$

where $\lambda = \lceil r/\gamma \rceil$ and $|B^{(h)}[i]| = \sum_g |B_g^{(h)}[i]|$.

Proof. The proposition can be easily derived by combining Propositions 3 and 7. □

Proposition 9. *For a point p in the bucket $B[u]$, if $\lceil r/\gamma \rceil - r/\gamma \geq \frac{1}{2}$, then the count of its r-neighbors has an upper bound*

$$|N(p,r)| \leq \max \left(\sum_{i=u-\lambda}^{u+\lambda-1} \sum_{h=1}^{\rho} |B^{(h)}[i]|, \sum_{i=u-\lambda+1}^{u+\lambda} \sum_{h=1}^{\rho} |B^{(h)}[i]| \right),$$

where $\lambda = \lceil r/\gamma \rceil$ and $|B^{(h)}[i]| = \sum_g |B_g^{(h)}[i]|$.

Proof. The proposition can be easily derived by combining Propositions 4 and 7. □

Proposition 10. *For a point p in the bucket $B[u]$, if $r \geq \gamma$, then the count of its r-neighbors has a lower bound*

$$|N(p,r)| \geq \sum_{i=u-\ell+1}^{u+\ell-1} |B[i]| \geq \sum_{i=u-\ell+1}^{u+\ell-1} |B^{(\rho)}[i]|,$$

where $\ell = \lfloor r/\gamma \rfloor$ and $|B^{(\rho)}[i]| = \sum_g |B_g^{(\rho)}[i]|$.

Proof. The proposition can be easily derived by combining Propositions 5 and 7. □