**Identify major design space axes that could be explored for your implementation.**

CDC:

|  |  |
| --- | --- |
| **Axis** | Throughput; window size; parallelism; data size; II; memory |
| **Challenge** | Improving the throughput of CDC; applying pipelining strategy |
| **Opportunity** | Generating the chunks at multiple starting points |
| **Continuum** | Range from 1 to *Input\_size* *-Windows\_size* windows |
| **Equation for Benefit** | Running *N* match-ups in parallel: *T(N)=T(1)/N* |
| **Equation for Resources** | Resources(*R*)= *R\*single\_chunk\_resource* |

SHA:

|  |  |
| --- | --- |
| **Axis** | Throughput; chunk size; number of chunks; input size; II; memory |
| **Challenge** | Improving the throughput; optimizing the II |
| **Opportunity** | Parallelism with LZW; pipelining; multiple-threads parallelism |
| **Continuum** | Range from 1 to *number\_of\_chunks* hashes |
| **Equation for Benefit** | *N* threads parallelism: *T(N)=T(1)/N;*  Parallelism with LZW: *T=max(T\_SHA, T\_LZW)* |
| **Equation for Resources** | Resource(*R*)=*R\*single\_threshold\_resource* |

Deduplication:

|  |  |
| --- | --- |
| **Axis** | Throughput; input size; memory; chunk size; hashmap; hashes |
| **Challenge** | Improving the throughput; dependency on SHA256; |
| **Opportunity** | Parallelism |
| **Continuum** | Range from 1 to *number\_of\_chunks* hashes |
| **Equation for Benefit** | Running *N* storing in parallel: *T(N)=T(1)/N* |
| **Equation for Resources** | Resources(*R*)= *R\*single\_storage\_resource* |

LZW:

|  |  |
| --- | --- |
| **Axis** | Throughput; input size; memory; pointers; encoding; II |
| **Challenge** | Improving the throughput; optimizing the II |
| **Opportunity** | Parallelism with SHA; apply parallelism in LZW; pipelining |
| **Continuum** | Range from 1 to input\_size comparisons and memory lookup |
| **Equation for Benefit** | Running *N* comparison and lookup in parallel: *T(N)=T(1)/N* |
| **Equation for Resources** | Resources(*R*)= *R\*(single\_comparison\_resource+single\_lookup\_resource)* |

Communication/integration:

|  |  |
| --- | --- |
| **Axis** | Data rate among operations; input size; memory; hashmap; chunk size; pipeline |
| **Challenge** | Improving the data rate; data dependency |
| **Opportunity** | SHA and LZW may run in parallel |
| **Continuum** | All the data transmission processes |
| **Equation for Benefit** | Parallelism with LZW: *T=max(T\_SHA, T\_LZW)* |
| **Equation for Resources** | Resource = *Resource\_SHA*+*Resource\_LZW* |