

Family name: ..... Given name: .....

Family name: ..... Given name: .....

Family name: ..... Given name: .....

Consider a table stored in HDFS with the following characteristics:

Size (T) = 256MB; |T| = 64 rows; Size (Row) = 4MB

Cols (T) = 4; Size (Col) = 1MB;  $Size(Footer)=0$ ;  $Size(Header)=0$

- a) *Given the configuration parameters, how much space would you need to store it in a horizontal layout (i.e., Avro)?*

$Size (MetaHRow) = 0.1MB$ ;  $Size (MetaHBody) = 0.1MB$

$Size (Avro) =$  .....

- b) *Given the configuration parameters, how much space would you need to store it in a hybrid layout (i.e., Parquet)?*

$Size (MetaYCol) = 0.2MB$ ;  $Size (MetaYRowGroup) = 0.5MB$ ;  $Size (RowGroup) = 8MB$

$Used_{RG}(Hybrid) =$  .....

$Size(Parquet) =$  .....

Family name: ..... Given name: .....

Family name: ..... Given name: .....

Family name: ..... Given name: .....

Consider a table stored in HDFS with the following characteristics:

Size (T) = 256MB; |T| = 64 rows; Size (Row) = 4MB

Cols (T) = 4; Size (Col) = 1MB; *Size(Footer)=0; Size(Header)=0*

- c) *Given the configuration parameters, how much data would you need to retrieve it in a hybrid layout (i.e., Parquet) to project two columns?*

*Size (MetaYCol) = 0.2MB; Size (MetaYRowGroup) = 0.5MB; Size (RowGroup) = 8MB*

$Used_{RG}(Hybrid) =$  .....

$Used_{rows} (RowGroup) =$  .....

Size (RefCols) = .....

Size (Project<sub>Parquet</sub>) = .....

Family name: ..... Given name: .....

Family name: ..... Given name: .....

Family name: ..... Given name: .....

Consider a table stored in HDFS with the following characteristics:

Size (T) = 256MB; |T| = 64 rows; Size (Row) = 4MB

Cols (T) = 4; Size (Col) = 1MB; *Size(Footer)=0; Size(Header)=0*

- d) *Given the configuration parameters, how much data would you need to retrieve it in a hybrid layout (i.e., Parquet) to select one row if the table is not sorted?*

*Size (MetaYCol) = 0.2MB; Size (MetaYRowGroup) = 0.5MB; Size (RowGroup) = 8MB*

$Used_{RG}(Hybrid) =$  .....

$Used_{rows} (RowGroup) =$  .....

$P(RGSelected) =$  .....

$UsedRG(Select_{Parquet}) =$  .....

$Size (Select_{Parquet}) =$  .....