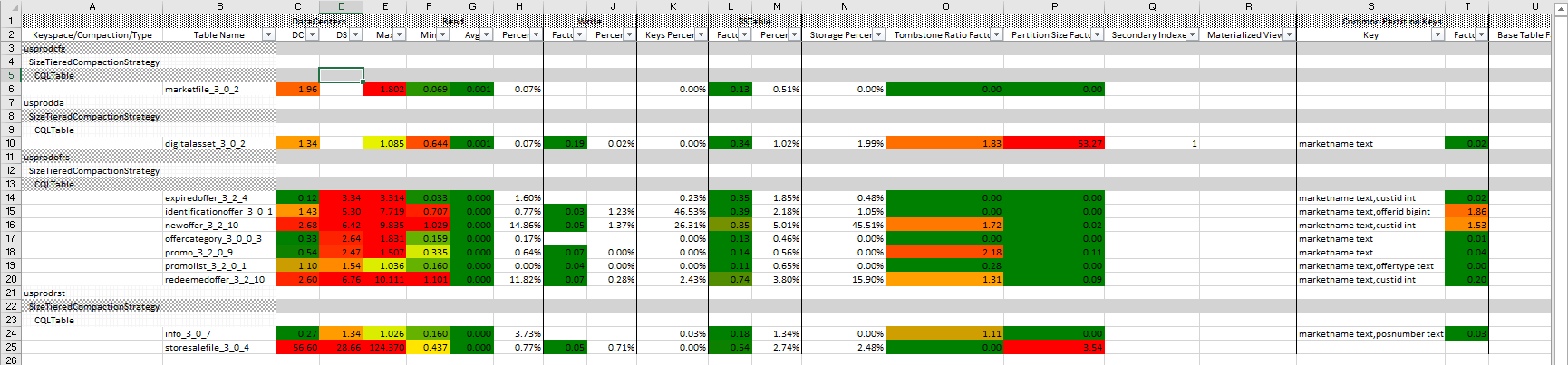
# Release Notes on the Upgraded Health Assessment Application (4.0)

1. New weighted contributing factor engine. This new “factor” will provide insights on how likely or probable an issue(s) and/or contributing issue(s) is. This “factor” is NOT a percent of probability but a factor from 0 and up. A value under 0.75 is typically not seen as a probability cause of an issue or contributing factor. A value above one, is considered more probable.   
     
   There are six new worksheets that utilize the new weighted contributing factors. These new worksheets are:
   * CommonPartitionKey
   * Read Latency
   * Write Latency
   * SSTables
   * Tombstones
   * Large Partition Wide Rows

These new worksheets (except CommonPartitionKey) will only provide tables (or indexes, views, etc.) that have probable issue(s) with their contributing factors. These will greatly reduce the number of worksheets that needed to be reviewed and interrupted in prior workbooks. The prior worksheets with their vast amount of detailed data are still present for additional review.   
  
Below is an example of a “Read Latency” worksheet. All tables (or views, indexes, etc.) listed in this worksheet have a probable read latency issues. All the new “factor” worksheets are similar. Each cell is color coded to provide an understanding of how probable the factor is. If a cell is empty, that indicates that the factor does not apply. If green, that factor is not probable. Yellow to Red indicates a higher degree of probability. There will be also a factor value. This value will give an idea about how probable the issue/contributing factor is. Care must be considered when comparing these values. For example, if a contributing factor is a “5” for tombstones and a “10” for SSTables (note that each of these influence the read latency). This does not mean that SSTables are twice the influence as compared to tombstones. But it does mean that SSTables is more probable then tombstones as influencing the read latency for that table.   
  
Also, each worksheet will have a “Max”, “Min”, “Avg” columns. These columns represent the cluster wide attribute of that table (e.g., for read latency this will be the maximum, minimal, and average of the read latency for that table in milliseconds). These columns are color coded represents how “far” they are over the typical guidelines. They range from green (good), to yellow (borderline), to red (over guidelines). There will also be a “percent” column (e.g., columns H, K, M, etc.). This column represents the overall (cluster wide) activity, size, storage, etc. of the associated table. Belwo is an example of the “Read Latency” worksheet:  


* Column “Keyspace/Compaction/Type” (column A) – This will consist of the keyspace name, compaction strategy, CQL type (Table, Index, Materialized View, etc.)
* Column “Table Name” (column B) – This will be the name of the CQL Item (e.g., table name, index name, etc.)
* Column group “Datacenters” (columns C and D, one column for each datacenter) -- There will be one column for each datacenter in the cluster. The header cell will be an abbreviation of the datacenter name. The comment associated to the cell will be the full name. In this example there are two DCs (abbreviated “DC” => “DC1-Cassandra”, “DS” => “DC1-Spark”). The cell values (e.g., cell C6, C14, D14, etc.) represent how probable a DC is having an issue with the associated table (e.g., A6, A14, etc.). In this example, read latencies within each DC for that table. These values are factors.
* Column group “Read” (columns E to H) – This grouping represents the maximum, minimal, and average (cluster wide) with read latencies in milliseconds for these tables. This grouping also includes the percent of read activity for these tables cluster wide.
* Column group “Write” (e.g., columns I and J) – This grouping represents the write latency factor and the overall write activity of the cluster.
* Column “Key Percent” (column K) – This column is the percent of keys this table represents cluster wide.
* Column group “SSTable” (column L and M) – This grouping is a possible contributing factor to poor read latency. The factor column (column L) represents how probable SSTables are contributing to poor read latencies. The percent column (column M) is the percent of SSTables that this table makes up for all SSTables in the cluster.
* Column “Storage Percent” (column N) – This column represents the storage utilization of the associated table cluster wide.
* Column “Tombstone Ratio Factor” (column O) -- This grouping is a possible contributing factor to poor read latency. The factor column represents how probable tombstones are contributing to poor read latencies.
* Column “Partition Size Factor” (column P) -- This grouping is a possible contributing factor to poor read latency. The factor column represents how probable large partitions are contributing to poor read latencies.
* Columns “Secondary Indexes” and “Materialized Views” (columns Q and R) – These are the number of indexes and/or materialized views associated with this table.
* Column group “Common Partition Keys” (columns S and T) -- This grouping is a possible contributing factor to poor read latency. The factor column represents (column T) how probable hot spotting of a “common partition key” is contributing to poor read latencies.
* Column “Base Table Factor” (column U) – This column is only present when the associated item (column B) has an associated base table. This is the factor value of the base table.
* Interpretation of the read latency example by table (BTW, the example workbook is simulated data to produce different edge cases for testing):
  + marketfile\_3\_0\_2 – DC1\_Cassandra is orange, there is a degree of probability that this table is having a read latency issue. DC1\_Spark either has no data to determine the factor or is not replicated to this DC (the DC factor cell is blank). All contributing factors are green (no probability) or blank. This is an “edge case” where read latency is underdetermined. After review of the “TaggedTableDetails” worksheetone node in DC1\_Cassandra had “super high” read latencies.
  + digitalasset\_3\_0\_2 -- DC1\_Cassandra is orange, there is a degree of probability that this table is having a read latency issue. DC1\_Spark either has no data to determine the factor or is not replicated to this DC (the DC factor cell is blank). This is another “edge case” (missing read latency data for three nodes in that DC). There are two contributing factors making that are contributing tombstones and large partitions. Large partitions seem to be influencing read latencies “more” than tombstones.
  + newoffer\_3\_2\_10 – Both DC have a higher degree of possible read latency issues where tombstones seem to be influencing read latencies “more” than hot spotting.

1. Two new factor-based Excel tables (worksheets TaggedTableDetails and TaggedTables) used to provide additional detail around tables, indexes, views, etc. This detailed weight factor values can be used down to a table-node review.
2. New collection of worksheets around flushing, compactions, schema changings, sharding, SSTables, tombstones, large partitions, non-DSE events, back-to-back GCs, commit log pauses, etc.
3. New log-based rules for DSE 6.x
4. Many new features and bug fixes
5. New health assessment report format based on the new factor based analysis