Hello,

First of all, I am very glad that you liked my application.

So, here are my answers/solutions :

- **Task1:**
  Input data should be stored in Simple Storage Service (S3). It is easy to access and security (access key) can be configured. For data processing we can use amazon cluster, which will be optimized according to a job we perform. In this particular situation I would suggest two types of clusters, one for processing raw data (used for job processing) and making database tables which should contain (amongst job_id) timestamp for each column, and the other for creating the daily report (making .CSV file). Processing cluster should put timestamp from a log into the DB table (start/end time of a job), and the reporting cluster should be able to read that timestamps so in that way we can limit our .CSV for one day, or hour,.. etc. Also reporting cluster should be able to sort these reports by timestamp. Every cluster should shut down after his job is done to save money. Our outputs (.CSV) should be stored on S3 also, and this can allow sync on some external terminal server (customer for example). Cluster number can be scalable and it is a cloud solution, so we don't have to worry about or connection or something like that, and AWS is good platform for big data, because it is robust, and secure.
- **Task2:**
  a: select sum(AP Amount (£)) , Department family from <table> group by Department family
  b: select sum(AP Amount (£)) , Department family, Expense Type from <table> group by Department family, Expense Type order by Department family
- **Task3:**
  I'll use some pseudocode

  Create table <new table> as
  select expense type, expense area, AP amount from <old table>

  ;

  Select *

  From

  (

  Select expense type, expense area from <new table>

) src

  Pivot

  (

  Sum(AP amount)

  For (expense area)

) piv;