RL for OE Project plan

To do:

* Update done + action function (execute inventory fully at the end of order)
* Split time down to different orders

Configurations/Tweaks to consider after running:

* No of orders and time points within an order: scaling in comparison to the paper
  + Paper: One order = 3600 data points (60 seconds x 60 mins = 1hour). No of orders = 750 (1 year = 250 days x 3 orders per day)
* Adjust alpha in our reward function to match our time frames of execution?