This summary explains the method used to assign the clusters to the scenarios listed. These scenarios are:

- High load rate during peak hours
- Shut down of generator for maintenance
- Low load rate during night
- Disconnection of a line for maintenance

Initially, it was expected that these were general scenarios corresponding to multiple different precipitating events, i.e. different generators offline or different lines disconnected. This would have meant that four clusters (for the four states) may not have been the optimal number of clusters. However, using the Java program produced for the assignment, it was found that the cluster number elbow point clusters was indeed 4. Further examining the data using the "data export" tool in the program, it was seen that all the states within the clusters were very similar, lending weight to the existence of only 4 of the same event (i.e. same line, same generator).
To determine the scenarios some assumptions need to be made regarding how the data was generated. It appears the data was determined using a load flow calculation with some randomisation. Further interrogation :

- Generator 2 and 3 Buses are PU buses which don't change generate power on the basis of operating state.
- Generator 1 Bus is the slack bus which will either produce or consumer power depending on operating state.
- The remaining buses are modelled as PQ buses.
- An increase or decrease in load is modelled through a factor change in all the loads.

While impedance values for the network are not available, the DC power flow approximation [1] can be used to gain some insight into the flows occuring within the network. Since network resistance is typically low in transmission lines and for small angles $\sin(\theta_k - \theta_j) \approx (\theta_k - \theta_j)$, the approximation states the general power flow equation can be reduced as so:

$$P_k = \sum_{j=1}^{N} |V_k||V_j|(G_{kj}cos(\theta_k - \theta_j) + B_{kj}sin(\theta_k - \theta_j)) \tag{1}$$

$$\Rightarrow P_k \propto (\theta_k - \theta_j) \tag{2}$$

The centroids of the four clusters (note that the numbers will change in the program depending on the randomised initialisation points but the clusters are consistent):

	CHAR	AMHE	WINL	BOWM	TROY	MAPL	GRAN	WAUT	CROSS
A	1	1	1	0.991	0.994	1.014	1.011	1.01	0.983
B	1	1	1	0.972	0.945	1.006	0.982	0.989	0.935
C	1	1	1	0.99	0.977	1	0.983	0.995	0.962
D	1	1	1	0.896	0.85	0.954	0.904	0.938	0.81

Table 1: Finalised centroid voltage magnitudes in pu

	CHAR	AMHE	WINL	BOWM	TROY	MAPL	GRAN	WAUT	CROSS
A	0	24.605	19.358	4.218	7.455	16.543	17.036	18.817	7.721
B	0	18.54	19.677	-2.699	-7.827	16.838	11.989	12.63	-1.736
C	0	1.832	-9.241	-5.288	-9.831	-9.241	-8.619	-4.042	-8.968
D	0	-20.756	-25.103	-15.215	-27.518	-28.096	-32.915	-26.989	-29.861

Table 2: Finalised centroid voltage angles in degrees

From the topology of the system, we can determine the voltage angle difference along the lines for each cluster:

Line	A	B	C	D
1-4	-4.218	2.699	5.288	15.215
2-8	5.789	5.91	5.874	6.234
3-6	2.815	2.839	0	2.992
4-5	-3.237	5.128	4.542	12.302
4-9	-3.503	-0.963	3.68	14.646
5-6	-9.088	-24.665	-0.59	0.578
6-7	-0.492	4.849	-0.622	4.819
7-8	-1.781	-0.641	-4.576	-5.925
8-9	11.096	14.365	4.926	2.872

Table 3: Angle difference along the lines in degrees (source-destination)

Using the approximations from Table 3, two of the scenarios can be immediately observed:

- For the **low load case**, the slack bus must be consuming power, given the other generation is assumed to be constant. If the slack bus is consuming power the flow will be from Bus 4 to Bus 1, which occurs for the data in **Centroid A**. With a bit more difficulty, it is also possible to see that there is less power being consumed by the load buses. For example, the flow along lines 6-7 and 8-9 result in proportionally less net flow than in the other clusters.
- For the **generator disconnected case**, there must be zero power flow from the generator. This occurs for Generator 3 for the data in **Centroid C**, where there is no flow along Line 3-6. The flow along Line 1-4 shows the additional power being produced by Generator 1 to compensate.

The remaining two scenarios can be split by examining the flows on Line 6-7 and 7-8. The flow on 6-7 is similar in both cases, however, the flow on 7-8 is greater for Centroid D. Additionally, the flow out of Bus 6 on Line 6-7 and into Bus 6 on Line 3-6 are approximately equal in both Centroid B and D data, however, the flow on Line 5-6 is disproportionally great for Centroid B. This suggests that the disconnected line is Line 5-6 and this leads to a reasonable interpretation of the remaining flows in the circuit (i.e. Generator 1 supplying the majority of the power to Load 5). Hence, the **line disconnected case** corresponds to **Centroid B** and the **high load case** to **Centroid D**.

Another interpretation can be made based on only the flows from the generator. From the assignment specification, the total load is 315MW and total power supplied from Gen 2 and 3 is 248 MW. This means that Gen 1 must supply some power in the base case. When Gen 3 is offline, Gen 1 is covering both the normal Gen 1 load and the Gen 3 shortfall. Therefore, the flow from Gen 1 must be less for the line disconnected case than for the generator disconnected case.

In the program, it is the slack bus power flow described above that has been used to generate the classification. If the power flow (voltage angle difference) is negative it corresponds to a low load case, if it is between 0 and 3.5 degrees it is a line disconnected case, if it is between 3.5 and 10 degrees it is a generator disconnected case and if greater than 10 degrees then it is a high load case. This allows the cases to be generalised for multiple clusters (different causes) as long only the four scenarios are present.

References

[1] Biggar, D.R. and Hesamzadeh, M.R. (2014) *The Economics of Electricity Markets* Wiley - IEEE