1. As you can see from the graphs above and below comparing PA3 and PA4, that PA4 is slightly faster. The drop to lower execution times is a slight bit steeper. This makes sense since the request are being pushed and handled simultaneously. For PA4 to be slower there would need to be a small enough buffer size to create a bottleneck on how fast requests can be processed.
2. The performance regardless of buffer size is relatively the same, but slightly slower with a smaller buffer size. This makes sense because the buffer size acts as a secondary bottleneck for being able to process requests.