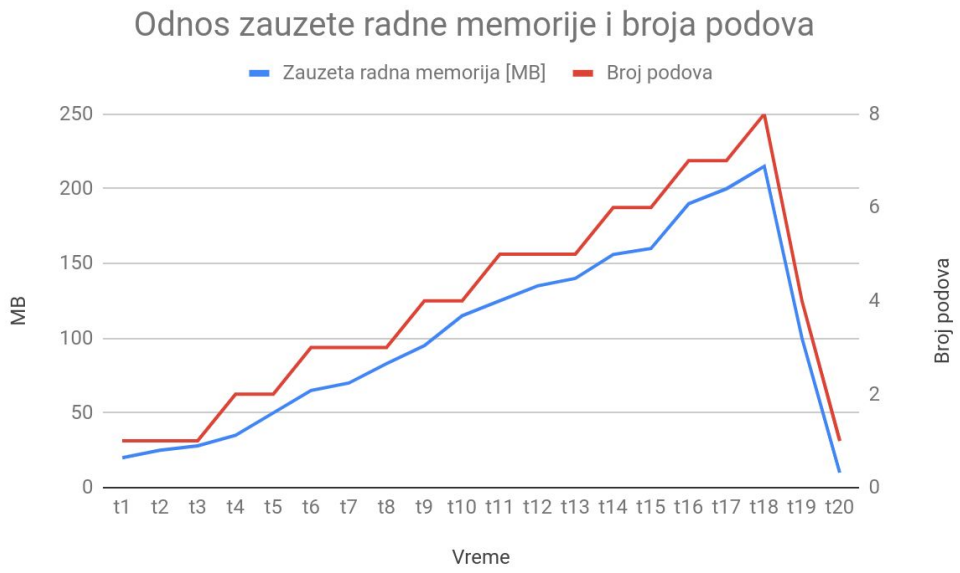
Potrebno je kreirati *web servis* koji implementira jednu GET metodu:

1. http://host:port/metod1?kolicinaMemorije=10 - Zauzeti onoliko MB radne memorije kolika je vrednost parametra ***kolicinaMemorije***. Kada se zahtev obradi, potrebno je oslobotiti zauzetu memoriju.

Web servis je potrebno hostovati na Kubernetesu i podesiti automatsko horizontalno skaliranje, između jedne i osam instanci *web servisa*, prema prosečnoj upotrebi radne memorije od 30 MB.

Za potrebe testiranja autoskaliranja je potrebno napisati program koji će simulirati pozive prema *web servisu*, a zatim napraviti grfik koji u vremenu prati zauzetost radne memorije i broj podova koji je Kubernetes podigao. Kada Kubernetes pokrene 8 instanci, potrebno je zaustaviti slanje zahteva i zabeležiti vreme kada je broj podova smanjen na 1. Primer jednog grafika je dat u nastavku.



Program koji šalje zahteve prema *web servisu*, prati promenu količine memorije i broja podignutih podova i upisuje ih u CSV fajl, na osnovu koga je moguće kreirati gore prikazani grafik (Korisna naredba: **kubectl get hpa**).