

# Virtualizált környezetek teljesítményanalízise intelligens adatfeldolgozási módszerek alkalmazásával

---

Az e félévi önálló laboratórium során a virtualizált adatközpontok teljesítménymenedzsmentjének egy új aspektusával ismerkedhettem meg, különös gondot fordítva a virtuális gépek terhelésének becslésére. Feladatomban egy valós virtualizált infrastruktúrából monitorozással kinyert nagyméretű adathalmaz feldolgozása volt oly módon, hogy az intelligens módszerek segítségével kinyert információ a magas szintű kapacitástervezés támogatására kell hogy szolgáljon.

Mivel a virtuális gépekkel kapcsolatos adatfeldolgozás első néhány lépését (adattisztítás, aggregátumok előállítás) szakmai gyakorlatomban az R matematikai keretrendszerben végeztem, emiatt a virtuális gépek viselkedésének vizsgálata is elsősorban R szkriptek formájában jelent meg, jelen beszámoló további részeiben néhány jellemző kódrészlet bemutatásra is kerül.

A kapacitástervezés feladata, hogy egy, már felépített infrastruktúra további bővíthetőségét legalább becsülni tudjuk. Az infrastruktúra ugyanakkor 40-50 fizikai hosztot tartalmaz, amelyek együtt kb. 800 virtuális gépet futtatnak. Ilyen méretek mellett lehetetlenné válik a manuálisan elvégzett becslések elvégzése, ezek helyett indokoltak adatbányászati módszerek alkalmazása. Ezek célja, hogy egyfajta virtuális gép típusokat alakítsunk ki. Jelen dolgozat bemutatja, melyek az általam jelenleg ismert lehetőségek erre, ezek hogyan valósíthatók meg R környezetben, milyen problémákkal szembesültem munkám során és milyen eredményeket értem el a tárgyalt módszerek alkalmazásával.