

## 4. gyakorlat – Teljesítménymodellezés

**Dimenzióanalízis.** A teljesítménymodellezés feladatok megoldása során érdemes a fizikából ismert dimenzióanalízist<sup>1</sup> elvégezni. Idézzük fel a négyzetes úttörvény:  $s = v_0 t + \frac{a}{2} t^2$

Dimenziókkal:  $s[\text{m}] = v_0 \left[\frac{\text{m}}{\text{s}}\right] t[\text{s}] + \frac{a}{2} \left[\frac{\text{m}}{\text{s}^2}\right] t^2[\text{s}^2] = v_0 t[\text{m}] + \frac{a}{2} t^2[\text{m}]$

A dimenzióhasználat fő motivációja, hogy ha a dimenziók nem stimmelnek, akkor a képletet is biztosan elrontottuk valahol.<sup>2</sup> A dimenzióanalízis gyakran segít a megfelelő képlet kiválasztásában. Fontos, hogy a „darab”, „kérés” stb. jellegű mértékegységek nem számítanak külön dimenzióknak, ezért pl. a  $\frac{\text{kérés}}{\text{s}}$  és az  $\frac{1}{\text{s}}$  dimenziók megegyeznek.

**Alapképletek.** Little-törvény:  $N = X \cdot T$ ,  $N [1] = X \left[\frac{1}{\text{s}}\right] \cdot T [\text{s}]$

Kihasználtság intuitíven és a Little-törvényből *egyetlen kizárólagos* erőforráspéldány esetén:

$$U = \frac{X}{X_{\max}} = \frac{T_{\text{busy}}}{T_{\text{measured}}} = N = X \cdot T$$

Átbocsátóképesség végrehajtási időből *egyetlen kizárólagos* erőforráspéldány esetén (az átbocsátóképesség az elérhető legnagyobb átbocsátás, vagyis ilyenkor a kihasználtság 100%):  $X = \frac{U}{T} \Rightarrow X_{\max} = \frac{1}{T}$

### 1. Zárthelyi javítása

A zárthelyik megtekintése során a hallgatónak lehetőségük van reklamálni esetleges javítási hibák miatt. Sikeres reklamáció esetén a pontszámuk módosításra kerül. Az első nagyfeladatból (F1) óránként 10 darabot képes átnézni egy javító, a második nagyfeladatból (F2) pedig 20 darabot. Mindkét feladathoz tartozik 1-1 javító, akik az adott feladatot javították. A továbbiakban készítsünk minden kérdéshez egy-egy folyamatmodellt és határozzuk meg, hogy óránként hány hallgató dolgozatát sikerül átnézni az egyes esetekben!

- A hallgatók először az F1, majd az F2 feladatot nézetik át a javítóval.
- A leleményes hallgatók a két feladatot külön-külön már egyszerre két javítónak adják oda, mivel külön lapra voltak írva. Mit nyerünk a párhuzamosítással?
- A nagy tömeg miatt a hallgatók csak az egyik feladatukat nézetik át, mégpedig azt, amelyiknek a javítója éppen szabad.
- Híre ment, hogy a második feladat javítója sokkal kevésbé szigorú, így a hallgatók 80%-a inkább kivárja ennél a javítónál a sort. A maradék 20% a másik javítónál reklamál az első feladattal kapcsolatban.
- A hallgatók 10%-ának a reklamáció után már csak 1-2 pont kellene a jobb jegyhez, ezért újra és újra megpróbálkoznak a reklamációval. Feltételezhetjük, hogy a hallgatók az a) részben leírt reklamációs stratégiát használják.

### 2. Diszk teljesítménye

Egy diszk 50 kérést szolgál ki másodpercenként. Minden kérés kiszolgálása 0,005 másodpercet vesz igénybe. A rendszerben nincs átlapolódás.

- Mekkora a kihasználtság?
- Mekkora a maximálisan kiszolgálható terhelés (érkezési ráta)?

### 3. Szerver teljesítménye

Egy szerveren az alábbi teljesítményjellemzőket mértük:

Mintavétel időpontja [ms]	500	600	700	800	900
Utolsó 100ms alatt feldolgozott kérések száma [darab]	11	12	21	18	20
Utolsó 100ms átlagos kiszolgálási ideje [ms]	15	20	21	25	27
Utolsó 100ms CPU kihasználtság [%]	12	13	16	17	19
Utolsó 100ms HDD I/O kihasználtság [%]	55	63	87	61	73

- A rendelkezésre álló adatok alapján a szerver melyik erőforrása tűnik a szűk keresztmetszetnek?
- Az első mintavétel idején mekkora az átbocsátási ráta értéke? Az 5 mintavétel alapján mekkora az átbocsátási ráta átlaga?
- Ezen 5 mérés alapján milyen becslést tudunk adni az egyszerre kiszolgálás alatt lévő kérések átlagos számára?

<sup>1</sup>Dimenzióanalízis (Wikipédia), <http://hu.wikipedia.org/wiki/Dimenzió>

<sup>2</sup>Ajánlott olvasmány: what if? – Droppings, <http://what-if.xkcd.com/11/>

## 4. Kétrétegű architektúra

Adott egy webszerver (WS) és két fürtözött adatbázisszerver (DB1, DB2). A két adatbázis szerver közt súlyozott round robin terheléelosztás alapján választunk, 1:2 arányban. Minden felhasználói kérés kiszolgálása során mindkét fajta erőforrást használjuk. A csúcsidőszakban 30 percig monitorozzuk a rendszert, ezalatt 9000 kérést szolgál ki. A szerveken mért foglaltsági idők: WS – 1350 s CPU idő; DB1 – 810 s, DB2 – 1320 s diszk IO idő.

- Készítsünk folyamatmodellt a kérések feldolgozásáról a szöveg alapján!
- Mekkora az egyes szerverek jelenlegi átbecsülése?
- Mennyi időt töltenek egy-egy kérés kiszolgálásával a szerverek?
- Mekkora a rendszer maximális átbecsülőképessége?
- Miért nem egyféle foglaltsági időt vettünk figyelembe a két erőforrástípusnál?
- Hol csal még így is a modell?

## 5. Közösségi oldal

Internetes közösségi oldalt működtetünk. Az utóbbi időben számottevően népszerűbb lett az oldal, de ezáltal a válaszidő is kellemetlenül megnőtt. Az üzleti cél, hogy csúcsidőszakban egyszerre 1500 felhasználót átlagosan négy másodperces válaszidővel szolgáljon ki a honlap.

- Minimálisan mekkorára kell tervezni a kiszolgáló infrastruktúra átbecsülőképességét, ha az azon kívüli késleltetés (hálózati forgalom, HTML megjelenítés a kliensoldalon) egy másodpercrek becsülhető?
- Az újratervezett weboldalon a mérések szerint egyetlen kérés kiszolgálása átlagosan 20 ms CPU-időt igényel a webszerveren, és 12,5 ms erejéig foglal le egy adatbázisszerveret. Jelenleg 15 web-szerver fogadja a kéréseket és az adatbázis 5 kiszolgálóra van replikálva. Lineáris skálázhatóságot feltételezve, milyen számítógépből és mennyit kell még legalább venni, hogy a fenti cél teljesülhessen?
- (\*) A kibővített rendszerben mekkora lesz az egyes szerver típusok kihasználtsági aránya? Ha az a cél, hogy még a csúcsidőszakban is legfeljebb 50%-os legyen a kihasználtság, meddig kellene még bővíteni a rendszert?
- Tekintsünk csak 2 db webszervert és 3 db adatbázis szervert. Készítsünk állapot alapú modellt(ek)e(t), amely(ek) az infrastruktúra erőforrásait modellezi(k) az elérhetőségeik (szabad/foglalt) szerint. Milyen tervezői döntésekkel szembesülünk? Mik az egyes lehetőségek előnyei és hátrányai?

## 6. Sziget közlekedési hálózata (\* korábbi zárthelyi feladat)

Egy sziget lakói minden reggel munkába menet átkelnek a szigetet ölelő tavon. Észak felé híd vezet, dél felé autós-komp. Az irányonként egysávos híd 200 m hosszú, és 60 km/h sebességgel szabad rajta haladni, a követési távolság (hátsó lámpától hátsó lámpáig 30 m) betartása mellett. A négy komphajó egyenként 15 percenként teszi meg a sziget-szárazföld-sziget kört, és így óránként négyen együtt legfeljebb 800 autót tudnak átvinni a szárazföldre.

- Mekkora a híd átbecsülőképessége (észak felé)?
- Hány autó fér el egy komppal?
- A reggeli csúcsforgalomban mekkora a szigetet elhagyó két útvonal együttes átbecsülőképessége?
- Ha délben a szárazföldi főutat baleset miatt lezárták, és a szigeten keresztül (a hídon, majd a kompon átkelve) terelik a forgalmat, mekkora a terelőútvonal átbecsülőképessége?
- Valamelyik reggel 7:00 és 8:30 között 900 autó hagyta el a szigetet komppal. Mennyi volt ebben az időszakban a kompok átbecsülése és kihasználtsága?
- A fenti mérésben átlagosan hány autó állt sorba egyszerre a parton, ha az autók jól időzítve, átlagosan fél perccel a beszállásuk előtt érkeztek kompkikötőhöz?

## 7. Tudásbázis (\*)

Vállalatunk nyilvános szakmai tudástára egymásra is hivatkozó szócikket kínál a cég termékeit világszerte használó ügyfeleknek. Egyetlen szócikk lekérésének kiszolgálásához a szervert átlagosan 60 ms-ig veszi igénybe. A szócikk megtekintése után az olvasó csak az esetek 30%-ában hagyja el az oldalt, többnyire ugyanis egy újabb szócikkre mutató hivatkozásra kattint.

- Egy olvasó összes tudásszomjának kielégítéséhez átlagosan mekkora szervertípus szükséges?
- Tekintsük úgy, hogy az egyes kérések a szerveren nem párhuzamosíthatóak. Óránként hány egyedi látogatót képes kiszolgálni a szerver?