

1. Feladat

Adott egy Hadoopban tárolt adatfájl, amely kiskereskedelmi boltok eladási adatait tartalmazza a 2015 és 2016-os évben a következő adatstruktúra szerint:

Date	Time	StoreID	Location	ProductCat	Price	PaymentMethod
20150103	8:12	12389	Budapest	Toys	1250	Cash
20150103	14:27	12389	Budapest	Electronic	89990	Visa
20150103	16:44	2323	Pécs	Food	4500	Mastercard

Tervezze meg a *map*, a *reduce* és *combiner* lépéseket optimális módon, úgy, hogy az alábbi lekérdezések, minél hatékonyabbak legyenek futási időben:

- Összesen hány vásárlás történt és mekkora értékben a 2015-ös évben?
- Mikor és melyik boltban történt a legnagyobb értékű vásárlás?
- Melyik boltban a legnagyobb a szórása az eladott termékek árának?
- Melyik nap volt a legnagyobb a készpénzzel történt vásárlások összege?

2. Feladat

- Ismertesse a Bloom Filter működését. Adjon példát az alkalmazására.
- Ha $n=1000$ elemű halmazba való tartozást szeretnénk vizsgálni, $p=0.05$ false positive valószínűséggel, akkor hogyan válasszuk meg optimálisan a Bloom filter m és k paramétereit, ahol m bites a Bloom filter segéd tömbje és k hash függvényünk van?
- Hogyan tudná úgy kiegészíteni a Bloom Filtert, hogy ne csak a beszúrást, hanem a törlést is támogassa.

3. Feladat

- Ismertesse a weighted average predictor és a Follow the Perturbed Leader algoritmust.
- Az optimális tanulási paraméter (learning rate) értéke a weighted average predictor esetén függ n -től, a predikciós lépések számától. Javasoljon megoldást a tanulási paraméter megválasztására arra az esetre, ha n nem ismert.
- Konstruáljon egy olyan hiba sorozatot, amikor Follow the Leader stratégia nem konzisztens.
- Ismertesse a Multi-Armed Bandit problémát és mutassa be az alapötletét a konzisztens algoritmusnak.