# Big Data

## Célkitűzés

Az előadás célja bemutatni a hallgatóknak milyen mechanizmusok állnak rendelkezésükre Azure-ban nagy mennyiségű strukturált adat feldolgozására, illetve elemzésére. A konkrét eszközök mellett hallgatók betekintést nyerhetnek a gépi tanulás és a Hadoop alapjaiba is.

## Előadás áttekintése

### Motiváció

Az adattárolás költségeinek és a számítási kapacitásunk fajlagos költségeinek folyamatos csökkenése mára már lehetővé teszik, hogy nagy méretű adathalmazokat is költséghatékonyan tudjunk tárolni, feldolgozásukkal pedig olyan következtetéseket vonhassunk le, melyek túlmutatnak az egyszerű legyűjtéseken, vagy statisztikai elemzéseken. Az előadás a determinisztikus elosztott adatfeldolgozó környezetek felől indul, s végül eljut a különböző mesterséges intelligencia alapú megoldásokig.

### HDInsight, Cortana Analyitics, Power BI

Az HDInsight egy Azure alapú Hadoop keretrendszer. Bemutatásának (és a MapReduce folyamat részletes tárgyalásának) elsősorban az a célja, hogy a hallgatók betekintést nyerjenek az elosztott adatfeldolgozó rendszerek egy népszerű architektúrájába, s el tudják képzelni, hogyan lehet lekérdezéseket és kimutatásokat hatékonyan előállítani és futtatni nagyméretű elosztott adathalmazokon.

A Cortana Analytics és a Power BI ezt követően gyakorlatilag már csak alkalmazási példák. Előbbi bemutatja, miként illeszkedhet például egy HDInsight szolgáltatás az alkalmazásunk architektúrájába, míg utóbbi egy általános jelentéskészítő alkalmazás, mely egyebek mellett a korábban bemutatott adatforrásokra is tud építeni.

### Gépi tanulás

A big data témakör egyik érdekes eleme a gépi tanulás. A gépi tanulás alapötletét a neurális hálókon és a szupport vektort gépeken keresztül szemléltetjük. Az alapötlet közös tanulási fázisok során a rendszer valamilyen irányított tanítási megoldással megpróbálja úgy behangolni a saját paramétereit, hogy az a vizsgált adatokat a megadott példaválaszainknak megfelelően osztályozza. Minél nagyobb számítási kapacitásunk van a rendszerhez és minél több adatunk, annál pontosabb eredményeket érhetünk el, ezáltal a felhő kiváló alapokat tud nyújtani gépi tanuló algoritmusok fejlesztéséhez.