

Na základě analýzy datových vstupů přináší datové profilování sumarizaci hodnot, která je vypočtena pomocí různých typů statistik. S růstem velikosti dat uložených v datových zdrojích se však nejedná o výpočetně snadnou operaci. Pro velká data není možné uložit všechny hodnoty v rychlé paměti RAM. Tudíž nelze použít jednodřuchodové a zároveň přesné algoritmy bez využití pomalejšího úložiště počítače. Diplomová práce má za cíl implementovat, porovnat a následně vybrat vhodné aproximativní algoritmy pro profilování objemných dat. Díky aplikaci aproximativního přístupu lze následně omezit paměť potřebnou pro výpočet natolik, že celou jednodřuchodovou analýzu dat je možné spočítat pouze v paměti RAM. Tím se proces profilování dat značně urychlí. Nástroj pak dokáže produkovat výsledky frekvenční analýzy, výpočtů kardinality, kvantilů nebo histogramů a dalších jednosloupcových statistik v krátkém čase s chybou řádově v desetinách procent.