

Datové toky jsou důležitým aspektem dnešního korporátního prostředí. MANTA Flow je platforma pro analýzu datových toků, která funguje na základě extrakce a analýzy zdrojových souborů zákazníka. Často se však stává, že zákazník chce aktualizovat graf datových toků pouze kvůli malé změně v jeho zdrojových souborech. Současně se ale v takovém případě všechny vstupní zdrojové soubory analyzují znovu a většina času se tak promarní analýzou nezměněných souborů. V této práci jsme představili způsob, jak lze analyzátor datových toků vylepšit pomocí inkrementálních updatů tak, aby analyzoval pouze zlomek všech vstupních souborů a přitom stále produkoval správné datové toky.

V rámci práce jsme změnili způsob provádění celé analýzy tím, že jsme změnili granularitu analýzy na mnohem menší části. Také jsme vylepšili mergeovací algoritmus, aby rozpoznal, kdy by nezměněný soubor mohl vygenerovat rozdílný datový tok pomocí nových konceptů, jako jsou zdrojové segmenty, nebo nové akce při odstranění či vytvoření vrcholu. Na základě toho nyní upravený algoritmus analyzuje pouze všechny změněné soubory a několik málo nezměněných souborů, které by mohly vygenerovat odlišný datový tok ve srovnání s předchozí analýzou. Implementovali jsme také prototyp pro MANTA Oracle skener, který využívá těchto nových konceptů. Prototyp byl testován jak na správnost, tak na výkon.