

Nástroje pro integraci dat často používají vložený kód pro manipulaci s daty. Mezi příklady takových nástrojů patří služba pro integraci dat AWS Glue, platforma Databricks, datový cloud Snowflake nebo integrační služby SQL Serveru (SSIS). Vestavěný kód je obvykle psán v programovacích jazycích jako je Python, Java, C# nebo JavaScript. Manta Flow je automatizovaná platforma pro analýzu datových linií v databázových modelech, datových kanálech nástrojů pro integraci dat, a ve zdrojovém kódu aplikací, ale chybí jí schopnost analyzovat vložený kód.

V této práci jsme zhodnotili potenciální způsoby rozšíření možností platformy Manta Flow o schopnost analyzovat datové linie ve vloženém kódu. Vytvořili jsme obecný návrh opakovaně použitelné služby pro vestavěný kód, která využívá stávající potenciál analýzy datových toků ve zdrojovém kódu, a používá jej pro analýzu vestavěného kódu. Implementovali jsme specializaci této služby pro programovací jazyk Python a abychom demonstrovali její užitečnost, navrhli jsme a implementovali prototyp skeneru datových linií pro službu integrace dat AWS Glue. Tento skener rozsáhle využívá dotčenou službu k analýze datových linií ve vložených skriptech v jazyce Python, což jsme předvedli na realistickém příkladu.