

Notebook-based technologies, like Databricks and Jupyter notebooks, have gained popularity in recent years due to their adaptability and convenience. A notebook is an interactive computational environment that allows users to create documents that contain code, visualizations, and explanatory text in one place. Notebooks provide a space for data exploration, analysis, and documentation, enabling users to easily develop and present their work. The ability to combine code execution with explanations and visualizations within a single document promotes reproducibility, enhances collaboration among team members, and motivates data scientists to efficiently work with data. In this work, we analyzed the Databricks technology in order to extend the Manta Flow platform, a highly automated data lineage analysis tool, to support this technology. We designed and implemented a new scanner that provides basic support for analyzing Databricks notebooks written in Python and Databricks SQL languages. We also provide an implementation of a so-called shared context that can be used for passing information between different scanners in the Manta Flow platform. To visualize the interactions between languages and scanners we extended the Manta graph with a new node type that represents the shared context. Alongside this, we implemented a so-called language context to enable scanners to store information useful for analysis of further cells written in the same language in a given notebook. Finally, we demonstrate the functionality of the scanner and the result graphs it produces on example Databricks notebooks.