Oct 27, 2025

## ADF pipeline Impact Analyzer

Invited [Balakrishnan Murthy](mailto:balakrishnan.mur@tigeranalytics.com) [Ranjit Kanamarlapudi](mailto:ranjit.kanamarla@tigeranalytics.com) [Aravind Ragunathan](mailto:aravind.ragunath@tigeranalytics.com) [Sathish Kanna Nagasundaram](mailto:sathish.nagasund@tigeranalytics.com) [Anand Jha](mailto:anand.jha@tigeranalytics.com) [Prasad Behere](mailto:prasad.behere@tigeranalytics.com) [Srinivasula Yeruva](mailto:srinivasula.yeru@tigeranalytics.com)

Attachments [ADF pipeline Impact Analyzer](https://www.google.com/calendar/event?eid=NjFsOXBmM2hkamY4MzByaDF2M3VxaTh2b2dfMjAyNTEwMjdUMDczMDAwWiByYW5qaXQua2FuYW1hcmxhQHRpZ2VyYW5hbHl0aWNzLmNvbQ)

Meeting records [Transcript](?tab=t.5jsbnw8tlc69)

### Summary

Ranjit Kanamarlapudi initiated the meeting to review the accelerator's output and the Streamlit application's implementation, with Anand Jha presenting the ADF Analyzer and its comprehensive Excel report. Anand Jha detailed various report sheets, including trigger and integration runtime information, data lineage, and activity counts, while also demonstrating the Streamlit application integration and discussing orphaned resources and impact analysis. Sathish Kanna Nagasundaram provided feedback on SQL code extraction, missing table names, and report sheet prioritization, which Anand Jha committed to addressing. Ranjit Kanamarlapudi decided the team would present the accelerator to Sachin this week, requiring Anand Jha to prepare a guideline document and simplify the functionality and data flow diagram.

### Details

* **Accelerator Output Review** Ranjit Kanamarlapudi initiated the meeting to review the output generated by the accelerator and later discuss the Streamlit application implementation ([00:00:00](?tab=t.5jsbnw8tlc69#heading=h.y9qm6u4bt2wq)). Anand Jha confirmed that they had made significant progress and would share the entire screen to present the output ([00:01:05](?tab=t.5jsbnw8tlc69#heading=h.viuj1svi9cje)). Sathish Kanna Nagasundaram noted that they had reviewed the output and found some points for improvisation ([00:00:00](?tab=t.5jsbnw8tlc69#heading=h.y9qm6u4bt2wq)).
* **ADF Analyzer Overview** Anand Jha presented a comprehensive overview of the ADF Analyzer, emphasizing that the `ADF\_analyzer.py` is the main Python file that performs a full analysis, including deep parsing and dependency tracking, and outputs a comprehensive Excel report ([00:03:20](?tab=t.5jsbnw8tlc69#heading=h.sshv5ygjbok7)). They also explained that the tool processes ARM templates, extracting details like the total number of activities, pipelines, data flows, and other services ([00:04:16](?tab=t.5jsbnw8tlc69#heading=h.oooj6knjo2w5)).
* **Detailed Report Breakdown** Anand Jha elaborated on the various sheets within the Excel report, starting with a summary page and then moving to specific details. They explained that the activity sheet tracks pipeline names, activity types (web, copy, etc.), roles, linked datasets, and SQL code details ([00:05:18](?tab=t.5jsbnw8tlc69#heading=h.5k1jzdbgapur)). Anand Jha also covered pipeline-level information, data flow linkages, data flow transformations, and comprehensive dataset information ([00:06:32](?tab=t.5jsbnw8tlc69#heading=h.7xb25o41gmrv)).
* **Trigger and Integration Runtime Information** Anand Jha provided a detailed explanation of the trigger-level information captured, including trigger name, type, state, frequency, interval, schedule, and associated pipelines ([00:09:04](?tab=t.5jsbnw8tlc69#heading=h.m55mm4hi98mf)). They also discussed the integration runtime details, noting that the ARM template included various integration runtimes like Azure, Oracle, and Salesforce ([00:11:53](?tab=t.5jsbnw8tlc69#heading=h.q9w8ofucimxk)). Aravind Ragunathan suggested incorporating integration runtime details into the linked services tab for better clarity ([00:13:24](?tab=t.5jsbnw8tlc69#heading=h.uld9l8x8qq0f)).
* **Data Lineage and Activity Counts** Anand Jha demonstrated the data lineage sheet, which outlines the pipeline name, activity type, source, sink, and transformations applied ([00:12:31](?tab=t.5jsbnw8tlc69#heading=h.fs8wq9jsadrq)). They also presented the activity count sheet, which summarizes the usage frequency of each activity type across all pipelines, such as copy activities and set variable activities ([00:13:57](?tab=t.5jsbnw8tlc69#heading=h.90eafy55fyq1)).
* **Streamlit Application Integration** Anand Jha showed how the generated Excel output can be loaded into a Streamlit application for a more intuitive and visual representation of the insights ([00:14:44](?tab=t.5jsbnw8tlc69#heading=h.firsda89c6jr)). Ranjit Kanamarlapudi suggested automating the process of linking the output file to the Streamlit app instead of manual placement ([00:16:22](?tab=t.5jsbnw8tlc69#heading=h.llv77qk6magc)).
* **Orphaned Resources and Impact Analysis** Anand Jha introduced the concept of "orphaned pipelines," which are pipelines that exist but are not referenced or used by any triggers or other pipelines ([00:16:22](?tab=t.5jsbnw8tlc69#heading=h.llv77qk6magc)). They also presented the impact analysis feature, which shows the upstream and downstream effects of deleting a particular pipeline ([00:17:57](?tab=t.5jsbnw8tlc69#heading=h.so13xqblpr2w)).
* **Feedback and Future Improvements** Sathish Kanna Nagasundaram provided feedback, highlighting issues such as incomplete SQL code extraction, missing table names in the SQL section, and the absence of stored procedure activity capture ([00:21:54](?tab=t.5jsbnw8tlc69#heading=h.4x6zonr4i4eg)). They also pointed out discrepancies in sequence numbering and suggested reordering the report sheets to prioritize pipeline-level information ([00:24:28](?tab=t.5jsbnw8tlc69#heading=h.cxeeme18bvsg)). Anand Jha acknowledged the feedback and committed to addressing these points ([00:22:43](?tab=t.5jsbnw8tlc69#heading=h.8uq15nga0i2l)) ([00:25:19](?tab=t.5jsbnw8tlc69#heading=h.9exsx863ruxr)).
* **Next Steps and Presentation Preparation** Ranjit Kanamarlapudi informed the team that they would be presenting the accelerator to Sachin during the current week ([00:18:47](?tab=t.5jsbnw8tlc69#heading=h.fwic3svh66v2)). They requested Anand Jha to prepare a separate document outlining the guidelines on how to run the accelerator, once the final version is ready ([00:20:20](?tab=t.5jsbnw8tlc69#heading=h.vtce4grkl5kc)). Ranjit Kanamarlapudi also emphasized the need to simplify the functionality and data flow diagram for better understanding ([00:20:55](?tab=t.5jsbnw8tlc69#heading=h.qfo9wfqrktwd)). The team scheduled a brief demo session before the main presentation to ensure everyone is prepared ([00:31:22](?tab=t.5jsbnw8tlc69#heading=h.nb2278qmwl9u)).

### Suggested next steps

* Anand Jha will incorporate the IR column, include the table or file name in the source and sync sections of copy activity and add the sync table name in data flow information.
* Anand Jha will automate placing the output file in the Streamlit application.
* Anand Jha will extend orphaned resources to include triggers, linked services, and data sets.
* Anand Jha will prepare a separate document with guidelines on how to run the accelerator.
* Anand Jha will check and fix the SQL part to (1) capture maximum text and (2) list table names consumed.
* Anand Jha will include stored procedure activity and a field for SP name triggered by that activity.
* Anand Jha will verify if sequences are getting populated correctly, as there are misalignments, and switch the order of sheets to have pipeline information as the first page.
* Sathish Kanna Nagasundaram will share the updated ARM template with Anand Jha.
* The group will work on making the functionality and data flow diagram clearer.

*You should review Gemini's notes to make sure they're accurate.* [*Get tips and learn how Gemini takes notes*](https://support.google.com/meet/answer/14754931)

*Please provide feedback about using Gemini to take notes in a* [*short survey.*](https://google.qualtrics.com/jfe/form/SV_9vK3UZEaIQKKE7A?confid=7XKMuX_l-2XiR2PIBqljDxIUOAIIigIgABgECA&detailid=unspecified)