

Technical Disclosure Commons

Defensive Publications Series

April 2022

ROBUST DATA SENDING IN PRINTER TELEMETRY

HP INC

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation

INC, HP, "ROBUST DATA SENDING IN PRINTER TELEMETRY", Technical Disclosure Commons, (April 22, 2022)

https://www.tdcommons.org/dpubs_series/5086



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

Robust data sending in printer telemetry

Abstract

Large Format Printers are IoT devices that send data to the cloud on event basis. The type of data sending is, for example, the total ink consumed by the printer up to this day or information about the cartridges.

Data is sent based on a data schema. Data schemas are a common interface for communication between printer and cloud data platforms. Printer could be operating under certain modes not supported by the schema. This causes data schema validation to fail, provoking data quality issues.

The solution proposed creates a second schema, tagging the files which do not have the minimum required fields and storing its file version. Using this feature, printer can request to send a specific schema version ensuring correct data is sent. Therefore, if printer sends data files with issues causing a schema failure, this will be detected, and printer will retry data sending with another schema until validation pass.

Invention

Printer sends data to cloud and schema is validated. If cloud detects schema issue or wrong data is tagged, then cloud informs back the schema failure to the printer and, afterwards, printer sends a second data event with a different schema not affected by first issue. There will be two files sent by the printer: one with the initial data schema sending and the other with this data reformed based on a different schema (previous versions) when the schema was validated correctly.

If data schema validation fails, the printer will send two files. First file with the data issue, and second file, with the schema that was validated before and worked properly. Both files will be classified in the cloud based on their reception timestamps and metadata using only the correct data event but not losing the additional details that the first file might have.

This second file, or data event, could be easily detected using metadata. File header will contain specific printer identifiers, timestamp of file sending, and the identifier of previous file sent, which was containing the schema bug.

When cloud detects the second file, it will replace the wrong information of the first file by the second one. Therefore, wrong data will be corrected and ETL (extract, transform and load) will be done only for validated schema events.

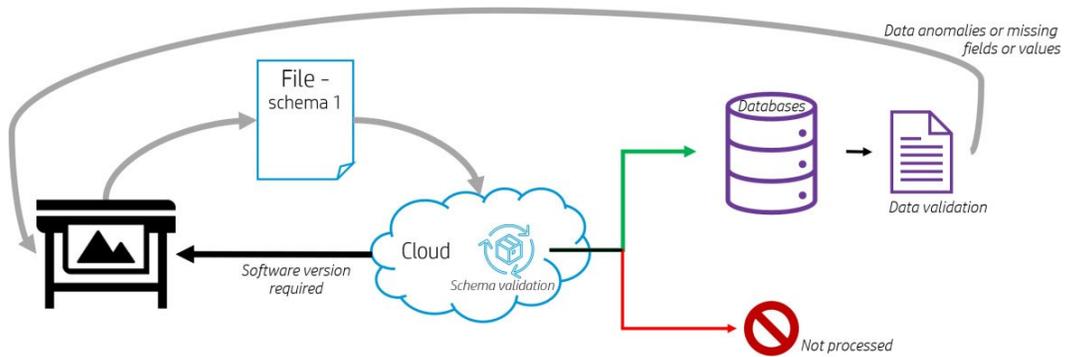


Figure 1: Schema of data pipeline. The printer sends data using certain SW version. A schema validation is done to accept to process the file or not, even if it is incomplete. In case the data processed has missing values or with wrong data type, then an alert is triggered to the printer.

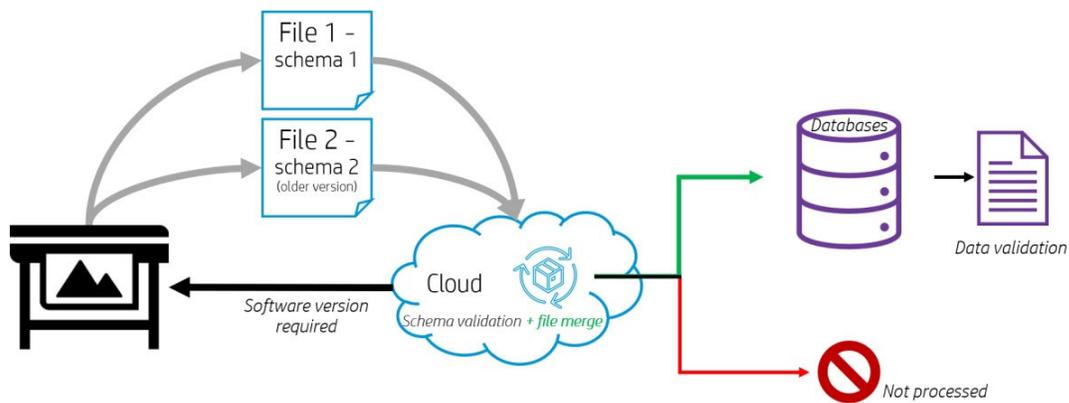


Figure 2. Schema of proposed solution. Once an alert is triggered, the printer sends the same data twice:

1. File 1 with the version it is usually required.
2. File 2 with a previous version where similar failures were not seen.

The cloud then merges both schemas using the most data possible from file 1 and the rest of data from file 2.

The advantages of this solution are:

- Quality Cost Reduction: by having the option of sending previous data schema versions, quality test for new printer firmware / software version could be reduced and only test incremental data schema fields.
- Good data quality: trust on data will increase for business decisions, due the decrease of data quality issues.
- Robustness: adding the second schema while sending data will increase robustness of the data retrieving solution, as data can be recovered.
- Reducing computational cost in the cloud: data cleaning will be reduced, giving more time to perform other relevant actions.

Disclosed by Carmen Pardo, Aida Aranda and Ana Oropesa, HP Inc.