

Fredrik Hellman*, Janne Hakala, William Antus, André Norrback, Andreas Lundell, and Jerker Björkqvist

Automated Life Cycle Assessment in Scalable and Flexible Manufacturing

Abstract: Traditional Life Cycle Assessment (LCA) procedures are time-consuming and rely heavily on manual input. In collaboration with Mirka, a Finnish manufacturing company producing abrasives, we integrate real-time data flows between production databases and LCA tools. This collaboration is ideal due to Mirka’s extensive product portfolio, containing thousands of different products, providing a strong foundation for demonstrating the scalability and versatility of the automated LCA framework. We utilize a data lakehouse architecture to enable seamless handling of diverse datasets often required for LCA, supporting near real-time updates to sustainability metrics. This automated solution minimizes manual effort, adapts dynamically to operational data changes, and delivers a continuous overview of environmental performance.

Keywords: Life Cycle Assessment, Data Lakehouse, Real-time Data Integration, Automated LCA, Manufacturing

***Corresponding Author: Fredrik Hellman:** Åbo Akademi University, Information Technology, E-mail: fredrik.hellman@abo.fi

Janne Hakala: Åbo Akademi University, Information Technology, E-mail: janne.hakala@abo.fi

William Antus: Mirka Ltd, E-mail: william.antus@mirka.com

André Norrback: Mirka Ltd, E-mail: andre.norrback@mirka.com

Andreas Lundell: Åbo Akademi University, Information Technology E-mail: andreas.lundell@abo.fi

Jerker Björkqvist: Åbo Akademi University, Information Technology, E-mail: jerker.bjorkqvist@abo.fi

1 Introduction

Life Cycle Assessment (LCA) evaluates the environmental impacts of a product throughout its entire life cycle. LCA plays a key role in driving sustainability, supporting frameworks like Environmental Product Declarations (EPDs) and Digital Product Passports (DPPs), both essential for complying with circular economy (CE) principles. Yet, traditional LCA methods remain manual, time-consuming, and static [11], which limits their usefulness for current sustainability needs. With the European Union’s Corporate Sustainability Reporting Di-

rective (CSRD) requiring transparent reporting on sustainability [6], companies are pushed towards more sustainable production, and could benefit from implementing more dynamic and automated processes in calculating sustainability metrics.

For companies with extensive product portfolios, often containing over thousands of products, automation is essential to make the LCA process more effective. Additionally, automating LCA calculations unlocks further benefits, such as real-time visualization of emissions during production, supporting data-driven management and continuous performance monitoring.

In this study, we describe an automated life cycle assessment (ALCA) framework implemented at Mirka. The goal is to demonstrate how automating LCA calculations can reduce manual effort and enhance the scalability of sustainability assessments. By leveraging a data lakehouse architecture, we enable near real-time integration of production data with LCA tools, allowing for dynamic updates to sustainability calculations. This approach could improve efficiency and provide Mirka with continuous insights into environmental performance, promoting data-driven decision-making for more sustainable production.

2 Background

2.1 Automated Life Cycle Assessment

LCA is becoming increasingly relevant, especially for manufacturing companies, both to comply with regulatory requirements, as well as, reducing emissions as lower emissions are becoming a competitive advantage. A traditional LCA consists of four steps in accordance with ISO14040 [3] and ISO 14044 [4]: goal and scope, life cycle inventory analysis, life cycle impact assessment, and finally the interpretation. LCA require a substantial amount of data, and current LCA practices are not scalable and rely heavily on manual input.

As the authors are aware, there is currently no framework with full support for automated LCA calculations. However, commercial LCA software solutions,

such as SimaPro Synergy [12], and Sphera [13], have recently released application programming interfaces (APIs), enabling programmatic communication with their systems, underlining the need for automated solutions in industry. Additionally, the framework proposed by [11] presents an ALCA process for manufacturing companies, and [9] solved their defined challenges by developing a life cycle inventory (LCI) modeling approach, allowing automated assignments of a component’s attributes, to pre-defined LCI models using commercial LCA tools; all of these indicate an active research area. These advancements, both in industry and academia, highlight a clear trend toward automation in LCA. However, fully scalable and integrated solutions remain an ongoing challenge.

2.2 Data Lakehouse Architecture

According to [9], automated LCA for complex products faces several significant challenges, e.g., balancing accuracy with efficiency, managing the heavy manual workload of mapping input data to LCI datasets, dealing with inflexible LCI modeling, and the absence of automation support in many commercial LCA software packages. Further, [10] also highlights that LCI is the most resource-intensive part of the assessment.

Given that the main challenges are inherently data-related, a robust and scalable data management solution is essential for enabling automation. A data lakehouse architecture offers a promising approach by integrating structured and unstructured data within a governed, query-optimized, and scalable environment. Originally introduced by Armbrust et al. [5], a data lakehouse combines the advantages of data lakes and data warehouses, supporting ACID (atomic, consistent, isolated, durable) transactions, metadata management, and data sharing.

These features align well with the requirements of automated LCA, as they enable programmatic data ingestion, continuous updates to LCA models, and support for AI-driven analysis of environmental impacts. Cloud-based implementations further enhance the scalability and efficiency of LCA workflows by providing cost-effective, high-performance storage and compute capabilities. Leveraging a data lakehouse approach enables organizations to overcome data scalability and accuracy challenges in automated LCA, resulting in more efficient, real-time assessments of environmental impact, and better alignment with evolving sustainability regulations.

2.3 Mirka Ltd

Mirka is an essential partner for this project due to several key factors. First, their broad product portfolio provides a good test case for demonstrating the scalability of the ALCA framework. With thousands of distinct products, the system can be tested for its ability to handle diverse materials, processes, and environmental impacts at scale.

Second, Mirka’s established data lakehouse infrastructure supports integration capabilities for LCA tools. Real-time in this context is defined at least once per day down to on a minute basis. It is worth noting that there are no technical limitations preventing even more frequent calculations, however, the chosen update frequency aligns with Mirka’s specific requirements and operational needs. Their existing systems ensure that the automated framework can access reliable and well-organized data for processing. Additionally, Mirka’s strong sustainability focus, including initiatives like transitioning to fossil-free electricity to reduce emissions, shows clear willingness to adapt to sustainability efforts.

Therefore, the combination of product portfolio complexity, data readiness, and sustainability alignment makes Mirka an ideal collaborator for implementing and showcasing the ALCA system.

3 ALCA Framework

The objective with this study is to describe the ALCA framework that integrates real-time production data into LCA tools, assess its scalability and flexibility by applying it to Mirka’s extensive product portfolio, and demonstrate continuous sustainability monitoring through real-time visualizations of key environmental indicators.

The ALCA workflow implemented in this study is illustrated in Figure 1. Mirka integrates real-time data from Enterprise Resource Planning (ERP), Manufacturing Execution Systems (MES), and sensor systems into a centralized data lakehouse, which acts as a storage repository to enable automated, dynamic LCA updates. Python scripts then query, process, and map the data to OpenLCA models, with the obtained results fed back into the repository, for example, for usage in dashboards that support near real-time monitoring.

The ALCA framework is here applied to two distinct cases:

1. **Real-time monitoring of manufacturing:** This approach visualizes how fluctuations in electricity consumption and raw material usage affect emissions during a single production process.
2. **Batch-level LCA calculations:** This method calculates the LCA for an entire production batch after completion, using a gate-to-gate approach that automatically generates sustainability metrics and supports the preparation of an EPD.

By applying the ALCA framework in these two scenarios, this study demonstrates both the real-time functionality of monitoring sustainability metrics and the ability to perform batch-level LCA calculations across multiple production processes. Since certain processes depend on the completion of others before LCA calculations can be performed, the batch-level approach ensures accurate assessment of the entire production cycle.

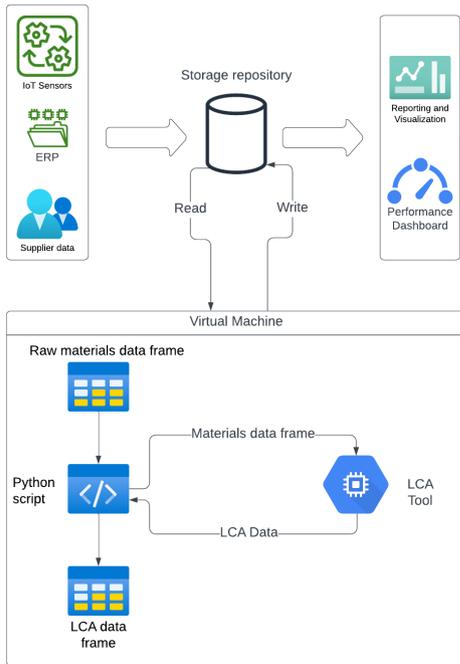


Fig. 1. An illustration of the workflow in the ALCA framework

4 Implementation

4.1 Data structure

Firstly, to automate LCA practices, a standardized data structure is required. Databricks [1], a cloud-based data management platform, serves as the centralized stor-

age repository. The data lakehouse architecture within Databricks provides a unified environment that integrates data from various sources, such as production systems, ERP databases, and supplier datasets. For a single process, all relevant inputs should be systematically documented alongside the corresponding process name, ensuring scalability and applicability to other processes, as illustrated in Table 1. This table is updated dynamically with new values during the production, and sent to the LCA software for calculation, which allows for real-time monitoring of relevant sustainability metrics. In this table layout, the primary flow for each process is defined from the process name. It should be noted that the names and data presented in the tables are illustrative examples and do not represent actual values or proprietary information.

Once all processes for a product are completed, the table structure, as illustrated in Table 2, consolidates the data for the final product. All relevant process inputs are aggregated within the table and sent for calculation. The selected impact assessment method dictates the specific environmental impact values that are calculated and returned. The upstream process column defines the sequence of processes, ensuring that outputs from one process serve as inputs for the next. The process with no downstream dependencies represents the final stage, while the process with a null upstream value marks the starting point. Based on this structure, we can systematically trace material and energy flows and ensure accurate impact calculations.

4.2 Data processing

When data is sent for processing, it should be correctly mapped in the LCA tool. In this study, OpenLCA is used as it provides an API for inter-process communication (IPC). OpenLCA is a free, open-source LCA software that can be used for professional LCAs [8]. It is designed to be flexible and customizable, allowing for adoption of the software to suit the users specific needs and conduct a wide range of sustainability assessments [8]. The IPC protocol is provided by an OpenLCA server that can be a running instance of the OpenLCA desktop application or a webserver with an OpenLCA backend that exposes this protocol. In our case, the IPC server is running on a virtual machine, that fetches the data from Databricks using databricks-connect Python package [2].

OpenLCA's Python package `olca-ipc` [7] allows for programmatic communication with OpenLCA. The

Column	Description
Process ID	Unique identifier for each process step.
Process	Name of the process
Upstream	The previous process that serves as input to the current process
Raw Material	The material or resource used in the process
Unit	Measurement unit for the raw material
Chemical Classification	Categorization of the raw material based on its classification in the LCI
Amount	Quantity of raw material used in the process
Impact Method	Environmental impact assessment method applied to the process

Table 1. Description of table columns

Process ID	Process Name	Raw Material	Unit	Chemical Classification	Amount	Impact Method
P004	Production3	Final paper product	kg	Paper, coated, printing	1000	ReCiPe 2016 Midpoint (H)
P004	Production3	Ink	kg	Pigment, organic, printing	10	EF 3.0
P004	Production3	Polyethylene	kg	Polyethylene, low density	15	EF 3.0
P004	Production3	Electricity	kWh	Electricity, high voltage	200	ReCiPe 2016 Midpoint (H)

Table 2. Example table layout for a single process

Process ID	Process	Upstream	Raw Material	Unit	Chemical Classification	Amount	Impact Method
P000	Transportation	NULL	Paper roll	kg	Pulp, mechanical, bleached	1000	ReCiPe 2016 Midpoint (H)
P001	Storage	P000	Paper roll	kg	Pulp, mechanical, bleached	1000	ReCiPe 2016 Midpoint (H)
P002	Production1	P001	Paper roll	kg	Pulp, mechanical, bleached	980	EF 3.0
P002	Production1	P001	Starch	kg	Starch, from maize	20	EF 3.0
P002	Production1	P001	Electricity	kWh	Electricity, medium voltage	150	ReCiPe 2016 Midpoint (H)
P003	Production2	P002	Coated paper	kg	Coating, polymer-based	970	ReCiPe 2016 Midpoint (H)
P003	Production2	P002	Calcium carbonate	kg	Calcium carbonate, precipitated	50	EF 3.0
P003	Production2	P002	Steam	MJ	Heat, from steam, in chemical industry	300	ReCiPe 2016 Midpoint (H)
P004	Production3	P003	Final paper product	kg	Paper, coated, printing	1000	ReCiPe 2016 Midpoint (H)
P004	Production3	P003	Ink	kg	Pigment, organic, printing	10	EF 3.0
P004	Production3	P003	Polyethylene	kg	Polyethylene, low density (LDPE)	15	EF 3.0
P004	Production3	P003	Electricity	kWh	Electricity, high voltage	200	ReCiPe 2016 Midpoint (H)

Table 3. Example Table Layout for Multiple Processes for a Finished Product

Python script in the ALCA framework will dynamically create, or update product systems, processes, and flows based on incoming data. For existing product systems, flows will be updated, while new models will be generated for new products.

Once each LCA calculation completes, impact assessment results are extracted from OpenLCA through the same programmatic interface. These outcomes are then written back to Databricks as newly versioned records, where they become available for visualization or further analysis. In an effort to ensure transparency and support a continuous improvement cycle, Mirka's monitoring dashboards can track changes in impact indicators in near real-time, enabling stakeholders to respond promptly to unexpected variations in energy use, raw material consumption, or transportation routes.

4.3 Results

In collaboration with Mirka, we tested the proposed ALCA framework on a single process in the manufacturing process and on multiple processes once the product is finished. The framework is adaptable to other products as we have a defined standard way of storing the data going into an LCA. For the single process, we dynamically updated the model with different values for the measured electricity consumption, and visualized the results using a dashboard in Databricks. For the batch process, we aggregated results for one product from all processes in a gate-to-gate approach for 1000 meters of final product.

By adopting an automated approach to LCA, we achieve a higher level of automation while reducing the risk of manual errors. Moreover, this method promotes reusability and scalability across different products and processes by using a standardized data structure. In the future, it could also be extended into a simulation tool to explore strategies for minimizing emissions.

5 Conclusions

This study demonstrates a structured, scalable, and automated way to calculating LCA. The proposed framework handles both real-time visualizations of emissions from a single process and a batch calculation once a product is finished. The implementation was successful for the tested product, both for a single process and multiple processes.

This study reaches the same conclusion as previous research, identifying data mapping as a primary challenge in automated LCA. The current implementation maps process flows to their corresponding entries in the LCI database in OpenLCA using flow names. While functional, this approach introduces potential difficulties, as naming conventions may vary across datasets. A more robust solution would involve mapping flows using unique identifiers, such as Chemical Abstracts Service (CAS) numbers. However, this would require an enriched LCI database containing CAS numbers, as well as ensuring that the company's internal databases also include CAS information. Integrating such identifiers would improve the accuracy and consistency of automated LCA calculations by minimizing discrepancies in flow matching.

Overall, the proposed framework successfully enables automated LCA calculations, demonstrating its applicability for both real-time emissions visualization and batch-level assessments. While some challenges remain, particularly in data mapping, the implementation proves effective for the tested product and provides a foundation for further refinement and scalability.

5.1 Limitations

The ALCA framework demonstrated scalability by being tested on multiple single processes. However, it was only applied to one aggregated process. Potential challenges, such as long calculation times, may arise when calculating sustainability metrics across multiple processes with extensive data points. While the framework enables near real-time LCA calculations, actual real-time processing depends on multiple factors, including data transfer speeds, computational resources, and software limitations. The capability of the framework should be further evaluated to determine the extent to which LCA calculations can be performed in real-time.

LCA databases are typically large, which can compromise performance in LCA software. Optimizing database performance by limiting the dataset to only the necessary flows could be an area for improvement.

Future research could explore alternative tools beyond OpenLCA. However, as of this writing, the availability of open-source LCA tools with API functionality is limited. While commercial alternatives exist, they may introduce limitations in terms of flexibility and accessibility.

References

- [1] Databricks. <https://www.databricks.com/>. Accessed: 2025-03-06.
- [2] Databricks connect for Python. <https://docs.databricks.com/aws/en/dev-tools/databricks-connect/python/>. Accessed: 2025-03-06.
- [3] Environmental management — Life Cycle Assessment — Principles and Framework. ISO 14040:2006, 2006. Available at <https://www.iso.org/standard/37456.html>.
- [4] Environmental management — Life Cycle Assessment — Requirements and Guidelines. ISO 14044:2006, 2006. Available at <https://www.iso.org/standard/38498.html>.
- [5] Michael Armbrust, Ali Ghodsi, Reynold Xin, and Matei Zaharia. Lakehouse: A new generation of open platforms that unify Data Warehousing and advanced Analytics. *Conference on Innovative Data Systems Research (CIDR)*, 2021.
- [6] European Commission. Directive (EU) 2022/2464 of the European Parliament and of the Council on Corporate Sustainability Reporting (CSRD). Technical Report L 322/15, Official Journal of the European Union, 2022. Accessed: January 30, 2025.
- [7] GreenDelta. olca-ipc Documentation, n.d. Accessed: 2025-03-06.
- [8] GreenDelta. OpenLCA 2 Manual: Introduction, n.d. Accessed: 2025-03-06.
- [9] Patrick Haun, Philipp Müller, and Marzia Traverso. Improving automated Life Cycle Assessment with Life Cycle Inventory model constructs. *Journal of Cleaner Production*, 370:133452, 2022.
- [10] Pascal Köck, Jürgen Karl, and Daniel Lintl. Automation of Life Cycle Assessment—a critical review of developments in the field of Life Cycle Inventory analysis. *Sustainability*, 15(6):5531, 2023.
- [11] Daniel Schneider, Patrick Jordan, Jonathan Dietz, Michael F. Zaeh, and Gunther Reinhart. Concept for automated Life Cycle Assessment of Manufacturing Processes. *Procedia CIRP*, 116:59–64, 2023.
- [12] SimaPro. Simapro Synergy: Collaboration and Data Sharing for LCA. <https://simapro.com/synergy/>, n.d. Accessed: January 30, 2025.
- [13] Sphera Solutions. LCA Automation: Life Cycle Assessment software and data. <https://sphera.com/solutions/product-stewardship/life-cycle-assessment-software-and-data/lca-automation/>, n.d. Accessed: January 30, 2025.