CASE STUDY

How an Industrial Group Surveyed SAP ECC and SAP BW in 4 Weeks Instead of 8 Months

Two live SAP systems read in a single run. 26 objects end-to-end in 72 minutes. 97% fewer person-hours, and every statement traceable to a real read.

Powered by CLU | cluagents.com

Industry: Manufacturing / multi-company industrial group | Source systems: SAP ECC + SAP BW | Target: Google BigQuery | Period: July – August 2026

−97%	2 systems	4 weeks	72 min
Person-hours	ECC + BW, one run	Instead of 6–8 months	Per 26-object run

KEY RESULTS

- 97% reduction in person-hours: from ~600 PH to ~15 PH across a 15-object scope
- Both SAP systems — ECC and BW, separate servers and separate credentials — read in a single run, with evidence cross-referenced between them
- AS-IS documentation cycle compressed from 6–8 months to 4 weeks
- 26 objects surveyed in a 72-minute run: 11 minutes harvesting against SAP, 61 minutes of synthesis
- 208 pages delivered, 14 fixed sections per object, zero hand-written facts
- Findings the manual method never surfaced: 60 customer DataSources hidden in $TMP and process chains idle since 2017

Executive Summary

A multi-company industrial group running SAP ECC and SAP BW set out to migrate its analytical layer to Google BigQuery. Before a single table could be modelled, someone had to answer a deceptively simple question: what exactly is in there, and how does it work?

That question cannot be answered in one system. The data originates in ECC — in ERP tables, extractors and customer ABAP — and only then travels to BW. A survey that looks only at BW documents half the journey and leaves out precisely the part nobody remembers.

The established answer is a manual AS-IS survey: a specialist consultancy documents object by object — one Word document per object, written by two consultants over roughly a week each — from SAP GUI

screenshots, interviews and institutional memory. For a scope of this size that path runs 6 to 8 months before the migration team can design anything.

CLU deployed a read-only harvesting agent inside the customer's network, connected to BOTH production systems. It reads metadata directly over RFC — transformations, ABAP routines, extraction exits, process chains, ECC table structures — and writes the document from what it measured. The first consolidated deliverable landed in weeks, not quarters: 208 pages, every claim traceable to the table it came from.

The Challenge

The obstacle was never the migration. It was that nobody — not the customer, not the consultancy — could describe the current state with confidence, least of all across the two systems.

Operational Reality

- The manual method documented one object per pair of consultants per week, measured from the start and end dates the consultancy stamps on its own deliverables
- Traceability broke at the system boundary: BW names the DataSource that feeds an object, but what that extractor actually does lives in ECC, in ABAP nobody had read in years
- The inventory was unknown even to its owners — 1,135 DataSources, of which 530 sat in the $TMP local package and 60 of those were customer-namespaced
- Scope was chosen blind: the customer had to name technical objects before anyone could show them a map of what existed
- Risk was invisible: process chains that had not executed since 2017 were still wired into flows slated for migration

The compounding cost is not the writing — it is the elapsed time. Every week the AS-IS is unfinished is a week the target model cannot be designed, estimated or committed to.

The Solution

CLU deployed an autonomous survey agent built on its GRID orchestration framework. It runs as a local runtime inside the customer's network, holds read-only credentials, and never moves business data — only metadata about how the system is built.

How It Works

- Multi-system by design: one runtime serves every SAP server it is given — ECC, BW and whatever comes next — each with its own credential set, and cross-references the evidence between them
- Census first, survey second: a fast inventory answers "what do I have" before anyone commits to the expensive question, "how does this work"
- Recursive lineage to point zero: it walks from a BW InfoCube back through DSOs, InfoSources and DataSources, crosses into ECC and reaches the ERP table where the data originates — rather than reporting one hop and stopping
- It follows the code in ECC: where a customer exit dynamically dispatches to another ABAP program, the agent resolves the name at read time and retrieves the real source
- Fixed document structure: 14 numbered sections identical for every object, rendered from measured evidence — the language model writes only bounded explanatory prose, never the structure

- Target-platform fit: every object is scored against the published Google Cortex Framework standard — which replicates ECC, not BW, which is exactly why the cross-system read is what makes the verdict computable
- Measured versus inferred is declared on every page, and a gap that could not be read is reported naming the table that was consulted

Deployment ran against the customer's production systems with a single credential attempt per system and no retries — a deliberate constraint, because locking an account on a productive ERP is a far more expensive incident than a slow survey.

Results in Detail

All figures come from production runs against the customer's own SAP systems and from the delivered documents. Nothing here is a projection.

1. Person-Hours: The Consolidated Benefit

This is the metric that orders all the others. The manual method spends consultant hours per documented object; the agent spends machine minutes and leaves the person reviewing rather than transcribing.

Item	Manual AS-IS survey	CLU
Writing effort per object	40 – 80 PH (2 consultants × 5 business days)	Zero — generated by the agent
Machine time per object	—	≈2.8 minutes
Review and validation per object	Included above	≈1 PH
Total effort — 15-object scope	**600 – 1,200 PH**	**≈15 PH**
Person-hour reduction	**—**	**97% – 99%**

Basis of calculation: 40 PH per object assumes two consultants at 50% dedication over the 5 business days the consultancy itself records between the start and end date of each document; 80 PH assumes full dedication. CLU-side review hours are an operating estimate, not a measurement.

2. Reach: Both SAP Systems in the Same Run

The survey does not stop at the BW boundary. In the same run, with separate credentials, the agent reads both servers and cross-references the evidence — which is exactly where the manual method loses the thread.

What is read	SAP BW	SAP ECC
Object structure	InfoCubes, DSOs, InfoObjects with their keys	Extractor structure and its fields
Business logic	Transformations, field-level rules, start/end/field routines	Extraction exits, CMOD projects, dispatched programs
Data origin	InfoSources and DataSources	ERP tables and how they join to each other
Operations	Process chains, InfoPackages and DTPs with last execution	Extraction method and delta capability

What that means in practice: for an accounting flow, the document does not merely say “fed by DataSource 0FI_GL_10”. It names the ERP table where the balance originates, the master tables that complete it, the field they join on and with what cardinality, and the customer ABAP program that runs during extraction — with its source in the appendix.

That crossing is also what makes the BigQuery target computable: Google Cortex Framework replicates ECC, not BW, so fit can only be assessed against ERP tables. Without the jump into ECC there is no target-platform verdict — only an opinion.

3. Cycle Time: Months to Weeks

Metric	Manual AS-IS survey	CLU
Elapsed per object	≈5 business days	≈2.8 minutes
Objects per run	1 per consultant pair	26
Systems covered per run	1 at a time	2 (ECC + BW)
Full survey cycle	**6 – 8 months**	**4 weeks**
Speed per object	**—**	**≈850× faster**

The 72-minute figure is the full run measured on 7 August 2026 across 26 objects: 11 minutes of metadata harvesting against SAP, then 61 minutes of synthesis. Deeper runs over the same scope have taken up to four hours — still a working session rather than a quarter.

4. Depth: What the Manual Method Never Found

Reading the systems instead of interviewing them changes what the survey can see. Across the engagement the agent surfaced:

- 60 customer-namespaced DataSources inside the $TMP local package — non-transportable objects feeding production flows
- Process chains with no execution since 2017, still present in the documented data flows
- One InfoProvider with 326 direct lineage edges, a hub no interview would have described accurately
- In ECC, the real extraction logic behind a generic dispatcher exit: the include holds no logic — it builds another program’s name and calls it. A keyword search reports “no customer code” there, confidently and wrongly
- The actual joins between the ERP tables feeding an extractor, with the join field and its provenance: whether SAP’s dictionary declares it, or whether it is a shared key that was measured

All of this surfaced while examining a scope the customer had already selected. A landscape census would have shown it on day one.

5. Evidence Quality: Defects Burned Down Between Versions

Each delivered version was audited fact by fact against the previous one. The measurable result of that loop:

Defect measured in the delivered document	Earlier version	Latest version
Failed metadata reads against the wrong table	71	0

Defect measured in the delivered document	Earlier version	Latest version
Invalid-field read errors surfaced to the reader	24	0
“Re-run required” statements left for the customer	1	0
References to objects that do not exist in the system	6	0

An object that genuinely does not exist is now reported as exactly that — naming the directories consulted — instead of receiving a chapter of its own.

6. Consistency: One Structure, Every Object

The document carries 14 numbered sections identical for every object: identification, field-level mapping, transformation routines, extraction exits, load units, process chains, ECC source tables and target-platform fit. A section with nothing to report says which of three things happened: it was read and empty, it does not apply to this object type, or it could not be read — and why.

Every object is named with its system and type (BW · InfoCube · 0FIGL_C10, ECC · Table · FAGLFLEXT), in the text and in the diagrams, so nobody has to guess which side of the landscape they are looking at. Prose per object dropped from roughly 3,300 words to about 750, with the difference moved into tables generated from measured evidence.

What Made the Difference

This was not a document generator pointed at a database. Four decisions drove the outcome:

Cross the System Boundary

Data does not originate in BW. A survey that reads only BW documents the transport and not the origin, and leaves the most expensive part of the migration — extractor logic and customer ABAP — unopened. Reading ECC and BW in the same run is what makes it possible to trace from a query to the balance in an ERP table.

Read the System, Never the Room

Every field, rule, routine and table in the deliverable came from a live read of the customer’s SAP dictionary. Where a value was deduced rather than read, the document says so. That distinction separates an auditable inventory from a list of assumptions — and it is why the customer can hand the document to a developer.

Never Guess a Column Name

SAP metadata reads fail silently: a single non-existent column in a request empties the entire result, and an empty result reads exactly like “this object has nothing”. The agent resolves the real column layout on the customer’s own release before reading. An unglamorous rule that removed an entire class of confident, invisible errors.

Structure Owned by Code, Not by the Model

The language model contributes explanation, never architecture. Section titles, numbering and tables are generated deterministically, so the same section carries the same name and number in every chapter. When the model is unavailable the document still ships — it loses commentary, never structure or evidence.

Looking Ahead

With the AS-IS survey in hand, the customer's team moved from reconstructing the past to designing the target model. The next phase extends the same evidence into the migration itself: automated reconciliation between the SAP figures and the BigQuery result, per-object packaging so each component travels with the team that owns it, and a full census of both systems so the next scope decision is made from a map rather than from memory.

Intellectual Property and Confidentiality

Results notice: figures come from production runs and delivered documents in a specific engagement executed between July and August 2026. Manual-method timings are taken from the start and end dates recorded on the comparable manual survey deliverables for the same landscape; person-hours are derived from those dates under the assumptions declared in section 1. Individual outcomes vary with landscape size, system access and scope. Customer and third-party identities anonymized at the request of the parties.