PROPOSAL

Data Layer Foundation

Prepared for

alfatec GmbH & Co. KG

2026-07-07

Valid for 30 days

EXECUTIVE SUMMARY

Gemma Analytics proposes to build alfatec's foundational data layer – a dedicated infrastructure that extracts data from Business Central and makes it available for downstream applications, analytics, and AI-powered workflows.

The engagement delivers a **production-ready data layer** connected to Business Central, with automated data loading, a lightweight transformation layer, and clean documentation. The architecture is designed to serve alfatec's near-term needs (Sales Agent, third-party data provision) while remaining extensible for future use cases.

We present **two infrastructure options** – a self-hosted PostgreSQL instance and a cloud-hosted Snowflake warehouse – with a clear recommendation and trade-off analysis so alfatec can choose the path that fits best.

YOUR SITUATION

CURRENT STATE

alfatec is a 220-person electronic components distributor operating across Germany, the US, and Shanghai. The group includes alfatec GmbH & Co. KG (distribution, key account, area sales) and AI Energy GmbH (design-in engineering). All entities run on shared infrastructure with **Business Central on-prem** as the central ERP system, backed by a Microsoft SQL Server database.

Today, all data access happens directly through Business Central. Reporting has been handled by a legacy BI server, now partially replaced by Power BI. There is no separate data layer or analytical infrastructure outside the ERP.

Product master data quality is a known challenge: articles are duplicated with inconsistent naming conventions across text fields, making reliable analysis and automation difficult.

GOALS

1. **Decouple data access from the ERP** – reduce load on Business Central and provide a stable, performant foundation for applications that need alfatec's data
2. **Enable the Sales Order Agent** – provide the data infrastructure that the planned AI-powered sales agent will connect to
3. **Support third-party data provision** – create a clean, controlled path for sharing product data with external partners and e-commerce platforms
4. **Lay the groundwork for data quality** – establish infrastructure that enables systematic product master data cleanup over time
5. **Build the foundation for AI-first analytics** – create the data infrastructure that enables agentic analytics, natural language data exploration, and AI-native tooling as a next step
6. **Stay flexible** – choose an architecture that supports alfatec's likely migration of Business Central to the cloud without requiring a rebuild

OUR APPROACH

- **Lean and incremental** – we start with one data source (Business Central) and a focused scope, then expand based on real needs rather than speculative requirements
- **Future-proof foundations** – even in a lean Phase 1, we include a transformation layer (dbt) from day one so that data models, cleaning logic, and business rules have a proper home from the start
- **Code-first, AI-assisted** – all pipelines, models, and configuration are version-controlled in Git. We use AI-assisted development tooling (Claude Code) as standard practice for accelerated delivery
- **Your infrastructure, your control** – whether cloud or on-prem, alfatec owns and controls the data layer. Gemma builds it, documents it, and hands it over

INFRASTRUCTURE

The data layer consists of several components working together. Each plays a distinct role in moving data from Business Central into a usable, governed analytical environment.

Component	Role	Technology
Data Warehouse	Central storage for extracted and transformed data. Single source of truth for all downstream consumers.	Snowflake (recommended) or PostgreSQL – see Infrastructure Options below
Transformation	Cleans, structures, and models raw data into analytically useful tables. Enforces data quality and business logic.	dbt (data build tool)
Orchestration	Schedules and monitors all pipeline jobs: BC extraction, dbt runs, data quality checks.	Apache Airflow on a dedicated VM
Data Extraction	Pulls data from Business Central (and future sources) into the warehouse on a regular schedule.	Python-based pipelines (dlt or custom), triggered by Airflow
Data Access (future)	How downstream consumers query and interact with the data layer.	To be determined based on use case: MCP server for AI tools via semantic layer, visualization tools (Power BI, Lightdash), direct SQL access, or APIs

Airflow serves as the central scheduler for the entire platform – not only for BC data extraction, but also for running dbt transformations and any future data source connections. This gives alfatec a single operational surface for monitoring all data jobs.

INFRASTRUCTURE OPTIONS

We recommend choosing one of the following options for the data layer. Both are fully viable – the right choice depends on alfatec's priorities around operational overhead, future extensibility, and cloud readiness.

OPTION A: POSTGRESQL (SELF-HOSTED)

A PostgreSQL database running on a virtual machine in alfatec's own infrastructure (or a light cloud VM, e.g. Hetzner). Data stays fully within alfatec's network perimeter.

Option A: PostgreSQL

Self-hosted PostgreSQL database on an alfatec-managed VM. Data extracted from BC via scheduled pipelines, transformed with dbt, and served from Postgres.

Running costs: Minimal (VM resource costs only)
Maintenance: alfatec IT responsible for VM, OS updates, backups, monitoring
Data residency: Fully within alfatec infrastructure

Advantages

- Zero additional vendor costs beyond the VM
- Complete data sovereignty – data never leaves alfatec's infrastructure
- Simplest possible setup, fastest to deliver
- PostgreSQL Foreign Data Wrappers can provide direct read-only access to BC's SQL Server – potentially eliminating the need for a separate extraction pipeline in Phase 1
- TISAX compliance straightforward (data remains on-prem)

Limitations

- alfatec IT carries the operational burden: patching, backups, availability monitoring, disk management
- No native Azure AD / Entra ID SSO integration (requires additional configuration)
- Limited collaboration features for data cleaning workflows that involve manual review steps
- If Business Central migrates to the cloud, data flows cloud → on-prem (architecturally backwards)
- Harder to extend for external data sharing with third parties

OPTION B: SNOWFLAKE (CLOUD – GERMANY)

A fully managed cloud data warehouse hosted in Snowflake's Frankfurt region (Germany). Accessed via web UI and standard SQL tooling. Managed service with no infrastructure operations for alfatec.

Option B: Snowflake (Recommended)

Cloud data warehouse on Snowflake (Frankfurt region). Data extracted from BC via scheduled pipelines, loaded into Snowflake, transformed with dbt, and served to downstream consumers.

Running costs: EUR 50–200/month at current data volumes (pay-per-query model, scales with usage)
Maintenance: Fully managed – no patching, no backups, no disk management
Data residency: Frankfurt, Germany

Advantages

- Fully managed – no infrastructure maintenance for alfatec IT
- Native Azure AD / Entra ID SSO with MFA (alfatec's existing identity stack)
- Industry-standard security certifications: SOC 2 Type II, ISO 27001, HIPAA
- Purpose-built for analytics, transformation, and data sharing workloads
- Clean path for BC cloud migration (cloud → cloud)
- Secure data sharing features for third-party data provision (no API or file export needed)
- Ideal foundation for AI-first analytics tooling (see Phase 2 outlook)
- Scales transparently as data volumes and use cases grow

Limitations

- Monthly running costs (small at current volumes, but a new budget line item)
- Requires a dedicated data loading pipeline (e.g. Airflow) from day one – with PostgreSQL, Foreign Data Wrappers can defer this to a later stage
- New vendor relationship and contract
- Requires internet connectivity (not air-gapped)

COMPARISON

Dimension	PostgreSQL (Self-Hosted)	Snowflake (Cloud)
Monthly cost	EUR 0 (VM only)	EUR 50–200
Setup effort	Slightly less	Slightly more
Maintenance	alfatec IT	Fully managed
Azure SSO	Manual setup required	Native support
Data residency	On-prem (alfatec)	Frankfurt, Germany
Security certs	n/a (self-hosted)	SOC 2, ISO 27001
BC cloud migration	Requires rework	Seamless transition
External data sharing	Custom API needed	Built-in features
AI/analytics readiness	Basic	Strong
Scalability	Manual (VM sizing)	Automatic

Our recommendation: We recommend **Option B (Snowflake)** for alfatec. The marginal running cost (EUR 50–200/month) is offset by zero maintenance burden, native Azure SSO, and a significantly stronger foundation for the use cases alfatec is planning – Sales Agent integration, third-party data sharing, product master data cleanup, and AI-first analytics. Snowflake also avoids an architectural dead end if Business Central moves to the cloud.

When PostgreSQL is the right choice: If the priority is a fast, minimalist working solution with the lowest possible initial complexity, PostgreSQL is a strong option. Foreign Data Wrappers can provide immediate read access to BC data without building a separate extraction pipeline – the simplest possible starting point. The trade-off is that a migration to a cloud warehouse may be required downstream as use cases expand (analytics, data sharing, AI tooling). PostgreSQL is also the natural choice if on-prem data residency is a hard requirement rather than a preference, or if dbt scheduling will be handled via a simple cron job rather than a managed orchestration platform.

SCOPE OF WORK

Regardless of the infrastructure option chosen, the engagement delivers:

Data Layer Foundation

End-to-end setup of alfatec's data layer: Business Central data extraction, automated loading, transformation framework, and documentation. Delivered as a turnkey system ready for downstream use cases.

Start: Upon agreement
Duration: 4 weeks
Model: Fixed price

DELIVERABLES

Data Extraction from Business Central

- Pipeline connecting to Business Central (API or direct database access – see Technical Notes below)
- Initial data sets: articles/product master data, sales prices, purchase prices, order backlog, and customer master data per the fields identified with alfatec's sales and digitalization teams
- Automated scheduled refresh (configurable frequency – daily is typical)

Data Layer Setup

- Infrastructure provisioning (Snowflake account or PostgreSQL VM, per chosen option)
- Schema design: raw landing zone + clean staging layer
- Access controls and user setup
- Azure AD / Entra ID SSO configuration (Snowflake option)

Transformation Layer (dbt)

- dbt project setup with best-practice structure (staging / intermediate / marts)
- Staging models for all loaded BC tables (rename, cast, deduplicate)
- Lightweight data quality tests (not-null, uniqueness, referential integrity)
- Documentation generation

Orchestration

- Scheduled pipeline execution (Airflow or equivalent)
- Failure alerts and monitoring
- Pipeline hosted on an alfatec-provided VM (on-prem) or a light cloud server

Handover

- Full technical documentation (architecture, pipeline configuration, dbt models)
- Handover session with alfatec IT

TECHNICAL NOTES

BC Data Access: API vs. Database

alfatec's Business Central runs on-prem with a Microsoft SQL Server backend. Two extraction paths are available:

Approach	Advantages	Disadvantages
Direct DB access (SQL Server read user)	Full access to all tables immediately. No schema limitations. Fastest to build. Most flexible if new data points are needed later.	Tied to on-prem SQL Server. Will not work after BC cloud migration (no direct DB access in BC SaaS).
BC API (OData / REST)	Works with both on-prem and cloud BC. Migration-safe. Uses alfatec's existing DMZ server.	API surface is more limited than the full database. Adding new data points may require BC API customization (AL development). Slower for bulk loads.

> **Our recommendation:** If BC cloud migration is **more than 6 months away**, start with **direct database access** – it is faster to deliver and more flexible. When the migration happens, the extraction layer is swapped (the rest of the stack stays unchanged). If migration is **imminent**, start with the **BC API** to avoid rework.

TIMELINE

- **Kickoff & Access** Week 1

  Credentials provisioned, infrastructure access granted (VM, BC database/API, Snowflake account if applicable). Kickoff call with alfatec project team.

- **Infrastructure & Extraction** Weeks 1–2

  Data layer provisioned. BC extraction pipeline built and tested. Initial data sets loaded. Orchestration and monitoring set up.

- **Transformation & Quality** Weeks 2–3

  dbt project set up. Staging models for all loaded tables. Data quality tests. Schema documentation.

- **Handover & Documentation** Week 4

  Technical documentation finalized. Handover session with alfatec IT. Access review and cleanup.

INVESTMENT

The project is delivered at a fixed price. Both infrastructure options are priced identically – the effort difference is negligible.

Data Layer Foundation -- Fixed Price	**EUR 10,000**

This includes:

Deliverable	Included
BC data extraction pipeline (API or DB)	✓
Data layer setup (Snowflake or PostgreSQL)	✓
dbt transformation layer with staging models	✓
Orchestration and monitoring	✓
Data quality tests	✓
Technical documentation and handover	✓

Not included in this scope:

- Sales Order Agent development
- Write-back to Business Central
- Dashboards or reporting
- Product master data deduplication
- Third-party data sharing APIs
- Additional data source connections beyond BC

These are addressed in Phase 2 and beyond (see outlook below).

Ongoing costs after delivery (borne by alfatec):

- Snowflake: EUR 50–200/month at current volumes
- PostgreSQL: included in existing VM infrastructure
- Orchestration server: alfatec-provided VM (existing infra) or light cloud VM (EUR 10–20/month)

All prices net (excl. VAT). Fixed-price items are billed upon completion. Time & materials items are billed monthly based on actual days worked. Travel expenses, if any, are billed at cost.

PREREQUISITES

- **BC access** – Read access to the Business Central database (SQL Server read user) or API credentials, depending on chosen extraction approach. alfatec IT to provision.
- **VM for orchestration** – Linux VM (alfatec on-prem or cloud) for running the extraction pipeline. alfatec IT to provision. Minimum specs: 2 vCPU, 4 GB RAM, 50 GB disk.
- **Snowflake account** (if Option B) – Gemma will set up the account; alfatec to provide billing ownership. Azure AD / Entra ID admin access for SSO configuration.
- **Git repository** – Gemma to host during delivery; transferred to alfatec at handover.
- **Point of contact** – Designated project lead on alfatec side. Weekly sync cadence during delivery.

LEGAL FRAMEWORK

The engagement is governed by the following contracts:

Document	Purpose
Framework Service Agreement (FSA)	Master terms governing the consulting relationship, liability, IP, and general obligations.
Non-Disclosure Agreement (NDA)	Mutual confidentiality agreement. Must be signed before any access is granted.
Data Processing Agreement (DPA)	Per GDPR requirements, covering any personal data processed during the engagement.
Legal Offer	A short, signable summary of the scope, deliverables, timeline, and pricing defined in this proposal.

WAYS OF WORKING

Area	Approach
Communication	Email or shared channel (Slack/Teams) for day-to-day coordination
Sync Cadence	Weekly 30-minute call to review progress
Delivery	All work in Git with pull requests and documentation
Documentation	All pipelines and models documented to handoff-ready standard
Point of Contact	Designated project leads on both sides

PHASE 2 & BEYOND

The data layer delivered in Phase 1 is a foundation. Once live, it unlocks a series of high-value use cases that build on the same infrastructure. Below is an overview of where the journey can go – each of these would be scoped and priced as a separate engagement.

PRODUCT MASTER DATA QUALITY

The most immediate follow-on: systematic deduplication and standardization of product master data in Business Central.

- **Automated matching** – similarity analysis (e.g. Levenshtein distance, fuzzy matching) to identify duplicate articles with different naming conventions
- **LLM-assisted classification** – use large language models to extract structured attributes (dimensions, capacities, categories) from unstructured text fields
- **Human-in-the-loop review** – a lightweight web interface where team members confirm or correct automated matches, with confidence scoring to prioritize manual review
- **Continuous monitoring** – ongoing process that flags new duplicates or inconsistencies as they are created in BC

SALES ORDER AGENT DATA FOUNDATION

alfatec's planned Sales Order Agent requires reliable, structured access to the company's data. Phase 1 delivers the read-side foundation. Phase 2 would extend this to:

- Optimized data models tailored to the agent's query patterns (fast lookups on articles, prices, stock, customer history)
- API layer or direct query access for the agent
- Write-back pathways to Business Central for order creation

Scoped jointly once the agent architecture is defined.

THIRD-PARTY DATA PROVISION

alfatec's initiative to provide product data to external partners and e-commerce platforms. The data layer makes this significantly cleaner:

- Curated data views (product catalogs, pricing, availability) exposed via Snowflake's secure data sharing or a lightweight API
- Data quality gates ensuring only validated data reaches external partners
- Access controls and audit logging per partner

REPORTING MODERNIZATION

alfatec's current reporting stack (Power BI connected directly to BC) has limitations in scalability and maintainability. With a proper data layer in place, reporting can be rebuilt on a more sustainable foundation:

- dbt semantic layer defining business metrics once (revenue, margins, order pipeline) with consistent logic
- Dashboards powered by the data layer rather than live ERP queries
- Self-service analytics for the team without impacting BC performance

AI-FIRST ANALYTICS

The most transformative opportunity. With a clean, well-structured data layer and a semantic layer on top, alfatec can move beyond traditional dashboards toward AI-native data access:

- **Natural language data exploration** – tools like Claude Code or Claude Cowork allow team members to query and analyze data in natural language, generating SQL, visualizations, and insights on demand. No dashboard maintenance, no report backlog.
- **Semantic layer via MCP** – the Model Context Protocol (MCP) allows AI assistants to connect directly to alfatec's data infrastructure. Define your business metrics once in the semantic layer; every AI tool – internal chatbots, Sales Agent, external integrations – queries them consistently and correctly.
- **AI-native dashboards** – instead of maintaining static reports, generate visualizations dynamically based on the question being asked. Always current, always relevant, zero maintenance overhead.
- **Conversational data access** – a Slack or Teams bot that answers business questions from live data. Team members ask questions in plain language and receive answers grounded in alfatec's actual data.

This represents a fundamentally different approach to business intelligence – one where the data layer and semantic layer are the product, and the presentation layer is generated on demand by AI rather than manually built and maintained.

This is an emerging capability that Gemma is actively deploying with other clients. We would scope this jointly once the data layer is live and alfatec has seen it in action.

ABOUT GEMMA ANALYTICS

Gemma Analytics is a Berlin-based data consultancy specializing in modern data platform implementation. We work with growth-stage companies to build production-grade data infrastructure – from warehouse setup and pipeline engineering through to KPI dashboards and team enablement.

Our team brings deep experience across the modern data stack, with 70+ completed data platform projects. We operate code-first and leverage AI-assisted tooling to deliver at speed without compromising on quality or maintainability.

Over these projects we have built extensive internal documentation, reusable patterns, and engineering best practices that carry over to every new engagement – accelerating development significantly from day one.

Gemma Analytics GmbH – Chausseestraße 17, 10115 Berlin
gemmaanalytics.com