

CrateDB Factory Intelligence Dashboard — PoC Documentation

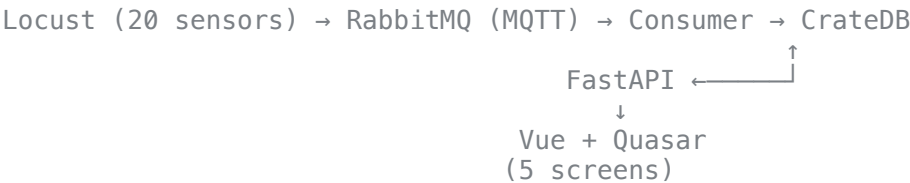
1. Purpose of the PoC

This Proof-of-Concept demonstrates how a **single CrateDB database** can replace an entire stack of specialized systems in an industrial environment. Instead of maintaining separate databases for time series, geospatial data, full-text search, and vector similarity, everything resides in one cluster with a single SQL interface.

The system simulates a smart factory with CNC machines, industrial robots, and autonomous guided vehicles (AGVs), collecting sensor data in real time and providing operators with a unified view of production status.

2. What the PoC Covers

2.1 System Architecture



- **Sensor simulator** generates data for 8 machines and 3 AGVs (temperature, vibration, RPM, power consumption, GPS positions, anomalies)
- **MQTT broker** transfers data in real time
- **Consumer** batch-inserts records into CrateDB every 5 seconds
- **FastAPI backend** serves REST API and WebSocket for live data
- **Vue/Quasar frontend** displays 5 interactive screens

2.2 Five Demonstration Screens

#	Screen	Purpose
1	Live Operations Overview	Real-time metrics, sparkline charts, AGV floor map, AI alerts
2	Time Series Deep Dive	Interactive sensor data analysis with window functions and downsampling
3	AI Predictive Maintenance	Predictive maintenance based on vector similarity to historical failures
4	Semantic Maintenance Search	Hybrid search across documentation (SOPs, manuals, incident reports)
5	Digital Twin	3D factory visualization with live machine status data

3. Demonstrated CrateDB Capabilities

3.1 Time Series

CrateDB natively supports time series without special extensions.

Key features:

- `PARTITIONED BY (month_partition)` — automatic partitioning by month for efficient management of millions of records
- `DATE_TRUNC()` + `GROUP BY` — downsampling to any granularity (1 min, 5 min, 1 hour)
- Window functions (`AVG()` `OVER`, `LAG()`, `ROW_NUMBER()`) within the same SQL query
- Fast queries: one million data points → results in milliseconds

PoC example: Screen 2 displays CNC machine temperature with a moving average, min/max range, and compression ratio — all from a single SQL query combining aggregation, window functions, and a `JOIN` with machine JSON metadata.

3.2 Geospatial Data

CrateDB supports `GEO_POINT` and `GEO_SHAPE` types with built-in spatial functions.

Key features:

- `GEO_POINT` for machine and AGV coordinates
- `GEO_SHAPE` for defining factory zones (polygons)
- `ST_WITHIN()` function for spatial queries
- Built-in indexing without additional plugins

PoC example: Screen 1 displays AGV robots in real time on the factory floor plan, with automatic zone detection for each AGV. Screen 5 (Digital Twin) visualizes positions in 3D space.

3.3 Full-Text Search

CrateDB includes a built-in full-text search engine based on Apache Lucene.

Key features:

- `MATCH()` function with configurable analyzers (`english`, `standard`, `custom`)
- `_score` for ranking results by relevance
- Indexing at write time — no separate indexing system required
- Support for multiple languages and custom analyzers

PoC example: Screen 4 enables searching across 15 maintenance documents (SOPs, manuals, incident reports, training videos) using natural language. A query such as "bearing failure high temperature" returns relevant documents ranked by `_score`.

3.4 Vector Search

CrateDB supports the `FLOAT_VECTOR(N)` type for storing embeddings.

Key features:

- `FLOAT_VECTOR(384)` columns for storing embeddings directly in the table
- Vectors reside in the same table as all other data — no need for a separate vector system
- Ability to combine with other SQL operations (JOIN, WHERE, GROUP BY)

PoC example: Screen 3 uses vector similarity for predictive maintenance. The current sensor pattern is converted into an embedding and compared against historical failures. Result: "This pattern is 82% similar to a bearing failure on CNC-003 — predicted failure in ~4.3 hours."

Note: CrateDB 5.10.2 stores `FLOAT_VECTOR` natively. The KNN operator for server-side search is available from version 6.x. In this PoC, vector similarity is computed at the application layer, while CrateDB serves as the centralized store for all vectors.

3.5 JSON / Object Data

CrateDB supports the `OBJECT(DYNAMIC)` type, enabling flexible schemas within structured tables.

Key features:

- Storage of unstructured metadata (machine configuration, raw sensor payload)
- Access to sub-fields in SQL queries: `metadata['axes']`, `metadata['spindle_rpm_max']`
- Dynamic addition of new fields without ALTER TABLE
- JOIN with other columns in the same query

PoC example: Each machine has a JSON metadata object with arbitrary fields (number of axes, max RPM, coolant type, payload capacity). Screen 2 displays this data alongside sensor values in the same query. The Digital Twin uses metadata for 3D visualization of each machine's specific characteristics.

3.6 Real-Time Ingest

CrateDB is optimized for high-throughput writes with low-latency data availability.

Key features:

- Batch insert with configurable `refresh_interval`
- Data available for queries within one second of insertion
- Support for parallel writes from multiple consumers
- Automatic partition and replica management

PoC example: The consumer continuously writes sensor data from the MQTT broker. Screen 1 displays the live ingestion rate and updates charts every 3 seconds. A WebSocket relay forwards messages directly to the frontend without an additional messaging system.

4. Key CrateDB Benefits Demonstrated by This PoC

4.1 One Database, Five Data Types

All factory data resides in a single database:

Data Type	Table	Feature
Sensor time series	sensor_readings	Time Series + Partitioning
AGV positions	agv_positions	GEO_POINT
Factory zones	factory_zones	GEO_SHAPE
Maintenance documentation	maintenance_docs	Full-Text + FLOAT_VECTOR
AI anomalies	anomaly_alerts	FLOAT_VECTOR + Time Series
Machine metadata	machines	OBJECT(DYNAMIC) + GEO_POINT

Benefit: No ETL pipelines, no cross-system synchronization, no data consistency issues.

4.2 One SQL Query, Multiple Features

Each screen demonstrates that multiple capabilities can be combined in a single SQL query:

- **Time series + JSON + Window functions** → downsampling with moving average and machine metadata
- **Geospatial + Time series** → AGV position with latest sensor data
- **Full-Text + Vector** → hybrid document search
- **Vector + Time series** → predictive maintenance from the same tables

Benefit: Operators and analysts use a single language (SQL) for all analytics, without switching between different APIs and systems.

4.3 Simplified Infrastructure

Instead of maintaining 4+ specialized systems, the PoC uses:

- 1 CrateDB cluster (single-node for PoC, horizontally scalable for production)
- Standard SQL protocol (PostgreSQL wire compatible)
- Standard REST/HTTP API

Benefit: One system for backup, monitoring, security policies, user privileges, and scaling.

4.4 Real-Time Without Compromise

CrateDB combines OLAP performance with near-real-time data availability:

- Write → available for query within one second
- Aggregations over millions of records in milliseconds
- No batch processing or ETL delays

Benefit: Operators see factory status in real time, while analysts can run complex queries on the same data without waiting.

4.5 Scalability

Although the PoC uses a single-node configuration, CrateDB is designed for horizontal scaling:

- Distributed sharding by key (CLUSTERED BY)
- Automatic rebalancing when adding nodes
- Time-based partitioning for managing data growth

Benefit: The same system running the PoC can scale to production by adding nodes, with no changes to application code or schema.

5. Conclusion

The Factory Intelligence Dashboard PoC proves that CrateDB can serve as a **unified data platform for industrial intelligence**. Five system screens demonstrate all key capabilities — from real-time sensor data ingestion, through geospatial visualization and predictive maintenance, to semantic documentation search and a 3D digital twin.

The key message: **everything an industrial factory needs from a database, CrateDB delivers in a single system, with a single SQL interface, with no compromise in performance or functionality.**