
Tech4Biz
Excellence Guaranted!

Tech4Biz
Excellence Guaranted!

CASE STUDY 06

AUTONOMOUS INDUSTRIAL DIGITAL TWIN

A Live Model of the Plant, Built on Streaming Data and Governed on a Databricks Lakehouse

We built a continuously updated model of the assets, the lines and the supply chain, fed by the controllers and the historian, corrected against what actually happened, and used to tell the plant what to do next. It is not a three dimensional picture of the factory. It is the plant's own data, contextualised well enough that a model can reason about it.

The Business Problem

Where the data is trapped

A manufacturing site already generates more data than it uses. The controllers know the state of every valve and every drive. The historian holds years of process values. The manufacturing execution system knows what was made and when. Quality results sit in the laboratory system, maintenance in the computerised maintenance management system, cost and planning in the enterprise resource planning system.

Each of those systems is correct on its own and none of them can answer the question that matters. When a batch fails final inspection, nobody can say which process window produced it, which machine was drifting at the time, and whether the same conditions are present on the line right now.

What the site was doing instead

- **Export to a spreadsheet.** A process engineer pulls a historian trend, joins it by hand to a quality result and reaches a conclusion that cannot be repeated next month.
- **Run to failure or run to calendar.** Maintenance is done on a fixed interval because condition data exists but has never been connected to failure outcomes.
- **Inspect at the end.** Defects are found after the value has already been added, so scrap is discovered at the most expensive possible moment.
- **Optimise one line at a time.** Because nobody has a view that spans lines, sites and the supply that feeds them.

The result is a site that is heavily instrumented and still managed by experience.

Where the cost sits

Cost	Driver
Unplanned downtime	Failures that had a detectable signature weeks earlier in data nobody was reading.
Scrap and rework	Process drift detected at inspection instead of at the point it started.
Over maintenance	Components replaced on a calendar while healthy, and others missed while degrading.
Yield loss	Lines run at a conservative setpoint because nobody can prove the wider window is safe.
Energy per unit	Utilities treated as a fixed overhead rather than as a controllable input.
Changeover time	Sequencing decided by habit rather than by the actual changeover matrix.
Knowledge loss	Failure mode experience leaving the site with retirement.

> **THE REQUIREMENT IN ONE SENTENCE**
>
> Bring controller, historian, execution, quality, maintenance and planning data into one governed lakehouse, contextualise it against a single asset model, keep a continuously corrected twin of every significant asset and line, and use it to predict failures, hold quality inside the window and schedule production against what the plant can actually do today.

Solution Overview

Data is collected at the edge, contextualised against one asset hierarchy, curated into a lakehouse, and used to maintain a twin state for every asset that matters. Models run against the twin rather than against raw tags, which is what makes them portable between lines and between sites instead of being rebuilt for each one.

Layered architecture

LAYER 7 DECISION

Operations command centre

Plant, line and enterprise views, with the same numbers at every level rather than three reports that disagree.

LAYER 6 AGENTS

Operations and maintenance copilots

Advisory workflows for shift handover, fault diagnosis, work order preparation and scheduling scenarios.

LAYER 5 OPTIMISATION

Simulation, setpoint advice and scheduling

Reinforcement learning and constrained optimisation run against the twin, never directly against the plant.

LAYER 4 TWIN

Asset state, physics and calibration

A live state object per asset, a model of how it should behave, and the residual between the two.

LAYER 3 CONTEXT

Asset hierarchy and unified namespace

Every tag bound to an equipment, a line, a site, a product and a production order.

LAYER 2 CURATION

Delta Live Tables, bronze to gold

Time series conditioned, resampled and joined to events, with quality expectations at write time.

LAYER 1 EDGE

Acquisition and buffering at the plant

OPC UA, MQTT and historian collectors with local store and forward across a link that will drop.

The advisory boundary

The platform reads the plant continuously and writes to it only through a path the control engineers own. That boundary is what makes the system approvable by the operations technology security team and by the process safety authority.

The platform does	The platform never does
Read tags, batches and events	Write to a safety instrumented system
Predict a failure and rank it	Stop a line on its own authority
Recommend a setpoint change	Move a setpoint outside the approved envelope
Propose a production sequence	Release a sequence without planner approval
Raise a quality alarm	Release or reject a batch
Prepare a work order	Bypass a permit to work

CLOSED LOOP IS A LATER PHASE, AND IT IS A SEPARATE CONVERSATION

Advisory setpoint recommendation and closed loop control are different systems with different assessment burdens. Where closed loop is genuinely wanted, it is introduced later, on one loop, inside a bounded envelope enforced in the control system rather than in the platform, and only after the advisory version has been trusted for months.

Edge Acquisition and Protocols

Protocol coverage

Protocol	Typical source
OPC UA	Modern controllers and the plant's structured data layer, including the type information that makes contextualisation possible.
OPC DA and classic	Older cells where a bridge is needed rather than a rewrite.
Modbus TCP and RTU	Drives, instruments, protection relays and legacy equipment.
Siemens S7 and Rockwell	Controller families where a native driver is faster and lighter than a gateway.
MQTT with Sparkplug	Existing telemetry brokers, and the preferred shape for new edge devices because state and birth certificates come with it.
Historian query	Bulk history for model training and for trend context that the live tags cannot give.
MES and ERP	Production orders, materials, batches and cost, through APIs or change capture.
Vision and audio	Cameras and acoustic sensors, processed at the edge with only the result and the exception image sent upstream.

What the edge layer has to survive

CONNECTIVITY

Store and forward is mandatory

The plant link will drop. The collector buffers locally with enough capacity for the longest realistic outage and replays in order, so a network problem never becomes a data gap.

SAMPLING

Rate chosen per signal, not per plant

Vibration for bearing analysis needs kilohertz sampling and edge feature extraction. A tank level needs one sample a minute. Sending everything at the highest rate is the fastest way to make the platform unaffordable.

COMPRESSION

Deadbands understood, not inherited

Historians already apply compression. Training a model on compressed data and then serving it on raw data produces a silent mismatch, so the deadband settings are recorded as part of the signal definition.

TIME

One clock across the site

Controllers, cameras and servers synchronised, because correlating a defect image with a process value is impossible if the timestamps disagree by seconds.

TAG NAMING IS THE HIDDEN PROJECT

Most sites have tag names that encode meaning in a convention that was consistent in 1998 and has drifted since. Reconciling tags to equipment is usually the single largest task in the first phase. Survey it before committing to a date, because it is entirely knowable in advance and it is the most common cause of a slipped first milestone.

Context and the Asset Model

A raw tag means nothing on its own

A value of 71 degrees is not information. The same value becomes information when the platform knows it came from the drive end bearing of pump six, that pump six is part of the cooling loop on line three, that line three was running product B on order 44817 at the time, that the ambient was 38 degrees, and that this reading has risen from 58 over six weeks.

The asset model is what carries all of that. It follows the standard equipment hierarchy of enterprise, site, area, line, unit and equipment, and every signal is bound to a node in it. Models are then written against the node type rather than against a tag list, which is what allows a pump model built on line three to be deployed on line seven without rewriting it.

What context is joined

- Equipment hierarchy
- Production order
- Product and recipe
- Batch and lot
- Shift and crew
- Material genealogy
- Maintenance history
- Quality result
- Ambient conditions
- Energy meter

The contextualised state object

```
{
  "asset"        : "SITE-A/LINE-03/PUMP-06",
  "type"         : "centrifugal_pump",
  "read_at"      : "2026-07-29T14:22:09Z",
  "order"        : { "id": "44817", "product": "B", "since": "06:10" },
  "signals": [
    { "tag": "VIB-01", "value": 4.2, "unit": "mm/s",
      "normal": "< 2.8", "trend_6w": "rising from 1.8" },
    { "tag": "TE-02",  "value": 71,  "unit": "degC",
      "normal": "45 to 60", "trend_6w": "rising from 58" } ],
  "twin": { "expected_vib": 2.4, "residual": 1.8,
            "health_index": 0.42, "model_rev": "pump-v7" },
  "state"        : "running, no active alarm",
  "runtime_hours": 41280,
  "last_maint"   : "2025-12-14"
}
```

BUILD THE HIERARCHY ONCE AND MAKE IT THE ONLY ONE

Sites usually already have three asset hierarchies, one in the maintenance system, one in the control system and one in finance, and they do not agree. Reconciling them into a single model, with the others mapped onto it, is unglamorous work that decides whether every later analytic is trivial or impossible. Do it in the first phase and give it a named owner.

The Digital Twin Engine

What the twin actually is

The twin is three things held together. A current state assembled from the plant, a model of how the asset should behave under those conditions, and the residual between the two. The residual is where all the value sits, because a machine that is degrading looks normal against a fixed threshold and abnormal against its own expected behaviour.

COMPONENT 1

State

Live and recent signals, contextualised, held at a resolution that suits the asset class.

COMPONENT 2

Behaviour model

First principles relationships where the physics is known, such as pump curves, heat balance and motor load, combined with a learned model for the parts that are not worth deriving.

COMPONENT 3

Residual and health index

The gap between expected and observed, normalised so that assets of different sizes and duties can be ranked against each other on one screen.

Keeping the twin honest

A twin that is not corrected drifts away from the plant within weeks. Parameters are recalibrated on a schedule against recent normal operation, and every recalibration is versioned, so a change in the twin can be distinguished from a change in the machine.

Fidelity is chosen per asset, not per plant

Tier	Where it is worth it
Statistical	Large populations of similar low criticality assets. Cheap, and enough to rank them.
Behavioural	Assets with a known operating curve. Expected value derived from load and conditions, residual tracked.
Physics based	Critical or expensive equipment where a mass and energy balance is worth building and can be validated.
Line level	A discrete event model of the line, used for throughput, buffer and changeover questions rather than for equipment health.
Network level	Supply, inventory and multi site allocation, run as scenarios rather than continuously.

SIMULATION IS FOR QUESTIONS, NOT FOR DECORATION

A twin justifies itself by answering a question somebody was going to answer anyway. What happens to throughput if we run this order first. How much life is left on this bearing at the current duty. What does the energy cost look like if we shift the compressor load. Start from the question and build the fidelity that question needs, because a high fidelity model of something nobody asks about is expensive to build and expensive to keep true.

EVERY TWIN NEEDS A MAINTENANCE PLAN OF ITS OWN

Models decay when the plant is modified, when a component is replaced with a different make, or when the product mix changes. Recalibration cadence, ownership and a rollback path are part of the deliverable, not an afterthought.

The Model Estate

What runs, and what it is for

Model	Job and shape
Predictive maintenance	Remaining useful life and failure probability per asset class, trained on residuals and on the maintenance record rather than on raw tags, so it transfers between sites.
Anomaly detection	Multivariate detection on the twin residual, which catches combinations that no single tag threshold would raise.
Visual inspection	Defect detection and classification at the line, running at the edge with only results and exception images sent upstream.
Quality prediction	Predicting the laboratory result from process conditions, so a drift is visible hours before the sample is taken.
Root cause analysis	Ranking the process variables most associated with a quality excursion, with the correlation stated as correlation.
Energy and utilities	Forecasting demand and identifying the load that can be shifted without touching output.
Throughput and scheduling	Sequencing against real changeover times, real availability and real yield rather than against the planning assumptions.
Supply and inventory	Demand sensing and buffer sizing across the network.

Training data is the constraint

Every predictive maintenance programme meets the same wall. Failures are rare, which is the point of the plant, and a model needs examples. Three things make it workable.

- **Label from the maintenance record, carefully.** A work order tells you something failed. It rarely tells you when degradation started, so the labelling window is agreed with the reliability engineers rather than assumed.
- **Pool across identical assets.** Forty pumps of the same type on one site give a usable population where one pump gives nothing.
- **Use the residual, not the raw signal.** Because the residual is already normalised for duty and conditions, which is most of the variance that would otherwise swamp the failure signal.

Lifecycle

REGISTER
Data, code and parameters logged per run

VALIDATE
Tested on assets and periods held out entirely

SHADOW
Predictions recorded and compared before anyone acts on them

MONITOR
Drift on inputs, and precision measured against closed work orders

Agents and the Reasoning Layer

Bounded copilots, not a general assistant

Each agent has a defined question set, a defined set of tools and a defined authority limit. It reads the twin, searches the plant document corpus, queries the lakehouse and produces an answer with references that the engineer can open.

Agent	Scope and limit
Maintenance copilot	Diagnoses a symptom against failure modes and asset history, drafts the work order with parts and duration. It cannot release the work order.
Operations copilot	Explains why a line is behind, which constraint is binding and what recovering it is worth. It cannot change the plan.
Quality copilot	Traces an excursion back through process conditions and material genealogy. It cannot release or reject a batch.
Scheduling copilot	Runs sequence scenarios against the twin and presents the trade off. The planner chooses.
Shift handover	Writes the handover from events, alarms and interventions, for the outgoing supervisor to correct and sign.
Energy copilot	Identifies shiftable load and states the effect on output before it states the saving.

The plant corpus

- Equipment manuals
- P&ID and wiring
- Standard procedures
- Maintenance history
- Failure mode analyses
- Vendor bulletins
- Change records
- Shift logs

Grounding contract

Condition	Required behaviour
Supported by documentation	State the hypothesis and cite the section.
Signal without documentary support	Report the anomaly, say the documentation does not cover it, and do not speculate on cause.
No relevant document	Say so, and record the gap for the next ingestion cycle.
Safety relevant condition	Give the documented instruction without softening it, and prompt for the permit to work requirement.
Several plausible causes	Rank them and state the discriminating test for each.
Conflicting documents	Prefer the most recent vendor bulletin and say the guidance was updated.

CITATIONS ARE MANDATORY

Every substantive claim carries a document, a section and a data reference the engineer can open. That is what makes the answer checkable before someone acts on it, and it is the single feature that most determines whether experienced plant staff trust the system or ignore it.

CORRELATION IS STATED AS CORRELATION

A root cause agent that presents an association as a cause will be right often enough to be believed and wrong often enough to be dangerous. The wording is part of the design, and the agent is required to name the experiment or the check that would confirm it.

Worked Example, End to End

TRIGGER

Quality flags a rise in surface defects on line three. The rate has gone from under one percent to four percent over two shifts and inspection is catching it at the end of the line.

STEP 1 SCOPE

Defects joined to genealogy

Every flagged unit traced back to the order, the material lot, the shift and the machines it passed through.

STEP 2 NARROW

The pattern is not uniform

The defect rate is concentrated in units produced while station four was running, and absent on the parallel station.

STEP 3 RESIDUAL

Twin checked for station four

Spindle vibration residual has been rising for eleven days while the absolute value stayed inside the alarm limit.

STEP 4 CORRELATE

Process and quality aligned in time

Defect onset matches the point where the residual crossed the level seen before the last two spindle bearing replacements.

STEP 5 ACT

Work order drafted, sequence rerun

Parts checked in stores, the intervention costed against continued scrap, and the schedule replanned around the window.

PROCESS ENGINEER

"Why did nothing alarm if this started eleven days ago?"

MAINTENANCE COPILOT

"The alarm limit on SPN-04-VIB is 6.0 millimetres per second and the reading is 4.4, so it never tripped. Against the twin the expected value at this load and this product is 2.1, so the residual is 2.3 and it has been rising steadily since the fourteenth. The same residual profile preceded the bearing replacements in March 2025 and in September 2025, both on this station. See the spindle maintenance manual, section 4.2, and work orders WO-20194 and WO-24471."

PROCESS ENGINEER

"Could it be the material lot instead?"

QUALITY COPILOT

"Unlikely. Lot 8841 ran on both stations for six hours on Tuesday. The defect rate on station four was 3.8 percent and on station five it was 0.4 percent over the same lot, which separates the machine from the material. If you want to rule the lot out completely, the incoming inspection record for 8841 is attached and the retained sample is still in the laboratory."

WHAT MADE THIS WORK

No new sensor and no new model. It needed the vision result, the process values, the genealogy and the maintenance history in one governed place, bound to the same asset model, with a twin that gave an expected value rather than only a fixed limit.

Security and OT to IT Convergence

Network position

The platform's plant footprint is an operations technology asset and it is treated as one. It sits where a monitoring device sits, not where a control device sits, and the flow of data is outward.

Control	Implementation
Zone and conduit	Collectors placed in a defined zone with declared conduits, rather than introducing a new trust boundary across the site.
Read only clients	No write function compiled into the acquisition layer for any signal outside an explicitly approved write list.
Unidirectional option	A data diode where the site requires physical enforcement rather than configuration.
Outbound only	The plant initiates the connection to the cloud. No inbound path from the enterprise into the control network.
Authentication	Certificate based, with OPC UA sessions signed and encrypted rather than left on the anonymous endpoint.
Patching	Edge software updated through the site's existing controlled change process, on the site's maintenance window.
Segregation of duties	The team that can change a control system is not the team that can change the platform.

Why this is the whole approval argument

An operations technology security team assesses a new system on two questions. Can it change anything in the plant, and can anything reach the plant through it. This architecture answers no to both by construction rather than by configuration, which is what makes the assessment tractable and short.

> **STANDARDS CONTEXT**
>
> The design follows the segmentation and zone and conduit model that industrial control system security standards expect, with the platform placed as a monitoring asset inside an existing zone. Exact placement, conduit definitions and the assessment evidence are agreed with the site security team during Phase 0 rather than presented at deployment.

Governance on the enterprise side

- **One catalogue across sites**, with a site seeing its own data by row policy and the group seeing the aggregate, so multi site analytics does not require copying data around.
- **Recipes, formulations and process windows treated as sensitive**, masked by default and released by role, because this is usually the most commercially valuable data the company owns.
- **Lineage from a board level number back to a tag**, which is what makes an operations claim auditable.
- **Retention by data class**, since high rate vibration data and monthly cost data do not need the same treatment or the same cost.

How We Delivered It

The phases we ran

Phase	Scope	What it produced
Phase 0 Assessment	Asset and tag survey, protocol survey, historian audit, network placement, security engagement, one asset class chosen.	Tag to asset mapping quality measured and network design agreed.
Phase 1 Foundation	Edge collectors, streaming into bronze, asset hierarchy built, silver and gold for one line.	Data reconciled against the historian and the execution system for a full month.
Phase 2 Twin	Behaviour models and residuals for the chosen asset class, calibrated on recent normal operation.	Residual stable on healthy assets and elevated on assets known to be degrading.
Phase 3 Advisory pilot	Predictive maintenance and quality prediction in shadow, with a small reliability and process group reviewing every output.	Precision and lead time accepted by the reliability engineers.
Phase 4 Line rollout	Additional asset classes and lines, copilots released, command centre live.	Downtime and scrap targets held for a quarter.
Phase 5 Multi site	Model and pipeline packaging standardised so a new site is a deployment rather than a project.	Repeatable site onboarding package.

How it runs now

EVERYTHING AS CODE

Pipelines, twins and dashboards in source control

A new site is configured, not rebuilt, which is the difference between a programme that scales and one that stalls after the pilot plant.

MODEL OPERATIONS

Recalibration on a cadence with rollback

Twin parameters and models versioned together so a change in behaviour can be attributed to the machine or to the model.

FEEDBACK

Every prediction closed out against the work order

Precision measured against what was actually found when the machine was opened, not against a proxy.

OWNERSHIP

A named site owner from Phase 1

Plant staff involved in building it, because a platform only the vendor can operate does not survive the second year.

THE PILOT TRAP

Most industrial artificial intelligence programmes succeed on one line and never leave it, because the pilot was built by hand against one tag list. Designing for the second line during the first one costs a little more in Phase 1 and is the difference between a case study and a capability.

Benefits and Measurement

Benefits realised

Failures seen earlier

Degradation detected against expected behaviour rather than against a fixed alarm limit.

Less scrap

Process drift visible while the material can still be saved.

Maintenance on condition

Work done when the asset needs it instead of when the calendar says so.

Yield from a wider window

Setpoints moved with evidence rather than left conservative out of caution.

Schedules that hold

Sequencing against real availability, real changeover and real yield.

Energy as a variable

Load shifted where it does not cost output.

Institutional memory

Prior failures on the same asset surface automatically rather than living in one person's head.

One number per site

Plant, region and group reading the same figure from the same table.

KPI framework

Measure	What it tells you
Overall equipment effectiveness	The headline number, split into availability, performance and quality so the movement can be attributed.
Unplanned downtime hours	The outcome the maintenance case rests on.
Prediction lead time	How much warning the site actually got, which decides whether the warning was useful.
Prediction precision	Share of alerts confirmed when the machine was opened. Low precision destroys trust faster than low recall.
Scrap and rework rate	Quality outcome, measured per line and per product.
First pass yield	Whether the process is being held inside the window rather than corrected after.
Energy per unit produced	Utilities normalised for output, which is the only comparable form.
Tag to asset coverage	How much of the plant the platform can actually reason about.

BASELINE FIRST

Capture current overall equipment effectiveness, unplanned downtime, scrap rate and maintenance cost per asset class before anything is deployed. Most sites already record these, which makes manufacturing one of the easier domains in which to build a defensible before and after picture. Agree the definitions in writing, because two departments rarely calculate effectiveness the same way.

Scale and Performance

Scale the platform was built to carry

Metric	Enterprise scale
Manufacturing sites	20 to 100
Connected devices	250,000 to 1 million
Sensor readings	150,000 to 500,000 per second
Daily data ingestion	50 to 150 TB
Video streams	1,000 to 4,000
Robots connected	500 to 5,000
Digital twins maintained	20,000 to 100,000

Engineering targets

Measure	Target
Streaming latency	Under 2 seconds
Predictive maintenance inference	Under 1 second
Computer vision processing	25 to 35 frames per second
Data availability	99.9 percent
Model deployment frequency	Weekly
Equipment data synchronisation	Near real time

Business improvement ranges

Measure	Expected improvement
Unplanned downtime	20 to 35 percent reduction
Maintenance cost	15 to 25 percent reduction
Production throughput	10 to 18 percent increase
Overall equipment effectiveness	8 to 15 point improvement
Product defect rate	18 to 30 percent reduction
Energy consumption	10 to 18 percent reduction
Spare parts inventory	15 to 20 percent reduction
Maintenance planning time	40 to 60 percent faster

HOW TO READ THESE FIGURES

The scale column describes the volumes the platform was engineered to carry. The target column is the service level it is built and operated to, and it is measured continuously. The improvement column is the range this class of platform delivers, and it is confirmed against a client's own baseline during assessment rather than claimed in advance. We capture that baseline before the first pipeline is written, because a benefit without a baseline is an argument rather than a result.

THE NUMBER THAT GOVERNS THE DESIGN

Sensor readings at half a million a second is what forces feature extraction to the edge. Streaming full rate vibration and video to the cloud would work technically and would make the platform unaffordable within a quarter. The edge does the reduction, the lakehouse does the correlation, and raw waveform is retained only in the window around an event.

Platform Capability Across the Estate

The three platforms are built on the same Databricks foundation and are operated to the same envelope. The figures below describe the capability of that foundation across the estate rather than the load of any single engagement, and they are the numbers we size and design against when a new platform is scoped.

Platform metric	Typical target
Data ingestion	20 to 150 TB per day
Structured streaming throughput	20,000 to 500,000 events per second
Historical lakehouse	2 to 30 PB
Delta tables	5,000 to 25,000
Models under management	100 to 500
Feature store features	10,000 to 100,000
Vector embeddings	100 million to 2 billion

WHY WE PUBLISH THE ENVELOPE RATHER THAN ONE NUMBER

Sizing is driven by event rate and retention, not by headcount or revenue. Quoting a single figure invites a comparison that does not hold. The range is what we design and cost against, and the point inside it is settled during assessment from the client's own event volumes.

Platform metric	Typical target
Daily inference requests	10 to 100 million
Enterprise users	5,000 to 50,000
Platform availability	99.9 to 99.95 percent
Automated data quality checks	Above 95 percent of published tables
Governance coverage	100 percent of production datasets
Mean time to detect a data issue	Under 15 minutes

THE TWO THAT MATTER MOST

Governance coverage and mean time to detect are the two we hold hardest. A production dataset outside the catalogue cannot be audited, and a data issue that is found by a business user rather than by a monitor has already cost the client the thing the platform was bought to protect.

What We Learned

The hard problems, and what we did about them

Problem	What we did
Tag to asset mapping	Surveyed and measured in Phase 0 with an explicit budget. This is the dominant driver of first phase duration and it is knowable before anything is committed.
Historian compression	Deadband and compression settings recorded per signal, and training data drawn the same way the live path will see it.
Few failure examples	Pool across identical assets, label with the reliability engineers, and start with anomaly detection where supervised learning is not yet possible.
Plant network constraints	Edge buffering sized for the longest realistic outage, and bandwidth calculated from the real sampling plan rather than assumed.
Operations technology approval	Security team engaged in Phase 0, not at deployment, with read only enforced in the client and in the network.
Operator trust	Advisory pilot with a small group, mandatory citations, and every prediction closed out against what was found.
Model decay after plant changes	Change records fed into the platform so a modification invalidates the affected twin rather than silently degrading it.
Cost of high rate data	Feature extraction at the edge for vibration and acoustic signals, with raw waveform kept only around events.

What we settled before writing any code

- **The first asset class.** Which line and which asset class would go first, chosen so the value was measurable inside two quarters.
- **Control network exposure.** What was available, through which protocol, and whether an OPC UA server already existed.
- **The historian.** Which product, what retention, and whether it could be queried in bulk without affecting operations.
- **The asset hierarchy.** Whether a maintained hierarchy existed anywhere, and who owned it.
- **Maintenance history.** Whether the records were detailed enough to label failures, and who could authorise their use.
- **Security ownership.** Who owned operations technology approval, and what their assessment process actually was.
- **Closed loop intent.** Whether closed loop control was an eventual objective, and whether the control system supported a bounded envelope.
- **Site count.** How many sites were in scope, and how similar their control systems were.

> **WHAT WE WOULD TELL THE NEXT CLIENT**
>
> A data readiness assessment on one line. It costs little, it produces a measured picture of tag coverage, historian quality and maintenance record usability, and it tells you honestly whether a predictive programme is supportable before any model is trained. If the data is not there, that finding is worth having in six weeks rather than after a pilot underperforms.

Contact

Website

https://tech4bizsolutions.com

Contact details

If you would like to know more about Tech4Biz Solutions
and our products please contact us via email contact@tech4biz.io

Address

Bangalore
NGR Arcade,
HSR Layout Sector 1,
Bangalore, Karnataka

Surat
306-307,
Millionaire Business Park,
LP Savani Road, Adajan,
Surat, Gujarat - 395009





Tech4Biz is a provider of comprehensive engineering and IT solutions for businesses of all sizes. From cloud computing and data management to security, networking and embedded intelligence, we have the expertise and experience to help your business stay ahead of the curve.

Our mission is to provide businesses with the best possible technology solutions. We understand that a reliable and efficient technology foundation is crucial in an increasingly competitive marketplace. That is why we offer a wide range of services, from managed IT and cloud computing to data platform, applied AI and embedded engineering.

This communication contains general information only, and Tech4Biz Solutions Private Limited is not, by means of this communication, rendering professional advice or services. This document describes work delivered by Tech4Biz Solutions Private Limited. The client has requested that it not be identified, so the client name, commercial terms and any figures that could identify the organisation have been withheld. Architecture descriptions, performance characteristics, sizing figures and integration details are specific to the environment in which the work was delivered and are confirmed against the specific customer platform, workload and data volumes in any subsequent engagement. Third party product names are the property of their respective owners. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser. Tech4Biz Solutions Private Limited shall not be responsible for any loss whatsoever sustained by any person who relies on this communication.

© 2026 Tech4Biz Solutions Private Limited