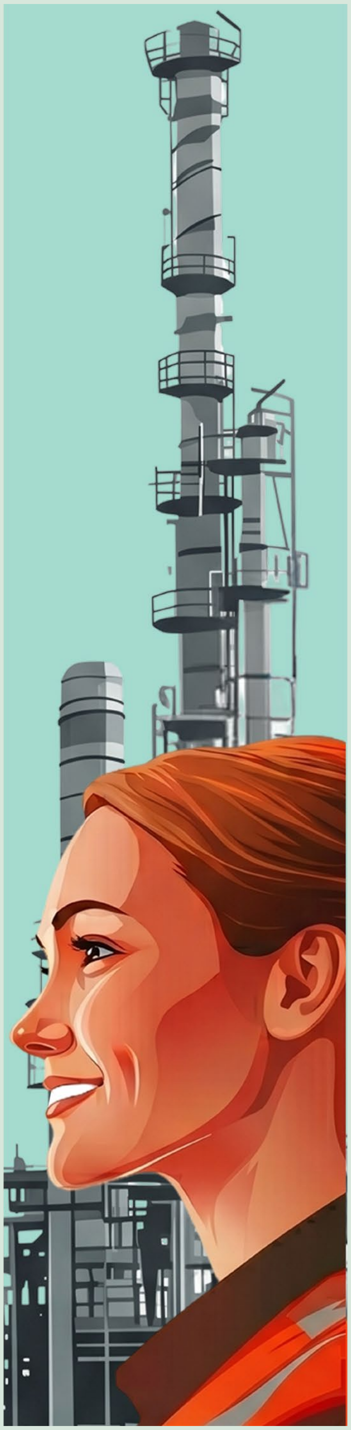
WHITE PAPER



AI-DRIVEN, HUMAN-CENTRIC PRODUCTION OPERATIONS AND OPTIMIZATION ON DATABRICKS

databricks

Infosys®
Navigate your next

Contents

Contents 2
Executive Summary 3
1. Introduction 4
2. Illustrative Outcome Envelope 5
3. Structural Bottlenecks: The Predicament of the Energy Industry 6
4. AI-driven, Human-centric Production Operations Workflow 7
5. Core Production Optimization Workflows 8
    5.1 Production optimization 8
    5.2 Fleet-scale decline curve analysis 8
    5.3 Production surveillance 8
    5.4 Flow assurance 9
    5.5 Intervention prioritization 9
    5.6 Operating cost and chemical optimization 9
6. Reference Architecture on Databricks 10
    Architecture Layer Summary 11
7. Why Databricks for Production Optimization 11
8. Governance and Human Authority 12
9. Implementation Roadmap 13
10. Conclusion 14
About the Author 15

Executive Summary

Upstream operators in the oil and gas industry are custodians of large quantities of diverse data. This includes reservoir and fluid data, drilling and completion data, and geological data. However, the true value of this valuable data is significantly constrained by how quickly it can be used to drive well-informed and reliable production decisions. An exploration and production (E&P) landscape brings together multiple factors such as supervisory control and data acquisition (SCADA) standards, data historians, and enterprise resource planning (ERP) systems on one side, with volumes of well completion and surveillance data on the other. The complex nature of the E&P environment, along with the large volumes of data acquired, can cause delays at various levels—in loss detection, flow assurance response, and intervention planning.

This is where Infosys' Production Optimization solution can make a difference. Powered by the Databricks, it is rooted in Infosys' deep expertise in the upstream energy domain. The solution creates a governed production intelligence layer that unifies exploration and production data, AI/ML-based forecasting, surveillance analytics, and flow assurance into a single operating stack.

This white paper looks at the challenges of the oil and gas industry and explores how such a solution can transform upstream operations.

1. Introduction

In the highly capital-intensive business of oil and gas exploration and production, every erroneous or delayed decision can prove costly. Unlocking the potential of a sea of valuable data lying dormant within multiple enterprise systems has become both a key challenge and a top priority for upstream energy businesses today. Insights hidden in this data can help exploration and production (E&P) companies make the right decisions at speed to avoid losses and ensure successful outcomes. Multiple individual systems functioning simultaneously in silos make it difficult to obtain these insights swiftly.

Infosys' Production Optimization solution unifies data with precise forecasting models, surveillance analytics, and flow assurance methodology by creating a governed production intelligence layer. While Databricks strengthens the solution with data intelligence for valuable insights, the Infosys team's in-depth expertise in upstream energy helps extract value and drive optimization. By bringing these varied elements together harmoniously on one operating stack, Infosys empowers upstream operators to leverage artificial intelligence/ machine learning (AI/ML) for stronger results.

To drive measurable value, the solution leverages AI for automated data engineering, deviation detection, opportunity ranking, evidence-pack generation, and forecast improvement. This enables production teams to shift their focus from reactive monitoring to continuous, AI-assisted optimization.

The Infosys Production Optimization solution strengthens four high-impact capabilities for upstream operators:

Automated integration

Probabilistic fleet forecasting

Real-time anomaly surveillance

Predictive flow assurance

Engineers receive faster insights that enable prioritized actions while retaining control over reserves, health, safety, and environment (HSE) practices, capital expenditure, and operational setpoints.

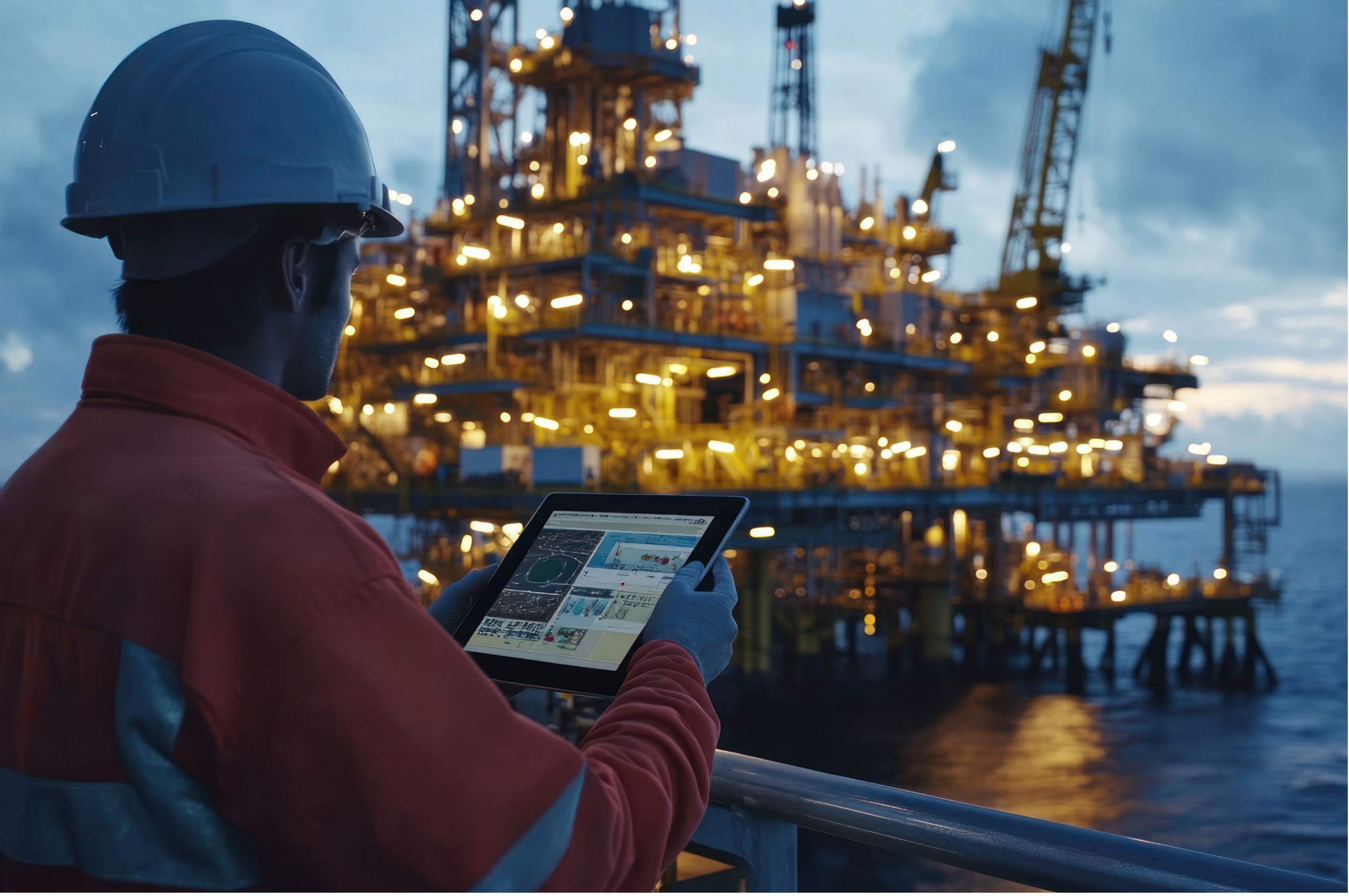
The governed production operations workflow on Databricks delivers a multifaceted outcome with reduced deferment, faster interventions, stronger production efficiency, improved lifting - cost performance, and greater reserves confidence.

2. Illustrative Outcome Envelope

This outcome framework translates the production optimization stack into measurable business value for upstream operators through faster loss detection, shorter intervention cycles, broader surveillance coverage, stronger forecast confidence, and improved lifting-cost efficiency.

BUSINESS METRIC	BASELINE/CURRENT STATE	TARGET	BUSINESS RELEVANCE
Production Loss Detection Latency	30+ days through monthly reviews	**2–4 days through daily AI-assisted surveillance**	• Detect underperformance earlier • Reduce deferred production exposure
Intervention Cycle Time	3–6 weeks from issue identification to action	**1–2 weeks with ranked opportunities and evidence packs**	• Accelerate decision making for engineers • Improve response discipline
Production Forecast Accuracy	~20–25% error in medium-term well forecasts	**~10–13% error with hybrid physics + ML forecasting**	• Improve production planning • Increase reserves confidence • Enhance asset performance visibility
Engineering Productivity/ Surveillance Coverage	Manual review of limited well sets	**3–5 times broader well surveillance coverage**	• Enable engineers to manage larger portfolios • Cut proportional headcount increase
Lifting Cost/Operating Efficiency Impact	Reactive optimization and broad operating actions	**3–7% controllable cost-efficiency improvement**	Support lower operating cost per barrel of oil equivalent (BOE) through • Faster interventions • Better chemical targeting • Reduced production losses

Note: These ranges are illustrative planning envelopes and actual values should be calibrated based on asset type, data maturity, well count, operating model, and change adoption.



3. Structural Bottlenecks: The Predicament of the Energy Industry

The delay in obtaining the required data has been the most common reason for bottlenecks in the traditional E&P scenario. Today, most upstream operators have digitalized their production data using supervisory control and data acquisition (SCADA), data historians, production accounting systems, engineering workflows, and cloud data platforms. Therefore, the key challenge is no longer delays in data availability. Now the bottleneck is around converting the fragmented data into reliable, prioritized operating decisions.

Data from multiple systems such as SCADA, data historians, enterprise resource planning (ERP) systems, production accounting, subsurface, workover, and chemicals often remains disconnected. This obscures a single view of well performance, deferment, constraints, and opportunity value, which is critical for accurate decision-making.

Traditional deterministic analysis and monthly surveillance cycles tend to delay the detection of underperformance. This slows down the process of recognizing flow assurance risks and intervention needs. Dashboards may show deviations, but they rarely translate them into ranked actions, evidence packs, and workflow-ready recommendations.

This is where AI can be leveraged to close the gap by enabling continuous surveillance, physics-informed forecasting, anomaly detection, and engineer-assisted intervention planning. With rapid developments in AI technology, upstream operators today have the opportunity to create a governed production intelligence layer that improves deferment reduction, intervention speed, lifting-cost efficiency, and forecast confidence while ensuring human authority over operational decisions.

4. AI-driven, Human-centric Production Operations Workflow

The operations workflow begins with the ingestion of trusted production data from SCADA, data historians, ERP, wells, facilities, and external sources. The Databricks Lakehouse architecture then standardizes this data into a governed production intelligence layer, while AI/ML models continuously compare expected and actual performance.

Surveillance agents detect deviations and classify their likely causes. Based on this analysis, the agents generate evidence packs and rank intervention opportunities. Engineers review the recommendations within governed workflows while retaining complete authority over decisions related to reserves, HSE, capital expenditure, chemical programs, and operating setpoints.

This creates a closed-loop production optimization model that connects various points in decision-making: data capture, diagnosis, decision support, and human-approved action across wells, facilities, and assets.

5. Core Production Optimization Workflows

The production optimization stack of the Infosys Production Optimization solution encompasses six workflows that convert data into operational action—from portfolio forecasting and surveillance to flow assurance, intervention prioritization, cost optimization, and human-approved execution. Let us look at these workflows in the context of some real-life scenarios and challenges in the upstream energy industry.

5.1 PRODUCTION OPTIMIZATION

Problem

When production teams manage well performance, facility constraints, artificial lift, operating envelopes, and deferment across disconnected workflows, they find it challenging to pinpoint where exactly value is being lost.

Use case

In a mature offshore asset, a group of high-rate wells begins underperforming against expected deliverability.

The system correlates multiple parameters such as wellhead pressure, choke settings, separator backpressure, and recent operating changes to determine whether the loss is related to the well, flowline, or facility.

Approach

Integrate real-time telemetry, production history, allocation data, operating conditions, workover context, and facility constraints. Use the data to continuously identify production gaps and recommend the next-best engineering action.

Outcome

Continuous production optimization that reduces deferment, improves production efficiency, accelerates intervention decisions, and strengthens lifting-cost performance.

5.2 FLEET-SCALE DECLINE CURVE ANALYSIS

Problem

Traditional single-well decline curve analysis (DCA) often misses crucial factors such as completion design, operating strategy, offset-well behavior, parent-child well interference, facility constraints, and uncertainty.

Use case

In an unconventional shale program, wells with similar lateral length and proppant intensity show different decline profiles because of choke management, spacing, and parent-child effects.

The model learns from offset-well behavior and produces uncertainty-banded forecasts for planning and reserves review.

Approach

Run portfolio-scale hybrid DCA using physics-anchored Arps DCA, ML residual correction, completion and operating features, and calibrated P10/P50/P90 forecasts.

Outcome

Probabilistic forecasts that improve production planning, assessment of net asset value (NAV), capital prioritization, and reserves-supporting confidence.

5.3 PRODUCTION SURVEILLANCE

Problem

Offshore production losses build silently because well behavior, subsea conditions, separator trends, and facility constraints are monitored across fragmented systems.

Use case

In a deepwater offshore asset, a high-rate subsea well shows gradual oil-rate decline while flowing tubing pressure, choke position, and separator backpressure suggest a developing constraint rather than reservoir decline.

The surveillance model flags the deviation early, correlates it with subsea flowline pressure build-up as well as facility backpressure, and then prepares an engineer-ready investigation pack.

Approach

Continuously compare actual production against the expected performance of the well and the facility using these criteria: real-time rates, flowing pressure, choke position, electric submersible pump (ESP) or gas-lift parameters, separator data, subsea pressure-temperature trends, and operating constraints.

Outcome

Earlier detection of offshore production losses, faster root-cause diagnosis, better prioritization of high-value wells, and reduced deferment through timely action.

5.4 FLOW ASSURANCE

Problem

Risks related to hydrate, wax, scale, emulsion, sand, and slugging are often managed reactively or through broad chemical treatment programs.

Approach

Use data related to pressure, temperature, composition, water cut, flow regime, and operating envelope with physics-informed models to predict flow assurance risk and recommend targeted responses.

Use case

In a subsea tieback, falling temperature and changing pressure profile indicate rising hydrate risk during a low-flow operation.

The model flags the risk window, identifies the affected operating envelope, and recommends proactive inhibitor review before production restriction occurs.

Outcome

Earlier risk detection, more disciplined chemical usage, fewer avoidable constraints, and improved production continuity.

5.5 INTERVENTION PRIORITIZATION

Problem

Engineers often receive multiple alerts but lack a consistent way to assess production impact, urgency, feasibility, cost, and expected uplift.

Approach

Combine production gap, likely root cause, operating history, intervention cost, uplift potential, and execution complexity to prioritize the highest-value actions.

Use case

Ten offshore wells show deviation from expected production, but only three combine high uplift potential, clear diagnostic signals, and low operational complexity.

The system ranks those wells first and generates evidence packs for engineer review.

Outcome

Sharper work queues, faster engineering decisions, better capital and operating discipline, and a more secure linkage between surveillance and execution.

5.6 OPERATING COST AND CHEMICAL OPTIMIZATION

Problem

Chemical dosing, lift energy, compression usage, and routine operating actions are often governed by static rules rather than live production risk and economic value.

Approach

Analyze chemical treatment history, production response, flow assurance exposure, lift performance, energy intensity, and operating constraints to optimize controllable production costs.

Use case

In an offshore field, scale-inhibitor dosing remains fixed across wells even though only a subset shows elevated risk based on water chemistry, pressure–temperature conditions, and production trend.

The system identifies wells best suited for targeted dosing review.

Outcome

Better chemical discipline, lower controllable operating cost, reduced lifting-cost pressure, and improved reliability without compromising production integrity.

6. Reference Architecture on Databricks

The production optimization architecture of the Infosys Production Optimization solution is designed as a governed data intelligence platform that connects production data, domain models, AI/ML, agentic workflows, and business-facing applications into one operating stack. It is anchored in the Databricks Lakehouse architecture for scalable data management, governance, ML lifecycle management, and serving. The architecture is extended through Infosys Topaz Fabric, de.ai, do.ai, the Infosys Energy Data Model, Infosys AI Factory, and domain accelerators.

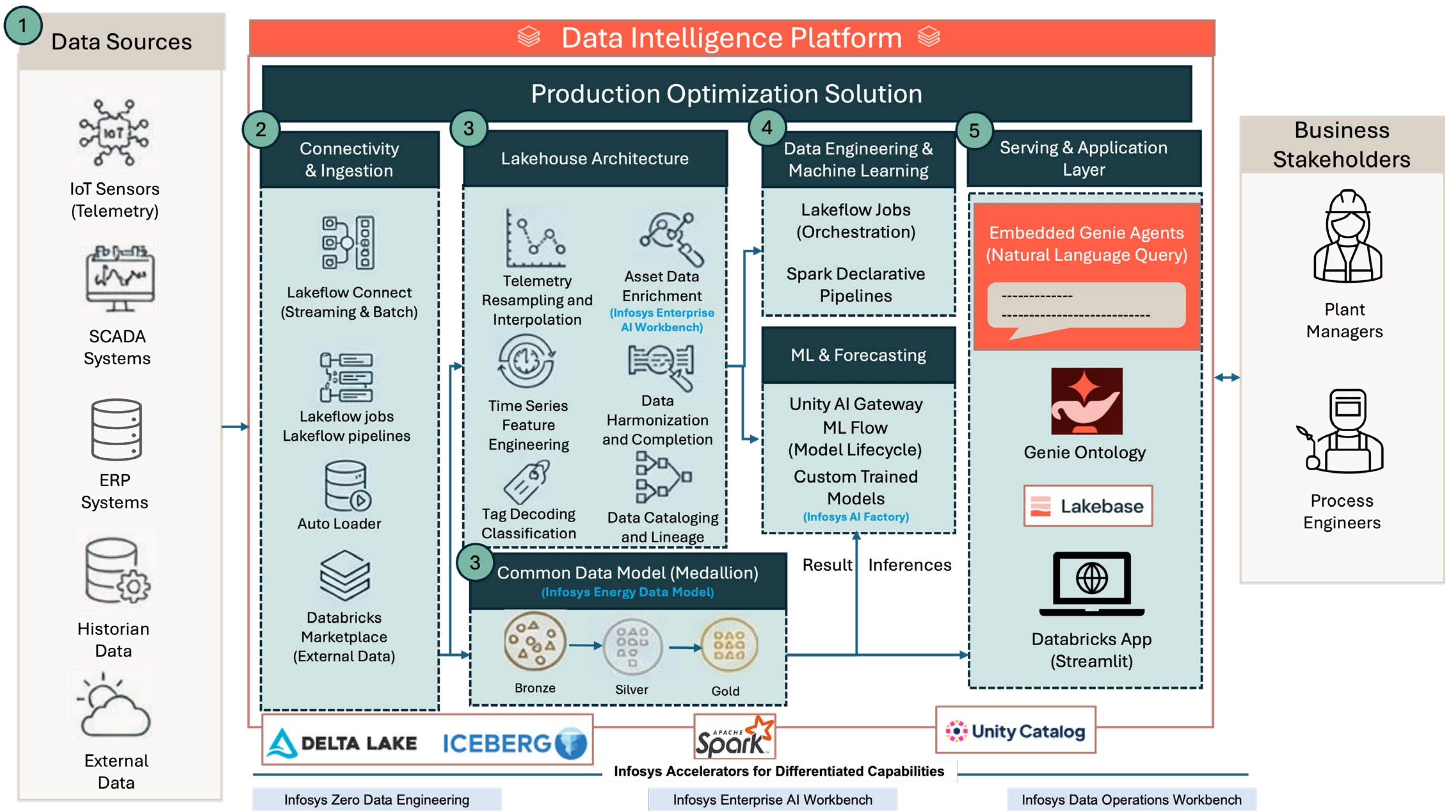


Figure 1. Infosys Production Optimization solution data intelligence platform reference architecture

ARCHITECTURE LAYER SUMMARY

COMPONENT	DESCRIPTION
Layer 1 DATA SOURCES	• Ingests Internet of Things (IoT) telemetry, SCADA, historian data, ERP, and curated external datasets • Keeps operational technology (OT) data protected behind the Purdue Level 2/3 demilitarized zone (DMZ), with controlled unidirectional ingestion into the data platform
Layer 2 CONNECTIVITY AND INGESTION	Uses Lakeflow Connect, Spark Declarative Pipelines, Lakeflow Jobs, Auto Loader, Kafka/Azure Event Hubs/Java Database Connectivity (JDBC), and PI ingestion patterns to securely move operational and enterprise data into the Lakehouse
Layer 3 LAKEHOUSE AND COMMON DATA MODEL	• Structures data through Bronze, Silver, and Gold layers normalized by the Infosys Energy Data Model across well completion, production, and surveillance entities • Unity Catalog governs lineage, access, and auditability
Layer 4 DATA ENGINEERING AND ML	Powers hybrid decline curve analysis (DCA), production surveillance, flow assurance models, anomaly detection, and agent-assisted workflows using Lakeflow Jobs, MLflow, Unity AI Gateway, Infosys AI Factory models, LangGraph, and Lakebase
Layer 5 SERVING AND APPLICATIONS	Delivers role-based production intelligence through Genie Agent, Databricks Apps, artificial intelligence/business intelligence (AI/BI) dashboards, serverless SQL warehouses, and governed experiences for plant managers, process engineers, reserves, and finance teams

Architecture Outcome: The architecture converts fragmented production and enterprise data into governed operational intelligence. It enables faster data onboarding and earlier anomaly detection. This leads to stronger forecast confidence, better engineering productivity, and more disciplined intervention execution across wells, facilities, and assets.

7. Why Databricks for Production Optimization

Databricks provides the governed data, AI, and application foundation required to industrialize production optimization. Delta Lake and the Medallion architecture create trusted production data layers. Unity Catalog, a unified governance and security layer built into Databricks, provides lineage, access control, and auditability. Lakeflow and Spark Declarative Pipelines, which form a framework for creating batch and streaming data pipelines, support scalable ingestion and transformation. MLflow manages model lifecycle and traceability, while Genie, Databricks Apps, Lakebase, and AI/BI dashboards expose governed insights to engineers, plant managers, reserves teams, and finance users. Together, these capabilities elevate production optimization from isolated analytics to an integrated operating workflow.

8. Governance and Human Authority

When considering AI technology adoption, upstream operators must consider human accountability in light of the multiple safety and environmental standards they need to follow. Recognizing this aspect, the solution applies AI in areas where it safely accelerates production workflows: data engineering, anomaly detection, forecasting support, opportunity ranking, evidence-pack generation, and workflow preparation. While AI agents assist engineers by preparing recommendations and routing decisions, they do not autonomously approve changes that impact critical areas such as reserves, HSE, capital expenditure, chemical dosing, or operational setpoints. Human authority remains explicit, auditable, and deeply embedded in the operating model.

9. Implementation Roadmap

PHASE	FOCUS	OUTCOME
Pilot Asset	• Ingest selected SCADA/historian/ERP data • Stand up common data model • Validate surveillance and forecasting	• **Prove data readiness, model accuracy, and engineer workflow fit**
Shadow Mode	• Run AI recommendations alongside current operating process	• **Build trust** • **Tune thresholds** • **Validate intervention logic**
Workflow Integration	• Connect evidence packs, work-order context, and approval paths	• **Move from insight to governed action**
Portfolio Scale	• Expand across assets, wells, facilities, and production teams	• **Realize measurable impact on deferment, intervention cycle time, surveillance coverage, and cost efficiency**

10. Conclusion

With breakthrough developments taking place in the field of AI/ML, AI-driven production optimization is no longer a digital experiment for upstream operators. It is swiftly becoming a core operating capability for protecting barrels, reducing deferment, and improving cost per BOE.

This white paper shows how a Databricks-powered production intelligence stack—Lakebase and Genie—can connect fragmented SCADA, historian, well, facility, ERP, and external data into a governed workflow for forecasting, surveillance, flow assurance, and intervention planning.

The value is not created by AI models in isolation, but by the precise combination of components. The real game-changer is the compression of the production decision cycle—from data capture to anomaly detection, engineering diagnosis, ranked intervention, and, finally, human-approved action.

Infosys strengthens this workflow with upstream domain expertise. Infosys' AI platform, Topaz Fabric, de.ai, do.ai, the Infosys Energy Data Model, AI Factory models, and governed agentic workflows accelerate engineering effort without removing the human role in operational accountability.

For business owners, the impact of the solution is both practical and measurable. They can detect production loss earlier, improve forecast confidence, accelerate intervention cycles, and broaden surveillance coverage. Enhanced discipline in chemical and operating-cost management, along with clearer reserves-supporting visibility, contributes to greater optimization.

The shift is operational for production teams that were earlier engaged in periodic reviews and reactive troubleshooting. They can now move to continuous, AI-assisted optimization across wells, facilities, and assets.

This is how AI can drive the optimization of upstream operations—through reduced deferment, stronger production efficiency, faster intervention, improved lifting-cost performance, and greater reserves confidence. With its built-in intelligence and recognition of the critical role of human authority in upstream operations, the Infosys Production Optimization solution is the right fit for the oil and gas industry.

About the Authors

Dr. Arunkumar Ranganathan

Vice President and Global Domain Consulting Leader, Infosys

Focused on Energy, Utilities, and Energy Transition, Dr. Ranganathan serves as the domain anchor for the Infosys SURE AI Council and holds a PhD in Oil & Gas Engineering.

Ankit Sharma

Principal Data Scientist - Data, Analytics and AI, Infosys

Ankit leads the Databricks partnership and drives solution development on the Databricks platform within the Infosys Data & Analytics practice.

Achyut Bajpai

Digital Solution Specialist - Data, Analytics and AI, Infosys

A member of the Databricks CoE, Achyut provides AI and data analytics thought leadership for the Energy & Utilities sector in North America.

Vishal Vibhandik

Partner Solutions Architect, Databricks

Vishal helps partners grow their practices through reference architectures, technical enablement, and joint customer success on the Data Intelligence Platform.

Herman Lombard

Lead – Ecosystem Solutions Strategy (Manufacturing, Energy, Automotive & Transportation), Databricks

Herman helps customers and partners translate complex IT/OT/ET challenges into scalable, AI-driven industry solutions and build partner ecosystems (GSIs, ISVs, and hyperscalers) that enable high-value outcomes.

Shiv Trisal

Global Industrials GTM Leader, Databricks

Shiv is responsible for the worldwide adoption of Databricks in the manufacturing, transportation and energy sectors and for the development of new industry solutions, ecosystem partnerships and GTM programs that deliver differentiated value to CxO priorities. Shiv has a proven track record of leading diverse teams and delivering game-changing data & AI-led innovation across Manufacturing, Transportation and Energy industries, with roles at EY, Strategy & and Raytheon. He regularly connects with senior executives to unfold data and AI strategies and unlock competitive advantage in their businesses.

Databricks is the Data and AI company. More than 20,000 organizations worldwide — including adidas, AT&T, Bayer, Block, Mastercard, Rivian, Unilever, and 70% of the Fortune 500 — rely on Databricks to build and scale data and AI apps, analytics and agents. Headquartered in San Francisco with 30+ offices around the globe, Databricks offers a unified platform that includes Lakebase, Genie, Agent Bricks, Lakeflow, Lakehouse, and Unity Catalog. To learn more, follow Databricks on LinkedIn, X, YouTube, and Instagram.

Infosys Topaz is an AI-first set of services, solutions and platforms using generative AI technologies. It amplifies the potential of humans, enterprises, and communities to create value. With 12,000+ AI assets, 150+ pre-trained AI models, 10+ AI platforms steered by AI-first specialists and data strategists, and a responsible by design approach, Infosys Topaz helps enterprises accelerate growth, unlock efficiencies at scale and build connected ecosystems. Connect with us at infosystopaz@infosys.com

Infosys® Navigate your next

For more information, contact askus@infosys.com

Infosys.com | NYSE: INFY

Stay Connected