

# Blueprint

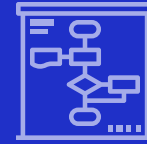
Delivering Intelligence That Matters

|  databricks Hadoop Migration **QuickStart**

# We are Blueprint We Build

We are a data intelligence company committed to delivering the right data to the right person, at the right moment – because we believe intelligence matters.

We connect strategy to delivery. With expertise in cloud, infrastructure, data science, analytics, and application development, **our promise is that together, we can solve anything.**



## **BUSINESS STRATEGY**

---

We help you redefine what's possible with our innovative business and technology strategies for exceptional, scalable, and repeatable results.



## **ACCELERATORS**

---

Start with our stress tested pre-built solutions for speed to value. Our accelerators provide the base for your data loading, cataloging, querying, and Lakehouse needs.



## **TECHNOLOGY SOLUTIONS**

---

Capitalize on cutting-edge technology to plan smarter, act faster, and maintain your competitive edge while decreasing costs, maximizing return on investment, and increasing revenue.



## **SERVICES AT SCALE**

---

Leverage Blueprint's expert teams, processes, products, and support infrastructure to reduce cost and increase agility.



## CASE STUDY

# Major U.S. Oil Drilling Company

## Challenge

- Siloed data in disparate systems and Hadoop environment — not highly performant
- Drill bit sensors collected data every second, but delivered to data scientists only once every 24 hours (not quick enough for BI to inform decisions)
- Timing delays in data availability resulted in slow response times and drilling adjustments (inefficient)



## Solution

- Stream siloed data (rig, sensor, oil sample, financial, HR, marketing) into Azure Databricks Data Lake
- Process, normalize, and organize data into tables,
- Remove need for third-party legacy engineering tools to structure raw data
- Data modeling via Azure SQL & Azure Data Factory
- Share data to Power BI for real-time analytics, dashboards, and reports

**Blueprint**

## Impact

**~100TB**  
data migrated to Data Lake

**80%**  
increase in rig state  
data processing  
(from 24 to every 4 hours)

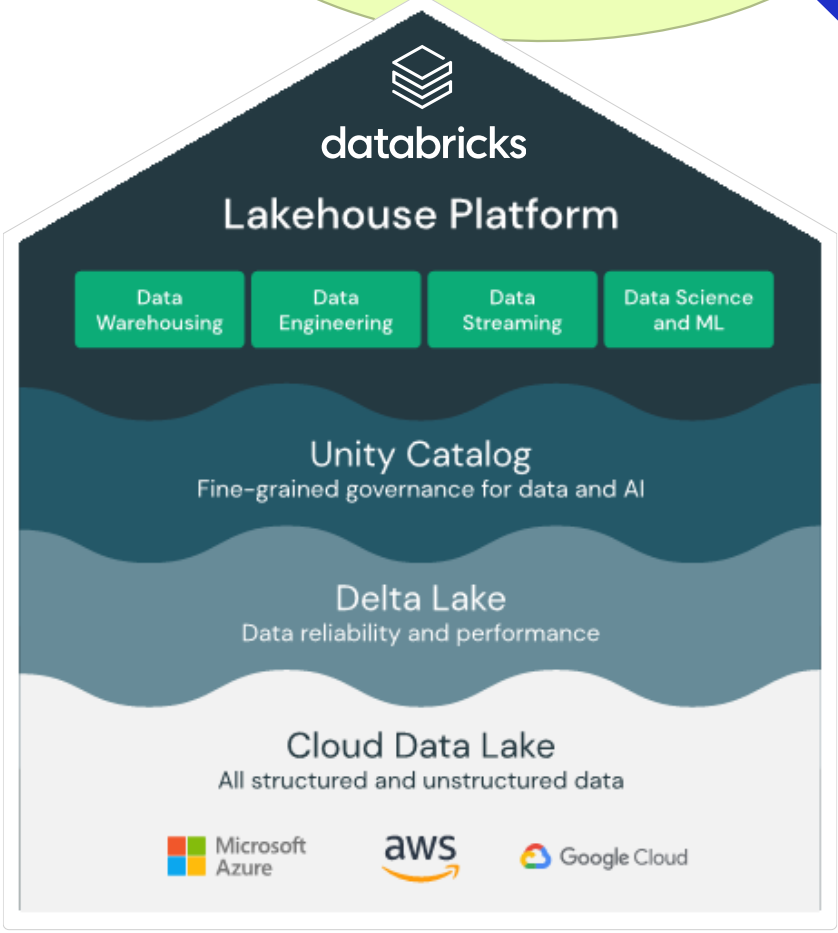
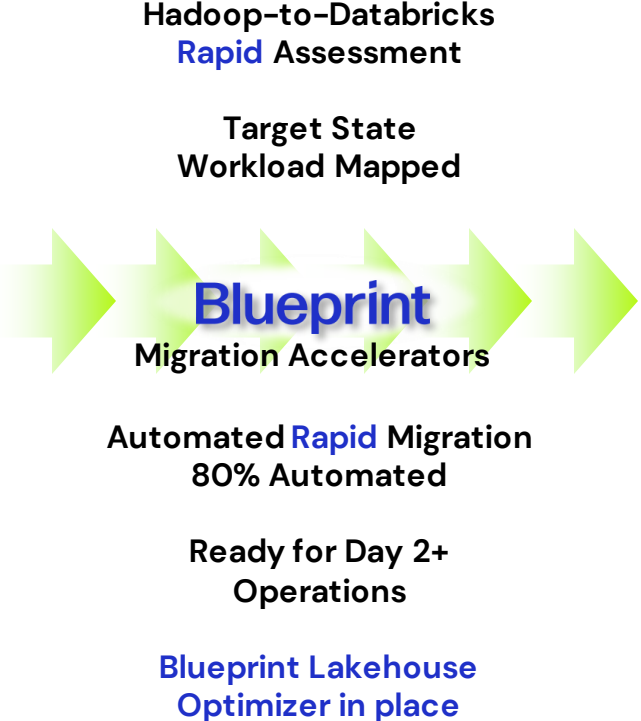
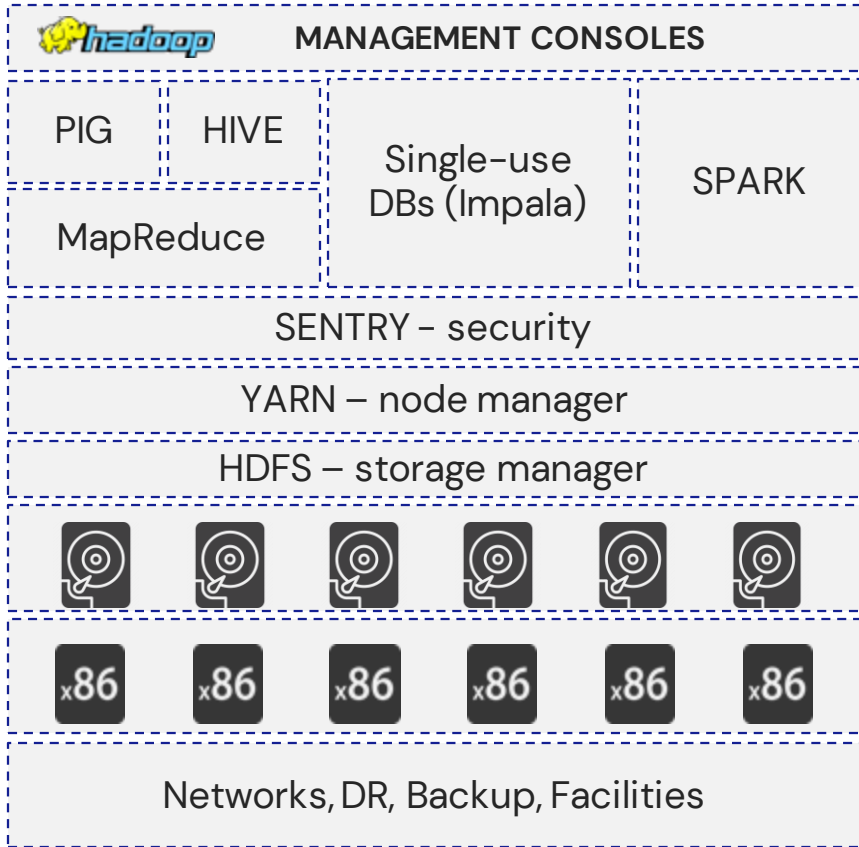
**95%**  
increase in speed of  
OFT data processing  
(45 days of OFT data processed in  
1 hour instead of 24 hours)

**Real-time drilling data  
available in 1 second!**



# Lakehouse Migration

## Simple. Optimized. Fast.



**LEGACY STATE: Complex, Inefficient**

**FUTURE STATE: Databricks + Blueprint = Simple. Optimized. Fast.**

# QuickStart | Overview

## 10 DAYS

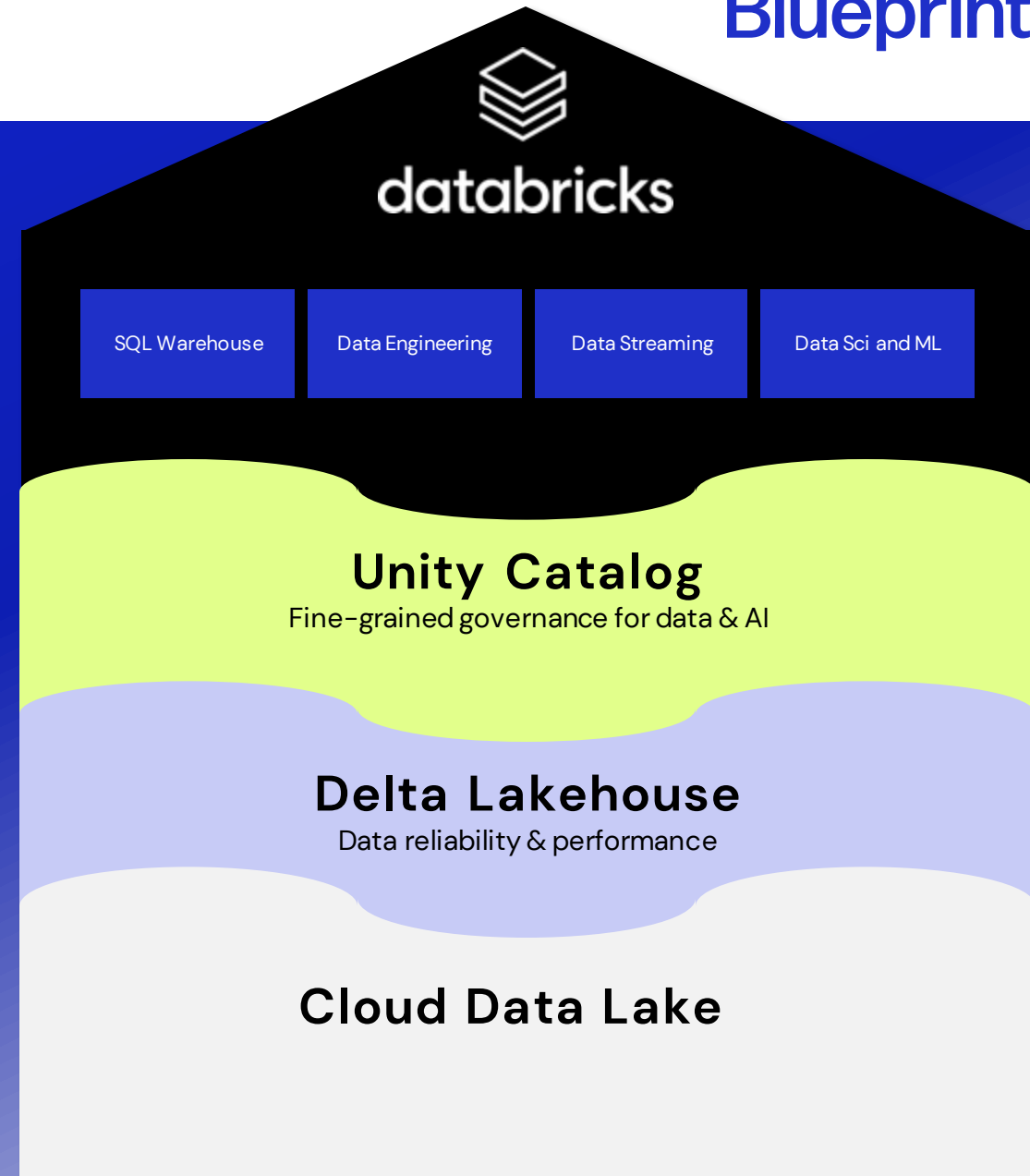
- **INTRODUCE** a data team to analytics on Databricks
- **PREPARE** Databricks & data services, non-prod env.
- **COMPLETE** 1 Hadoop workload migration
- **ANALYZE** with SQL Analytics and BI tools

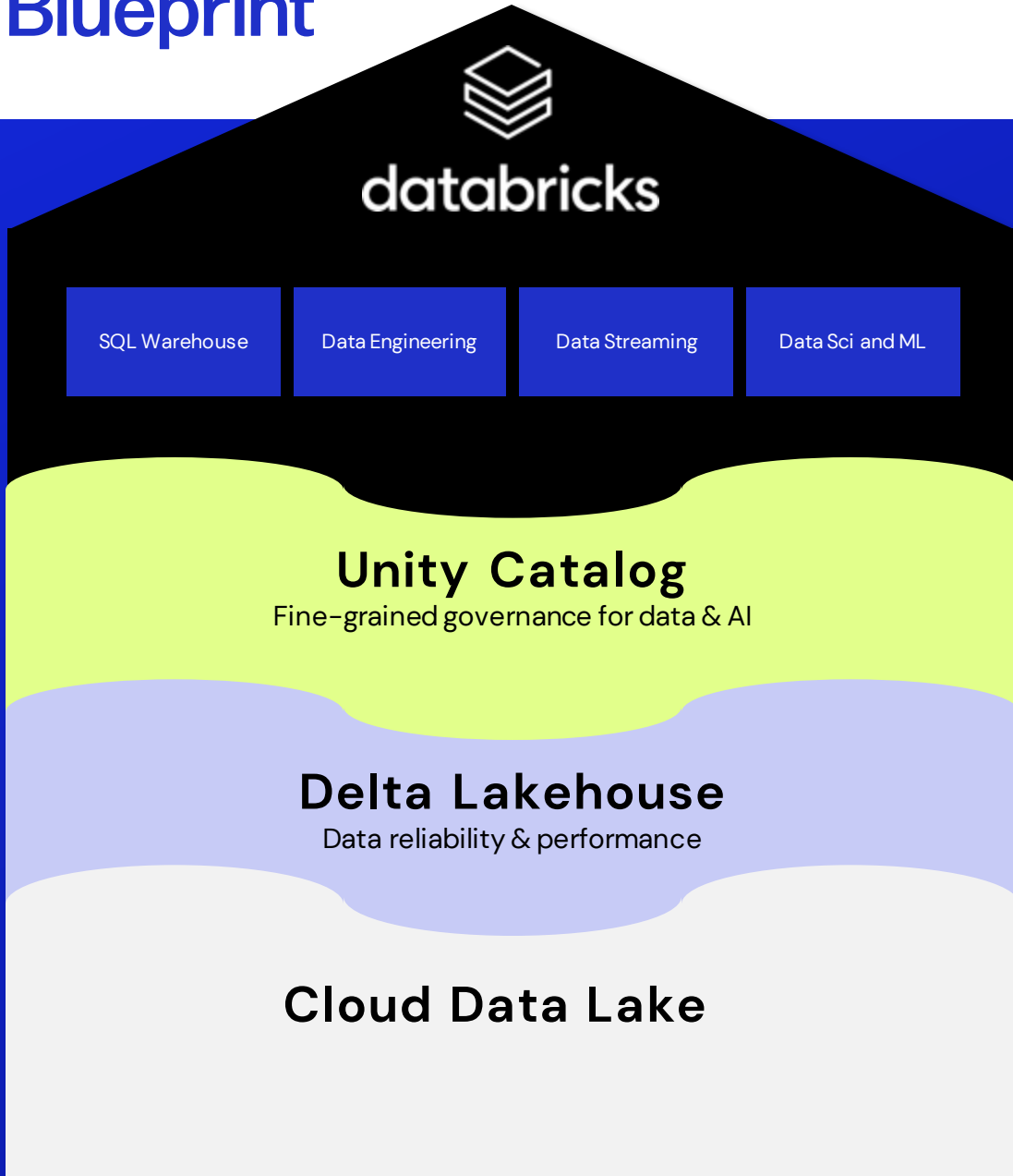
## Day 1 Lakehouse 101

- ▣ Data Acquisition
- ▣ Simple Data Transformations
- ▣ Organizing Data
- ▣ Security
- ▣ BI, Reporting
- ▣ Optimizing Costs & Management

## Day 2-3 Up & Running

- ▣ Implementation → Powered by Infra-as-Code
- ▣ Security Config → Blueprint Security Rapid Config
- ▣ Lakehouse Optimization → Blueprint Lakehouse Monitor





## Day 4-8 Data Pipelines

- ▣ Identify Data Sources
- ▣ Historical & Current Data
- ▣ Data Transformations
- ▣ Data Quality
- ▣ Data Set Creation, Scheduled
- ▣ Tables in the Lakehouse, Ready!

## Day 8-10 BI & Analysis, Optimize, & Roadmap

- ▣ Power BI or Tableau
- ▣ DBSQL for Ad hoc Analysis
- ▣ Dashboards & Reports
- ▣ Utilization Management
- ▣ Roadmap the Future

# QuickStart | Lakehouse Components



Streamed or Scheduled Ingestion



- Data as Sourced
- Retained
- Stream | Batch



- Enriched / Enhanced
- Validation / Quality
- ML Ready Data



- DW & Modeled
- Secured
- BI / Analytics Ready



Data Validation & Observability



Continuous Processing



Automatic Recovery



Security & Compliance



Automated Deployments



Reverse ETL & Sharing



BI & Analytics



AI/ML



Event & Alert

ID Access Mgmt



Key Management



Provisioning



Operational Readiness



Lakehouse Optimizer

Blueprint



# QuickStart | Deployment Journey

1

## AWS or Azure Readiness

- ✓ Account with create-rights
- ✓ Terraform installed

## QuickStart Data Sources

- ✓ Cloud accessible srcs. identified
- ✓ Read-only account

2

## Scripted Build

- ✓ Resource / Admin Groups
- ✓ Storage, Databricks Workspace, Clusters
- ✓ Unity Catalog / Metastore / Permissions
- ✓ Blueprint Lakehouse Optimizer

## Data Ingestion Pipeline

- ✓ 1 Hadoop workload migrated

## Workflows Active

- ✓ Enable Job/Workflows, Scheduled

3

## Databricks is LIVE!

- ✓ SQL Analysis Workshop
- ✓ Power BI / Tableau
- ✓ Engineering & Workflow Demos

## Lakehouse Optimized

- ✓ Monitor jobs
- ✓ Understand costs
- ✓ Identify orphaned workloads

# QuickStart | Deliverables

Build a net-new use case and validate cost vs performance and usability of Databricks platform.

- TCO and performance projection report
- Established Databricks Lakehouse environment
- Data ingestion pipeline
- Platform utilization monitoring app
- Working end-to-end use case/business process deployed to non-production environment
- Results report and demonstration

**Cost & Timeline: \$25K, 2 weeks**  
(tool subscriptions complimentary)

# Get Started Now

[Email us](#) to schedule a call

[Visit our website](#) to learn more  
about using Databricks

Talk to your Databricks sales rep