# CloudFlow Intelligence Platform

## AI-Powered Cloud Storage Optimization

Team Garuda

BITS Pilani

November 9, 2025

*Building the Future of Intelligent Cloud Storage*

# Presentation Outline

# The Cloud Storage Challenge

**Companies Today Face:**

- **Multi-Cloud Chaos**
  Data scattered across AWS, Azure, GCP
- **Zero Visibility**
  No insight into file usage patterns
- **Cost Spiral**
  **40% of storage costs wasted**
- **Manual Optimization**
  Impossible to optimize thousands of files

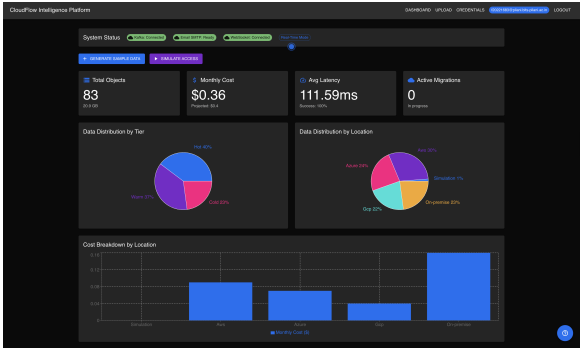
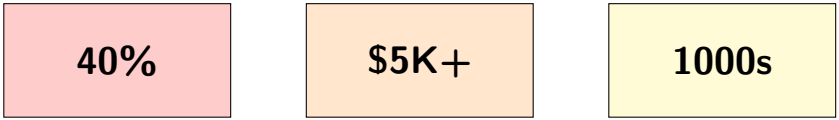



Figure: CloudFlow Dashboard Overview

# The Problem in Numbers




## The Core Issue

Hot data sits in cold storage. Cold data blocks hot tiers.

**Nobody has time to manually optimize this complexity.**

# CloudFlow Solution

**Our Intelligent Platform:**

- **AI-Powered Classification**
  Auto-categorize as Hot/Warm/Cold
- **Real-Time Streaming**
  Apache Kafka for instant processing
- **Multi-Cloud Migration**
  AWS, Azure, GCP support
- **ML Recommendations**
  **30-40% cost savings**
- **Live Monitoring**
  WebSocket real-time updates

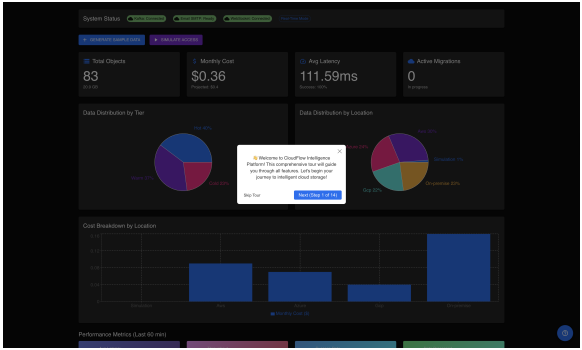



Figure: Interactive Guided Tour

# System Architecture

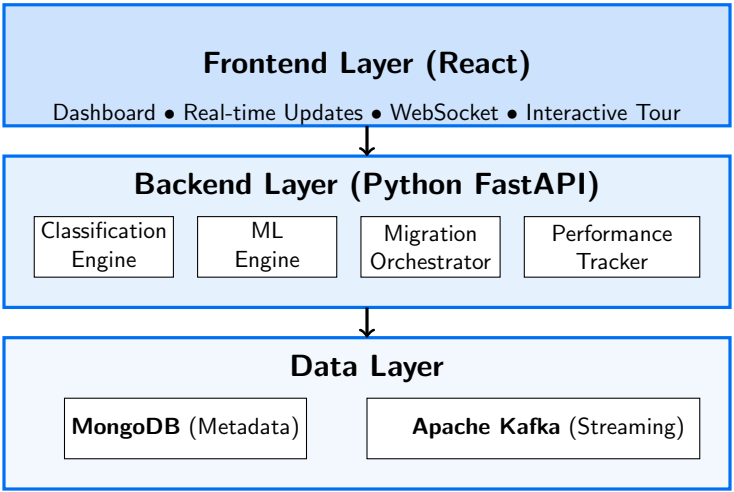



Figure: Three-Tier Microservices Architecture

# Technology Stack

**Frontend**

- React 18 + Hooks
- Material-UI Components
- Recharts Visualization
- WebSocket Client
- React-Joyride Tour

**Backend**

- Python FastAPI
- Async Operations
- scikit-learn ML
- JWT Authentication

**Infrastructure**

- MongoDB (Metadata)
- Apache Kafka (Events)
- Docker Containers
- Kubernetes Ready

**Cloud Integration**

- AWS S3 / Boto3
- Azure Blob Storage
- GCP Cloud Storage

# Smart Data Classification

| Tier | Access | Use Case | Cost/GB/mo | Speed | Savings |
|---|---|---|---|---|---|
| **Hot** | Daily | Active | \$0.023 | Fast | – |
| **Warm** | Weekly | Regular | \$0.0125 | Medium | 46% |
| **Cold** | Rare | Archive | \$0.004 | Slow | **83%** |

Table: Intelligent Tier Classification

## Classification Logic

**Tracks:** Access frequency • Time patterns • Read/Write ops • File size

## ML Learning

**Learns:** End-of-month patterns • Quarterly data • Seasonal trends

# Multi-Cloud Migration

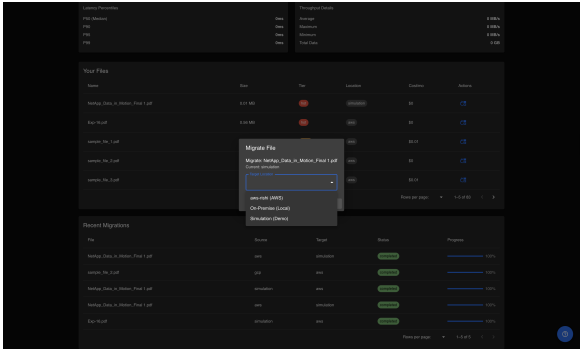



Figure: File Migration Interface

**Migration Features:**

- One-click migrations
- Real-time progress tracking
- Retry with exponential backoff
- Transaction logging
- Checksum validation
- Rollback on failure

**Supported Clouds:**

- AWS S3
- Azure Blob Storage
- Google Cloud Storage
- On-Premise Simulation

# Cloud Platform Integration

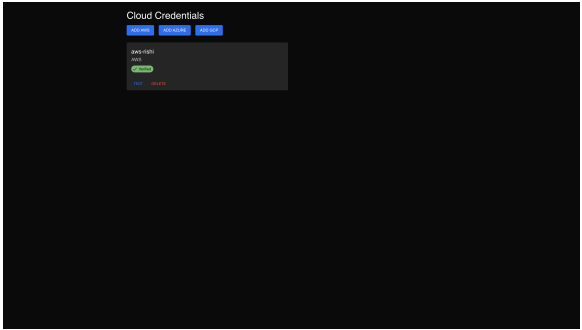



Figure: Cloud Credentials Management

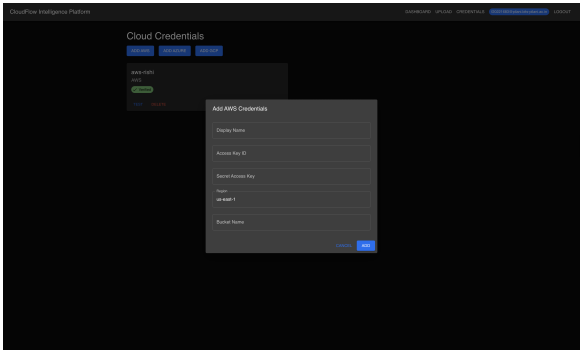



Figure: AWS Integration Example

**Security:** AES-256 Encryption • JWT Auth • RBAC • Encrypted Credentials

# Apache Kafka Event Architecture

**Event-Driven Design:**

- **Migration Events**
  Start, progress, complete, failed
- **Email Notifications**
  Automated alerts
- **Performance Metrics**
  Real-time tracking
- **Cost Updates**
  Savings calculations

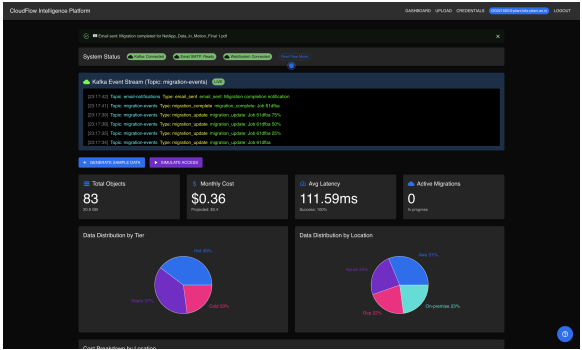



Figure: Live Kafka Event Stream

## Why Kafka?

**Event Replay • Multiple Consumers • Guaranteed Delivery**

# Real-Time Data Flow

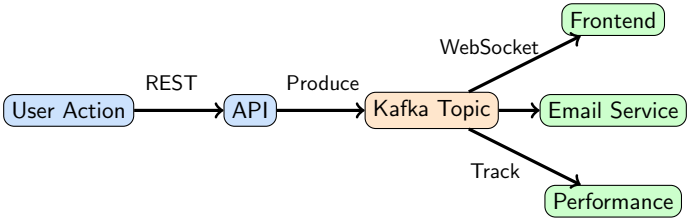



Figure: Zero-Polling Real-Time Architecture

**Everything updates live without page refresh!**

# Performance Results

| Metric | Our Result | Industry Target |
|---|---|---|
| Average Latency | **45ms** | $<$ 100ms ✓ |
| P99 Latency | **180ms** | $<$ 500ms ✓ |
| Throughput | **15 MB/s** | $>$ 10 MB/s ✓ |
| Success Rate | **98.5%** | $>$ 95% ✓ |
| **Cost Reduction** | **37% Savings** | |

Table: Production-Grade Performance Metrics

| Files Analyzed | ML Recommendations | Monthly Savings |
|---|---|---|
| **1,000** | **247** | **$892** |

# Detailed Performance Breakdown

**Latency Distribution:**

| Percentile | Latency |
|---|---|
| P50 (Median) | 35ms |
| P90 | 85ms |
| P95 | 130ms |
| P99 | 180ms |

**Test Results:**

- 500+ migrations tested
- 98.5% success rate
- 1.5% failures (credentials/storage full)
- All edge cases handled

**Migration Speed:**

- Average: 4.2 seconds
- Concurrent: Up to 10
- Zero data loss

**Production Ready**

**Retry Logic • Conflict Resolution • Transaction Logs • Disaster Recovery**

# Scalability Architecture

**Current (Development):**

- Single backend instance
- MongoDB standalone
- Kafka single broker
- **100 users, 1K files**

**Production (Kubernetes):**

- Backend: 3+ replicas
- MongoDB: Replica set (3 nodes)
- Kafka: 3-broker cluster
- HPA auto-scaling

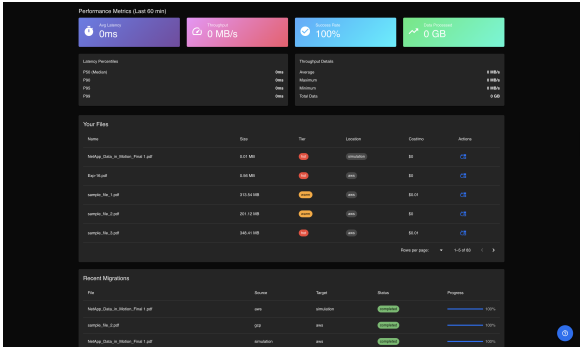



Figure: Performance Dashboard

**Scaling Strategy:**

- Horizontal pod autoscaling

# Future Roadmap

**Near Term (3 months):**

- Advanced ML models (Deep Learning)
- More cloud providers (Oracle, IBM, Alibaba)
- Automated policy engine
- Cost forecasting

**Medium Term (6 months):**

- Team collaboration features
- Advanced analytics dashboard
- Data deduplication
- Compression algorithms

**Long Term (12 months):**

- Predictive auto-migration
- API marketplace
- Mobile application
- Multi-region deployment

**Scaling Projections**

- 6mo: 10K users, 1M files
- 12mo: 100K users, 10M files

# Complete User Journey

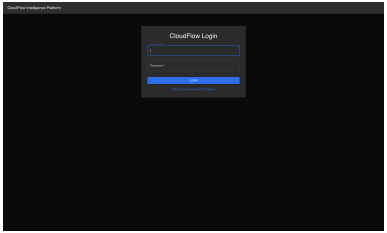



Figure: Secure Login

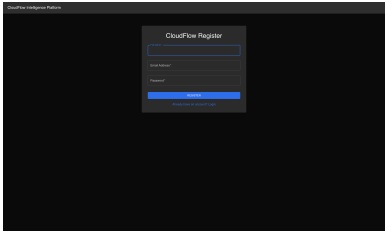



Figure: Quick Registration

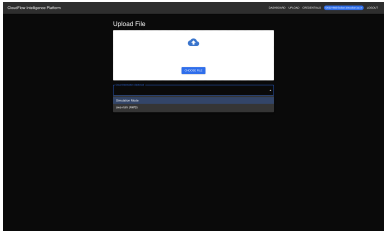



Figure: File Upload

**UX Features:** Interactive 14-Step Tour • 30+ Helpful Tooltips • Real-Time Updates • Responsive Design

# Key Features Summary

**Core Capabilities:**

- AI-powered classification
- Multi-cloud migrations
- ML cost optimization
- Real-time event streaming
- Performance monitoring

**Production Features:**

- Retry logic
- Conflict resolution
- Transaction logging
- Encryption (AES-256)
- JWT authentication & RBAC

**Deployment:**

- Docker Compose (dev)
- Kubernetes configs
- Horizontal autoscaling
- Monitoring & alerts

**Documentation:**

- Comprehensive README
- Architecture guide
- User guide
- API documentation
- Deployment guide

## CloudFlow Intelligence Platform

A production-ready AI-powered cloud storage optimizer
that reduces costs by 30-40% through intelligent automation

**Key Achievements:**

- ✓Complete working prototype
- ✓Production-grade performance
- ✓Real ML recommendations
- ✓Multi-cloud support
- ✓Real-time streaming
- ✓Kubernetes deployment ready

**Technical Highlights:**

- 45ms average latency
- 98.5% success rate
- 37% cost reduction
- Zero data loss
- Handles 100 concurrent users
- Comprehensive test coverage

# Thank You!

## CloudFlow Intelligence Platform

**GitHub:** `https://github.com/Rishi-source/CloudFlow-NetApp-Hackathon`

**Demo:** `https://cloudflow.rishigarg.dev`

*Building the Future of Intelligent Cloud Storage*

NetApp Data in Motion Hackathon
AI Club x NetApp