



KUBERNETES FAILURE PREDICTION SYSTEM

Presented by Jugad Geeks



INTRODUCTION

- Machine learning-based system for predicting failures in Kubernetes clusters
- Helps maintain cluster stability and reliability
- Reduces downtime and improves system performance



WHAT PROBLEMS DOES IT SOLVE?

- Proactively identifies potential failures before they occur
- Enables preventative maintenance and resource allocation
- Improves operational efficiency and reduces manual monitoring



TECHNICAL IMPLEMENTATION

Model Type

Training/Testing

Accuracy

Model Storage



TYPES OF FAILURE PREDICTIONS

Node Failures

- Predicts complete node failures
- Uses metrics: CPU usage, memory usage, disk pressure, pod restart count

Resource Exhaustion

- Detects when resources are approaching limits
- Uses metrics: CPU usage, memory usage, disk usage, network I/O

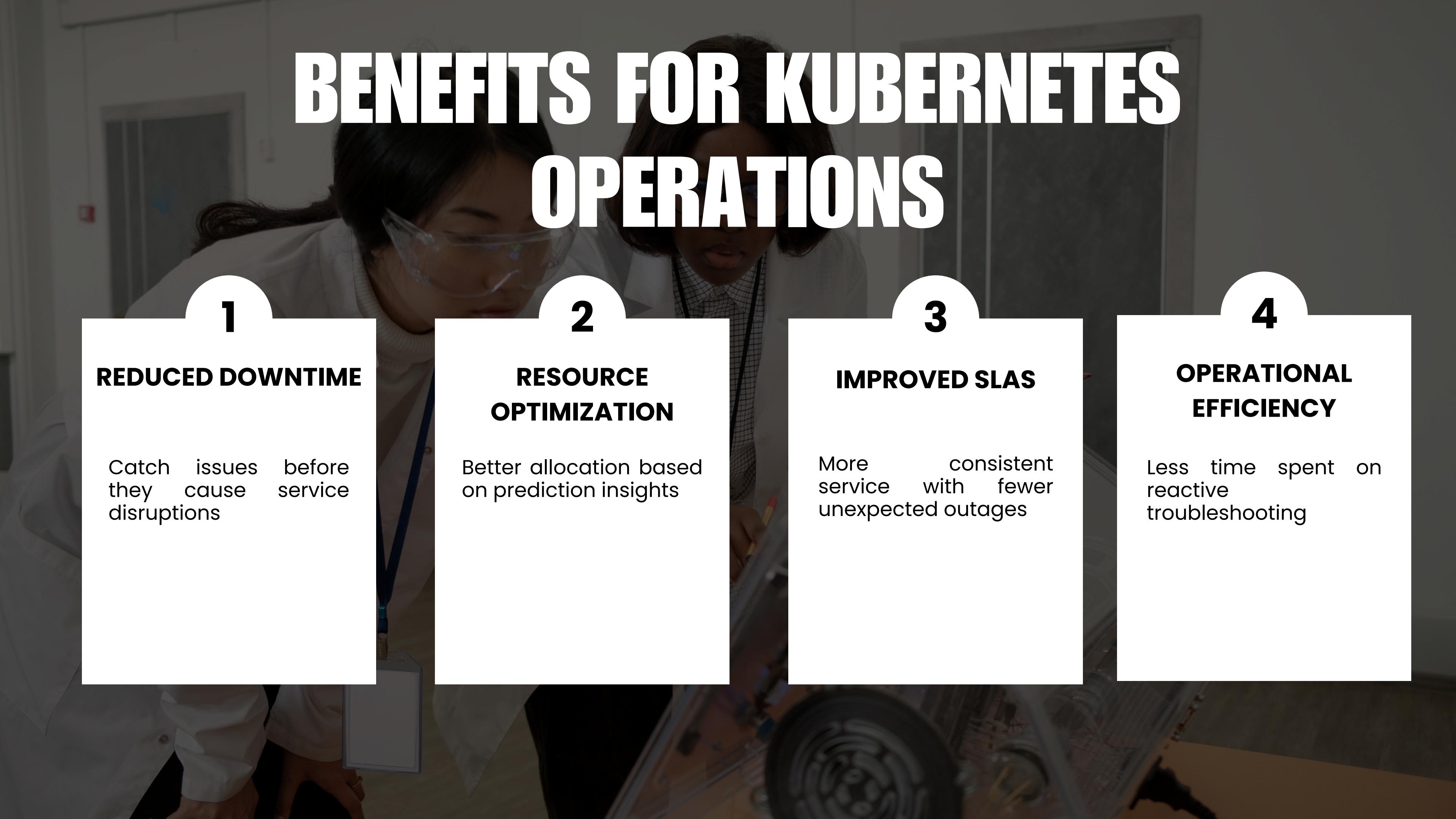
Network Issues

- Forecasts connectivity problems and latency spikes
- Uses metrics: network latency, packet loss, DNS failures, connection resets

Scheduling Issues

- Predicts pod scheduling failures
- Uses metrics: pending pods, node capacity, resource requests, autoscaler latency

BENEFITS FOR KUBERNETES OPERATIONS

A black and white photograph of two female scientists in a laboratory. They are wearing white lab coats, safety glasses, and hairnets. One scientist is holding a tablet computer, and they both appear to be focused on the screen. In the background, there are shelves with various laboratory equipment and supplies.

1

REDUCED DOWNTIME

Catch issues before they cause service disruptions

2

RESOURCE OPTIMIZATION

Better allocation based on prediction insights

3

IMPROVED SLAS

More consistent service with fewer unexpected outages

4

OPERATIONAL EFFICIENCY

Less time spent on reactive troubleshooting

FUTURE ENHANCEMENTS

- Integration with alerting systems
- Automated remediation based on predictions
- Additional prediction categories
- Deep learning models for improved accuracy



THANK YOU

