

DATA

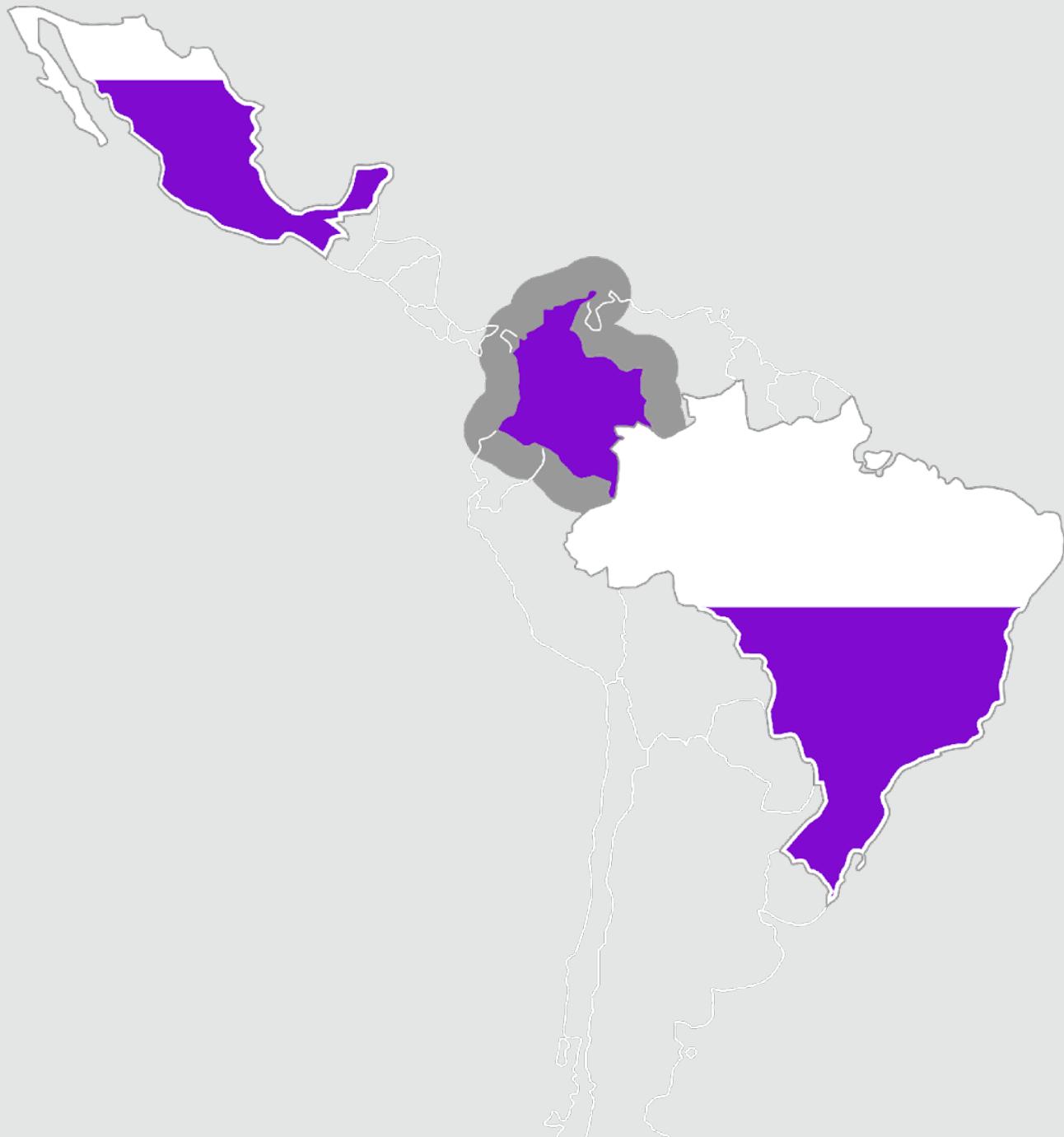
THE LAND DEVOPS FORGOT

Michael Nygard
VP, Data Engineering, Nubank

Introducing Nubank

A woman with curly hair is sitting on a couch, looking down at her laptop. She is wearing a light-colored hoodie and dark pants. A large green plant is next to her, and a wine bottle is on a shelf behind her. A small dog is lying on the couch next to her.

Fight against
complexity
to return the
power to people.



**Almost half of
Brazilians**

78% of Mexicans

Half of Colombians

100 million

92+ million customers in Brazil

7+ million customers in Mexico

1 million customers in Colombia

A lot of customers...

Making a *lot* of transactions...

Producing a *whole lot* of data

OPERATIONAL

Transactions

3,500 microservices

20 long-lived K8s clusters

800+ Datomic databases

10,000 DynamoDB tables

ANALYTIC

Batch

50+ PB ingested nightly

180,000+ jobs

Dozens of short-lived Spark clusters

15,000+ EC2 instances

OPERATIONAL

Transactions

3,500 microservices

20 long-lived K8s clusters

800+ Datomic databases

10,000 DynamoDB tables

Work unit: microseconds

Deploy time: minutes

Deploy lag: minutes

ANALYTIC

Batch

50+ PB ingested nightly

180,000+ jobs

Dozens of short-lived Spark clusters

15,000+ EC2 instances

Work unit: hours

Deploy time: hours

Deploy lag: days

Long lead time

Centralized operations

Unpredictable results

Low accountability

Accumulating technical debt

Cost trajectory unsustainable

Frustration in all parties

EVEN MORE CHALLENGES



MONOLITHIC
CODE: 7.1M LOC
SCALA



100 PULL
REQUESTS / DAY



BUILD & TEST
TAKES 6+ HOURS

NEW PARADIGM REQUIRED

NEW PARADIGM REQUIRED

Independent, autonomous teams

Moving at their own pace

Supported by a self-service platform

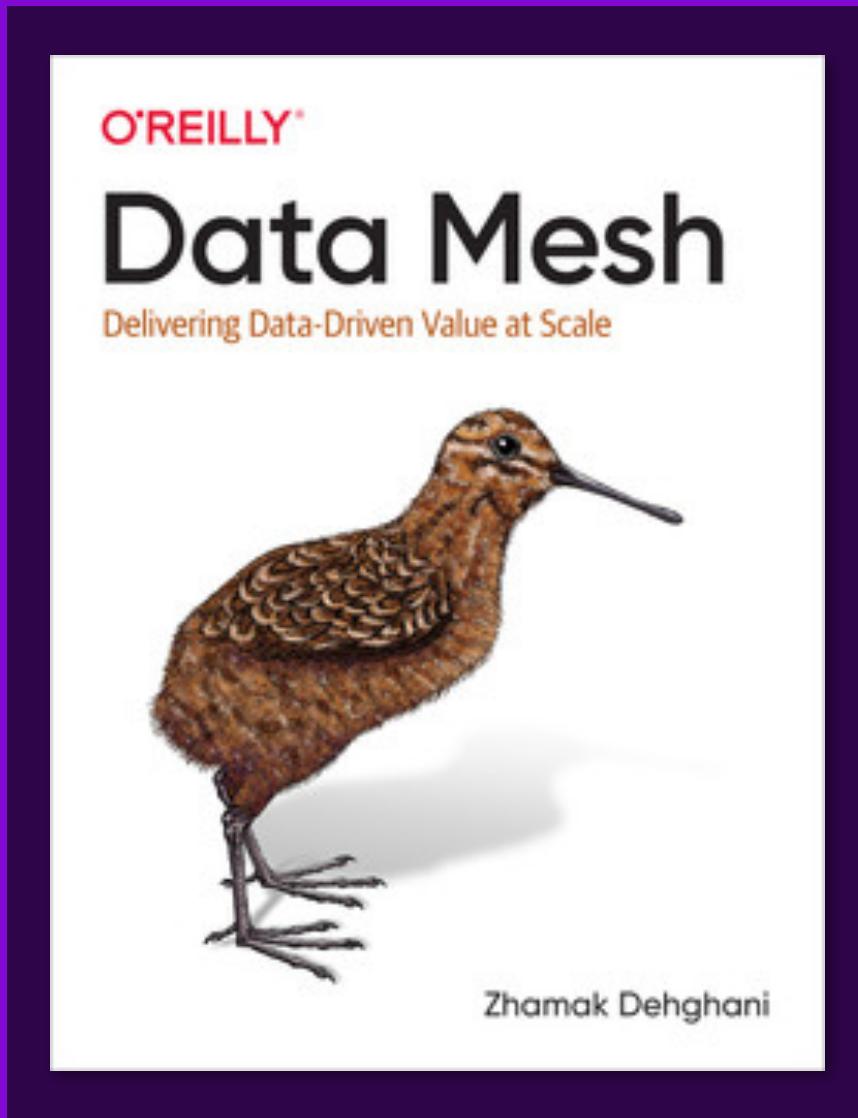
With high discoverability and visibility



The “Microservices of Data”



Data Mesh



<https://martinfowler.com/articles/data-mesh-principles.html>

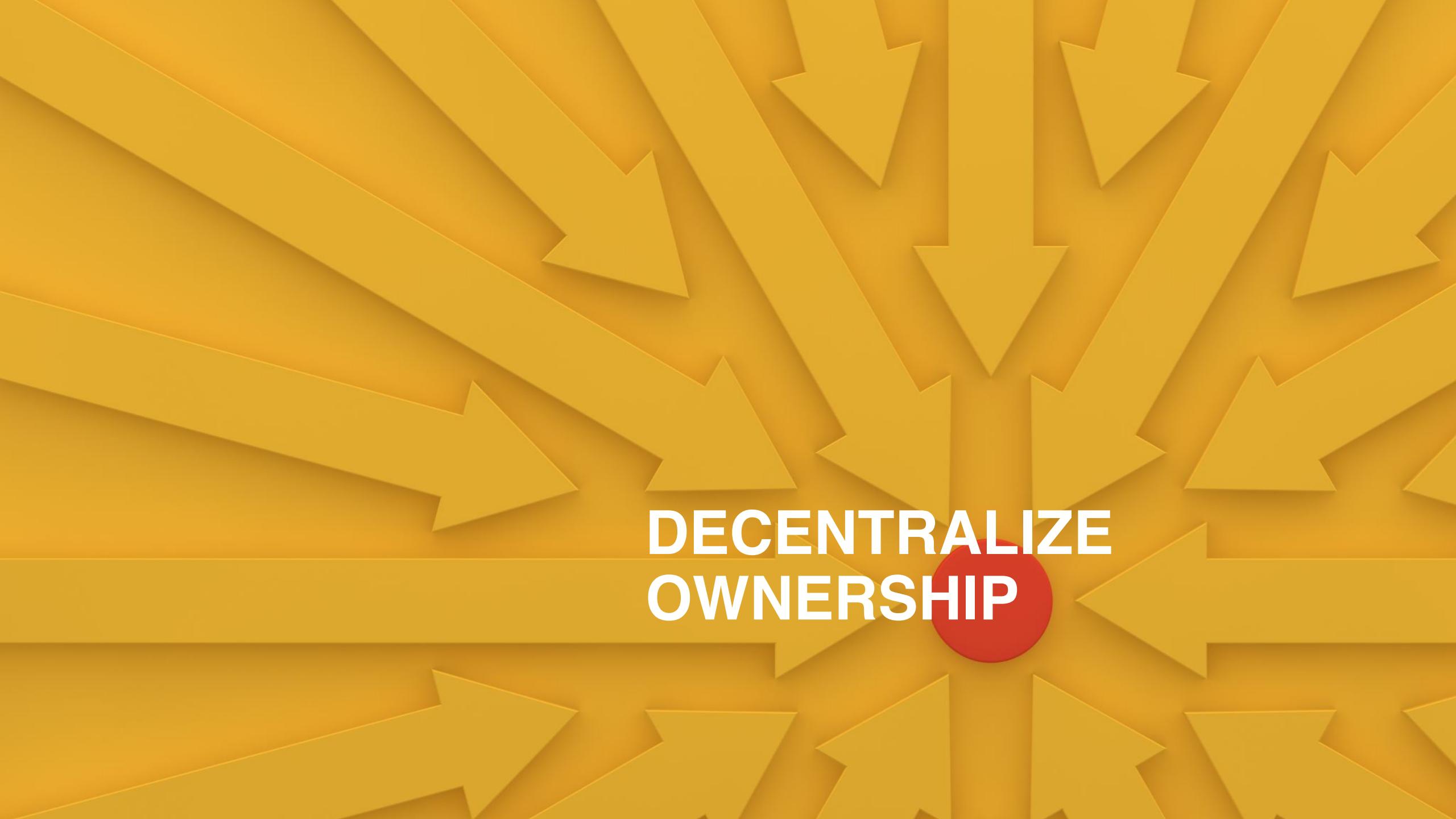
4 KEY PRINCIPLES

Domain ownership

Data as a product

Self-service platform

Federated governance



**DECENTRALIZE
OWNERSHIP**

A close-up, slightly blurred view of a chain-link fence against a bright sky. The fence is made of dark metal wires and is set against a backdrop of a clear, light blue sky. The perspective is from behind the fence, looking out towards a field.

GOOD FENCES
MAKE GOOD
NEIGHBORS

GOOD FENCES ALLOW FLEXIBILITY

KEEP PRIVATE

Transformation

Intermediates

Storage

MAKE PUBLIC

Schema & Classification

Lineage

Output ports

SLOs & performance

Quality rules

SOUNDS GOOD...

SIMPLE IDEA ... NOT SO EASY TO IMPLEMENT

Technical hurdles

Organizational change

Why should data owners care?

SIMPLE IDEA ... NOT SO EASY TO IMPLEMENT

Technical hurdles

Organizational change

Why should data owners care?

Who even are the “data owners”?

DATA AS A PRODUCT

Quantum of sharing

Self describing

Discoverable

Understandable

Viable & Valuable

SELF-SERVICE PLATFORM

Manage cognitive load

Support the work, make the simple things easy

Provide flexibility with automated guardrails

Give users a lot of power, but help them understand the consequences

SUPPORTING THE DATA OWNER

SUPPORTING THE DATA OWNER

Observability

Automation

Optimization

Guardrails

SUPPORTING THE DATA CONSUMER

Discover

Trust

Understanding

Use

HOW TO GET THERE?

HOW TO GET THERE?



INCREMENTALLY

ARCHITECTURE INCREMENTS

Execution & storage options

Publish to catalogs

Automated classifiers

Quality assertions

Public data products as interfaces

PR bots

CHANGE MANAGEMENT BASICS

Explain the "why"

Communicate relentlessly

Create incentives & be careful with feedback loops

Create tracking & visibility of progress

Top-down & bottom-up

BUILD INTERNAL SUPPORT TOO

Bring everyone along

Don't create a “legacy world” and a “promised land”

Lessen burdens on the implementers as well as the users

Provide slack for change

SUMMING UP

“Data Mesh” is a promising paradigm

Decentralize production and consumption

Centralize platform and governance rules

Federate governance execution

SEEKING HELP

There is no “Swagger for Data Products”... there are a couple of competing candidates but not a large amount of traction. Anyone with experience using a data contract specification?

Vendor lock-in is high. Interop is a fledgling prospect, mostly starting with table formats. I’d love to hear from people using the *data product* as a vehicle for interoperability.

DATA

THE LAND DEVOPS FORGOT

Michael Nygard
VP of Engineering