DATA+AI SUMMIT BY databricks

EFFICIENTLY SCALING YOUR DATA TEAM

Mike Xu
10 June 2024

©2024 Databricks Inc. — All rights reserved

Data Teams

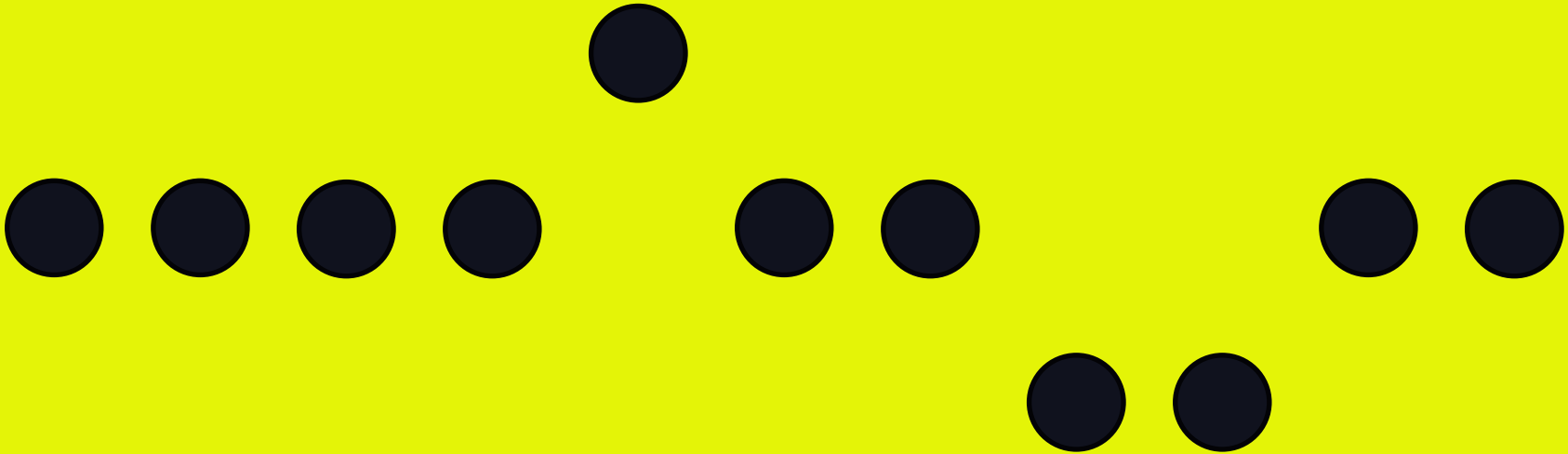
Scaling, Not Growing

A Modern Data Community

Data Mesh, The Pragmatist Way

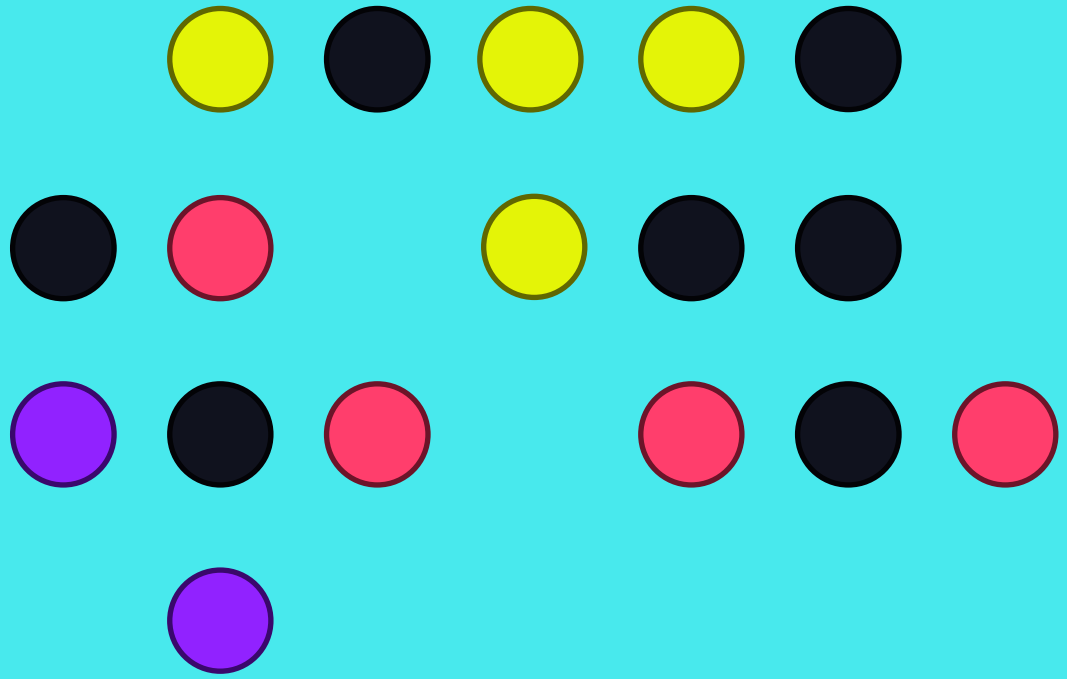
©2024 Databricks Inc. — All rights reserved

GROWING A TEAM



©2024 Databricks Inc. — All rights reserved

SCALING A TEAM



©2024 Databricks Inc. — All rights reserved

THE DATA SCIENCE UNICORN

Evolving data specializations

Data Scientists

Deploying Models as Services

Developing Apps & Software

Data Specializations

ML Research

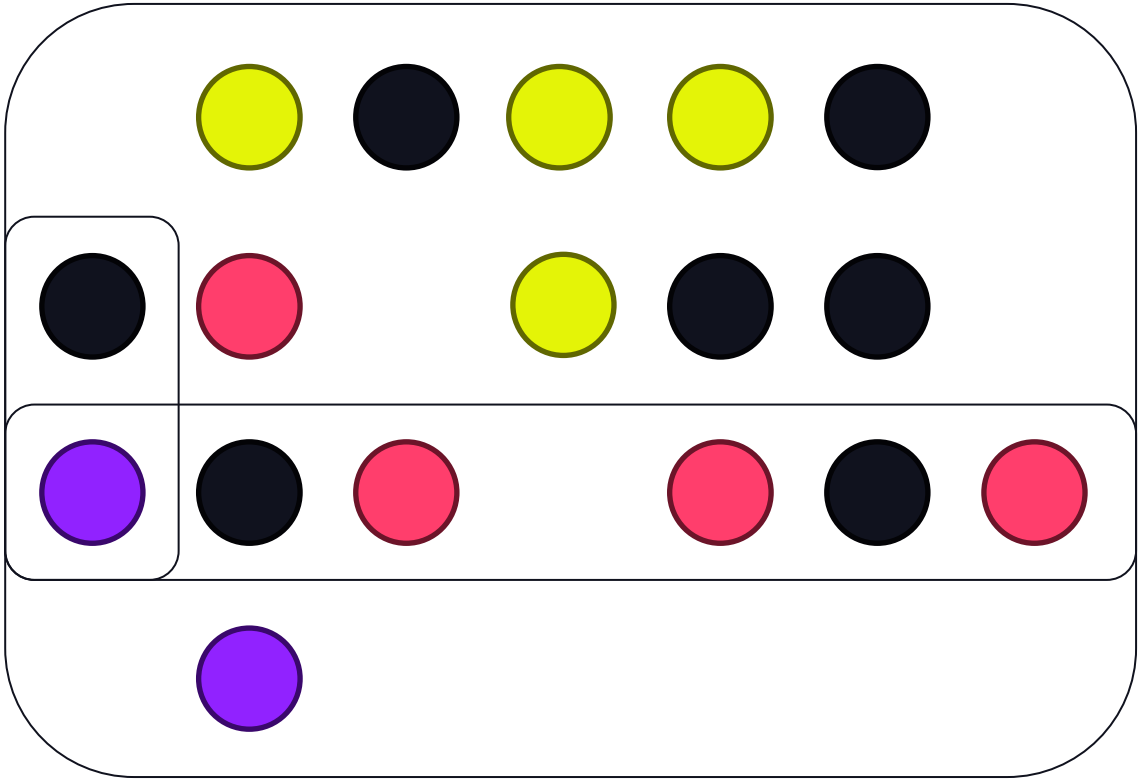
Wrangling & Transforming Data

Implementing Business Logic

©2024 Databricks Inc. — All rights reserved

ORGANIZATION → COMMUNICATION

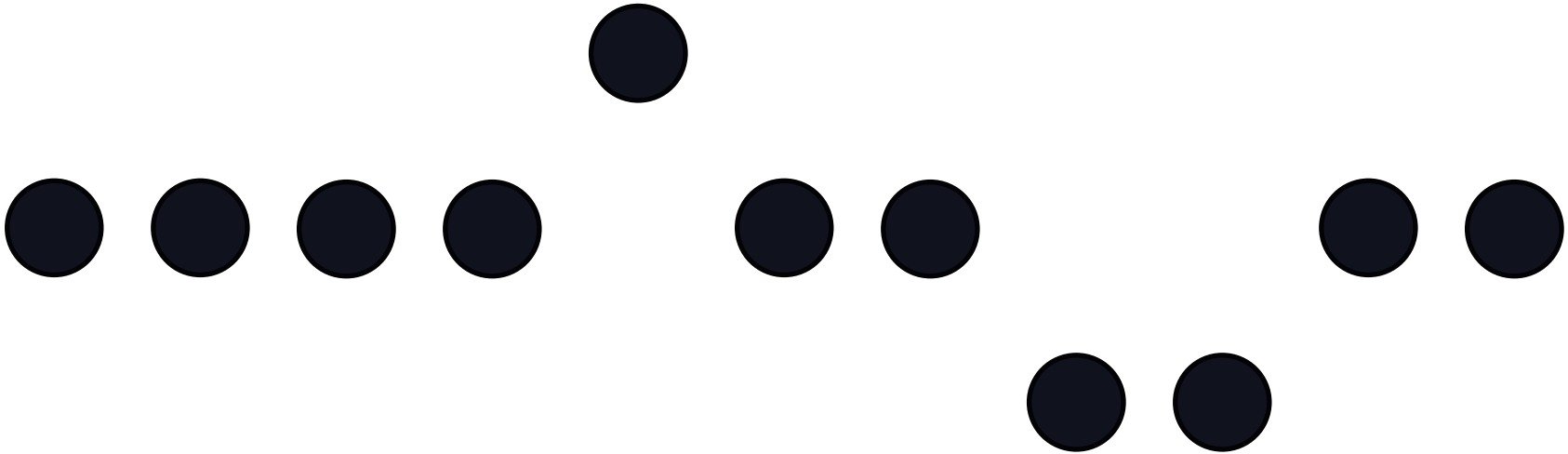
Conway's Law



©2024 Databricks Inc. — All rights reserved

COMMUNICATION WITHOUT DIVERSITY

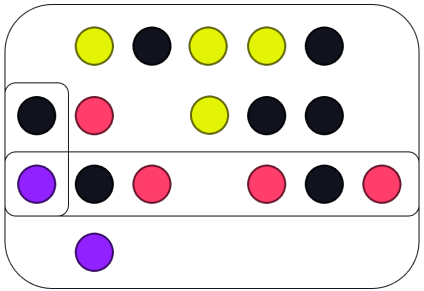
What communication will we get here?



©2024 Databricks Inc. — All rights reserved

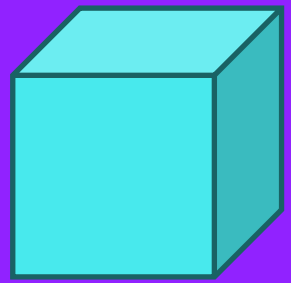
COMMUNICATING WITH NON-DATA TEAMS

They may not even care about data



©2024 Databricks Inc. — All rights reserved

A MODERN DATA COMMUNITY



©2024 Databricks Inc. — All rights reserved

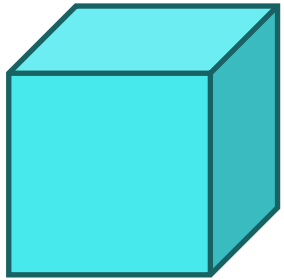
A MODERN DATA COMMUNITY

Data Producers

Generate and/or share data

Data Technology

Enables the sharing and consumption of data

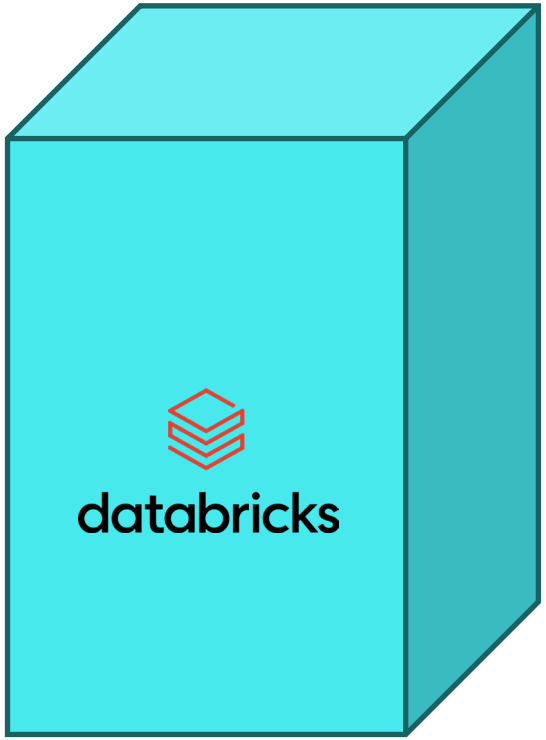


Data Consumers

Leverage data to produce insights driven by business demands

©2024 Databricks Inc. — All rights reserved

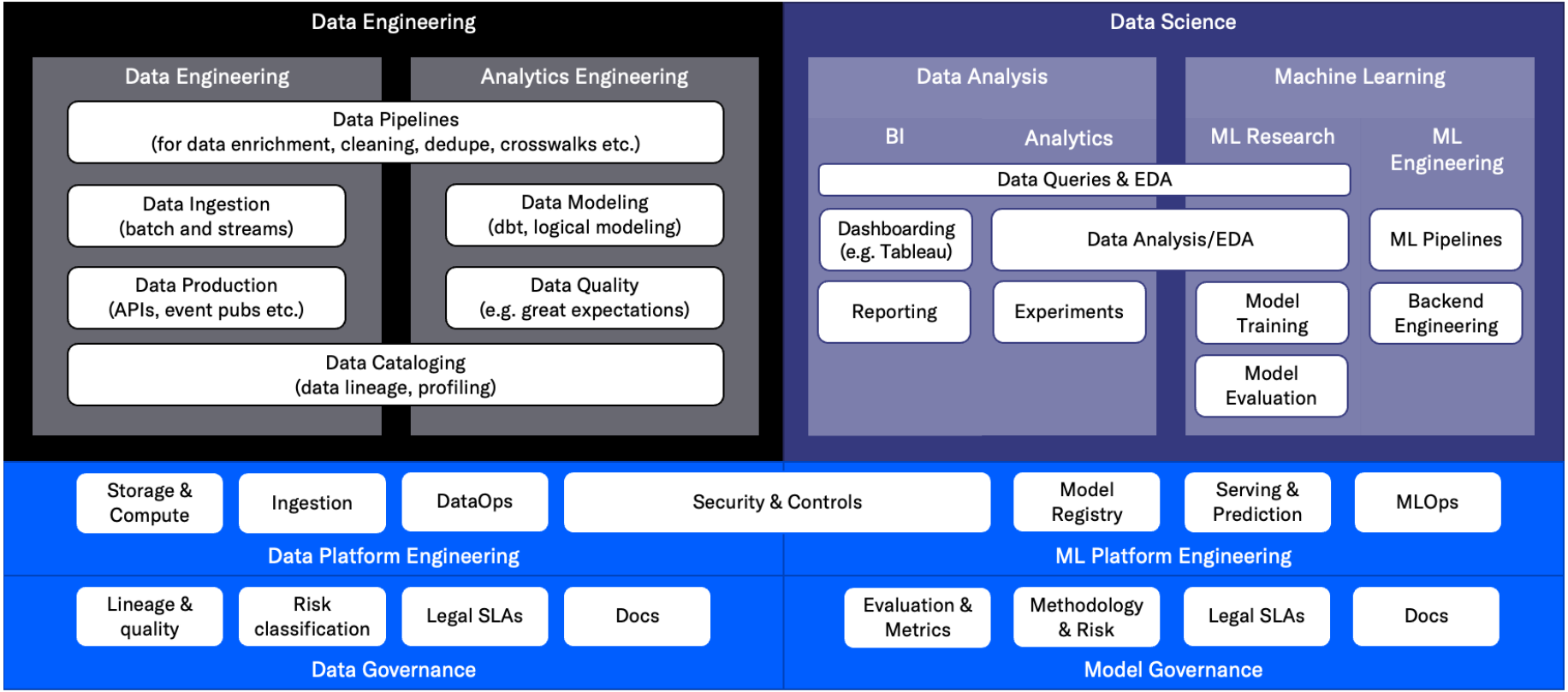
OUR DATA COMMUNITY @ MOODY'S KYC



- Data Scientists
- ML Engineers
- Data Engineers
- Data Analysts
- Risk Analysts

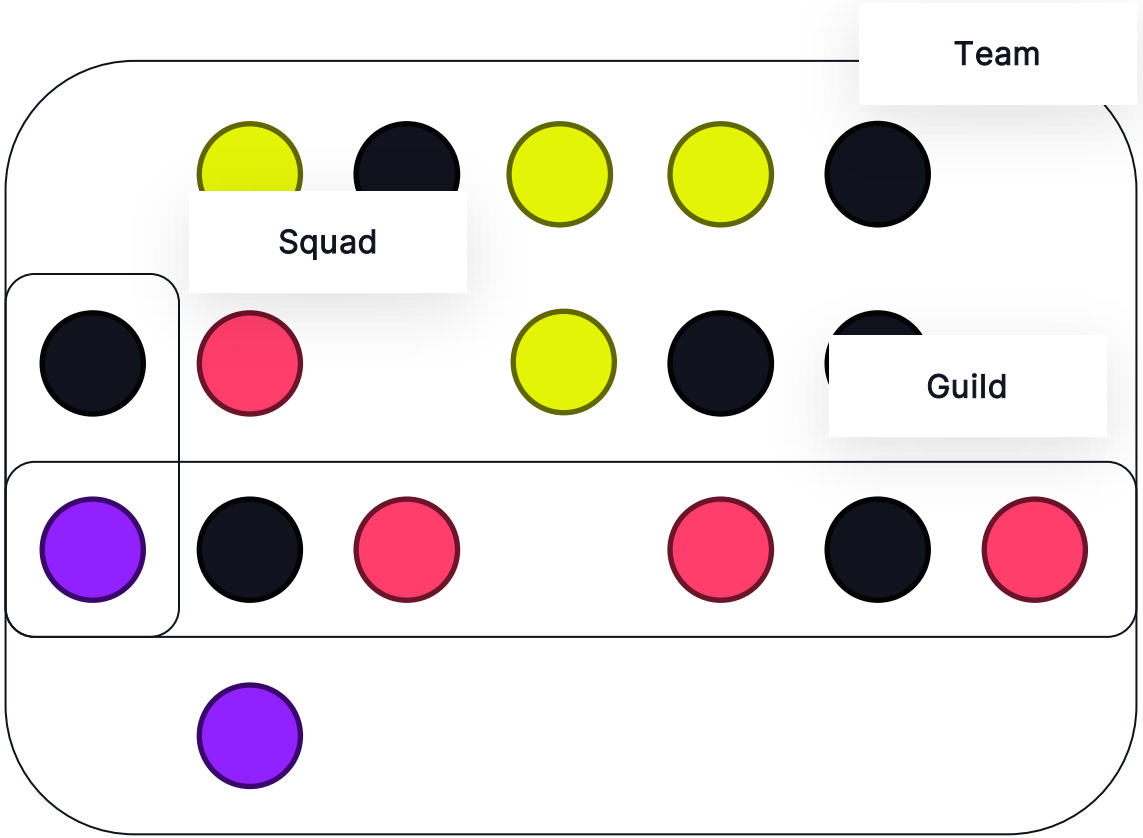
©2024 Databricks Inc. — All rights reserved

FUNCTIONAL ROLES & SPECIALIZATIONS



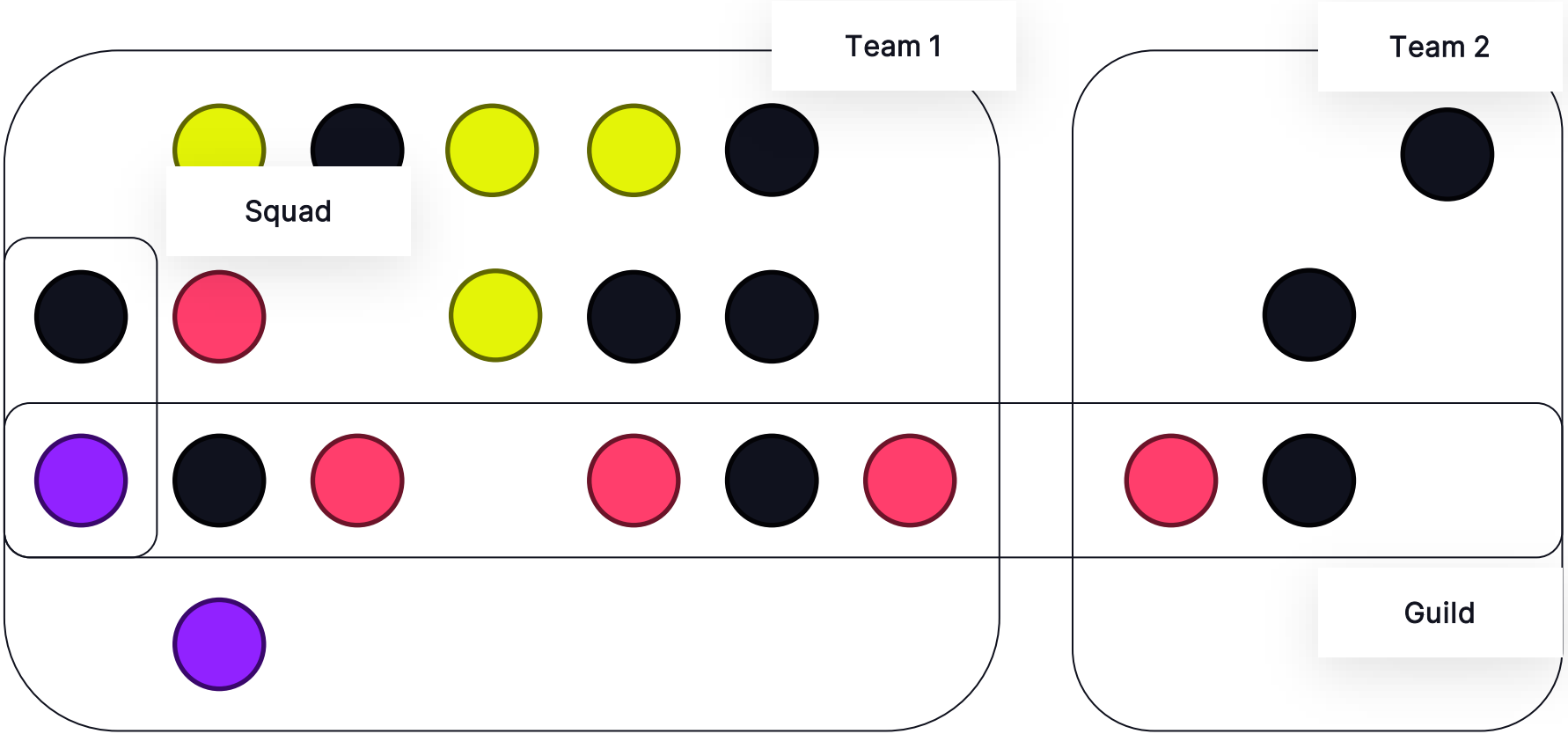
©2024 Databricks Inc. — All rights reserved

HOW WE COMMUNICATE



©2024 Databricks Inc. — All rights reserved

SCALING COMMUNICATION



©2024 Databricks Inc. — All rights reserved

EMPOWER THE RIGHT COMMUNICATION

Having all of these makes a community

Teams

- Centralized to have an opinion on how things should be done
- Shares a reporting line into a leader

Squads

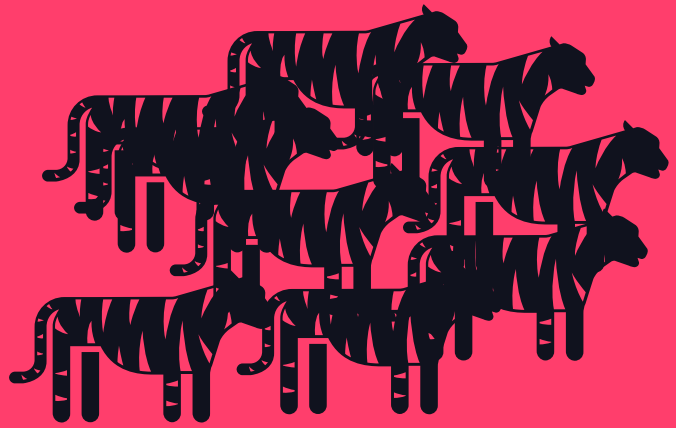
- Federated to operate with guidance from the Functional Guild as well as Data Team
- Shares a project focus, and is cross-functional

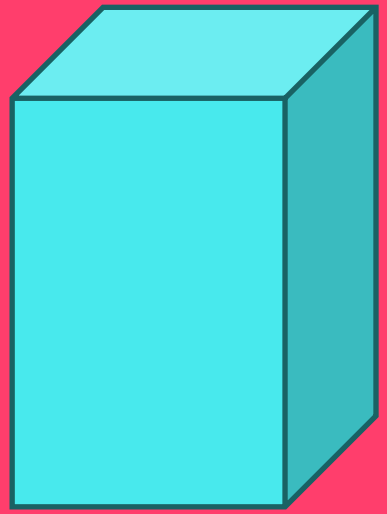
Guilds

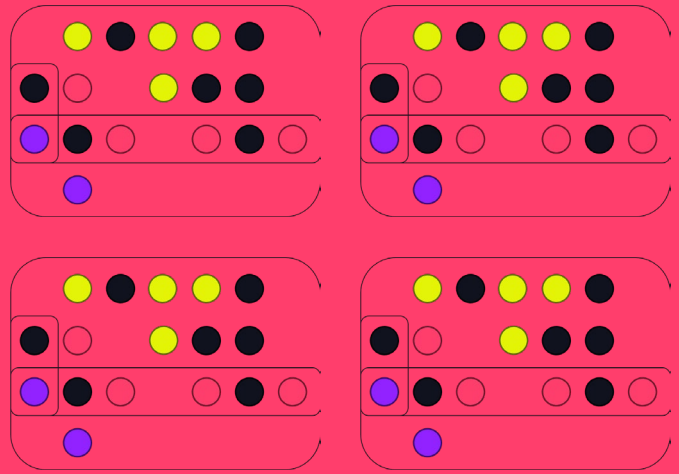
- Decentralized for individuals to share opinions, tech stack, and best practices
- Shares a technical depth and domain knowledge of a data specialization

©2024 Databricks Inc. — All rights reserved

DATA MESH, THE PRAGMATIST WAY



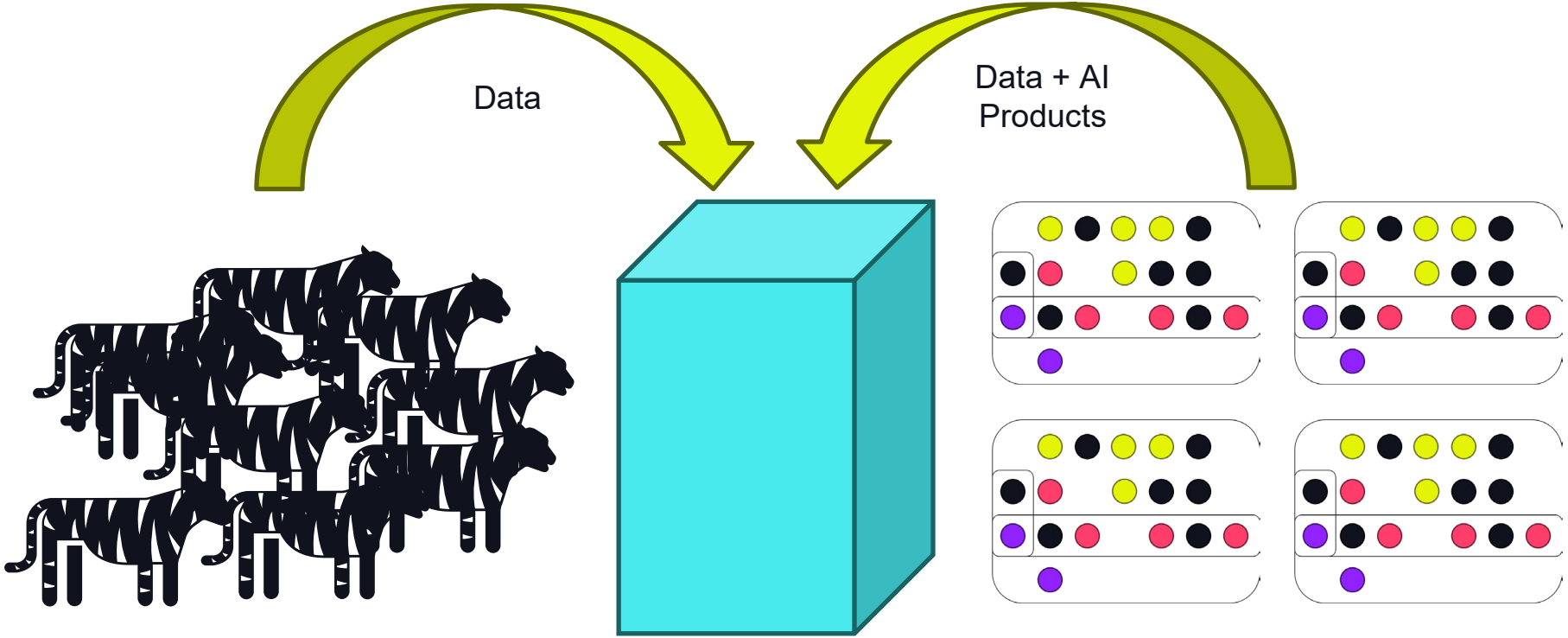




©2024 Databricks Inc. — All rights reserved

DATA PRODUCTION & CONSUMPTION

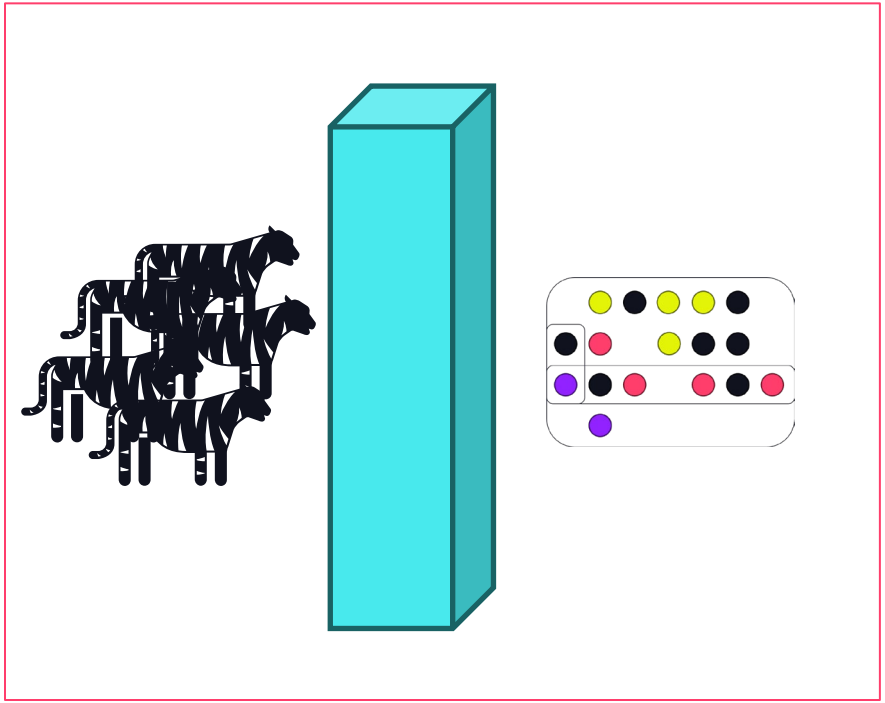
After Democratization



©2024 Databricks Inc. — All rights reserved

SOLUTION 1

Centralized

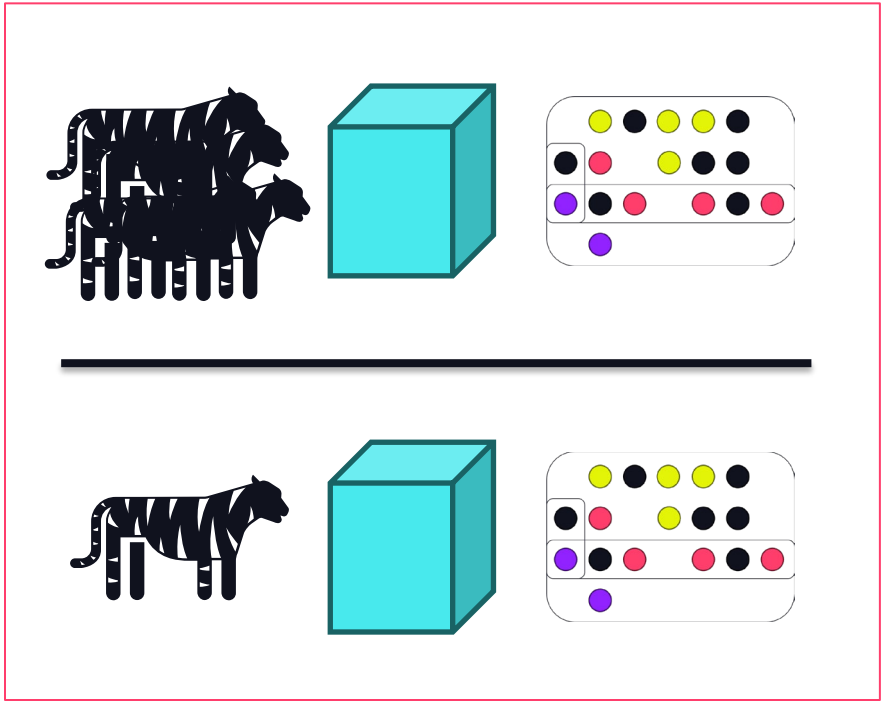


- One data team and tech stack
- High trust, low friction in data production & consumption
- Everyone has access to everything

©2024 Databricks Inc. — All rights reserved

SOLUTION 2

Decentralized

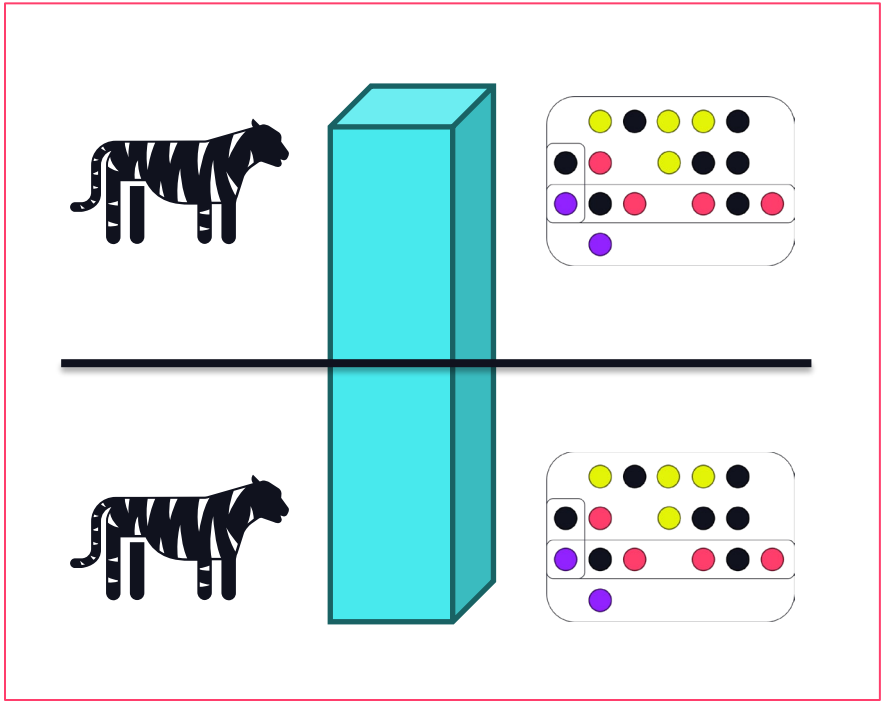


- Individually, all the benefits of centralization
- Overall, fragmented & redundant tech stack/costs across teams
- Data sharing & production across teams is difficult or simply not permitted
- Low trust across boundaries

©2024 Databricks Inc. — All rights reserved

SOLUTION 3

Data Mesh built on Databricks



- Centralized technology with federated ownership
- Decentralized access and permissions
- Medium trust, some light friction sharing data
- All-in-one platform with minimal contracts and solution-stitching

©2024 Databricks Inc. — All rights reserved

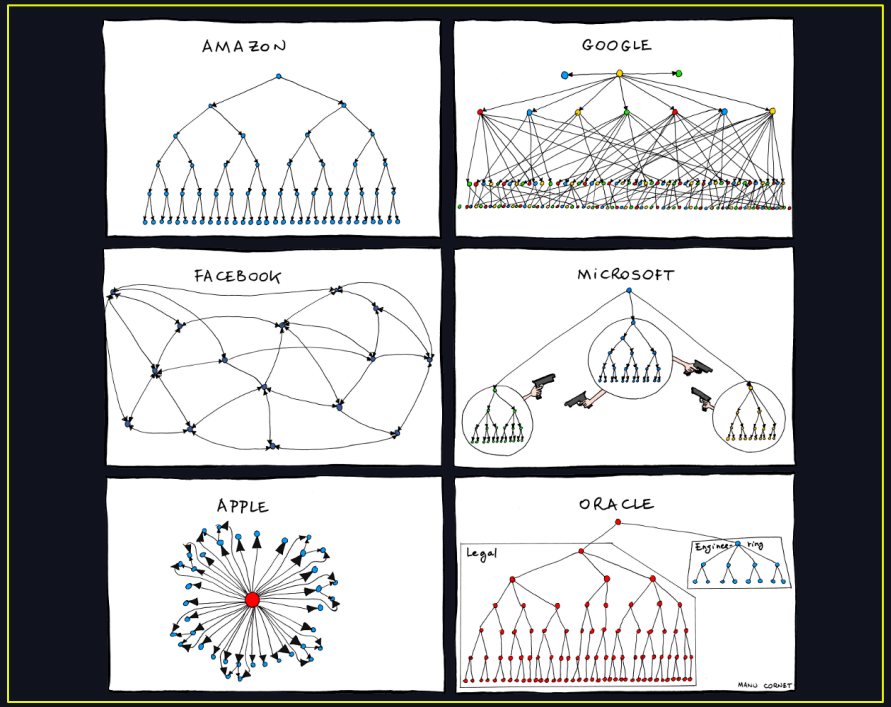
ROLE CLARITY IN DATABRICKS

Data Scientists	Notebooks	Compute		
ML Engineers	MLOps	Feature & Vector Stores	Model Serving	
Data Engineers	Workflows	DBSQL	Delta Live Tables	
Data Analysts	Lakeview	Notebooks		
Risk Analysts	Unity Catalog			

databricks

©2024 Databricks Inc. — All rights reserved

WHAT'S PRAGMATIC HERE?



- No idealistic right or wrong towards centralization, federation & decentralization
- Scale efficient collaboration & communication, not grow empires
- Go with how the business is divided, don't fight the business. (De-)centralization is always relative to something

©2024 Databricks Inc. — All rights reserved

Scale, not grow	Don’t hunt for unicorns	Databricks has mostly everything
Data role clarity	Squads, Guilds, Teams	Communicate above all else
Centralize tech	Federate content	Decentralize ownership

©2024 Databricks Inc. — All rights reserved