DATA+AI SUMMIT 2022

MC^2 Platform:

Enabling Learning on Confidential Data

Rishabh Poddar
Co-Founder & CEO
Opaque systems

ORGANIZED BY databricks

The Problem

Organizations often

wish to learn from cross-organization data

but

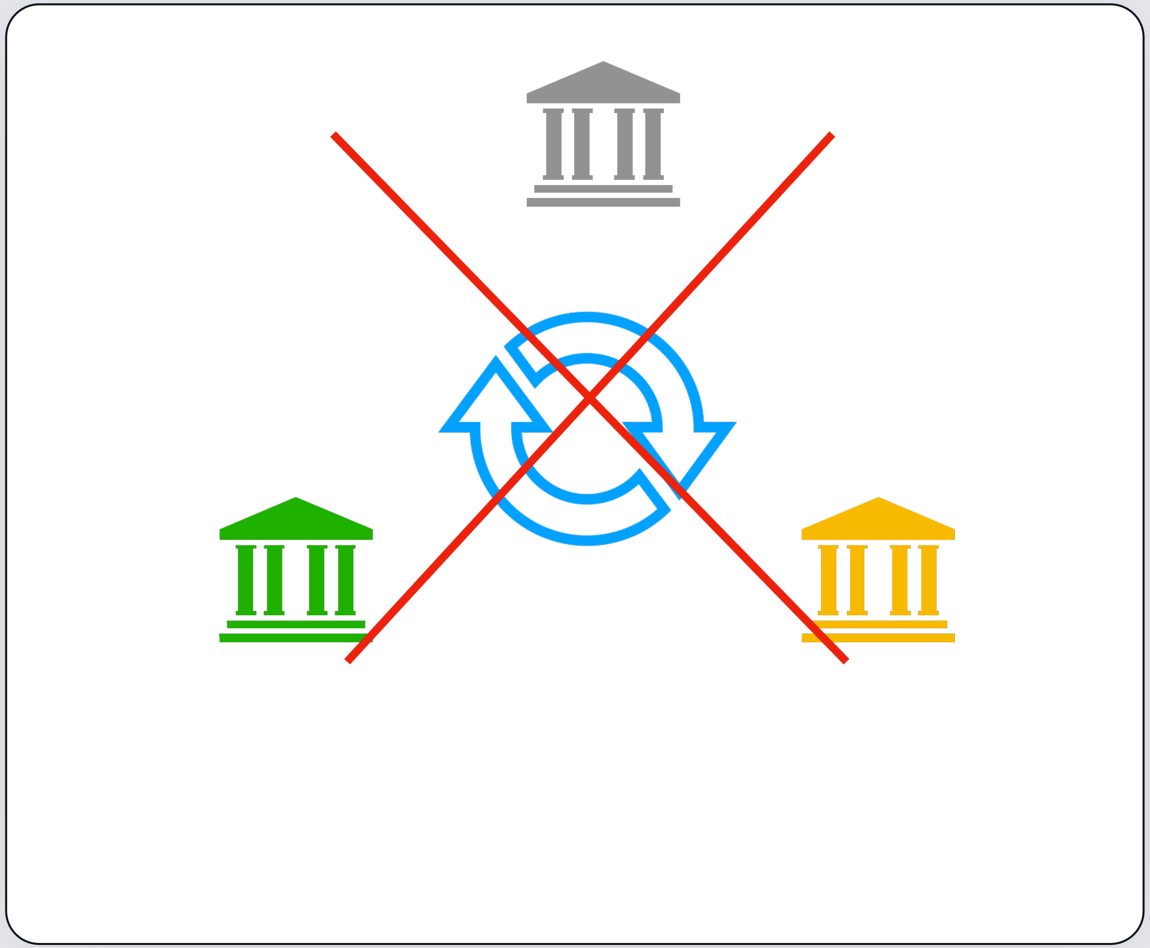
have confidential data they cannot share

Example: Anti-money laundering

- Banks want to detect money laundering
- Criminals hide their traces across different banks

Example: Anti-money laundering

- Banks want to detect money laundering
- Criminals hide their traces across different banks
- To detect money laundering, one needs to learn from multiple banks
- But banks can't share data due to competition / data confidentiality restrictions



"So In the future, ***collaboration will be vital***: across the financial-services industry, government, and law enforcement. The ability to put together our data sets and collaborate on typologies of attack — and the use of both advanced-encryption methods and analytics methods to mine the data — ***will enhance yields by orders of magnitude***."

Chief Risk Officer, Scotiabank

Many use cases across industries

Confidential data locked down in silos, but holds tremendous value

Financial crime

Human trafficking, money laundering, fraud

Healthcare

Patient profiling, disease prediction, clinical studies

Customer insights

Marketing campaigns, cross-selling opportunities

How to solve without trusted third parties?

Requirement: Protecting data *in use*

Existing encryption

Encryption at Rest

Encrypted data in storage (databases, blob storage, etc.)

Encryption in Transit

Encrypted data sent over the network

Requirement: Protecting data *in use*

Existing encryption

Encryption at Rest

Encrypted data in storage (databases, blob storage, etc.)

Encryption in Transit

Encrypted data sent over the network

Requirement

Encryption in use

Unencrypted data in memory during processing introduces risk

MC2: Multi-party Confidential Computing

github.com/mc2-project/mc2

Analytics and machine learning on confidential data

“Sharing without showing the data”

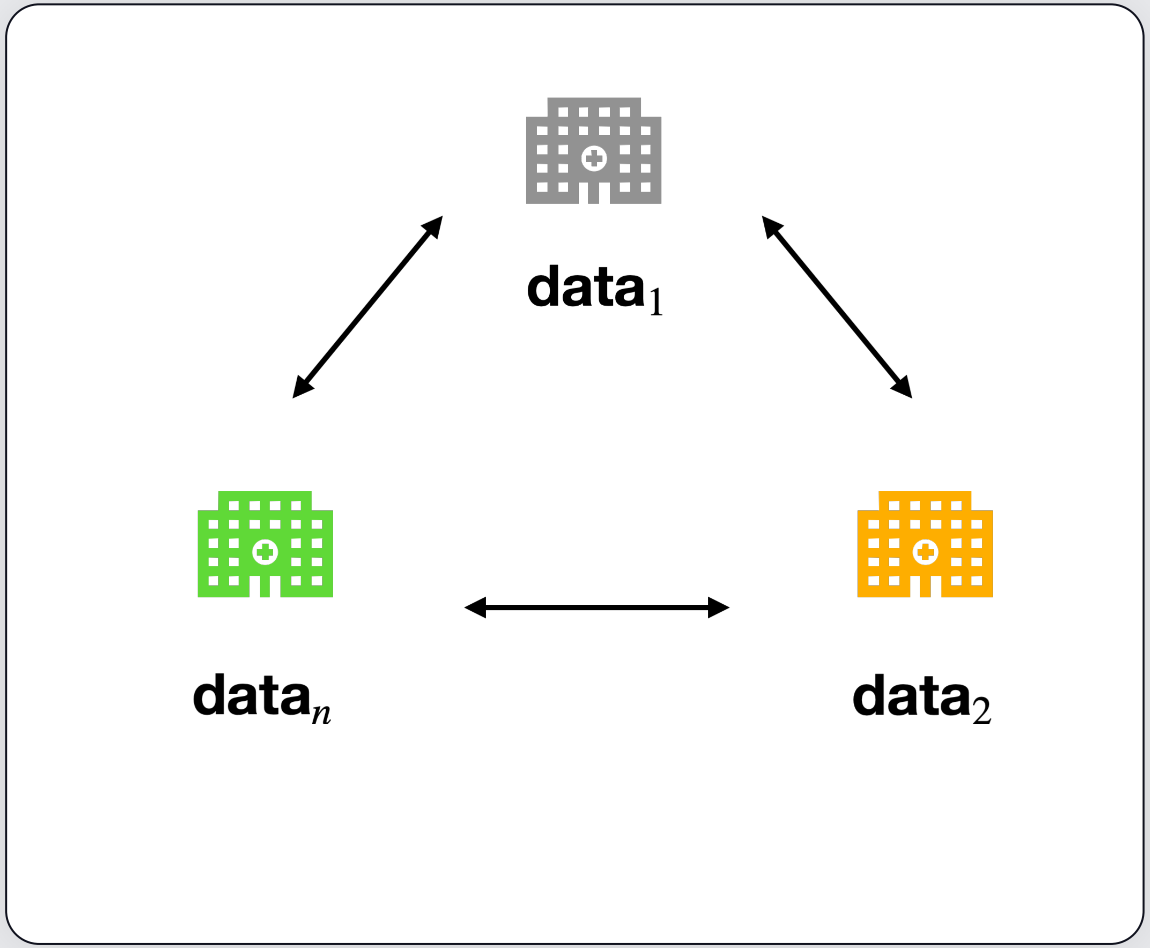
Two primary approaches

Each with its own tradeoffs

Two primary approaches

1 Cryptographic protocols: MPC / Homomorphic encryption

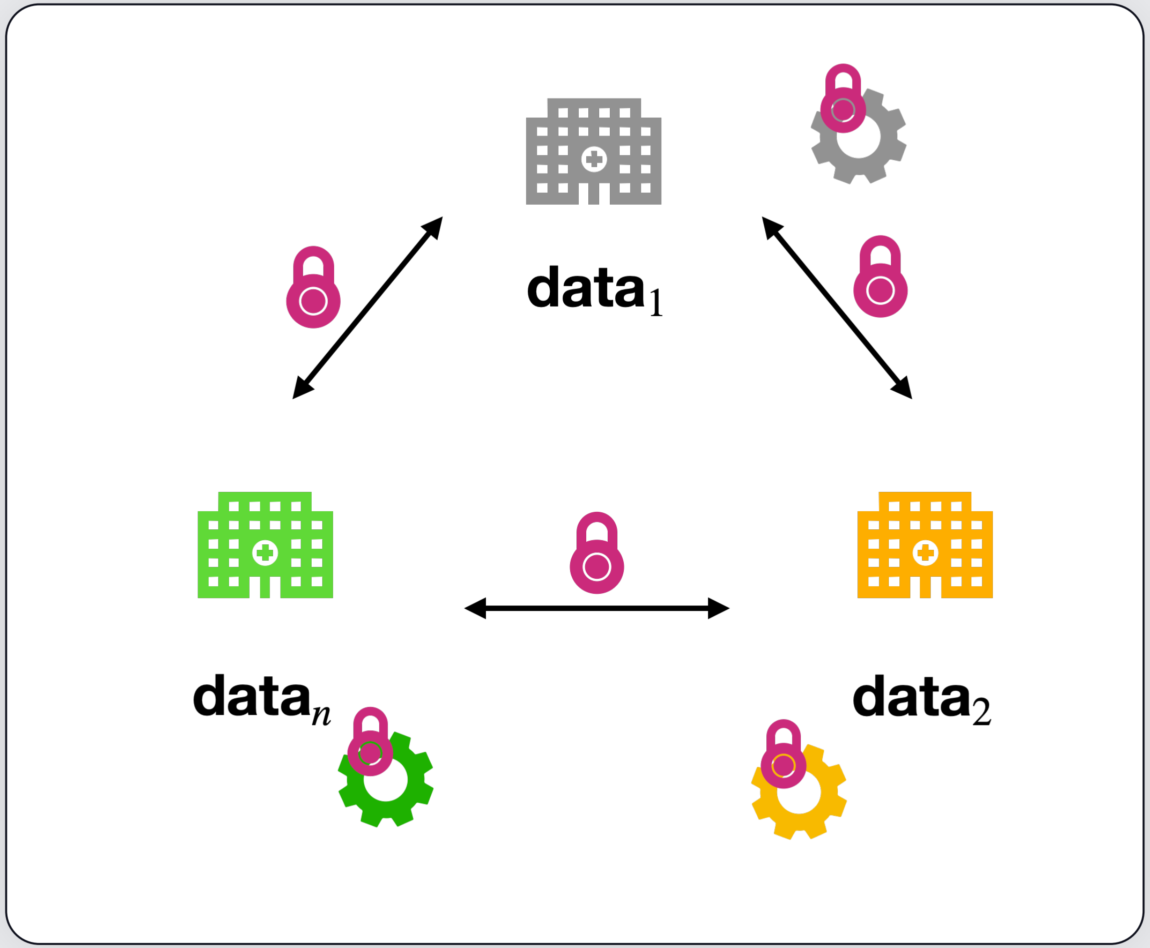
- Parties compute *F(data_1, … , data_n)* without any party learning the data of another beyond the function result



Two primary approaches

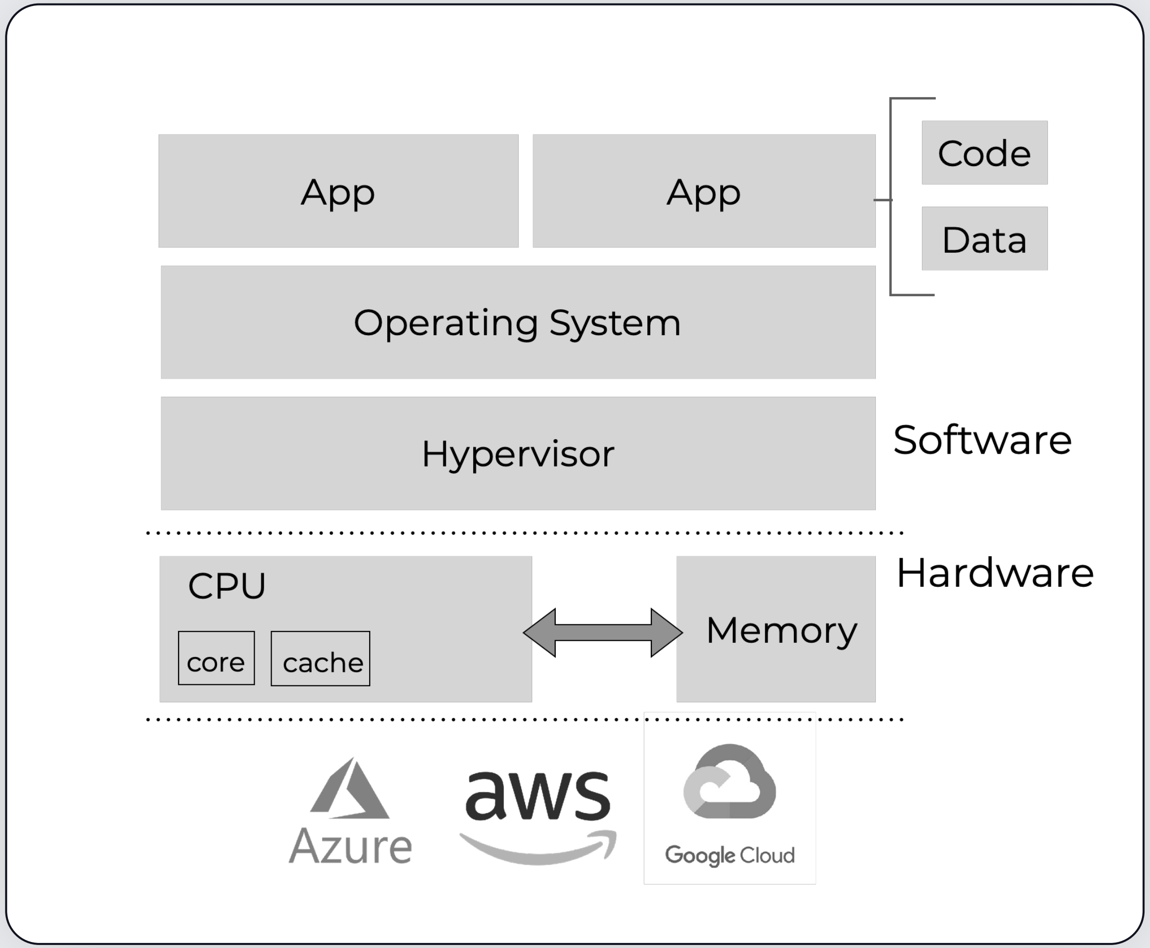
1 Cryptographic protocols: MPC / Homomorphic encryption

- Parties compute *F(data_1, … , data_n)* without any party learning the data of another beyond the function result
- They exchange encrypted data and compute on encrypted data



Two primary approaches

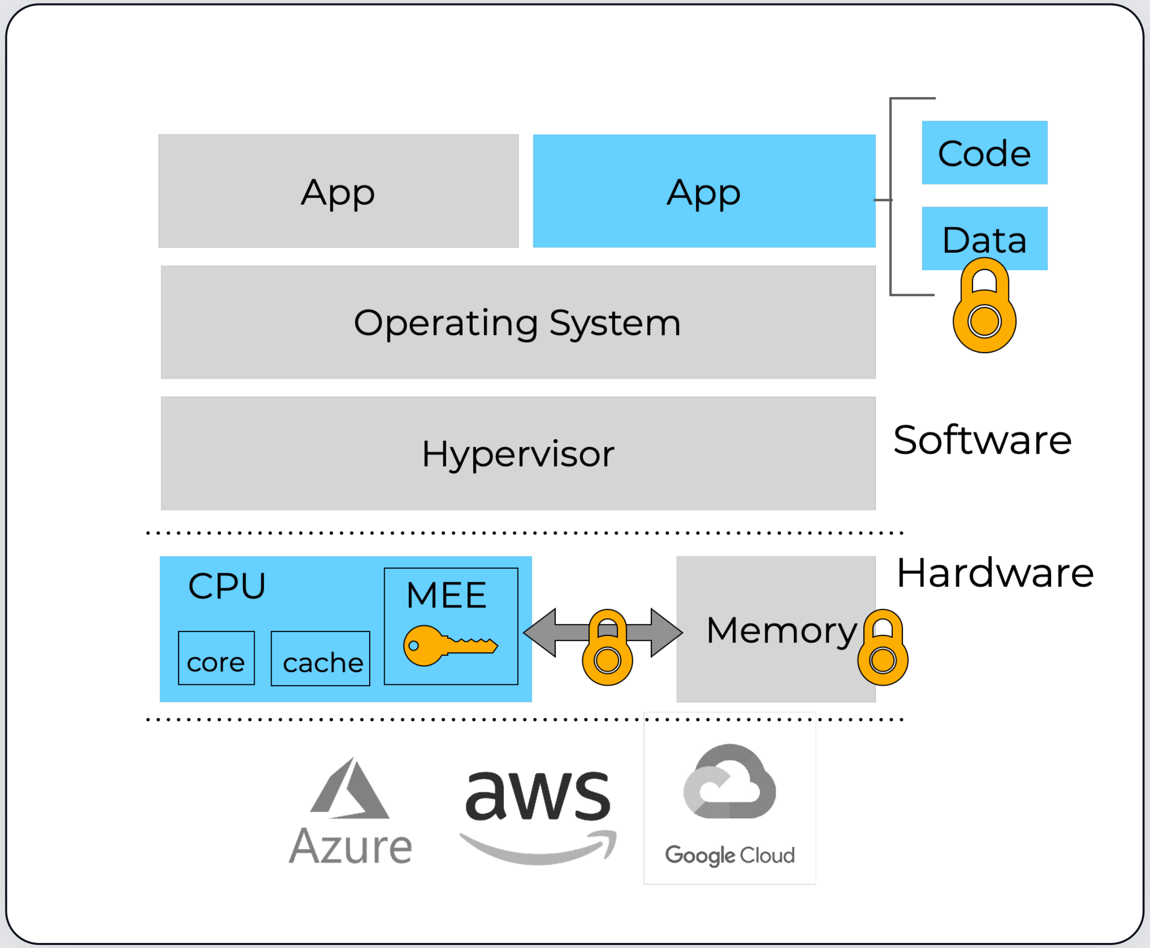
2 Secure hardware enclaves (e.g. Intel SGX)



Two primary approaches

2 Secure hardware enclaves (e.g. Intel SGX)

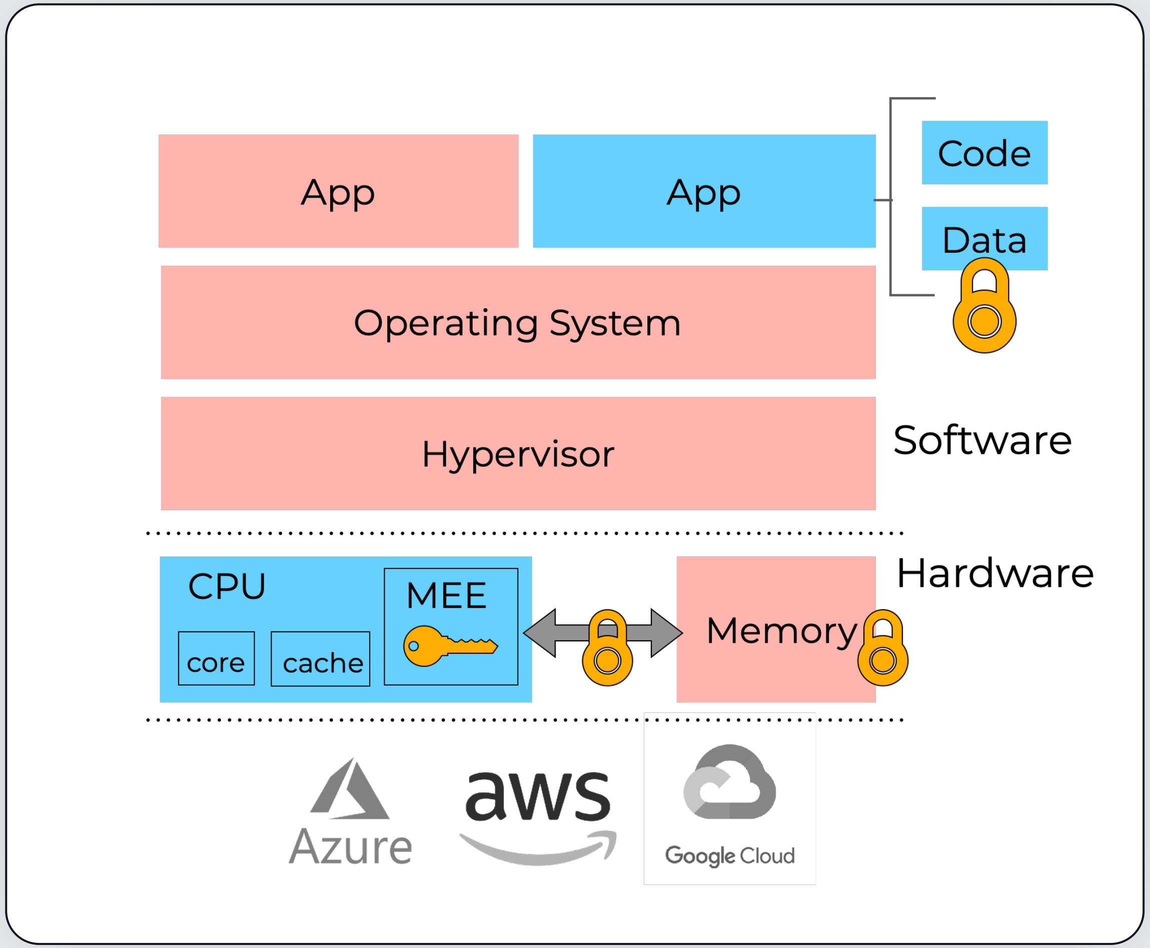
- Hardware-enforced isolated execution environment — protects against attackers with root access or compromised OS



Two primary approaches

2 Secure hardware enclaves (e.g. Intel SGX)

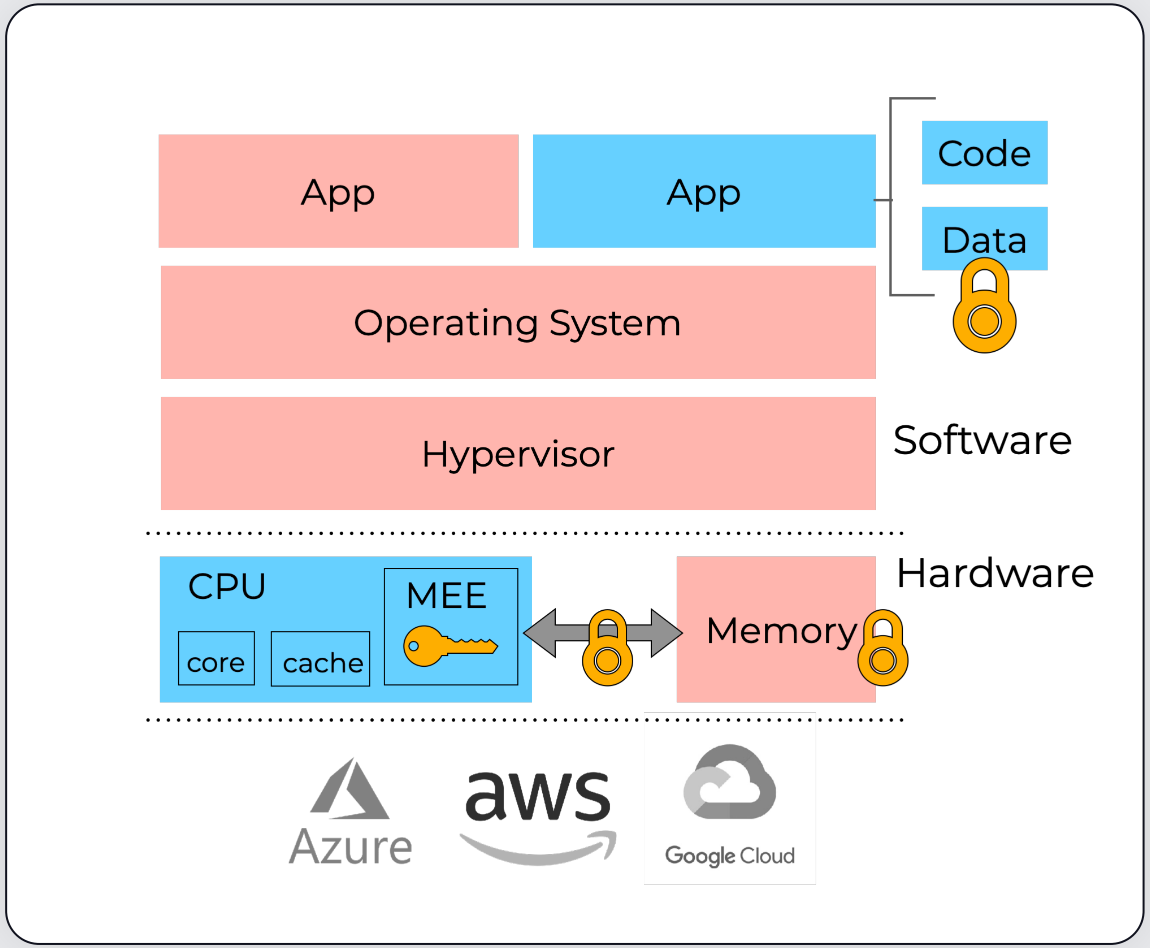
- Hardware-enforced isolated execution environment — protects against attackers with root access or compromised OS



Two primary approaches

2 Secure hardware enclaves (e.g. Intel SGX)

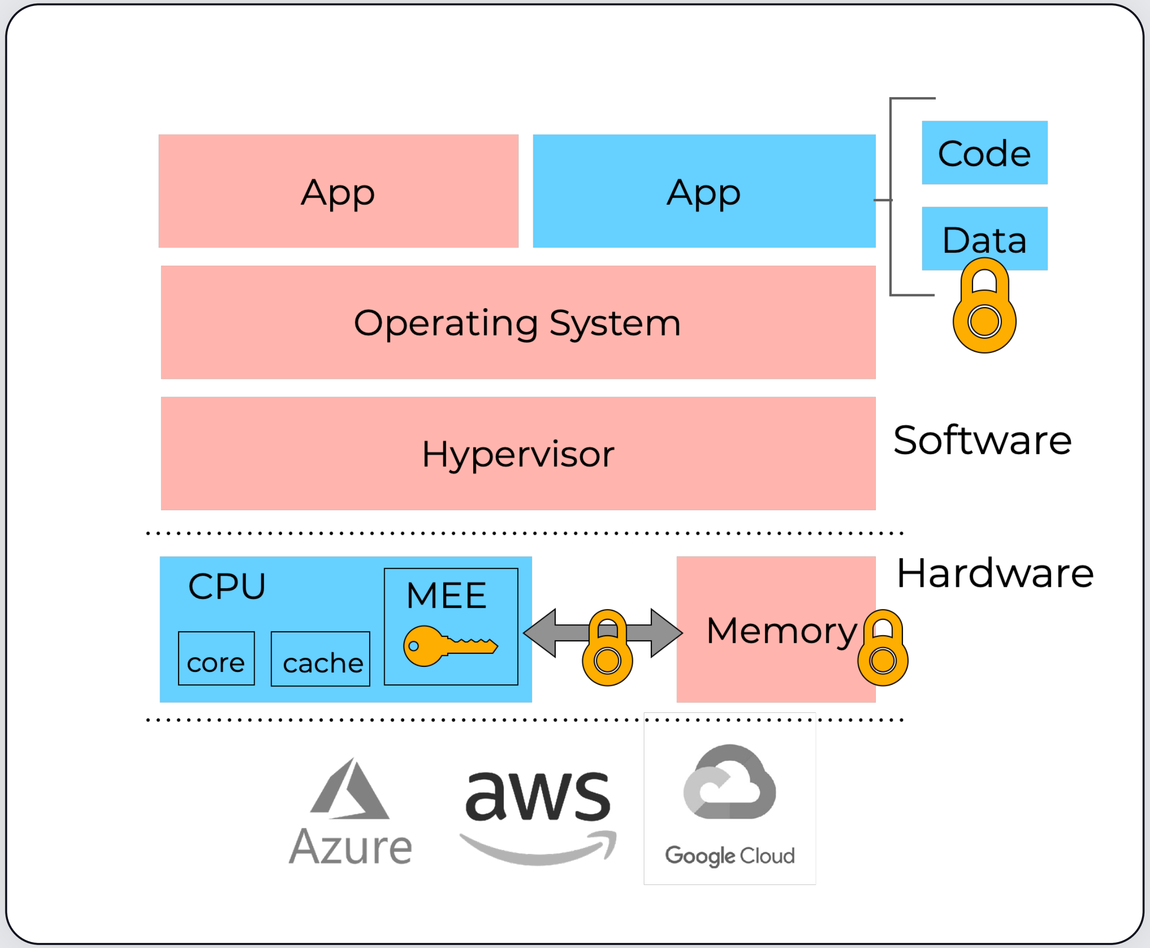
- Hardware-enforced isolated execution environment — protects against attackers with root access or compromised OS
- Remote client can verify enclave code via remote attestation



Two primary approaches

2 **Secure hardware enclaves (e.g. Intel SGX)**

- Hardware-enforced isolated execution environment — protects against attackers with root access or compromised OS
- Remote client can verify enclave code via remote attestation
- Supported by major CPU vendors and cloud providers



Two primary approaches

Each with its own tradeoffs

	Cryptographic Protocols (FHE, MPC)	Secure hardware enclaves (e.g. Intel SGX)
Efficiency	Prohibitively slow for complex analytics / ML training	Can support arbitrary workloads nearly as scalable as plaintext computation
Security	Private data always remains encrypted, but FHE does not provide integrity of data and computation	Private data and models remain encrypted in memory but can be vulnerable to side-channels

Two primary approaches

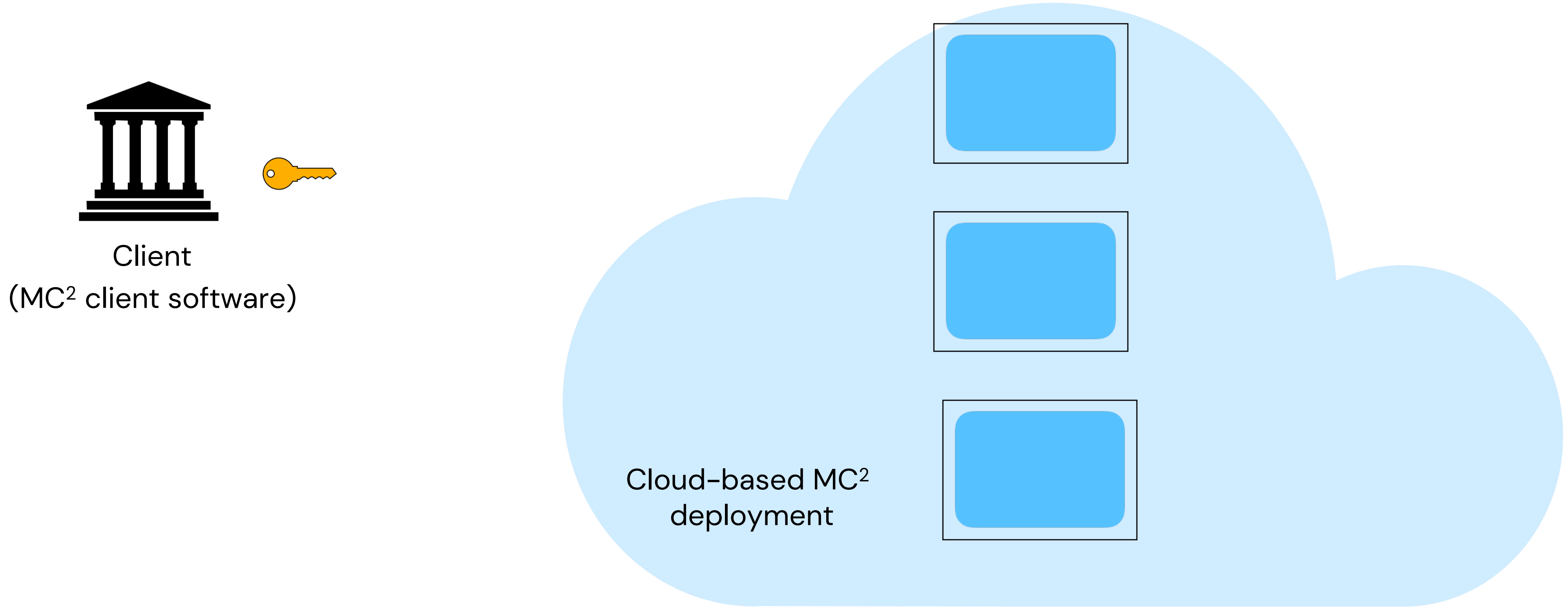
Each with its own tradeoffs

	Cryptographic Protocols (FHE, MPC)	Secure hardware enclaves (e.g. Intel SGX)
Efficiency	Prohibitively slow for complex analytics / ML training	Can support arbitrary workloads nearly as scalable as plaintext computation
Security	Private data always remains encrypted, but FHE does not provide integrity of data and computation	Private data and models remain encrypted in memory but can be vulnerable to side-channels

Addressed via cryptographic fortification in MC2

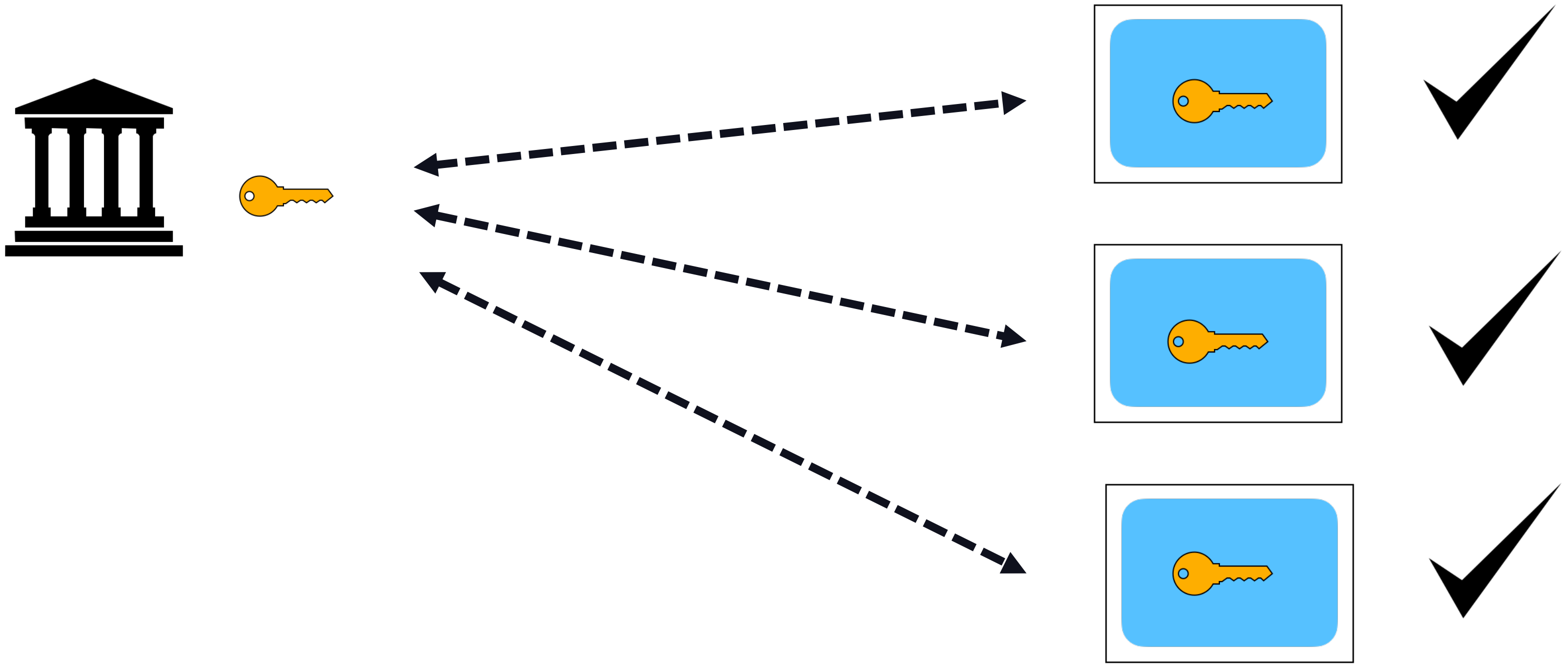
Example workflow

Setup: Cluster of secure hardware enclaves in the cloud



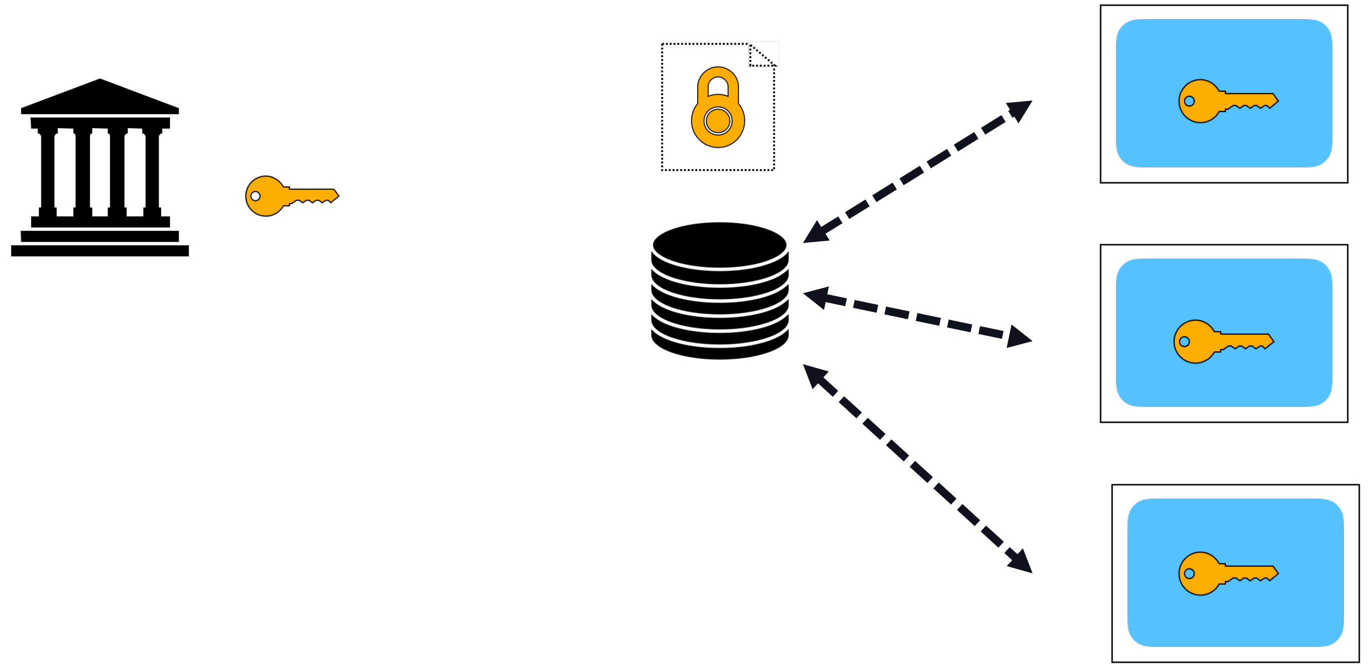
Example workflow

1 Client verifies enclave cluster via remote attestation



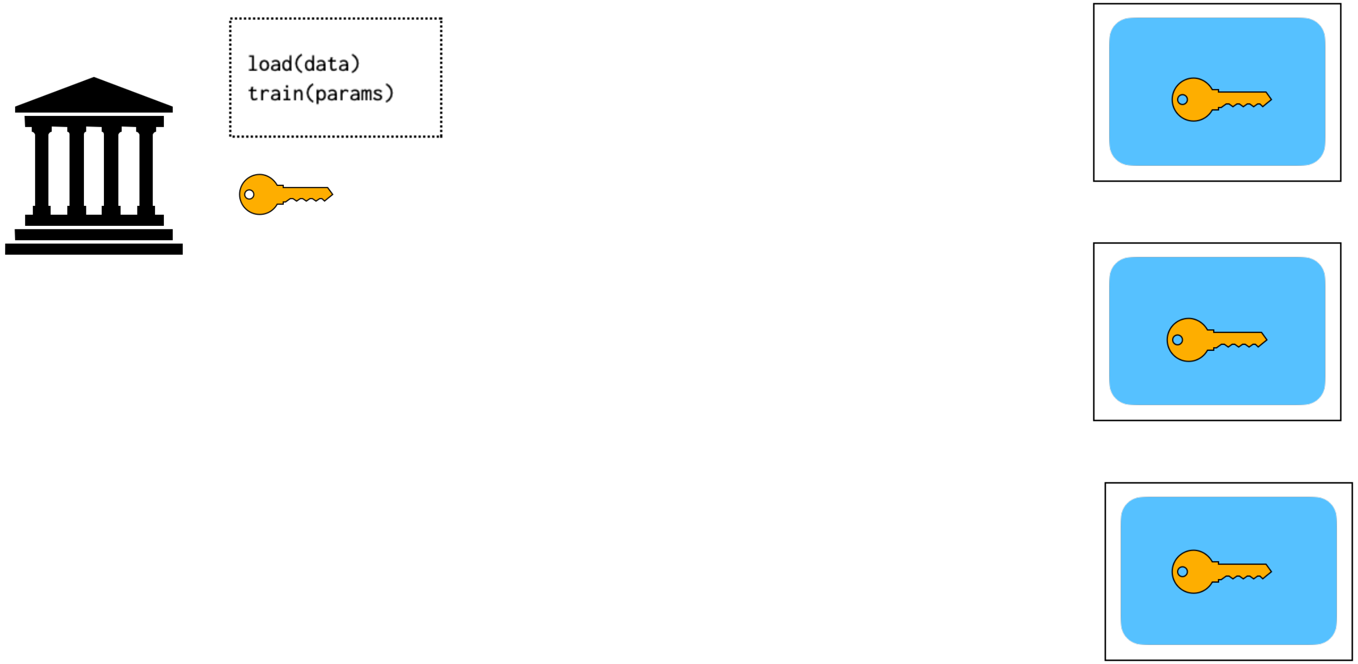
Example workflow

2 Client transfers encrypted data to the cloud



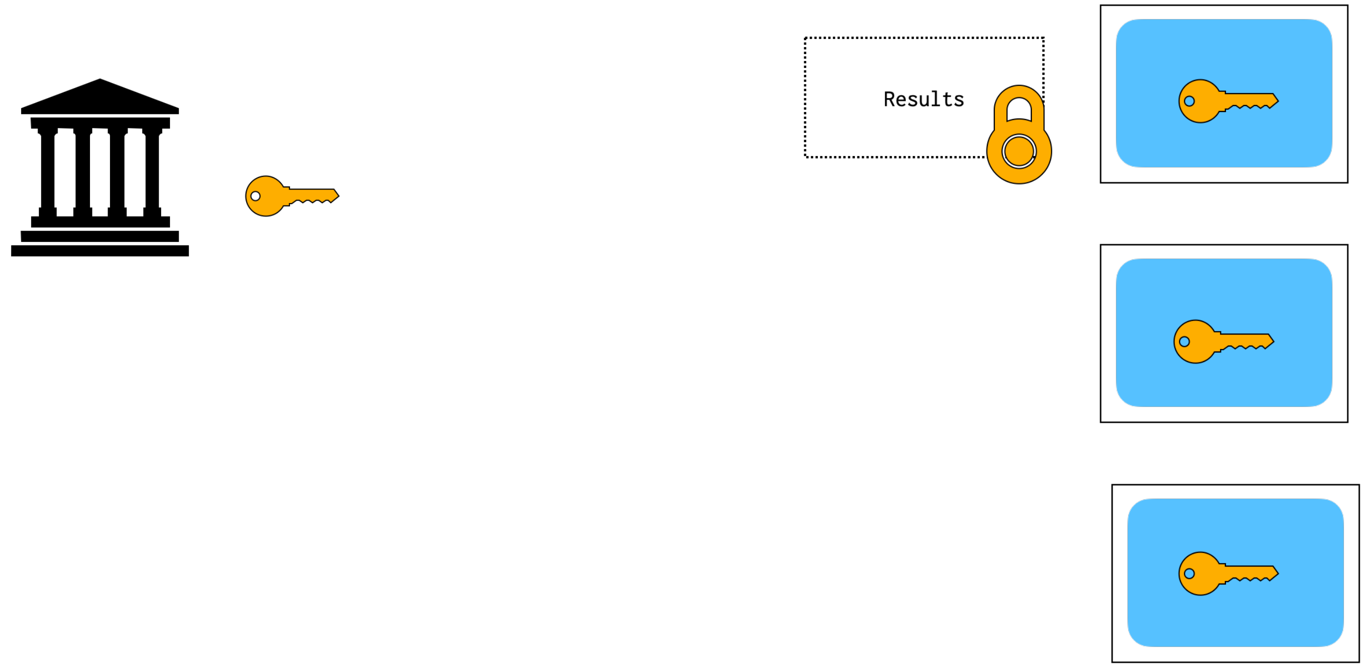
Example workflow

3 Client submits job / script



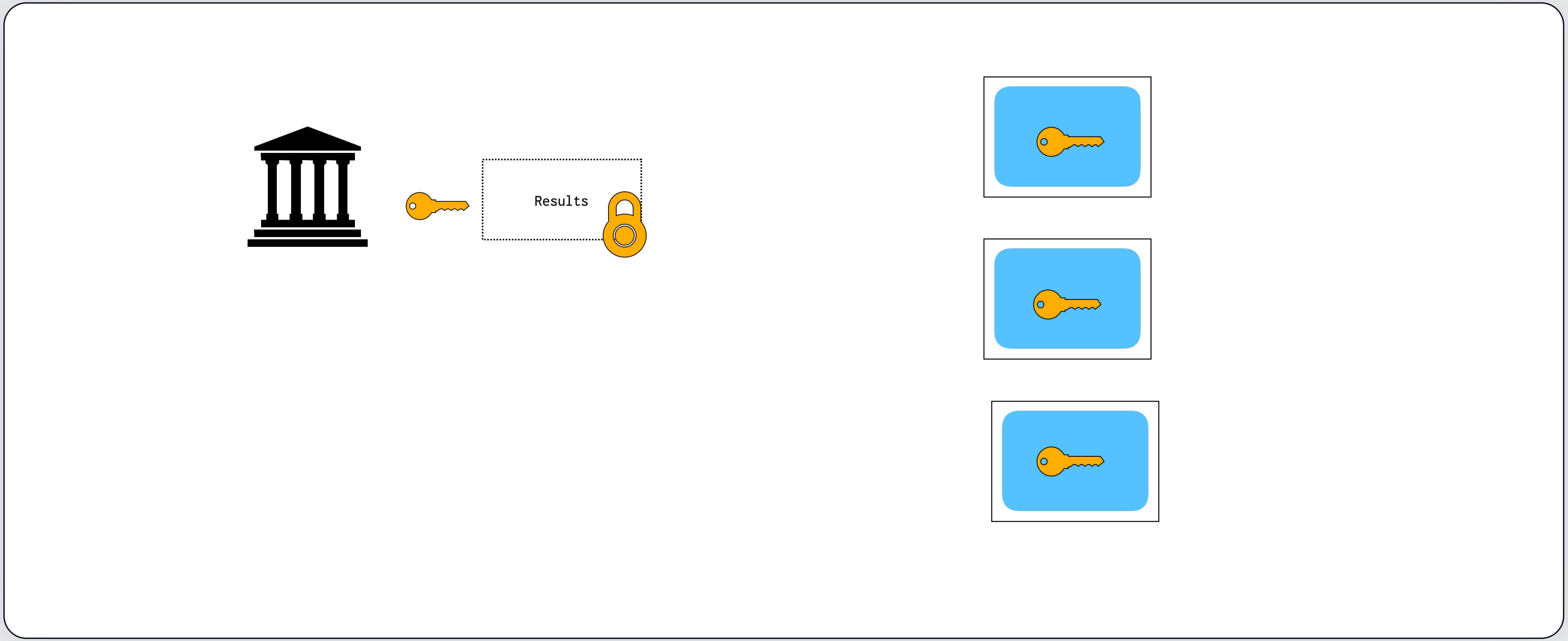
Example workflow

3 MC² processes the data and outputs encrypted results



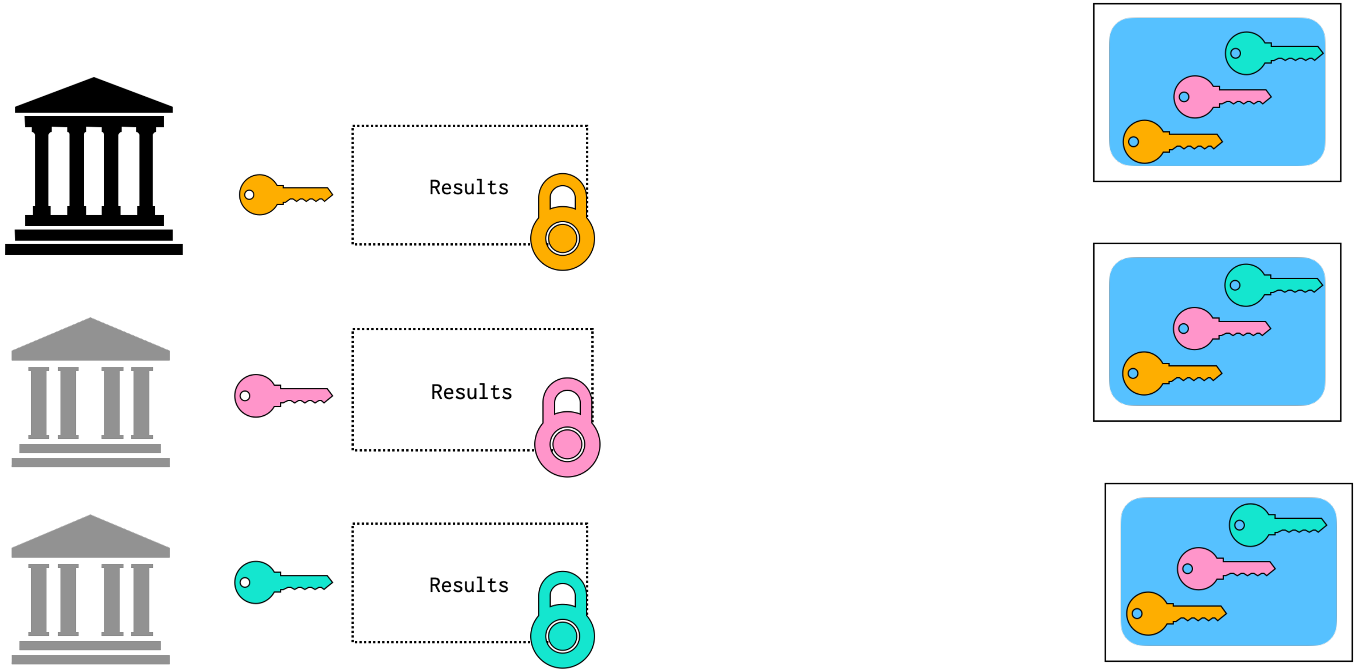
Example workflow

3 MC² processes the data and outputs encrypted results



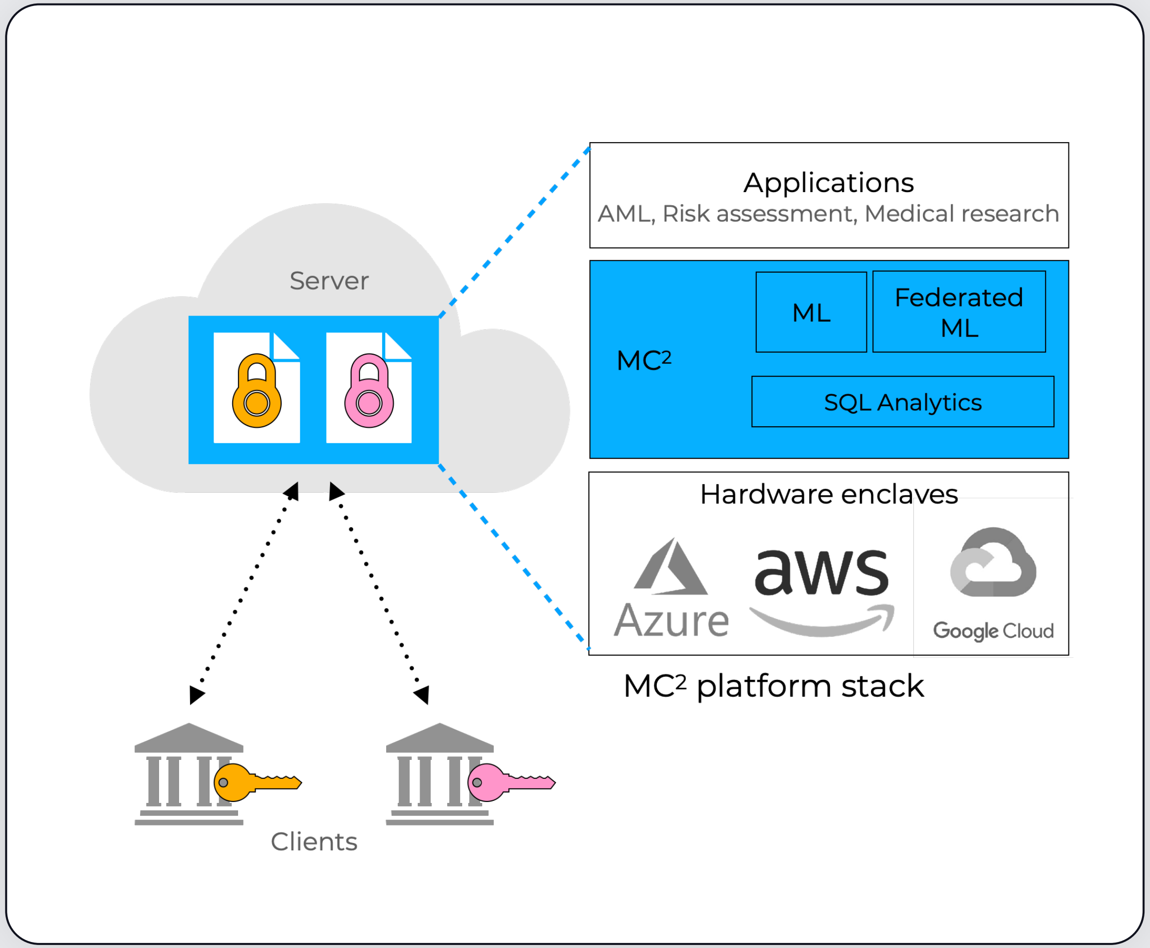
Example workflow

3 MC^2 processes the data and outputs encrypted results



Platform

- Easy-to-use, efficient
  - Spark SQL
  - Machine learning (e.g. XGBoost)
  - Federated learning
- Adoption / collaborators



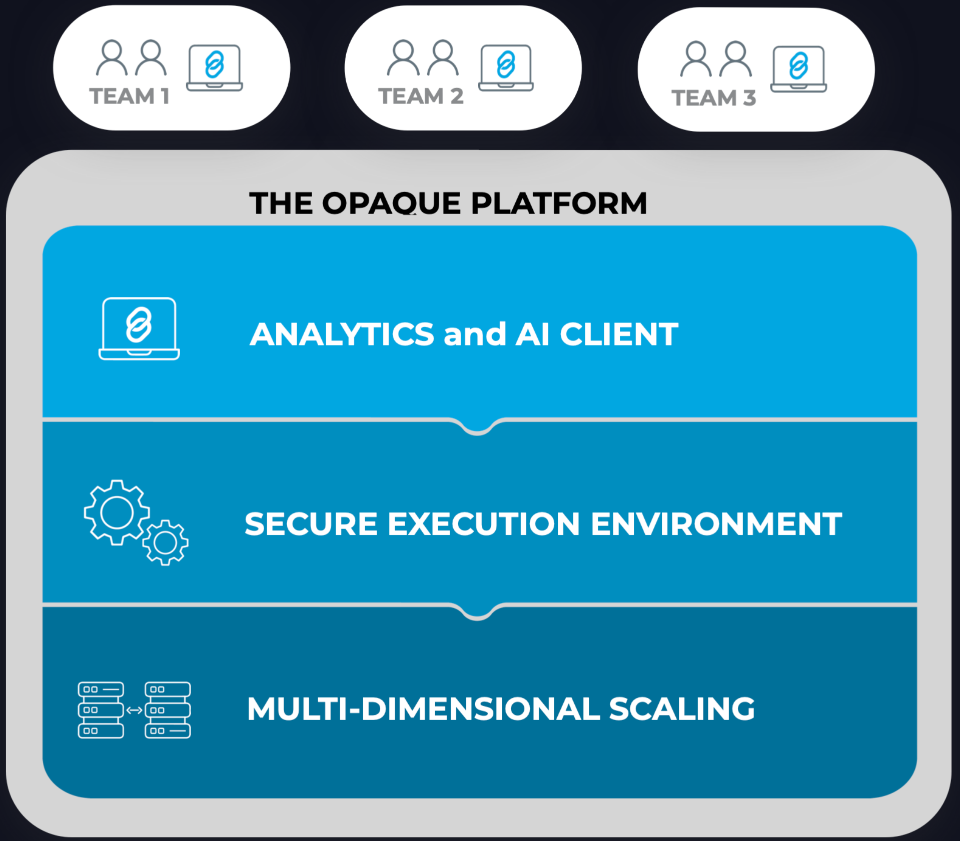
Demo

Demo: MC^2 on Azure

Opaque : The Confidential Computing Platform for Collaborative Analytics and AI at Scale

https://opaque.co

Opaque : The Confidential Computing Platform for Collaborative Analytics and AI at Scale



Instantiate clusters, set policies, enable SQL-based analytics and AI / ML models using standard tools

Execute confidential collaborative analytics, AI / ML and data sharing on encrypted data

Enable secure inter-enclave communication, orchestration and multi-cloud operations

https://opaque.co

MC^2 Summary

Contact us if you want to collaborate!

https://github.com/mc2-project/mc2

mc2-project.slack.com

mc2-dev@googlegroups.com

rishabh@opaque.co

DATA+AI SUMMIT 2022

Thank you

Rishabh Poddar

Co-Founder and CEO

Opaque systems

rishabh@opaque.co