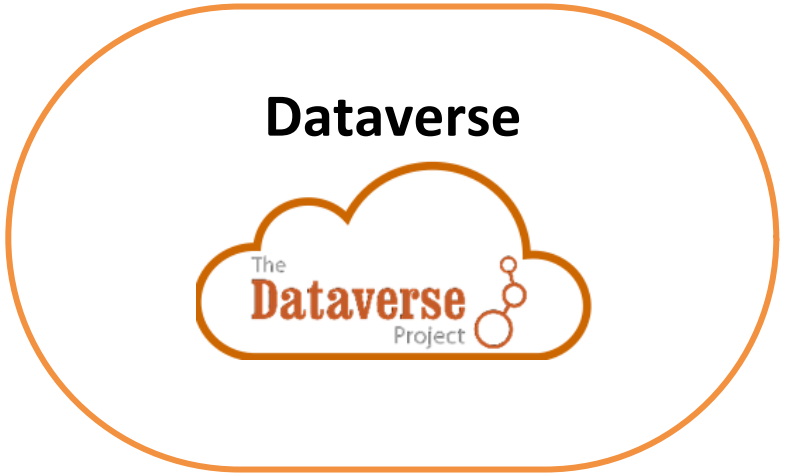
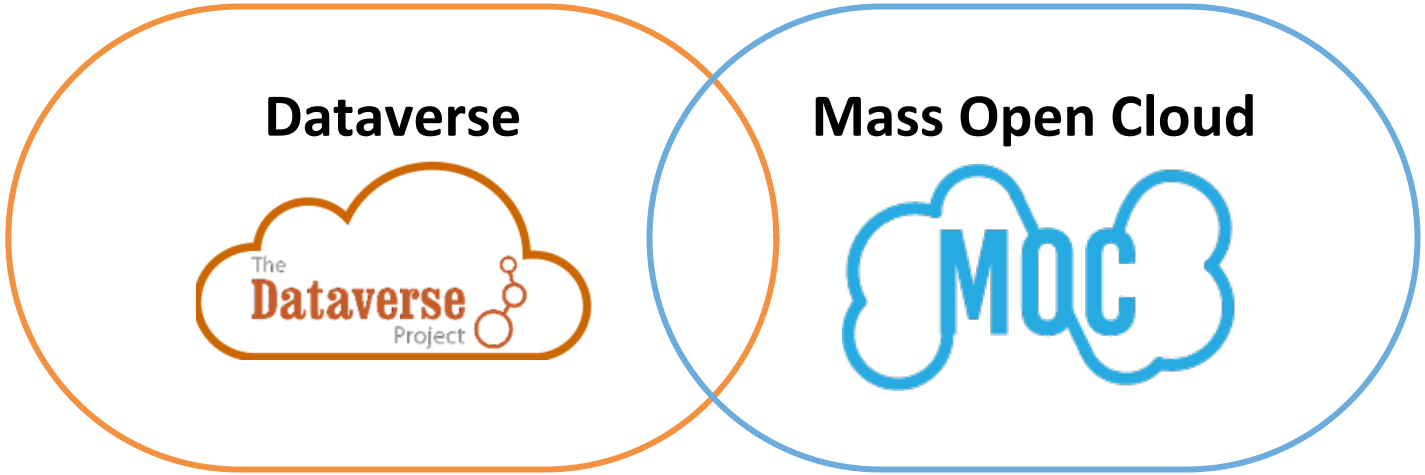
Privacy-Preserving Scientific Data Analysis in an Open Cloud

Mayank Varia, Boston University

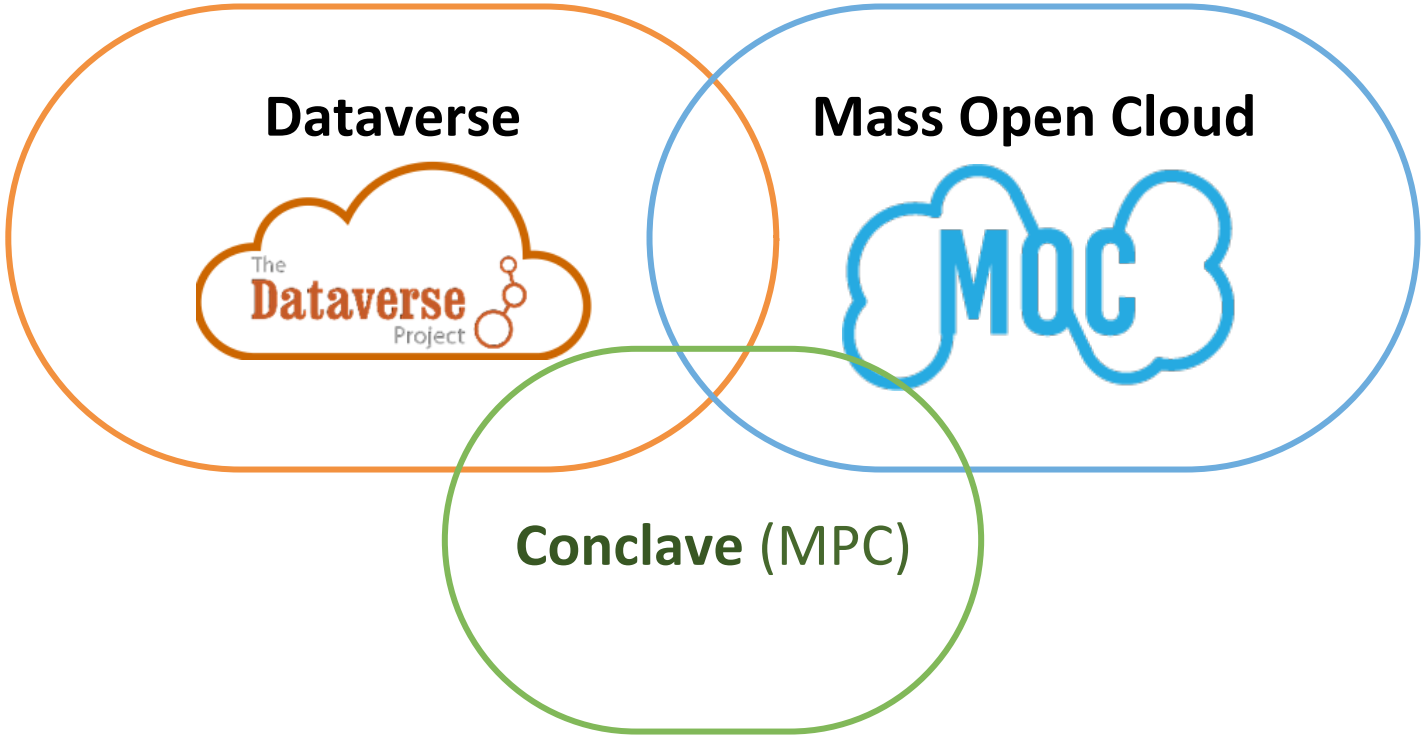
Privacy-Preserving Scientific Data Analysis in an Open Cloud



Privacy-Preserving Scientific Data Analysis in an Open Cloud



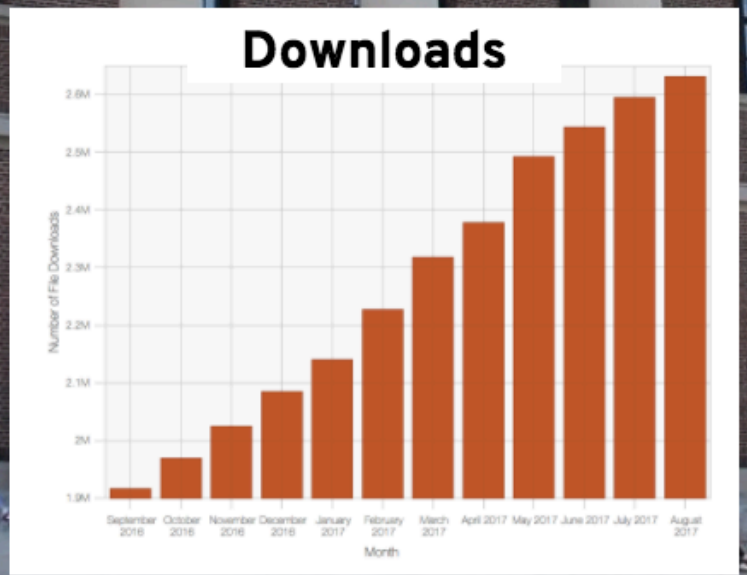
Privacy-Preserving Scientific Data Analysis in an Open Cloud



HOW RESEARCHERS SHARE & USE DATA WITH DATAVERSE

Datasets Added

Downloads



Harvard Dataverse Repository

A public repository for research data

> 70,000 datasets total
> 49,000 datasets uploaded to Harvard Dataverse repository
200 datasets/month

> 340,000 files
4,000 files/month

> 2.5 M downloads
60,000 downloads/month

dataverse.harvard.edu

Cloud Dataverse

Cloud Dataverse

- Extends Dataverse
- Store datasets in Object Store (Swift)
- Adds a compute button next to each dataset that enables on-site computation
  - No need for download

Cloud Dataverse combines the power of cloud computing and storage with access to thousands of datasets from a feature-rich data repository platform


MOC

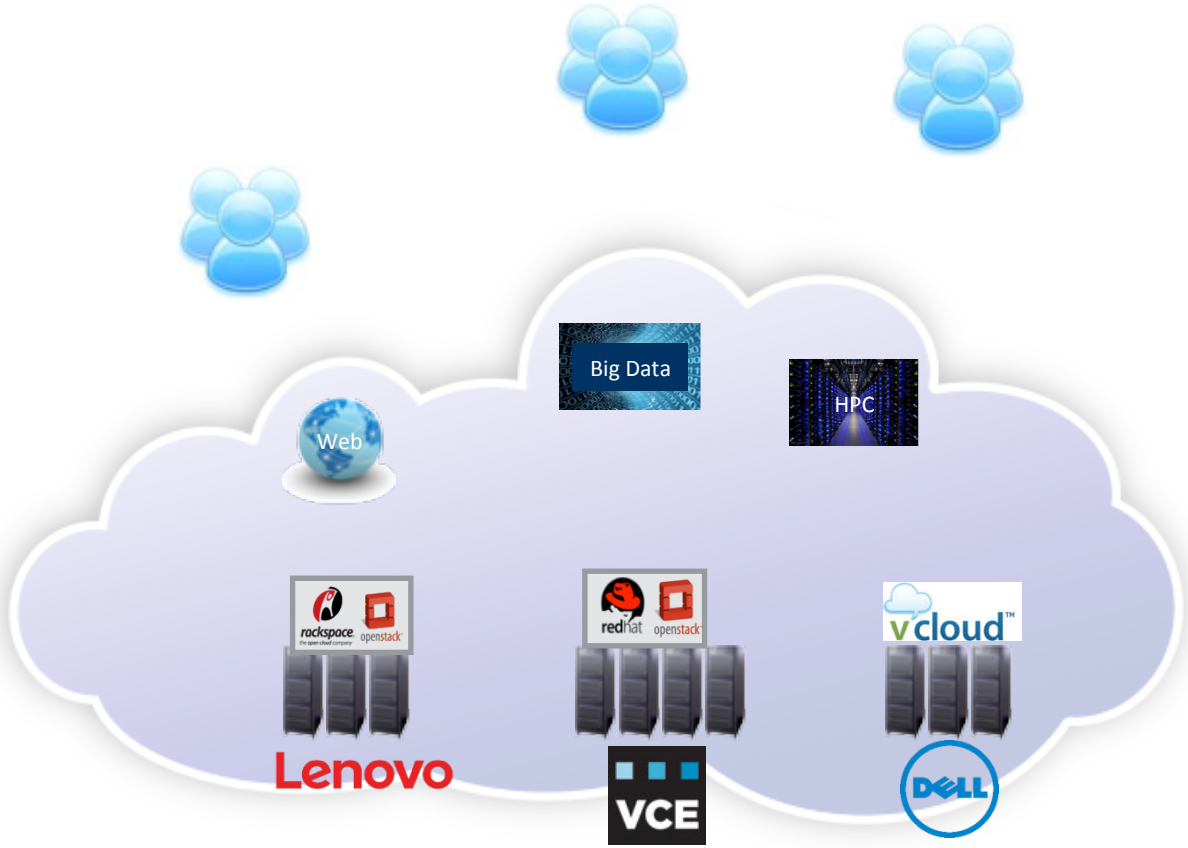
Today's clouds: monolithic

- Owned, operated, and controlled by a single provider
- Limits research & innovation
- One-size-fits-none security
- Vendor lock-in

We are in the equivalent of the pre-Internet world, where AOL and CompuServe dominated on-line access

Open Cloud Exchange model: federated

- Multiple providers in a level playing field
- Users control which services they use
- Domain specific intermediaries
  - Enable optimization
  - Provide customers with simple model



The Massachusetts Open Cloud

ADVERTISEMENT

Governor Patrick Announces Funding to Launch Massachusetts Open Cloud Project

Mon, 04/28/2014 - 12:07pm

by Mass Open Cloud Project

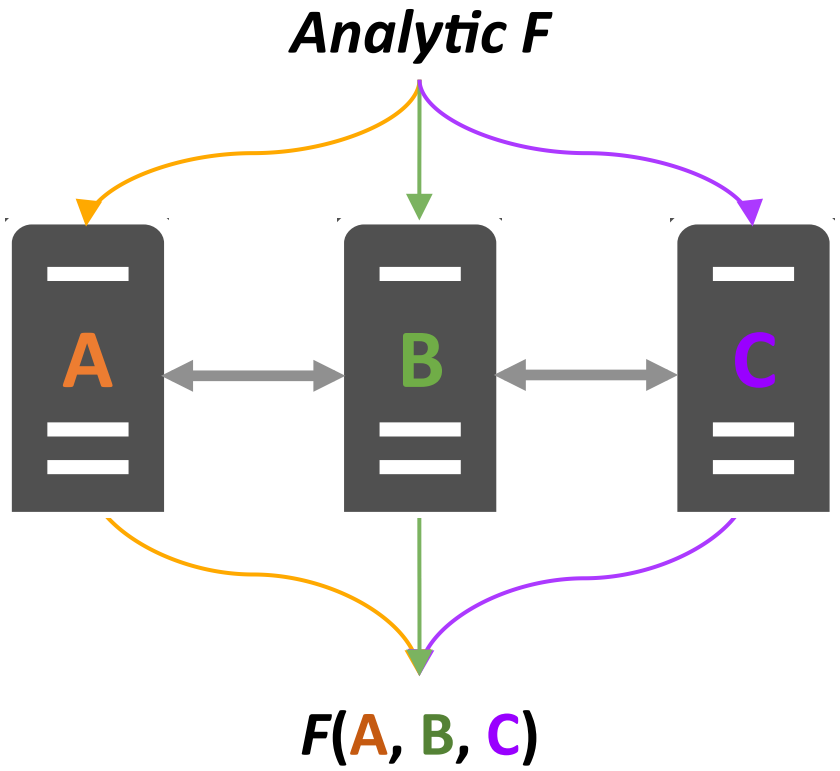
Get the latest news in High Performance Computing, Informatics, Da
more - Sign up now!

The Massachusetts Open Cloud

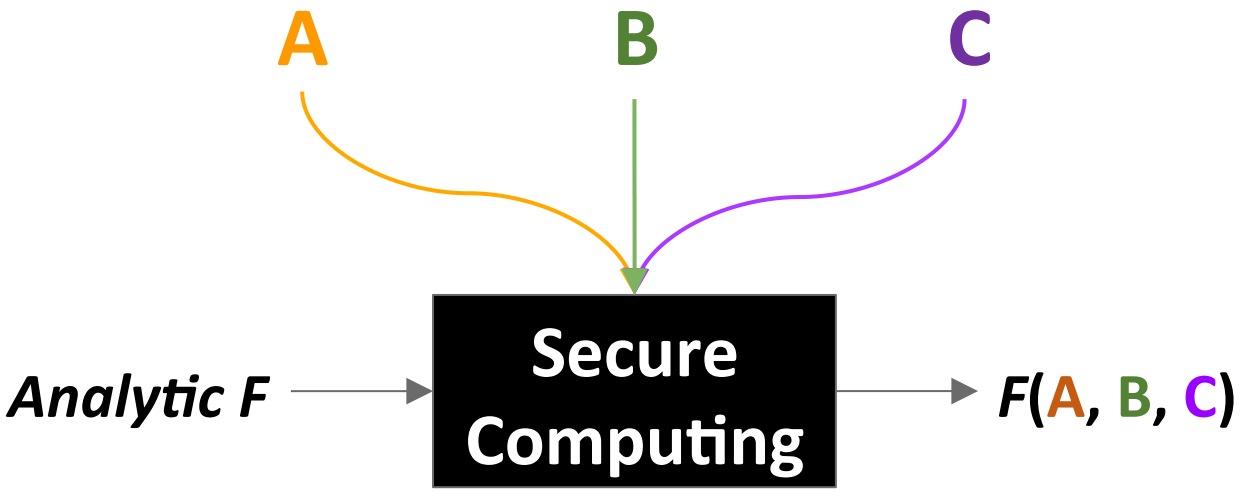
Conclave

Secure Multi-Party Computation

Makes this…



…Look as if it were this



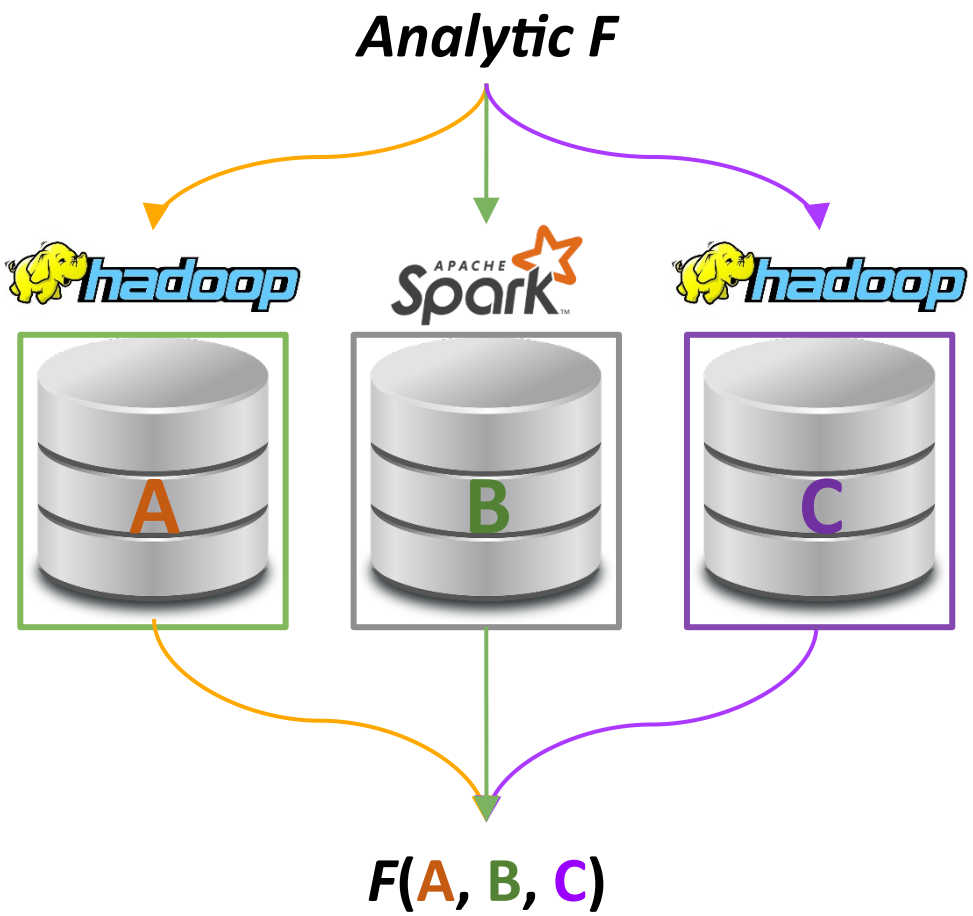
Objective: Federate trust in data & computing among several compute entities

Tradeoff: Gain security at expense of networking

Conclave system for MPC at scale

Dispatcher executes jobs on available backends

⇒ ***No new infrastructure***



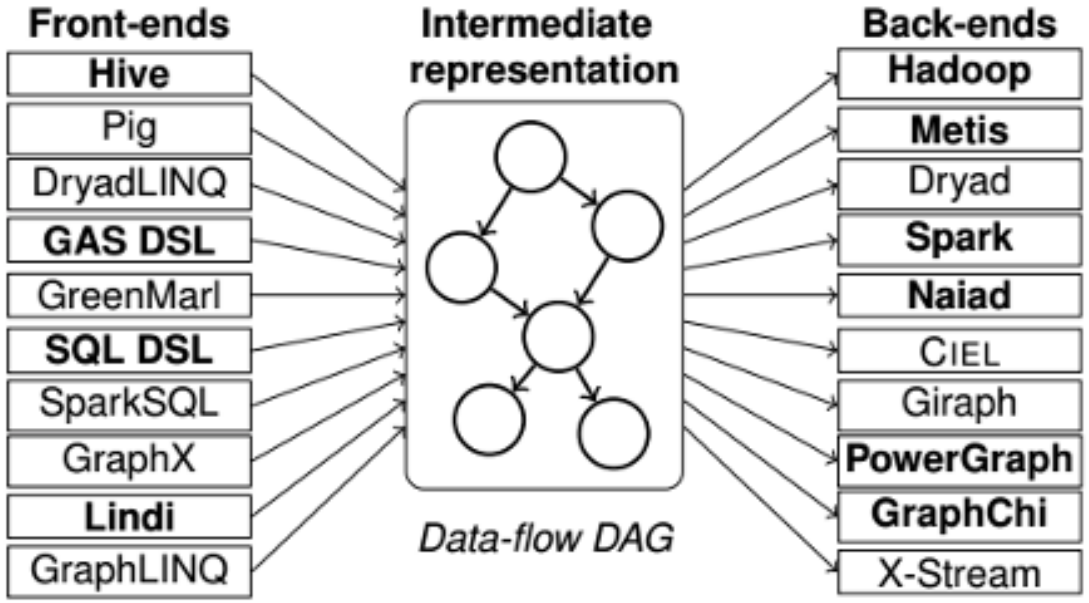
Conclave system for MPC at scale

Dispatcher executes jobs on available backends

⇒ ***No new infrastructure***

SQL-like programming language

⇒ ***No MPC experience needed***



Conclave system for MPC at scale

Dispatcher executes jobs on available backends

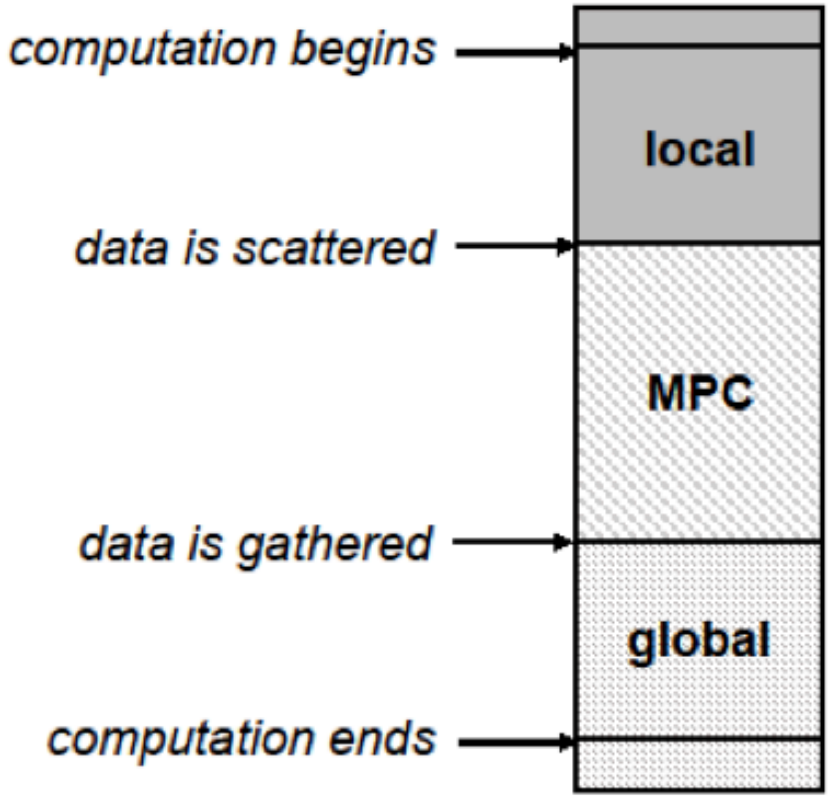
⇒ ***No new infrastructure***

SQL-like programming language

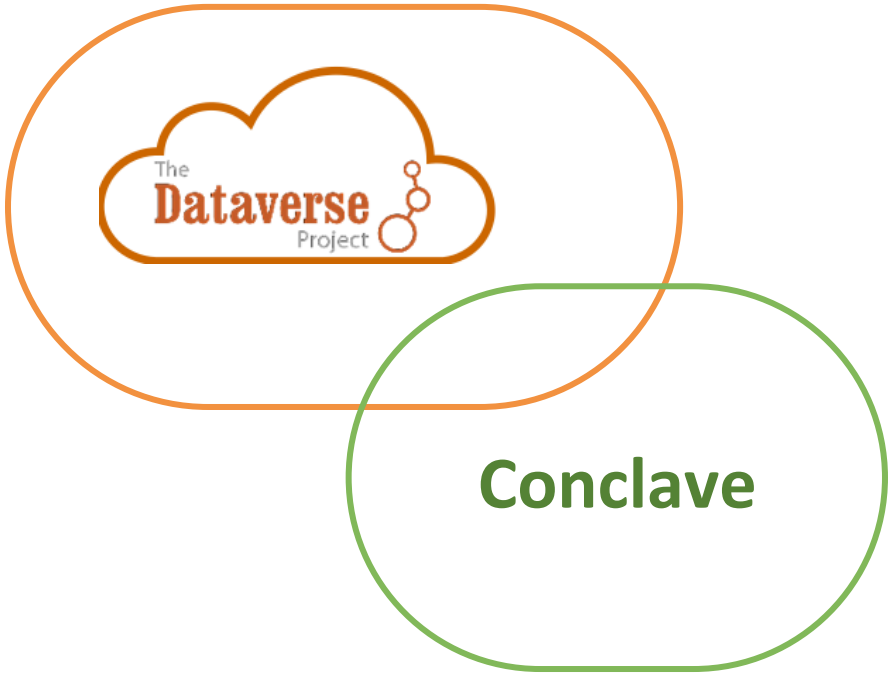
⇒ ***No MPC experience needed***

Static analysis to discern boundaries of secure computing

⇒ ***No need for privacy experts***

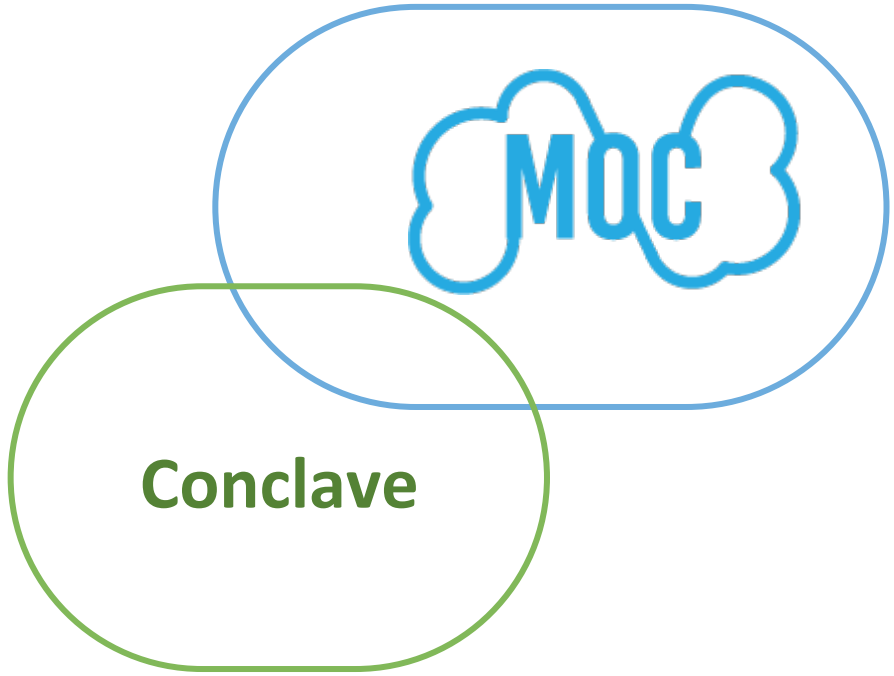


The Synergistic Payoff



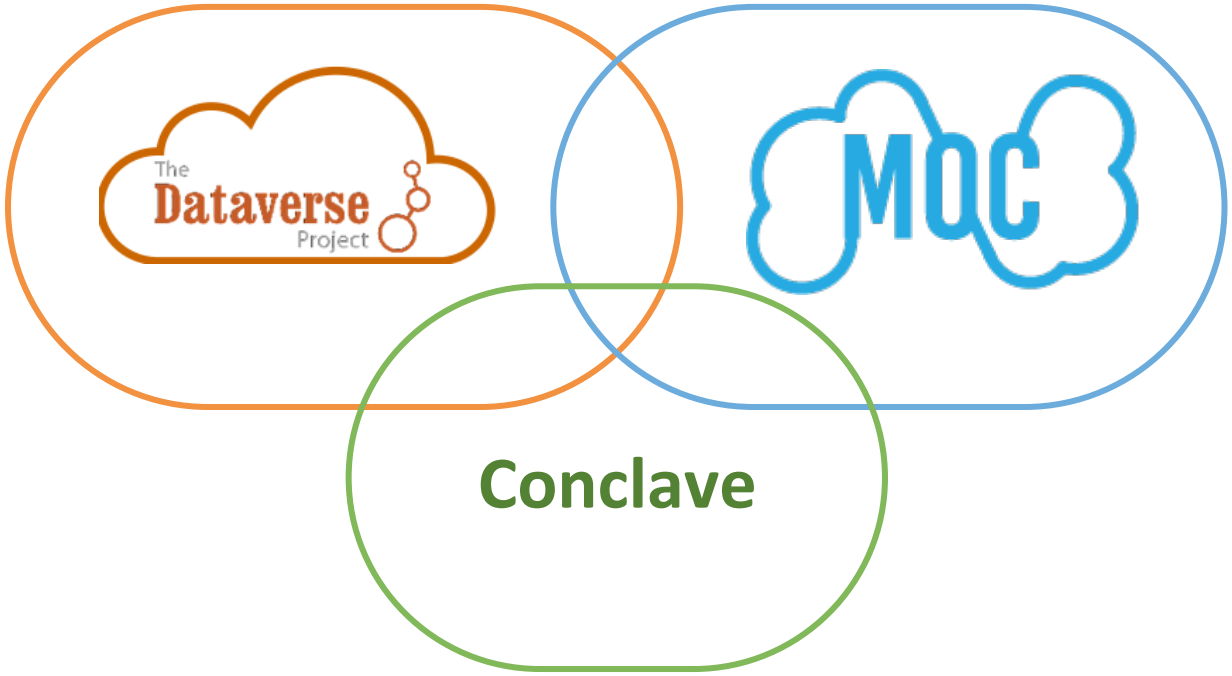
- Bring secure computing technology to the data
- Mechanized tagging controls on derived data

The Synergistic Payoff



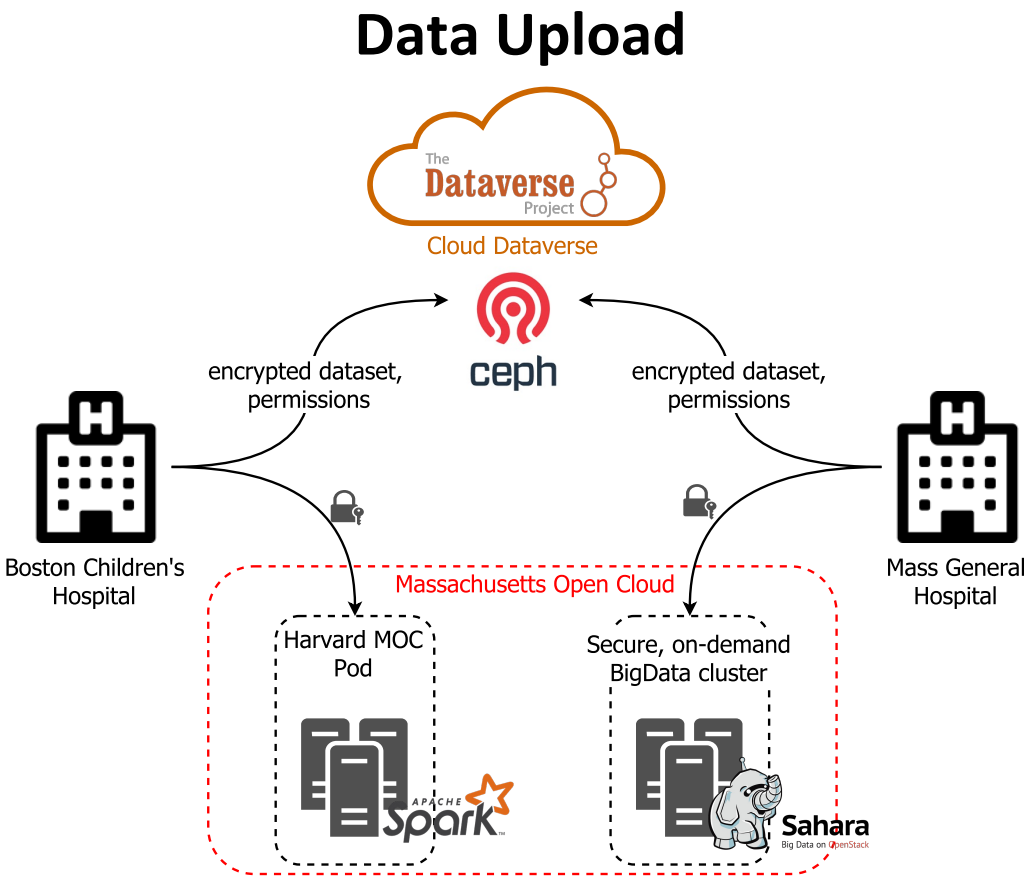
- Reduce trust in the cloud provider
- Low-latency MPC within the datacenter

The Synergistic Payoff

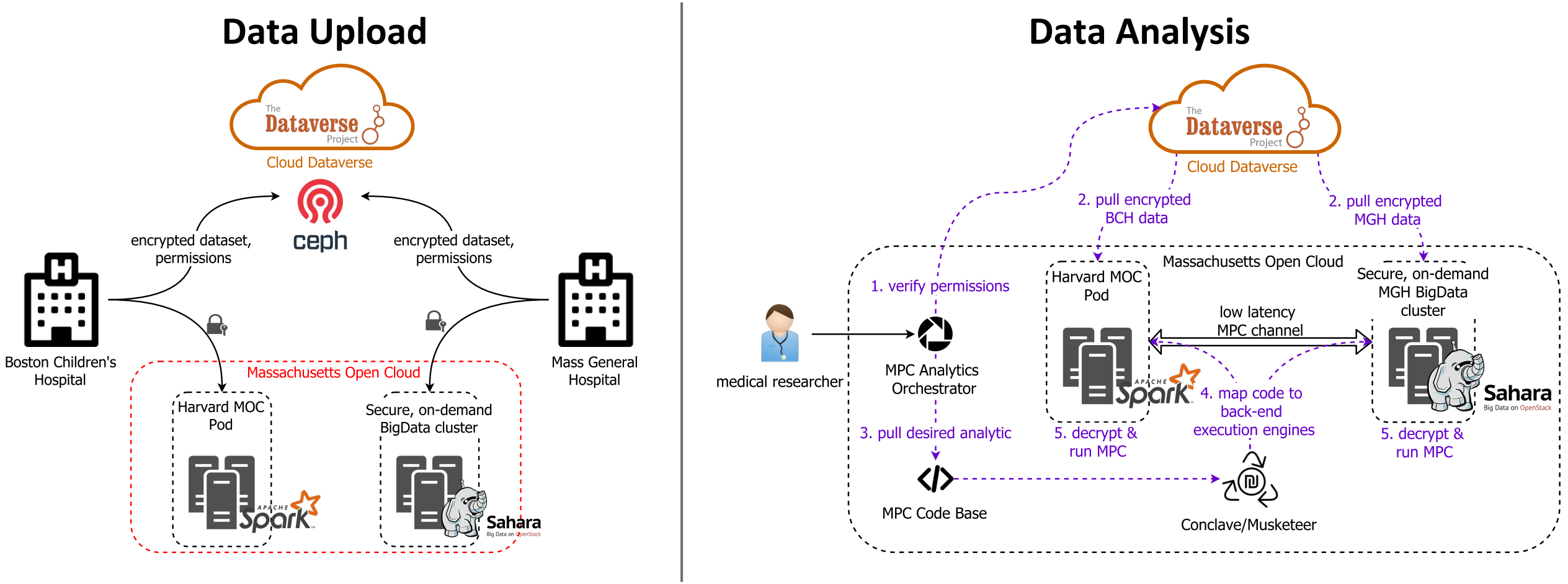


- Separation of responsibilities
- Amortization of effort

Envisioned Workflow



Envisioned Workflow



Thanks!