Domino: A platform for data science

How it works,
How it's built,
Where we're going

DominoDataLab.com

nick@DominoDataLab.com

@DominoDataLab

Domino does three things

1. Provides easy access to cloud compute resources

    Get large, long-running jobs off your computer

2. Provide version control that *actually works* for data science workflows

    Supports large data files

    Tracks artifacts / outputs of your code — reproduce past results
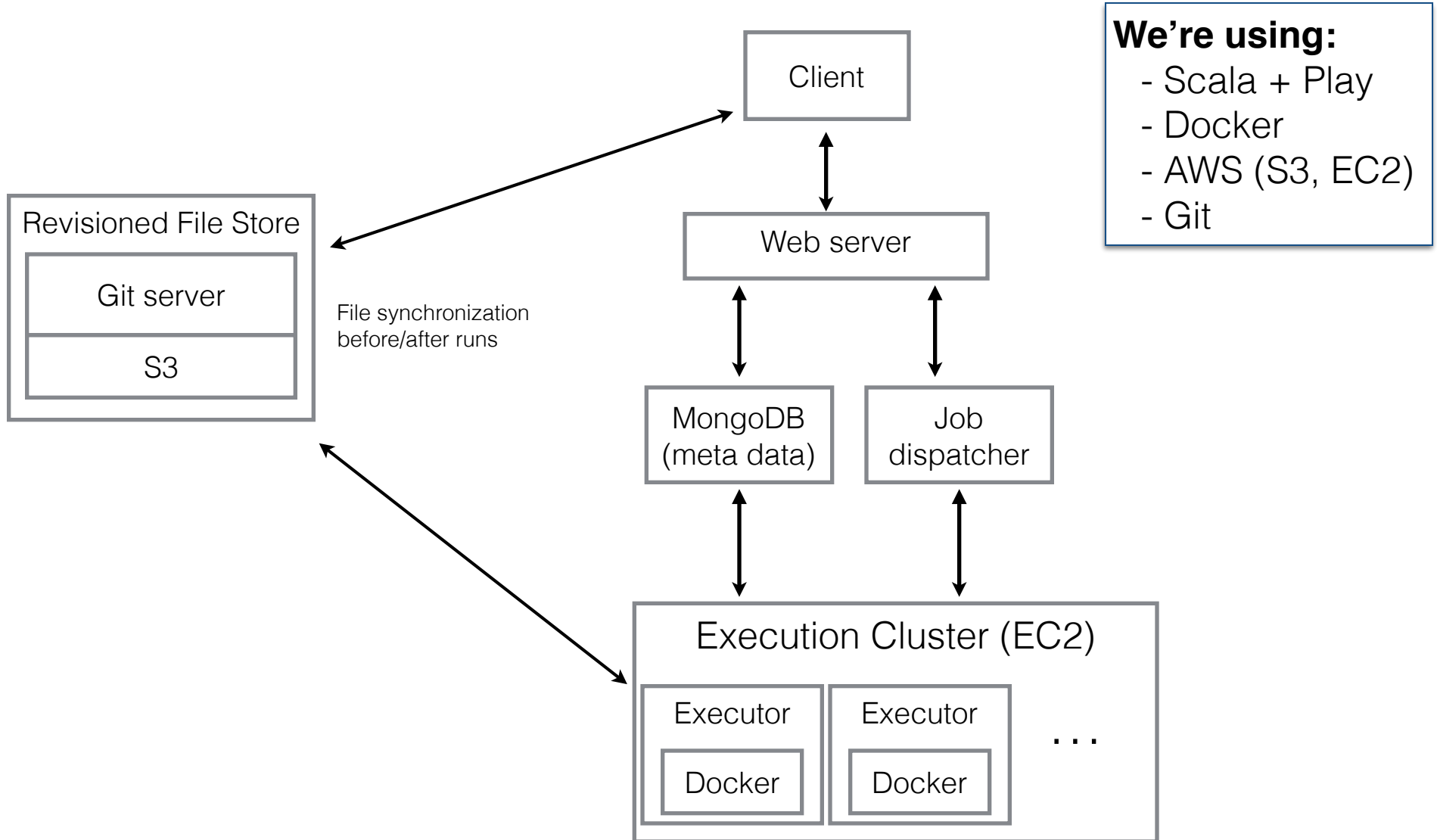
3. Enables collaboration with others

    Share files, get change notifications, etc

Demo

Architecture

Client

Web server

Revisioned File Store

Git server

S3

File synchronization
before/after runs

MongoDB
(meta data)

Job
dispatcher

Execution Cluster (EC2)

Executor

Docker

Executor

Docker

. . .

We're using:
- Scala + Play
- Docker
- AWS (S3, EC2)
- Git

Roadmap

- Better collaboration: Lab notebook, results diffing, branching

- Allow power users to customize machine class

- Easier run parallelization; "parameter sweeps"

- Facilitate rich, interactive visualization of results

We love feedback

Please sign up for our beta and tell us what you think

DominoDataLab.com

nick@DominoDataLab.com

@DominoDataLab