DATA SCIENCE IN PRODUCTION

Scio+ toolchain & our ways of working

SEPTEMBER 5TH 2019

Scio+

MARIA SOLER

AGENDA

1. INTRODUCTION TO SCIO+
2. WE ARE ALL DEVELOPERS
3. THE TOOLBOX

Scio+

WHO AM I

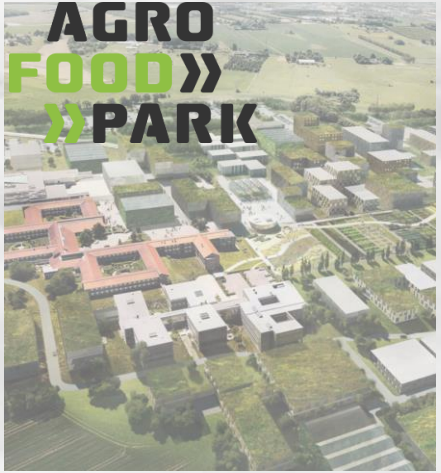
- Born in Barcelona, but lived 13 years in Denmark
- Math, Electrical Engineering, Computer Science, Project Management
- Software Developer, Scrum Master, Continuous Delivery, Team Lead, Project Manager
- Kamstrup 11 years… And now Scio+, 1 year and counting!
- Bonus info: Married a dane and together we have 7 year old twin boys

WHO IS SCIO+

- Founded in 2017. Sister company to SKOV A/S and Big Dutchman in the Big Dutchman group
  - Egg, poultry, pig, climate/ventilation systems, farm management
- Scio+ is the Data Science center of excellence in the group
- Scio+ is primarily focusing on broiler production…for now
- 20 people (7 data scientists + 6 sw developers)

SKOV
Climate for Growth

Big Dutchman

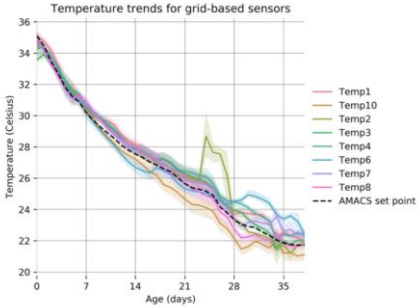


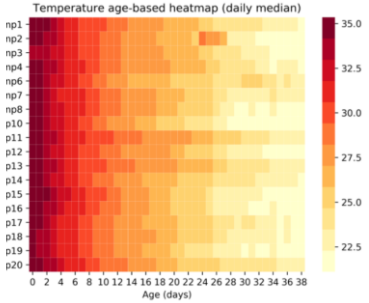
WHAT DATA DO WE HAVE ACCESS TO?

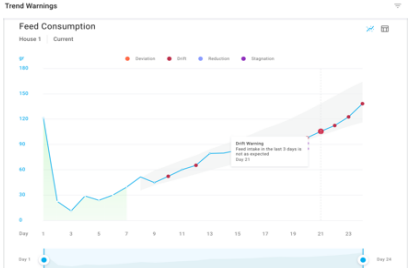
- Information collected from sensors installed in broiler farm houses
  - SKOV sensors already installed in farms
  - Research on new types of sensors that could be relevant
- Other info from broiler production e.g. #birds placed, breed, parent stock, avg. slaughter weight, logistics, hatchery, feedmill/nutrition, …

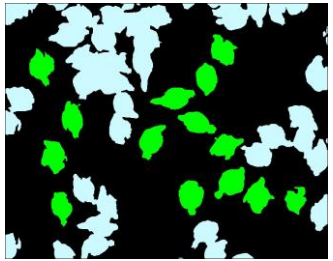
PROJECTS AT SCIO+

- Services for the big dutchman group
  - Pre-studies
  - Slaughter weight prediction
  - Trend-based warnings
- Research-based activities
  - Identify next-generation sensors for farm houses
  - Flock uniformity using broiler images
  - Animal behaviour









RUMOURS SAY…

WE ARE A TEAM

WE ARE ALL PEOPLE

(AKA DEVELOPERS)

Scio+

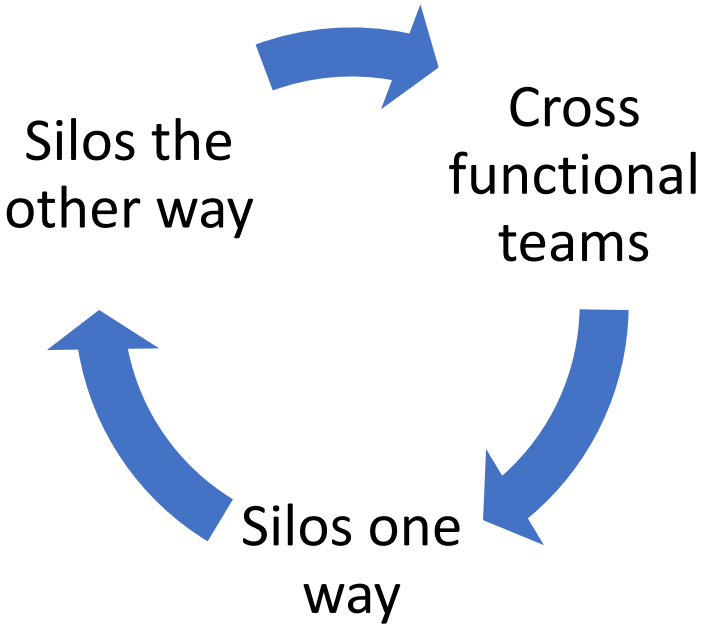
WE ARE ALL DEVELOPERS

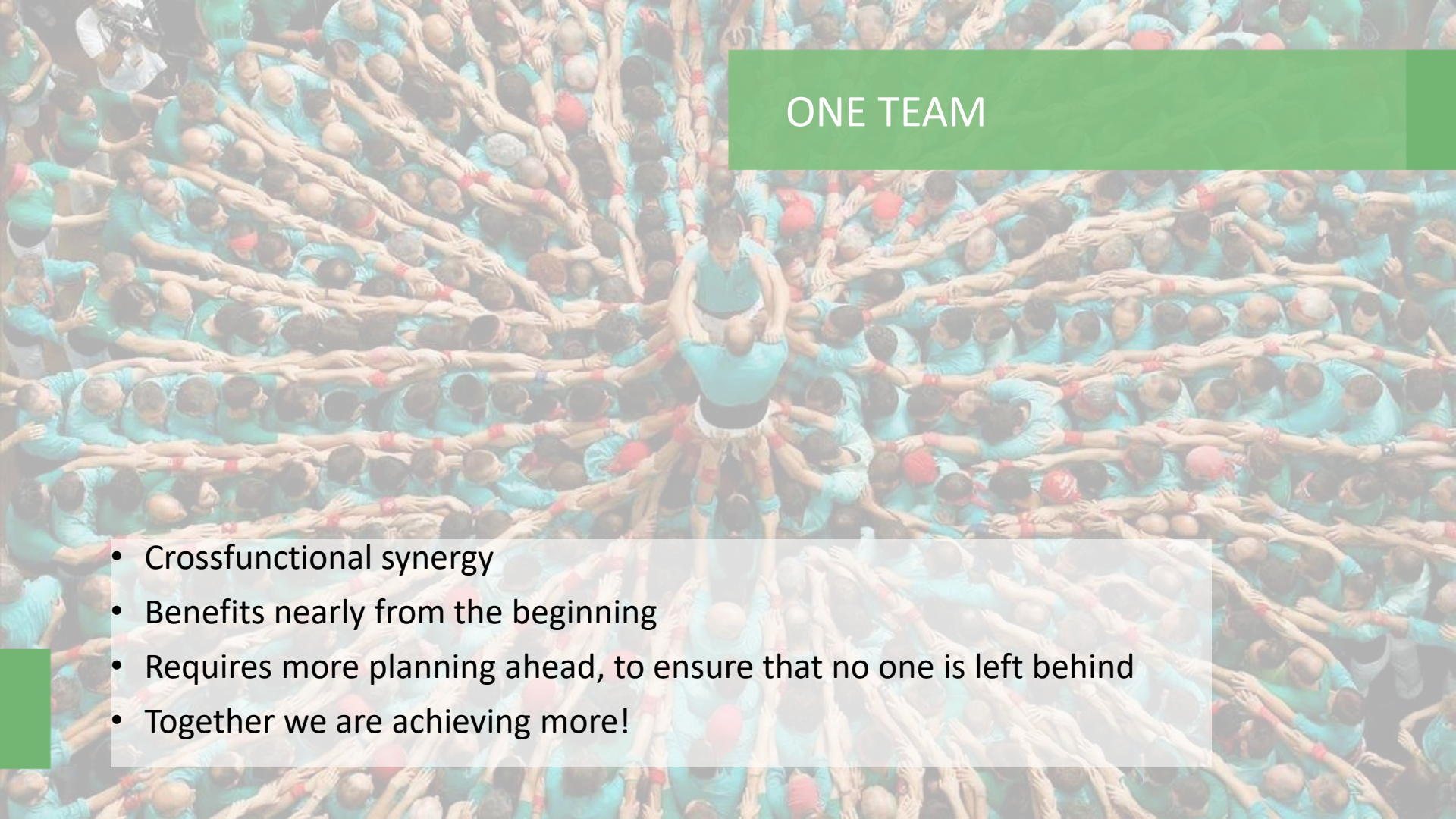
- The development team, as defined by the Scrum Guide:
  - Scrum Development Team consists of professionals who do the
    potentially releasable Increment of “Done” product at
  - The team is self-organizing
  - Development Teams are
    the skills as a team necessary to
    create a produ
  - S
    for Development Team members, regardless of the work
    by the person;
  - rum recognizes no sub-teams in the Development Team
  - Accountability belongs to the Development Team as a whole.

Disclaimer: We don’t use scrum ;-)

WORK IN PROGRESS…

WE'LL FIND OUR WAY





ONE TEAM

- Crossfunctional synergy
- Benefits nearly from the beginning
- Requires more planning ahead, to ensure that no one is left behind
- Together we are achieving more!

THE TOOLBOX

LANGUAGE, TOOLS & INFRASTRUCTURE

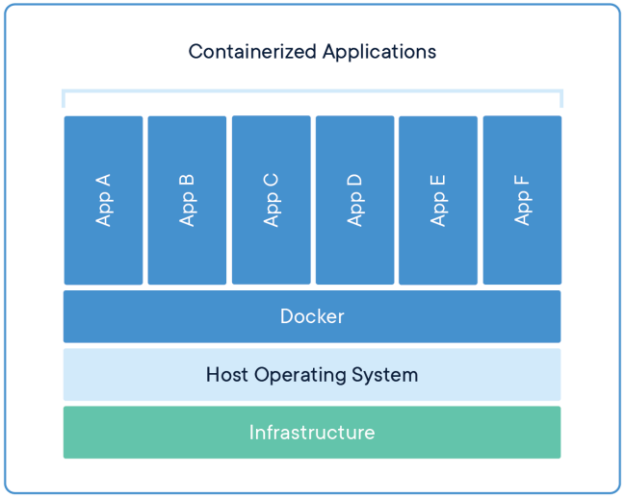
Scio+

R AND PYTHON

- R is a programming language for statistical computing developed in the early 1990's at the University of Auckland
- R's main selling point is the massive amount of libraries allowing you to perform almost any stastictical procedure in one command
- R is great for performaing analysis on a dataset, and presenting findings in a static set of visualization plots
- Python is a widely used high-level programming language for general-purpose programming released in 1991
- Python gains traction within data science …
- Python enables more synergy with software and infrastructure development

DOCKER

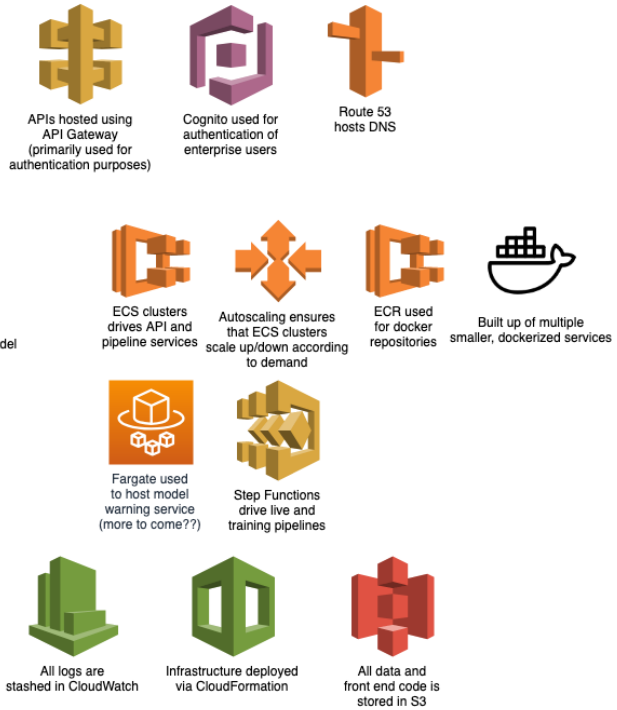
- Models are deployed as REST-enabled web services on AWS - bundled in docker containers
- Docker packages software into standardized units called containers that have everything the model needs to run including libraries, system tools, code, and runtime
- Using Docker, you can quickly deploy and scale applications into any environment and know your code will run in the same way
- Benefit: the software in production is the exact same that is run in development



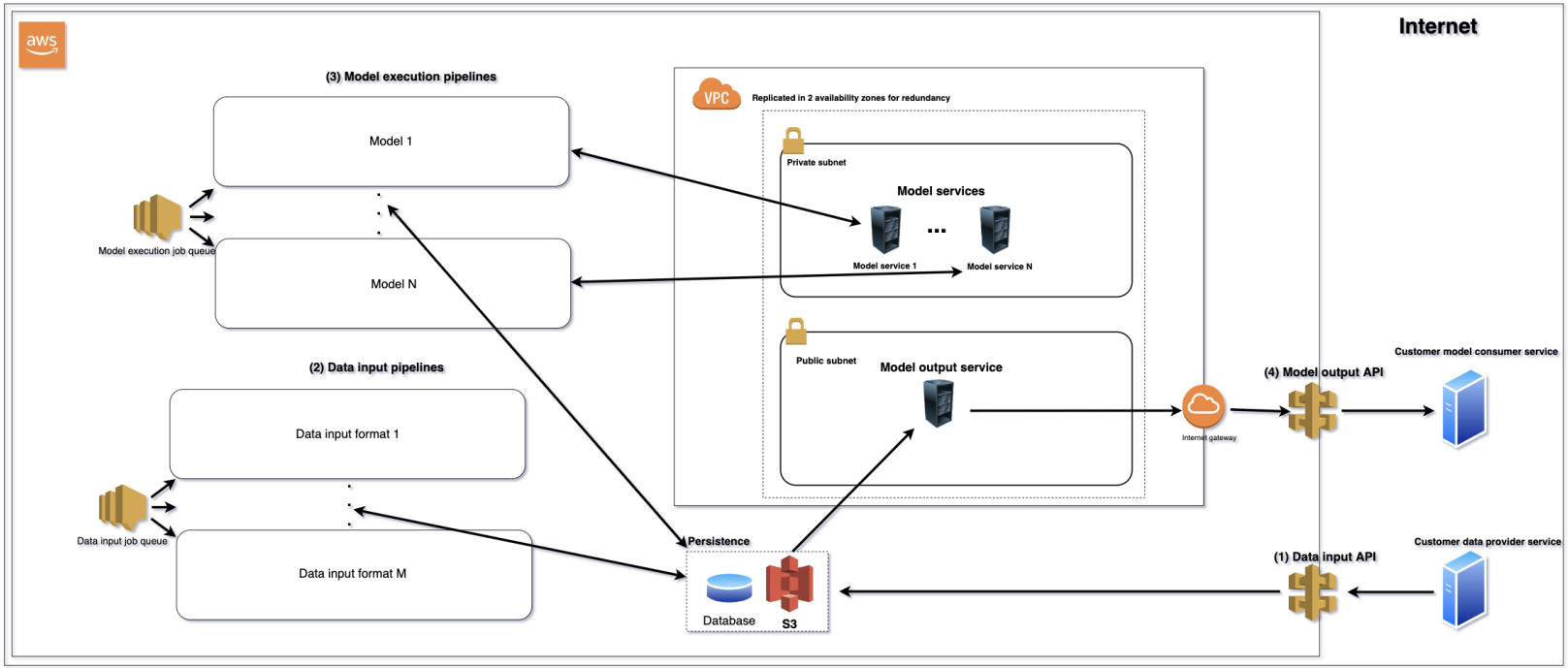
docker

CURRENT

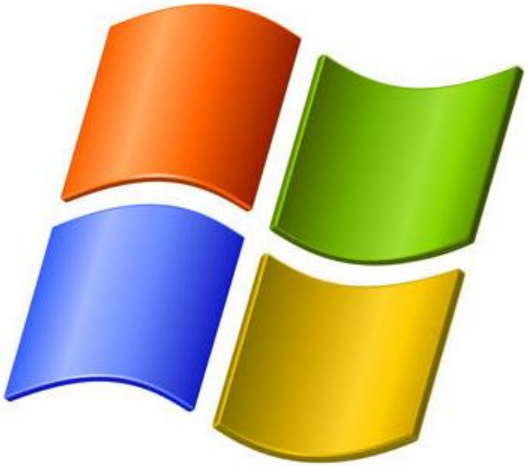
AWS TECHNOLOGY STACK



DECOUPLING MODELS AND INFRASTRUCTURE

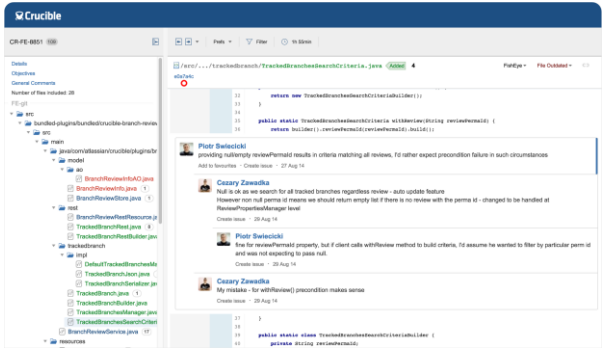


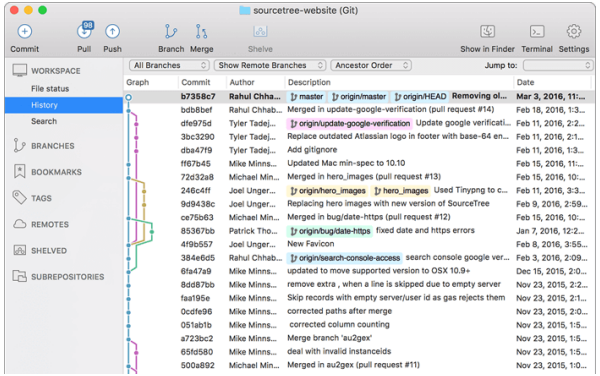
FREEDOM OF CHOICE



VERSION CONTROL

- Code is versioned by Git
- Bitbucket Server used for hosting git repositories and for code review
- Customer datasets and trained models are stored on AWS S3
- Models are versioned !

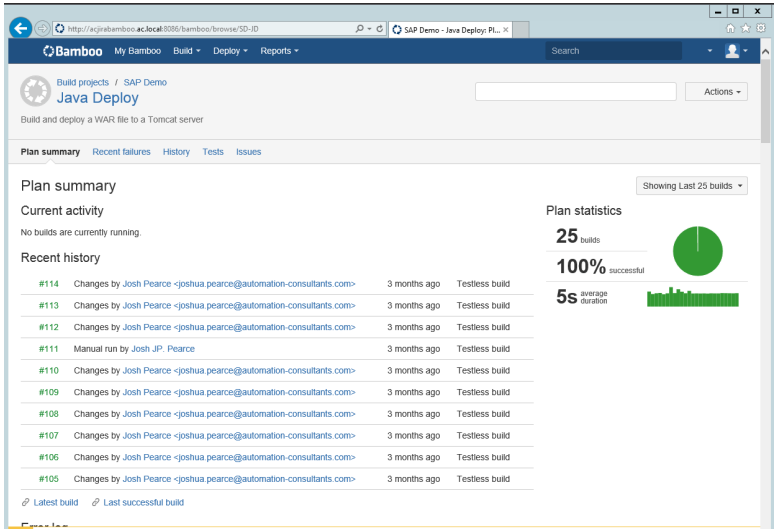




Bitbucket git

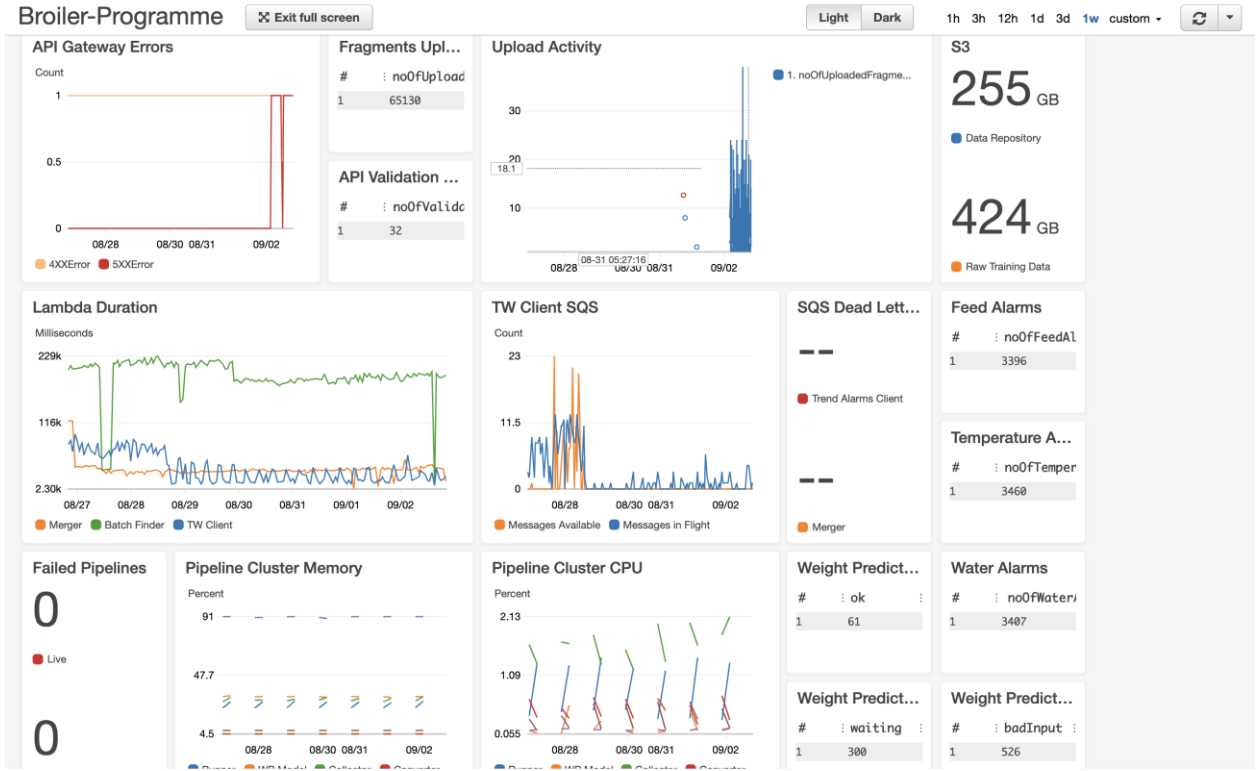
CI & TESTING

- Data science code is production code!
- Using gradle as build tool
- Bamboo as build server
- Staging and production environments
- Unit testing on commits to master and feature branches
- End-to-end testing on staging



gradle Bamboo

MONITORING & LOGGING



Amazon Cloudwatch

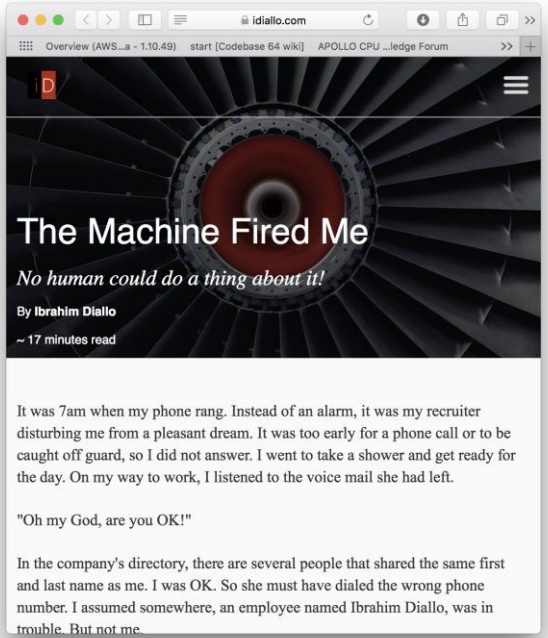
Other options…

Humio

elastic

DEPLOYMENT DOES NOT MEAN THAT YOU ARE DONE

Pictures of the 28 Congress members that were incorrectly matched against a database of mugshot pictures.
Image credit: ACLU



ANY QUESTIONS?

THANKS!

Maria Soler Climent

Team lead @ Scio+

maria.soler@scioplus.com

mariasoler

Scio+