Cyber Analytics Service Constraints and Solutions

Tristan Vanderbruggen

CISC850
Cyber Analytics

Range of Internet Services

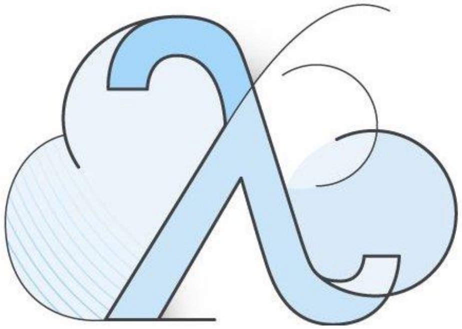
Monolithic

Until early 2000s
Scaling: Larger computer
Not reliable
Weekly update => Debugging Hell

Increased Traffic

Micro-services

Serverless
Scales
Low down time
Emergence



What the really large players do:

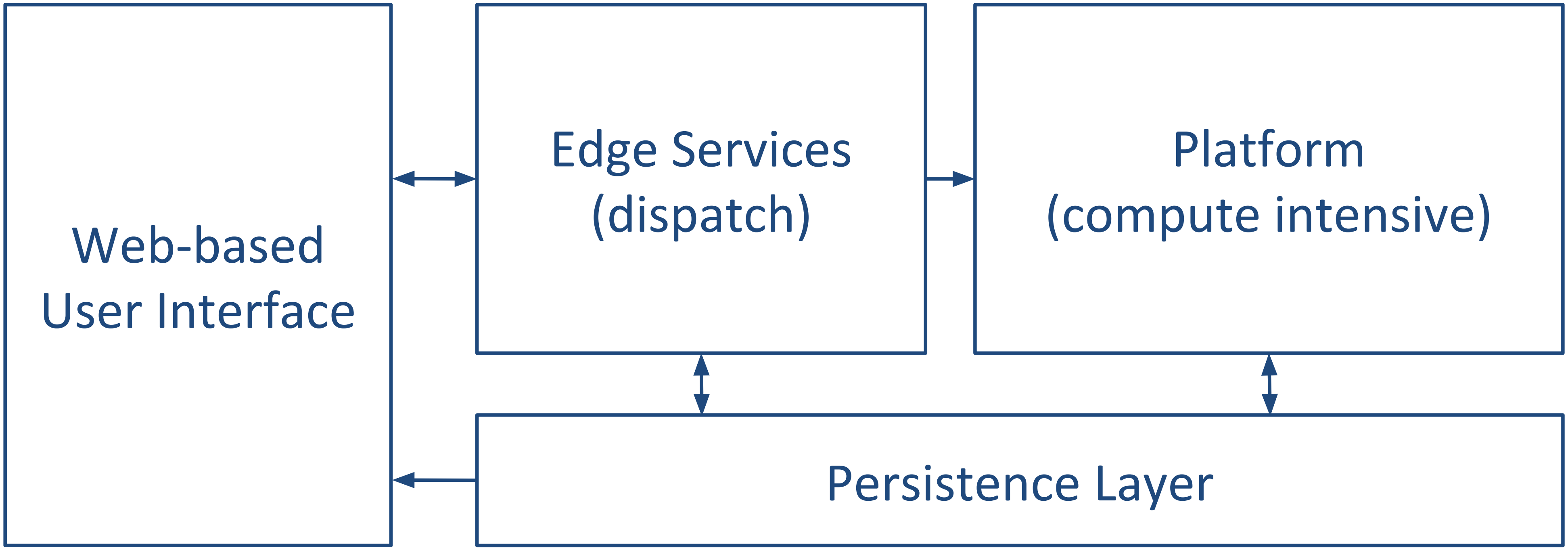
UBER

NETFLIX

What I Wish I Had Known Before Scaling Uber to 1000 Services

Mastering Chaos - A Netflix Guide to Microservices

Middle Ground Solution



WebUI

- *Static*:
  - HTML
  - JavaScript
  - CSS
- *Content*:
  - REST API: **edge services**
  - Media: **persistence layer**
- Short lifecycle

Edge

- Implement: **transaction logic**
    - REST API
- Micro-services
    - Serverless: *AWS Lambda*
    - Lightweight: *AWS Elastic Beanstalk*
        - WSGI application (Flask)
- Short Lifecycle

Platform

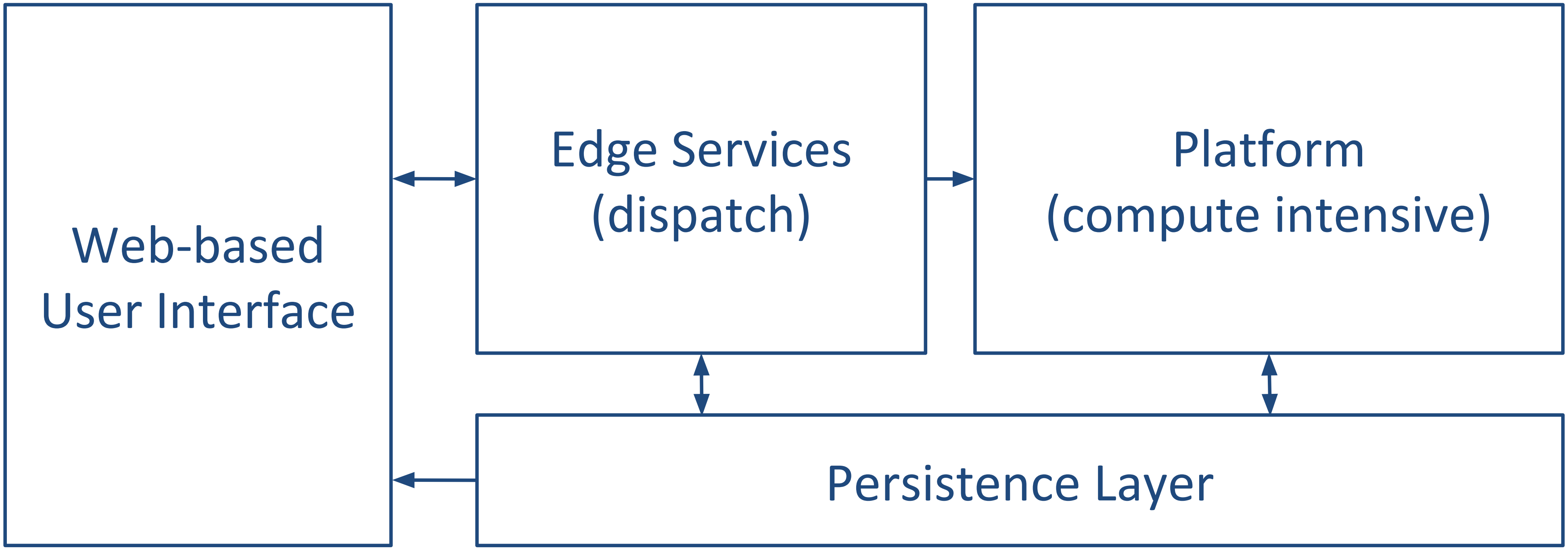
- Compute hungry
    - Actual application
    - Independent tasks
    - Embarrassingly parallel
- Somewhat monolithic
    - Large code base
    - Many dependencies
- Long lifecycle

Persistence Layer

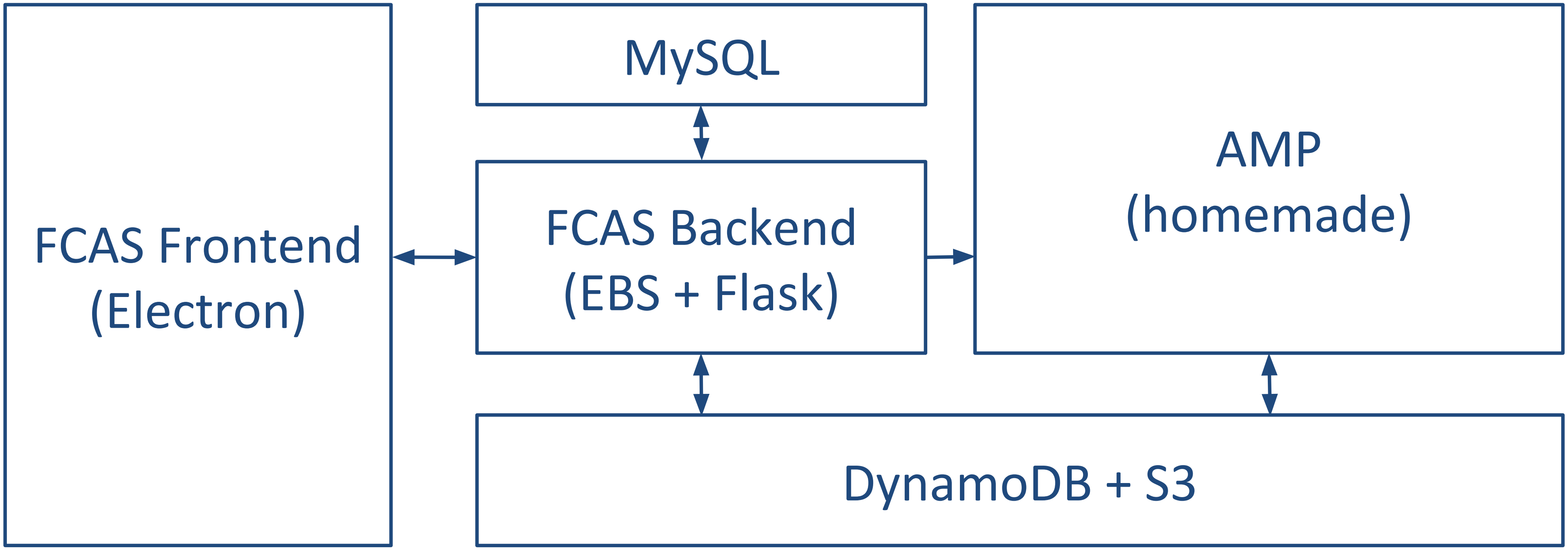
- Your data!
    - Relational Database
    - NoSQL Database
        - Key-value stores
            - Storage: AWS S3
            - Database: AWS DynamDB

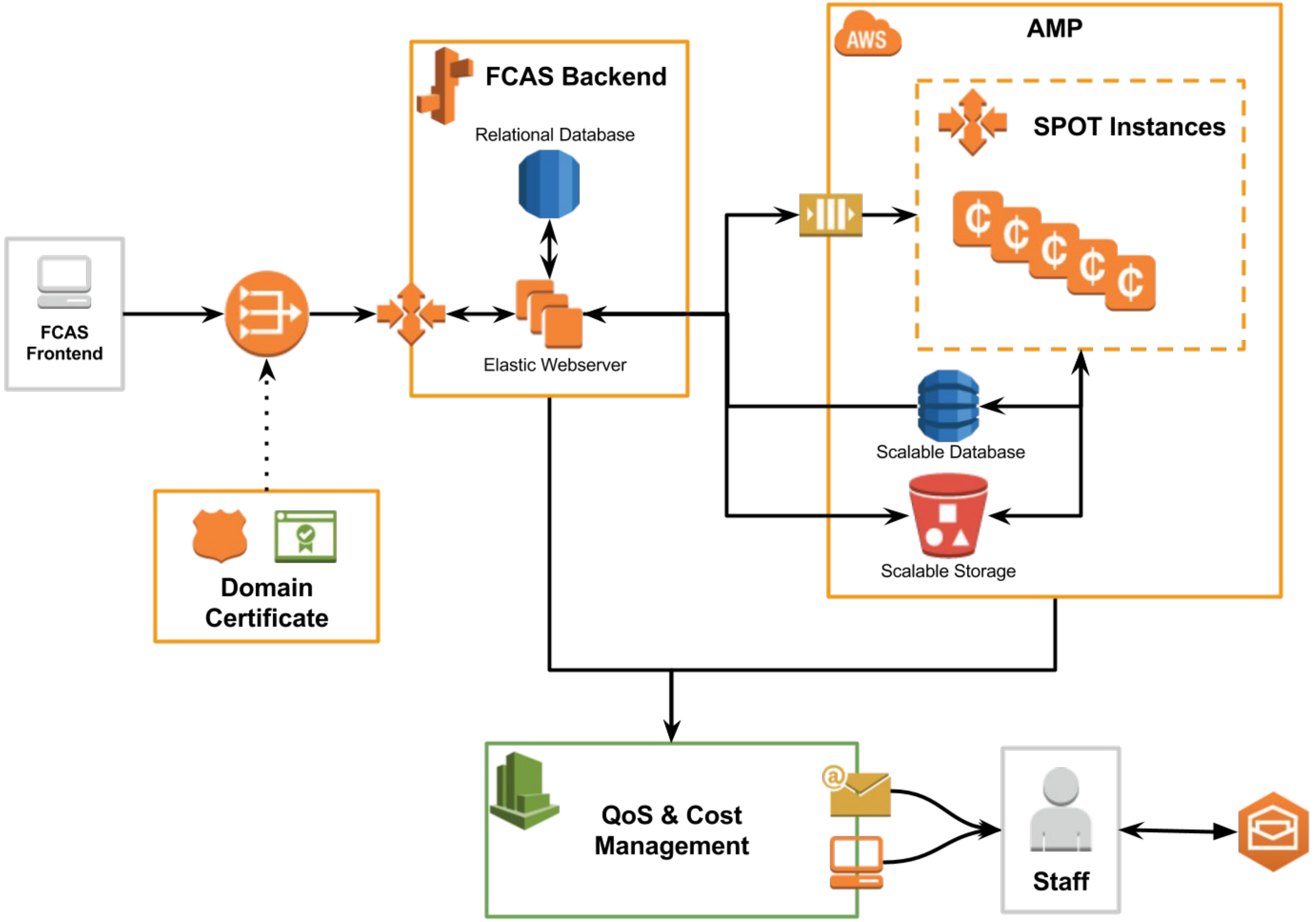
Introduction to NoSQL - Martin Fowler

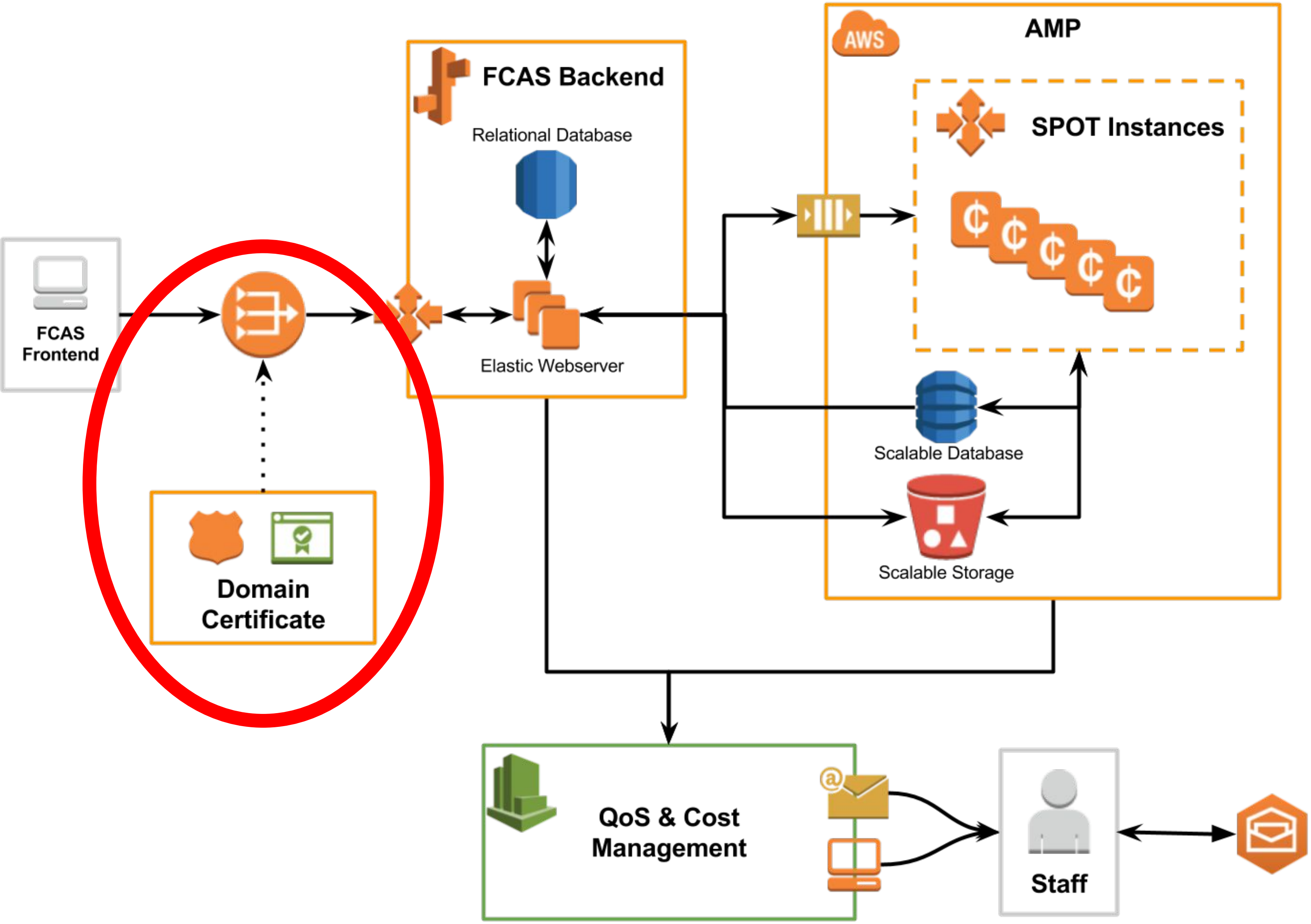
Middle Ground Solution

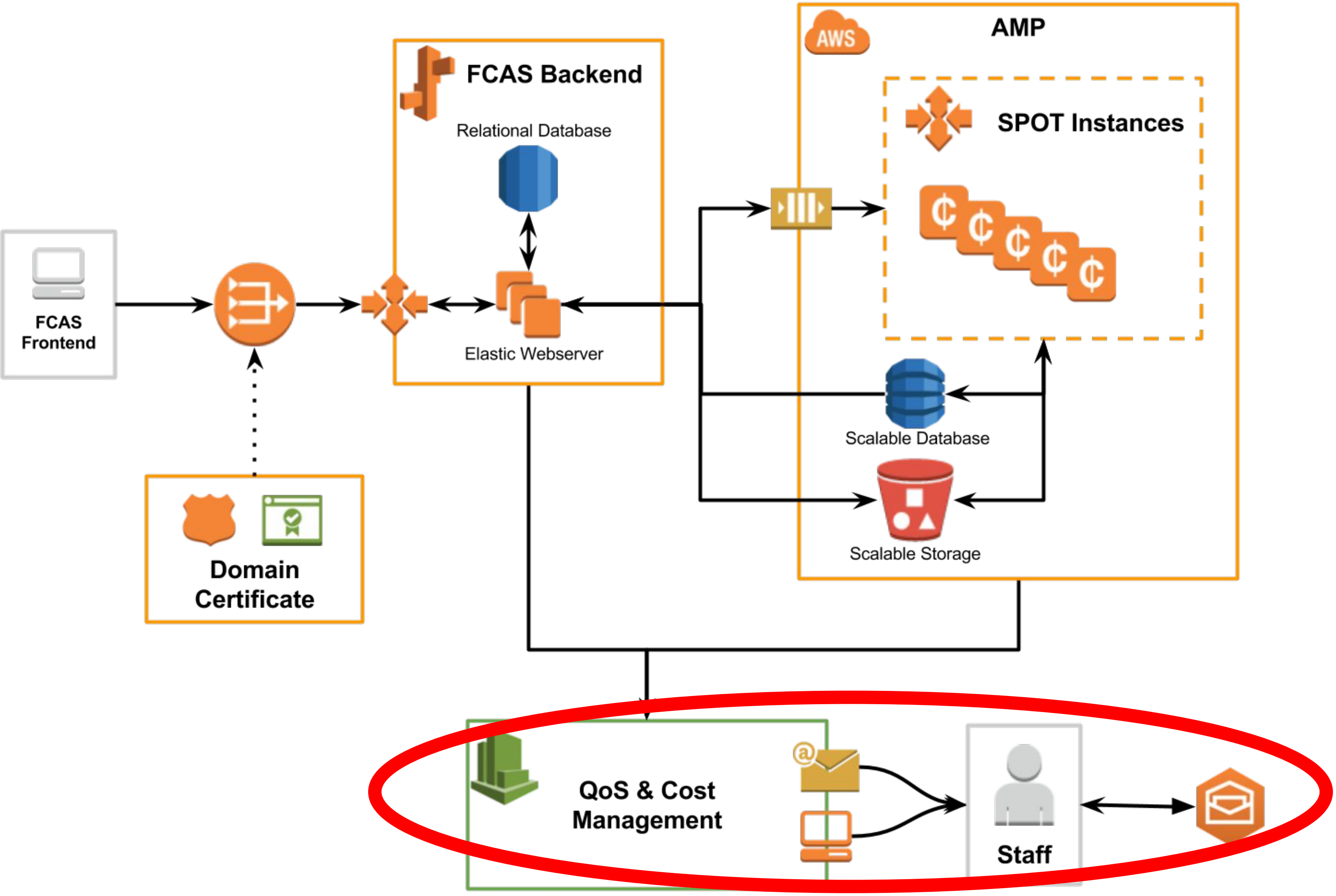


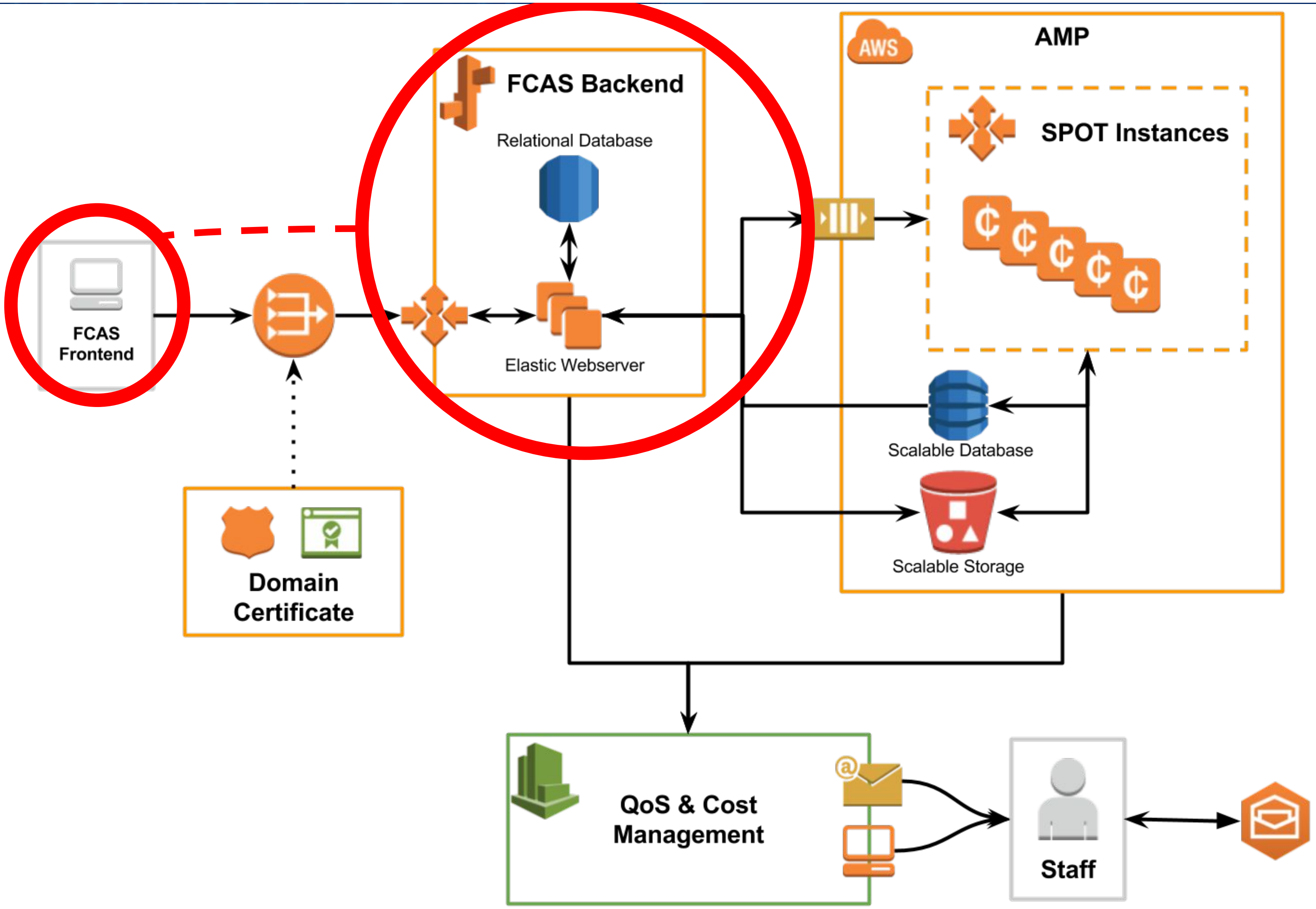
Cyber 20/20 Analytics Service




FCAS Frontend
FCAS Backend
Relational Database
Elastic Webserver
Domain Certificate
AWS
AMP
SPOT Instances
Scalable Database
Scalable Storage
QoS & Cost Management
Staff


AWS
AMP
FCAS Backend
Relational Database
SPOT Instances
FCAS Frontend
Elastic Webserver
Scalable Database
Domain Certificate
Scalable Storage
QoS & Cost Management
Staff


AMP
AWS
FCAS Backend
Relational Database
SPOT Instances
FCAS Frontend
Elastic Webserver
Scalable Database
Scalable Storage
Domain Certificate
QoS & Cost Management
Staff


AWS
AMP
FCAS Backend
Relational Database
Elastic Webserver
SPOT Instances
Scalable Database
Scalable Storage
FCAS Frontend
Domain Certificate
QoS & Cost Management
Staff

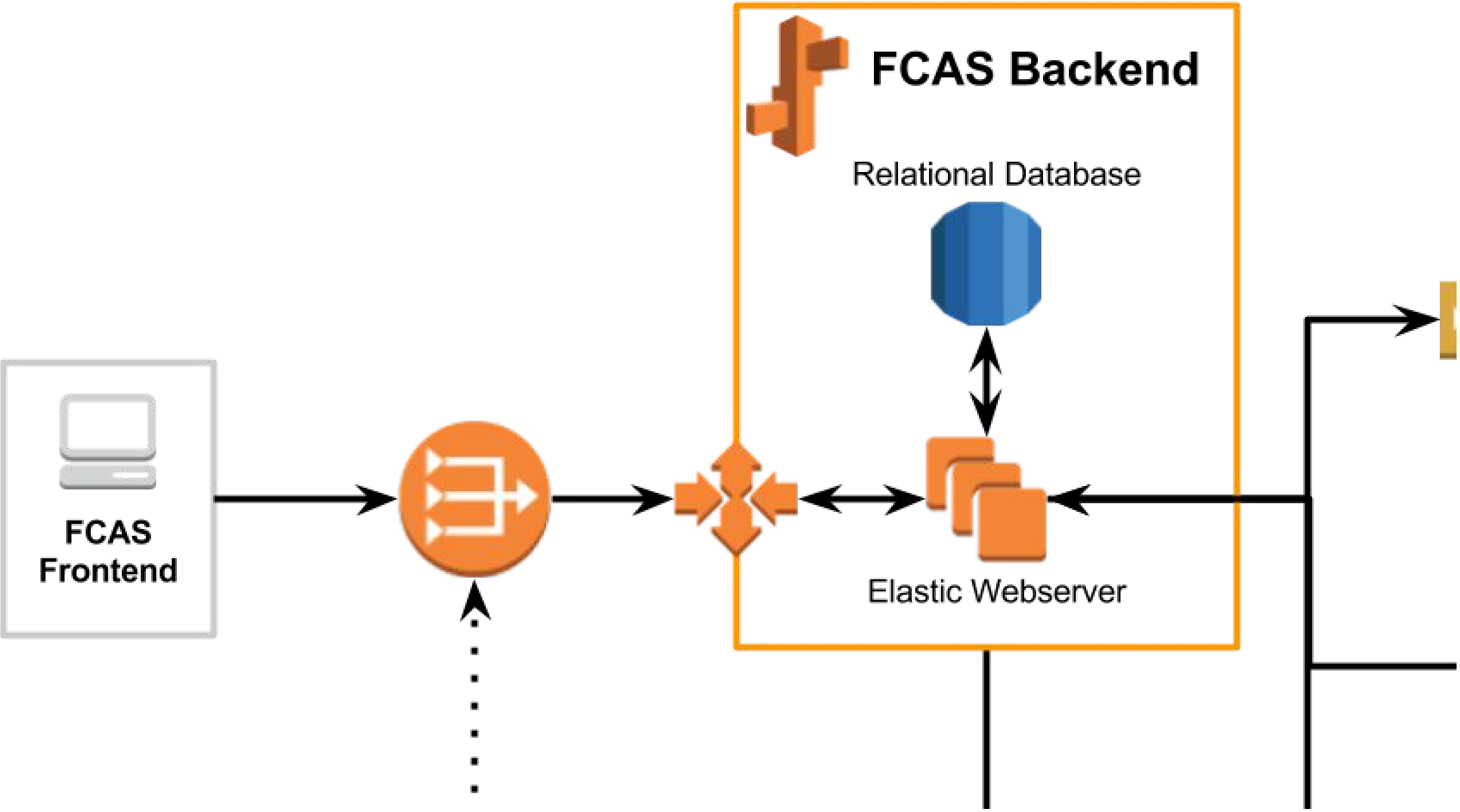
File Capture and Analysis Service

- Tightly coupled frontend and backend
  - Web-based UI: Electron
    - Presents analysis and prediction results
    - Use D3 to provide visual insights
  - REST server: Flask + MySQL
    - Dispatch analysis and prediction workload
    - Gather results in relational DB

FCAS Frontend

- Constraints
  - Controlled Environment
  - Visually Appealing

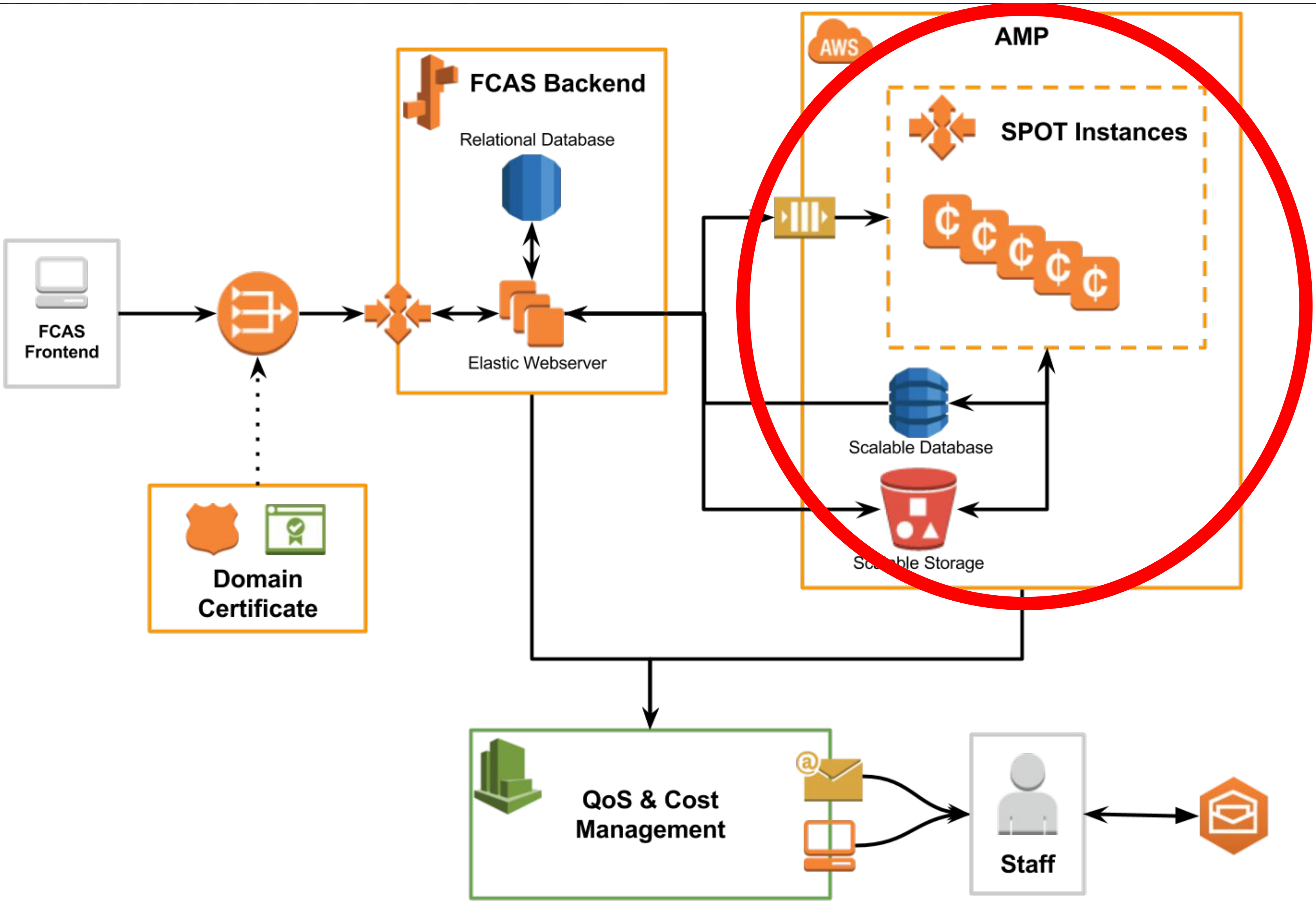
- Solutions
  - Electron
  - D3js


FCAS Backend
Relational Database
Elastic Webserver
FCAS Frontend

FCAS Backend

- Constraints
  - Deploy and Scale
  - Complex Queries
  - Short Lifecycle

- Solutions
  - Elastic Beanstalk
  - Relational Database
  - Python + Flask


AMP
AWS
FCAS Backend
Relational Database
Elastic Webserver
SPOT Instances
FCAS Frontend
Domain Certificate
Scalable Database
Scalable Storage
QoS & Cost Management
Staff

Analysis and Machine Learning Platform

- Analyses files
  - Basic: crypto hash, strings, PE, …
  - Bytes-Entropy Histograms
  - Reverse Engineering with Radare2
- Make predictions
  - DNN applied to various analysis results
  - DNN ensemble for consensus

=> Lots of dependencies <=

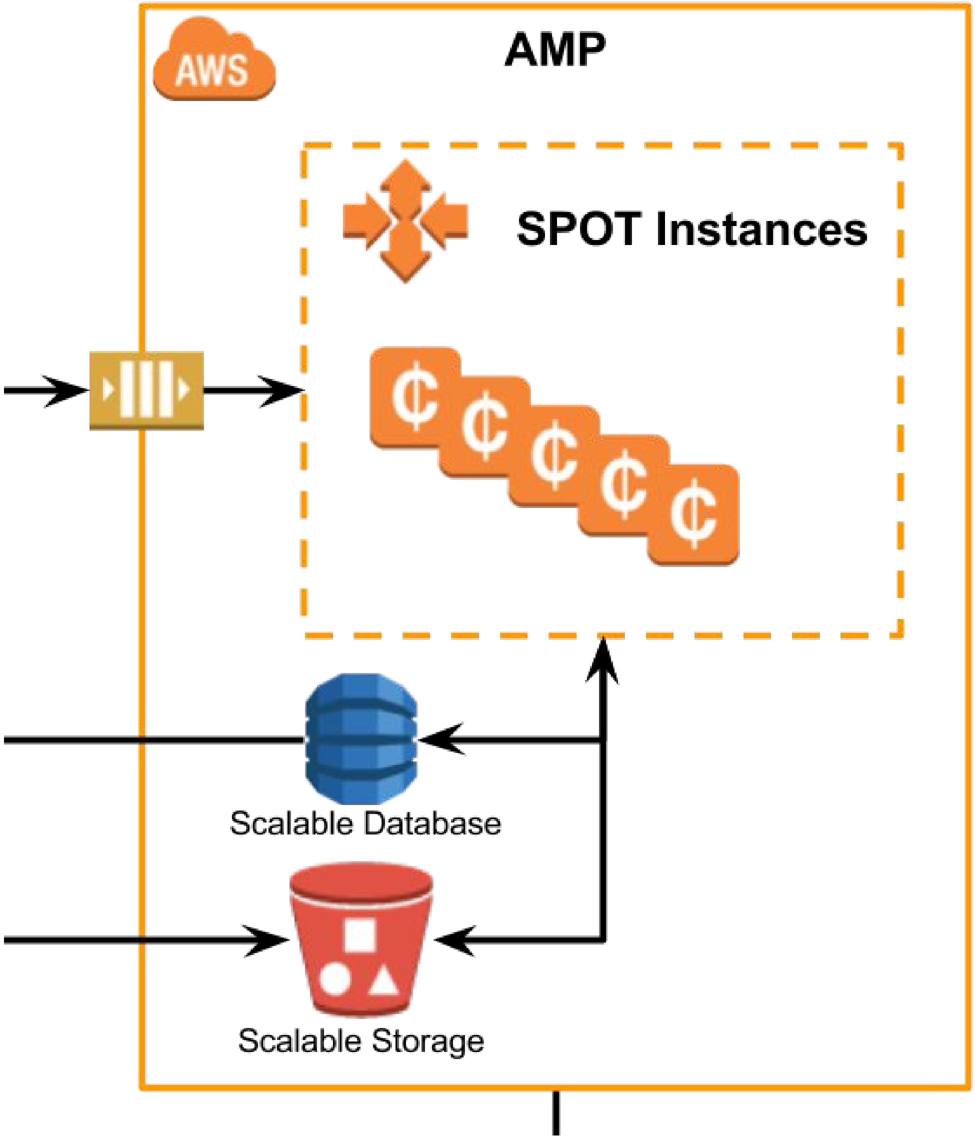
Analysis and Machine Learning Platform

- Analysis Tools
  - Independent
  - Lots of dependencies (Radare2, ssdeep, pefile, scipy, ...)
- Machine Learning (Theano + Scikit Learn)
  - Handle big data (training)
  - Fast inception (predictions)
- Glue code
  - receive workload
  - dispatch to subprocesses

Analysis and Machine Learning Platform

- Constraints
  - Highly scalable
  - Cheap
  - Reliable
  - Low latency

- Solutions
  - ASG + S3 + DynamoDB
  - SPOT instances
  - Simple Queue Service
  - Hard work !!!


AWS
AMP
SPOT Instances
Scalable Database
Scalable Storage