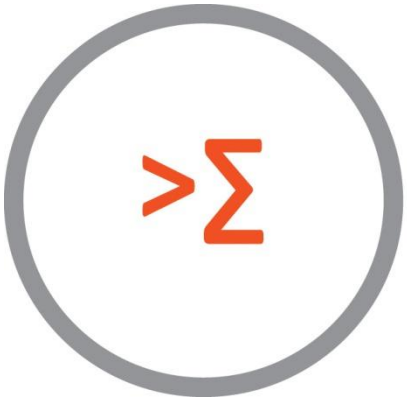
SC16 Presentation



ARISTOTLE
CLOUD FEDERATION

PI David Lifka, Cornell University, lifka@cornell.edu
Co-PI Tom Furlani, U. at Buffalo, furlani@buffalo.edu
Co-PI Rich Wolski , UC Santa Barbara, rich@cs.ucsb.edu

https://federatedcloud.org

Award No. 1541215

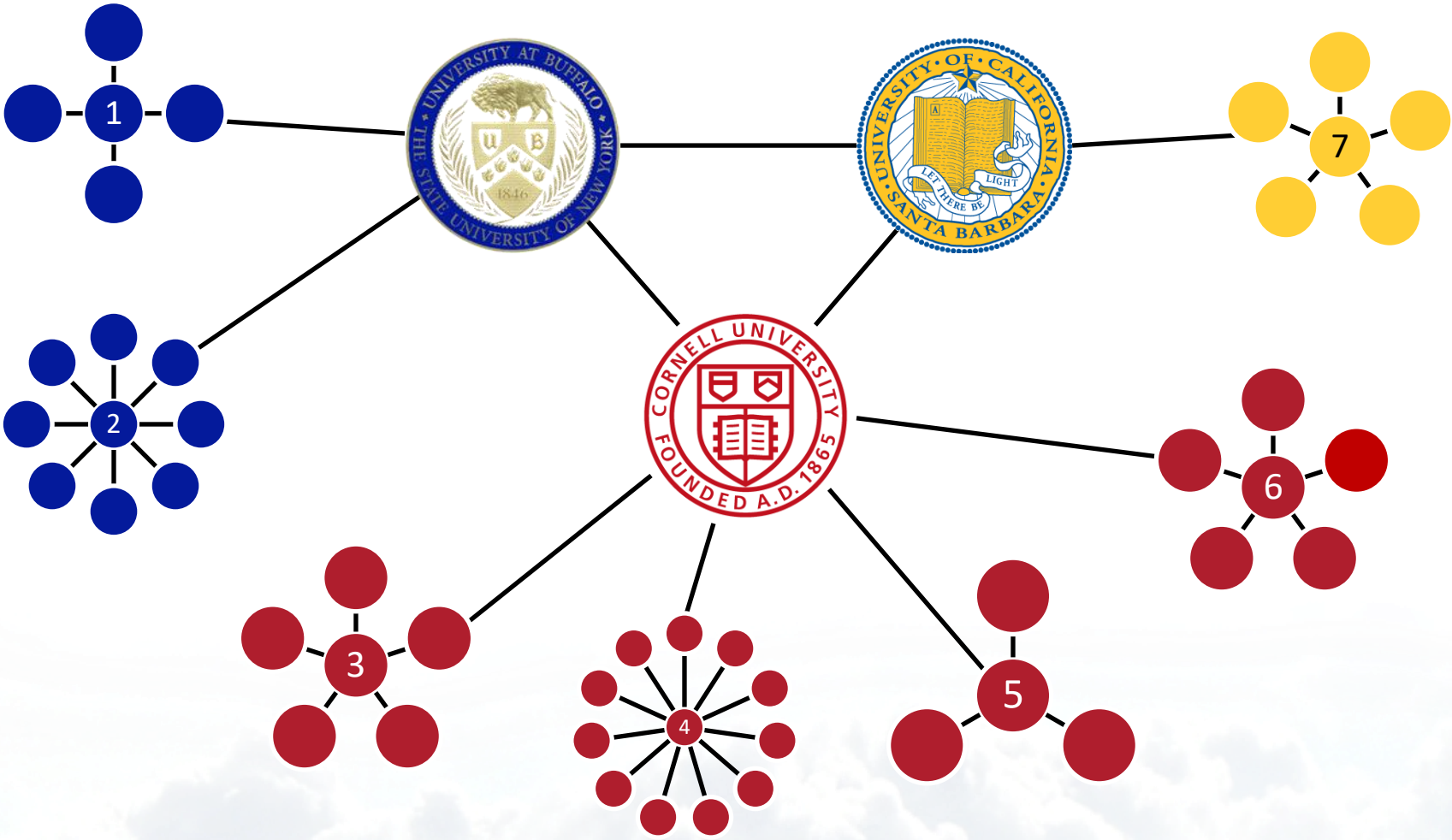
Aristotle Overview & Goals

- $5 million, 5-year (2015-2020) NSF award to deploy a federated DIBBs cloud at Cornell, U. Buffalo, UC Santa Barbara
  - 7 science teams, over 40 global collaborators
    - Requiring flexible workflows/analysis tools for large-scale data
    - Representing diversity of analysis requirements and cloud usage modalities
      - Earth and atmospheric sciences, finance, chemistry, astronomy, civil engineering, genomics, food science
- Overarching project goals
  - Optimize "time to science"
  - Demonstrate the value of sharing resources and data across institutional boundaries

Award No. 1541215

ARISTOTLE CLOUD FEDERATION

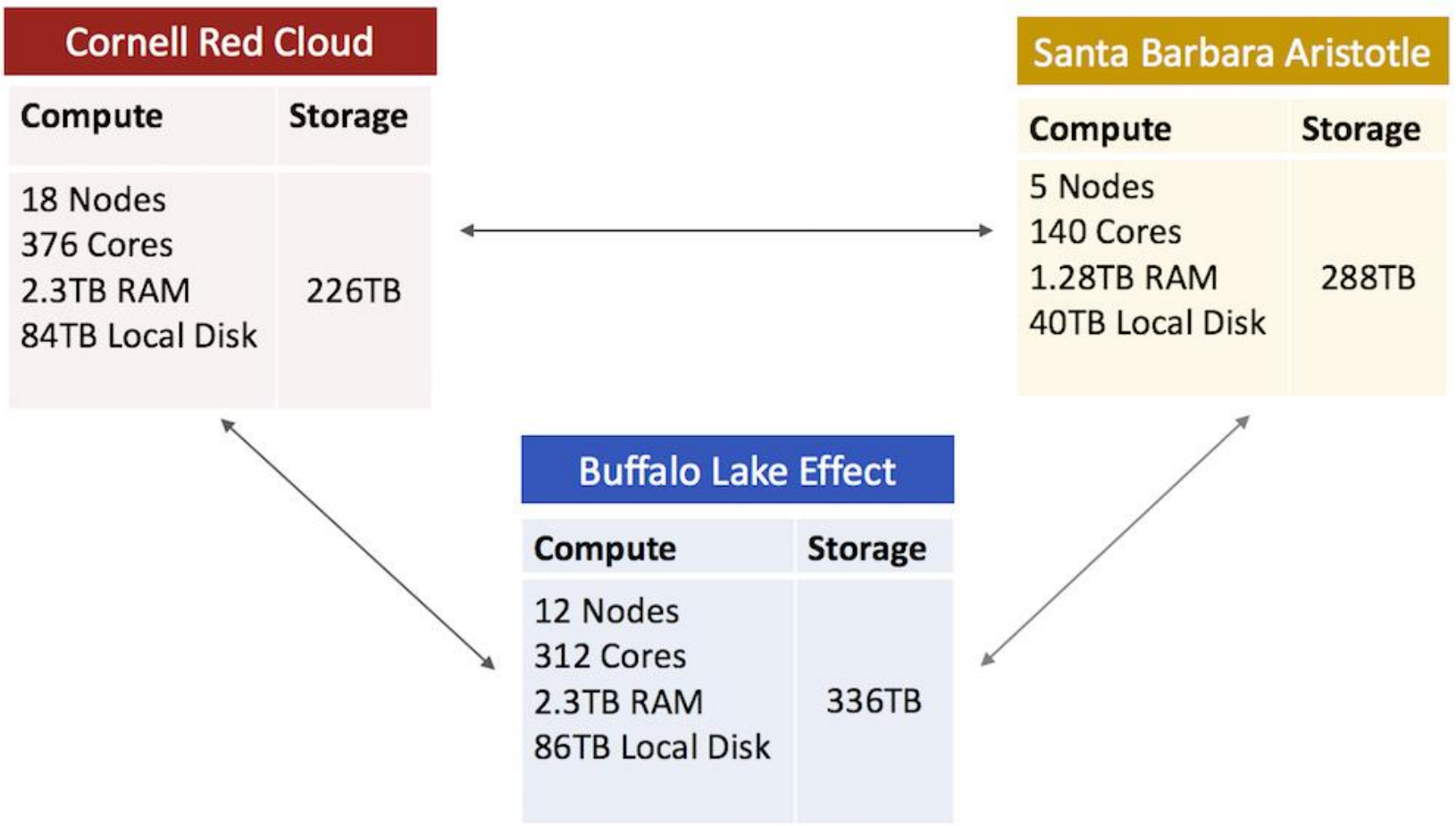
Multi-Disciplinary Collaborations in a Multi-Campus Cloud Federation



Award No. 1541215

ARISTOTLE CLOUD FEDERATION

Year 1 Infrastructure



Award No. 1541215

Year 2 Plans

- Infrastructure and Portal
  - Install 2nd year storage assets at each site
  - Implement OAuth 2.0 support for single sign-in (working with HPE Helion Eucalyptus team)
  - Continue to share resources and transition from local to federated accounting and allocations system
  - Update user documentation to be federation-specific and reflect new software, e.g., MATLAB MDCS
  - Develop how-to guide for sites who wish to develop similar systems
  - Hold local training and webinars for remote researchers
- Metrics and Usage
  - Implement working version of Federated Open XDMoD (see Federated XDMoD Requirements and Job Reporting for Cloud)
  - Implement cloud metrics in XDMoD which will require re-engineering the XDMoD data warehouse
  - Refactor QBETS and DrAFTs analytic web services
    - QBETs will make on-line forecasts of future performance and cost levels available to Aristotle.
    - DrAFTS (Durability Agreement from Time Series) developed by Wolski under the Aristotle project predicts "bid price" an AWS user should bid in the spot market to ensure minimum duration of execution before AWS terminates the instance
  - Continue testing Cornell CS Supercloud with other sites (see Supercloud demo migrating VM across 4 platforms: Aristotle (Cornell Red Cloud), AWS, Microsoft Azure, and Google Compute Engine)

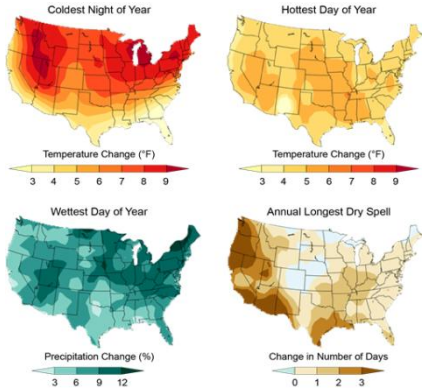
Award No. 1541215

Use Case: A Cloud-Based Framework for Visualization and Analysis of Big Geospatial Data

Professor Varun Chandola
Computer Science and Eng. Dept., U. at Buffalo, chandola@buffalo.edu
with R. Vatsavai (NC State), P. Hogan (NASA Ames), B.B. Bhaduri (ORNL)

Problem

- Understand the impact of climate change from climate simulation outputs
  - E.g., understand anomalous changes in future climate projections
- Challenges
  - How to identify changes?
  - How to scale? (single 200 yr. run can result in >2TB data)



Projected changes in hottest/coldest and wettest/driest day of the year (Source: EPA)

Solution

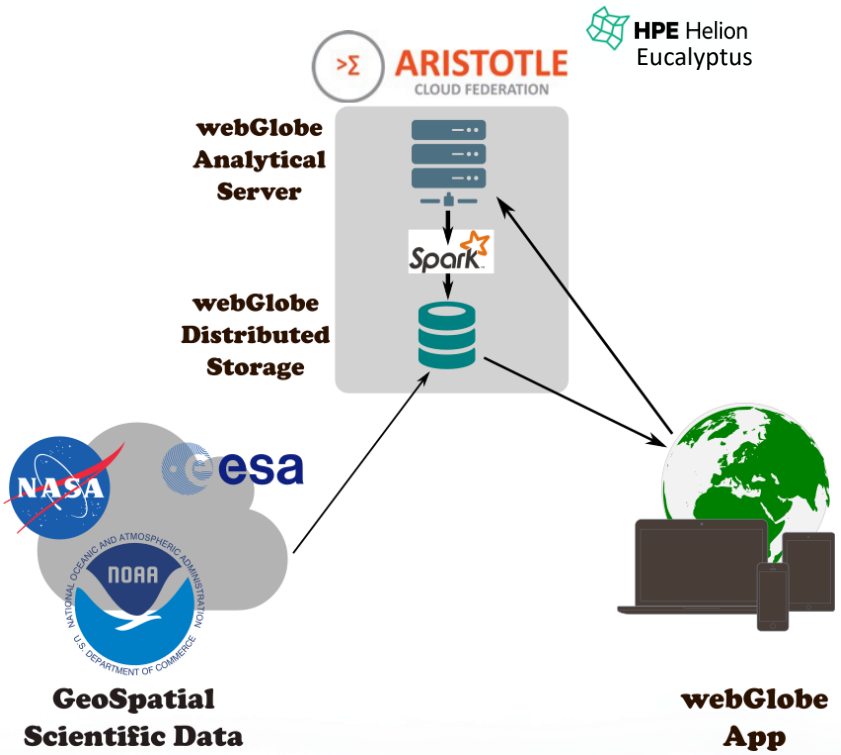
- Gaussian Process-based spatio-temporal change detection algorithm
- Distributed Apache Spark-based implementation in webGlobe enabled by Aristotle

Award No. 1541215

Visualization and Analysis of Big Geospatial Data

webGlobe

- Big geospatial analytics in the cloud
- Integrated and interactive analysis and visualization
  - Ran scalability tests on Cornell Red Cloud before migrating the app to Buffalo Lake Effect cloud



Architectural Overview of webGlobe

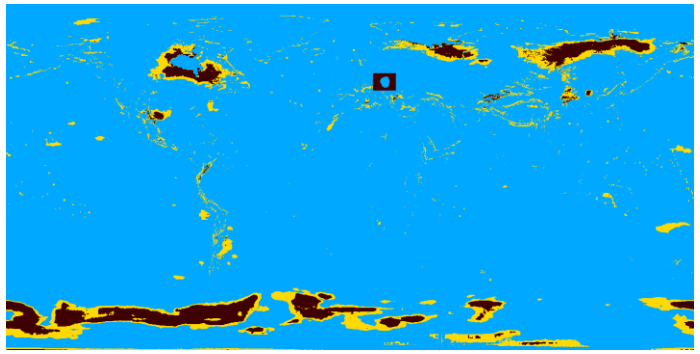
Award No. 1541215

ARISTOTLE CLOUD FEDERATION

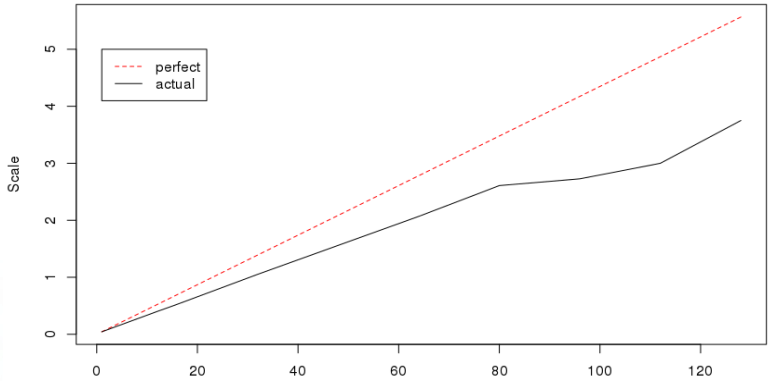
Visualization and Analysis of Big Geospatial Data

Impacts for the Scientific Community

- Analyzed 200 years of climate forecast data
  - NASA Earth Exchange Global Downscaled Projections
- Distributed method exhibits strong scaling
- Output change maps are currently being validated by climate experts
- Enables access to a wealth of scientific data
  - Climate/weather simulations
  - Remote sensing & other observed data
  - Other geospatial data products



Daily change output from the GP change algorithm



Speed up achieved by the distributed implementation on Aristotle

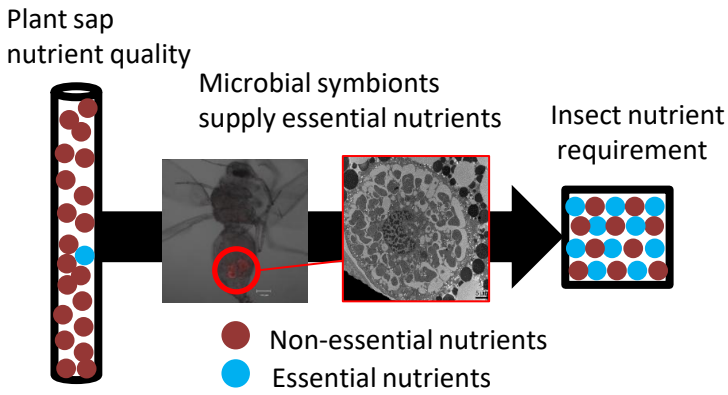
Award No. 1541215

Use Case: Mapping Transcriptome Data to Metabolic Models of Gut Microbiota

Professor Angela Douglas
Dept. Entomology, Cornell, aes326@cornell.edu
with J. Chaston (BYU), A. Moya (U. Valencia), G. Thomas (U. York), A. Heddi (INSA Lyon), B. Barker (Cornell CAC)

Problem

- Understand animal-microbe metabolic interactions and their influence on host fitness outputs
  - Important in identifying molecular targets for insect pest control (target microbial weaknesses instead of host insect directly)
  - Also important in developing a biomedical model for human health (gut microbiota)
- Challenge
  - How to scale for variety of algorithms, some still under development (complexity unknown)?
  - How to make computational pipelines reproducible?



Solution

- Used resizable virtual machines on Aristotle (Cornell Red Cloud) to handle difficult to predict computational demands (4GB to 196GB memory)
- Using Docker to build a new easy-to-use computational pipeline that can be redeployed across the Federation, on other clouds, or by other researchers

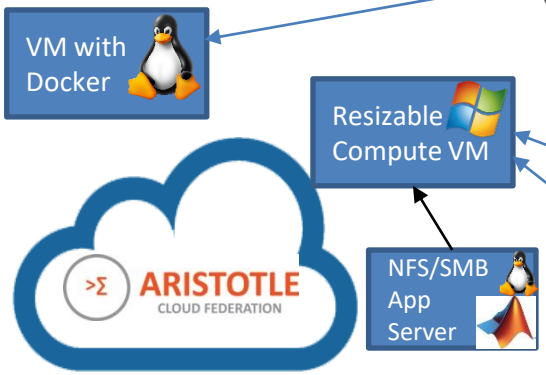
Award No. 1541215

Animal-Microbe Metabolic Interactions: A First Application

Collect biological data

Sequence genomes Measure transcript Measure metabolites

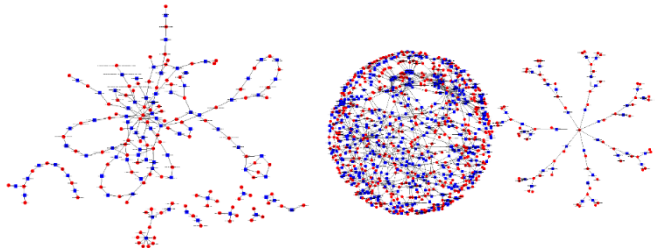
Uses Docker container with suite of tools on any Aristotle node



⇩

Create metabolic models

⇩



Analyze metabolic models to understand how nutrient exchange has evolved in insect-symbiont symbiosis

⇩

Validate models by comparing predictions to in vivo data

Award No. 1541215

Animal-microbe metabolic interactions

Impacts for the Scientific Community

- Docker containers for existing pipelines will improve the speed at which researchers may validate and continue their work
- Results presented at 6th ASM Conference on Beneficial Microbes. Seattle, WA using simulated data from Aristotle
  - Ankrah, Nana, Luan, Junbo and Douglas, Angela. 2016. Evolution of metabolite exchange in a three-partner symbiosis
  - Manuscript being written as well
- Developing general tools that will be of interest to researchers in both entomology and agriculture
  - Also will be of value to researchers studying health in humans as it relates to gut-microbial models

Plant sap feeding insects currently being studied

Postdoc associate Nana Ankrah is responsible for much of the initial model building and analysis work, utilizing Aristotle resources in the process

Award No. 1541215

Use Cases: Multi-Sourced Data Analytics to Improve Food Production and the Environment

Professor Chandra Krintz
Dept. of Computer Science, UC Santa Barbara
ckrintz@cs.ucsb.edu

Kate McCurdy
Director of UCSB Sedgwick Research Reserve
kate.mccurdy@lifesci.ucsb.edu

Professor Rich Wolski
Dept. of Computer Science, UC Santa Barbara
rich@cs.ucsb.edu

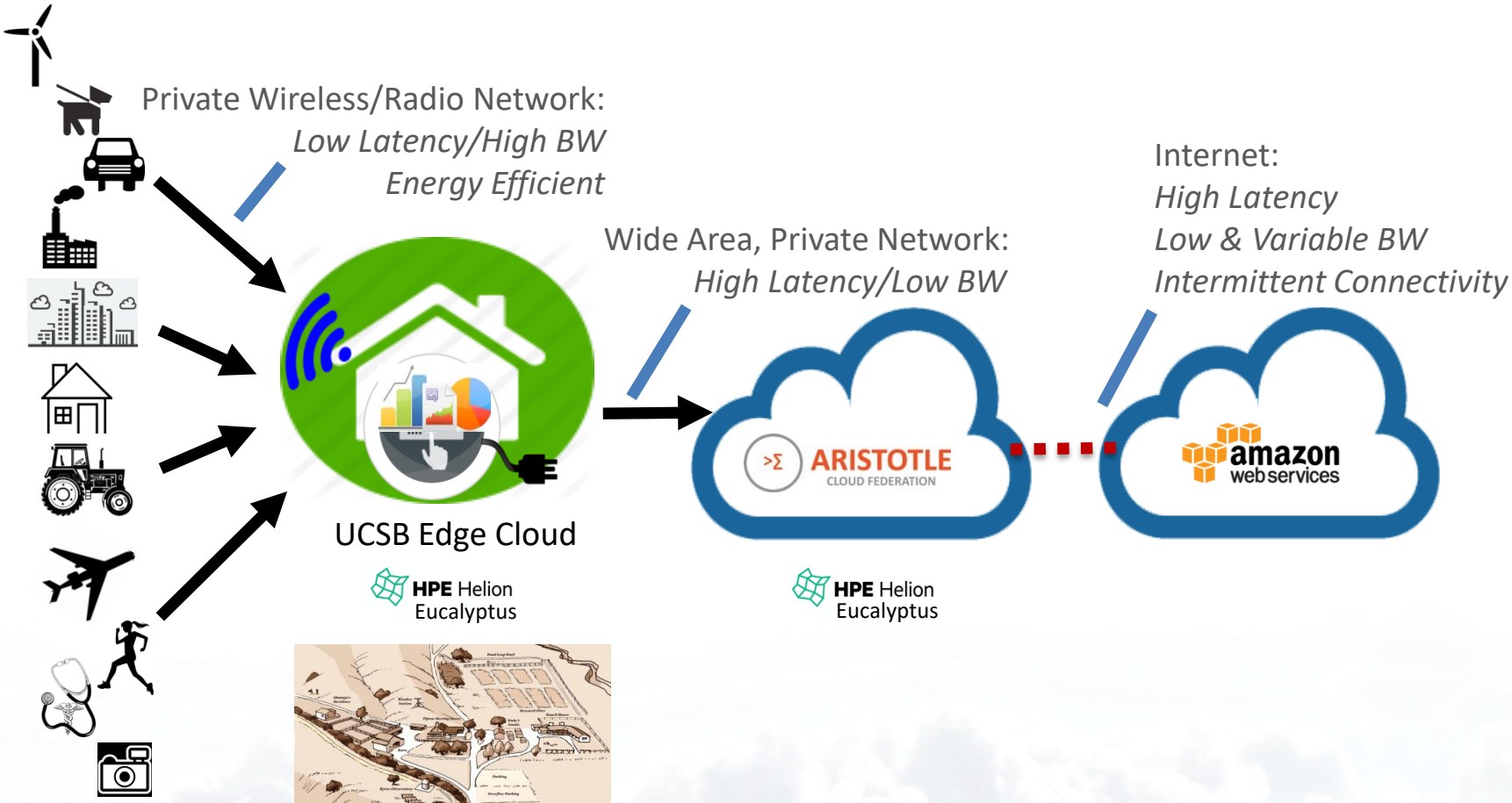
Graduate student researcher Benji Lampel testing IRROMETER soil moisture sensors

with B. Roberts, B. Sethuramasamyraja (California State University, Fresno), B. Liu (California State Polytechnic University, San Luis Obispo)

Award No. 1541215

Multi-Sourced Data Analytics

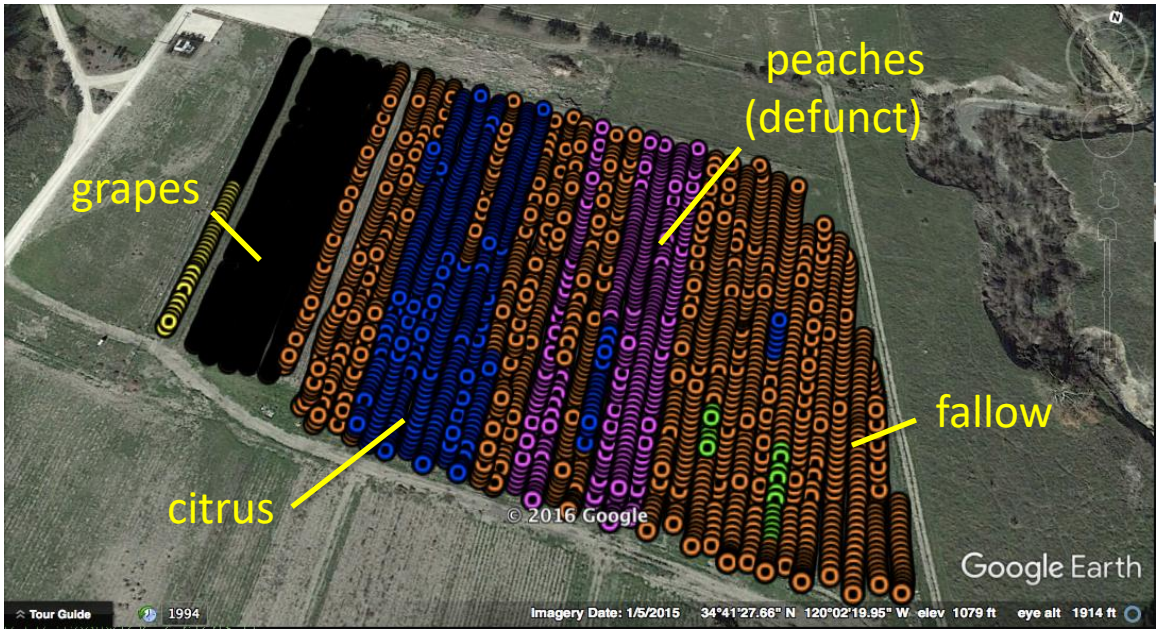
Infrastructure for Things (I4T): Edge and Federated Clouds



Award No. 1541215

Multi-Sourced Data Analytics

Food Security and Agriculture Productivity



- Developing IoT and analytics to improve yields
- Using I4T to interface from Sedgwick farm to Aristotle cloud
- Performing automated Soil Electrical Conductivity (EC) analysis
  - Segwick Reserve agricultural block colors = different levels of EC (moisture)
  - Based on analysis of soil moisture data, now using 66% less water for grape crops
  - Using new non-parametric time series technique

Award No. 1541215

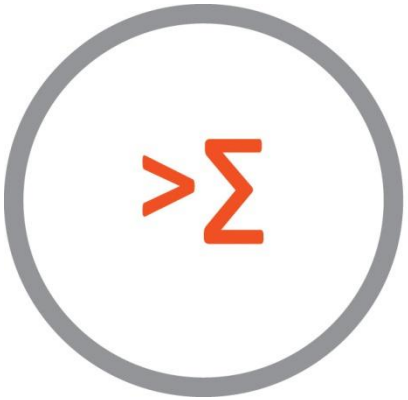
ARISTOTLE CLOUD FEDERATION

Multi-Sourced Data Analytics

Ecology and Citizen Science

- Camera Traps at Sedgwick
  - Generate 200,000 images per month
  - 12TB/year
- Where's The Bear (WTB) Application
  - Uses Google TensorFlow for image classification
  - Train on UCSB Aristotle cloud
  - Classify on UCSB Edge Cloud
  - Move only images that are interesting to researchers
  - Contribute appropriate images to Citizen Science projects such as eMammal
  - See Where's the Bear? – Automating Wildlife Image Processing Using IoT and Edge Cloud Systems (UCSB Tech Report: 10/12/16)

Award No. 1541215



ARISTOTLE

CLOUD FEDERATION

Award No. 1541215