|  |  |  |
| --- | --- | --- |
| Nathan Cook | Capt Troy Soileau | Maj Matthew McCormack |
| 96th Operations Group | 96th Cyberspace Group | USAF Test Pilot School |
|  | Air Force Test Center Department of the Air Force |  |

# Abstract

The United States Department of the Air Force’s (DAF’s) Air Force Test Center (AFTC) completed its third iteration of the AFTC Data Hackathon in August of 2022. “Hackathons” are low risk, high return investments that can jumpstart your test organization’s digital transformation. Hackathons for software engineering have existed for more than two decades in the disciplines of cryptography, web development, and apps. With the advent of data science, big data, machine learning, and artificial intelligence, “Data Hackathons” bring the hackathon model to these data-centric disciplines. A Data Hackathon allows your organization to: explore data infrastructure options, expose “data hackers” to your organization’s test and management data, evolve third-party and in-house scripts and apps to solve real-world problems, and expand awareness of the state of the art digital technologies within your organization. The future of test will require ever increasing data volume, variety, and velocity. The pace of improvement in tools and techniques will continue to accelerate. Data Hackathons can focus your test teams and provide momentum for your organization’s digital transformation.

# Acronyms, Abbreviations, Symbols

|  |  |
| --- | --- |
| AFTC | Air Force Test Center |
| API | Application Programming Interface |
| AWS | Amazon Web Services |
| CDAO | Chief Data and Artificial Intelligence Office |
| DAF | United States Department of the Air Force |
| EC2 | Elastic Compute Cloud |
| GFE | Government Furnished Equipment |
| S3 | Simple Storage Service |

# Disclaimer

Any mention of specific vendors, platforms, or tools is factual history of their use, and in no way an endorsement by the authors, the Department of the Air Force, the United States Government, or a sponsorship of the vendors, platforms, or tools themselves.

# Introduction

The United States Department of the Air Force’s (DAF’s) Air Force Test Center (AFTC) completed its third iteration of the AFTC Data Hackathon in August of 2022 [@http://zotero.org/users/7103471/items/T67Q4I5G; @http://zotero.org/users/7103471/items/T6TMBPY9]. “Hackathons” are low risk, high return investments that can jumpstart your test organization’s digital transformation. Hackathons for software engineering have existed for more than two decades in the disciplines of cryptography, web development, and apps. With the advent of data science, big data, machine learning, and artificial intelligence, “Data Hackathons” bring the hackathon model to these data-centric disciplines.

This paper will explain the benefits of implementing a Data Hackathon in your flight test organization. In addition, a framework for planning, provisioning, executing, and following up on a Data Hackathon is provided, which can then be modified as required to meet the opportunities and constraints of your organization.

# Benefits

A Data Hackathon allows your organization to:

* explore data infrastructure options,
* expose “data hackers” to your organization’s test and management data,
* evolve third-party and in-house scripts and apps to solve real-world problems, and
* expand awareness of the state of the art digital technologies within your organization.

## Explore

**Explore data infrastructure options.** Data infrastructure is a prerequisite for working with data. Monica Rogati, among others, made a comparison with Maslow’s hierarchy of needs [@http://zotero.org/users/7103471/items/8FKVC28U] and created “The Data Science Hierarchy of Needs” [@http://zotero.org/users/7103471/items/4LGMVBAM], depicted as a triangle or pyramid, [Figure 1](#fig-dshierarchy). Collection, movement, and storage are at the base of the pyramid. This representation is chronological, following the flow of data from source to working location, followed by techniques to explore, understand, and organize, before generating algorithms and models for inference, prediction, and influencing decisions.

|  |
| --- |
| Figure 1: The Data Science Hierarchy of Needs |

The exact sequence, though, is not essential for the bottom two layers. The basic components of infrastructure are:

* Store (sometimes called “data at rest”)
* Compute
* Transport (sometimes called “data in motion”)

A Data Hackathon directly stresses all three components, by design. Your organization may have been considering adding or modifying aspects of each infrastructure component. The hackathon is an opportunity to put these updates in action and uncover pain points that would remain invisible without direct application of focused effort. The hacker teams may use the components in ways that were not anticipated by the originators, as well.

### AFTC Data Hackathon Infrastructure

The AFTC Data Hackathon makes use of infrastructure already existing and available to DAF members, but which may not be used to their full potential. To date, the primary platforms used, in chronological order:

* DAF CDAO “VAULT” providing access to:
  + Amazon Web Services (AWS) Elastic Compute Cloud (EC2) compute and Simple Storage Service (S3) store in the cloud
  + Databricks notebooks running Python and R orchestrated with Apache Spark
  + Hue SQL assistant
  + Apache Zeppelin notebooks running Python orchestrated with Apache Spark
  + Plotly Dash dashboard front ends with Python visualizations
* Microsoft 365 including:
  + Microsoft Azure compute and store in the cloud
  + Excel
  + Teams
  + SharePoint
  + Power Automate
  + Power Apps
  + Power BI
* DAF CDAO “Envision” providing access to:
  + AWS EC2 and S3
  + Palantir Foundry
  + Python and R scripting
  + Data Application Programming Interfaces (APIs)

To the maximum extent, the hacker teams used government-furnished equipment (GFE)

# Lessons Learned

Knowledge learned along the way and of value to the next brave soul to work in this area.

# Conclusions

Points to be drawn from the material and data provided.

# Acknowledgement

To extend thanks in support of the paper.

# References

# Biography/Photograph

A brief paragraph(s) of the author(s). A photograph is highly recommended, but of course optional.