

Zika hackathon fights disease with big data

May 20 2016



More than 50 people from the Austin community met at Cloudera to fight Zika disease with big data at the Zika Hackathon. Credit: TACC

More than 50 data scientists, engineers, and UT Austin students gathered on Sunday, May 15, 2016 to use Big Data to fight the spread of Zika for the "Austin Zika Hackathon" at the Cloudera offices downtown.

Zika, a mosquito-borne disease that can cause fever and birth defects, threatens to spread to the United States. As of mid-May 2016, Mexico

had reported 272 cases of Zika, according to USA Today. The problem has grown so large that President Obama has requested \$1.9 billion to halt the spread of Zika. The U.S. Centers for Disease Control is now ramping up collection of [data](#) that tracks Zika spread. But big gaps exist in linking different kinds of data, and that makes it tough for experts to predict where it will go next and what to do to prevent it.

The Zika Hackathon participants investigated ways to pool together different sets of data, such as outbreak reports, stagnant [water sources](#), empty swimming pools and ponds that are potential mosquito breeding grounds, and even Facebook and Twitter feeds. The Texas Advanced Computing Center (TACC) plans to store all the data in one place, a new data-intensive supercomputer called Wrangler.

"We're trying to collect these disparate pieces of data, and there's not a good way for people to ask questions about that data—that's the big problem," said Ari Kahn, human translational genomics Coordinator at TACC.

Said Kahn: "TACC's role is providing an infrastructure and consulting to support this project. Wrangler is a specialized data-intensive system that runs an optimized version of Cloudera, and it really speeds up the process."

Cloudera is a [big data](#) company, according to its Chief Security Architect and Zika Hackathon organizer Eddie Garcia. "What we do is make Apache Hadoop enterprise-ready for organizations to do [big data analytics](#) and find new insights within their data sets," Garcia said.

"What we can do in a one-day hackathon is to focus on one data problem, for example, if there were an outbreak—where we would we first send support and kits to local communities and direct awareness programs on prevention by removing stagnant water or using repellents

that are effective against Aedes," Garcia said. "The Zika Hackathon is about bringing awareness and building a platform that is repeatable, not just for the Zika virus data analysis. Someone can basically take what we did here today and apply it to some other unknown outbreak or some other analysis for something even better than what we're doing today. It's really about getting people together, excited, bringing awareness, and building out a platform that is repeatable for others to collaborate, apply machine learning and perform analytics using Apache Hadoop."

"It's just great to see a roomful of people buzzing, talking about bringing these skills to bear either to build a consolidated data set, a little visualization, or even a little tool," said Jon Loyens, chief product officer and co-founder of Data.World, a new Austin startup. "Every little bit helps and everyone here realizes that."

The Zika Hackathon brought together an emerging kind of scientist, a data scientist. Data scientists specialize both in translating information from many different sources into data that can be used together and in using new technologies by which knowledge can be extracted from today's massive data collections.

Data scientist Juliet Houglund of Cloudera described what that is: "There are three classes of work that get put under the umbrella of data science. Data scrubbing—getting data in the right format, in the right place—is a huge part of any job where you're going to do something useful with that data. Investigative analytics looks at historic data and doing interesting, useful analysis on it. Operational analytics supports recommendation engines, fraud detection systems, and more."

The Zika hackers formed groups and worked on creating demo projects based off of sample CDC and other data available at this link. One project developed a working tensor flow model that used machine learning to search through aerial images for pools of stagnant water,

potential breeding ground for mosquitos that carry Zika. Another team developed a mobile app with node.js that would allow researchers to report developing cases of mosquito-borne illness. One demonstrated a way to map microcephaly occurrences in Brazil using an R maps interface to Leaflet. Another made headway into readying CDC data from Puerto Rico to layer with CIA Fact Book data for richer understanding of how Zika has progressed there.

Software developer David Walling of TACC's Data Intensive Computing group spoke of his current research extracting rich data from 'grey literature,' unofficial records that can be images inside PDF files, a bane of data scientists. His work uses natural language processing techniques to map occurrences in the grey literature of a given species such as fish at specific locations and dates. Progress on this problem would translate well to getting more information for researchers about Zika.

"If you can see where all the water sources are and then overlay how the reports of outbreaks are happening, then you can create a model for how it's spreading and how it will spread in the future based on where the water sources are. Then maybe you can come up with some plans to offset that so the spreading doesn't happen as fast or doesn't happen at all," Ari Kahn said.

The charitable arm of the data analytics company, Cloudera Cares, along with TACC and other local partners are planning to hold quarterly hackathons as part of a larger planned project to use Big Data to battle Zika and other threats. The project aims to make it easier for researchers to get answers and even help prevent outbreaks from happening.

Provided by University of Texas at Austin

Citation: Zika hackathon fights disease with big data (2016, May 20) retrieved 28 April 2024

from <https://medicalxpress.com/news/2016-05-zika-hackathon-disease-big.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.