
MMXVI
HACKATHON clt

WE ARE BACK

on the line

- wifiSSID: NinerWiFi Guest
- Password: no password (like really there is no password)

@charlottehacks

#HACKATHONCLT

agenda

friday, march 18

5p	Registration Table Opens
5:30- 7:30p	Kickoff & Party
7:30p	Hack Problem Presentation
8:30p	**Go Hack**
12a	**Midnight Snacks**

saturday, march 19

7a-8a	Breakfast
10a	Hacking Ends ｜ Judging Round 1 Begins
12:00p	Lunch
1:30p	Presentations & Awards Ceremony

the basics

- what is a hackathon?
- why charlotte?
- what's the motive?
- big data AGAIN?

what made it possible

- community
- community
- community

who made it possible

rules of engagement

- nothing illegal
- respect copyright
- keep it clean
- terms & conditions
- all work must be on site
- organizers reserve the right…

THE PROBLEM

business problem

HOW CAN THE SECOND HARVEST FOOD BANK OF METROLINA GROW –RAISE MORE DONATIONS, BECOME EFFICIENT IN DISTRIBUTION AND DO SOMETHING NO OTHER NON-PROFIT MAY HAVE DONE….

…THEN EVERYBODY WINS (LITERALLY)

key statistics

- DOUBLING FOOTPRINT
- SPACE - 23,000 sq. ft. - 61,000 sq. ft. of space
- 46MM lbs. of food - 75MM lbs. of food
- 35MM meals distributed
- 1 in 7 struggles with hunger
- Counties served – Anson NC, Burke NC, Cabarrus NC, Catawba NC, Cleveland NC, Gaston NC, Iredell NC, Lincoln NC, Mecklenburg NC, Montgomery NC, Rowan NC, Rutherford NC, Stanly NC, Union NC, Cherokee SC, Lancaster SC, Spartanburg SC, Union SC, York SC

what are you looking for

- understanding of the data (it is big)
- scale trumps complexity in algo
- Ability to tease out sophisticated insights
- Merge data we haven't given you
- Bottlenecks (in collection, distribution)
- PATTERNS (ANY pattern – checks, herringbone, tie'dye)
- Simple yet powerful algorithms

THE DATA

what you get

- real-world donation data + retail data
- 1.6 billion+ records
- scrubbed & de-identified
- field names & explanations in Data Dictionary on github

donations data

- Posting_Date | Date items were received
- Location_Code|Location of distribution center
- Lot_No | Internal set identifier
- Pallet_No | Internal sub-set identifier
- Document_No | Similar to receipt number
- Ext_Gross_Weight | Weight of items in lbs
- Unit | Unit to measure quantity
- Quantity | Number of items or cases
- Cost_Amount | Value of items if receipt provided
- Entry_No | Number of entry
- Category_Code | Item category identifier
- Category_Description | Type of product
- Donor_Category | High level description of the donor
- Donor_City | City
- Donor_State | State
- Donor_Zip | Zip
- Donor_ID | Unique identifier for donor

monetary donations data

- **ID** | Donation identifier
- **City** | Donor city
- **State** | Donor state
- **Zip** | Donor zip
- **Donation_Date** | Date donation was made
- **Donation_Amount** | Monetary donation amount
- **Donation_Type** | Code representing method of payment
- **Donation_Number** | Entry number for donation
- **Donation_Batch_Number** | Identifier for group donation
- **Donation_When_Added** | Date donation was accepted
- **Donation_When_Acknowledged** | Date donor was acknowledged

distributon center data

- Posting_Date | Date items were received
- Location_Code | Location of distribution center
- Lot_No | Internal set identifier
- Pallet_No | Internal sub-set identifier
- Document_No | Unique to each Donor's donation, bring together types of donations
- Ext_Gross_Weight | Weight of items
- Unit | Unit to measure quantity
- Quantity | Number of items
- Cost_Amount | Value of items if receipt provided
- Entry_No | Number of entry
- Category_Code | Item category identifier
- Category_Description | Type of product
- Agency_FBC_County_Code | County
- Agency_UNC_Activity_Status | Flag indicating agency is still active
- Agency_City | City
- Agency_State | State
- Agency_Zip | Zip
- Agency_FBC_Size_Code | Estimate of person count
- Agency_FBC_Category_Code | Internal program that is receiving items
- Agency_ID | Unique identifier for agency

ht transactions

- **field** | definition
- **customer** | unique id per customer household
- **receipt** | unique id per checkout
- **date time** |date and time of checkout in yyyy-mm-dd hh:mm:ss
- **zip** | 5 digit postal code where the transaction took place
- **Ean** | industry standard item id number
- **Plu** | item id number for weighted items like produce and meat
- **Mupc** | id number for small groups of items (different flavors of a particular brand and size of yogurt will have different eans but may have the same mupc)
- **Subcat** | item subcategory number
- **Cat** | item category number
- **Dept** | item department number
- **Ean description** | item description at the ean level
- **Subcat description**| item description at the subcat level
- **Cat description**| item description at the cat level
- **Dept description** | item description at the dept level
- **Promo id** |id identifying the promotion (if there is one) that applies to the transaction
- **Price** | gross $ sales for this ean from this receipt
- **Discount** | total $ discounts for this ean from this receipt
- **Quantity** | total number of units of the ean purchased on this receipt

how will you share results

- format – your call
- time – 5 minute presentations for shortlisted teams
- Creativity (yours), predictability (machines), scalability (your code) are all important
- judges will be revealed in 2 seconds… they are tough, but they are awesome

Shortlist Judges

Chase Cabanillas

Mike Keating

Pete Murphy

Tim Reagan

Kevin Ledford

Abhishek Mehta

Finals Judges

Abhishek Mehta, tresata

Jean Wright, CHS

Dr. Michael Dulin, UNCC

TECH

hardware stack

- At scale hadoop 5 node cluster (hat tip DataChambers)
- Node specifications:
  - 32 cores
  - 128GB RAM
  - 14TB storage

Data Chambers by NORTH STATE

dataset stored

- stored in HDFS
- For data dictionary location, visit here: http://www.github.com/tresata/hackathonclt2016
- DO NOT PULL DOWN THE ENTIRE SET
- DO NOT LEAVE WITH ANY DATA
- YES WE WILL SPOT CHECK

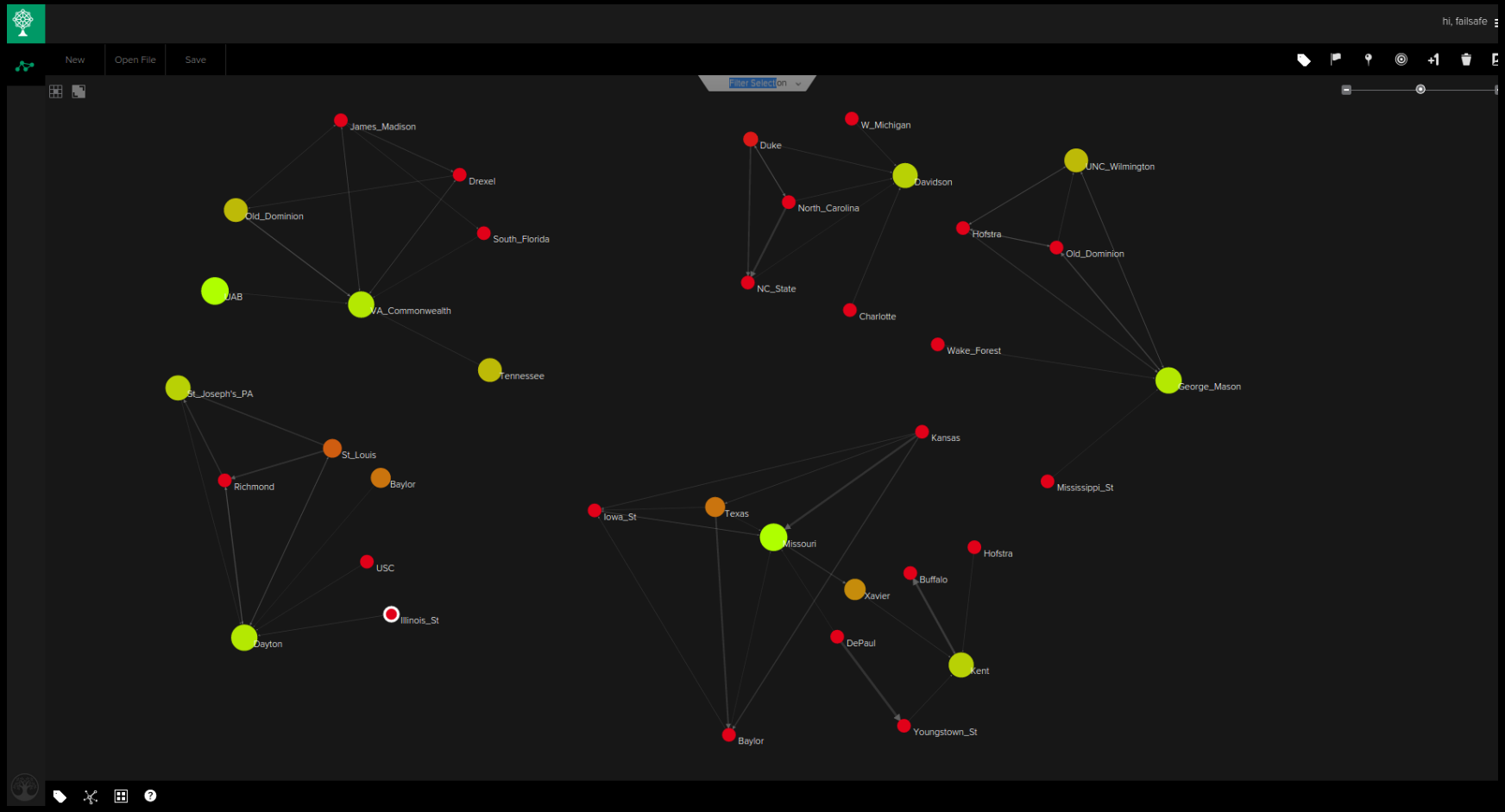
languages

- anything that can compile to a .jar (like java, scala, etc.)
- JDBC connection available through Hive
- Python, via pyspark and Anaconda
- Apache Spark

tresata ORION

- You asked…we listened…
- We've given you access to our real-time network graph engine + a summary explaining how to use it
- Not required
- To access location, visit here: http://www.github.com/tresata/hackathonclt2016

ORION


hi, failsafe
New
Open File
Save
Filter Selection
James_Madison
Drexel
Old_Dominion
South_Florida
UAB
VA_Commonwealth
Tennessee
St_Joseph's_PA
St_Louis
Richmond
Baylor
USC
Illinois_St
Dayton
W_Michigan
Duke
Davidson
North_Carolina
NC_State
Charlotte
UNC_Wilmington
Hofstra
Old_Dominion
Wake_Forest
George_Mason
Kansas
Mississippi_St
Iowa_St
Texas
Missouri
Hofstra
Buffalo
Xavier
DePaul
Kent
Youngstown_St
Baylor

THE PRIZES

in it to win it

$5,000 HACK PRIZE

$2,000 CODE PRIZE

$1,000 freeSTYLE PRIZE

not just about the cheddar

RAFFLE PRIZE

4 DAY, 3 NIGHT

SKI TRIP FOR 2

LOGISTICS

the basics

- restrooms – located on 5th and 4th floors
- building access – DO NOT LEAVE
- be reasonable – alcohol, noise, trash, etc.
- lounges – vote for your favorite tomorrow
- support – look for grey hoodie shirts (or all black onesies), help desks, and slack

slack

- General guidelines for use:
  - Ask questions in relevant channel
  - Keep it clean (not joking)
  - Remember, this is public to everyone at the event

help desks

- DATA
- ORION
- TECH
- BUSINESS PROBLEM
- GENERAL

if you haven't registered yet…

- COME FIND ME RIGHT AFTER THIS

- Also, for those of you who have not declared teams…no access until I have all your details !!

questions?

www.hackathonclt.slack.com

- Brittany Box– Event
- Jainin Shah – Tech/Infrastructure
- Caitlin Lohrenz – ORION
- Matthew Dix – Data
- Chase & Abhi – Business Problem

LET'S HACK

@charlottehacks | #HACKATHONCLT

http://www.github.com/tresata/hackathon2016