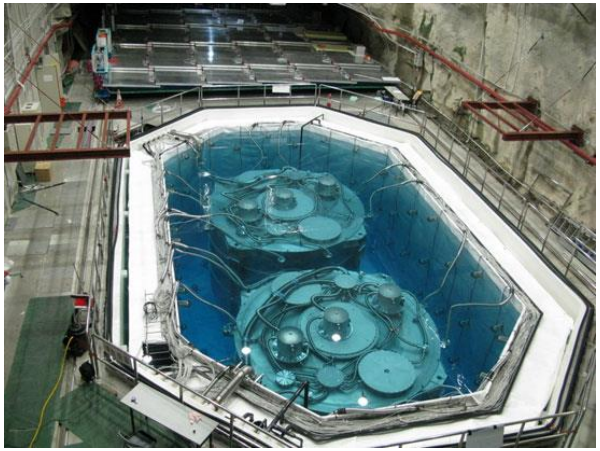
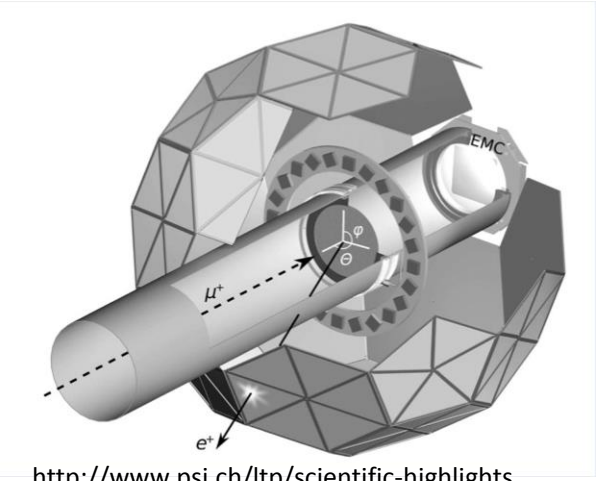
From Muons and Neutrinos to Big Data and the Internet of Things



http://www.psl.wisc.edu/projects/large/dayabay



http://www.psi.ch/ltp/scientific-highlights

David M. Webber

Chief Technology Officer,

Scanalytics, Inc.

November 7, 2014

Outline

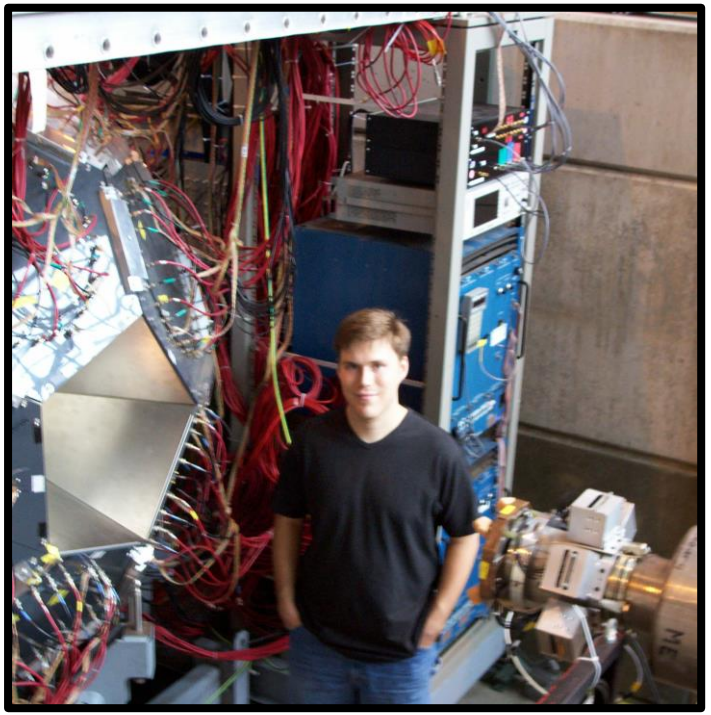
- A little about me
- A little about Scanalytics, Inc.
- Working in a startup
- Discussion of the Industry
  - Internet of Things
  - Big Data
- Making a career transition

Disclaimers

- Discussion of products or services not of Scanalytics, Inc is not a recommendation or endorsement of those products or services.
- I was not a member of the ATLAS collaboration

About David

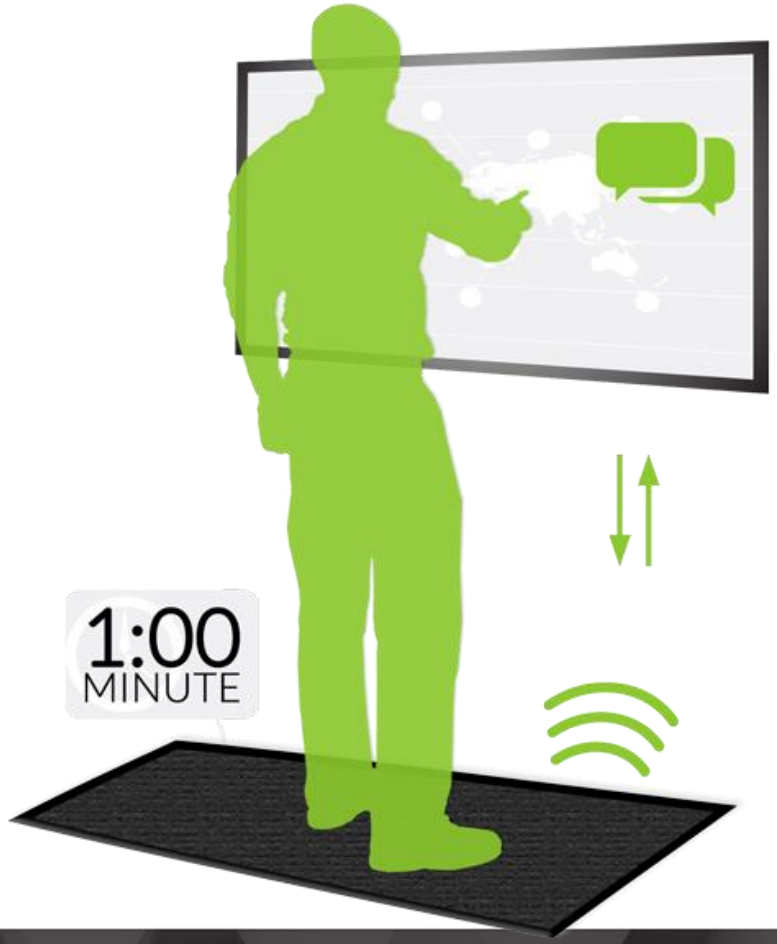
- Data scientist and physicist
- Measured particle properties 2003-2013
  - Muon (a “heavy electron”) lifetime
  - Neutrino mixing angle θ_{13}
- Scanalytics, Inc. 2013-Present
  - Internet-of-things startup
  - Developing algorithms using floor sensors to measure behavior in physical spaces



Standing next to a physics experiment (MuLan), circa 2005,

Scanalytics Inc.

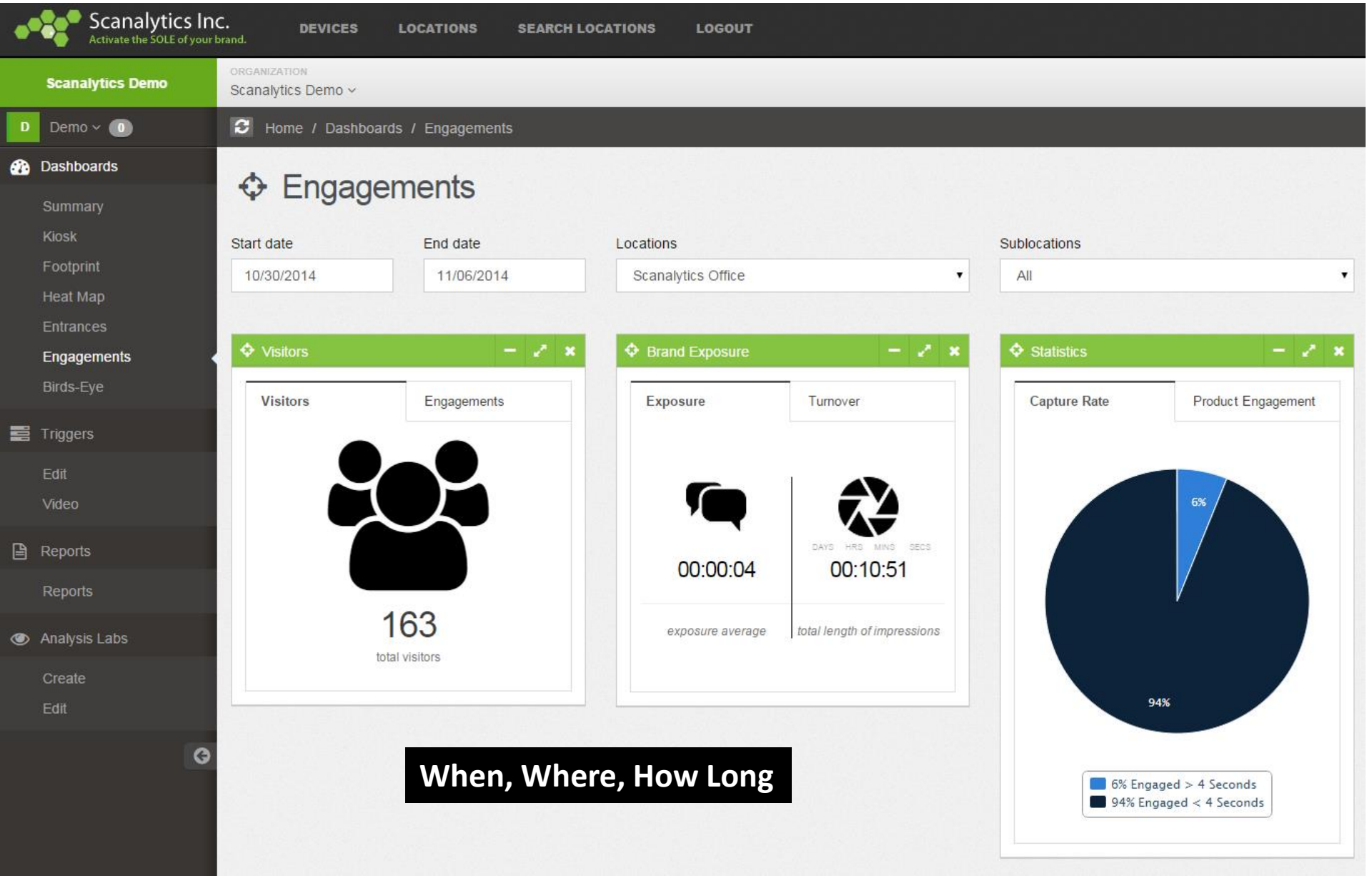
A sensor-based analytics and engagement platform for physical spaces



Features:

- Records time and duration of visits
- Trigger locally or remotely based on configurable thresholds
- Display real-time and historical data in online dashboard
- Platform for interactive spaces

Data Dashboard



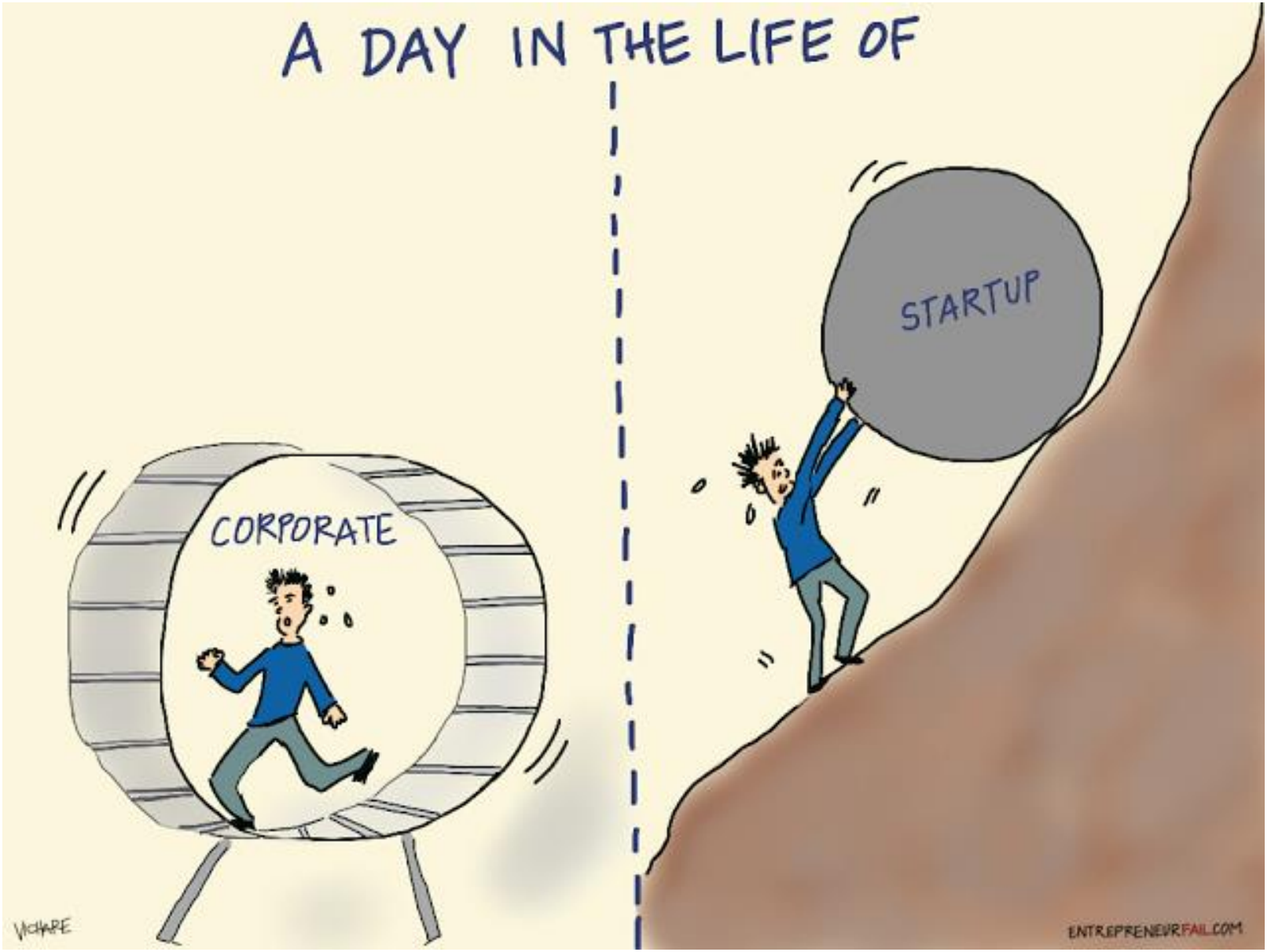
Scanalytics, Inc.

- Founded in 2012
- Based in Milwaukee
- Wisconsin Innovation Award, 2013
- Accepted into Microsoft Ventures Accelerator in Seattle, Fall 2014
  - Best Business Progression Award at Hackathon
- Top 5 Internet of Things companies to watch
  - http://www.techrepublic.com/article/5-internet-of-things-startups-to-watch/
- SoleSensor 2.0 Launched

My Day-to-Day

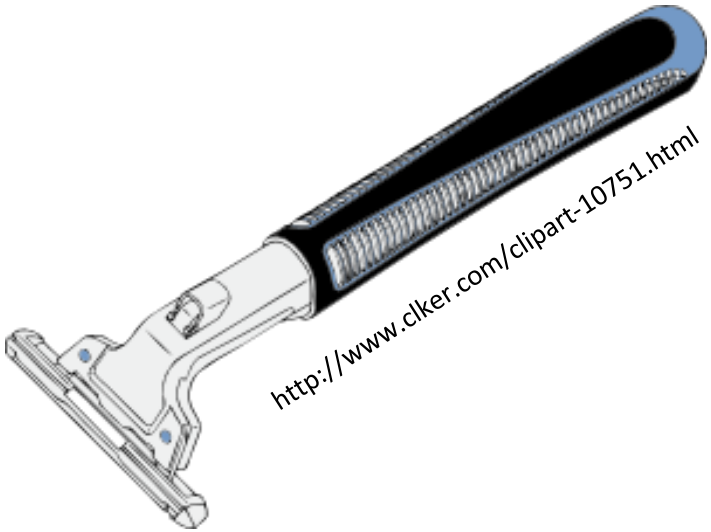
- Primary role: end-to-end data integrity
  - sensors, transmission, storage, analytics, display
- Every day is different
- Similarities to Grad School/Post-Doc
  - Work with a talented and motivated team
  - Analyze data
  - Write programs
  - Communicate technical information
  - Troubleshoot hardware and software issues
  - Coordinate work
- Sense of urgency

Working in a Startup



http://www.entrepreneurfail.com/2013/12/a-day-in-life-of-corporate-vs-startup.html

Case Study: New Razor



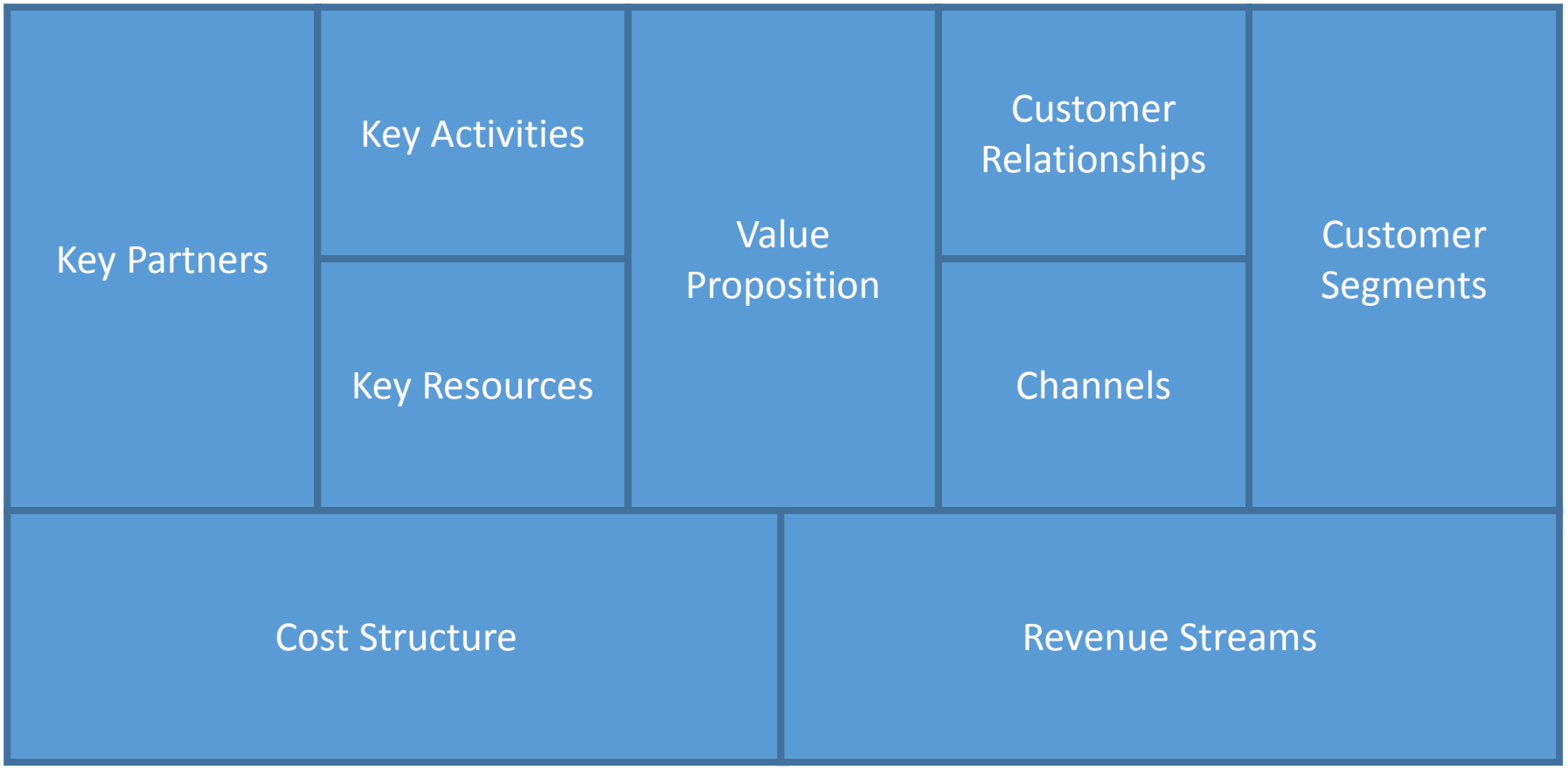
- 10-blade razor
- Lasts 4x as long
- Costs half the price
- Is it a superior product?
- Will it do well in the market?

Case Study: Medical Equipment

http://www.clker.com/clipart-293841.html

- In-home monitoring device
- Who uses the device?
- Who pays for the device?
- Is the customer the same as the end user?
- How do you market it?

Business model canvas: a tool for planning a business



See video at http://www.businessmodelgeneration.com/canvas/bmc

One of the metaphors that I use for startups is you throw yourself off a cliff and assemble an airplane on the way down.

-- Reid Hoffman, co-founder of LinkedIn

https://blog.kissmetrics.com/hoffmans-advice-for-entrepreneurs/

"You only ever experience two emotions: euphoria and terror. And I find that lack of sleep enhances them both."

-- Marc Andreessen, Co-Founder Loudcloud

http://www.bhorowitz.com/picking_a_general_partner

Advise for Entrepreneurs

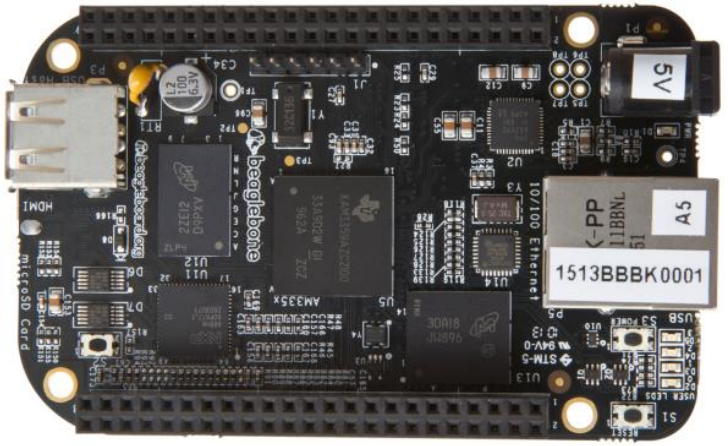
- Get one or more co-founders
  - Complement your skills
  - Share the ups and downs
- Focus on a minimum-viable-product (MVP)
  - Get feedback from customers
- Join an incubator
  - e.g. http://researchpark.illinois.edu/enterpriseworks
- Be critical of mentor whiplash
- You will have roles you didn't anticipate
  - You can't guarantee you'll be doing the same thing a year from now

The Internet of Things

A problem than the internet of things could solve

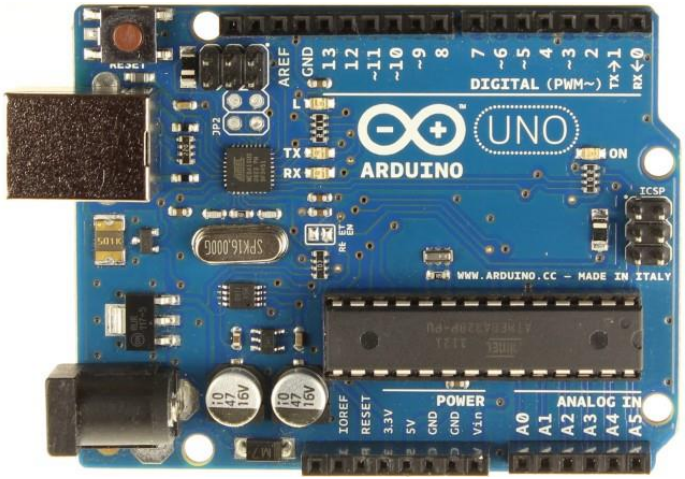
Computing Everywhere

Raspberry Pi, $35



Beaglebone Black, $45

- Credit-card-sized computers are small and affordable
- Easy to connect sensors to the internet
- What happens when it's $1?
- When they talk to each other directly?



Arduino Uno, < $35

images: http://makezine.com/magazine/how-to-choose-the-right-platform-raspberry-pi-or-beaglebone-black/, http://arstechnica.com/information-technology/2013/10/arduino-creator-explains-why-open-source-matters-in-hardware-too/

Components of the Internet of Things (IoT)

- Sensors & Control Systems
- Internet connectivity
- Integration of data, people, and processes
  - Security
  - Building & home energy management
  - Supply-chain management
  - Analytics
  - Lots of opportunities for people who know data!

Impact of IoT

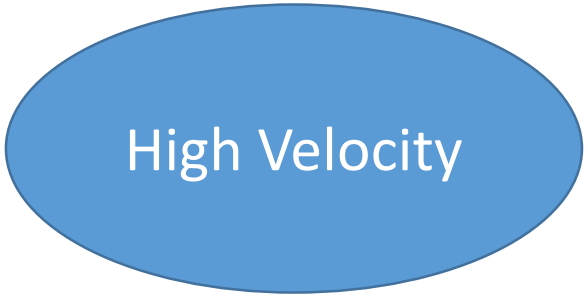
- Home
- Transportation
- Healthcare
- Buildings
- Cities

As the internet of things grows, vertical siloes systems will give way to horizontal networked platforms

- 2014
  - Over 6 billion installed devices
  - Over 180 Billion in revenue
- 2020
  - Over 27 billion installed devices
  - Over 1 trillion in revenue

Stats from http://postscapes.com/what-exactly-is-the-internet-of-things-infographic

Big Data



We live in a world of Big Data

- 90% of the world's data has been created in the last two years (IBM, 2013)
- Over 4.1 billion terabytes generated in 2013 (IDC)
- Big Data is characterized by
  - Volume
  - Velocity
  - Variety
- Examples
  - Scientific Data (Large Hadron Collider, Sloan Digital Sky Survey)
  - Social Data (Twitter, Facebook, LinkedIn, Foursquare…)
  - Sensor Data (Scanalytics!)

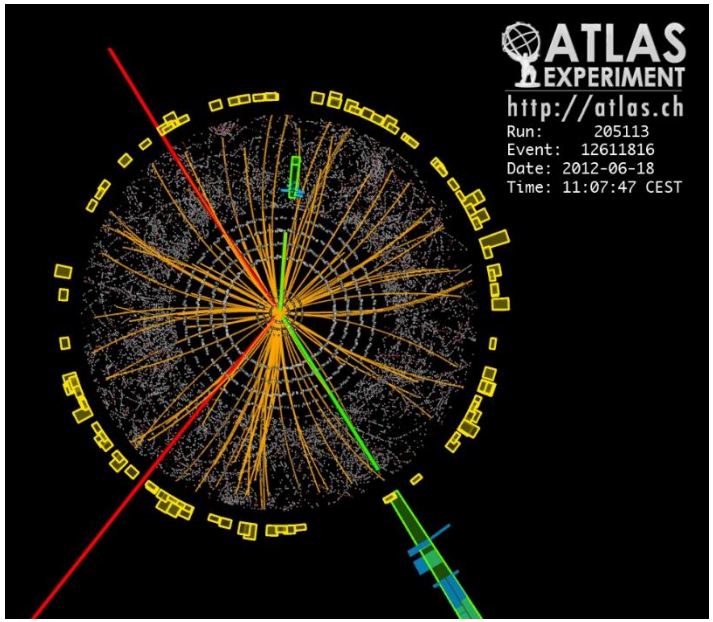
Challenges of dealing with Big Data

- Storage
  - Where do you keep it?
- Processing
  - Too much for a single machine
  - Multiple data sources
- Streaming
  - How do you keep up with real-time data
- Security
  - Weather vs. Financial vs. Medical Data
- Fault Tolerance
  - How do you recover from computer failures?

Big Science: Big Data

- 1 MByte per collision “event”
- 600 Million events per second
- 100 Million events per second selected for digital reconstruction
- 100-200 events selected for recording
- 1 GB/s raw data rate after event selection
- 15 petabytes (15,000 Terabytes) produced annually

A “sighting” of the Higgs Boson



H -> 2e2mu candidate event

http://home.web.cern.ch/about/computing/processing-what-record

http://www.bnl.gov/atlas/news/news.php?a=11512

Big Science: Big Computing

Cern Data Center

- 1450 square meters
- 3.5 MW electrical power capacity
- 30 Pbytes (30,000 TB) data storage on disk
- 90,000 processing cores
- Processes about 1 petabyte each day
- Tier 0 of a multi-tier system

Source: home.web.cern.ch/about/computing

Social Media: Big Data

- Real-time social network feeds
  - Twitter, Facebook, LinkedIn, etc.
- Systems capable of processing 50,000 tweets/second
- Sentiment analysis (happy/sad)
- Demographic analysis (gender, stage of life)
- Location analysis
- Identify Influencers

New models for computation: Hadoop

- Open-source (Apache project)
- Java-based
- Hadoop Distributed File System (HDFS)
  - Replicates data blocks in case of node failure
  - Storage is local to the computer (no shared network drive)
- Map-Reduce Algorithms
- Advantages:
  - Fault tolerant
  - Brings compute to the data
- Disadvantages:
  - Java is slower than native C
  - HDFS is write-only (can't update data)

New models for data storage: NoSQL Databases

- Document Database (e.g. MongoDB)
  - A “document” is a self-describing set of keys and values
  - Format can change document-by-document
- Graph Database (e.g. Neo4J)
  - Stores nodes and links
  - “Friends of Friends” -- type queries
- Columnar Database (e.g. Cassandra, Hbase)
  - Stores columns together instead of rows

New date formats: JavaScript Object Notation (JSON)

- A nested, self-describing data structure of key/value pairs.
- Understood by many programming languages

{
“name” : { “first” : “David”, “last” : “Webber” },
“company” : “Scanalytics, Inc”,
“background” : “physics”
}

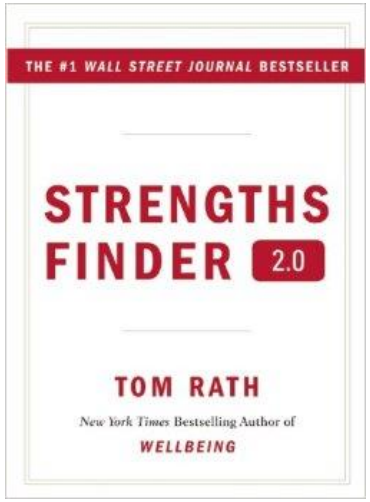
Big Data vs. Big Physics

- Both
  - Enormous data sets
  - High-velocity data
- Big Physics
  - Complex, highly structured
  - Combine multiple data sources
- Big Data
  - Structured and unstructured. (e.g. Natural language)
  - Fairly vertical/siloed (opportunities for cross-device analysis!)

Making the Transition

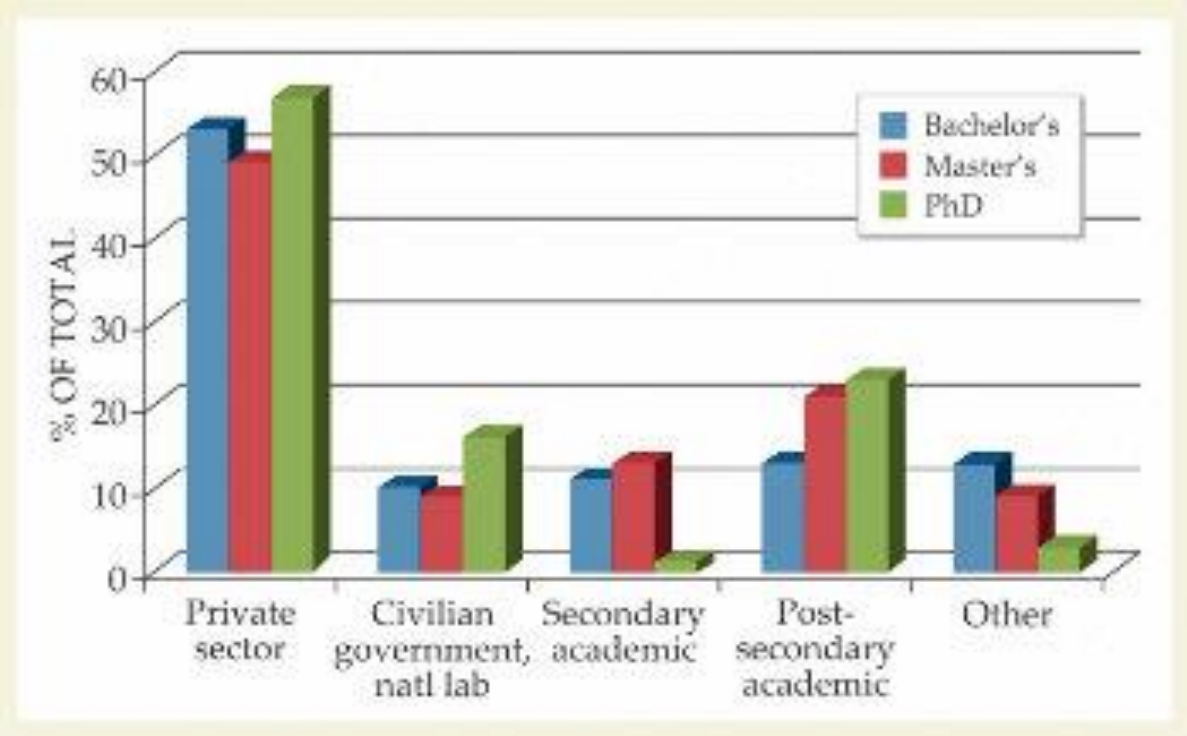
The jobs are out there

My Transition



- Soul Searching
- Self-Assessment
  - Strengths Finder 2.0
- Exploring Interests
  - Informational Interviews
  - Meetup.com (Linux, Programming, and Big Data)
  - Startup Weekend
  - Entrepreneurial Bootcamp
- Interviews
  - 2 startups, 2 Dev/Ops Linux positions
- Soft transition from Post-Doc to Startup
- Data Scientist → Director of Data Science → CTO

Non-Traditional Careers are More Common

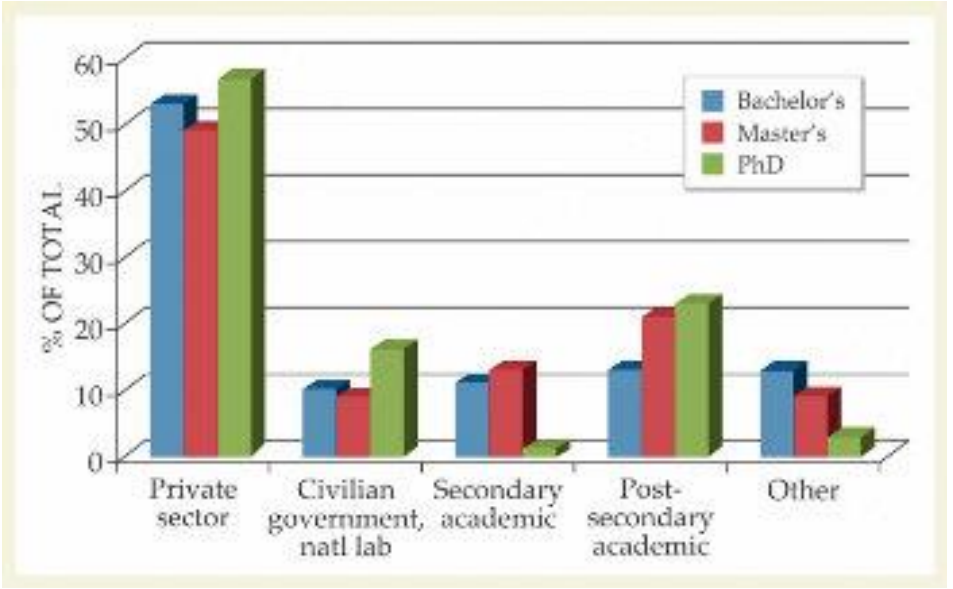


- Less than 20% of physics Ph.D.s are permanently employed in academia
- Physicist unemployment is low (less than 5%)

bullets and image from http://www.aps.org/careers/guidance/development/intro.cfm

The jobs are out there

- I personally know Physics Ph.Ds. that have gone on to
  - Defense and space
  - Bioinformatics
  - National labs
  - Science writing
  - Financial software
  - Linux Dev/Ops
  - Insurance analysis
  - Radiation screening of cargo
  - Optical silicon wafer analysis



- Other profiles
  - http://www.aps.org/careers/physicists/profiles/index.cfm

http://en.wikipedia.org/wiki/File:Rumsfeld1.jpg

As we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don't know we don't know.

-- Donald Rumsfeld

What about Unknown Knowns?

Physicists' Unknown Knowns

- Problem solving
  - Critical thinking
  - Know what's relevant and not relevant
  - Tackle a problem without a-priori knowing the solution
- Experimentation
- Model building
- Data analysis
- Public speaking
- Clear explain complex concepts
- Technical: Programming (C,C++,Python), Advanced Mathematics
- Even more:
  - http://www.aps.org/careers/guidance/development/skillsinventory.cfm

Summary

- A little about me
- A little about Scanalytics, Inc.
- Working in a startup
- Discussion of the Industry
  - Internet of Things
  - Big Data
- Opportunities are out there, especially for people who can work with Big Data

Resources

- APS career planning
  - http://www.aps.org/careers/guidance/development/intro.cfm
- Career profiles
  - http://www.aps.org/careers/physicists/profiles/index.cfm