PhDs in the Silicon Valley

Matt Pasienski

Illinois Physics Department September 6th 2012

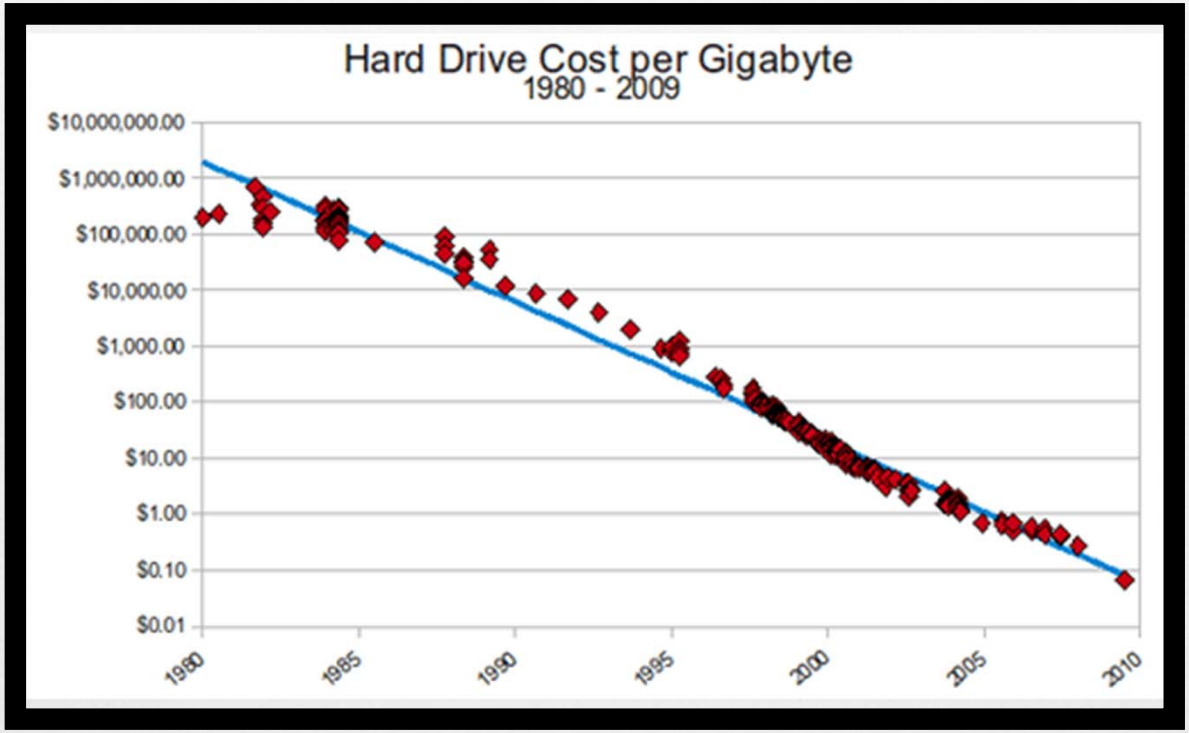
Who is this talk for?

- "As a student I would like a job that pays well and satisfies my intellectual and social needs."
- "As a professor I would like to secure influential and prestigious careers for my students."
- "As a scientist I would like to learn about a different industry that uses similar skills."
- "As a younger student I would like to learn about marketable skills in 'Big Data'."

Outline

- Section 1: What is “Big Data”?
- Section 2: Scientists in the Silicon Valley.
- Section 3: Securing a job in this industry.

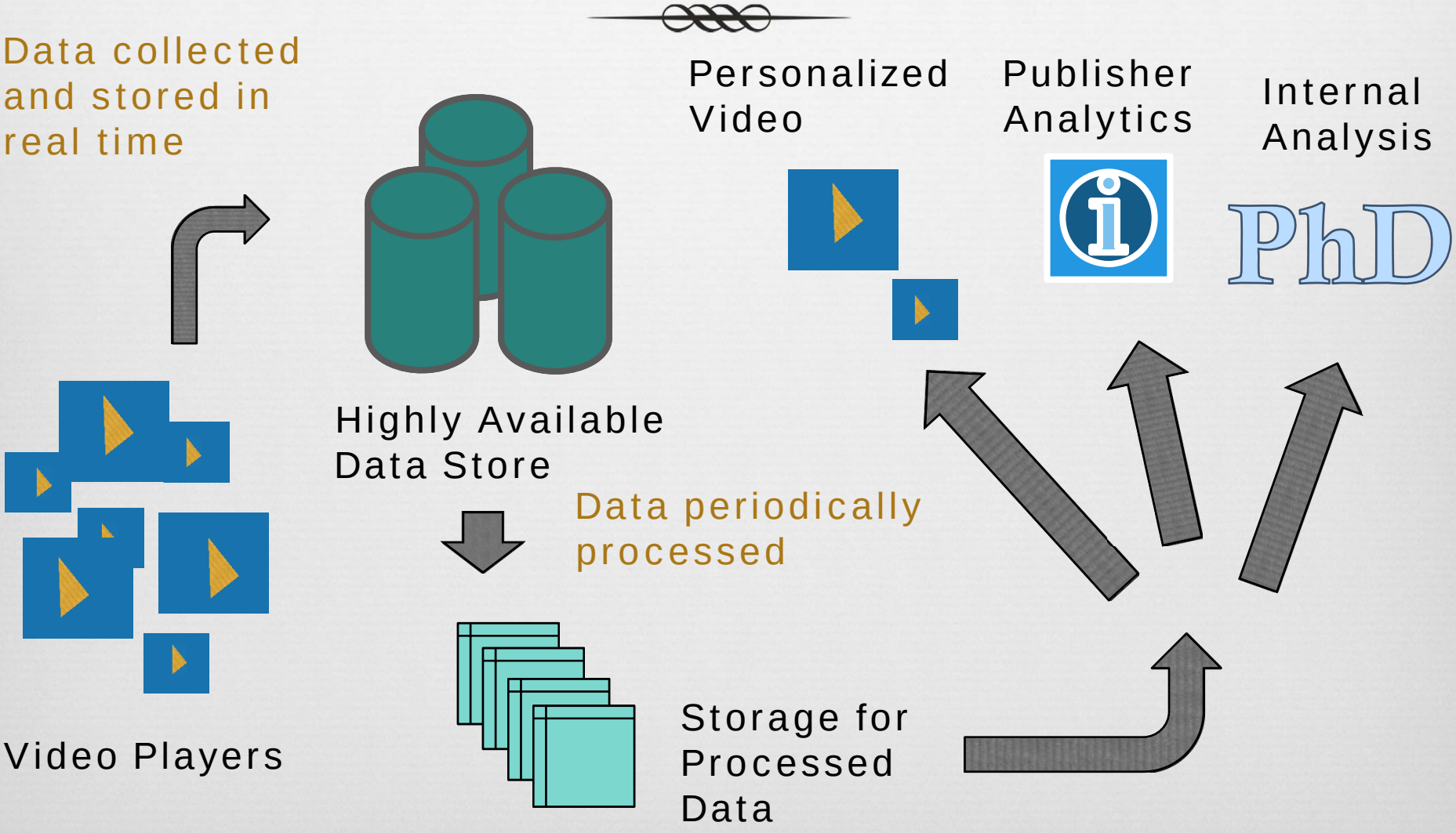
What is Big Data?



Memory is Cheap

So why not store everything?

A Big Data Example



Components of Big Data

Storage and Collection

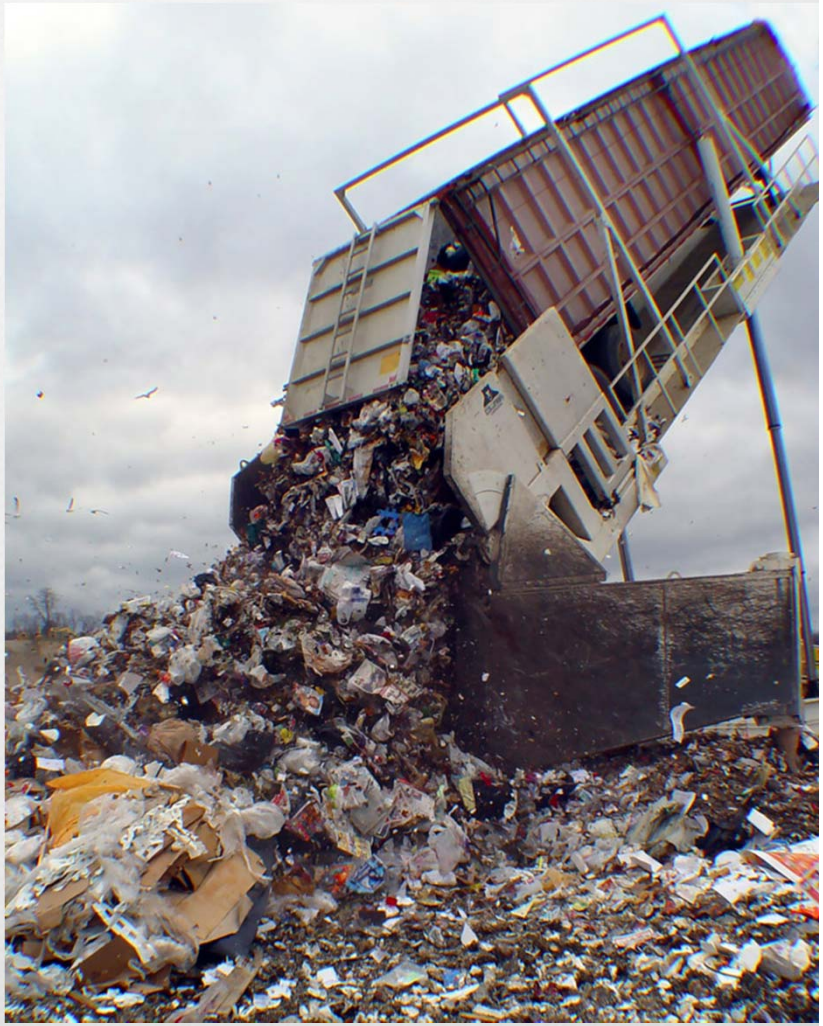
- High Availability
- Distributed Data Stores
  - Cassandra
- Automated data collection over the internet

Processing

- Distributed Computing
  - Hadoop - batched
  - Storm - streaming
- Analysts (this doesn't scale well)

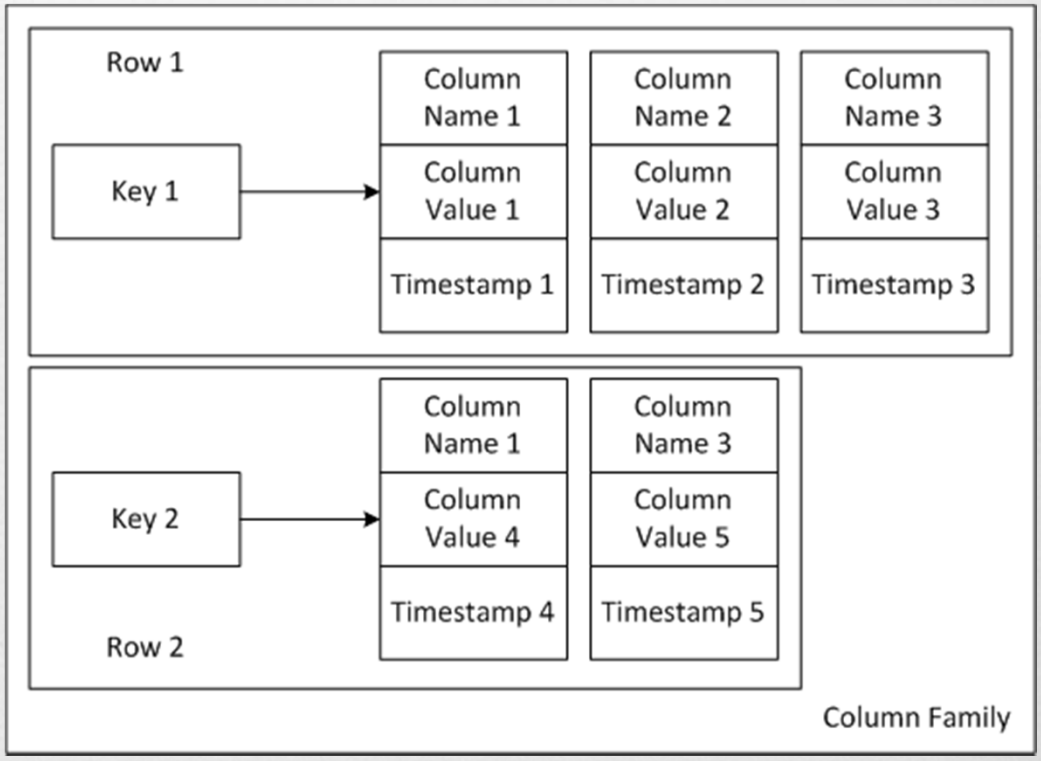
Hadoop

- Always writable
- Key, Value Pairs
- No structure
- Denormalized



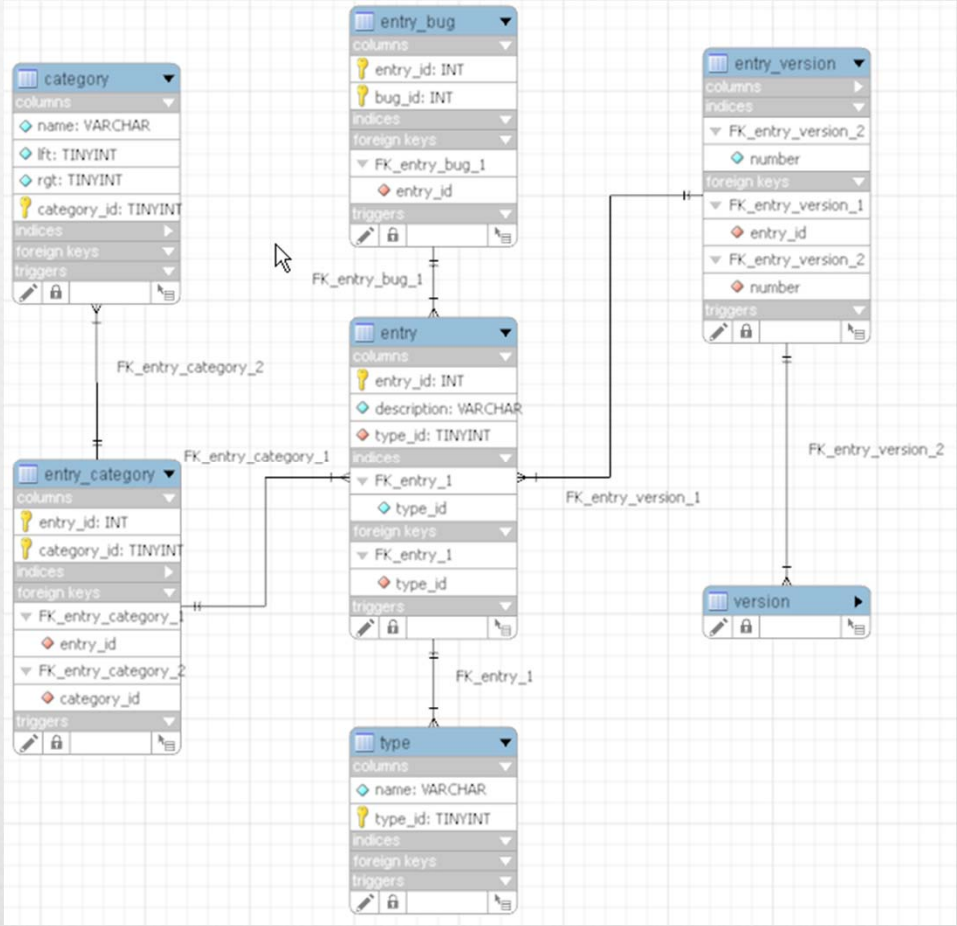
Schemaless Data Stores - Cassandra

- Structured around queries
- Distributed, denormalized
- Always readable
- Eventually consistant



RDBMS

- Difficult to scale
- Normalized
- Schema's allow complex data sctructures
- SQL – structured query language
- RDBMS – relational database management systems, e.g. MySQL



Digression

What do we mean by “Industry” and “Career”?

Scientists in the Silicon Valley

Job description

What do you do? What's the job like?

How I use my PhD

- Programming, math, stats every day
- Communicating complex information clearly to a diverse audience
- Work ethic
- People in tech respect an Illinois PhD

The Bay Area

Environment

- San Francisco
- Mountains
- Ocean
- Public Transportation
- Universities/Culture/Wealt
  People

Companies

- Big Data: Google, Twitter, Facebook, LinkedIn, Salesforce, Palantir, Seattle: Amazon
- Medicine: Genentech
- Hardware/Software: Apple, Seattle: Microsoft
- Startups

The Office

Google inspired work places are the norm

Free food!

Example Career Paths

- Management
  - Marketing, Pricing, Product Development, Recruiting
- Software Engineer
  - Building Products and Tools, Coding
- Entrepreneur
  - Fund Raising, Marketing, PR, Coding
- Actual Scientist
  - Fundraising, Turning Wrenches, Math, Physics

If I could go back to grad school…

- Record everything
- Learn to use databases, SQL is really useful
- ABCDE – Always be collecting data earnestly

Physicists could improve on…

- Shorter development cycles
  - Multiple papers per year, less time between starting a project and seeing the results
  - Place more emphasis on projects with quick returns, only do something big if it is really justified
- Do PR
  - More communication outside of papers and press releases
  - Really understand who you are trying to influence and why

Getting your first job

Making Connections

- Visit, make friends, get warm introductions
  - Talk to Matt or email him after the talk matt.pasienski@gmail.com
- Get an internship
  - Its ok to take time off
  - Less pressure on both employer and employee
- LinkedIn is effective at finding openings

Resume

- “Data Science”, “Machine Learning”
- Emphasize experience with programming languages, data bases, analysis techniques, distributed computing
- You’ve built amazing projects, emphasize the responsibility and scope of what you can accomplish without much supervision
- Statistics and programs like R

Consider the following

Physics

- Become the best in the world in an important subject
- Few job opportunities
- Many more years of investment before any payoff

Industry

- More pay, more flexibility
- Smaller faster projects
- More interaction with people
- Opportunities to branch out into other parts of business

Come Talk!

I'll be in Loomis 265 Thursday and Friday afternoons. Send me an email at matt.pasienski@gmail.com.