Understanding Cloud Data Tools and Complexity

John R Williams
http://Geospatial.mit.edu
Stockfish > 3500  Carlson ~2900
Transformation of Workforce

- Technology and Tools
- Managing Change
- Levers of change.
  - Command and Control vs Self Organization
Business Intelligence Pipeline – Modern Stack

Online Apps

Google Sheet

Connectors

BigQuery

Transform

GitHub

BI Tools & Reports

Google

Tableau

Power BI

Looker
ML Model Development Pipeline

Online Apps → Google Analytics → Connectors (Fivetran, DataBricks, Delta Lake) → Feature Engineering → Models → MLOps
# Chicago Taxi Earnings

**SUM of trip_total - Day of the week**

<table>
<thead>
<tr>
<th>Day</th>
<th>Sunday</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>24,031,466</td>
<td>10,988,478</td>
<td>9,364,401</td>
<td>9,634,370</td>
<td>11,306,717</td>
<td>14,868,929</td>
<td>22,157,487</td>
</tr>
<tr>
<td>1</td>
<td>22,742,268</td>
<td>7,393,945</td>
<td>6,045,628</td>
<td>5,989,475</td>
<td>7,241,932</td>
<td>10,288,987</td>
<td>20,237,402</td>
</tr>
<tr>
<td>2</td>
<td>19,658,139</td>
<td>4,610,916</td>
<td>3,572,900</td>
<td>3,540,671</td>
<td>4,431,481</td>
<td>6,645,357</td>
<td>17,084,502</td>
</tr>
<tr>
<td>3</td>
<td>16,024,603</td>
<td>3,409,553</td>
<td>2,448,541</td>
<td>2,556,626</td>
<td>3,223,240</td>
<td>4,583,376</td>
<td>12,491,021</td>
</tr>
<tr>
<td>4</td>
<td>11,370,029</td>
<td>5,035,806</td>
<td>3,181,744</td>
<td>3,222,854</td>
<td>3,623,088</td>
<td>4,587,944</td>
<td>7,909,446</td>
</tr>
<tr>
<td>5</td>
<td>8,364,281</td>
<td>8,830,371</td>
<td>5,611,930</td>
<td>5,554,700</td>
<td>5,797,426</td>
<td>6,374,355</td>
<td>5,519,383</td>
</tr>
<tr>
<td>6</td>
<td>7,286,898</td>
<td>11,915,338</td>
<td>8,913,596</td>
<td>9,007,495</td>
<td>9,209,259</td>
<td>9,348,182</td>
<td>5,679,418</td>
</tr>
<tr>
<td>7</td>
<td>7,638,419</td>
<td>15,439,263</td>
<td>14,373,713</td>
<td>14,810,192</td>
<td>14,960,192</td>
<td>14,340,212</td>
<td>7,095,267</td>
</tr>
<tr>
<td>8</td>
<td>8,809,245</td>
<td>20,796,907</td>
<td>21,582,768</td>
<td>21,763,421</td>
<td>21,815,308</td>
<td>19,925,585</td>
<td>9,105,243</td>
</tr>
<tr>
<td>9</td>
<td>11,496,373</td>
<td>24,480,081</td>
<td>25,116,680</td>
<td>25,451,283</td>
<td>24,850,716</td>
<td>23,125,604</td>
<td>12,325,742</td>
</tr>
<tr>
<td>10</td>
<td>14,949,368</td>
<td>23,747,909</td>
<td>23,801,653</td>
<td>24,244,814</td>
<td>23,581,076</td>
<td>23,181,796</td>
<td>14,989,227</td>
</tr>
<tr>
<td>11</td>
<td>18,017,531</td>
<td>23,951,869</td>
<td>24,198,592</td>
<td>25,239,104</td>
<td>24,997,237</td>
<td>25,619,831</td>
<td>17,163,651</td>
</tr>
<tr>
<td>12</td>
<td>19,687,628</td>
<td>24,727,287</td>
<td>25,251,578</td>
<td>27,072,914</td>
<td>27,367,479</td>
<td>28,683,365</td>
<td>18,927,695</td>
</tr>
</tbody>
</table>
ROBOT RIGHTS: US court rules AI machines ‘are not human beings’ and cannot claim rights over their inventions
MetaHumans by Epic
White House unveils ‘AI bill of rights’ as ‘call to action’ to rein in tool
Github Copilot - GPT3 Machine Learning Help

```html
<html>
<head>
  <title>My First Web Page</title>
</head>
<script>
  let api = "https://pollysnips.s3.amazonaws.com/users.json";
</script>
</html>

"https://pollysnips.s3.amazonaws.com/users.json"
Cloud Services

SaaS – Software
FaaS – Functions
DaaS – Data
PaaS – Platform
STaaS – Storage
IaaS – Infrastructure
Tooling has changed

AWS Ground Station Service [https://www.youtube.com/watch?v=JIbIYCM48to](https://www.youtube.com/watch?v=JIbIYCM48to)
Automated Robotic Labs – Strateos
GPT-4 Will Have 100 Trillion Parameters — 500x the Size of GPT-3
1900 – World knowledge was double that of 1800.
World knowledge was doubling in 25 years.
Today knowledge is doubling every 18 months
Knowledge Doubling Curve

Buckminster Fuller

Until 1900 human knowledge doubled approximately every century.

By 1945 knowledge was doubling every 25 years.

Currently, on average human knowledge is doubling every 13 months.

IBM predicts the build out of the “internet of things” will result in the doubling of knowledge every 12 hours.
Time to Value

DevOps

The fast companies are **440x** faster than the slow

We found that, compared to low performers, high performers have:

- 46 times more frequent code deployments
- 440 times faster lead time from commit to deploy
- 170 times faster mean time to recover from downtime
- 5.0 times lower change failure rate (1/5 as likely for a change to fail)

Months ⇔ Hours

If we gave every employee in MIT (~10,000) an app that generated a random deck of cards every second. How long would it be before we get a repeat of the same order?
1.85e+50 Ages, where an Age is the age of our universe
Hume’s “is” vs “ought” dichotomy

Science and Trans-Science
DATA but from Chaotic Systems
THE BLACK SWAN

The Impact of the Highly Improbable

“The most prophetic voice of all.”
—GQ

Nassim Nicholas Taleb
Market Crash May 6, 2010 - $1 Trillion lost in 10 mins
Speed of Learning
Time to value (2018 DORA Report)

The elite companies are 2555x faster than the slow

Compared to low performers, high performers:
- 46 times more frequent code deployments
- 2555 times faster lead time from commit to deploy
- 2604 times faster mean time to recover from downtime
- 7 times lower change failure rate (1/7 as likely for a change to fail)

The elite are widening the gap
RotiMatic Bread Maker - Designing MVP to Learn Fast
https://www.youtube.com/watch?v=fKhrlO0KGVI
Learn what your customers want – let them build it online
Organization Design Challenge - Human + Machine

Command & Control (Top-down) or Self Organizing (Bottom-up)
“I’m an Architect, not a Captain.”
Imitation
Stigmergy – leaving signals for others to follow

Open Source – Millions of Libraries
Companies as complex human systems

challenges

motivations

needs
View companies as Complex Adaptive Systems
What Framework?

What framework depends on knowledge of Problem and knowledge of Solution space.

- John Thomas and Pam Mantri, Cognitive Tools Inc.
Different systems require different management practices.

David Snowden
Cyber attacks for sale on the web (I)

- Hacking a Facebook or Twitter account: $130
- Hacking a Gmail account: $162
- Hacking a corporate mailbox: $500
- Scans of legitimate passports: $5 each
- Windows rootkit (installing malicious drivers): $292
- Winlocker ransomware: $10–20
- Unintelligent exploit bundle: $25
- Intelligent exploit bundle: $10–3,000
- Traffic: $7–15 per 1,000 visitors for the most valuable traffic (from the US and EU)

Note: Ars Technica, Wired, Russian Underground 101, Trend Micro, Max Goncharov
Case Study

Testing is part of a continuous pipeline
Testing in the cloud using Mabl

...but it’s bursty concurrency

Concurrency needed over the course of a day - by test window