Well We Don't Know …

What We Don't Know

Leveraging data to disrupt competitors, derive insight, and generate actionable results

Jim Fletcher

CMP Associates, LLC

May 2019

Jim.fletcher.cmp@gmail.com

We know what we know …

We know what we don’t know …

But we don’t know what we don’t know …

Donald Rumsfeld, US Secretary of Defense

Intelligence is the new Disruptor

- Effective use of data is the new differentiator across industries
- Data, and more importantly, the ability to effectively manage and index, and derive insights from that data is key
- Leading companies are moving to monetize their intelligence at new levels
- “As a service” driven by data is a game changer allowing companies to transition from CAPEX to OPEX driven models

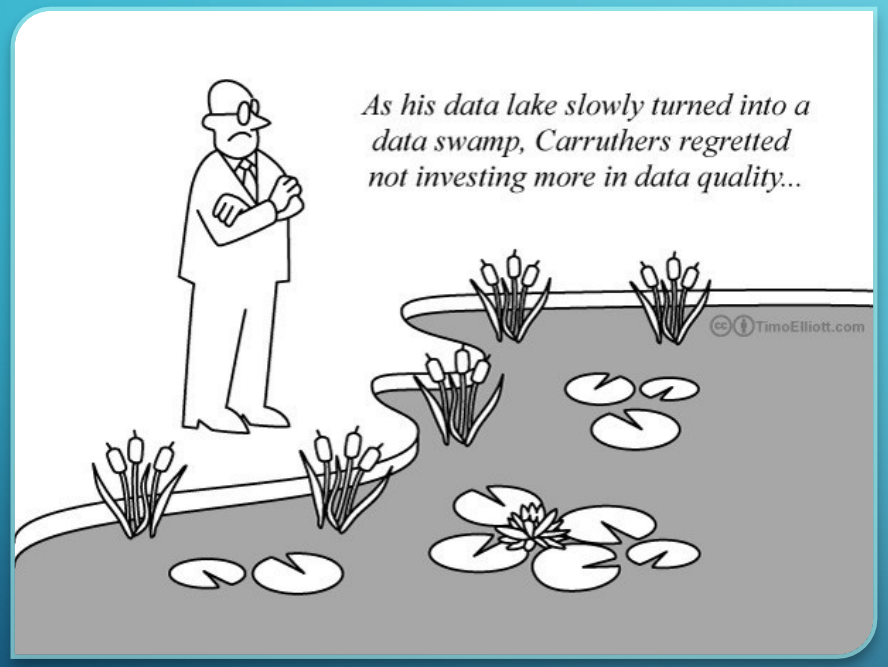
Data Labeling

- Data collection frequency rapidly growing
- Breadth of data collected is growing
- Most “legacy” data today continues to be raw and unformatted
- Data relationships are developed by humans, and data structure is overlaid
- Moving to era of self-labeling for “structured” data
- Haystack has developed a solid approach to labeling and data organization for structured data

Emerging Data Sources and Services

- Unstructured data is providing significant new insights
- Data from text, voice, audio, and video
- Sound patterns, NLP, sentiment analysis, image recognition, quality control analysis, etc.
- Rapid growth in cloud services for effective processing of unstructured sources
- Provide augmentation for traditional structured sources to derive additional data insights

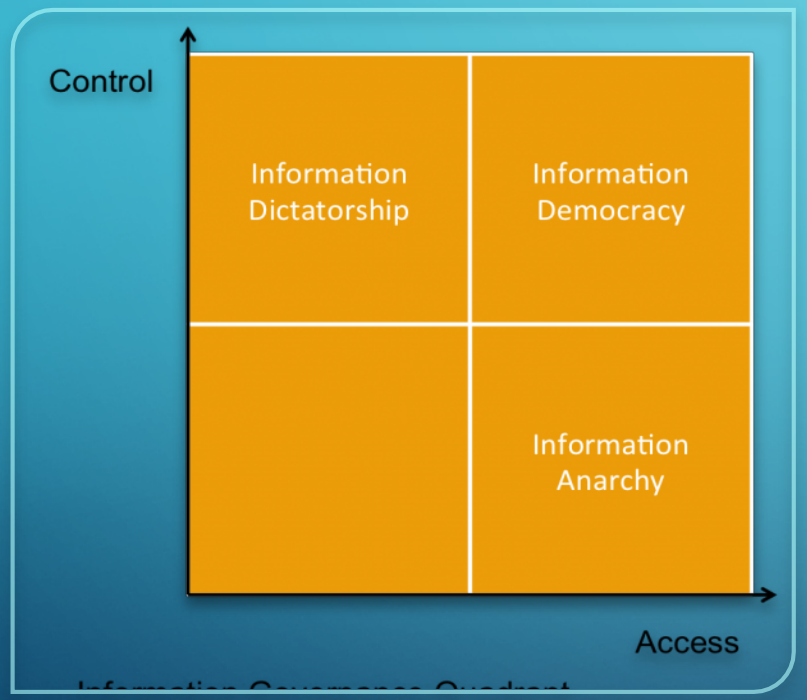
Data Lake or Data Swamp



- Rapid enablement of new era of data insights driving new requirements for data management
- Lower storage costs driving reduced focus on efficiency
- GPUs and TPUs providing more cost effective data mining for AI/ML solutions

- More data does not mean better insights
- Most data has a shelf life with decreasing value
- Data is the most valuable asset to any company yet Gartner estimates 90 percent of all data collected is unused

Data Democracy



- Finding the balance between open access and security controls is an increasing challenge
- Data siloes as a proxy for security significantly limit data value
- The value of the data is a direct function of how easily it can be leveraged by the average user, not the data scientist
- As we move from an era of business intelligence (BI) to one of Artificial Intelligence (AI) our approach to data must evolve
- Data quality must be a key element of the data ingest process, not a post-storage access level service

Data Monetization

- Data monetization should be a new revenue stream
- Most companies continue to build “things” and give away “insights”
- Data can be monetized through improved internal efficiencies, improved customers experience, and new product values from derived insights
- Data aggregation is critical to higher value monetization driven by richer insights
- Don’t narrow your thinking to just data you create!

Rules and Regulations

- Personally identifiable Information (PII)
- Country and state specific rules like GDPR
- Historical internal data rules
- Traditional Organizational boundaries
- ”Power centers”

Data Visualization

- Data should be visualized in a form factor that maximizes the derived value of the data
- Data visualization should encompass data context
- Data visualization must enhance the value of the data, not simply visualize the data itself
- Consider new use cases such as worker safety, less skilled worker, etc.

Digital Twin

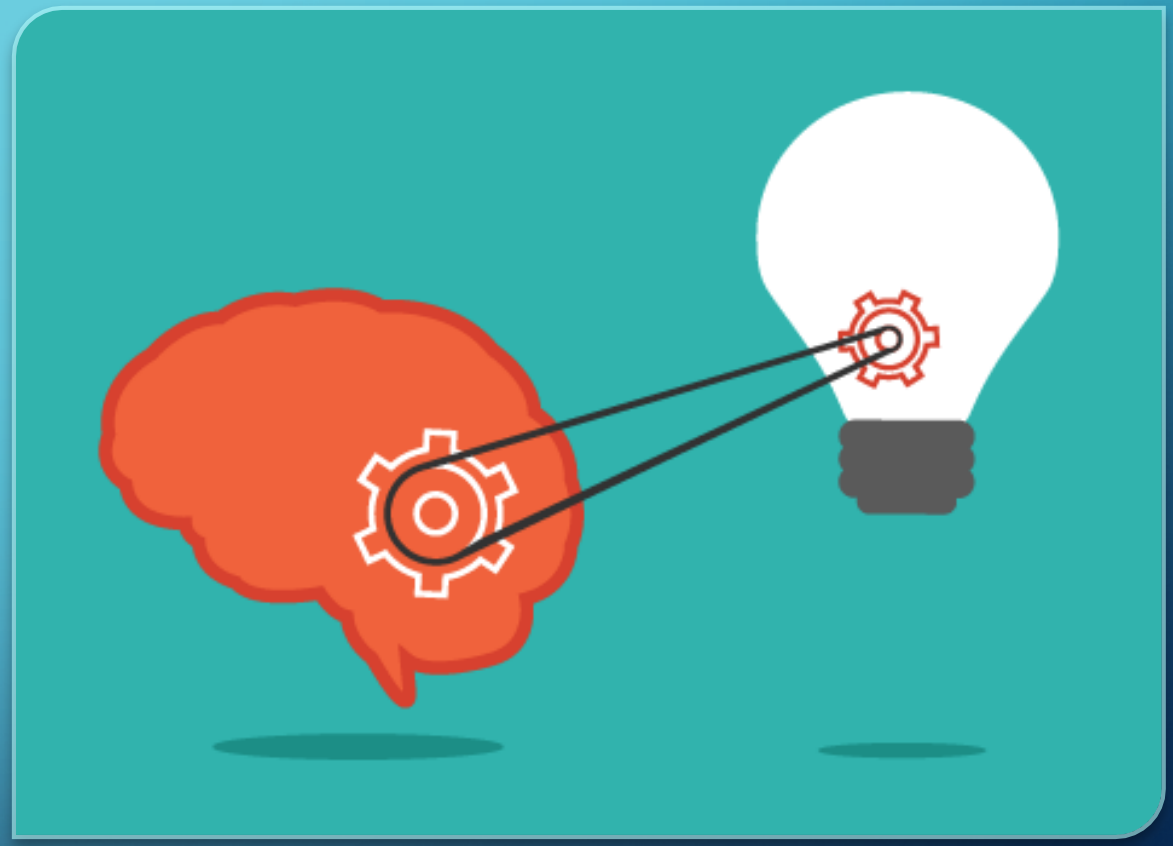
- Leveraging data to generate model of physical resource
- Powerful means to drive “what if”
- Provide virtualized lab environment
- Rich visualization via AR and VR technologies
- Identifying issues, and understanding problems before they happen in the “real world”

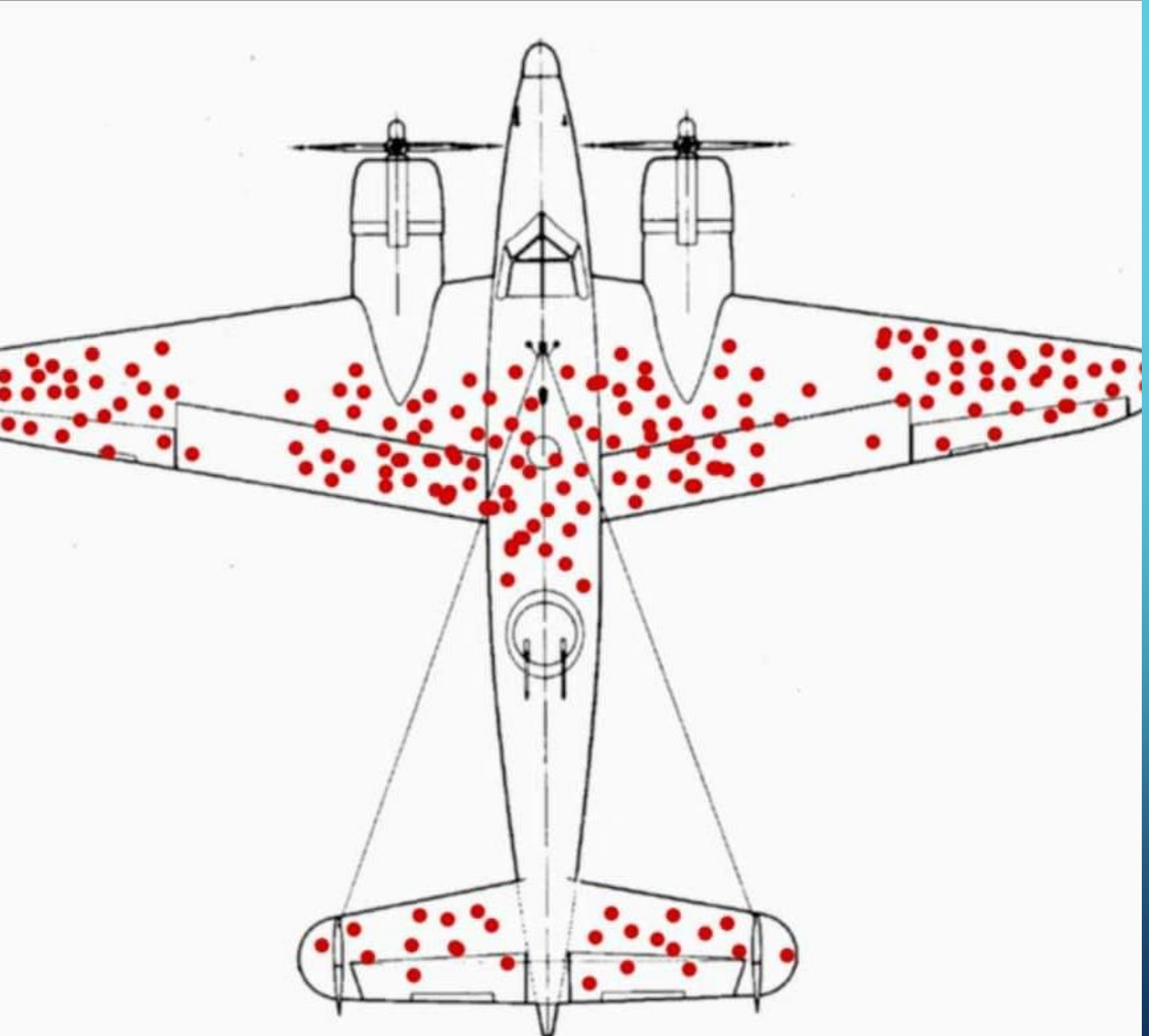
Deriving New Insights

- Moving from era of validation of assumed results, to deriving new, sometimes unexpected, insights
- Model training dependent on understanding what one believes the data actually represents
- Critical that training data represents breadth of expected possibilities otherwise the model will only reflect subset of reality

Autonomy, AI, and ML

- Successful autonomy requires **complete** understanding of **all** potential situations
- Early successes primarily around tight closed loop systems
- Success highly dependent on quality of data ingested, and quality/completeness of training data
- AI/ML success will be a continual journey





The Reality of Poor Training

- In World War II, airplanes that returned from battle were examined to determine where they needed additional armament
- Where do you think they recommended to improve the planes?
- What is missing from the training data?

The Data Persistence Paradox

At the edge

In the cloud

In the datacenter

Near the analytics services

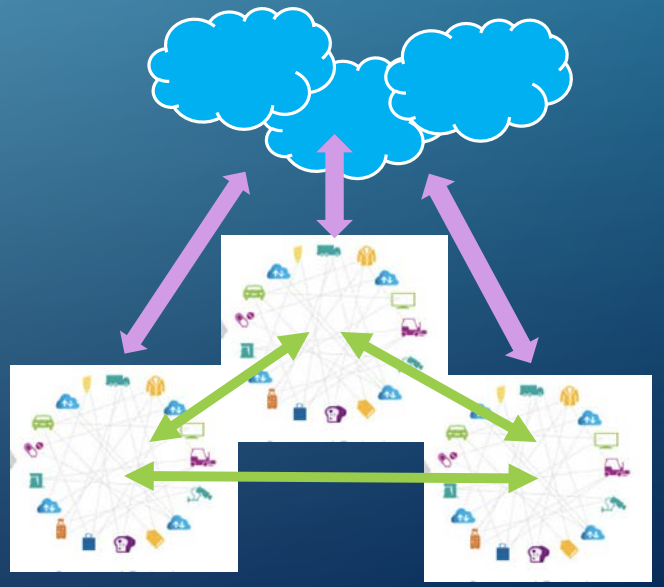
Point of lowest latency

Where it cost the least to store? To retain? To process?

Gartner predicts that by 2022, as a result of digital business projects, 75% of enterprise-generated data will be created and processed outside the traditional, centralized data center or cloud, which is an increase from today's less than 10%

Evolving Architectures

- We need an aggregate model, encompassing the isolation of the edge, with the power of selective aggregation and inclusion of services regardless of their physical location
- “Edge clouds” are aggregations of processes that can operate autonomously and also can share/learn with other hives and clouds
- Can be physical or logical aggregations of nodes or services connected as mesh, peer, or hierarchical
- Micro-service containers easily can be deployed and aggregated across nodes



Thoughts …

Our real value will only be seen as business transforms methods and processes by leveraging data to drive new “actionable insights”

Success is not measured in the “volume of data” collected

Focus must be maintained not on the technology being deployed but on the business value derived

New ”unexpected” insights will be found through aggregation of seemingly unrelated data sources