Future of BIG DATA at State of Michigan

MISA Conference, September 15, 2016
Bellaire, Michigan

DTMB

ENABLING MICHIGAN'S REINVENTION

Talking Points

- Why do I need to know about this?
- Big Data – Attributes and Review
- Public Sector Use Cases (Michigan)
- How do I get started?
- What mistakes should I avoid?
- Whom to trust?
- References

My sincere thanks to Paul Groll, Deputy CSO for the content of this presentation.

Big Data Attributes

Big data **is a term for data sets that are so large or complex that traditional data processing applications are inadequate to deal with them. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying, updating and information privacy. The term "big data" often refers simply to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set.[2][*need quotation to verify*] Accuracy in big data may lead to more confident decision-making, and better decisions can result in greater operational efficiency, cost reduction and reduced risk - Wikipedia**

Big Data Attributes - continued

- **The Rules of V's – Volume, Velocity, Variety, Veracity = Value**

Volume

The quantity of generated and stored data. The size of the data determines the value and potential insight- and whether it can actually be considered big data or not.

Variety

The type and nature of the data. This helps people who analyze it to effectively use the resulting insight.

Velocity

In this context, the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development.

Variability

Inconsistency of the data set can hamper processes to handle and manage it.

Veracity

The quality of captured data can vary greatly, affecting accurate analysis.

Big Data Review

- **What do we want from Big Data?**

Big Data Points

"Gartner forecasts that 6.4 billion connected things will be in use worldwide in 2016, up 30 percent from 2015, and will reach 11.4 billion by 2018,"

From 2013 through 2020, Gartner expects IoT endpoints to experience an annual growth rate of 32%, and for endpoint spending to be dominated by connected cars and machinery, such as commercial aircraft, as well as farming and construction equipment.

...Bombardier showcased its C Series jetliner that carries Pratt & Whitney's Geared Turbo Fan (GTF) engine, which is fitted with 5,000 sensors that generate up to 10 GB of data per second. A single twin-engine aircraft with an average 12-hr. flight-time can produce up to 844 TB of data.

Aviation Week Jan 2016

Public Sector Use Cases (Michigan)

- What's happening at State of Michigan?

✓ Transportation – Stress, Load, Volume

✓ State Police – Body & Vehicle Video

✓ Health and Human Services – Fraud Analysis

✓ Treasury – Revenue Recovery

✓ Information Security – Predictive Analysis

EIM Today @ SOM (Foundational)

EIM is a **Program Team** – Steering Committee, Brom, Rob, Virginia, PwC Team, GIS Team

6

1

EIM is a set of state-wide solutions

Identity Master | Location Master | Open Data | Business Glossary

2

Citizen Master Repository | Location Master Repository | Data Dictionary and Metadata Repository

EIM is building a set of statewide repositories

3

EIM is the State's 'official data integration' vehicle

EIM is the data foundation layer to multiple state projects

Traffic Fatalities Reduction | 21CN Infrastructure

Integrated Service Delivery | MiPage

Enterprise Data Services

Enterprise Information Management

Citizen Services

4

EIM is a methodology

The data component of all projects should be formalized

EIM proposes to embed itself in the State's SUITE methodology. In the interest of success of the project – the data component must be given the critical attention it deserves

5

EIM is a guidance and consulting service

Through its experiences and connections, EIM is able to provide guidance and consulting services to all departments, agencies or offices that need help with 'data matters':

- Could be about Analytics
- Could be about cross-agency data integration or data sharing.
- Could be about managing and organizing data
- Could be about setting up departmental data governance.
- Or setting up programs on any of the topics listed on this slide

Courtesy – Virginia Hambric

State Police Use Case (Possible)

- State Police – Body & Vehicle Video
- Big Data Tools:
  - Backhaul
  - Staging (Data Lake)
  - Analytics resident on Hadoop (many)
  - Less expensive Long Term legal retention
  - Solutions scale to Petabytes per year
  - Many Cloud solutions are appearing in this space – Beware Chain of Custody

How do I get started?

- Assemble a small cluster (3-5 nodes)
- Install and run I/O tools (Kafka, Flume, etc.)
- Partner with a vendor (free short-term trial)
- Try it in the cloud first (30-day free trials)
- Some can run natively on Isilon
- Lots of training on YouTube
- Be sure to experiment with various sources
  - Try a variety of data streams, formats

What mistakes I should avoid?

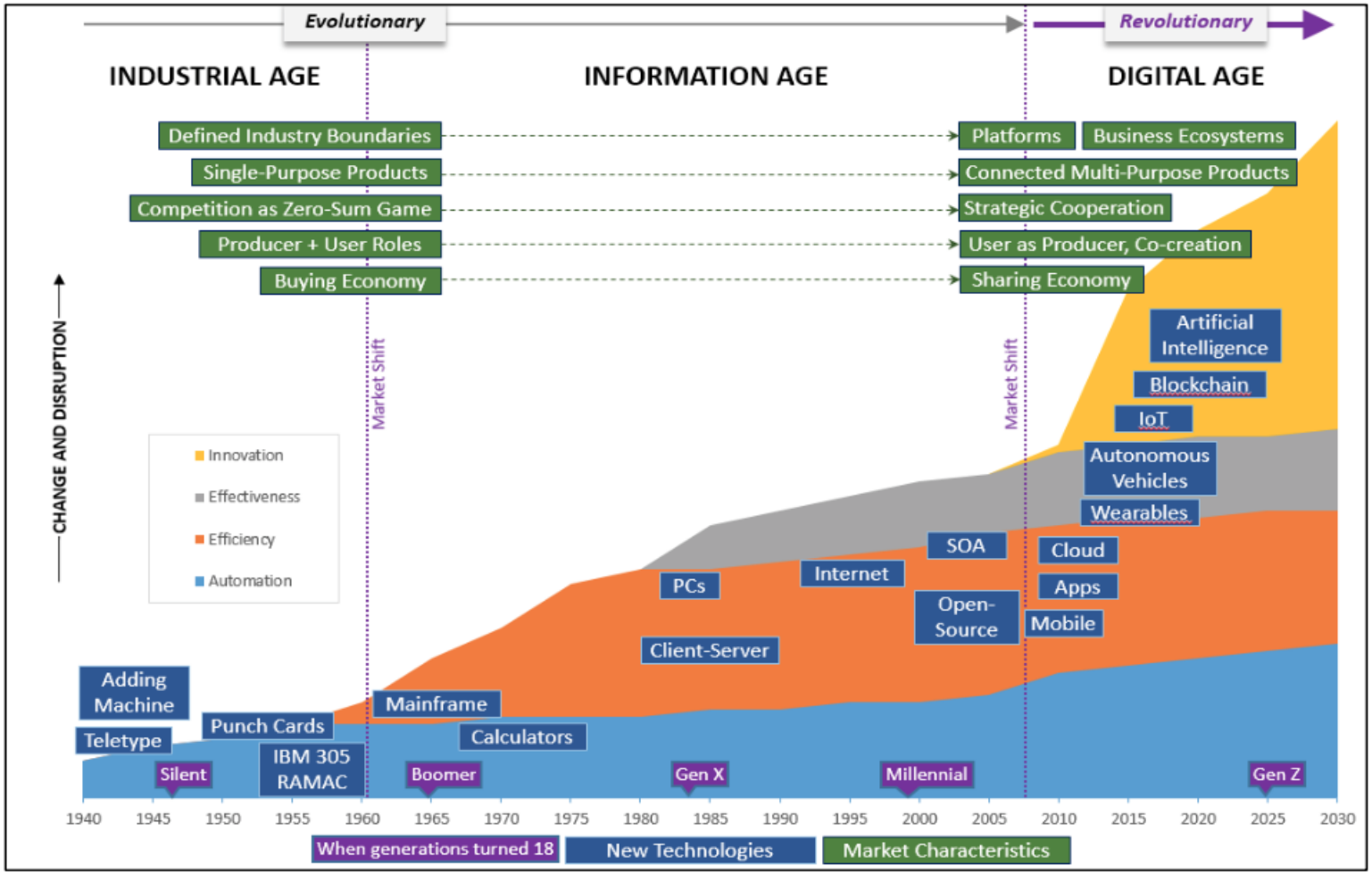
- Don’t start
- Start too big
- Start without executive support and KPIs
- Start without a Use Case, Pilot Charter, clear expectations, and success criteria
- Start without ODPI (next slide)
- Don’t be afraid to run a true Pilot – try things out for a few months, then shut it down. Make the most of Lessons Learned.

Resources

- Open Data Platform Initiative / www.odpi.org / “The open ecosystem of big data”
- The Data Warehousing Institute
- System Integrators and Software companies – IBM, SAS, Oracle, PwC, Deloitte, Accenture, etc.
- Local – Machine Learning Chapter @ Lansing


Q&A

References


Evolutionary
Revolutionary
INDUSTRIAL AGE
INFORMATION AGE
DIGITAL AGE
Defined Industry Boundaries
Single-Purpose Products
Competition as Zero-Sum Game
Producer + User Roles
Buying Economy
Platforms
Business Ecosystems
Connected Multi-Purpose Products
Strategic Cooperation
User as Producer, Co-creation
Sharing Economy
CHANGE AND DISRUPTION
Market Shift
Market Shift
Innovation
Effectiveness
Efficiency
Automation
Artificial Intelligence
Blockchain
IoT
Autonomous Vehicles
Wearables
SOA
Cloud
Internet
PCs
Apps
Open-Source
Mobile
Client-Server
Adding Machine
Mainframe
Punch Cards
Teletype
Calculators
IBM 305 RAMAC
Silent
Boomer
Gen X
Millennial
Gen Z
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
2030
When generations turned 18
New Technologies
Market Characteristics

References

4 MINUTE READ | LEADERSHIP

How Too Much Data Can Hurt Our Productivity And Decision-Making

Data findings are only valuable when they reliably change someone's behavior for the better.

"…data analytics is only valuable when it changes someone's behavior. "

http://www.fastcompany.com/3060945/how-too-much-data-can-hurt-our-productivity-and-decision-making - 16 June 2016