“Big” is a Relative Term (Context dependent)

IBM 5MB Hard Drive 1956

12th June, 2018,
Oslo, Norway
Presenter Profile - Dr. Bhuvan Unhelkar  
(BE, MBA, PhD, FACS, CBAP®, PSM)

- Professor and Lead IT Faculty, Univ. of South Florida, Sarasota-Manatee; Founder, MethodScience.com; PlatiFi.com;
- Courses: UML, Agile PM, Big Data; Designed and Presenter Australian Computer Society’s Business Analysis (BAS) and Agile PM Online
- Visiting Faculty at UTS (Agile SW Modeling) and WSU (UWS)
- Author: 20 Books (Including Big Data Strategies for Agile Business)
- Supervisor: 7 PhD Completions;
- Fellow of the Australian Computer Society; IEEE Sr. Member; Life Member, Computer Society of India & BMA
- President – Rotary Club of Sarasota Sunrise, USA (Paul Harris Fellow+4; AG); TiE ex-Mentor Director;
- www.unhelkar.com & www.methodscience.com

---

Today data scientist uses **Yottabyte** to describe how much government data the NSA or FBI have on people altogether.

In the near future, **Brontobyte** will be the measurement to describe the type of sensor data that will be generated from the IoT (Internet of Things)

![Diagram](www.theregister.co.uk)
Lean-Agile Tautology
by Bhuwan Unhelkar, Senior Consultant, Cutter Consortium

Agile Business Analysis:
Part II — Organizational Adoption with Centers of Excellence
by Bhuwan Unhelkar, Senior Consultant, Cutter Consortium

The Psychology of Agile:
Fundamentals Beyond the Manifesto
by Bhuwan Unhelkar, Senior Consultant, Cutter Consortium


Big Data Trends in 2018
by Arini Verma - January 23, 2018

Big data is no longer just considered as a big or large set of data. Instead, it is more about business-driven data in the use of analytics capabilities for long-term business value. However, the massive increase in the data volume, variety, and speed of data for analytics becomes clear that big data trends in 2018 will change a worth amount.
Big data: The next frontier for innovation, competition, and productivity

http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Big data can unlock significant value by making information transparent and usable at much higher frequency.

As organizations create and store more transactional data in digital form, they can collect more accurate and detailed performance information.

Big data allows ever-narrower segmentation of customers and therefore much more precisely tailored products or services.

Sophisticated analytics can substantially improve decision-making.

Big data can be used to improve the development of the next generation of products and services.
What is Big Data?

‘Big Data’ is still ‘Data’
- But difficult to process with the given tools and technologies
- Could be made up of much wider ‘types’ or ranges of data

Handling such Data requires Latest:
- Techniques, tools, and architecture
- Frameworks to guide adoption and reduce Risks

With an aim to:
- Solve new challenges in Business
- Improving the Value proposition
What is Big Data?

could dance on the head of a pin. Rob Lass, cotounder of Big Data pioneer Splunk, begins most presentations sharing results of a national poll asking people to respond to the question, “What is Big Data?” Respondents gave the following answers:

- A legitimate problem stemming from the growth of unstructured data (51%)
- A new catchphrase for an old data management challenge (23%)
- A meaningless marketing catch-all (9%)
- Another way to say “data warehouse” (5%)
- Another way to say “Hadoop” (3%)

What has Enabled Big Data to Happen?

Key enablers for the appearance and growth of Big Data are

- Increase in storage capacities in devices & on the Cloud
- Increase of processing power (CPU)
- Increase in Generation (and Availability) of data by businesses, individuals and machine sensors
- (Every day we create 2.5 quintillion bytes of data; 90% of the data in the world today has been created in the last two years alone)
Some interesting Facts About Big Data: (Not that it matters too much!)

100 hours of video uploaded to YouTube every minutes
Can Take 15 years to watch all those Videos
NSA analyze 1.6% of all global internet traffic around 30 petabytes or 30 million gigabytes every day
Retailer could increase profit by 60% through the full exploitation of big data analytics
(SeSense.com) > Nearly 2.3 trillion gigabytes of data created every single day and the data universe doubling every two years
Big data industry is 54.3 billion in 2017
The Key Business Challenge

Big or Small – Business is NOT interested in Data

Business interest are as follows:
- Can Data be used to “Make sense” of a situation?
- Can Data enhance Decision Making?
- Can Data be Managed? (i.e. Security & Privacy)
- Is there a Business “Value” in Data?
Figure 1.1 Big Data Strategies – Transcending Analytics and Technologies

Strategies (High Level):
- Business/End-User Value;
- Make sense of existing data;
- Spotting Patterns;
- Identifying new data;
- Mapping (Unstructured) to a Structure;
- Optimizing Business Processes;
- Contextualizing Enterprise Architecture;
- Sustainability Value;
- Balancing the ROI.

Analytics

Technologies

Sentiment, Predictive, Prescriptive, (NPS, Text Mining)

Strategies

Decision Maker; End-User; Manager; Owner; Worker; Government; Community; Environment

Agile Business

Application in Business
Building the Capabilities

Hadoop/Spark
NoSQL
Databases; MapReduce

Storage & Backup; Sharing; Interfacing; Security and Privacy;

SMAC-stack

Early literature on big data discusses the “three Vs”: volume, velocity, and variety. Adding the “fourth V” — versatility (or quality) — creates the contemporary description of big data. These characteristics of big data are made possible due to Hadoop-based technologies and corresponding sophisticated analytical algorithms. Big data technologies (or analytics) in themselves are not capable of providing strategic value to business. In fact, big data in itself does not guarantee business success and, as highlighted in my last Executive Update, can be described as something I call the “big data strategy lacuna.”
Analytics & Technologies

MACHINE LEARNING;
ARTIFICIAL INTELLIGENCE

Categories providing Agile Business Values (and form basis of Business Strategies)

- Informative: Presentation of Notifications and other information with minimal analytics
- Descriptive: Data analysis to ascertain and describe the current situation
- Diagnostic: Identification (spotting) and pinpoint description of a (hidden) Problem
- Predictive: Spotting Trends and Patterns in order to anticipate future events
- Prescriptive: Provide Suggested Actions (Decision Support)
- Exploratory: Analytics with Unknown Situations to arrive at possible New Revelations
- Reactive: Analytics only in response to a situation; no anticipation

Growth (External): Optimization (Internal):
Sustainability & Environment: Risk & Compliance:
Figure 3.10: Identifying Associations and mapping Clusters
(Based on Agarwal, A., “Predictive Analytics and Text Mining”) – An increasingly finer granular Analytics (Amazon’s Long-tail Marketing: Products (A,B,C..) and Customers (X1, X2..)

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
<th>X7</th>
<th>X8</th>
<th>X9</th>
<th>X10</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

A Framework for Descriptive Analytics

Ack. Dr. A. Agarwal, USFSM
Univariate Descriptive Analytics

- Structured Data
  - Uni-variant
    - Quantitative
      - Measures of Central Tendency
        - Mean, Median, Mode
      - Measures of Variation
        - Range, Variance, Std. Dev., IQR
      - Measures of Location
        - Quartiles, Percentiles, Max, Min
      - Measures of Shape
        - Skewness, Kurtosis
      - Charts
        - Histogram, Frequency Polygon, Ogive, Box-Whisker
  - Qualitative
    - Ordinal
      - Measures of Central Tendency
        - Mode
      - Charts
        - Frequency Polygon (Ordinal)
    - Nominal
      - Measures of Central Tendency
        - Mode
      - Charts
        - Column, Pie
      - Histogram (Ordinal), Frequency Polygon (Ordinal)

Ack. Dr. A. Agarwal, USFSM

Machine Learning...

1. Consider what a Machine learns..... \( A + B = C \)
2. We then add conditions: If \( A \geq 5 \) THEN..... \( A + B = C \)
3. Machine Algorithm: What are my options when \( A \geq 5 \); and when \( A < 5 \);
4. What had happened in the past? (Examining Historical Patterns)
5. When \( A \geq 5 \) and \( B = 'n' \), (where ‘n’ ranged from –10 to +25), C was Positive for 1 BILLION records
A Common Thread: Applying Hex Externalization in IoT Data Analytics
by Lin's Hsiang and Yi-Chen Lin

IoT Data Management and Analytics:

Realizing Value from Connected Devices

Opening Statement

By Mike Lohman and Ken Babcock

IoT Data Management Challenge: A View from the Trenches

By M. R. Case

A Common Thread: Applying the Hexadecimal

By Lin's Hsiang and Yi-Chen Lin
Figure 3.6: Hex-Elementization as a mechanism for Context of a Data Point (Based on research conducted by Girish Nair, WSU, Australia)

Figure 3.7: Journey of a Data Point via Context and Analytical Engines and the Granularity-driven Feedback Loop
Figure 3.8: The concept of Granularity in Analytics and the Factors in Ascertaining the Optimum Granularity Level (OGL)

Finer Granularity

- Business Decision Making
- Business Agility

Ch 1   30

Context

Volume – Velocity – Variety

Optimum Granularity Level (OGL)

- Costs
- Technology
- Risk
- Currency of Analytics
- Population of Users
- Volume
- Velocity of Data

Business Decision (Strategic)

Exploratory

Predictive

Prescriptive

Finer Granularity

[Business Decision Making: Business Agility]
Figure 3.10: Identifying Associations and mapping Clusters
(Based on Agarwal, A., "Predictive Analytics and Text Mining" Presentation at The Suncoast Technology Forum, Sarasota, FL; February 16th 2016)

<table>
<thead>
<tr>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
<th>X7</th>
<th>X8</th>
<th>X9</th>
<th>X10</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Rapidly Changing Business Ecosystem (External Impact)

Enterprise Agility
(Big Data, Cloud Computing, Mobile, Lean Processes)

Business Agility is the Ability of the Business to Rapidly Respond to external and internal Changes
~from the Art of Agile Practice

Business (Enterprise) Agility is the Rapidity and Accuracy of an Enterprise’s iterative and incremental Response to Fast changing External and Internal Situations

Business Agility is the Ability of the Business to Rapidly Respond to external and internal Changes
~from the Art of Agile Practice

Enterprise Response (Internal; Change Mgmt)
Sub-Module

Examples (Use Cases) in Practice

TOWARDS STRATEGIC USE OF BIG DATA

Farecast by Oren Etzioni
(Part of Bing now)

Crunches 200-billion price records
Saves $50 per ticket on average
  ◦ 75% of the time

Entirely based on Big Data technologies in the background
  ◦ Was first Major attempt at use of “Big Data” in 2004; Technology Followed Business
  ◦ Sophistication grew with Hadoop

Acquired by Microsoft for $110 M

Shift in Mindset in how Data is Used
23andMe.com (DNA exploration)

Gene sequencing as an Industry
- From a few thousand $ to $99

Hints at Predicting cancer, heart etc.

Need to sequence Billions of base-pair DNAs
- Not many are possible for this cost
- Steve Jobs’ sequenced entire Gene set ($N=All$)

Inrix.com (Traffic Scorecard)

Data is collected from 100 Million Vehicles

Real time Geo positions
- BMW, Ford and Toyota
- Collaborations for Value
- Fleet vehicles (Taxis)

Smartphone Apps
- Driver permits data collection
Predicting Crimes and Terrorist events

Big data has been used to predict crimes & disaster before it happens (e.g.
- Predictive policing; Enabling Evasive Actions;

Campaign Manager – 2008 elections
(Dan Siroker - Optimizely) 5 part strategy:
*For Success with Big Data*

1. Define Success
- What do we want from Big-Data for ‘business’

2. Question Assumptions
- About possibilities; correlations;

3. Divide and Conquer
- Data; processes; people

4. Take advantage of Circumstances
- Entire Data is DYNAMIC;

5. Start Today
- Truly Agile – Planning may be negative
Big Data Framework of Agile Business (BDFAB v2.5)

Examples of Technical, Analytical and Strategic Decisions related to Big Data

How many Nodes on a Cluster? How to distribute 1 Petabyte?
- MapReduce / HDFS / Spark (In-Memory)
- NoSQL (MongoDB) for non-Transactional Unstructured

What is the NPS of the new product launched last month?
- Why is the NPS so low? And which product parameters need to be changed to improve the score?

Customer Service Excellence
- Undertaking Business Process Integration with a NoSQL DB to enable decentralization in Decision Making
- Enhance Knowledge Sharing in the process to enable Business Agility
Figure 2.2 – Business Parameters and Maturity assessment of an organization provides an understanding of its Capabilities in Analytics and Technologies to create Big Data strategies that will handle Risks and provide Agile Value.

- **Sustainable**
- **Collaborative**
- **Directional**
- **Operational**

**Maturity**
- Initial

**Analytics Strategies Technologies**
- **Type (Product, Services)**
- **Size (Small, Medium, Large, Global)**
- **Nature (Primary, Secondary)**

Governance Functions around Risk Controls (e.g. maintenance of Risk Control Matrix, Reporting risks)

Risk Controls around operational, support & admin processes (e.g. User Profile creation & maintenance)

Risk Controls around the Analytical Services (e.g. context engine, Validations)

Value (Agility)

Insights

Visualize

Sense

Data Science supported by Enterprise Architecture (Big data based) is the key to Leadership in Business

Leadership (Strategies)

Big Data

Enterprise Architecture (Technologies)

- External: User Experience
- External: Market Expansion
- Internal: Operational Excellence
- Internal: Re-Engineering
- Both: Regulatory Compliance
- Both: Community, Environment
Approaching Big Data in a Strategic manner for Agile Business

### Envisioning a Holistic Big Data Strategy

- **Big Data Innovation (Business Models & Architecture; Enhanced Decision-Making)**
- **Big Data Analytics & Visualizations (Statistical Modeling; Probabilities & Causations)**
- **Big Data Technologies (Hadoop/Spark eco-system, NoSQL, Relational, Cloud)**
- **Big Data Management & Administration (Quality and Risks)**

Using Principles and Practices of Agility & Change Management for Business

- Customer-centric Agile; Incorporate Sentiments in Product/Service Development; Sustainability
- Developer-centric; Apply Agile (techniques) to develop solution
- Agile in Infrastructure; Maintenance; Operations; (DevOps)
- Agile in Business Functions (New & as Usual - BAU); Psychology of Agile; (Socio-Cultural)

Supporting Infrastructure, Operations and Tactics

The Strategy Cube: A 3-dimensional Big Data Opportunity Matrix based on Big Data technology and analytics, and Business Agility

- **Opportunities (External):**
  - Innovative Business Growth
  - Customer Satisfaction/ Experience
  - Innovative Products & Services
  - Mergers & Acquisitions; Collaborations

- **Optimization (Internal):**
  - Innovative Problem Solving
  - Internal Process Optimization & Maturity
  - Decentralizing Decision making (& Re-structuring)

- **Environment & Corporate Social Responsibility:**
  - Reduce Carbon Footprint
  - Enhanced Measurement & Analytics with Machine Sensors & Tools
  - Develop positive User Attitude

- **Risk & Compliance**
  - Business Risks (with, without) Big Data adoption
  - Security & Privacy Risks
  - Compliance & Documentation
  - Audits and Traceability

- **Value (Customer, User)**
- **Collaborate (Partners)***
- **Perceive, Anticipate**
- **Respond & Iterate**

**Hadoop/HDFS – NoSQL - “R”**

**Analytics & Fine Granularity**

**Social-Mobile (SoMo) – Internet of Thing (IoT)**

**Iterations & Increments (Agility)**
Opportunities (External):

Innovative Business Growth
Customer Satisfaction/ User Experience
Innovative Products & Services
Mergers & Acquisitions; Collaborations with External, Government, Third-party Organizations

Optimization (Internal):

Innovative Problem Solving
Internal Process Optimization & Maturity
Decentralizing Decision making (& Re-structuring of the organization to a Process-oriented flat, non-hierarchical structure)
**Environment & Corporate Social Responsibility:**

- Reduce Carbon Footprint
- Enhanced Measurement & Analytics with Machine Sensors & Tools to provide most current Data
- Incorporate Automated feedback loops
- Develop positive User Attitude

---

**Risk & Compliance**

- Business Risks (with, without) Big Data adoption
- Security & Privacy Risks
- Compliance & Documentation
- Audits and Traceability
The Big Data Manifesto

We will use Big Data Analytics and Big Data Technologies to provide Business Value to Users (Customers, Staff and Partners)

We will use Agile Principles to Dynamically create high Quality Business Processes that utilize Big Data and enable Business Agility

We will be Outcome-driven, Fail-fast, Collaborative, welcoming Change and Holistic in applying Big Data to Digital Business so as to Reduce Risks

We will use Big Data with due deference to Security, Privacy and Compliance needs of Individuals, Society and Government

We will embed Environmental and Social Responsibility in Big Data driven Digital Enterprise
Sub-Module

The BDFAB v2.5

A HOLISTIC RISK REDUCTION FRAMEWORK

© MethodScience; Material Designed and Presented by Dr. B. Unhelkar

Agile Business Values [7]

Key roles (Technical, Business) in Big Data & Agile [10]

Business Decision (Org. Maturity)
Data Science: Analytics, Context & Technologies
Business Processes & Fine Granular Decision Making
Enterprise Architecture, SMAC, Semantic
Quality, ORM, People Skills

Inputs and Outputs (Iterative Deliverables) [10]

Business Parameters (type, size, maturity)


X 12 Laned Adoption (Transformation) Process
Big Data Strategy Cube (The Opportunity Matrix)
Big Data Manifesto: x 5 Statements
Figure 2.3: SWOT Analysis (Example) of a Business Organization in the context of Big Data – and resultant Projects (Activities) with the backdrop of Business Architecture, Enterprise Architecture and Business Analysis
<table>
<thead>
<tr>
<th>Sub-Module</th>
<th>A Look at the Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>IN THE BIG DATA SPACE</td>
<td></td>
</tr>
</tbody>
</table>

© MethodScience; Material Designed and Presented by Dr. B. Unhelkar
Types of tools used in Big-Data

Where processing is hosted?
- Distributed Servers / Cloud (e.g. Amazon EC2)

Where data is stored?
- Distributed Storage (e.g. Amazon S3)

What is the programming model?
- Distributed Processing (e.g. MapReduce)

How data is stored & indexed?
- High-performance schema-free aggregated No-SQL databases (e.g. MongoDB)

What operations are performed on data?
- Analytic / Semantic Processing

https://www.whizlabs.com/blog/big-data-tools/

Top 10 Open Source Big Data Tools in 2018

By Ank Varna - March 12, 2018

Today almost every organization extensively uses big data to achieve the competitive edge in the market. With this in mind, open source big data tools for big data processing and analysis are the most useful choice for organizations considering the cost and other benefits. Here are the top open-source projects and the big data benchmarks in the industry. However, it is not the end! There are plenty of other...

https://www.softwareadvice.com/bi/big-data-comparison/
Excel – with

A. The imperative in Big Data is not “How to” compute but “What to compute?”: Please Agree/Disagree; Let’s further add “Why to Analyze?”

B. Your views on Strategic use of Big Data! (Especially in the Context of your Organization)

**Key Points of the Session**

- UNDERSTANDING THE BIG DATA FRAMEWORK FOR AGILE BUSINESS (BDFAB V2.5)
  - As a Holistic Framework for Big Data adoption in a Business
  - As a Risk reduction mechanism

- CONSIDERING THE KEY ELEMENTS OF THE BDFAB:
  - Starting with VALUE
Conclusions & Future Directions

- Focus of Big Data is Business VALUE!
- Business is interested in AGILITY
  - Accuracy and Rapidity of Decision making
- Refine BDFAB to 2.5 and make it tool-enabled

Do you see what we see?