Big Data and Analytics
Use Cases

Dr. Abzetedin Adamov
School of Information Technology and Engineering
ADA University
www.site.ada.edu.az/~aadamov
Objectives

**After completing this lesson, you should be able to:**

- Technological Landscape
- Define Data Science and what a Data Scientist does
- List Big Data and Data Science Use Cases
- Define Standard Parameters for Use Cases
Most Requested Uses of Big Data

• Log Analytics & Storage
• Fusion of multi-INT (multi-formats)
• RFID Tracking & Analytics
• Fraud Detection & Modeling
• Risk Modeling & Management
• 360° View of a Person, Place, or Thing
• Warehouse Extension (case patterns)
• Email / Call Center Transcript Analysis
• Call Detail Record Analysis
• IBM Watson
Do we have Big Data?

DO WE HAVE ANY ACTIONABLE ANALYTICS FROM OUR BIG DATA IN THE CLOUD?

YES, THE DATA SHOWS THAT MY PRODUCTIVITY PLUNGES WHENEVER YOU LEARN NEW JARGON.

MAYBE IN-MEMORY COMPUTING WILL ACCELERATE YOUR APPLICATIONS.

PLUNGE, PLUNGE, PLUNGE.
real-time Big Data use cases

- Ad Technology
- Digital Marketing
- Fraud Detection
- Internet of Things
  - Cyberthreat Security
  - Network Monitoring
  - Personalized Medicine
Big Data Everywhere!

BIG DATA

Data that is TOO LARGE & TOO COMPLEX for conventional data tools to capture, store and analyze.

The 3V’s of Big Data

- VOLUME
- VARIETY
- VELOCITY

90% of the world’s data was generated in the last two years.

Shares traded on US Stock Markets each day:
7 Billion

Data generated in one flight from NY to London:
10 Terabytes

Number of tweets per day on Twitter:
400 Million

Number of ‘Likes’ each day on Facebook:
3 Billion
What is Analytics?

Data on its own is useless unless you can make sense of it!

WHAT IS ANALYTICS?
The scientific process of transforming data into insight for making better decisions, offering new opportunities for a competitive advantage.
The Case for Business Analytics

<table>
<thead>
<tr>
<th>BUSINESS NEED</th>
<th>SOLUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The Business environment today is more complex than ever before.</td>
<td>• Organizations have been turning to the use of analytics.</td>
</tr>
<tr>
<td>• Businesses are expected to be diligently responsive to the increasing demands of customers, various stakeholders and even regulators.</td>
<td>• More than 83% of Global CIOs surveyed by IBM in 2010 singled out Business Intelligence and Analytics as one of their visionary plans for enhancing competitiveness.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GOAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>In most cases the primary objective of an organization that seeks to turn to analytics is:</td>
</tr>
<tr>
<td>• Revenue/Profit growth</td>
</tr>
<tr>
<td>• Optimize expenditure</td>
</tr>
</tbody>
</table>
Types of Analytics

1. Prescriptive Analytics
   - Enabling smart decisions based on data
   - What should we do?

2. Predictive analytics
   - Predicting the future based on historical patterns
   - What could happen?

3. Descriptive analytics
   - Mining data to provide business insights
   - What has happened?
Disrupt or Be Disrupted

41% of executives say digital disruption increases risk of being put out of business

“New technologies combine to create a business innovation platform, not just a technology platform, helping transform every industry on the planet.”

IDC (Dec. 2014)
Challenging Environment – 5Vs

5 Vs that best describe the nature of Big Data Problem

- **VOLUME**: Data at Rest
- **VELOCITY**: Data in Motion
- **VARIETY**: Data in Many Forms
- **VERACITY**: Data in Doubt
- **VALUE**: Data in Limbo
Billy Bean Transformed Baseball Scouting

The use of data analytics has changed the way many major league front offices do business – emphasizing data over traditional player metrics.
Google Transformed Advertising

Google search results for "beginner yoga classes" showing ads related to yoga, including:
- Beginners Yoga Classes - yogaglo.com
- Yoga Online - DailyBurn - Beginner To Advanced Yoga Workouts
- Yoga for Complete Beginners - Yoga Class 20 Minutes - YouTube
- DoYogaWithMe.com: Free Online Yoga Videos - Classes and Poses

Ads include:
- Zumba® Class by Zip Code
- Yoga Classes for Beginners
- Yoga Videos For Beginners
- Beginners Yoga Course
- Yoga Certification NYC
LinkedIn Transformed Network Growth
Netflix Transformed Streaming Video
Amazon Transformed Product Recommendations
Retail Transformed Market Basket Analysis

Question
Which products are very frequently purchased together?

Solution Approach
Determine frequent item sets from transaction logs
Use to design physical store layout accordingly
Use to design special offers and coupon strategy

Example
The urban legend of Diapers & Beer
Financial Use Cases Transformed Analytics

**Customer Profiling**

Financial firms use parameters about customers to determine risk

- Likelihood of customer repaying a loan
  - Collect data (FICO, net worth, etc.)
  - Build a predictive model
  - Use model to price loans

**Fraud Detection**

Credit card companies look at transaction factors to detect fraud

- Likelihood any given transaction is fraudulent
  - Track spending habits
  - Build a spending model
  - Alert when transaction falls outside model
Demand for Data Scientists

Job Trends from Indeed.com

"Data Scientist"
# The Data Science Skillset Continuum

<table>
<thead>
<tr>
<th>Role</th>
<th>Data Engineer</th>
<th>Applied Scientist</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Function</strong></td>
<td>Build production-grade data products</td>
<td>Find signal/meaning in the data; Applies statistical/ML models and tunes the algorithm</td>
</tr>
<tr>
<td><strong>Good at....</strong></td>
<td>Data and Systems architecture</td>
<td>Statistics, Machine learning; Text processing, NLP; R, Python, Matlab, SAS, SQL; Scripting; Visualization / telling the story</td>
</tr>
<tr>
<td></td>
<td>Hadoop, PIG/HIVE, MapReduce, ops/admin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Java, Python, Perl, SQL, C++, NoSQL (HBase, Cassandra, Mongo)</td>
<td></td>
</tr>
</tbody>
</table>
What is Data Science?

“Building software products (aka Data Products) whose core functionality relies on applying statistical or machine learning methods to data.”

What is a data scientist?
A person who does this

Nice historical perspective:
http://whatshebigdata.com/2012/04/26/a-very-short-history-of-data-science/
Data Science Skill Set

- Computer Science
- Machine Learning
- Math & Statistics
- Traditional Software
- Traditional Research
- Subject Matter Expertise

Unicorn
Data Scientist Skill Set

Our perspective:
• It’s a hybrid role: data engineer + applied scientist
• Combines many disciplines:
What is Big Data?
Harness the growing and changing nature of data

Structured
- Calendar
- Database
- Shopping cart
- User profile

Unstructured
- Microphone
- Facebook
- Location
- Bird
- Quote
- Email
- Elephant

Streaming
- Train
- Satellite
- Binary code
- Radio

Challenge is combining transactional data stored in relational databases with less structured data

Big Data = All Data

Get the right information to the right people at the right time in the right format
What is the Internet of Things?

IoT = sensor-acquired data
Modern Architecture

Using a Data Lake

All data sources are considered

Leverages the power of on-prem technologies and the cloud for storage and capture

Native formats, streaming data, big data

Extract and load, no/minimal transform

Storage of data in near-native format

Orchestration becomes possible

Streaming data accommodation becomes possible

Refineries transform data on read

Produce curated data sets to integrate with traditional warehouses

Users discover published data sets/services using familiar tools
A Broad Perspective To Set The Scene

- SaaS
- PaaS
- IaaS

Data Center

Cloud

Edge / IoT

© 2016 Cisco and/or its affiliates. All rights reserved. Cisco Confidential
The Explosion of Unstructured Data

- More than 90% is unstructured data
- Approx. 500 quadrillion files
- Quantity doubles every 2 years
- Most unstructured data is neither stored nor analyzed!

1.8 trillion gigabytes of data was created in 2011...

Source: Cloudera
Application Centric Infrastructure

- Simple
- Scalable
- Secure
- Open
- Manageable
- & Visible
- Cost Effective

Application Centric Infrastructure
Modern Data Warehouse

The Dream
Let’s set ON light bulbs in your head
Data Analytics is needed everywhere
CROSS-INDUSTRIAL USE CASES

Microsoft’s Vision
The Internet of Things – Manufacturing

MANUFACTURING PLANT
- Monitor production flow in near-real-time to eliminate waste and unnecessary work in process inventory.
- Aggregate product data, customer sentiment, and other third-party syndicated data to identify and correct quality issues.
- Implement condition-based maintenance alerts to eliminate machine down-time and increase throughput.

GLOBAL FACILITY INSIGHT
- Manage equipment remotely, using temperature limits and other settings to conserve energy and reduce costs.
- Aggregate product data, customer sentiment, and other third-party syndicated data to identify and correct quality issues.

CUSTOMER SITE
- Transmits operational information to the partner (e.g., OEM) and to field service engineers for remote process automation and optimization.
- Provide cross-channel visibility into inventories to optimize supply and reduce shared costs in the value chain.

GLOBAL OPERATIONS
- I can see my production line status and recommend adjustments to better manage operational cost.
- I gain insight into usage patterns from multiple customers and track equipment deterioration, enabling me to reengineer products for better performance.
- I know when to deploy the right resources for predictive maintenance to minimize equipment failures and reduce service cost.

THIRD-PARTY LOGISTICS
- Provide cross-channel visibility into inventories to optimize supply and reduce shared costs in the value chain.
- I can see my production line status and recommend adjustments to better manage operational cost.
The Internet of Things – Oil & Gas

1. Exploration
   - Find new hydrocarbon reservoirs quicker with seismic data uploaded to the cloud and prepared for analysis
   - Consolidate data from surveys, drill logs, and external sources to generate advanced reservoir models and production forecasts

2. Development
   - Integrate all upstream data onto a unified platform to facilitate analytics, information sharing, and organizational transition
   - Utilize advanced 3D and 4D visualizations based on analytic algorithms to model subsurface geology

3. Drilling
   - Establish near real-time communication and automatically publish events and alarms to the field to guide and protect onsite personnel and assets
   - Combine near real-time drilling and seismic data to optimize drilling trajectories and recovery potential, while minimizing environmental risk

4. Production
   - Maximize recovery by monitoring near real-time production data and generating alerts for conditional maintenance needs
   - Consolidate data from surveys, drill logs, and external sources to generate advanced reservoir models and production forecasts

Operations Control Center

NORTH SHORE PRODUCTION
The Internet of Things – Pharma

- **R&D**
  - Anticipate medical device maintenance needs, and alert patients to schedule a doctor visit for replacement or repair

- **Manufacturing**
  - Manage equipment remotely, using appropriate KPIs
  - Develop better products, faster, informed by a much larger data set based on patient outcomes
  - Reduce machine downtime with condition-based maintenance alerts

- **Distribution**
  - Enable advanced product tracking and authentication to prevent counterfeits

- **Customer Service**
  - Monitor medical device functionality for better customer service, reduced risk, and insight to improve product designs

- **Healthcare Provider**
  - Monitor device data to make more timely health decisions, such as adjusting dosages

- **Patient Home**
  - Aggregate and correlate data from disparate medical devices with medications and health outcomes for advanced insight

- **Aggregate** and correlate data from disparate medical devices with medications and health outcomes for advanced insight
The Internet of Things – Healthcare

**PATIENT HOME**
- Monitor patient conditions with in-home medical devices that alert care team staff when a health event occurs.
- Transform the vehicle into a smart environment that monitors health indicators.

**HOSPITAL**
- Enable an interactive experience between patients and collaborative care teams, and reduce response times by providing remote access to the latest patient data.
- Connect patient data to contextual data, so the latest patient data automatically displays on care provider devices based on their location and role.
- Combine data from various sources to uncover insights that enable an enhanced patient journey, improved operational efficiency, and better risk management.

**OUTPATIENT FACILITY**
- Make authorized patient data accessible from a unified point, enabling a holistic view of the patient’s journey so providers can optimize each care interaction.
- Make patient data visible and actionable in near real-time, enabling improved outcomes through data-driven decision making, better coordination and error reduction.

**HEALTHCARE ECOSYSTEM**
- Integrate data from existing and non-traditional sources to drive Big Data analytics, enabling care process innovation and healthcare transformation.
The Internet of Things – Retail

**IoT DATA FUELS CUSTOMER AND PRODUCT INSIGHTS**

**Marketing**
- Weather Data
- Online Behavior

**Merchandizing**
- Shopping Route
- Purchase History

**RIGHT OFFER, RIGHT TIME, RIGHT PLACE**

**MOBILE EXPERIENCE**
- INSPIRATION, DISCOVERY, PRE-SHOPPING

**IN-STORE SHOPPING**
- Weather Data
- Online Behavior

**STORE PURCHASE HISTORY:**
- Dog food

**REFLECTION**
- We’re ready for the rain! #ShoppingSuccess

**CHECKOUT**
- Retail
  - 200ft

**BEST DEAL**
- Right Offer, Right Time, Right Place

**Have you seen these!**

**#ShoppingSuccess**

**42**
CROSS-INDUSTRIAL USE CASES
based on Standard Parameters
USE-CASE PARAMETERS

1. Company
2. Data Sources
3. Techniques
4. Business Value
EXAMPLE: DATA SOURCE

• Server Telemetry
• Monitoring Logs
• Memory Consumption
• Processing (CPU) Consumption
• Network Flow
EXAMPLE: TECHNIQUES

• Pattern Recognition
• Proactive Monitoring
• Early Alert Delivery
EXAMPLE: BUSINESS VALUE

\[ \text{PROFIT} = \text{REVENUE} - \text{COST} \]

To increase this...  ... increase this...  ... or decrease this...
EXAMPLE: TELECOMMUNICATIONS

Problem: ETL Offload

Data Sources
- Customer Records
- Contract Data
- Purchase Orders
- Call Center

Techniques
- ETL
- Analytics

Business Value

$$\text{PROFIT} = \text{REVENUE} - \text{COST}$$

To increase this... ... increase this... ... or decrease this...

ETL - Hadoop + Data Analytics
EXAMPLE: CARD ISSUER

Data Sources
- Customer Purchase History
- Merchant Designations
- Merchant Special Offers

Techniques
- ETL
- Machine Learning

Business Value

Problem: Low Profit

\[ \text{PROFIT} = \text{REVENUE} - \text{COST} \]

To increase this...
... increase this...
... or decrease this...
EXAMPLE: WASTE & RECYCLING

Data Sources
- Truck Geolocation Data
  - 20,000 trucks
  - 5 sec interval
- Landfill Geo-boundaries

Techniques
- Realtime Streaming
- Batch Computation
- Graph Algorithms

Business Value

Problem: Idle Alerts

PROFIT = REVENUE - COST

To increase this... ... increase this... ... or decrease this...

- Immediate Alerts
- Tax Reduction
- Route Optimization
EXAMPLE: BANKING

Data Sources
- Anti-Money Laundering
- Consumer Transactions

Techniques
- Bayesian Learning
- Peer Group Analysis

Business Value

\[ \text{PROFIT} = \text{REVENUE} - \text{COST} \]

Problem: Fraud Detection

To increase this...  ... increase this...  ... or decrease this...

Suspicious Events

+ hadoop + =
EXAMPLE: DNA ANALYSIS

Data Sources

• Birth, Death, Census,
• Military, Immigration Records
• Search Behavior Activity
• DNA SNP

Techniques

• Record Linking
• Search Relevance
• Clickstream
• Security Forensics
• DNA Matching

Business Value

\[
\text{PROFIT} = \text{REVENUE} - \text{COST}
\]

To increase this...

... increase this...

...or decrease this...

PROFIT = REVENUE - COST

Problem: Search Relevance, DNA Matching
Q & A

Abzetdin Adamov, Assoc Prof.
Email me at: aadamov@ada.edu.az
Follow me at: @
Link to me at: www.linkedin.com/in/adamov
Visit my blog at: aadamov.wordpress.com