

Big Data and Analytics Use Cases

Dr. Abzетdin 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?



real-time Big Data use cases

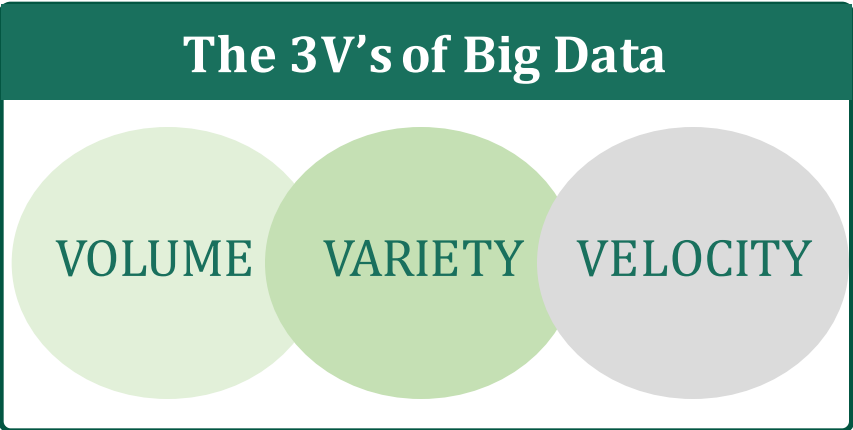
- Ad Technology
- Digital Marketing
- Fraud Detection
- Internet of Things
 - Cyberthreat Security
 - Network Monitoring
 - Personalized Medicine



Big Data Everywhere!

BIG DATA

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



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

90% OF THE WORLD'S DATA WAS GENERATED IN THE LAST TWO YEARS

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

BUSINESS NEED

- The Business environment today is more complex than ever before.
- Businesses are expected to be diligently responsive to the increasing demands of customers, various stakeholders and even regulators.

GOAL

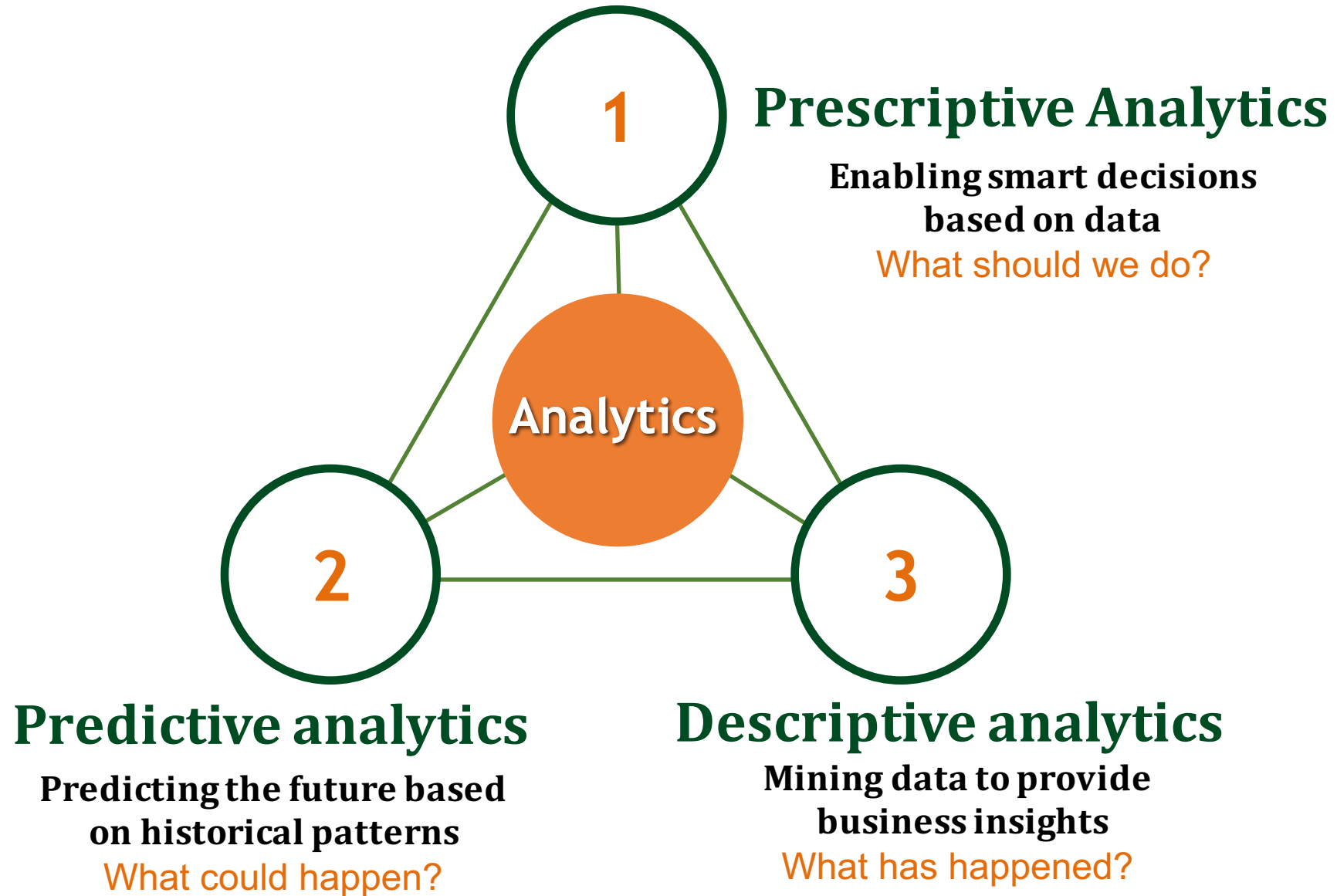
In most cases the primary objective of an organization that seeks to turn to analytics is:

- Revenue/Profit growth
- Optimize expenditure

SOLUTION

- Organizations have been turning to the use of analytics.
- 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.

Types of Analytics





Disrupt or Be Disrupted

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

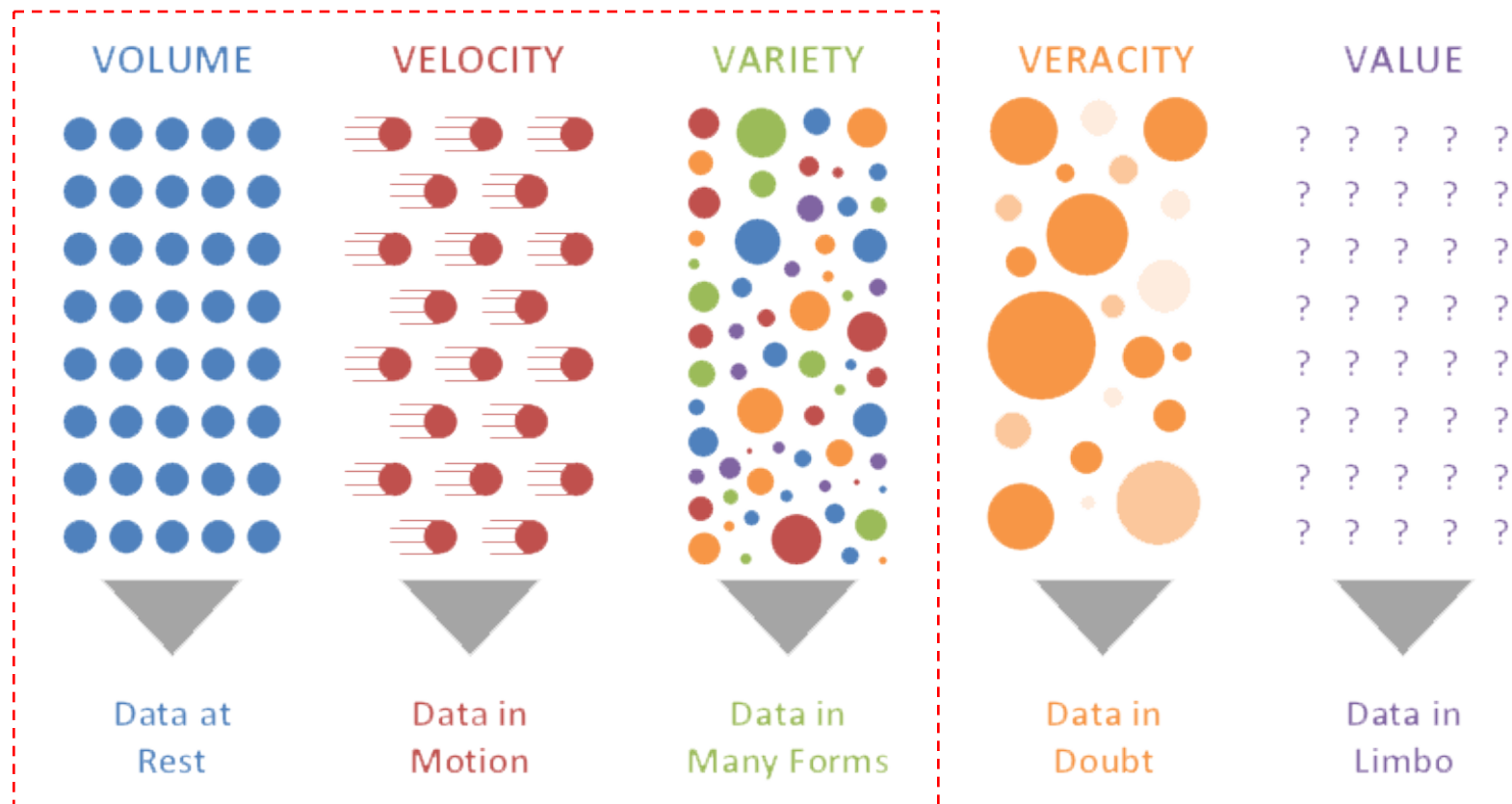
Digital Vortex 2015

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

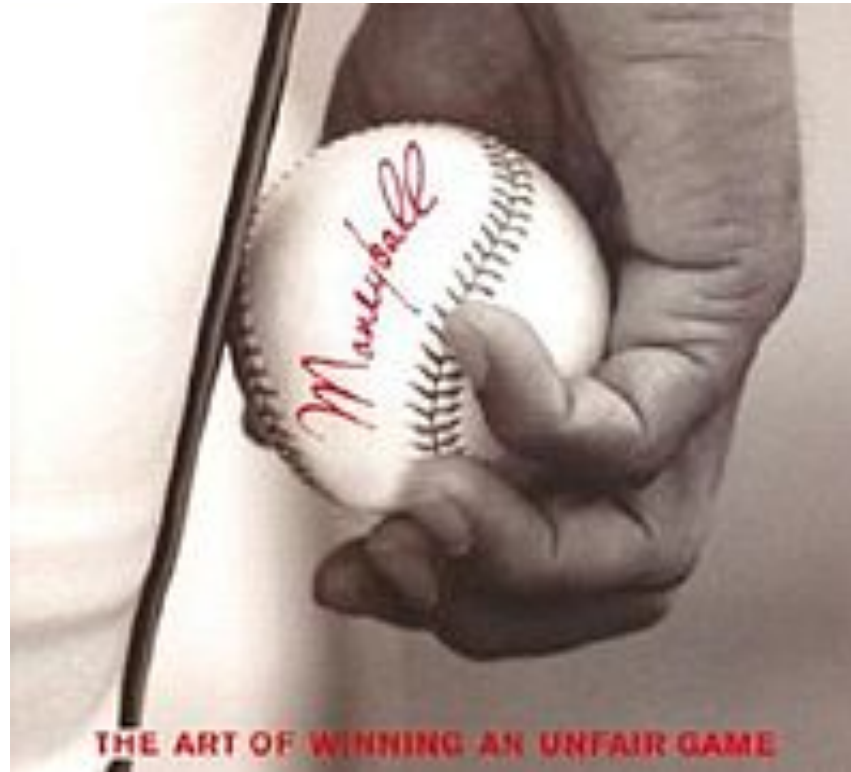
IDC (Dec. 2014)

Challenging Environment – 5Vs

5 Vs that best describe the nature of Big Data Problem





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  

[Web](#) [Maps](#) [Videos](#) [Images](#) [Shopping](#) [More ▾](#) [Search tools](#)

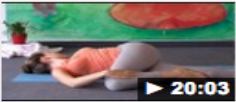
About 2,310,000 results (0.28 seconds)

Ads related to **beginner yoga classes** ⓘ

[Beginners Yoga Classes - yogaglo.com](#)
[www.yogaglo.com/BeginnerYoga](#) ▾
Find the Best **Beginner Yoga Classes** at YogaGlo® Today! Free Trial Offer
YogaGlo has 1,263 followers on Google+
[How It Works](#) [Start Your Free Trial](#)
[Beginners Center](#)

[Online Yoga Videos - Try Online Yoga For Beginners](#)
[www.myyogaworks.com/](#) ▾
Free 14 Day Trial - Start Today!

[Yoga Online - DailyBurn - Beginner To Advanced Yoga Workouts](#)
[www.dailyburn.com/Yoga](#) ▾
Start Your Free 30-Day Trial Today.
Be At Peace with Pricing - Namaste from DailyBurn - Discover the Yogi In You

[Yoga for Complete Beginners - Yoga Class 20 Minutes - YouTube](#)
 [www.youtube.com/watch?v...](#) ▾ YouTube ▾
Dec 6, 2010 - Uploaded by YogaVidyaEnglish
Yoga for complete **beginners**. 20 minute gentle **yoga class** to give you greater relaxation, more energy and ...

[DoYogaWithMe.com: Free Online Yoga Videos - Classes and Poses](#)
[www.doyogawithme.com/](#) ▾
Online yoga videos from DoYogaWithMe.com. ... Browse **Yoga Class** Videos By: ... This is a great transition class from **beginner** ashtanga to the full primary ...

Ads ⓘ

[Zumba® Class by Zip Code](#)
[www.zumba.com/FindAClass](#) ▾
A Fun, Fast & Effective Workout.
Find a Zumba® **Class** Near You Today!

[Learn Yoga for Free Today](#)
[www.alison.com/Free-Yoga](#) ▾
Improve Core Strength & Flexibility
Free, Online, Self-Paced Course

[Yoga Classes for Beginners](#)
[www.hulu.com/plus](#) ▾
Try Hulu Plus! More TV Shows & Movies. Get 1 Week Free Now.

[Yoga Videos For Beginners](#)
[www.gaiamtv.com/FreeTrial](#) ▾
Practice **Yoga** On Your Own Time.
Sign Up for Your Free Trial Today!

[Beginners Yoga Course](#)
[webcrawler.com/beginners+yoga+course](#) ▾
Find more facts, sites and tools.
Get more **beginners yoga** course







[Yoga Certification NYC](#)
[www.atmananda.com/](#) ▾
200 hr **Yoga Alliance**, all Summer
Inc Free **Yoga Classes**

LinkedIn Transformed Network Growth

People You May Know ^{beta} See people from different parts of your professional life

LinkedIn Google eBay amazon OMGPOP S Stanford University San Jose State University


All Suggestions / LinkedIn (12) Connect All


 <p>Brad Mauney ^{2nd} Senior Product Manager, Search & Social Graph at LinkedIn Mountain View, California</p> <p>Connect 5 shared connections</p>	 <p>Albert Wang ^{2nd} Senior User Experience Designer at LinkedIn Mountain View, California</p> <p>Connect 127 shared connections</p>
 <p>Sam Shah ^{2nd} Principal Engineer at LinkedIn Mountain View, California</p> <p>Connect 22 shared connections</p>	 <p>Tan Nhu ^{2nd} Senior Web Developer at LinkedIn Mountain View, California</p> <p>Connect 16 shared connections</p>
 <p>Vinodh Jayaram ^{2nd} Software Engineering Manager at LinkedIn Mountain View, California</p> <p>Connect 10 shared connections</p>	 <p>Andy Chen ^{2nd} Software Engineer at LinkedIn Mountain View, California</p> <p>Connect 78 shared connections</p>


Netflix Transformed Streaming Video


Because you added **The Hunger Games** Close X


The Hunger Games
has been added to My List


The Forger

[Play](#)
★★★★☆
Not Interested


Red Dawn

[Play](#)
★★★☆☆
Not Interested


Pirates of the Caribbean: Black Pearl

[Play](#)
★★★★☆
Not Interested


The Dictator

[Play](#)
★★★☆☆
Not Interested


Hansel & Gretel: Witch Hunters

[Play](#)
★★★☆☆
Not Interested

Olympus Has Fallen

[Play](#)
★★★★☆
Not Interested

The Grey

[Play](#)
★★★★☆
Not Interested

Mission: Impossible - Ghost Protocol

[Play](#)
★★★★☆
Not Interested

Skyfall

[Play](#)
★★★★★
Not Interested

Jack Reacher

[Play](#)
★★★★☆
Not Interested

Amazon Transformed Product Recommendations

Today's Recommendations For You

Here's a daily sample of items recommended for you. Click here to [see all recommendations](#)

Even Faster Web Sites: Performance... (Paperback) by Steve Souders
★★★★★ (7) \$23.10
[Fix this recommendation](#)

Simply JavaScript (Paperback) by Kevin Yank
★★★★☆ (19) \$26.37
[Fix this recommendation](#)

The Art & Science of JavaScript (Paperback) by Eric E. Sorenson
★★★★★ (5)
[Fix this recommendation](#)

Any Category Algorithms Boxed Sets Business & Culture Java
Graphic Design Microsoft Networking Networks, Protocols & APIs New SQL

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



The Data Science Skillset Continuum

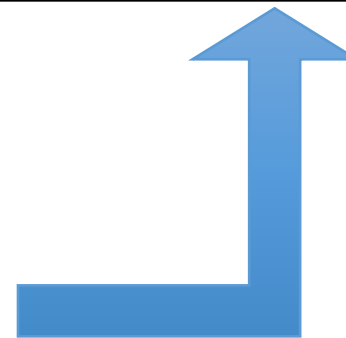


Role	Data Engineer	Applied Scientist
Function	Build production-grade data products	Find signal/meaning in the data Applies statistical/ML models and tunes the algorithm
Good at....	Data and Systems architecture Hadoop, PIG/HIVE, MapReduce, ops/admin Java, Python, Perl, SQL, C++, NoSQL (HBase, Cassandra,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

What is Data Science?

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

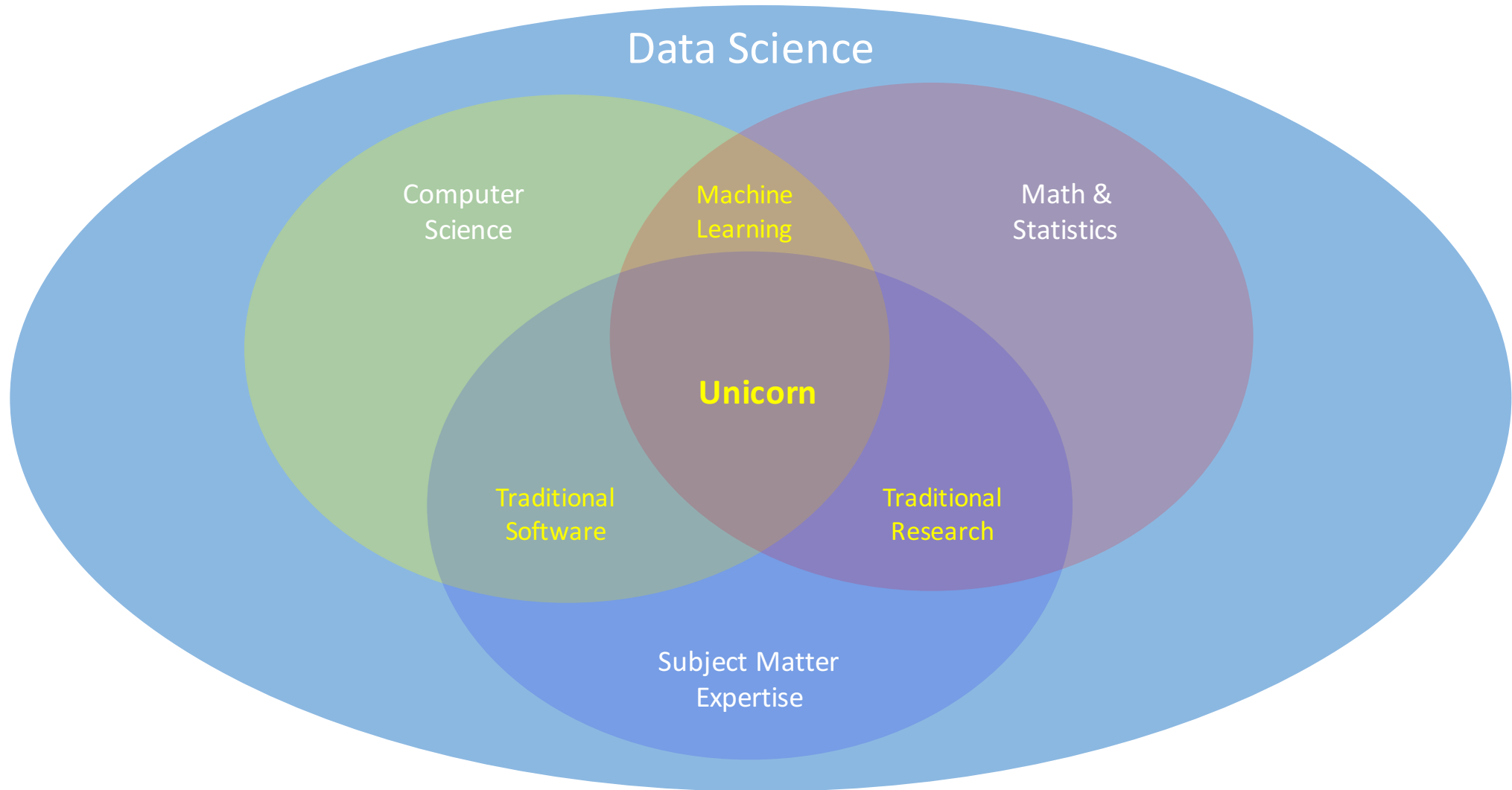
What is a data scientist?
A person who does **this**



Nice historical perspective:

<http://whatsthebigdata.com/2012/04/26/a-very-short-history-of-data-science/>

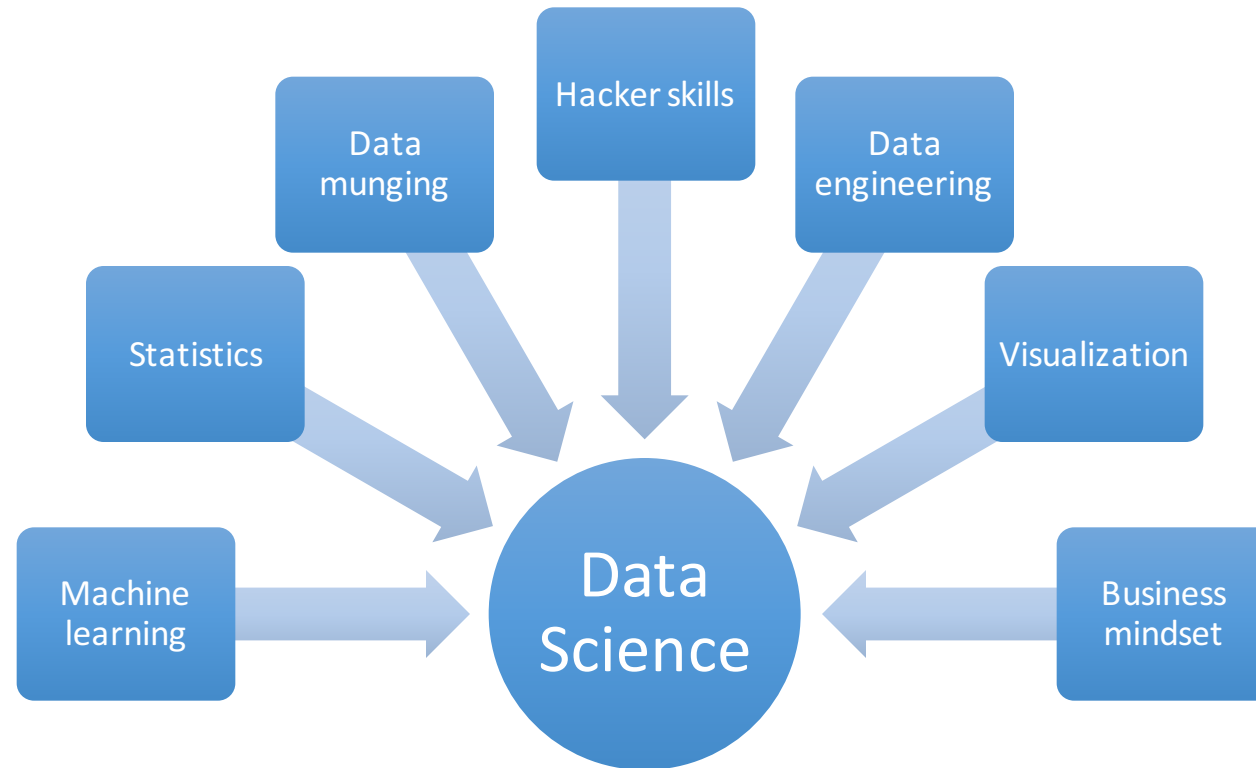
Data Science Skill Set



Data Scientist Skill Set

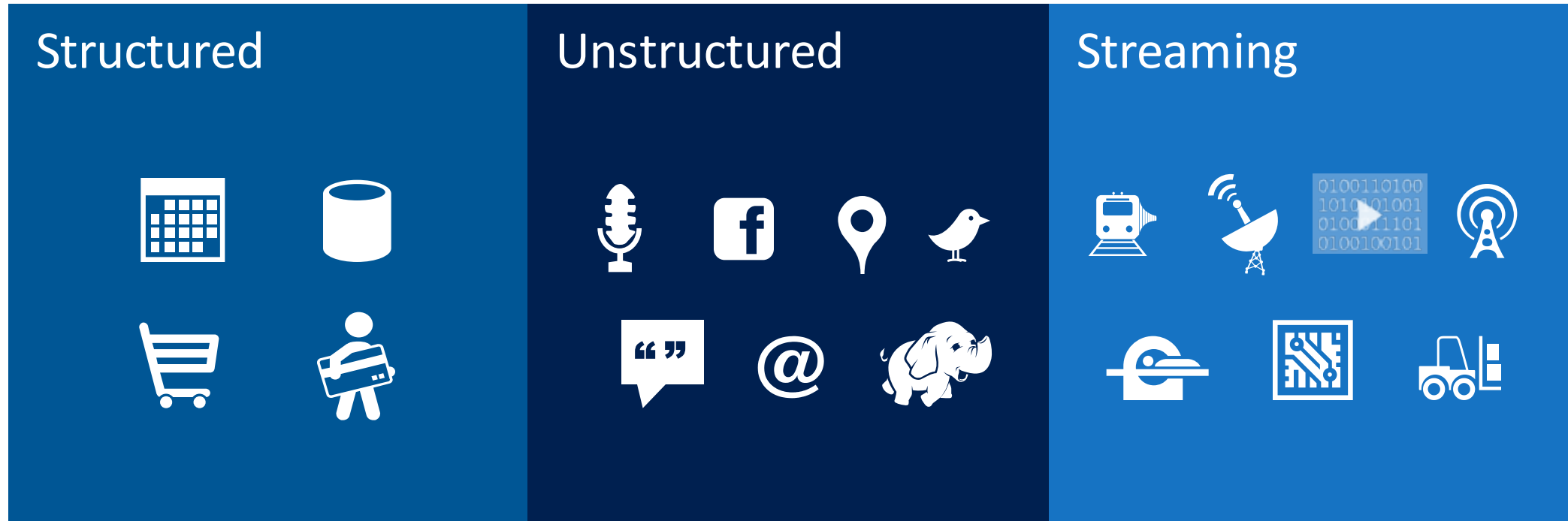
Our perspective:

- It's a hybrid role: data engineer + applied scientist
- Combines many disciplines:



What is Big Data?

Harness the growing and changing nature of data



- ▶ 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

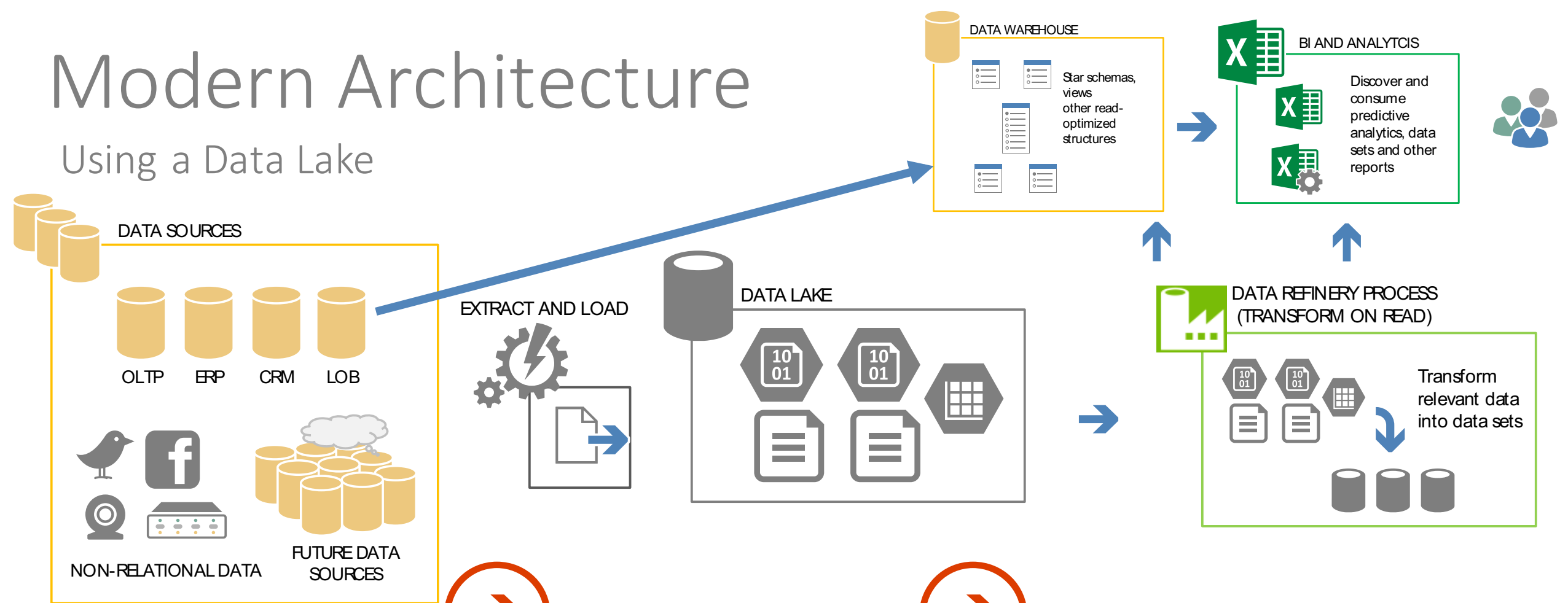
What is the Internet of Things?



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

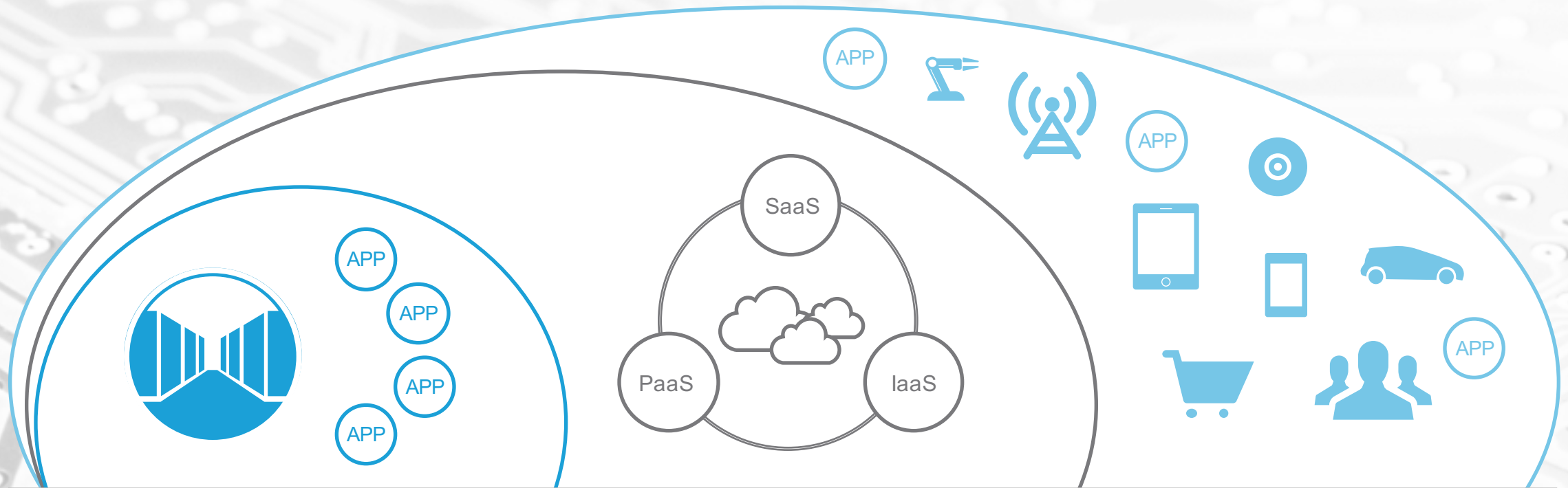


Refineries transform data on read

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



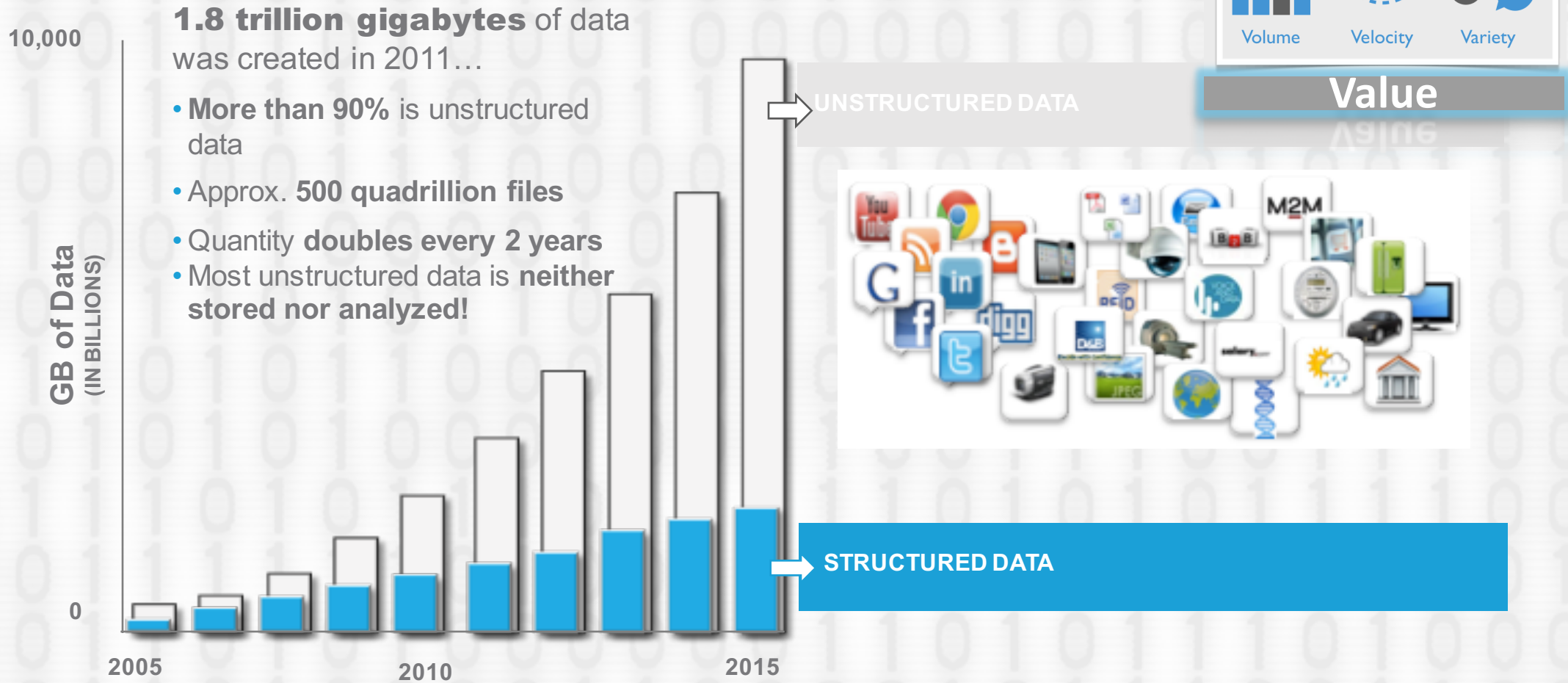
Data Center

Cloud

Edge / IoT



The Explosion of Unstructured Data



Source: Cloudera

Application Centric Infrastructure



Simple



Scalable



Secure

Open



Manageable
& Visible

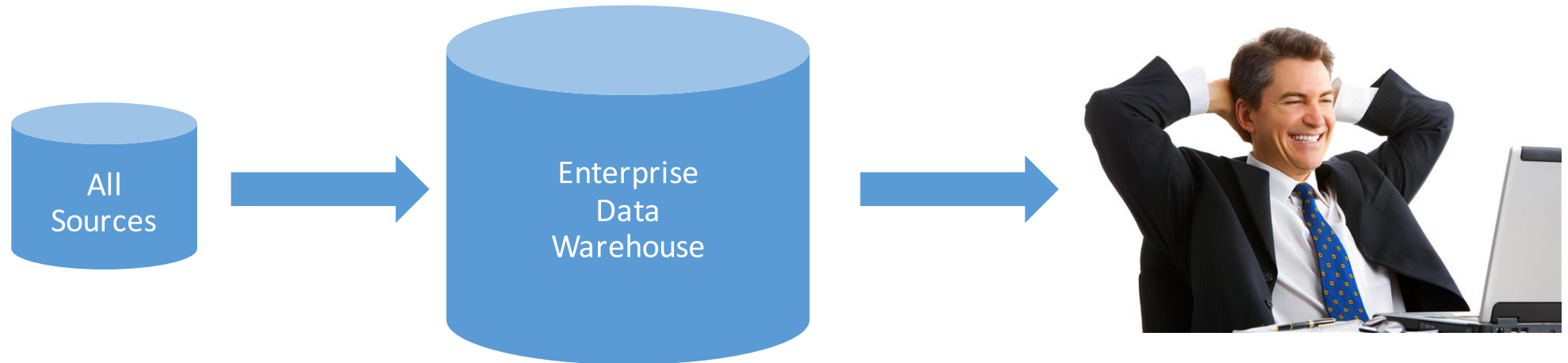


Cost Effective



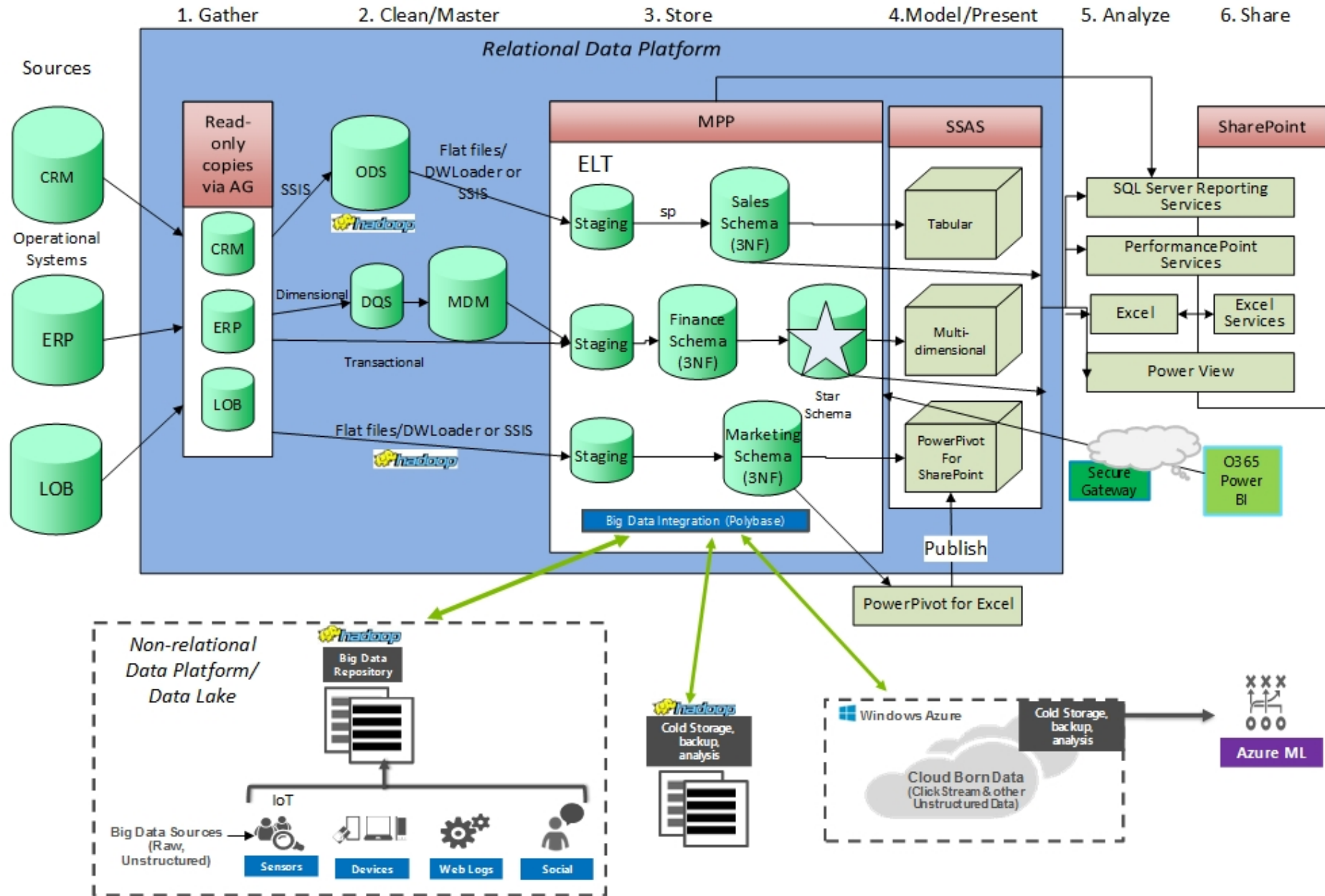
The
Dream

Modern Data Warehouse



Modern Data Warehouse

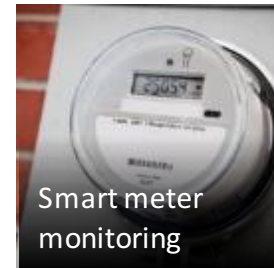
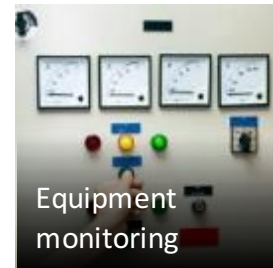
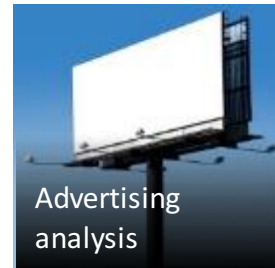
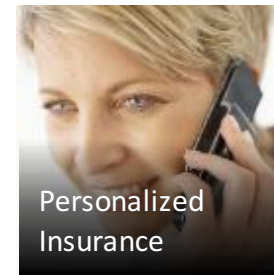
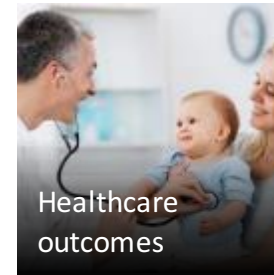
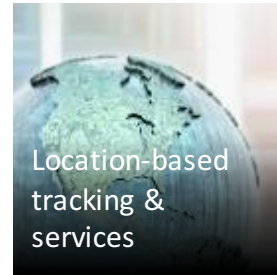
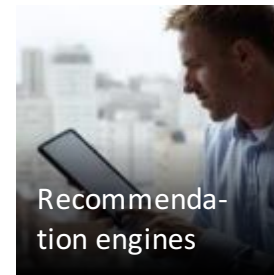
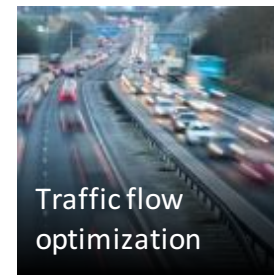
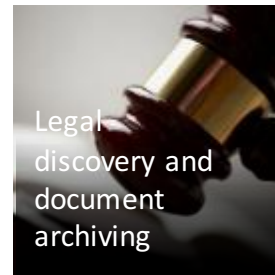
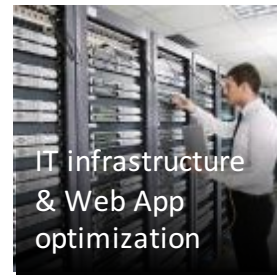
The Reality



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.

Manage equipment remotely, using temperature limits and other settings to conserve energy and reduce costs.

Implement condition-based maintenance alerts to eliminate machine down-time and increase throughput.

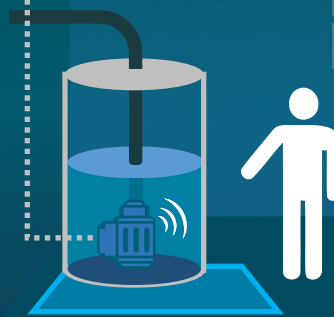
Aggregate product data, customer sentiment, and other third-party syndicated data to identify and correct quality issues.

GLOBAL FACILITY INSIGHT



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

Management



I can see my production line status and recommend adjustments to better manage operational cost.

R&D



I gain insight into usage patterns from multiple customers and track equipment deterioration, enabling me to reengineer products for better performance.

Field Service



I know when to deploy the right resources for predictive maintenance to minimize equipment failures and reduce service cost.



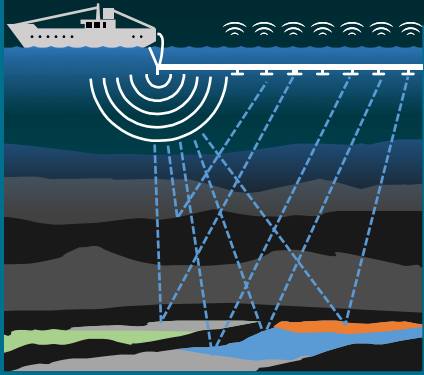
THIRD-PARTY LOGISTICS



The Internet of Things – Oil & Gas

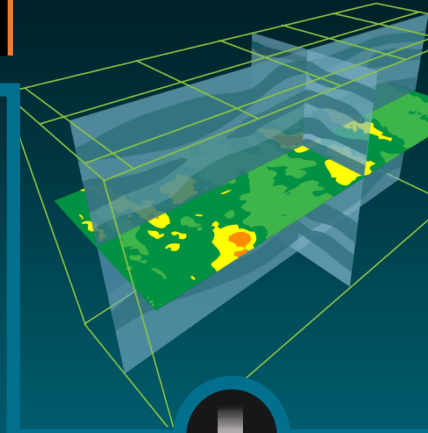
1. Exploration

Find new hydrocarbon reservoirs quicker with seismic data uploaded to the cloud and prepared for analysis

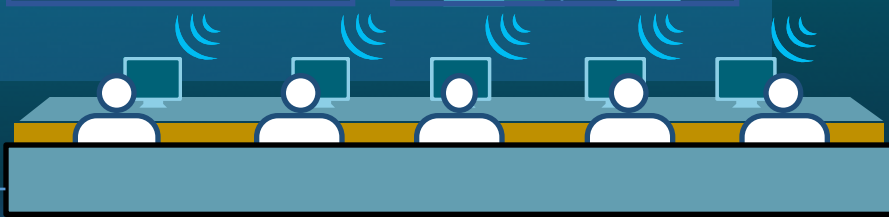
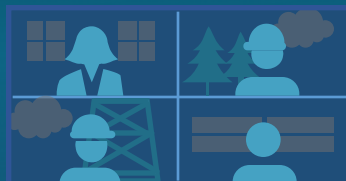
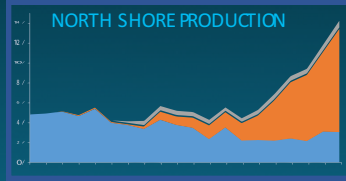


Consolidate data from surveys, drill logs, and external sources to generate advanced reservoir models and production forecasts

2. Development

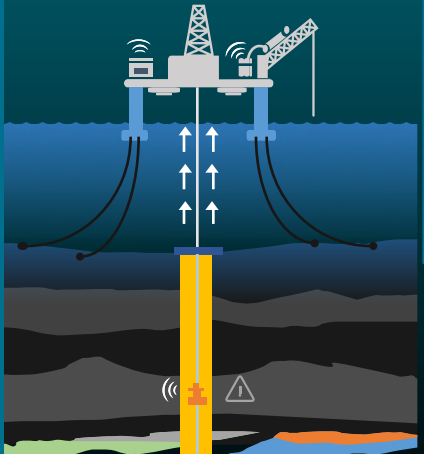


Operations Control Center



4. Production

Maximize recovery by monitoring near real-time production data and generating alerts for conditional maintenance needs



Combine near real-time drilling and seismic data to optimize drilling trajectories and recovery potential, while minimizing environmental risk

3. Drilling



Production Manager



Integrate all upstream data onto a unified platform to facilitate analytics, information sharing, and organizational transition

Geologist



Utilize advanced 3D and 4D visualizations based on analytic algorithms to model subsurface geology

Onsite personnel

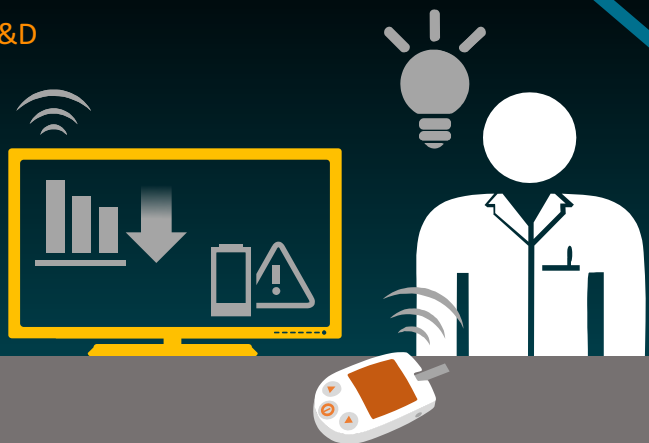


Establish near real-time communication and automatically publish events and alarms to the field to guide and protect onsite personnel and assets



The Internet of Things – Pharma

R&D



Manufacturing

Manage equipment remotely, using appropriate KPIs

Reduce machine downtime with condition-based maintenance alerts

Develop better products, faster, informed by a much larger data set based on patient outcomes

Distribution



Enable advanced product tracking and authentication to prevent counterfeits

Customer Service



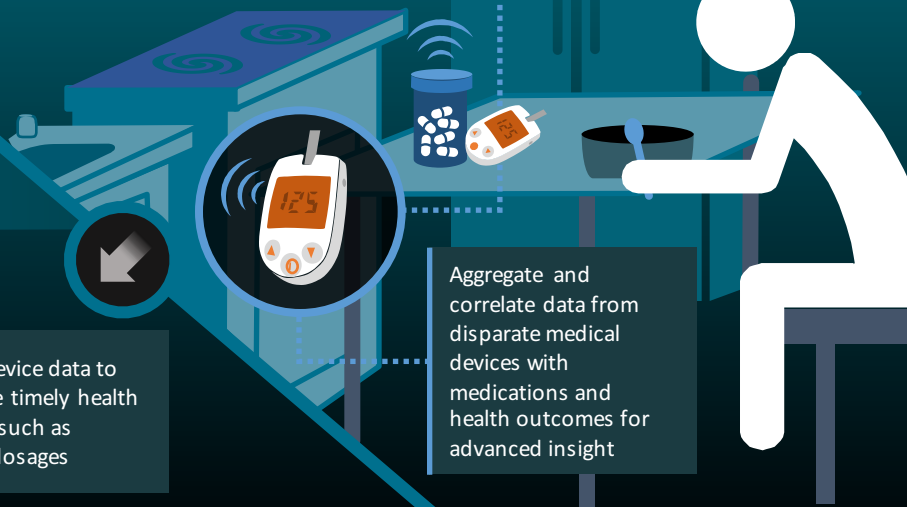
Anticipate medical device maintenance needs, and alert patients to schedule a doctor visit for replacement or repair

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



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



The Internet of Things – Healthcare



PATIENT HOME



HOSPITAL



OUTPATIENT FACILITY

Monitor patient conditions with in-home medical devices that alert care team staff when a health event occurs.



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.

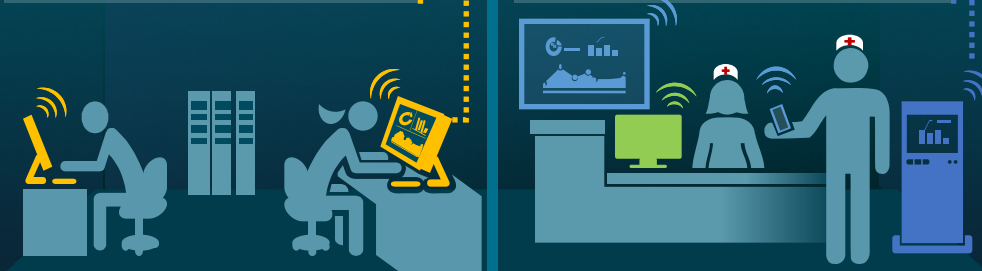


Make authorized patient data accessible from a unified point, enabling a holistic view of the patient's journey so providers can optimize each care interaction.



Combine data from various sources to uncover insights that enable an enhanced patient journey, improved operational efficiency, and better risk management.

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.

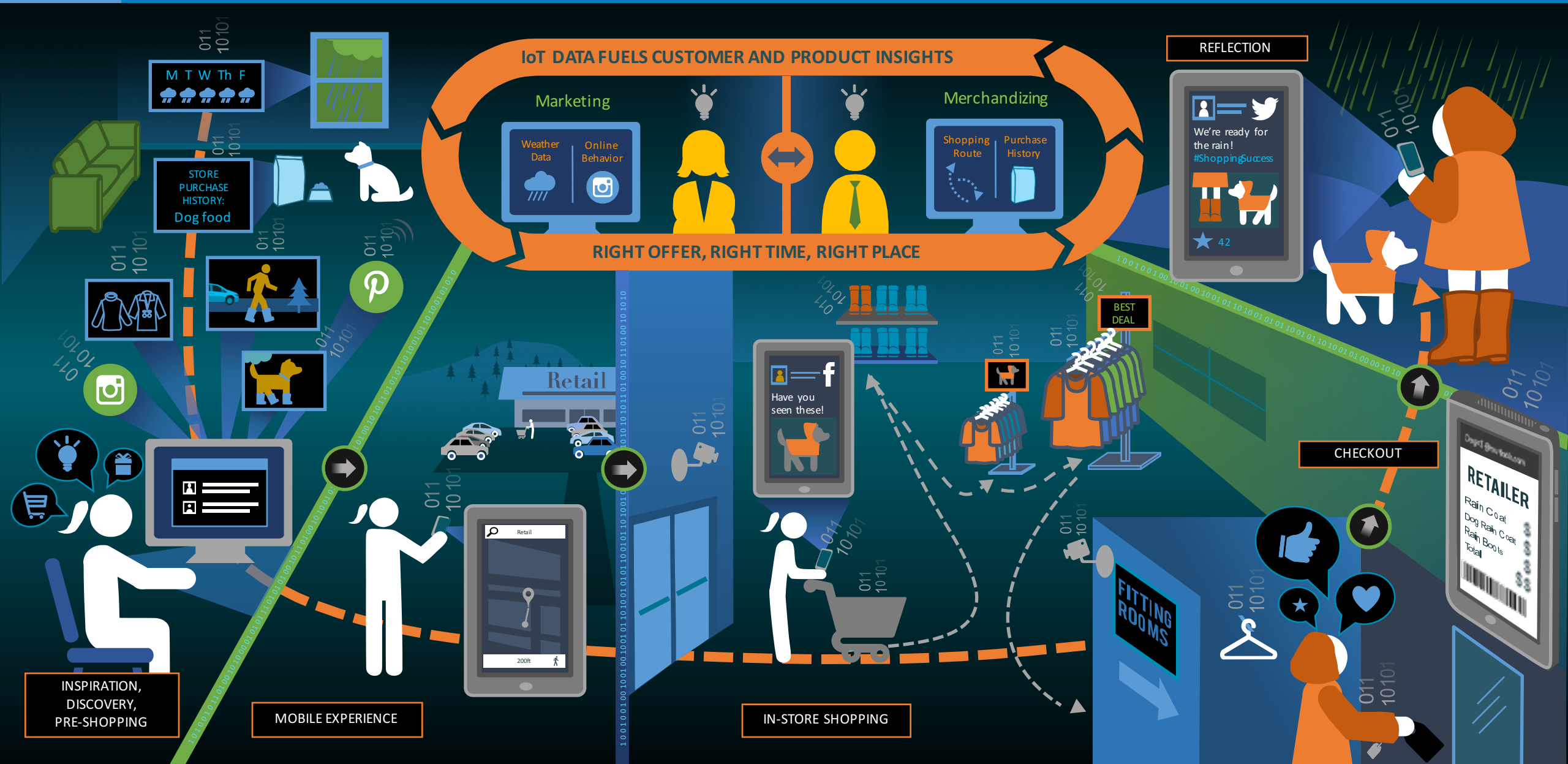


Transform the vehicle into a smart environment that monitors health indicators.





The Internet of Things – Retail



CROSS-INDUSTRIAL USE CASES

based on Standard Parameters

USE-CASE PARAMETERS

1. Company



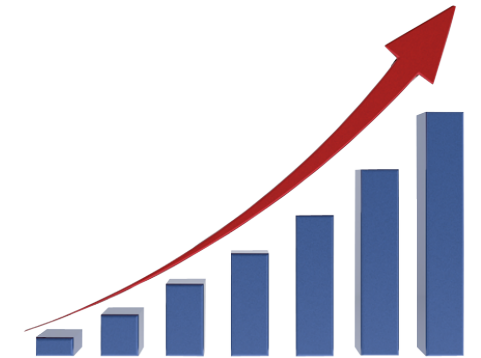
2. Data Sources



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



Problem: Low Profit

Data Sources

- Customer Purchase History
- Merchant Designations
- Merchant Special Offers

Techniques

- ETL
- Machine Learning

Business Value

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

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



EXAMPLE: WASTE & RECYCLING



Problem: Idle Alerts

Data Sources

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

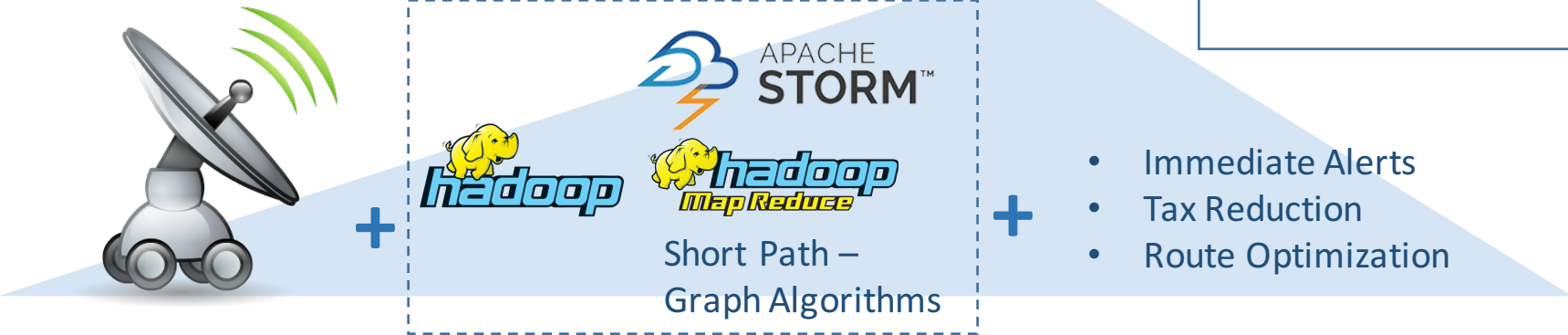
Techniques

- Realtime Streaming
- Batch Computation
- Graph Algorithms

Business Value

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

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



EXAMPLE: BANKING



Problem: Fraud Detection

Data Sources

- Anti-Money Laundering
- Consumer Transactions

Techniques

- Bayesian Learning
- Peer Group Analysis

Business Value

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

To increase this...

... increase this...

... or decrease this...



Suspicious Events

EXAMPLE: DNA ANALYSIS



Problem: Search Relevance, DNA Matching

Data Sources

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

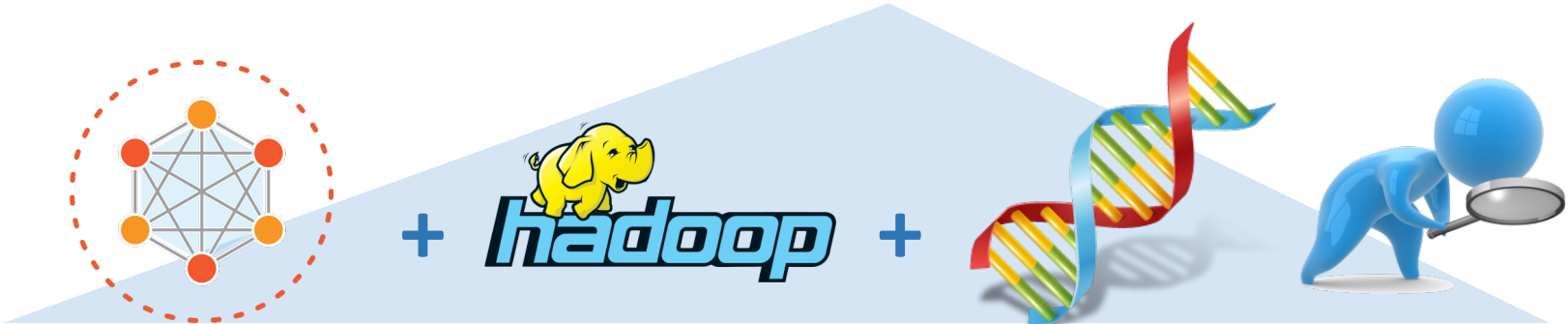
Techniques

- 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...



Q & A



Abzeturdin 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