latentview

Actionable Insights • Accurate Decisions

Analytics – Why, What, How & Beyond

May 2018

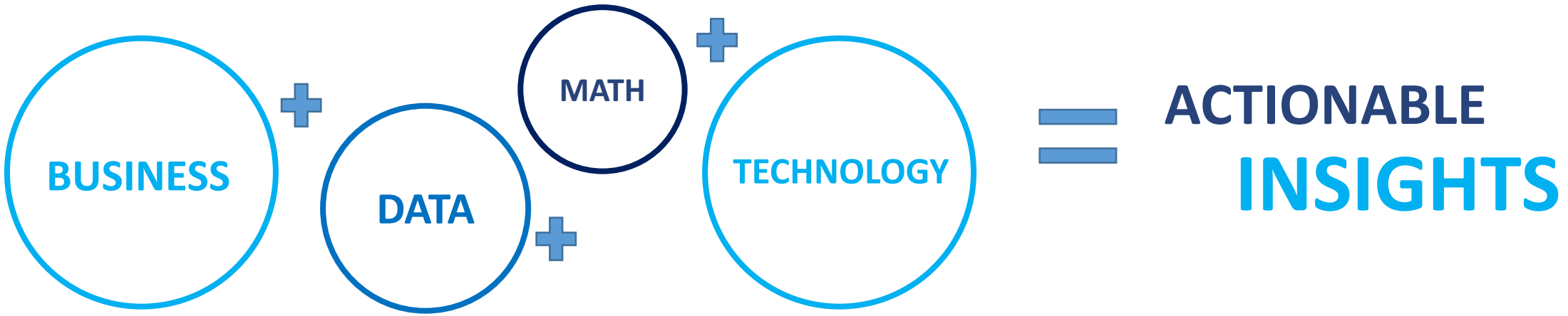
GREAT LAKES

INSTITUTE OF MANAGEMENT, CHENNAI

Global Mindset - Indian Roots

Q1: Analytics seems to be an abstract concept. What is the easiest way to think about it at a high-level?

Analytics – Highest Level View



Q2: Am interested to know a lot about Analytics but first can we start with some real-world examples?

Case Study 1:
Famous Automobile Manufacturer

<u>Starting Point:</u> 3 Years of vehicle sensor data collected across 108 countries along with data on warranty claims

Case Study 1: Customer Behavior Modeling

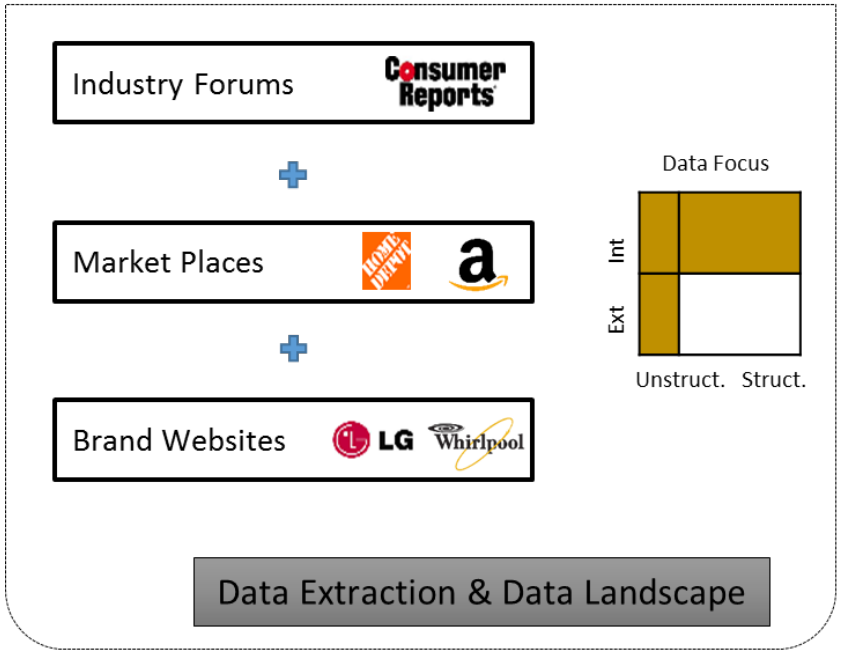
Salient Points from Analytics perspective:

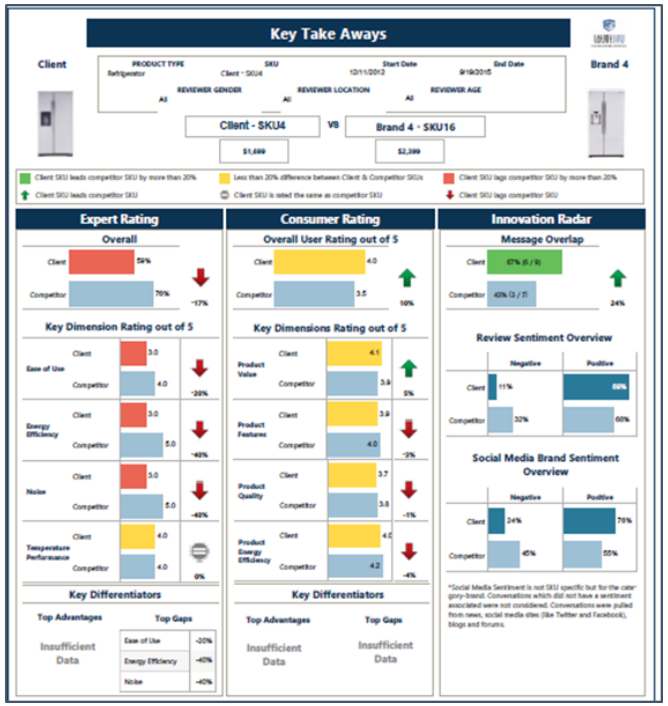
- **Business:** Warranty costs were high & rising. Urgent need to control costs & increase customer satisfaction
- **Data:** Sensor (semi-structured) data collected from cars running in 100+ countries
- **Math:** Clustering done on data to identify driving styles which is then correlated with warranty claims to predict defects
- **Technology:** Spark on the Cloud platform called Databricks, User Interface for self-service

Case 2:
Large Consumer Durables Company

<u>Starting Point:</u> Purchase drivers were determined by post-facto analysis of POS data at stores and survey data resulting in delays of up to eight months to get consumer feedback on product features.

Case Study 2: Social Data to Drive Innovation





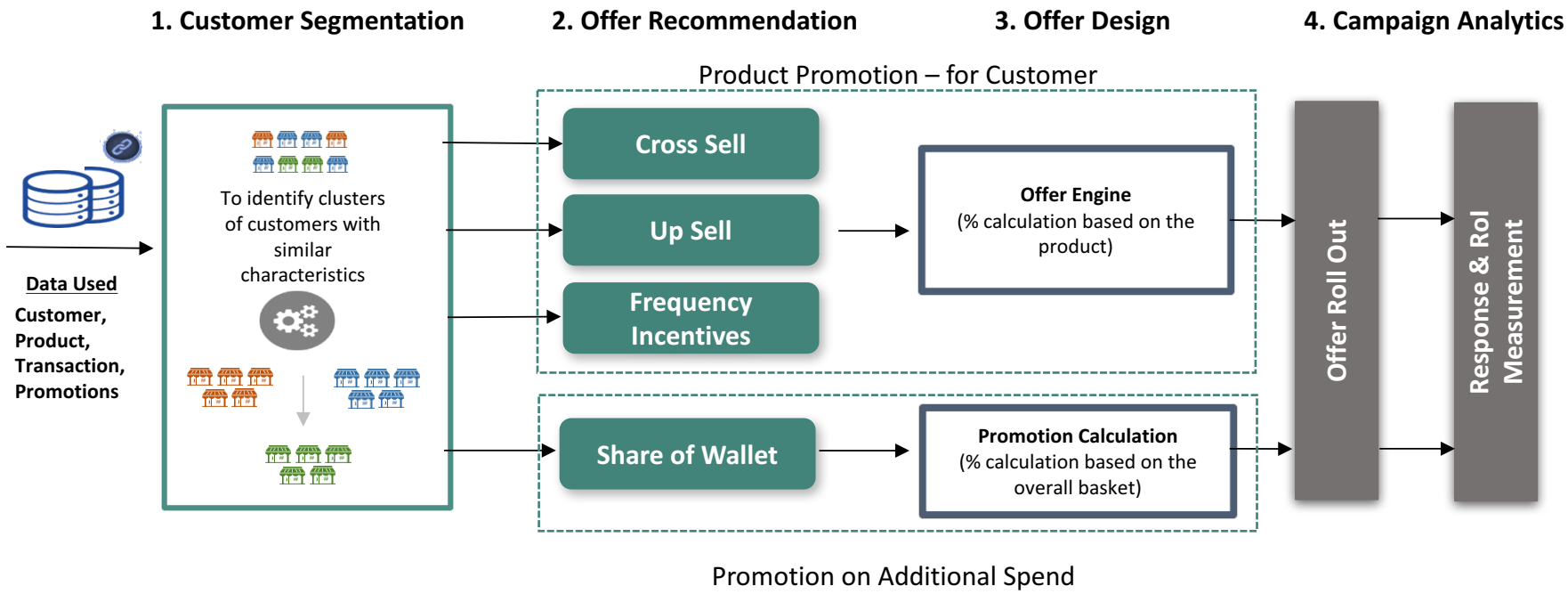
Salient Points from Analytics perspective:

- **Business:** Can we identify opportunities for Innovation using external data?
- **Data:** Reviews & Social Interactions captured across the globe. Unstructured data in the form of text
- **Math:** Sophisticated Natural Language Processing Techniques to extract insights from unstructured data
- **Technology:** Automated data pipeline to ingest & analyze data. Visualization using Tableau

Case 3:
World's Largest Food Distribution Company

<u>Starting Point:</u> Customer & transactional data related to sales of food products to over 500,000 customer locations including restaurants, healthcare & educational facilities and other food service customers. They wanted to drive high margin product sales through effective cross-sell & up-sell

Case Study 3: Recommendation Engine to Increase Sales



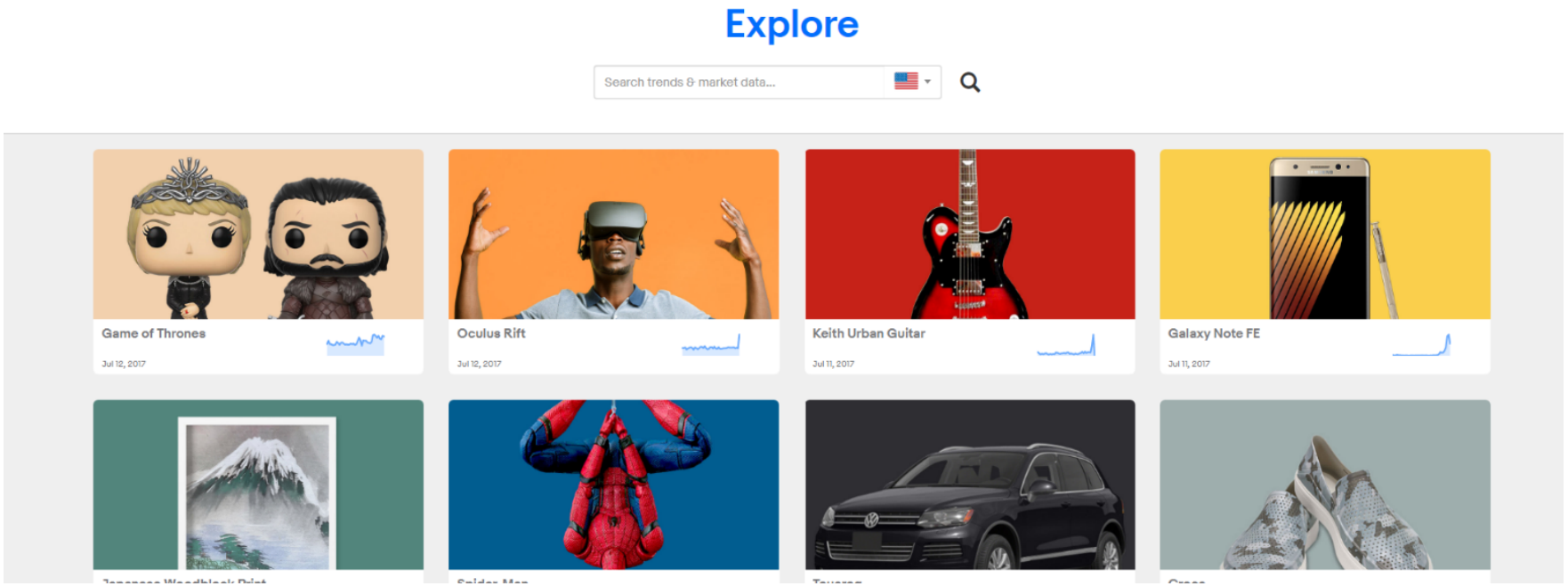
Salient Points from Analytics perspective:

- **Business:** Can we identify opportunities for cross-sell / up-sell to sell more of high margin products?
- **Data:** Customer, Product, Transactions and Promotions
- **Math:** Clustering followed by Collaborative Filtering (Recommendation Engine)
- **Technology:** Automated pipeline that generates recommendations for every sales person

Case 4:
One of the world's largest E-commerce marketplace

<u>Starting Point:</u> Millions of buy & sell transactions happen every minute in the marketplace. Can we utilize the data to inspire and enhance customer experience?

Case Study 4: Real-time Trends in E-Commerce Marketplace

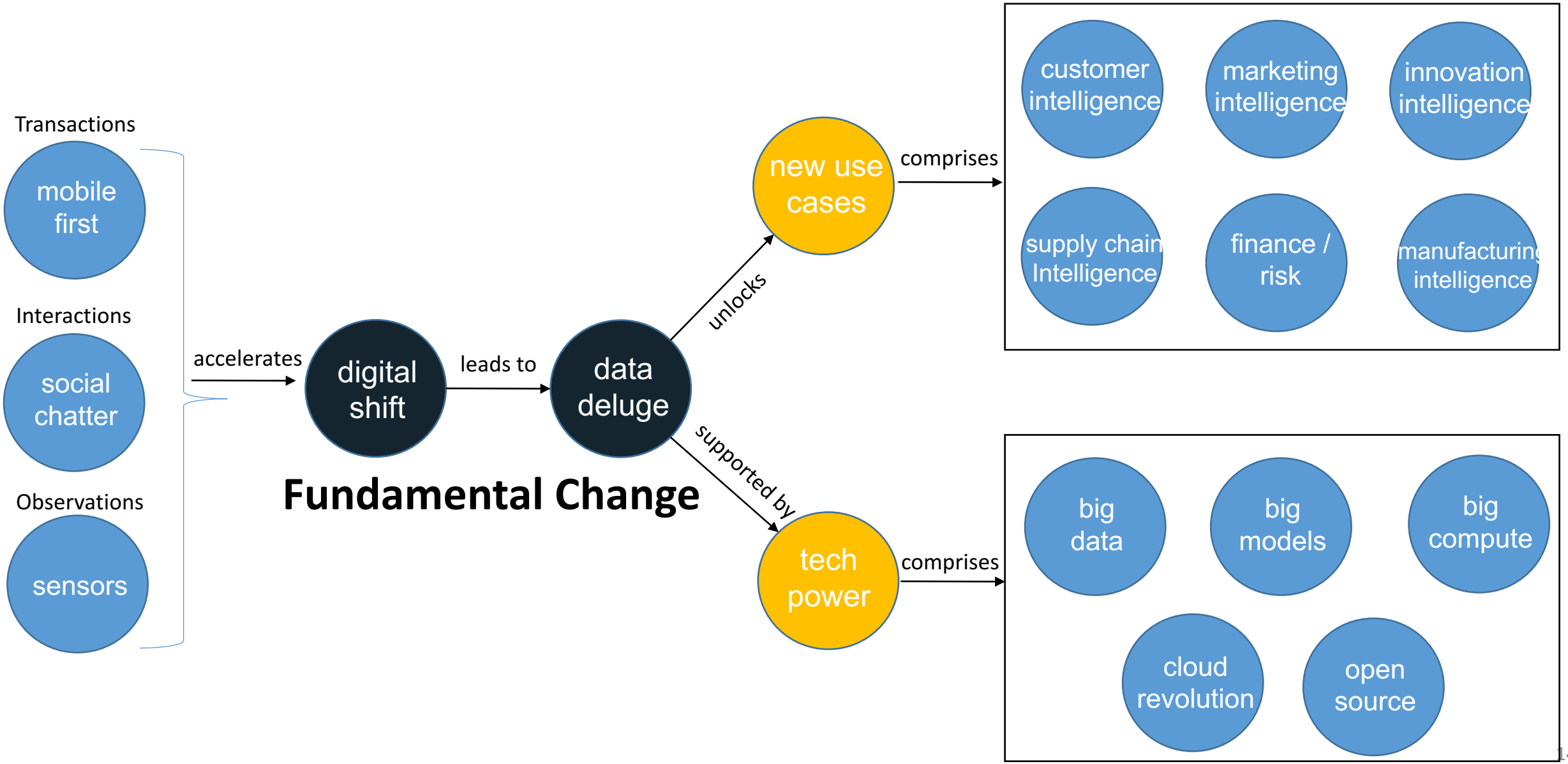


Salient Points from Analytics perspective:

- **Volume:** Hundreds of Terabytes of data needs to be stored & indexed for search
- **Variety:** Structured & Unstructured data processing
- **Velocity:** Real-time ingestion of transactions and millisecond response to search queries
- **Analytics:** Recommendations served in real-time

Q3: How are you sure that the application of analytics is for long-term and is not just a fad?

Digital Shift – Fundamental, Irreversible Change



Data Science & ML can have great impact on industries

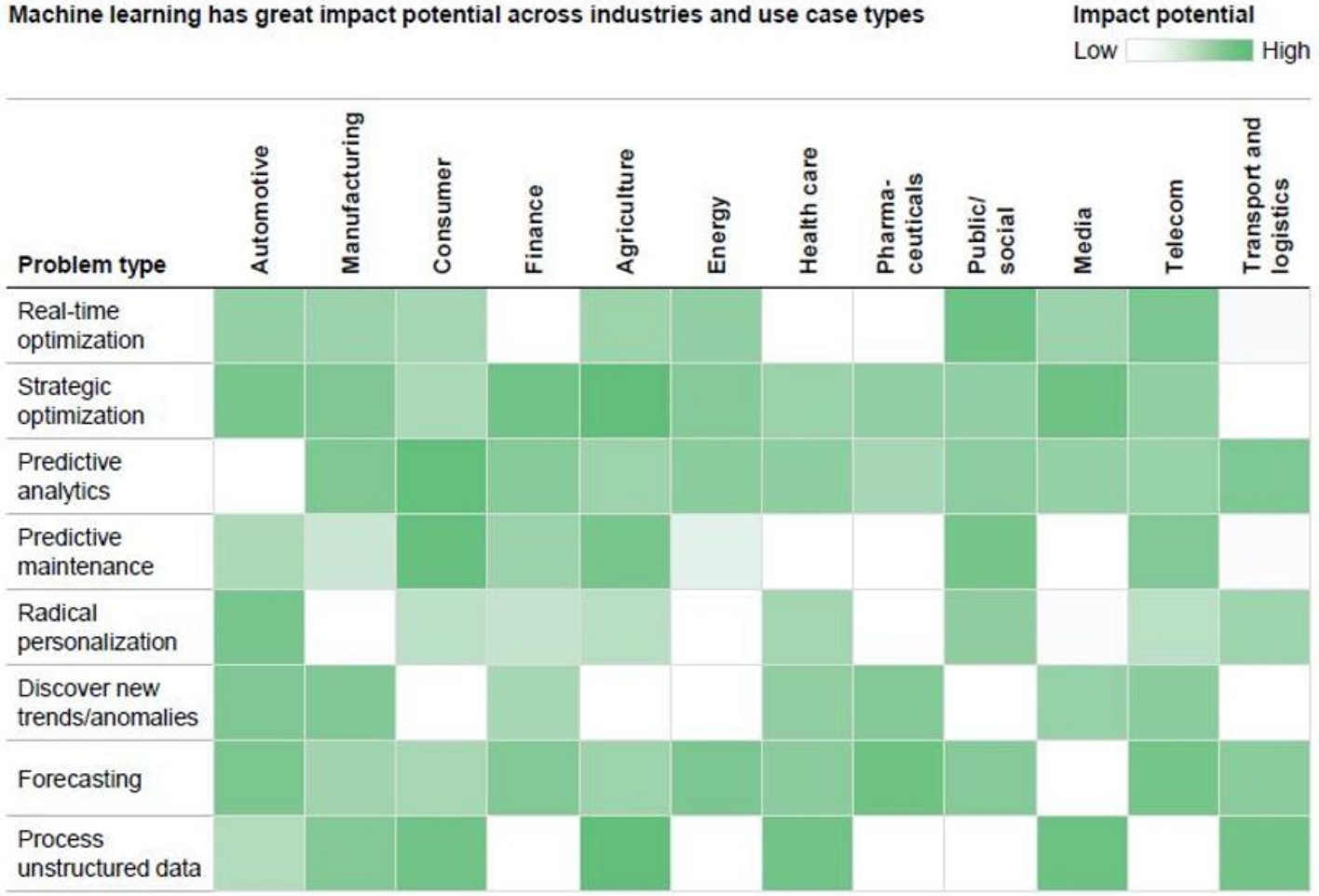
McKinsey&Company

MCKINSEY GLOBAL INSTITUTE

THE AGE OF ANALYTICS: COMPETING IN A DATA-DRIVEN WORLD

DECEMBER 2016

Machine learning has great impact potential across industries and use case types



SOURCE: McKinsey Global Institute analysis

More stories for inspiration…

- **Predictive Policing:** https://en.wikipedia.org/wiki/Predictive_policing
- **Genome Sequencing:** https://www.techemergence.com/machine-learning-in-genomics-applications/
- **Self-correcting Machines**: https://www.ge.com/reports/ge-takes-predix-cloud-edge/
- **AIphaZero:** https://www.extremetech.com/extreme/260215-alphazero-new-chess-champion-harbinger-brave-new-world-ai
- **Self-Driving Cars:** https://en.wikipedia.org/wiki/Autonomous_car

Q4: Business, Data, Math, Technology…hmm...that’s easy…so why should it take time & effort to acquire this knowledge?

Dimensions of Analytics

Business	Use Case Formulation	Interpret Analytics Output	Domain Expertise
Data	Acquisition & Wrangling	Data Visualization & Story Telling	Signals from data (subtract noise)
Math / Quant	Statistical Modeling vs ML	Select the right techniques & code	Evaluating the output of algos
Tech / Software	Understand the IT Ecosystem	Data Engineering / Pipelines	Software Engineering / SDLC

My Analytics Mindmap

- Global Trends in Society
- Macro-economy
- Business Fundamentals
- Specific Industry Domain
- Analytical use cases

- Data Management
- Reporting & Self-service
- Quantitative Techniques
- Performance Mgmt
- Insight Delivery

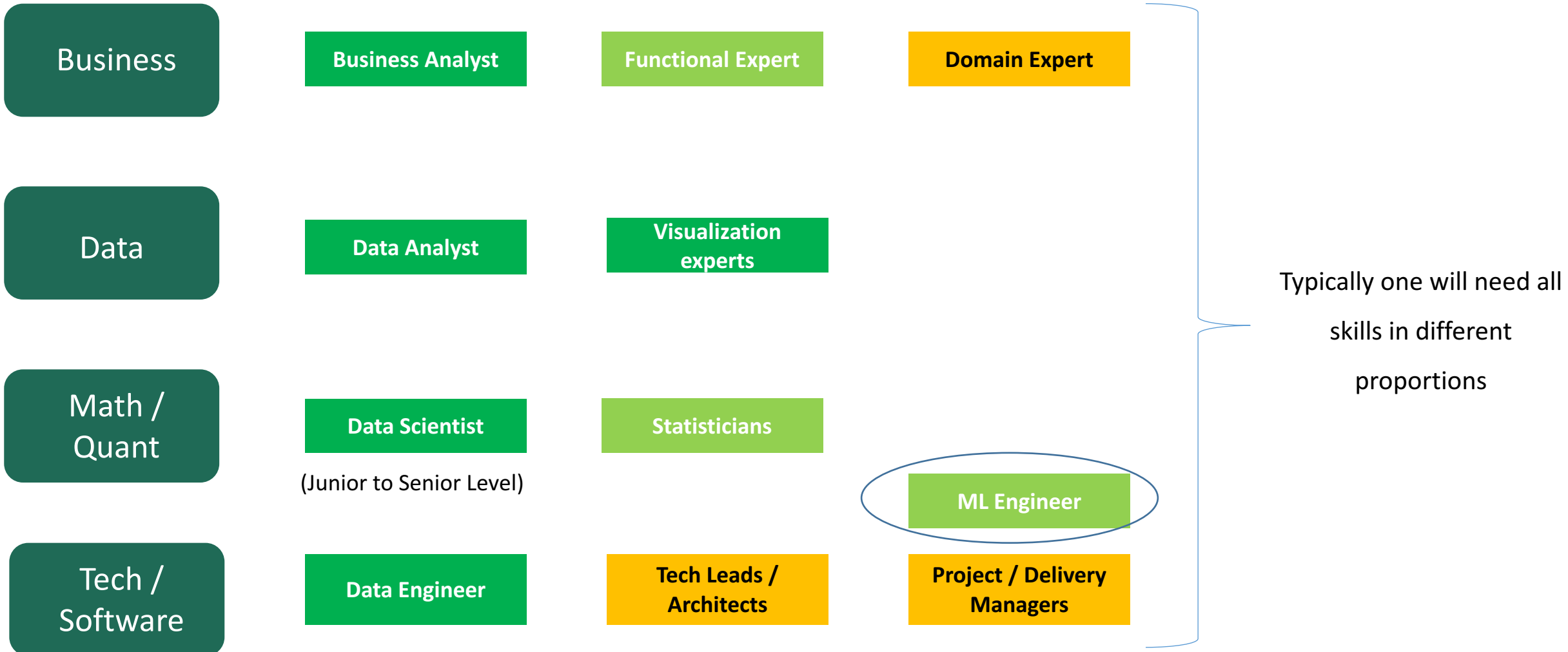
Analytics for Business Value
http://bit.ly/31KArT8

- Scan for New Products
- Evaluate Maturity

- Monitor Ecosystem
- Leverage Resources

Q5: What are the typical roles in the analytics space and what do they do on a daily basis?

Typical Roles in Analytics



Q6: What are the useful components in the data science toolbox?

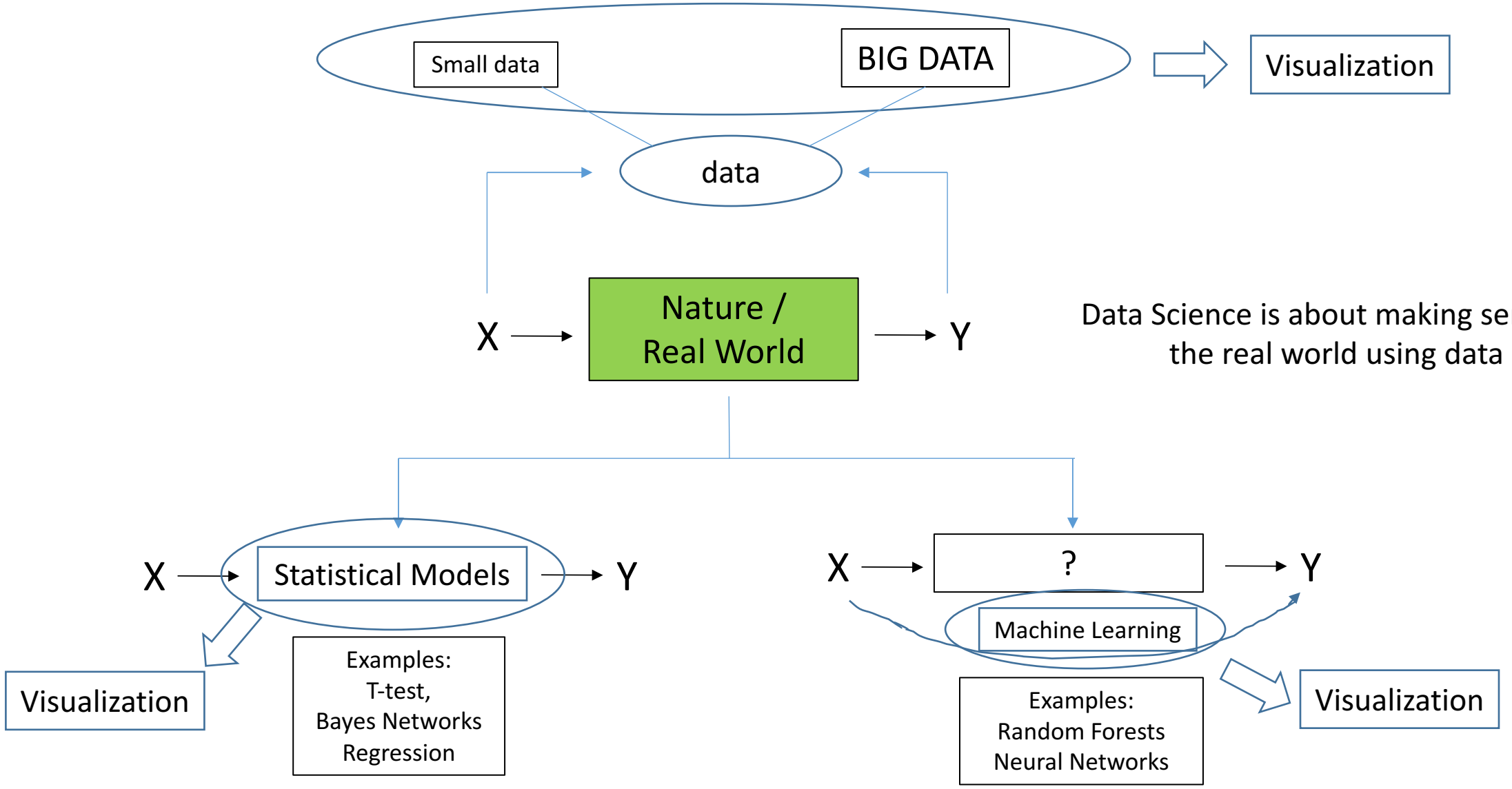
Data Science Toolbox

- Maths / Stats orientation (Not a tool but you know…)
- Atleast 1 programming language – Python (Jupyter notebooks), R
- Atleast 1 GUI based ML platform – H2o, Azure ML, BigML
- 1 Cloud based platform (Nice to have) – AWS, Databricks
- Github
- Kaggle (Competition & Kernels), AnalyticsVidhya
- Database / SQL knowledge (preferable)

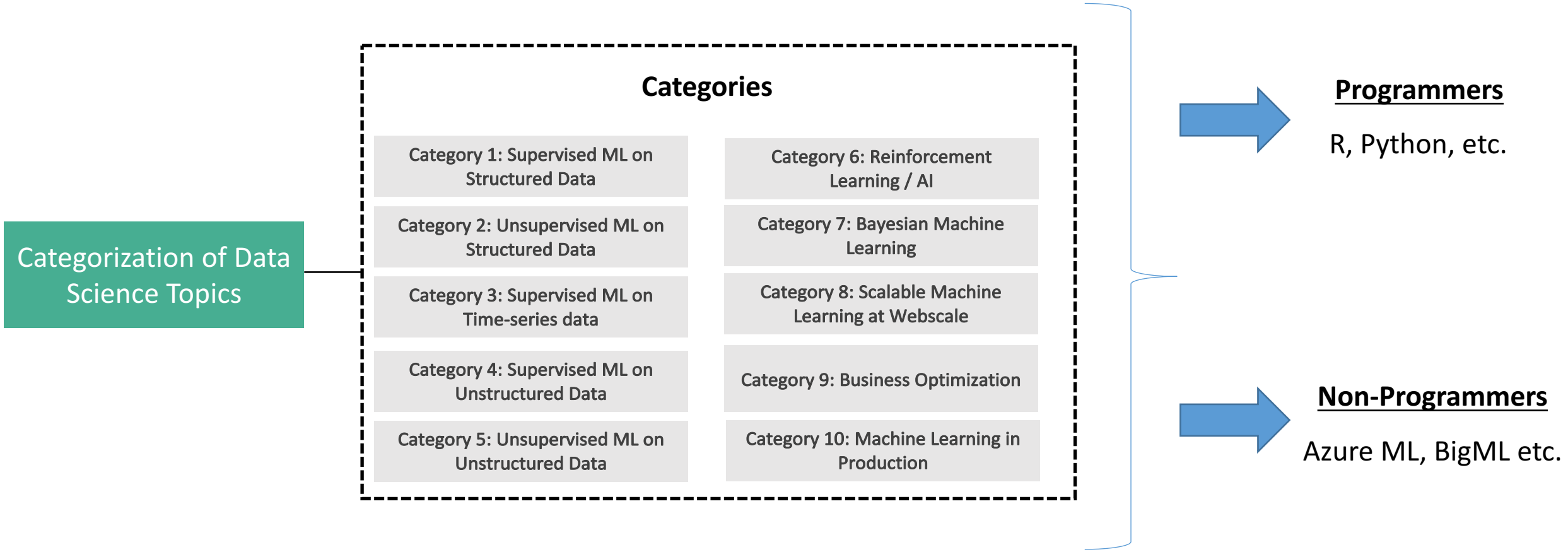
Q7: There could be many techniques and it is not possible to learn everything in a short timeframe? Any tips on how to keep track of them and learn as you go along?

Data Science Techniques – There are a lot of them!

Making Sense of the World (using Data)

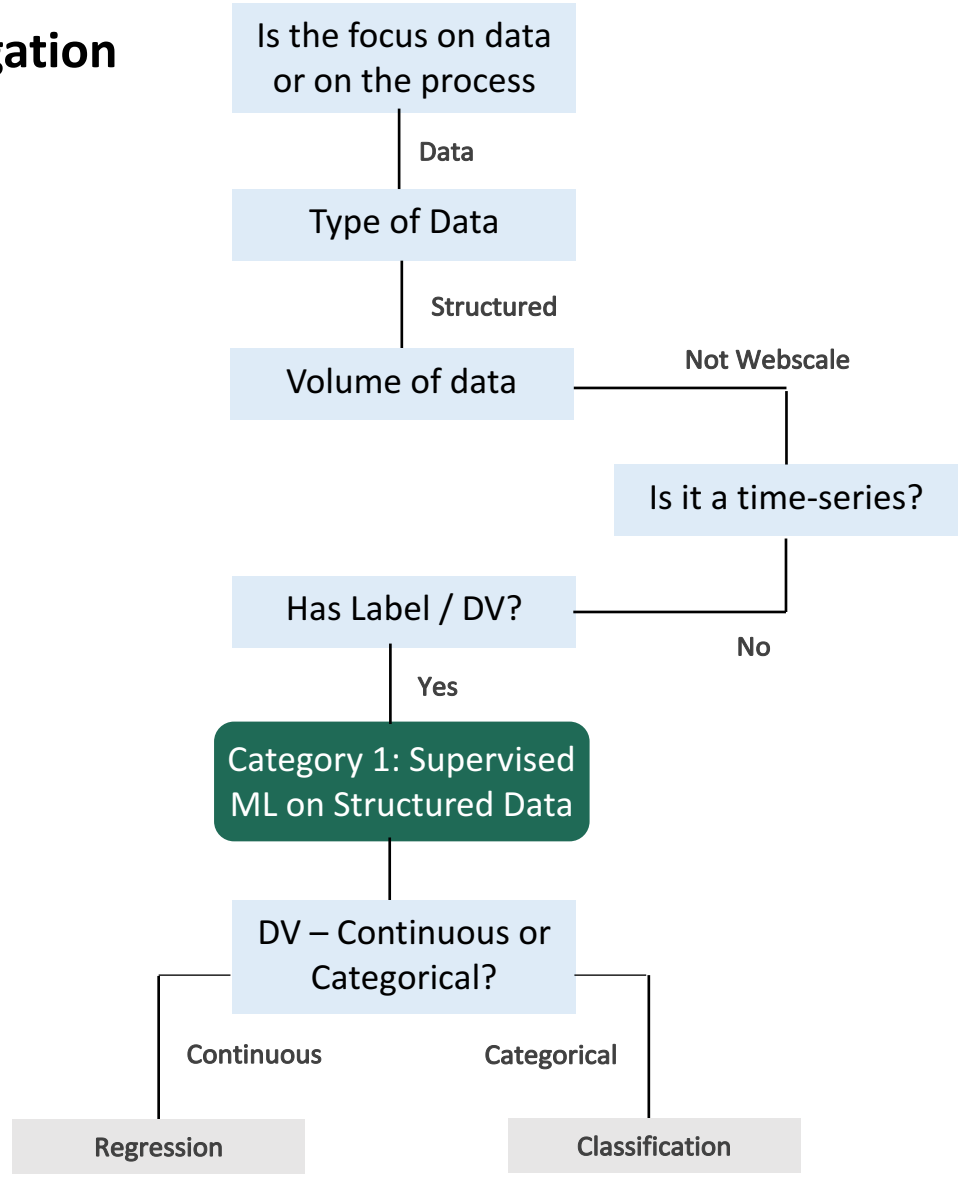


Develop your own personal map



Example: Category 1

Navigation



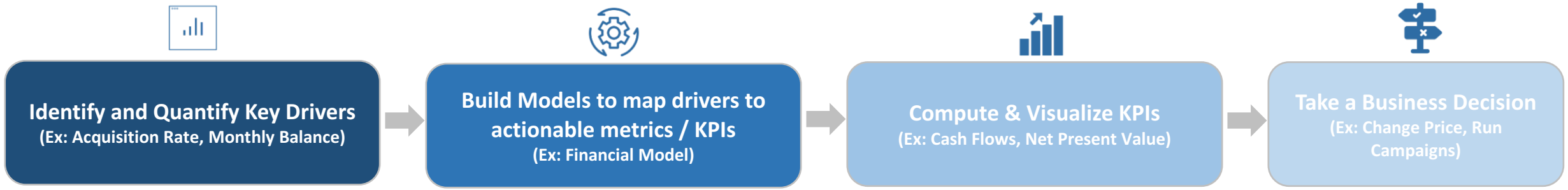
Details

- Exploratory Data Analysis (EDA)
- Data Pre-processing – Outliers, Missing data, Variable Transformations
- Feature Selection & Dimensionality Reduction
- Feature Engineering
- Algorithms – Standalone vs Ensembles
- Algorithms – Parametric vs Non-Parametric
- Algorithms – Linear vs Non-linear
- Cross validation
- Hyper-parameter Tuning
- Predict on Test set

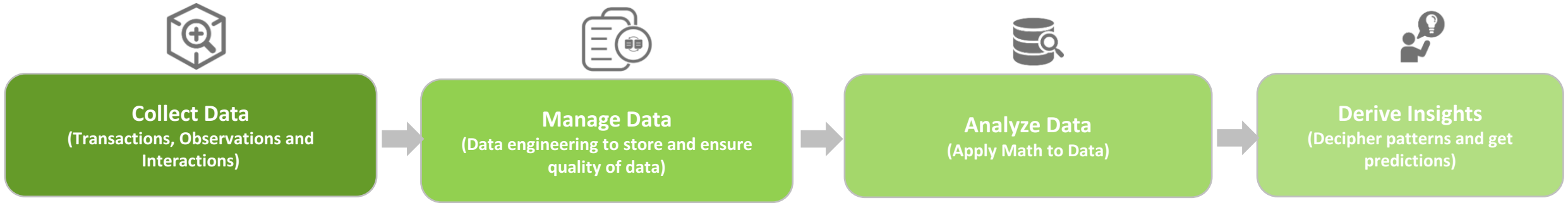
Q8: Other than the analytical techniques themselves, what are the top 2 skills that needs to be developed?

Business Orientation is the cornerstone of Analytics

Business Decision Making Pipeline



Gap that needs to be bridged to deliver maximum business value

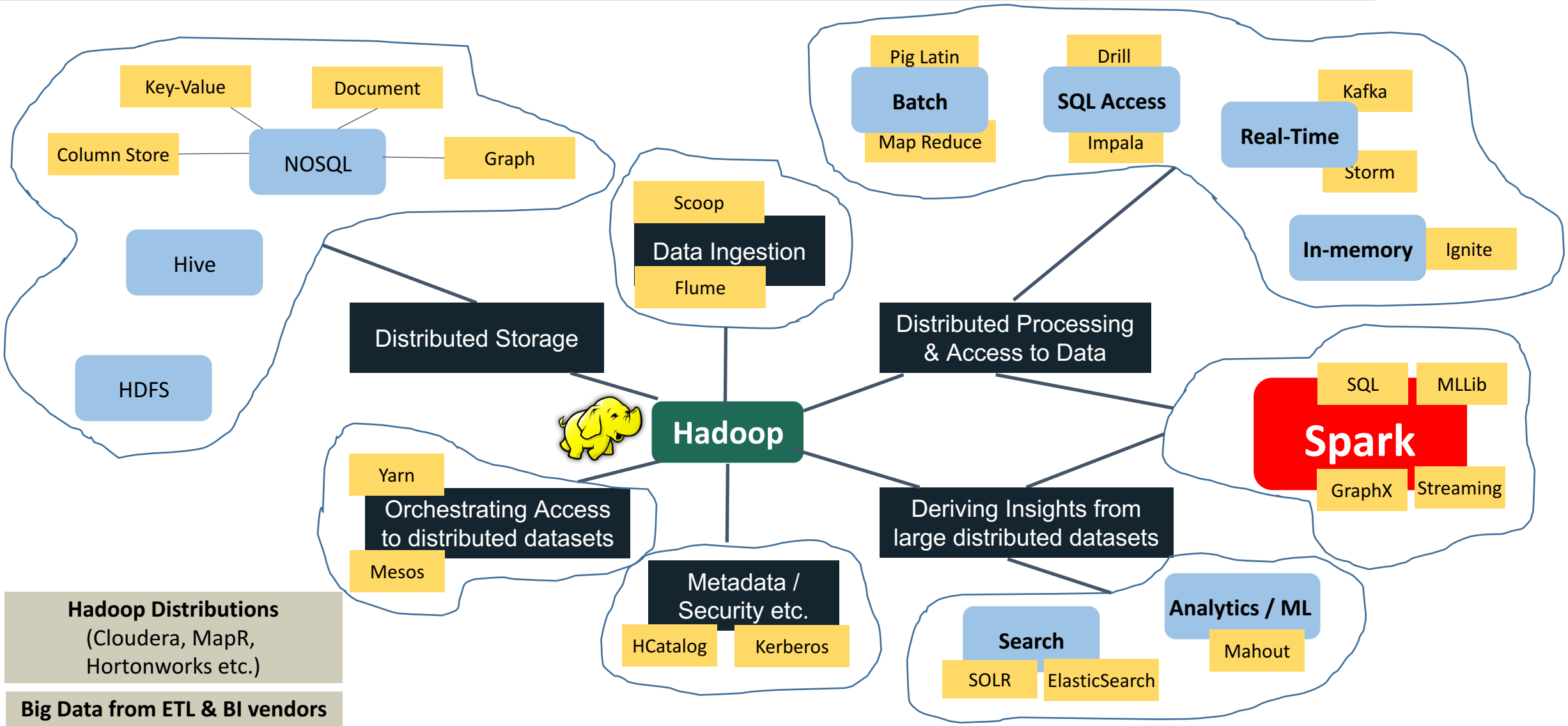


Data Science Pipeline

Think Technology Landscape

- Cloud
- Big Data
- Mobility
- Web Technologies
- Embedded Analytics in Applications
- Legacy Systems

Big Data – Key Technology Enabler



Hadoop Distributions
(Cloudera, MapR, Hortonworks etc.)

Big Data from ETL & BI vendors
(Infa, Teradata, MSTR, Tableau etc.)

Q9: Given the roles discussed earlier, how should one acquire the required skills?

How to make the transition? – Fresher / Developer

	Can Aspire to be	Skills to Acquire	How to Acquire
Fresher / Junior Developer	**Business Analyst**	➢ **Business Orientation** ➢ **Functional Knowledge (in 1 or 2 areas)**	➢ **Domain / Functional Certifications** ➢ **MBA**
	Data Analyst	➢ **SQL Skills / DB knowledge** ➢ **Translate business requirements to data needs** ➢ **Basic Stats knowledge**	➢ **Online Tutorials** ➢ **Technical certifications** ➢ **MOOCs**
	Big Data Engineer	➢ **SQL Skills** ➢ **Hands-on coding expertise in Big Data tools**	➢ **Focused Big Data Courses** ➢ **Online Courses** ➢ **MOOCs**

How to make the transition? – Experienced Techie (6-12 years)

	Can Aspire to be	Skills to Acquire	How to Acquire
Lead / Architect	**Data Engineer**	➢ **Strong SQL & Programming skills in Java, Scala, etc.** ➢ **Design data pipelines for analytics**	➢ **Technical certifications** ➢ **Online Tutorials** ➢ **Focused Courses**
	Big Data / Cloud Specialist	➢ **Design Big Data Systems** ➢ **Expertise in using databases / cloud platforms in the Big Data context**	➢ **Technical certifications in areas of specialization**
	Mid-Level Data Scientist	➢ **Good Stats / Math knowledge** ➢ **Intuitive understanding of algorithms** ➢ **Hands-on coding expertise in ML/Data Science (R, Python etc.)**	➢ **Specialized Analytics Programs** ➢ **MOOCs** ➢ **Build a portfolio of ML projects** ➢ **Online competitions (Ex: Kaggle, AnalyticsVidhya)**

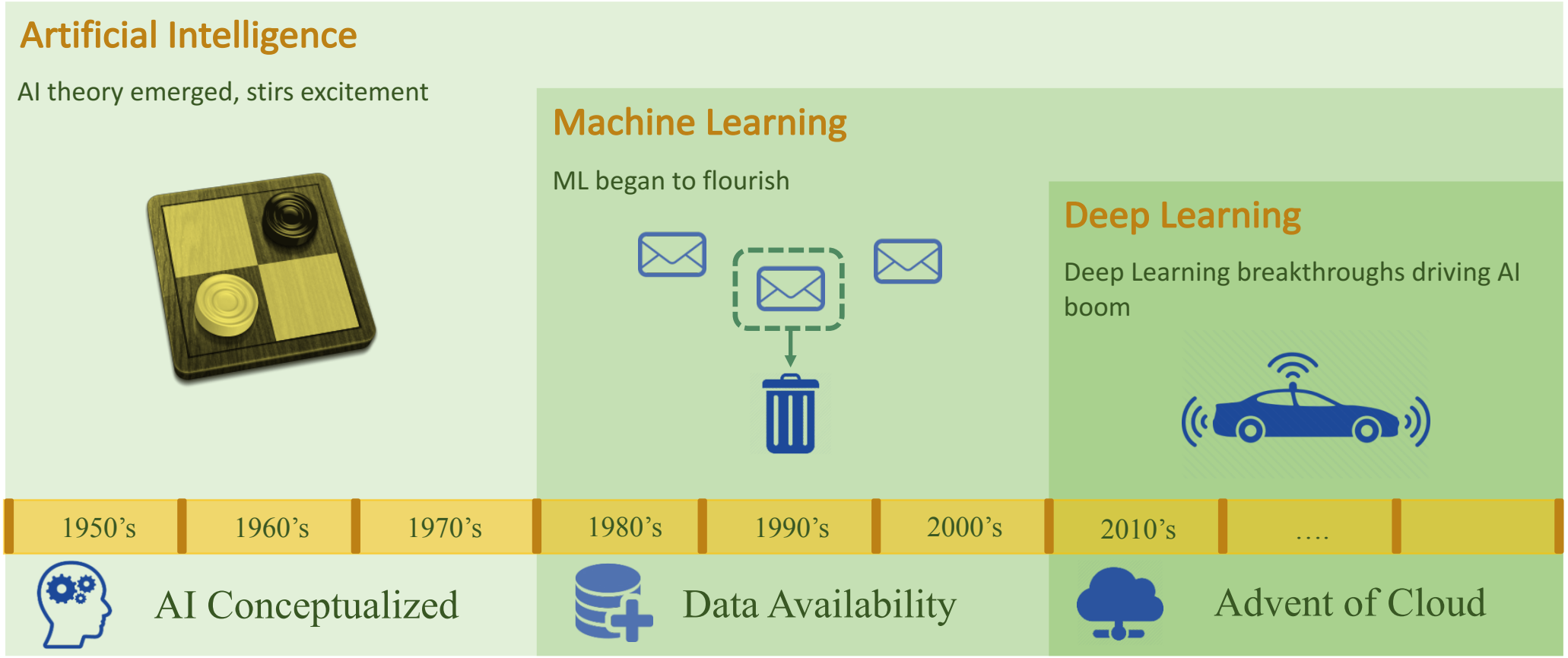
How to make the transition? – Senior Tech Professionals

	Can Aspire to be	Skills to Acquire	How to Acquire
Delivery Manager / Business Head	**Functional / Domain Expert**	➢ **Strong Functional / Domain Knowledge** ➢ **Conceptual Knowledge of Analytics**	➢ **Executive MBA programs** ➢ **Specialized Analytics programs (Online / Offline)**
	Project / Delivery Manager	➢ **SDLC as applicable to analytics / big data projects** ➢ **Conceptual knowledge of Business + Data + Math**	➢ **On the job** ➢ **MOOCs (Case study based approach)**
	Mid-Level Data Scientist	➢ **Good Stats / Math knowledge** ➢ **Intuitive understanding of algorithms** ➢ **Hands-on coding expertise in ML/Data Science (R, Python etc.)**	➢ **Specialized Analytics Programs** ➢ **MOOCs** ➢ **Build a portfolio of ML projects** ➢ **Online competitions (Ex: Kaggle, AnalyticsVidhya)**

Q10: In the industry whenever people speak of Analytics, they now talk about Artificial Intelligence. How is this course related to AI, if at all?

What is Artificial Intelligence?

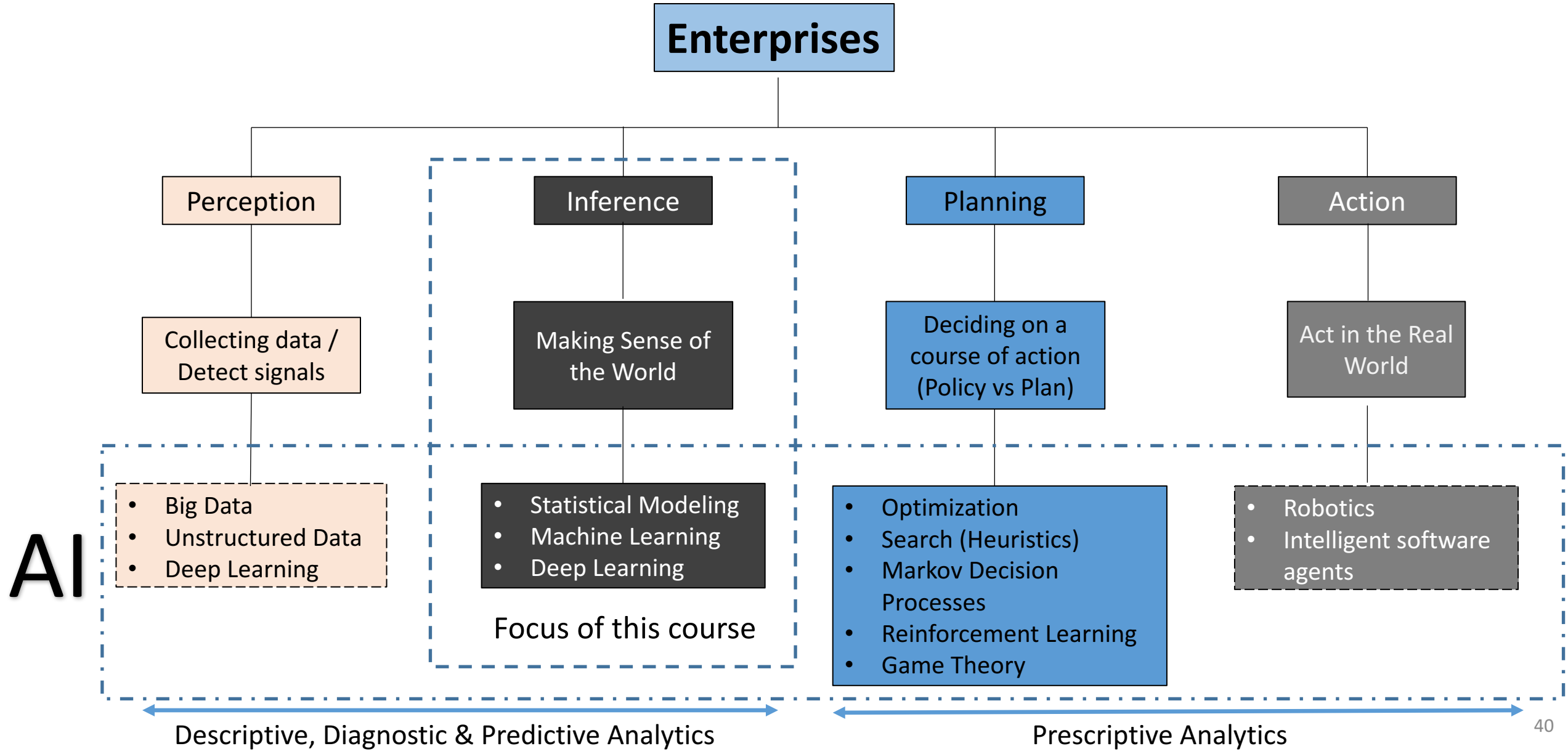
Artificial Intelligence refers to the theory and development of computer systems & machines with the ability to perform tasks normally requiring human intelligence



What constitutes Human Intelligence?

1. Perceive the world, detect signals and collect data
2. Make sense of the world using data (Insights, Inference, Predictions etc.)
3. Decide on the next course of action (Planning)
4. Act in the Real World

AI Techniques in Enterprises – Parallels to Human Intelligence



Yipee! Let's get going on Analytics

Strong Motivation – Data Science is a journey

Curiosity – Ask yourself, others & internet the right questions

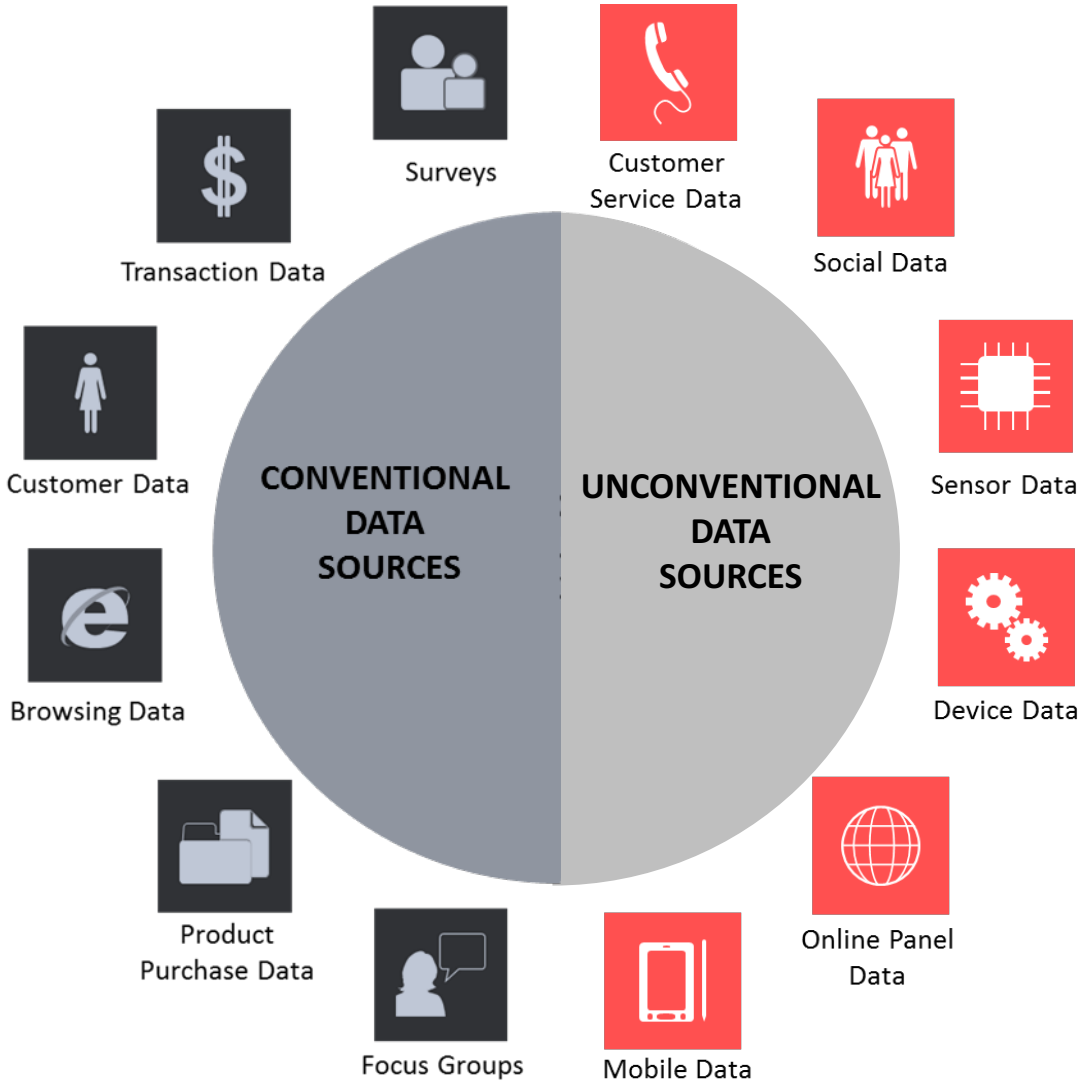
Connecting the Dots – Learn & Assimilate

Skill - Should enjoy working with numbers

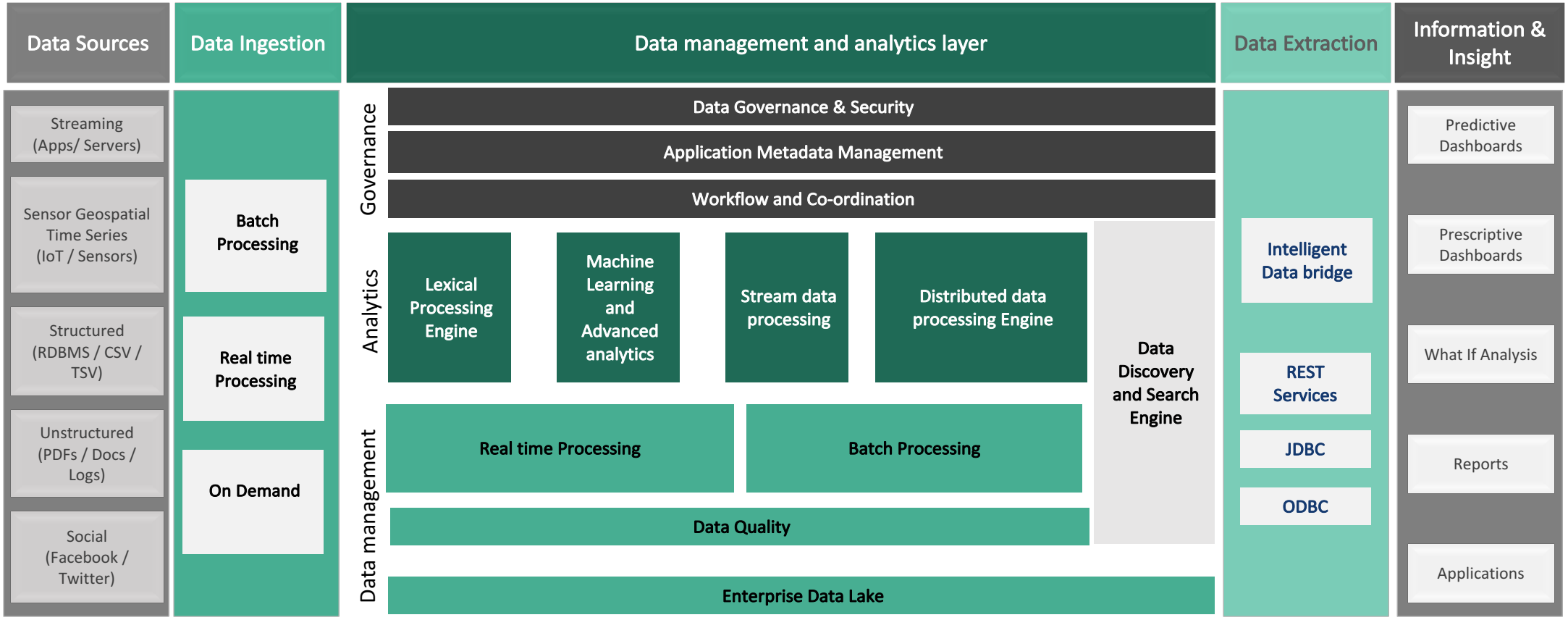
Q & A

- Karthikeyan Sankaran, Director, LatentView Analytics
- Email ID – Karthikeyan.Sankaran@latentview.com
- Mindmap – bit.ly/31KArT8
- LinkedIn – http://in.linkedin.com/in/karthikeyansankaran
- Tapchief - www.tapchief.com/karthik

Companies are looking beyond conventional data sources…



A Reference Architecture always helps…



A Big Data Platform Reference Architecture for Analytical Solutions

Why do we need a different technology paradigm?

Category	Traditional View	Issues	Big Data Paradigm
Volume	Data handled by scaling up SMP relational databases	• Limited Scalability (order of Terabytes)	➢ Distributed 'Scale-out' architecture ➢ Commodity Hardware (Cost effective) ➢ Robust software controlled data distribution, redundancy
	Data handled by using MPP relational databases	• Poor scalability / cost ratio • Specialized Hardware • Vendor lock-in	
Variety	Structured data handled with schema created at 'design' time	• Data types that do not fit the designed schema cannot be ingested	➢ Schema imposed at run-time allows all types of data to be ingested into the system ➢ Distributed high-performance processing capability (Ex:MapReduce, Spark etc.) on all types of data
	Unstructured data handled as blob objects inside RDBMS	• No processing capability	
Velocity	Utilize Messaging Infrastructure / Real Time ETL Flows	'True' Real-time is still not achievable (my personal experience)	Big Data based streaming platforms (Ex: Spark Streaming, Storm etc.) are built ground-up for real-time stream processing
Veracity	Only cleansed / processed data stored in enterprise systems	Lose out on interesting opportunities related to scenarios using messy, real-world data	Platform ingests processed data and also messy real-world data. Cost effective processing techniques helps to discover new insights

Why are we here?

Big Data - The Opportunity

IDC expects the Big Data and Analytics market to grow from $130 billion in 2016 to more than $203 billion in 2020

According to Forbes, Big Data & Business Analytics will grow from $122 billion in 2015 to more than $187 billion in 2019, an increase of more than 50%

Big Data - The Challenge (which makes it interesting)