

DATA-DRIVEN IMMERSION

Alex Campos | Digital Data Strategist
alex@scroll.cl

AGENDA

- Introduction
- Digital Transformation
- Data-Driven Culture
- Personas
- Methodology
- Data Governance
- How to Start

INTRODUCTION

CHAPTER TOPICS

Introduction

- **Introduction**
- Course Objectives

INTRODUCTION

- Alex Campos
- Computer Engineer, UDP
- Just moved from Chile/Latam to Iberia/Europa
- Enthusiast about data, analytics, agility and digitalization
- Solutions Engineer @ Cloudera



<https://www.linkedin.com/in/campossalex/>

CHAPTER TOPICS

Introduction

- Introductions
- **Course Objectives**

COURSE OBJECTIVES

During this course, you will learn:

- Why data is so important to Digital Transformation
- Main aspects to become a Data-Driven company
- How to build and develop analytical teams
- How Agile Methodology can improve analytical projects outcomes
- Important aspects to consider when implement Data Governance
- How to start analytics project from use cases perspectives

DIGITAL TRANSFORMATION

CHAPTER TOPICS

Digital Transformation

- **Why Transform?**
- Key Drivers for Digital Transformation
- Challenges

WHY TRANSFORM

Digitalization is a moving target

Business

Rethinking Business Model

Perhaps the first and most visible effects from Digitalization is the imperative of “Do Business Differently”. The capability to scale, in product and services, but at the same time be efficient without trade off quality and price.

Process

Respond quickly to market changes

Companies should find opportunities to automate business process to improve performance. Technologies will drive the company to be more agile, integrated, transparent, making easy to identify bottleneck that slowdown performance.

Customer

End-to-end experience

Clients are digital, using new technologies to interact with the companies. Analyzing all the data generated from those touch points leverage omnichannel and “segment of one”, improving customer journey.

CHAPTER TOPICS

Digital Transformation

- Why Transform?
- **Key Drivers for Digital Transformation**
- Challenges

KEY DRIVERS FOR DIGITAL TRANSFORMATION

Digitalization is a moving target



Data
Deluge



Digital
Customer



Global
Connectivity



Disruptive
Technologies

CHAPTER TOPICS

Digital Transformation

- Why Transform?
- Key Drivers for Digital Transformation
- **Challenges**

CHALLENGES

Introduction

The Digitalization process is not that easy. "Do business differently" require companies to shift from a traditional way of operation and introduce new techniques like lean startup, agile methodologies, culture changes and implement new disruptive technologies.

Business
Strategy

Technology

Analytics

Skilled
People

DIGITAL TRANSFORMATION

One Statement

Digital Transformation is the new way of **doing business** and engage with customers in this **digital era**. So, by **digital era** we understand of products and services that are **developed** in few weeks, scale very fast and have an **aggressively adoption** in the market. To keep the pace in the digital era, companies need a **data platform** flexible enough to ingest, store and process all the data to **better understand customers, monetize insights** and implements **cross and governed** data-driven culture.

DATA-DRIVEN CULTURE

CHAPTER TOPICS

Data-driven Culture

- **What they do well**
- Data Strategy
- Democratizing Data

WHAT THEY DO WELL

Implement a truly Data-driven Culture means:

- Be Data-driven is an **organizational capability**, not a technical practice
- Incorporate analytics at **different levels** of complexity and sophistication
- The capability to include data and analytics from **operational to strategical decisions**
- Intensive and extensively use of data for **research and new product/service development**
- Manage analytical projects **across the organization**, not as “individual” solutions
- Manage data as an **asset**
- Data is **democratized and accessible** across all company

WHAT THEY DO WELL

Five Key Success Factors

1. Build a Big Data Culture

Led by an enabled executive sponsor(s). Communication methodologies. Advocating change.

2. Assemble the right team

Tightly aligned team. Mix of seasoned experts and innovators

3. Adopt an agile approach for data engineering, data science, analysis

Successful projects start small, are hypothesis driven and iterate to success approach.
Roadmaps: Document expected direction, yet expect insights to create change

4. Efficiently operationalize insights

Analytics -> Reports, Big Data -> Actions. Create a bridge between Dev and Ops

5. Rightsize your data governance

Rightsize and iteratively building towards maturity.

CHAPTER TOPICS

Data-driven Culture

- What they do well
- **Data Strategy**
- Democratizing Data

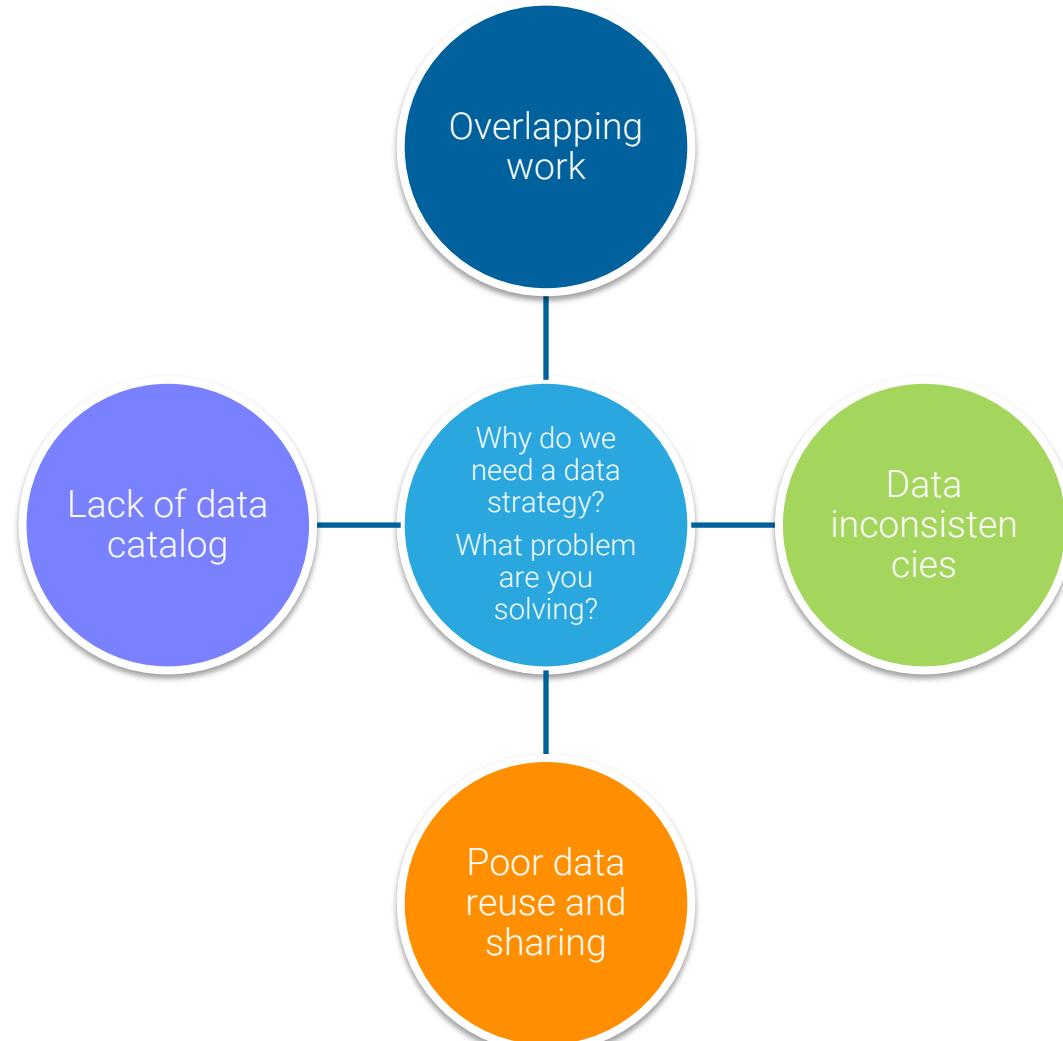
DATA STRATEGY



Why do we need a
data strategy?

What problem are
you solving?

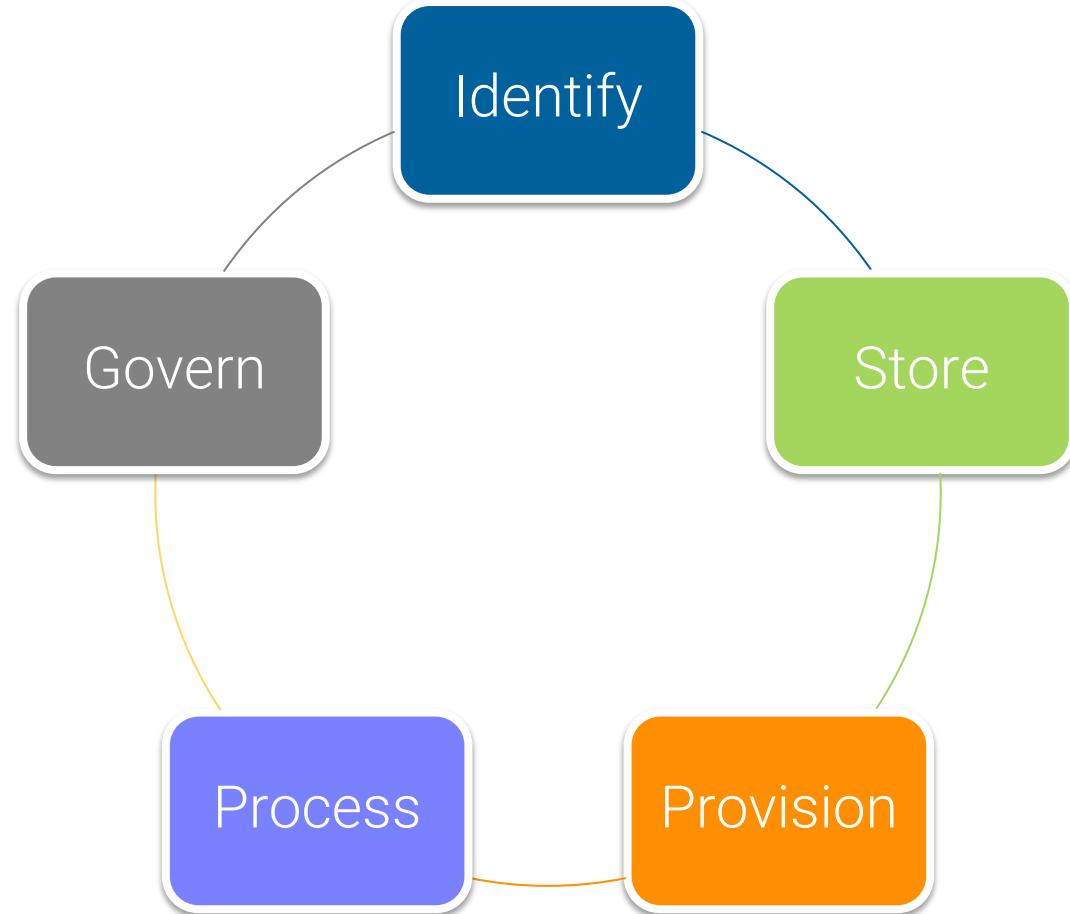
DATA STRATEGY



DATA STRATEGY

A data strategy is a plan designed to improve all the ways you **acquire, store, manage, share** and **use** data.

DATA STRATEGY



CHAPTER TOPICS

Data-driven Culture

- What they do well
- Data Strategy
- **Democratizing Data**

DEMOCRATIZING DATA

Data Democracy means that various teams can **use** data for innovation regardless of their **level or technical** skill.

DEMOCRATIZING DATA

Self service tools

- Visual tools
- SQL access
- From raw to aggregated
- Compliance, privacy and security

Supporting users

- Data Ops Teams
- Improve communication
- Build strong relationships
- Care about data products performance/usage

PERSONAS

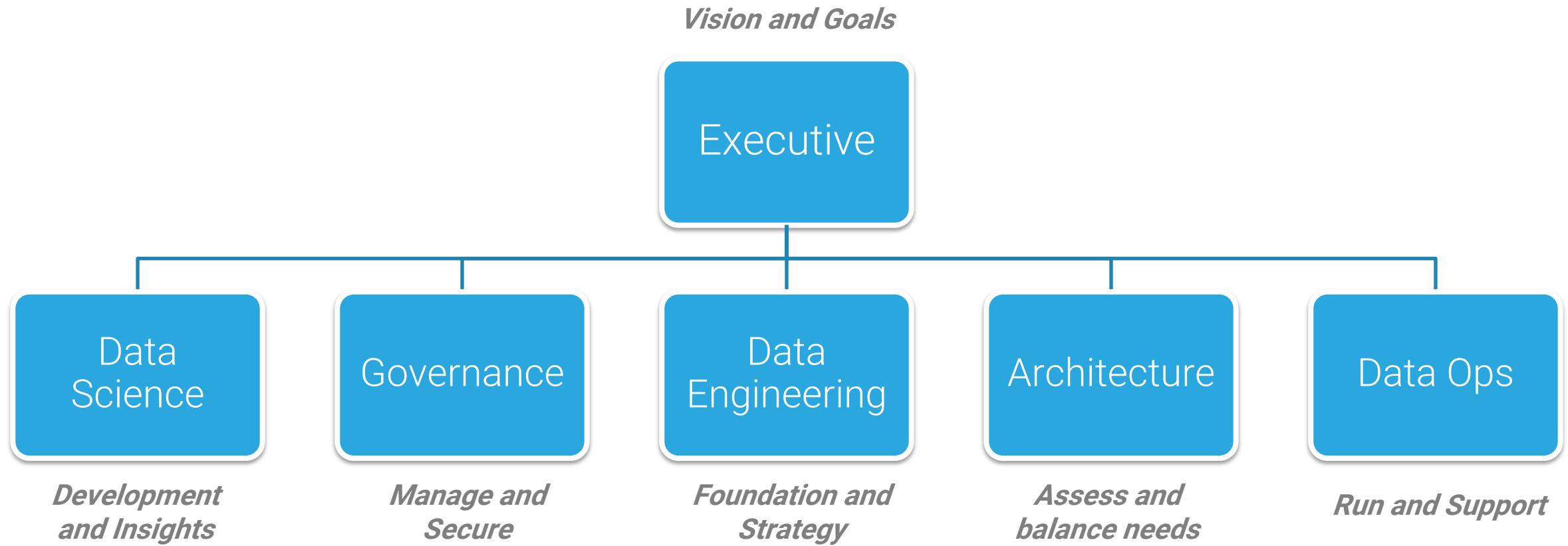
CHAPTER TOPICS

Personas

- **Introduction**
- Organizational Structure
- Roles

INTRODUCTION

Big Data Functional Areas of Responsibility

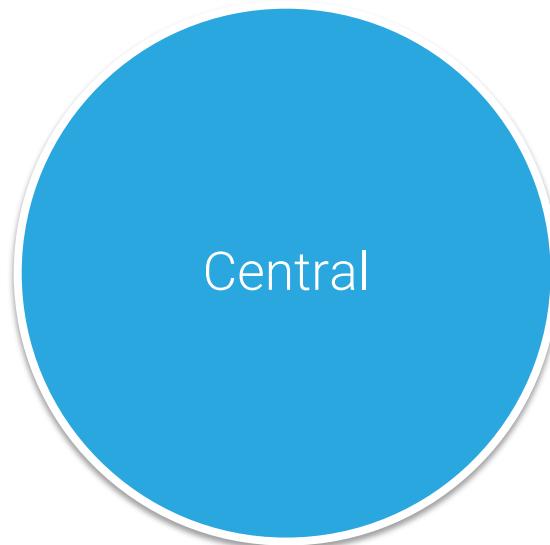


CHAPTER TOPICS

Personas

- Introduction
- **Organizational Structure**
- Roles

ORGANIZATIONAL STRUCTURE



Versus



ORGANIZATIONAL STRUCTURE

Hub-and-Spoke Structure



ORGANIZATIONAL STRUCTURE

Hub-and-Spoke Structure

- Central Data Team is responsible for publishing most important data sets, making sure to create a “single source of truth”.
- Support self service capabilities
- Avoid duplicate effort, such as ingesting, data prep and models
- Increase connectivity between departments and avoid data silos
- Data Analyst and Data Scientist are embedded in different departments helping to resolve specific challenges/questions
- Data Ops supporting projects maintenance and final user activity

CHAPTER TOPICS

Personas

- Introduction
- Organizational Structure
- **Roles**

ROLES

Platform Architecture

Professional certified in Big Data Technologies, with the experience and knowledge necessary to carry out activities related to the selection of Big Data platforms for the storage and processing of large volumes of data, both in batch process and real-time cases. Additional supports platform evolution and expansion, always taking into account the business cases, characteristics of the workloads and performance of the components.

Tools: Big Data concepts / scalability, Hadoop, NoSQL, Search Engine, Database, hardware, operating system, distributed computing

ROLES

Data Engineer

Professional certified in Big Data technologies, with the experience and knowledge necessary to perform data ingestion, integration, storage and descriptive analysis of structured and unstructured data. It also uses visualization tools to display reports for prototyping.

Tools: Hadoop, MR, Sqoop, Hive / Pig, Impala, SQL, Python, API, NoSQL, ETL, Database, Spark

ROLES

Data Scientist

Professional with experience and knowledge in statistics, programming and business. It builds, develops and implements statistical models that allow generating predictive insights to support decision making, using data mining techniques and machine learning. This profile is very relevant when the organization is in an important stage of maturity in analytical processes, and wishes to develop advanced analytics.

Tools: Python, Scala, Hadoop, R, Spark, SQL

ROLES

Business Analyst

Professional with experience in the business domain, with easy to understand the present challenges of the organizations and to translate these needs and opportunities to solutions that allow the business to take advantage of the data for the decision making

Tools: reporting, SQL, process modeling tools, flow diagrams, user interviews

ROLES

Domain Expert

Professional with large deep and understanding in a particular area, process or activity. They mix practical and theoretical experience in some fields.

They contribute in Analytical projects, delivering guide lines and best practices to resolve business issues.

ROLES

UX Engineer

Professional responsible to define best ways to deliver usability and final products and services, enhancing user experience and increasing consumption.

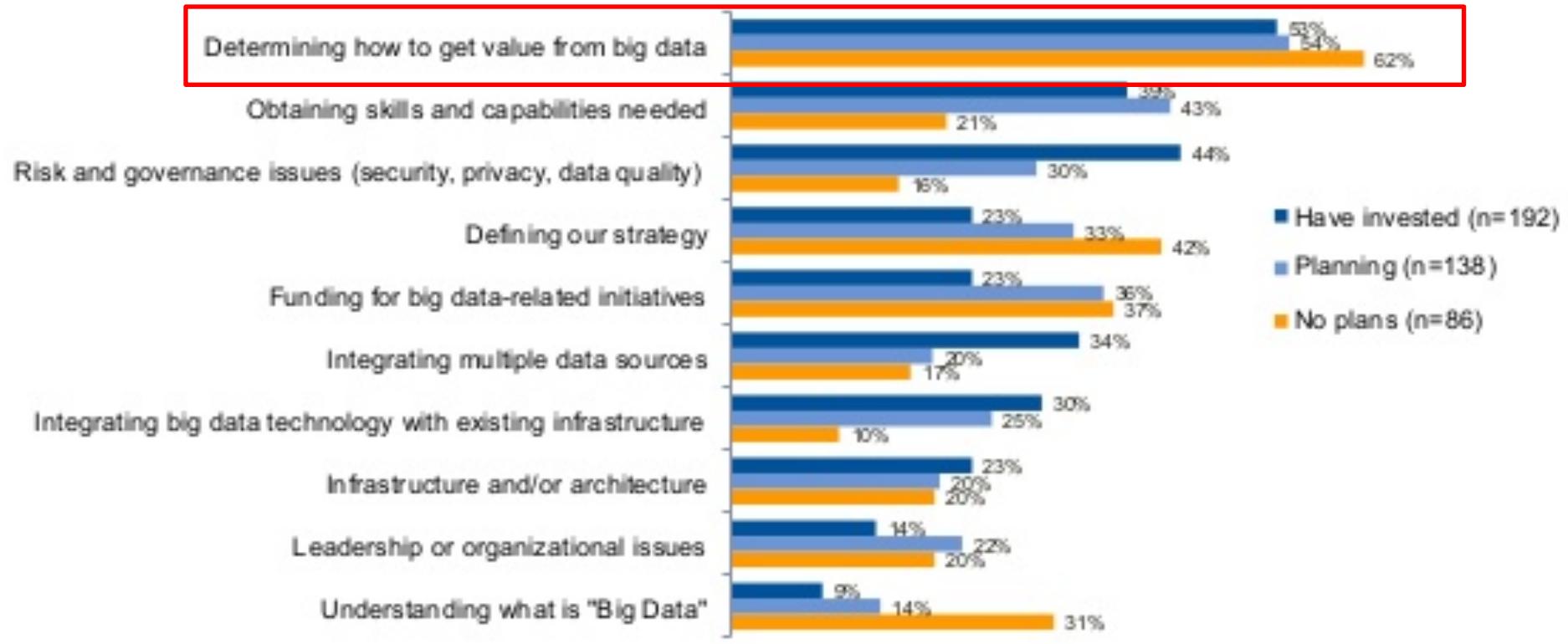
METHODOLOGY

CHAPTER TOPICS

Methodology

- **Why do we need a methodology?**
- CRISP-DM

WHY DO WE NEED A METHODOLOGY



Top Big Data Challenges. 2015

WHY DO WE NEED A METHODOLOGY

- Big Data and Analytics project are more like research
- Stay focus on the business (instead of data)
- An agile approach give flexibility to the process
- Prioritize simultaneous projects and requirements (create a roadmap)
- Understand that a “negative answer” is a possible analytics outcome
 - Fail fast

CHAPTER TOPICS

Methodology

- Why do we need a methodology?
- **CRISP-DM**

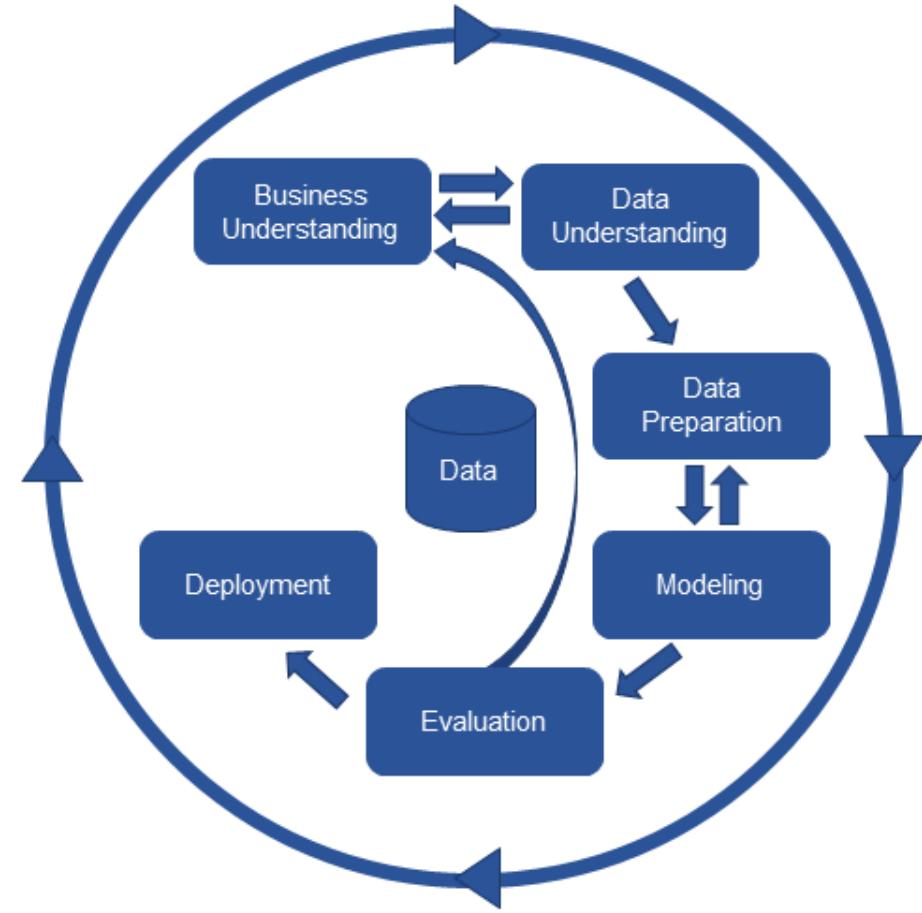
CRISP-DM

- **Cross Industry Standard Process for Data Mining**
- Developed in 1996 by big players in data analysis
- Most popular methodology for data-centric projects
- Agile approach
 - Introduces almost no overhead
 - Emphasizes adaptive transitions between project phases

CRISP-DM

Composed of six consecutive phases:

- **Step 1:** Business Understanding
- **Step 2:** Data Understanding
- **Step 3:** Data Preparation
- **Step 4:** Model Building
- **Step 5:** Testing and Evaluation
- **Step 6:** Deployment



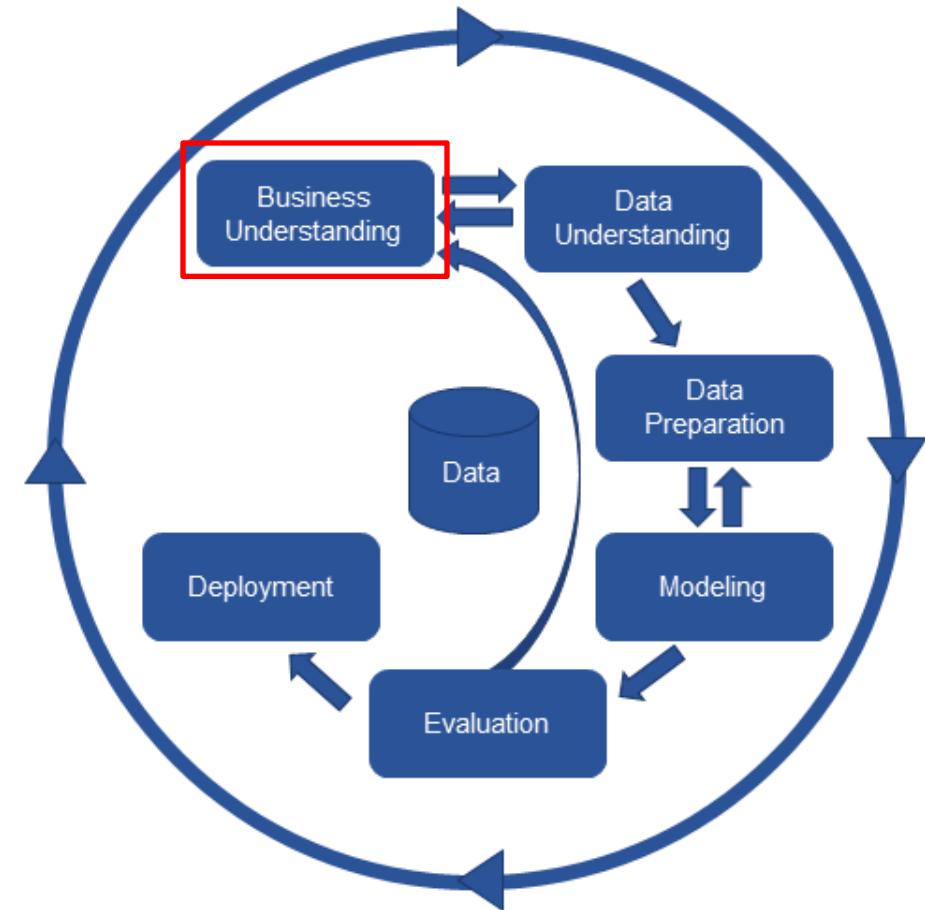
CRISP-DM

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
Determine Business Objectives <i>Background</i> <i>Business Objectives</i> <i>Business Success Criteria</i>	Collect Initial Data <i>Initial Data Collection Report</i>	Select Data <i>Rationale for Inclusion/Exclusion</i>	Select Modeling Techniques <i>Modeling Technique</i> <i>Modeling Assumptions</i>	Evaluate Results <i>Assessment of Data Mining Results w.r.t. Business Success Criteria</i> <i>Approved Models</i>	Plan Deployment <i>Deployment Plan</i>
Assess Situation <i>Inventory of Resources Requirements, Assumptions, and Constraints</i> <i>Risks and Contingencies</i> <i>Terminology</i> <i>Costs and Benefits</i>	Describe Data <i>Data Description Report</i>	Clean Data <i>Data Cleaning Report</i>	Generate Test Design <i>Test Design</i>	Review Process <i>Review of Process</i>	Plan Monitoring and Maintenance <i>Monitoring and Maintenance Plan</i>
Determine Data Mining Goals <i>Data Mining Goals</i> <i>Data Mining Success Criteria</i>	Explore Data <i>Data Exploration Report</i>	Construct Data <i>Derived Attributes</i> <i>Generated Records</i>	Build Model <i>Parameter Settings</i> <i>Models</i> <i>Model Descriptions</i>	Determine Next Steps <i>List of Possible Actions Decision</i>	Produce Final Report <i>Final Report</i> <i>Final Presentation</i>
Produce Project Plan <i>Project Plan</i> <i>Initial Assessment of Tools and Techniques</i>	Verify Data Quality <i>Data Quality Report</i>	Integrate Data <i>Merged Data</i>	Format Data <i>Reformatted Data</i>	Assess Model <i>Model Assessment</i> <i>Revised Parameter Settings</i>	Review Project <i>Experience Documentation</i>
		Dataset <i>Dataset Description</i>			Prescriptive Analytics

CRISP-DM

Business Understanding

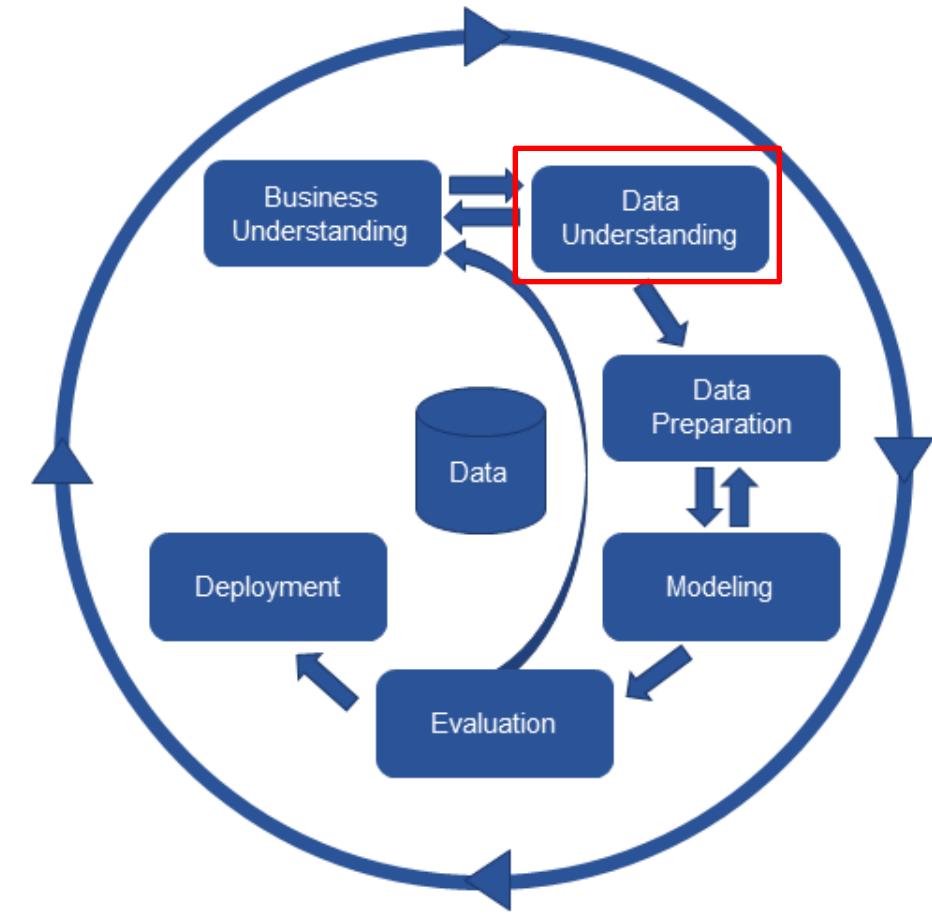
- Determine business objectives
- Assess situation
 - Resources, risks, costs, benefits
- Determine data mining goals
 - Ideally with quantitative success criteria
- Develop a project plan
 - Estimate timeline, budget, but also tools and techniques



CRISP-DM

Data Understanding

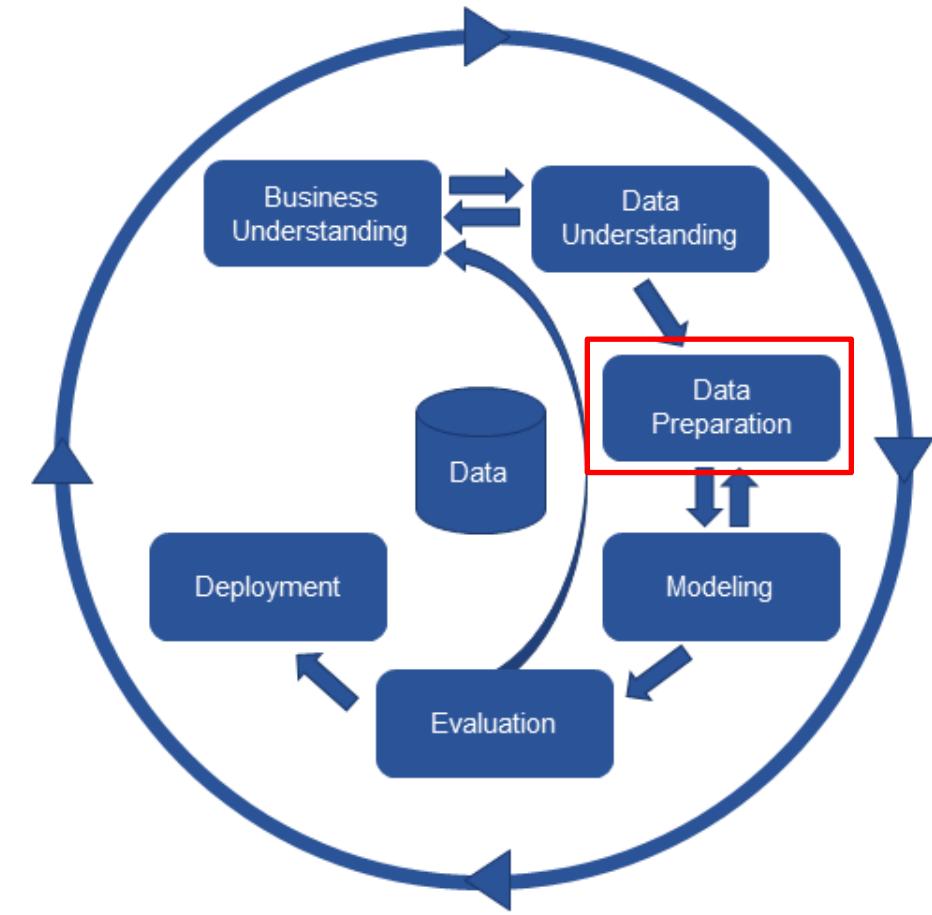
- Collect initial data
- Describe data
 - Important to count with a data catalog
- Explore data
 - Self service tools
- Verify data quality



CRISP-DM

Data Preparation

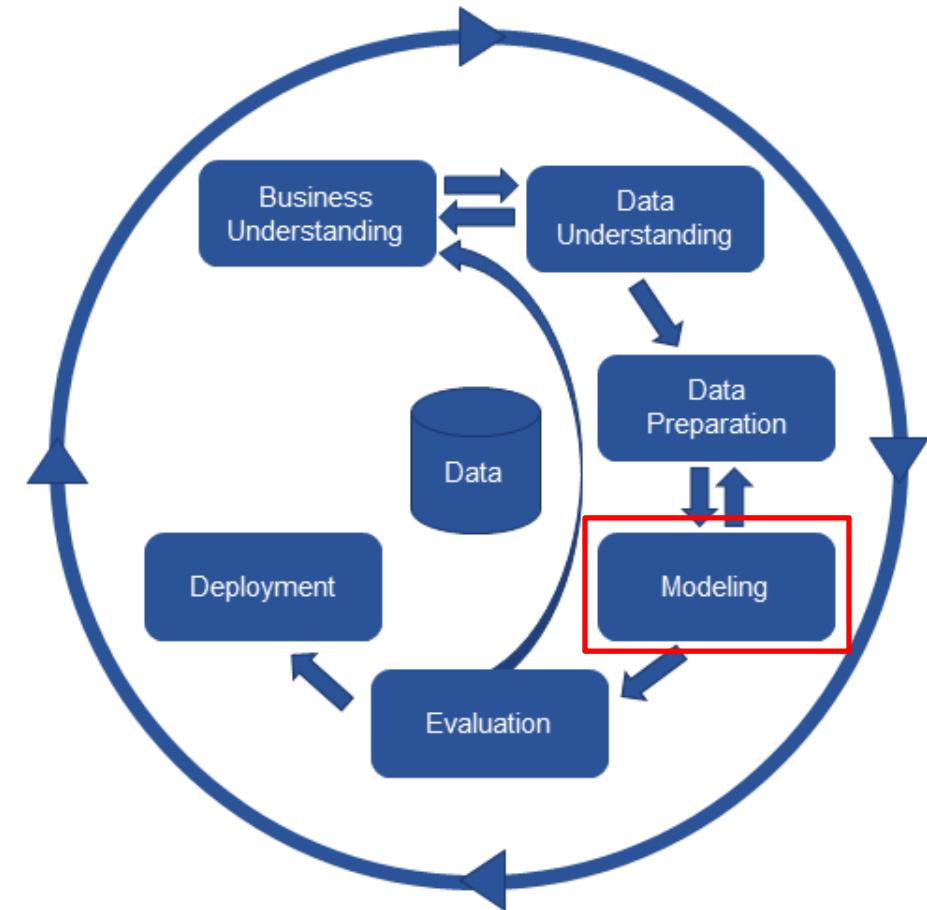
- Select data
- Clean data
- Construct data
 - Generate derived results
- Integrate data
 - Merge information from diverse sources
- Format data
 - Convert to format convenient for modelling



CRISP-DM

Modelling

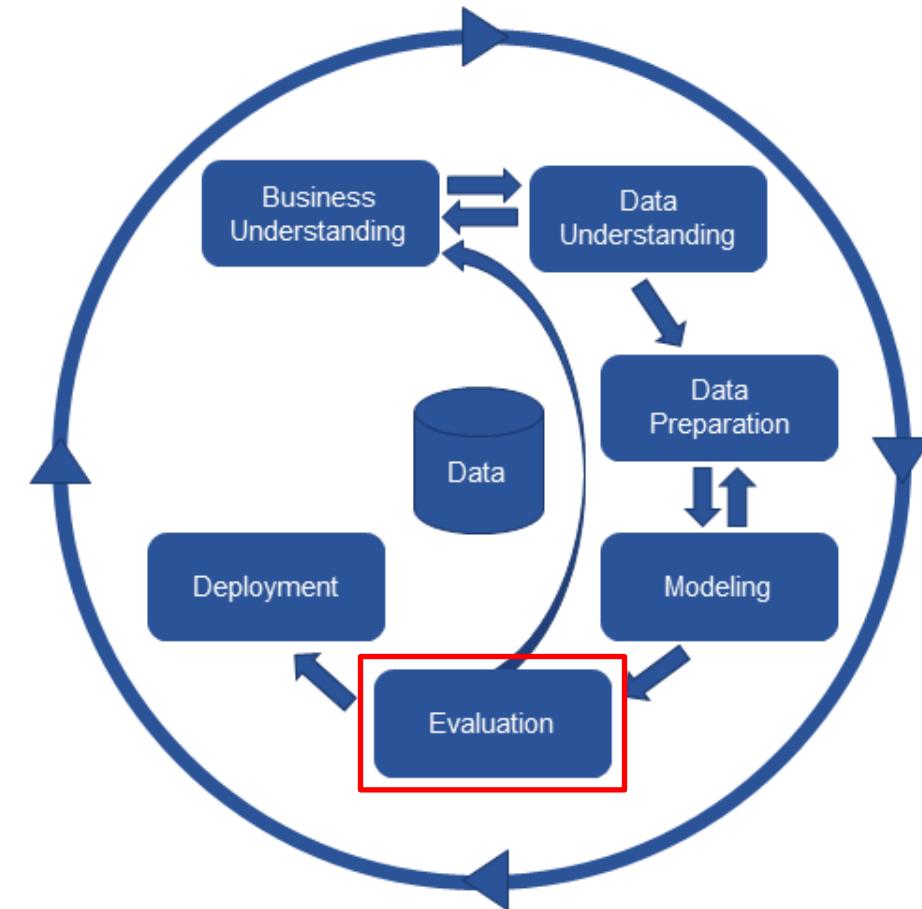
- Select modelling techniques
 - Assumptions, measure of accuracy
- Generate test design
- Build model
 - Feature engineering, optimize model hyper parameters
- Assess model



CRISP-DM

Evaluation

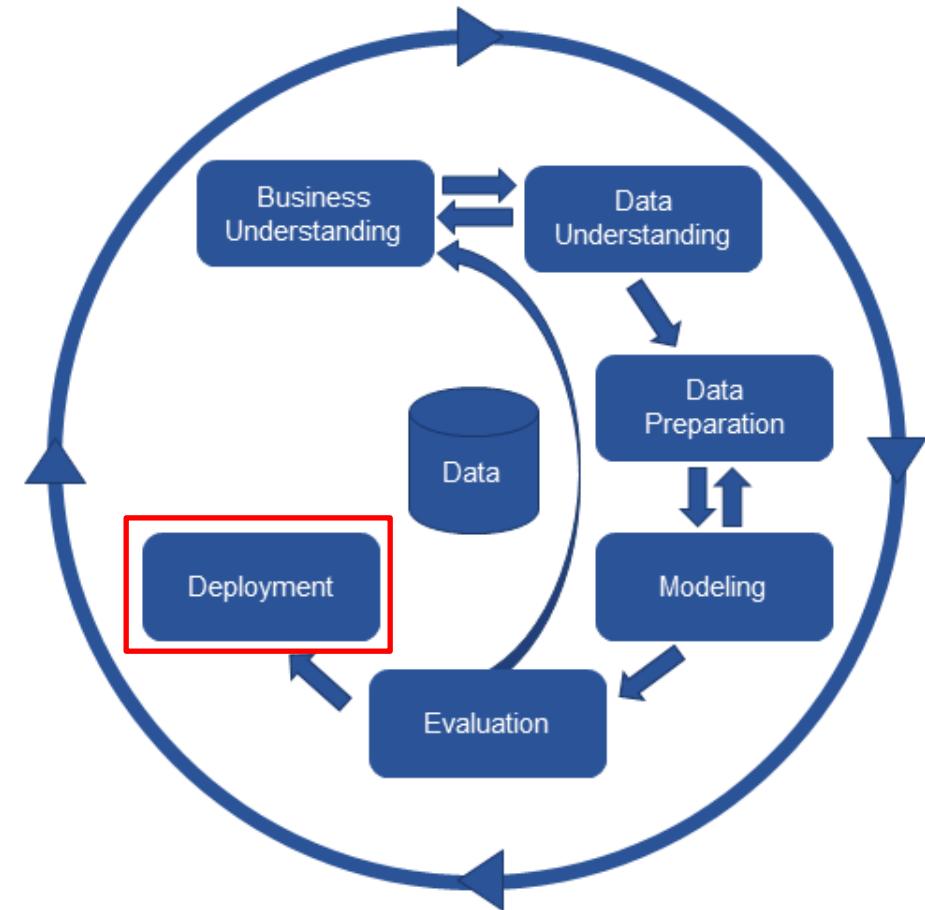
- Evaluate results
 - Business success criteria fulfilled?
- Review process
- Determine next steps
 - Feature engineering, optimize model hyper parameters
- Assess model



CRISP-DM

Deployment

- Plan deployment
- Plan monitoring and maintenance
 - Data Ops
 - Retraining models
- Produce final products
 - Reports, data outputs, integrations, etc
- Review project
 - Collect lessons learned
 - Share results



DATA GOVERNANCE

CHAPTER TOPICS

Data Governance

- **Introduction**
- What is Data Governance
- Data Management Framework
- CDO Role
- Right Size Data Governance

INTRODUCTION

Myths about Data Governance

- Data governance is a published repository of common definitions.
- Data governance is a concern of – and hence managed by – IT.
- Data governance is just data quality (DQ) and master data management (MDM).
- Data governance is siloed by business function.
- Data governance provides no value or participation for the data consuming community.

CHAPTER TOPICS

Data Governance

- Introduction
- **What is Data Governance**
- Data Management Framework
- CDO Role
- Right Size Data Governance

WHAT IS DATA GOVERNANCE

“The exercise of authority, control and shared decision-making (planning, monitoring and enforcement) over the management of data assets. Data Governance is high-level planning and control over data management”

WHAT IS DATA GOVERNANCE

Objectives of Data Governance are:

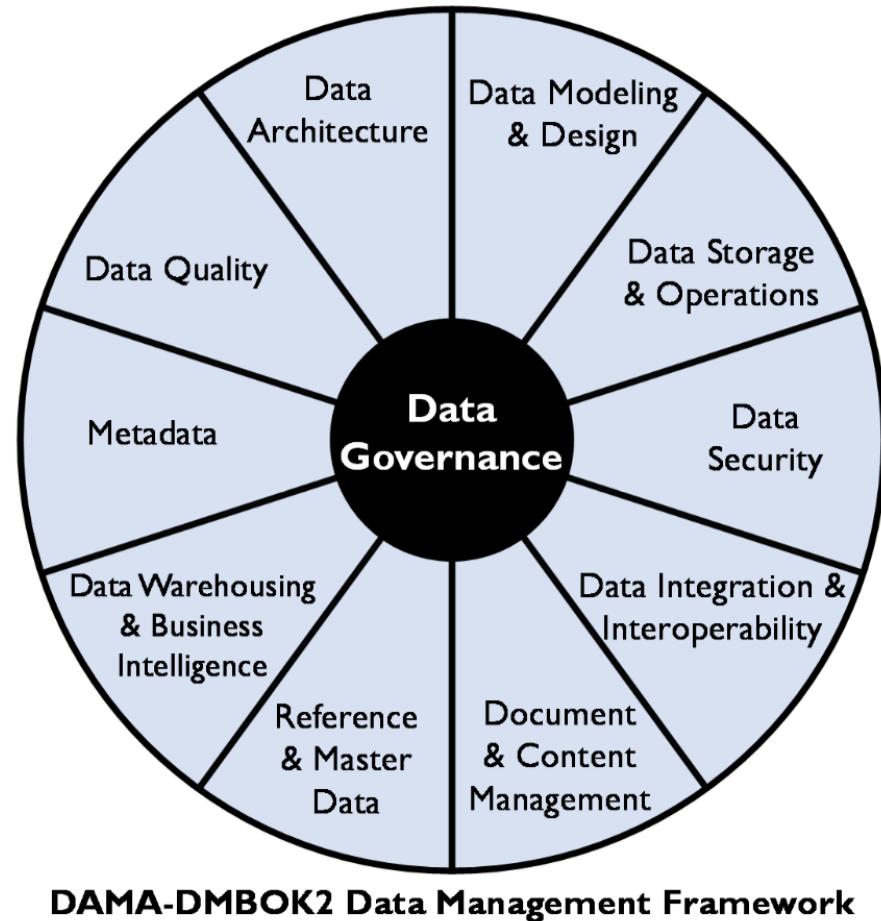
- Enable better decision-making
- Reduce operational friction
- Protect the needs of data stakeholders
- Train management and staff to adopt common approaches to data issues
- Build standard, repeatable processes
- Reduce costs and increase effectiveness through coordination of efforts
- Ensure transparency of processes

CHAPTER TOPICS

Data Governance

- Introduction
- What is Data Governance
- **Data Management Framework**
- CDO Role
- Right Size Data Governance

DATA MANAGEMENT FRAMEWORK



DATA MANAGEMENT FRAMEWORK

- Holistic approach to understand the information needs of the enterprise & its stakeholders
- Consistency for planning & process development
- 11 major functional areas, including governance
- Aligns data with business strategy (above) and technology (below)
- Takes into account the data lifecycle – creation through destruction
- Internationally recognized through Data Management Association International (DAMA)

DATA MANAGEMENT FRAMEWORK

Supporting components for Big Data

- **Data Governance** – Exercise of authority and controls over the management of data assets. Policies, processes, standards, definitions, metrics.
- **Data Architecture** - Defines data requirements, guides integration and control of data assets, aligns data investments with business strategy.
- **Master Data Management** – Control over master data values to enable consistent, contextual use across systems of the most accurate, timely and relevant version of truth about essential business entities.
- **Meta Data Management** – Descriptive tags about data, concepts, and connections between data and concepts.
- **Data Security** – Planning, development, and execution of security policies and procedures to provide proper authentication, authorization, access, and auditing of data and information assets.

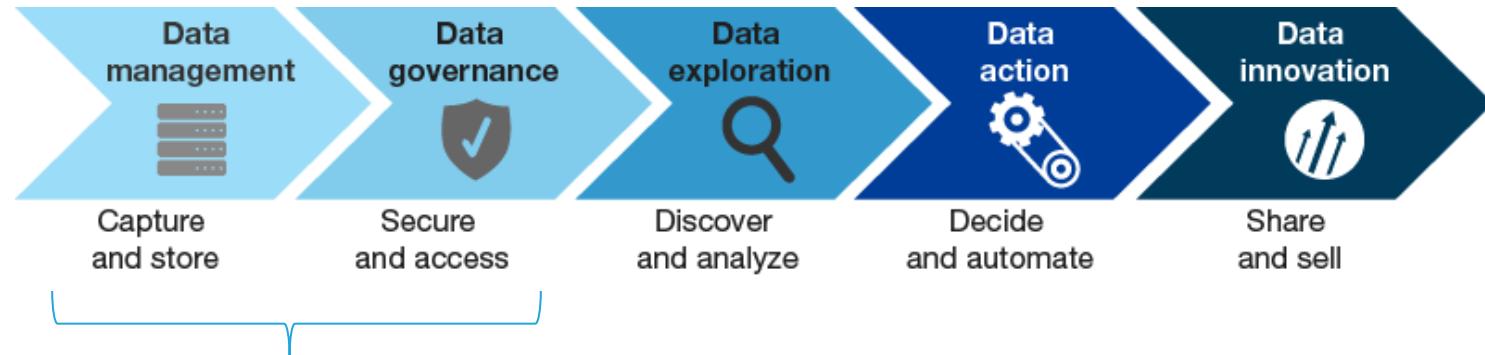
CHAPTER TOPICS

Data Governance

- Introduction
- What is Data Governance
- Data Management Framework
- **CDO Role**
- Right Size Data Governance

CDO ROLE

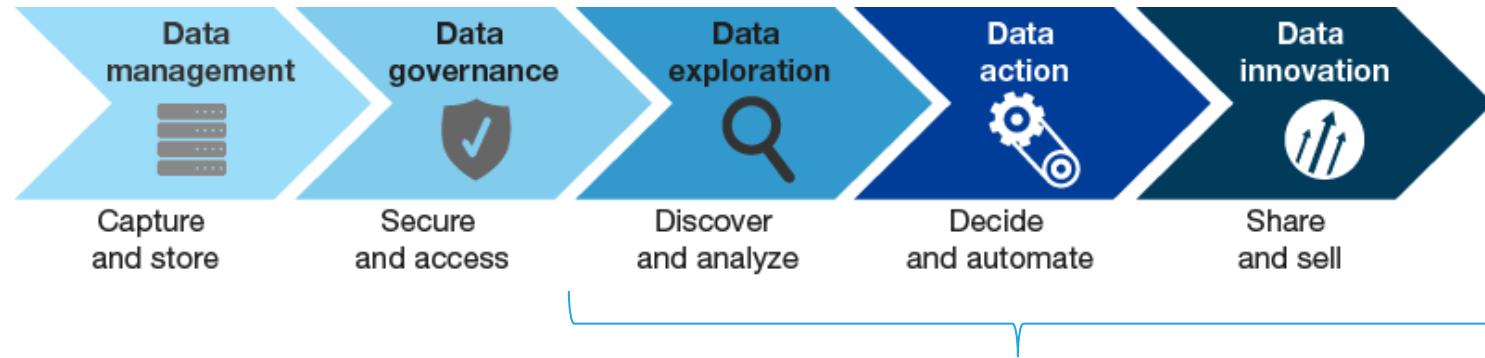
Changing the focus



Early CDOs focused on
first stages of
analytics pipeline

CDO ROLE

Changing the focus



Current CDOs focus on the analytics maturity, and consequently on the business needs

CDO ROLE

Scope

Functional Scope

- Governance
- Strategy
- Administration
- Quality & Stewardship
- Technology Partner

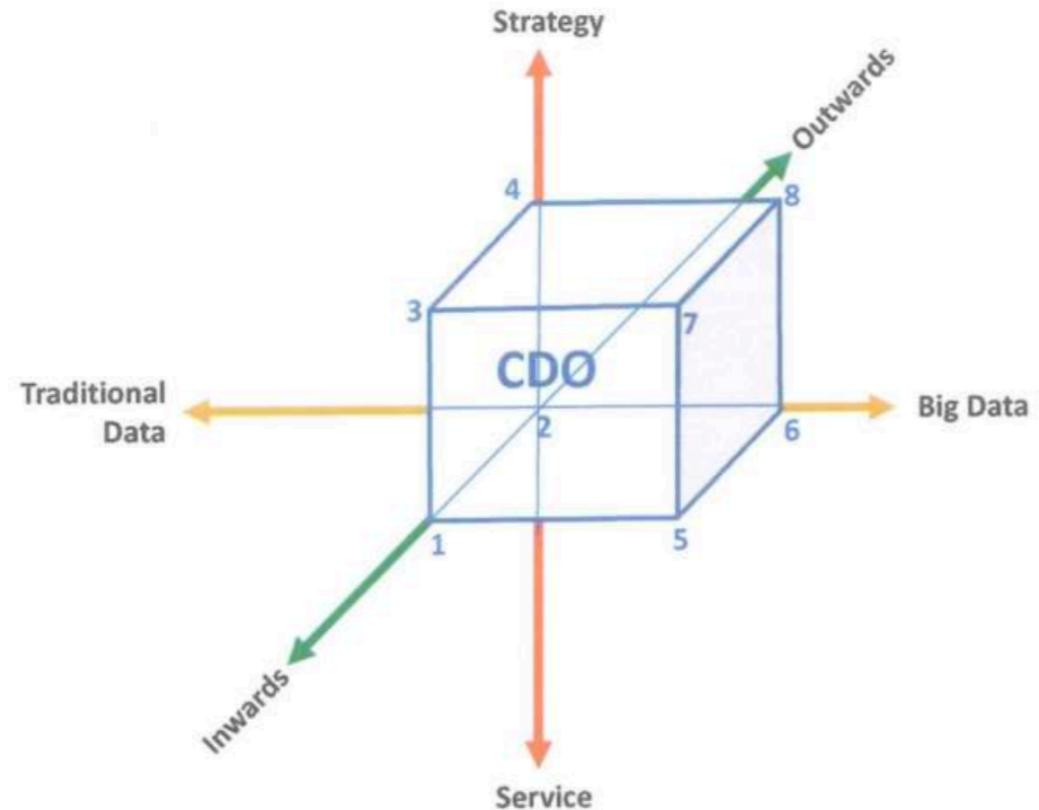
Data Scope

- Instrument
- Client & Account
- Assets & Liabilities
- Legal Entity/ Counterparty
- Pricing
- Corporate Actions
- Org Structure
- Unstructured Data
- Analytics

CDO ROLE

6 Dimensions

- **Collaboration:** inwards / outwards
- **Data Space:** traditional data / big data
- **Value Impact:** service / strategy



CHAPTER TOPICS

Data Governance

- Introduction
- What is Data Governance
- Data Management Framework
- CDO Role
- **Right Size Data Governance**

RIGHT SIZE DATA GOVERNANCE

Data Stewards

Owners and/or creators of the data

Responsibilities

- Providing knowledge about the data (e.g. privacy, use case concerns)
- Documenting and improving the raw data, with focus on link-ability

Data Engineers

Implement the data governance policies

Responsibilities

- Defining and driving the governance
- Organizing and hosting the Governance Council
- Delivering and utilizing tools (e.g. Navigator) to enforce governance

Data Governance Council

Business owners of the Data Governance

Responsibilities

- Communication about and enforcement of data governance
- Assigning data steward roles
- Improving the link-ability of data

HOW TO START

CHAPTER TOPICS

How to Start

- Introduction
- Design Thinking for Analytics

INTRODUCTION

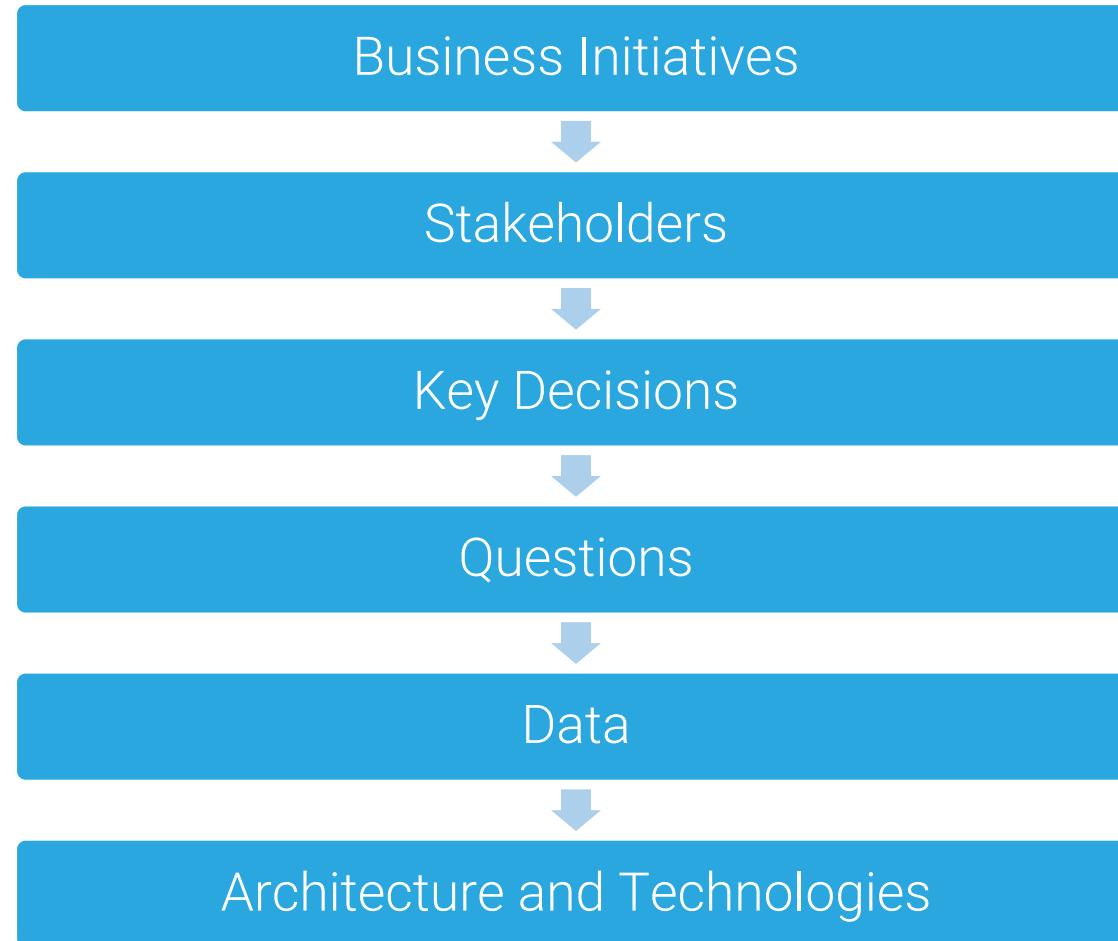
Where and how to start?

is such and open ended question!

..but it is a clear call to start think about use cases.

INTRODUCTION

Start with the end in mind



METHODOLOGY

Design Thinking and Analytics share many common values, including the power of “might”, learning through failure, and the disrupting that innovation brings in place by the hand of “diverge to converge” thinking.

Coincidentally, companies are investing in **data** (analytics capabilities) and **design** (generation capabilities). Their combination is really powerful, unlocking new ways to drive new sources of value and actionable outcomes.

The proposed methodology will drive companies to creatively design analytics solutions aligned with business initiatives.

METHODOLOGY

Empathize Phase

Objective: understand current business challenges, how they perform today analytics, what are the areas of opportunities, competitors, marketing situation, etc.

Activity: brainstorm to get a big picture of the company/industry you chose.

Output:

- List three main challenges of your company
- Generate a SWOT (FODA) analysis (2 to 3 items per area)

METHODOLOGY

Define Phase

Objective: after understand company/industry situation, propose/define drivers to improve business performance, operational efficiency or new business models. This is called by Business Initiatives, and will drive the exercise.

Activity: brainstorm to define Business Initiatives.

Output:

- Statement of 2 to 3 Business Initiatives that can have real impact in the company/industry. Describe those initiatives with short information.

METHODOLOGY

Ideate Phase

Objective: once Business Initiative are defined, it is time to details and transform them into specific use cases.

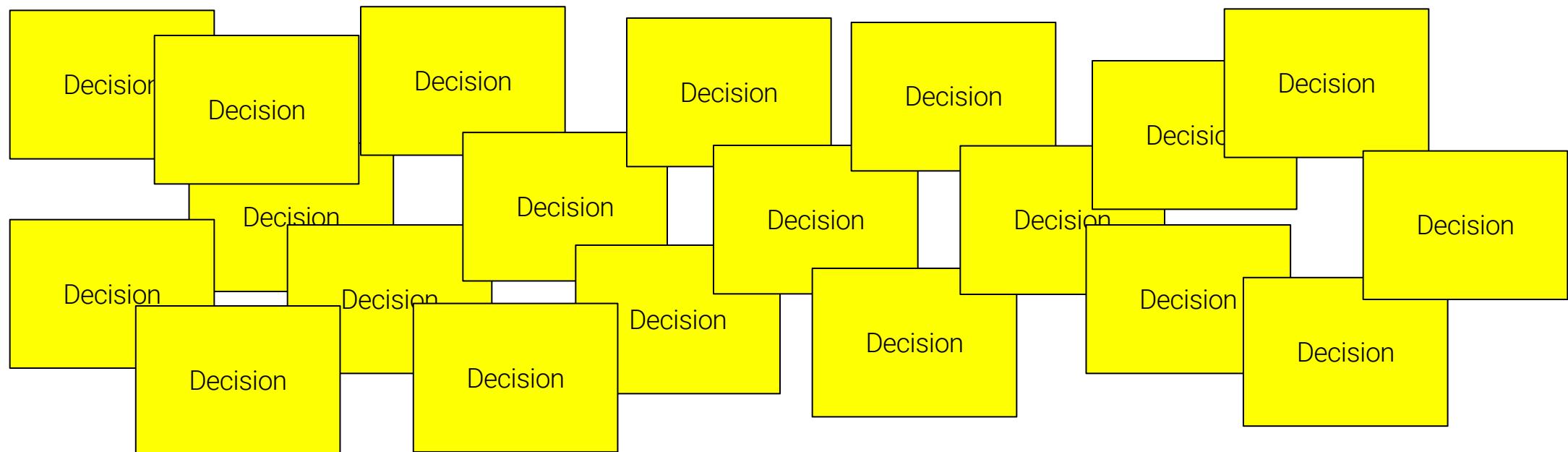
Activities: divided into two sessions:

Activity 1, Use Case co-creation: brainstorming to identify decisions that support Business Initiatives, grouping those decisions into common subjects or use cases.

METHODOLOGY

Ideate Phase – Brainstorming Decisions

Business Initiative

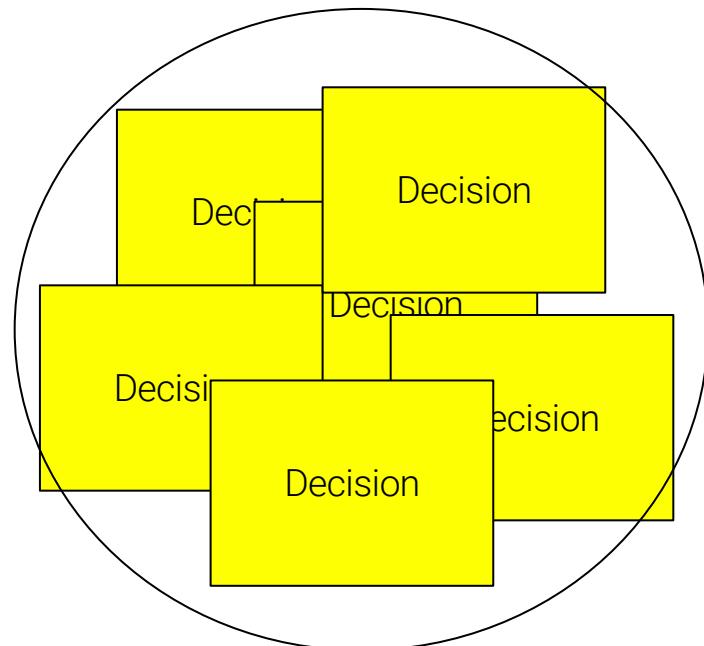


METHODOLOGY

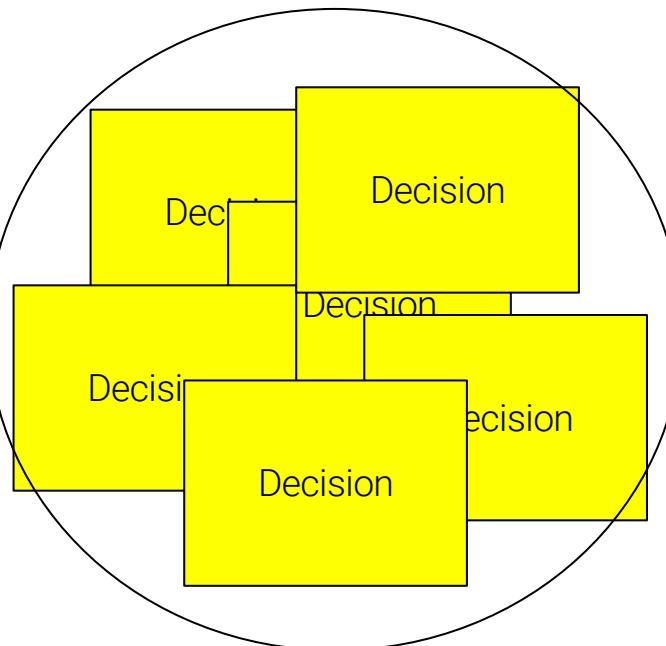
Ideate Phase – Group Decision into Use Cases

Business Initiative

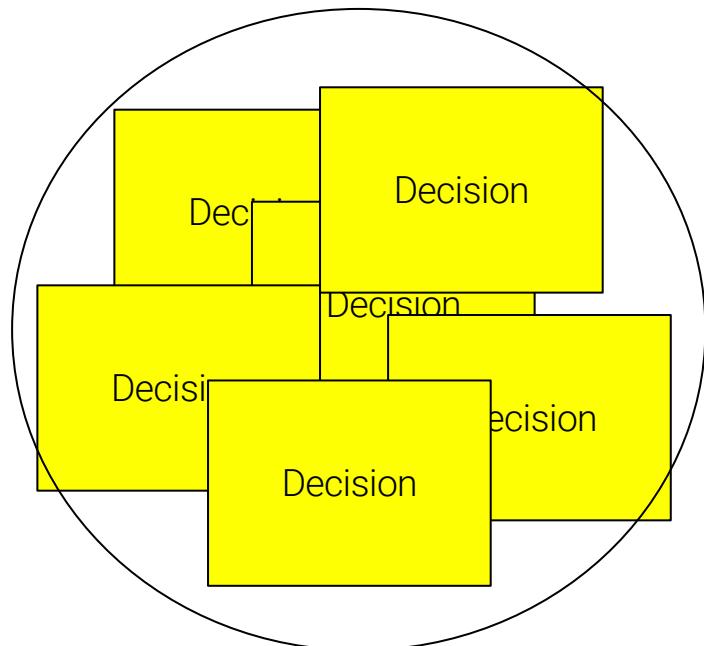
Use Case 1



Use Case 2



Use Case 3



METHODOLOGY

Ideate Phase

Activities: divided into two sessions:

Activity 2, Use Case Prioritization: once use cases are defined, prioritizing them based on implementation feasibility and business value. The main idea is to find quick-wins.

Output: complete list to use cases associated with Business Initiative, prioritized use cases, and also the intermediate artifacts used during the sessions.

METHODOLOGY

Ideate Phase – Data Value Assessment

Data Source	Increase Store Traffic	Increase Shopping Bag Revenue	Increase # Corporate Events	Increase Promotional Effectiveness	Improve NPI Effectiveness
Point of Sales Transactions	●	●	●	●	●
Market Baskets	●	●	○	●	●
Store Demographics (Zip Code)	○	○	○	○	○
Local Competitive Stores	○	○	○	○	○
Store Manager Demographics	○	○	●	○	○
Consumer Comments	○	○	○	○	○
Social Media	○	○	○	●	●
Key					
Worst...	○	●	○	○	○
Best	●	○	●	●	●
Zillow	○	○	○	○	○

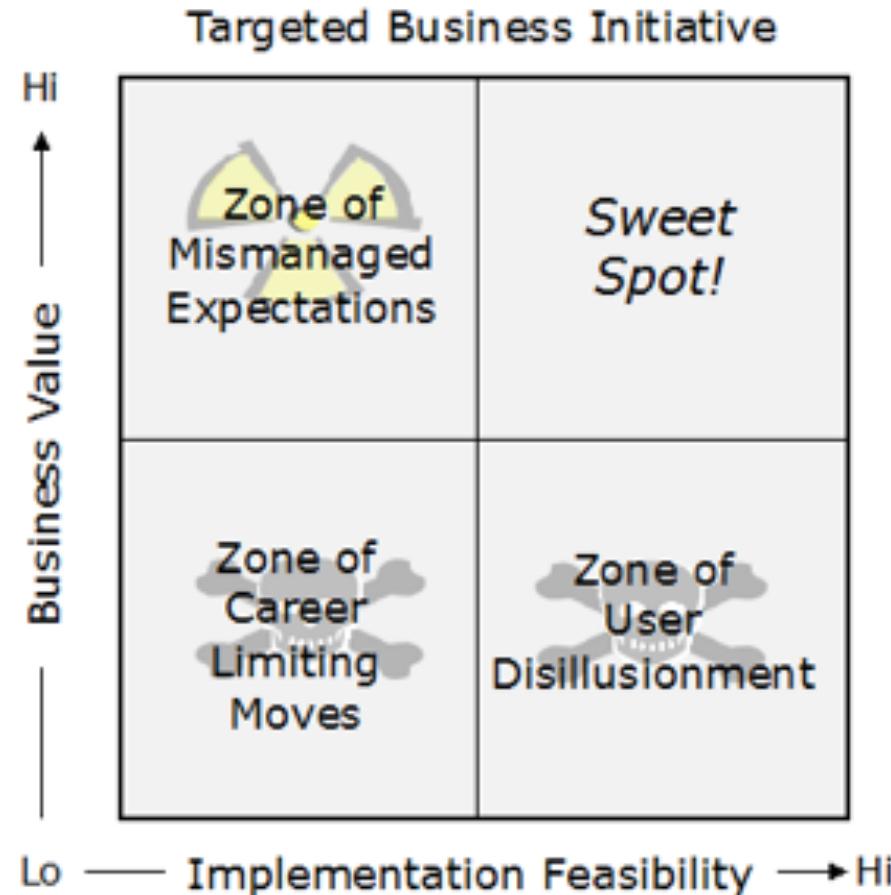
METHODOLOGY

Ideate Phase – Implementation Feasibility

Data Source	Ease of Acquiring	Cleanliness	Accuracy	Granularity	Cost
Point of Sales Transactions	●	●	●	●	●
Market Baskets	●	●	●	●	●
Store Demographics (Zip Code)	●	●	●	●	●
Competitive Stores Sales	○	○	○	○	○
Store Manager Demographics	●	●	●	●	●
Consumer Comments	○	○	○	○	●
Social Media	○	○	○	○	○
Key	●	●	●	●	●
Worst... Best	○	○	○	○	○
Zillow	○	○	●	○	○

METHODOLOGY

Ideate Phase – Prioritization Matrix



METHODOLOGY

Prototype Phase

Objective: with the quick-wins defined, it is time to idealize possible solutions to implement the use cases. Prototyping is the best ways to create solutions and verify impact. It can a app screen, a data pipeline or an architecture draft, prototypes will open the discussion around how the final solution looks like.

Activity: develop a prototype with the necessary detail to present them to audience.

Output: Prototypes and related documentation.

METHODOLOGY

Test Phase

Objective: verify, stress and defend prototypes. The main idea is to open the discussion, finding ways to improve prototype and increment success factor.

Activity: define future vision for use cases and possible spin off of the use cases.

Output: improved prototypes and documentation.

HOW TO START

Wrap Up



- Identify decisions that address the business initiative
 - Determine pains
 - Find opportunities to leverage data
 - Industry use cases
- Find the value of the current data
 - Discovery new data that should be generated
 - Data usage for the use cases
- Assess technical data aspects
 - Understand challenges and risks associated with each data source
- Identify use cases that can deliver fast business value with less effort to implement
 - Visualize short and medium term use cases

THANK YOU