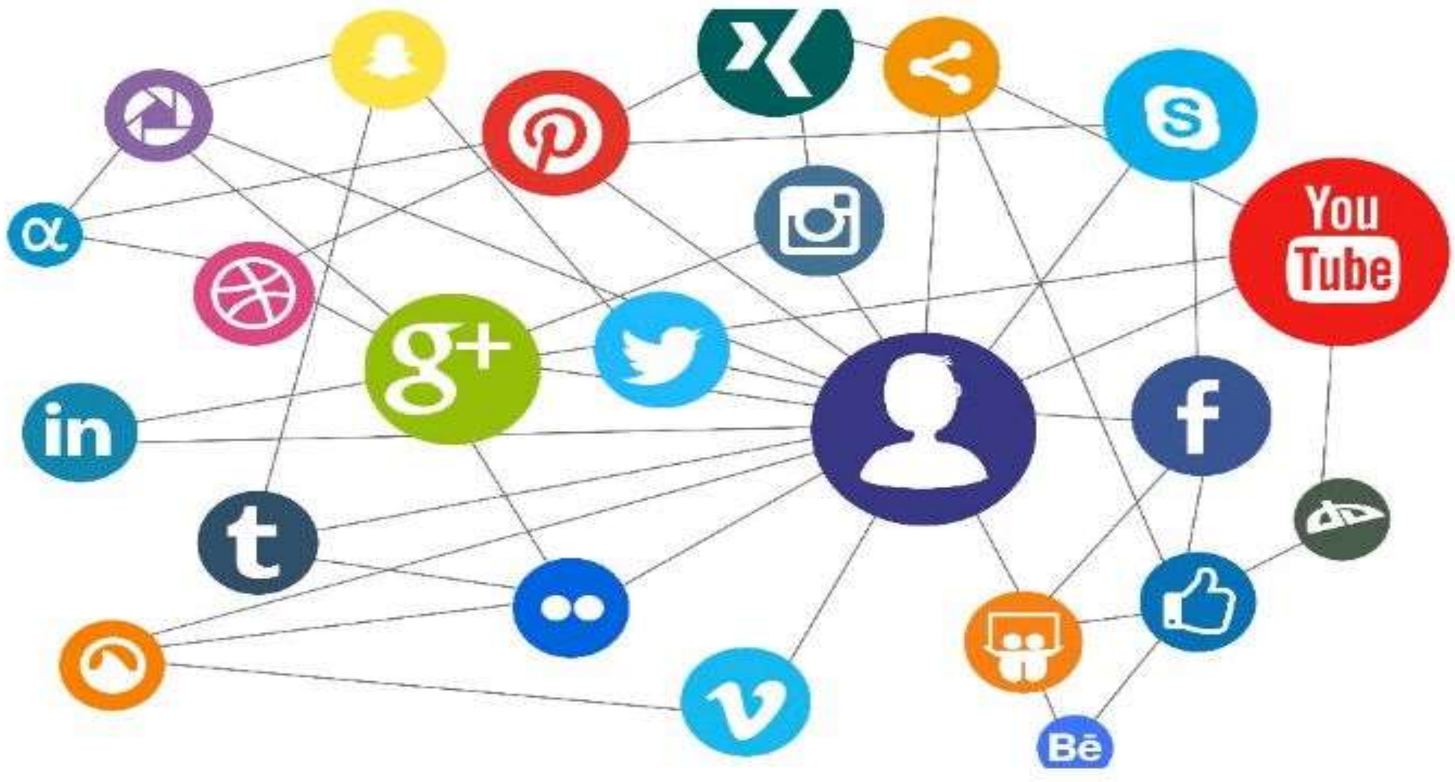
Unit 6

Big Data Technologies

Applications and Impact

Social Media Analytics

- Social Media is a platform that allows knowledge to be rapidly created and shared through social networks. It is changing the way people communicate and has thus become an important tool for business..



Social Media Analytics is concerned with developing and evaluating informatics tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data, usually driven by specific requirements from a target application.

Social media analytics research serves several purposes:

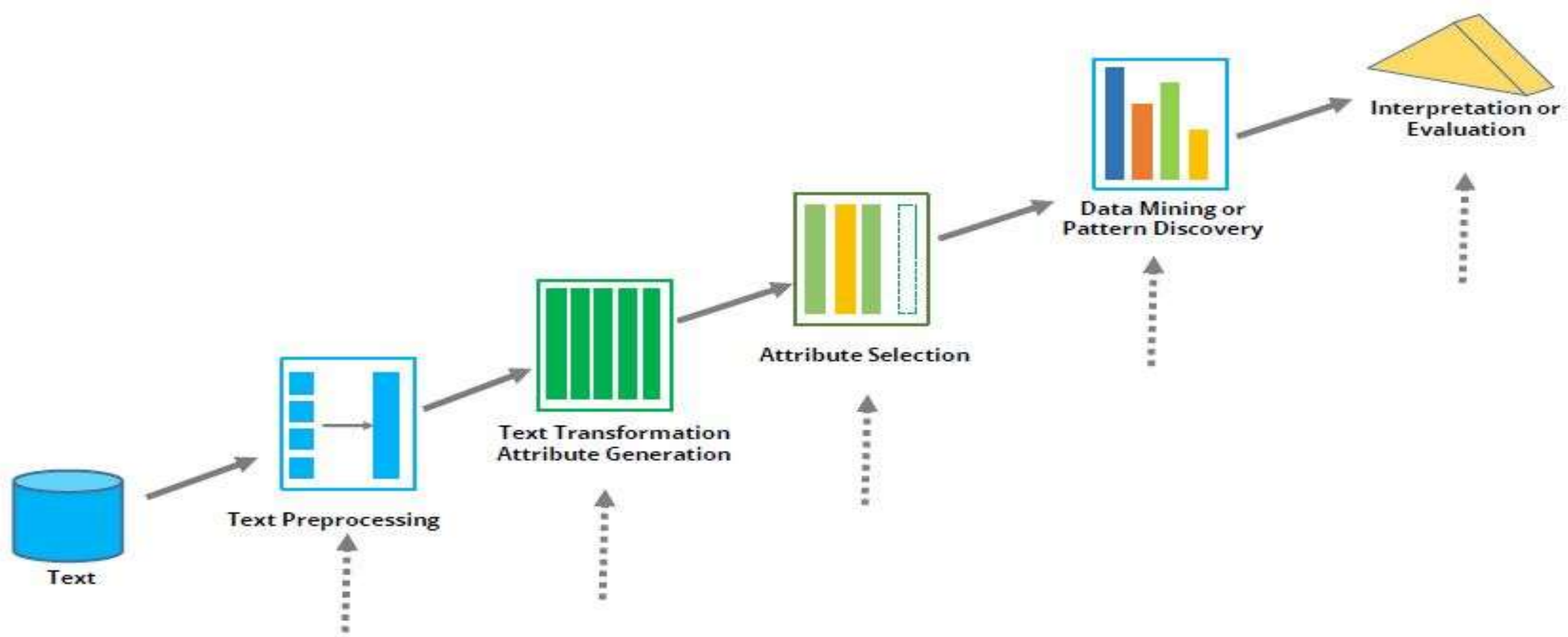
- ✓ facilitating conversations and interaction between online communities and
- ✓ extracting useful patterns and intelligence to serve entities that include, but are not limited to, active contributors in ongoing dialogues

Some of the key elements of social media are:

- Collect
- Curate
- Create
- Share
- Engage

Text Mining

- **Text Mining** is also known as text analysis, is the process of transforming unstructured text data into meaningful and actionable information.
- Text mining utilizes different AI technologies to automatically process data and generate valuable insights, enabling companies to make data-driven decisions.



Significance

DOCUMENT CLUSTERING

Clustering makes it easy to group similar documents into meaningful groups.
News sections are often grouped as business, sports, politics

PRODUCT INSIGHTS

Mining consumer reviews can reveal insights like most loved feature, most hated feature, improvements required, reviews of competitors' products

PATTERN IDENTIFICATION

Features such as telephone numbers, e-mail addresses can be extracted using pattern matches

SECURITY MONITORING

Text mining helps in monitoring and extracting information from news articles and reports for national security purposes

Applications

SPEECH RECOGNITION

Recognition and translation of spoken language into text and vice - versa

SPAM FILTERING

Automatic detection of spam mail

SENTIMENT ANALYSIS

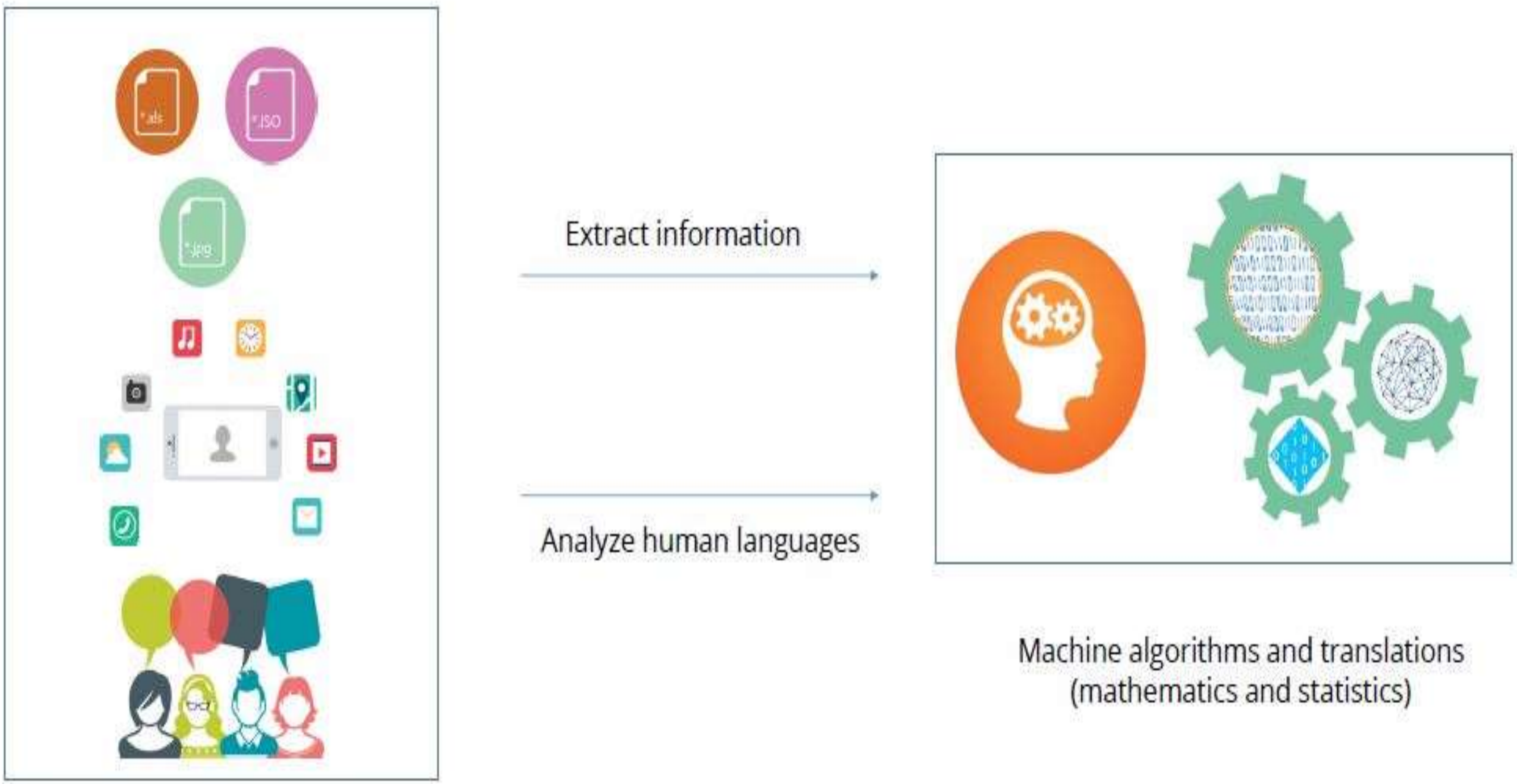
Determining if a given sentence expresses positive, neutral, or negative sentiment

E-COMMERCE PERSONALIZATION

Suggest products that fit into a user's profile

Natural Language Processing (NLP)

Natural language processing is an automated way to understand and analyze natural human languages and extract information from such data by applying machine algorithms.



Data from various sources

Text Mining Software

Commonly used Text Mining Software are:

- Active Point
- Attensity
- Crossminder
- Compare Suite
- IBM SPSS Predictive Analytics Suite
- Monarch
- SAS Text Minen
- Textalyzer

Text Mining Process

- **Text transformation**
  A text transformation is a technique that is used to control the capitalization of the text.
  Here the two major way of document representation is given.
  - Bag of words
  - Vector Space
- **Text Pre-processing**
  Pre-processing is a significant task and a critical step in Text Mining, Natural Language Processing (NLP), and information retrieval(IR). In the field of text mining, data pre-processing is used for extracting useful information and knowledge from unstructured text data. Information Retrieval (IR) is a matter of choosing which documents in a collection should be retrieved to fulfill the user's need.
- **Feature selection:**
  Feature selection is a significant part of data mining. Feature selection can be defined as the process of reducing the input of processing or finding the essential information sources. The feature selection is also called variable selection.

- **Data Mining:**
  Now, in this step, the text mining procedure merges with the conventional process. Classic Data Mining procedures are used in the structural database.
- **Evaluate:**
  Afterward, it evaluates the results. Once the result is evaluated, the result abandon.

Mobile Analytics

Mobile analytics is the practice of collecting user behavior data, determining intent from those metrics and taking action to drive retention, engagement, and conversion.

Mobile analytics involves measuring and analyzing data generated by mobile platforms and properties, such as mobile sites and mobile applications

The Primary goal of Mobile Analytics is to understand the following:

- ✓New Users
- ✓Active Users
- ✓Percentage of new Users
- ✓Sessions
- ✓Average
- ✓Average Usage Duration
- ✓accumulated users
- ✓Bounce rate
- ✓Users Retention

3 Major Types of Mobile Analytics

1.) Advertising/Marketing Analytics:

With a marketing analytics solution, you are able to attribute which ad a user clicked that led to the install. With data gathered from multiple attribution methodologies, you can recognize which ad networks and publishers are driving the right types of users, and use this knowledge to optimize ad campaigns and drive the highest possible return on investment (ROI) and lifetime value (LTV) of acquired users.

2.)In-App Analytics:

In-app analytics is essentially "in-session" analytics – what users are actually doing inside the app and how they are interacting with the app.

3.)Performance Analytics:

Performance analytics is generally concerned with two major measures:

a.) App uptime

b.) App responsiveness

Regardless of how well your app is coded, there are a number of factors that can impact the performance of your app.

Roles and Responsibilities of Big Data Person

Role: Chief Data Officer

The chief data officer will be responsible for acquiring, storing, enriching, and leveraging the company's data assets. This role is likely to be filled by people with an economics or finance background as they look at ways to put economic value on the data that they have and want to acquire.

Responsibility:

- **Data inventory:** Many organizations don't even know what data sources they have, so this role would be responsible for inventorying data (looking for unnecessary and redundant data purchases) and determining how that data is being used (to determine if the organization should continue to capture the data). This role would also have the critical responsibility for identifying and placing value on external data sources that could be acquired.
- **Data economic valuation:** Establish a framework around which to determine the economic value of the organization's data, especially as companies look to acquire more external, partner, and third-party data.
- **Data monetization:** Establish a process to continuously evaluate the organization's data assets for monetization opportunities through improved decision-making, integrating data into physical products, or packaging data for sale to other organizations.

- **Instrumentation:** Develop strategies to determine how to use tags, beacons, and sensors across operational, web, and mobile platforms to capture additional customer, product, and operational data.
- **Data governance:** Develop and enforce (audit) a set of processes that ensures that important data assets are formally and consistently managed across the enterprise to ensure the appropriate level of data cleanliness and accuracy.

Role: Chief Analytics Officer

The chief analytics officer will be responsible for capturing and tracking the analytic models and resulting analytic insights that are developed and deployed throughout the organization. The ideal chief analytics officer probably has a law degree to legally protect the organization's analytical intellectual property (IP) including—data models, analytic models, and analytic algorithms.

Responsibility:

- **Analytic assets:** Collaborate with the data science team to inventory analytic models and algorithms throughout the organization.
- **Analytics valuation**: Establish a framework and process for determining the financial value of the organization's analytic assets.
- **Intellectual property management**: Develop processes and manage a repository for the capture and sharing of organizational IP (check-in, check-out, versioning).
- **Patent application**s: Manage the patent application and tracking process for submitting patents to protect key organizational analytics IP.

- **Intellectual property protection:** Monitor industry analytics usage to identify potential IP violations, and then lead litigation efforts to stop or get licensing agreements for IP violations.
- **Intellectual property monetization**: Actively look for business partners and opportunities to sell or license organizational analytics IP.

Role: Big Data Engineer

Responsibility:

- Gather and process raw data at scale (including writing scripts, web scraping, calling APIs, write SQL queries, etc.).
- Process unstructured data into a form suitable for analysis – and then do the analysis.
- Work closely with engineering team to integrate your amazing innovations and algorithms into production systems.
- Support business decisions with ad hoc analysis as needed.

Roles and Responsibilities of Big Data Person

Role: **Data Analyst**

Responsibility:

- Coordinate with customers and staff and provide support to all data analysis.
- Perform data analysis on all results and prepare presentations for clients.
- Perform audit on data and resolve business-related issues for the customer base.
- Coordinate with engineering and product management team and ensure accuracy on all deliverables and prepare summaries.
- Perform data analysis and facilitate in delivery to all end use
- Supervise all client issues and coordinate with managers and supervisors and facilitate in deliverables.
- Organize all consumption anomalies and determine defects for data and prepare appropriate resolutions.
- Perform internal audit and prepare all invoices and determine quality improvement processes.

Roles and Responsibilities of Big Data Person

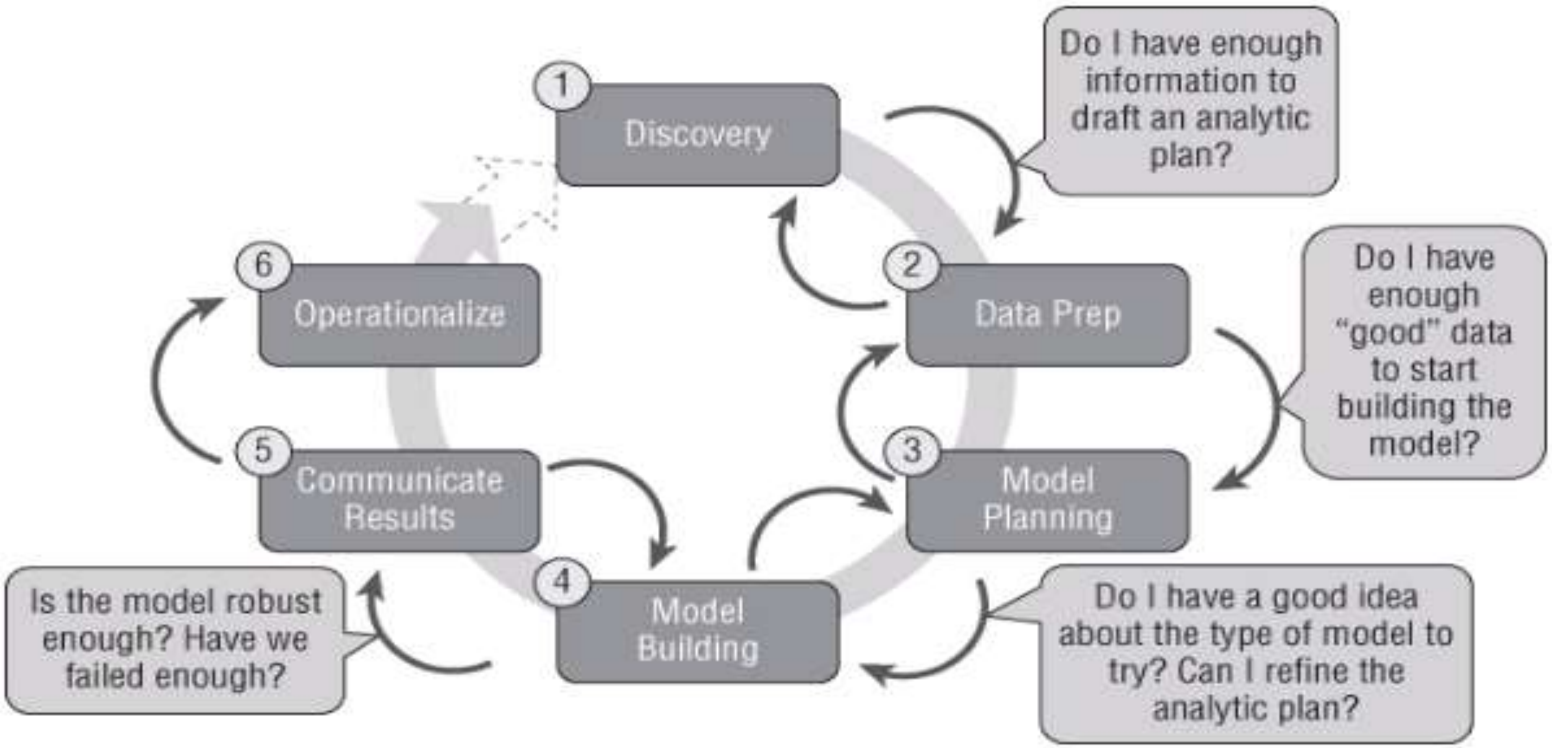
Role: Big Data Solutions Architect

Responsibility:

- Guide the full lifecycle of a Hadoop solution, including requirements analysis, platform selection, technical architecture design, application design and development, testing, and deployment.
- Provide technical and managerial leadership in a team that designs and develops path-breaking large-scale cluster data processing systems..
- Help extreme insights customers develop strategies that maximize the value of their data.

Data Scientist Roles and Responsibility

Data Scientists collect data and explore, analyze, and visualize it. They apply mathematical and statistical models to find patterns and solutions in the data.



The Data Scientist Lifecycle

Discovery focuses on the following data scientist activities:

- Gaining a detailed understanding of the business process and the business domain. This includes identifying the key metrics and key performance indicators against which the business users will measure success.
- Capturing the most important business questions and business decisions that the business users are trying to answer in support of the targeted business process. This also should include the frequency and optimal timeliness of those answers and decisions.
- Assessing available resources (for example, people skills, data management and analytic tools, and data sources) and going through the process of framing the business problem as an analytic hypothesis. This is also the stage where the data scientist builds the initial analytics development plan that will be used to guide and document the resulting analytic models and insights.

Data preparation focuses on the following data scientist activities:

- Provisioning an analytic workspace, or an analytic sandbox, where the data scientist can work free of the constraints of a production data warehouse environment. Ideally, the analytic environment is set up such that the data scientist can self-provision as much data space and analytic horsepower as required and can adjust those requirements throughout the analysis process.
- Acquiring, cleansing, aligning, and analyzing the data. This includes using data visualization techniques and tools to gain an understanding of the data, identifying (and eliminating as necessary) outliers in the data and assessing gaps in the data to determine the overall data quality; ascertaining if the data is “good enough.

- ”Transforming and enriching the data. The data scientist will look to use analytic techniques, such as logarithmic and wavelet transformations, to address potential skewing in the data. The data scientist will also look to use data enrichment techniques to create new composite metrics such as frequency (how often?), recency (how recent?), and sequencing (in what order?). The data scientist will make use of standard tools like SQL and Java, as well as both commercial and open source extract, transform, load (ETL) tools to transform the data.

- At the end of this step, the data scientist needs to feel comfortable enough with the quality and richness of the data to advance to the next stage of the analytics development process.

Model planning focuses on the following activities:

- Determining the different analytic models, methods, techniques and workflows to explore as part of the analytic model development. The data scientist might already believe that they know which analytic models and techniques are most appropriate, but it is always a good idea to have a plan to test at least one other to ensure that the opportunity to build a more predictive model is not missed.

- Determine correlation and collinearity between variables in order to select key variables to be used in the model development. As much as possible, the data scientist wants to quantify the cause-and-effect variables. Practical judgment will have to be used by the data scientist, and this may even be a good opportunity to re-engage with the BI analyst and the business users to ensure that the variables being selected "make sense." Remember, correlation does not guarantee causation, so care must be taken in selecting variables that not only make sense, but are also variables that can be measured going forward.

Model building focuses on the following activities:

- Massaging the data sets for testing, training, and production. New transformation techniques may have to be tested to see if the quality, reliability, and predictive capabilities of the data can be improved.
- Assessing the viability and reliability of data to use in the predictive models. Judgment calls will have to be made on quality and reliability of the data—is the data "good enough" to be used in developing the analytic models. Again, different transformation techniques may have to be tested to see if the quality of the data can be improved.
- Finally, developing, testing, and refining the analytic models. Testing is conducted to see which variables and analytic models deliver the highest quality, most predictive and actionable analytic insights.

The **communicate results** step is where the data scientist focuses on the following activities:

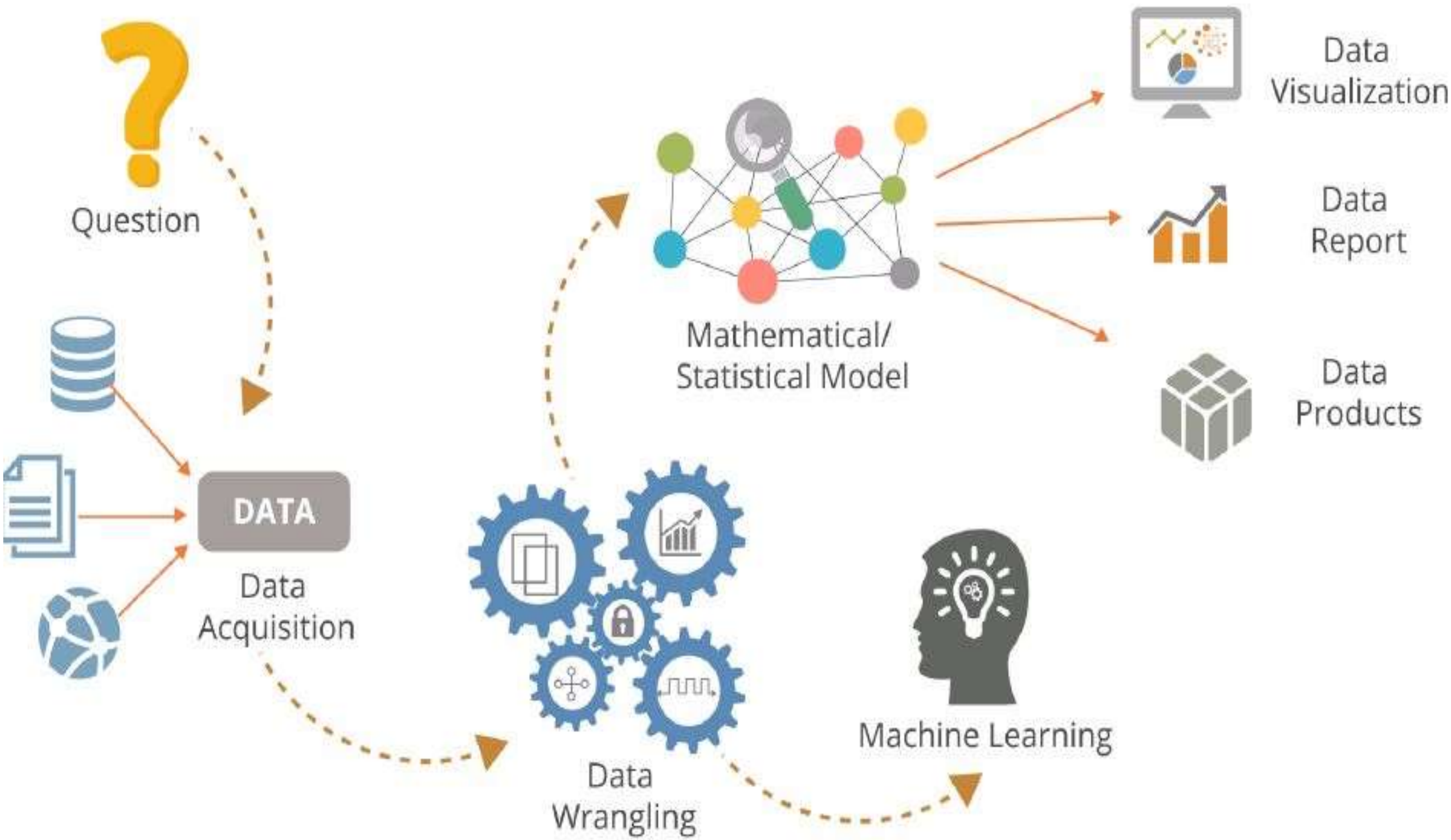
- Ascertaining the quality and reliability of the analytic model and statistical significance, measurability, and actionability of the resulting analytic insights. The data scientist needs to ensure that the analytic process and model was successful and achieved the desired analytic objectives of the project.
- Developing the charts and graphics to communicate the analytic model insights, results, and recommendations. It is critical that the business stakeholders—the business users, business analysts, and the BI analysts—understand and "buy into" the resulting analytic insights. If the business stakeholders do not have confidence in the results, then your work will have been for naught.

The **operationalize** step is where the data scientist focu

- Delivering the final recommendations, reports, briefings, code, and technical documents.ses on the following activities:
- Optionally, running a pilot or analytic lab to verify the business case, and the financial return on investment (ROI) and the analytic lift.
- Implementing the analytic models in the production and operational environments. This involves working with the application and production teams to determine how best to surface the analytic results and insights. The application and production teams can help determine how to "productionize" the analytic models so they run on a regular, scheduled basis, something that should have been covered in the analytics development plan.

- Integrating analytic scores into management dashboards and operational reporting systems, such as call centers, sales systems, procurement systems, and financial systems.
- The operationalization stage is another area where collaboration between the data scientist and the BI analysts should be invaluable. Many BI analysts already have experience integrating reports and dashboards into the operational systems, as well as establishing centers of excellence to propagate analytic learning and skills across the organization.

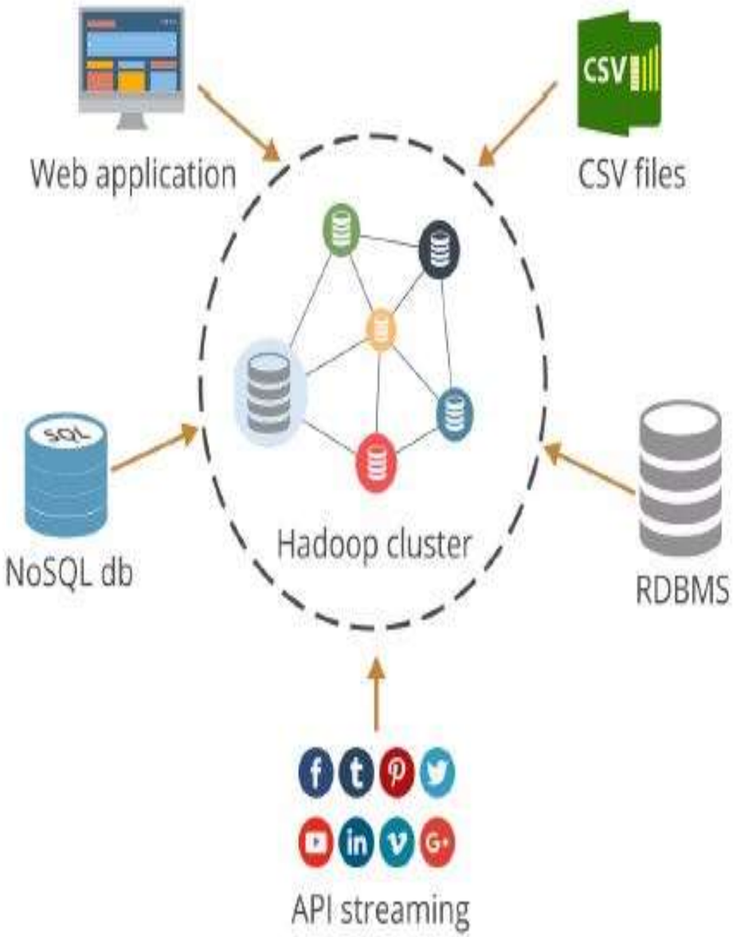
A Day in a Data Scientist's Life



The Real Challenge for Data Scientist’s

Some of the challenges Data Scientists face in the real world are listed here.

- Data quality doesn’t conform to the set standards.
- Data integration is a complex task.
- Data is distributed into large clusters in HDFS, which is difficult to integrate and analyze.
- Unstructured and semi-structured data are harder to analyze.



BI Analyst versus Data Scientist Responsibilities

Area	BI Analyst	Data Scientist
Focus	Reports, KPIs, trends	Patterns, correlations, models
Process	Static, comparative	Exploratory, experimentation, visual
Data sources	Pre-planned, added slowly	Chosen on the fly, on-demand
Transformation	Up front, carefully planned	ELT, on-demand, in-database, enrichment
Data quality	Single version of truth	Tolerant of "good enough"; probabilities
Data model	Logical / relational / formal	Conceptual / semantic / informal
Results	Report what happened	Predict what will happen
Analysis	Hindsight	Forecast, foresight Insight

Understanding Decision Theory,

- One interesting aspect of big data is how it is challenging the conventional thinking regarding how the non-analytical business user should be using analytics.
- Organizations are looking to use big data to become more analytics-driven in their business decision-making processes. However, there are several challenges that need to be addressed in order to make that transformation successful. One of those challenges is the very nature of how humans make decisions, and how our genetic makeup works against us in analyzing data and making decisions.
- The human brain is a poor decision-making tool. Human decision making capabilities have evolved over millions of years of survival on the savanna. Humans have become very good at pattern recognition: from “That looks like just a harmless log behind that patch of grass,” to “Yum, that looks like an antelope!” to “YIKES, that's a saber-toothed tiger!!” Necessity dictated that we become very good at recognizing patterns and making quick, instinctive survival decisions based on those patterns.
- Awareness of these human decision-making flaws is important if we want to transform our organization, and our people, into an analytics-driven business.

Traps in Decision Making

- ✓ Decision Trap #1: Overconfidence
- ✓ Decision Trap #2: Anchoring Bias
- ✓ Decision Trap #3: Risk Aversion
- ✓ Decision Trap #4: Don't Understand Sunk Costs
- ✓ Decision Trap #5: Framing

Other decision-making traps of which you need to be aware include:

- Herding (safety in numbers)
- Mental accounting
- Reluctance to admit mistakes (revisionist history)
- Confusing luck with skill
- Bias to the relative
- Overemphasizing the dramatic
- Regression to the mean
- Don't respect randomness

What Can One Do?

The key is to guide, not stifle, human intuition (think guardrails, not railroad tracks). Here are some things that you can do to guide your decision-making as you make the transformation to an analytics-driven organization:

- Use analytic models to help decision-makers understand and quantify the decision risks and returns. Leverage proven statistical tools and techniques to improve the understanding of probabilities. Employ a structured analytic discipline that captures and weighs both the risks and opportunities.
- Confirm and then reconfirm that you are using the appropriate metrics (think Moneyball). Just because a particular metric has always been the appropriate metric, don't assume that it is the right one for this particular decision.
- Consult a wide variety of opinions when you vet a model. Avoid Group Think, which is yet another decision-making flaw. Group Think is a trap where you surround yourself with people who think like you. Consequently, the group is already predisposed to like and agree with whatever ideas and decisions you make. Have someone play the contrarian (think Tom Hanks in the movie Big). Use facilitation techniques in the decision process to ensure that all voices are heard and all views are contemplated.
- Be careful how you frame decisions.

- Create business models that properly treat sunk costs. Ensure that the model and analysis only consider new incremental costs. Ensure that your models include opportunity costs.
- Use “after the decision” review boards and formal debriefs to capture what worked, what didn't, and why.
- Beware of counterintuitive compensation; humans are revenue optimization machines.

Making the transformation to an analytics-driven culture is a powerful business enabler, but more than technology needs to be considered in driving that transformation. Understanding, managing, and educating people on common decision-making traps will help ensure a successful transformation.

Creating Big Data Strategy

One of the key challenges IT organizations face in building support for a big data initiative is to ensure that the big data initiative is valued by, or of value to, the business stakeholders.

The document enforces a discipline that any organization can follow, as long as you truly understand and are focused on your organization's key business initiatives. The document is:

- Concise in that it fits onto a single page so that anyone can review it quickly to ensure they are working on the top priority items.
- Clear in defining what the organization and individuals need to do and accomplish in order to achieve the targeted strategic initiatives.
- Relevant to the business stakeholders by starting and focusing the process on supporting the organization's overall business strategy, and identifying the supporting business initiatives before diving into the technology, architecture, data, and analytic requirements.

- The Big Data Strategy Document is comprised of the following sections,

Business Strategy: The targeted business strategy is captured as the title of the document and clearly defines the scope upon which the big data initiative will be focused. The title should not be more than one sentence, but should still provide enough detail to clearly identify the overall business objective, for example: “Improve customer intimacy” or “Reduce operational maintenance costs” or “Improve new product launch effectiveness.”

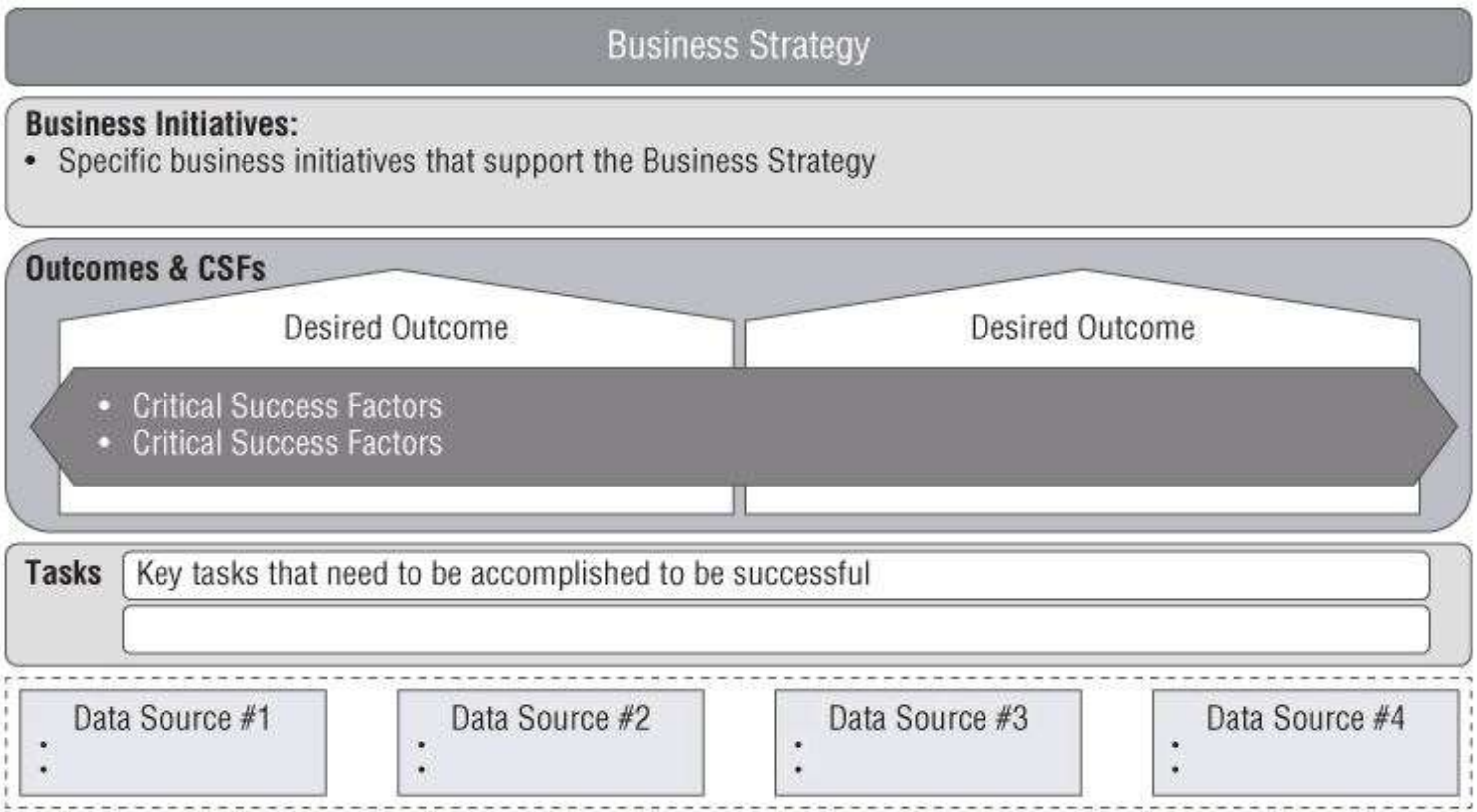
Business Initiatives: This section breaks down the business strategy into its supporting business initiatives. A business initiative is defined as a cross-functional project lasting 9 to 12 months in duration, with clearly stated financial or business goals against which success of the business initiative will be measured. Note that there should not be more than three to five business initiatives per business strategy. More than that and you have a wish list.

Outcomes and Critical Success Factors (CSF): This section captures the outcomes and critical success factors necessary to support the successful execution of the organization's key business initiatives. Outcomes define the desired or ideal end state. Critical success factors define “what needs to be done” for the business initiative to be successful.

Tasks: This section provides the next level of detail by documenting the specific tasks that need to be executed to perfection to be successful in support of the targeted business initiatives. These are the key tasks around which the different parts of the organization will need to collaborate to achieve the business initiatives. This is the “how to do it” section of the document, and it is at this level of detail where personal assignments and management objectives can be defined, assigned, and measured. One would normally expect 8 to 12 key tasks being identified and linked to the targeted business initiatives as part of the Big Data Strategy Document.

Data Sources: Finally, the document highlights the key data sources required to support the business strategy and the supporting key business initiatives. From the definition of the tasks, you should have a strong understanding of the key metrics and measures, important business dimensions, level of granularity, and frequency of data access.

The Big Data Strategy Document



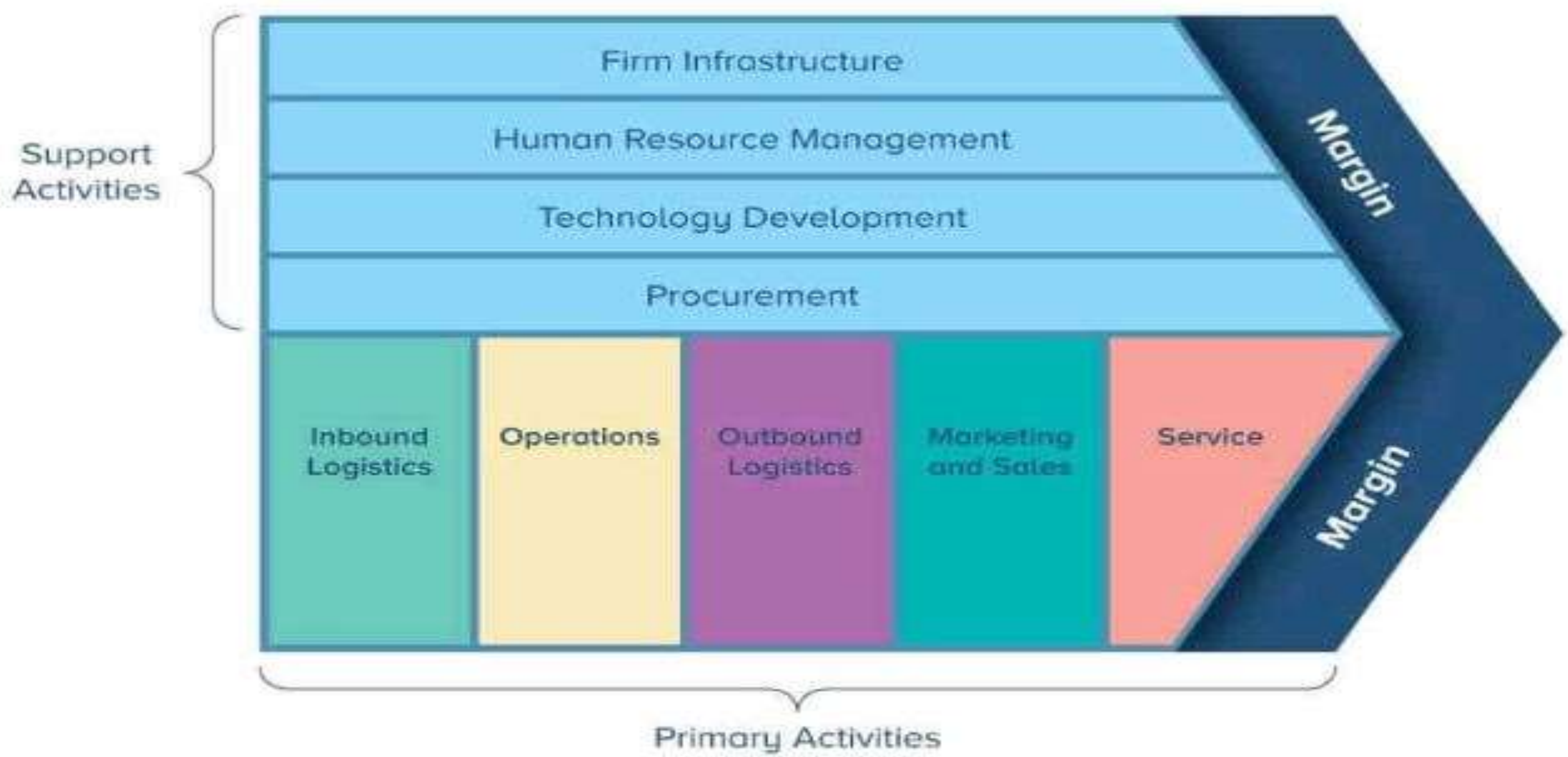
Big Data Value Creation Drivers

When analyzing the effectiveness of a value chain model, the economist Michael Porter introduced the following 10 cost drivers that help identify areas for improvement:

- **Economies of Scale:** A true picture of need includes cost analysis for the size of the demand, whether local, national, or global.
- **Learning:** Activities that change the environment for efficiency or improvement, such as scheduling, asset use, and office or warehouse layout.
- **Capacity Utilization:** Procedures that keep capacity at efficient levels to prevent under-utilization or the addition of unnecessary capacity.
- **Linkages among Activities:** Identifying areas of cross-functional improvement through coordination and optimization.
- **Interrelationships among Business Units:** Opportunities to share information and resources.
- **Degree of Vertical Integration:** Identifying areas of joint integration or, in some cases, de-integration.
- **Timing of Market Entry:** Driven by economic or world conditions and competitive position in the marketplace.
- **Firm's Policy of Cost or Differentiation:** Identified value integrated into the process.
- **Geographic Location:** This includes wages, climate, and raw materials.
- **Institutional Factors:** These include taxes, unions, and regulations.

Michael Porter's valuation creation models,

- **Michael Porter** explains Value Chain Analysis: that a value chain is a collection of activities that are performed by a company to create value for its customers.



Porter’s value chain involves five primary activities:

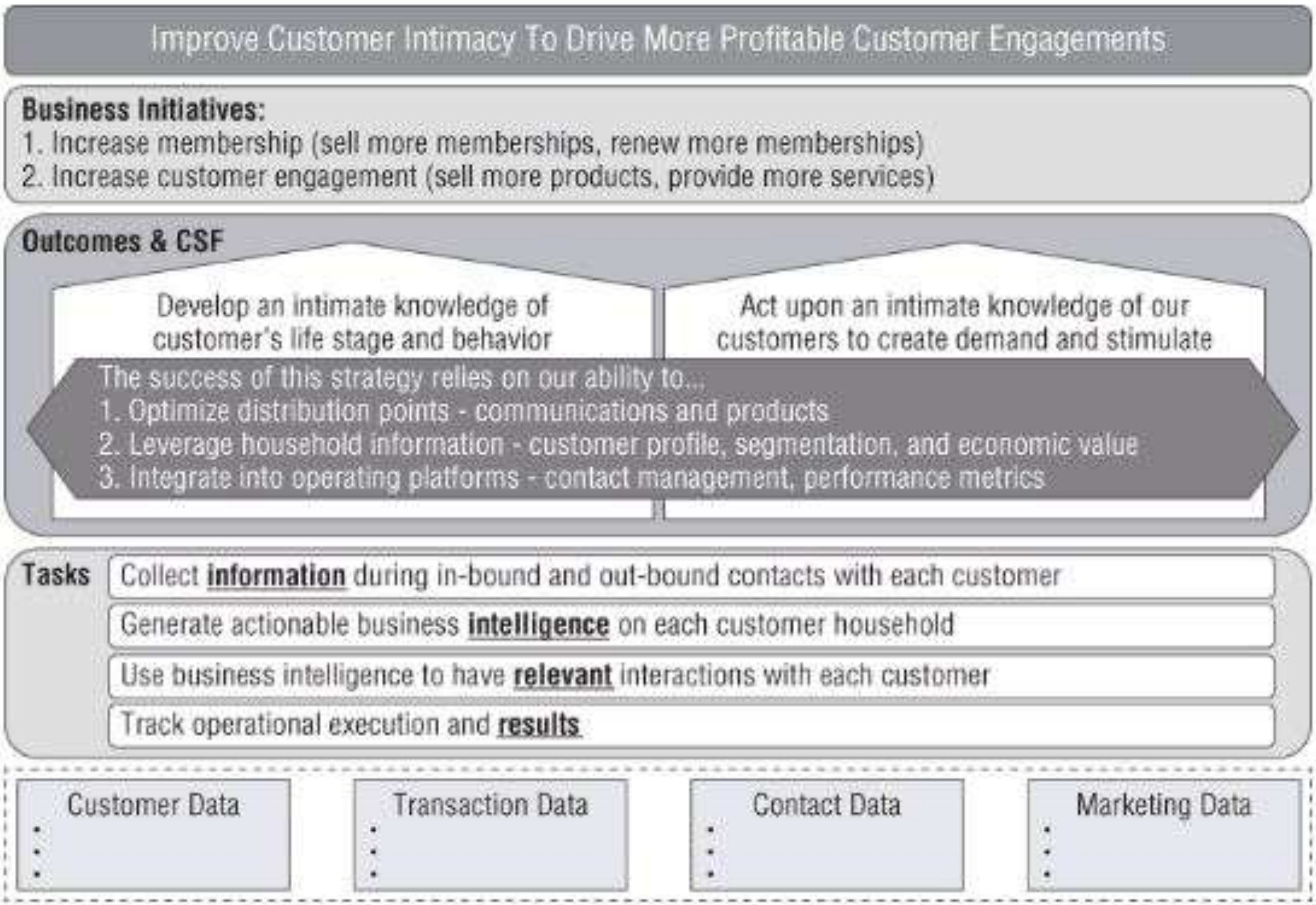
- inbound logistics,
- operations.
- outbound logistics.
- marketing and sales,
- and service.

Support activities are illustrated in a vertical column over all of the primary activities. These are procurement, human resources, technology development, and firm infrastructure.

The generic value chain model visually represents all activities with equal weight. However, value chain analysis emphasizes the real needs of the company. For example, a company that assists after the sale, such as for copiers or air conditioners, has a larger service activity set than a company that performs little follow-up action, such as FedEx or UPS.

When using Porter’s value chain, you must identify whether you are trying to differentiate or lower costs, prioritize the changes you identify during analysis, and consider how changes will benefit the entire organization.

Customer Intimacy Example



Benefits of Value Chain Modeling

Value chain modeling yields numerous benefits. In *Competitive Advantage: Creating and Sustaining Superior Performance*, Porter says, "Competitive advantage frequently comes from perceiving new ways to configure and manage the entire value system."

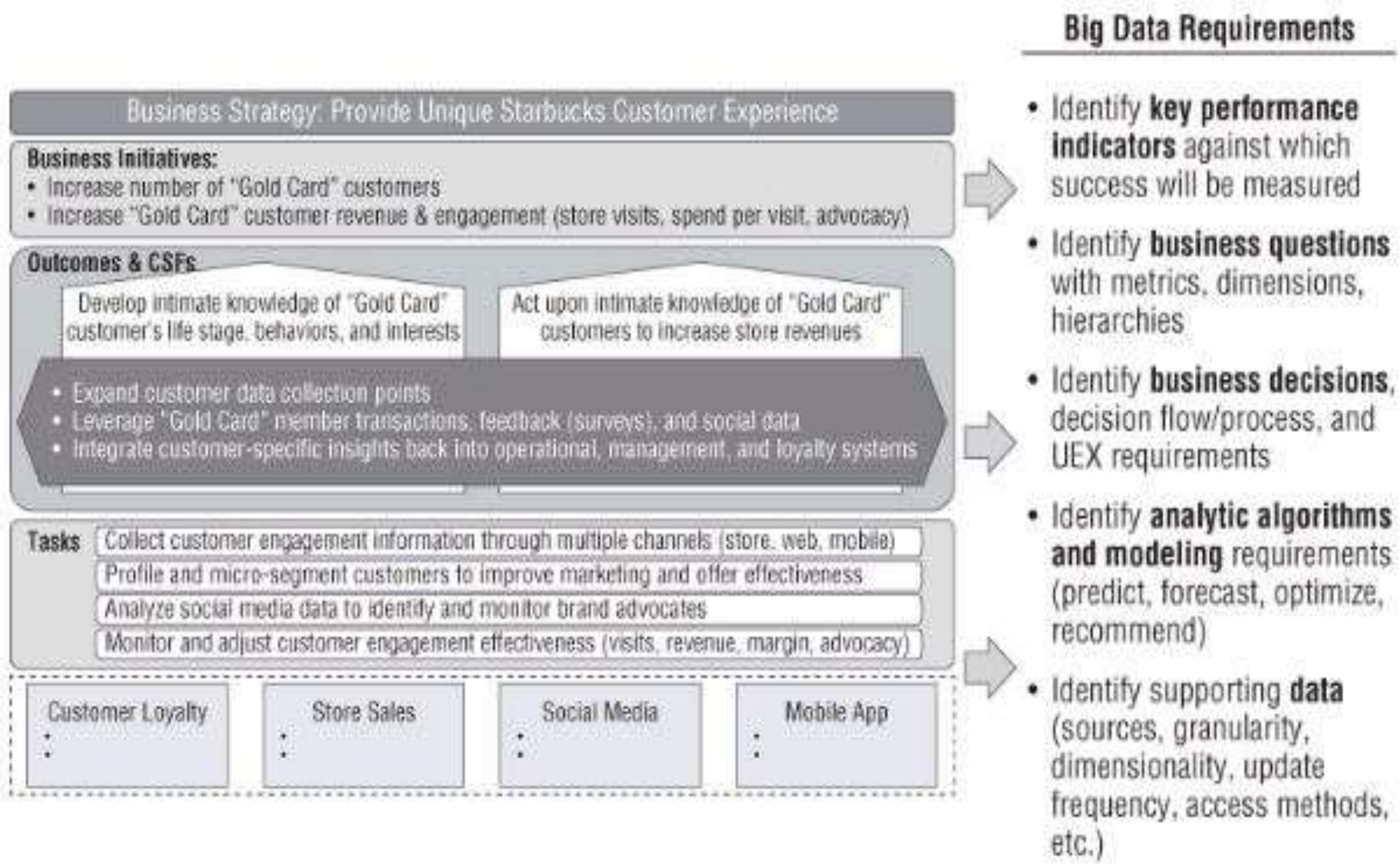
Ultimately, value chain modeling offers the following benefits:

- Cost reduction.
- Competitive differentiation.
- Increased profitability and business success.
- Increased efficiency.
- Decreased waste.
- Higher-quality products at lower costs.

- **Business Strategy:** The title of the document states the business strategy upon which the big data initiative is focused, in this case, “Improve customer intimacy to drive more profitable customer engagements.” The title sets the scope of the strategy—you're focused on improving customer relationships, not improving predictive maintenance of network components—but you can see that this is not yet enough detail to be actionable.
- **Business Initiatives:**This section captures the business initiatives that support the customer intimacy business strategy. These business initiatives capture the desired end goals, outline what the business hopes to achieve, and define how success will be measured. Examples of relevant business initiatives to support the customer intimacy business strategy could include:Increase memberships, such as sell more memberships, renew more memberships, or leverage advocacy to drive new memberships.Increase customer engagement, such as sell more products, provide more services, or co-market complementary services.
- **Outcomes and Critical Success Factors** This section contains the “what needs to be done” details to support the successful execution of the customer intimacy business initiatives. Examples of relevant CSFs could include:Develop an intimate understanding of your customer's life stage, behaviors, and areas of interests.Act upon an intimate knowledge of your customers to create demand and stimulate purchases.Optimize distribution or customer contact points through customer communications and customer-centric products and services.Acquire and leverage additional member and household information including customer profile, segmentation, and economic value.Integrate customer insights and actionable recommendations into operating platforms including contact management, performance metrics, and analysis tools.

- **Tasks:** This section provides the next level of detail regarding the specific tasks around which the different organizations will need to collaborate to successfully execute against the different business initiatives (or the “how to do it” stage). This could include the following tasks (again, there are likely 8 to 12 of these tasks).Collect information (via increased use of surveys, question asking, and online instrumentation) during in-bound and out-bound contacts with each customer Generate actionable intelligence on each customer household Use that actionable intelligence to have relevant interactions with each customerTrack operational execution and results.
- **Data Sources:** Finally, the document highlights some of the key data sources required to support the key business initiatives. In this case, you would need the following data sources to start: Customer data (demographic, behavioral, psycho-demographic)Transaction data (purchases, returns)Contact data (consumer comments, e-mail threads, social media dialogues)Marketing data (campaign spend, leads, conversions)

Big Data Strategy Document identifies supporting big data requirements



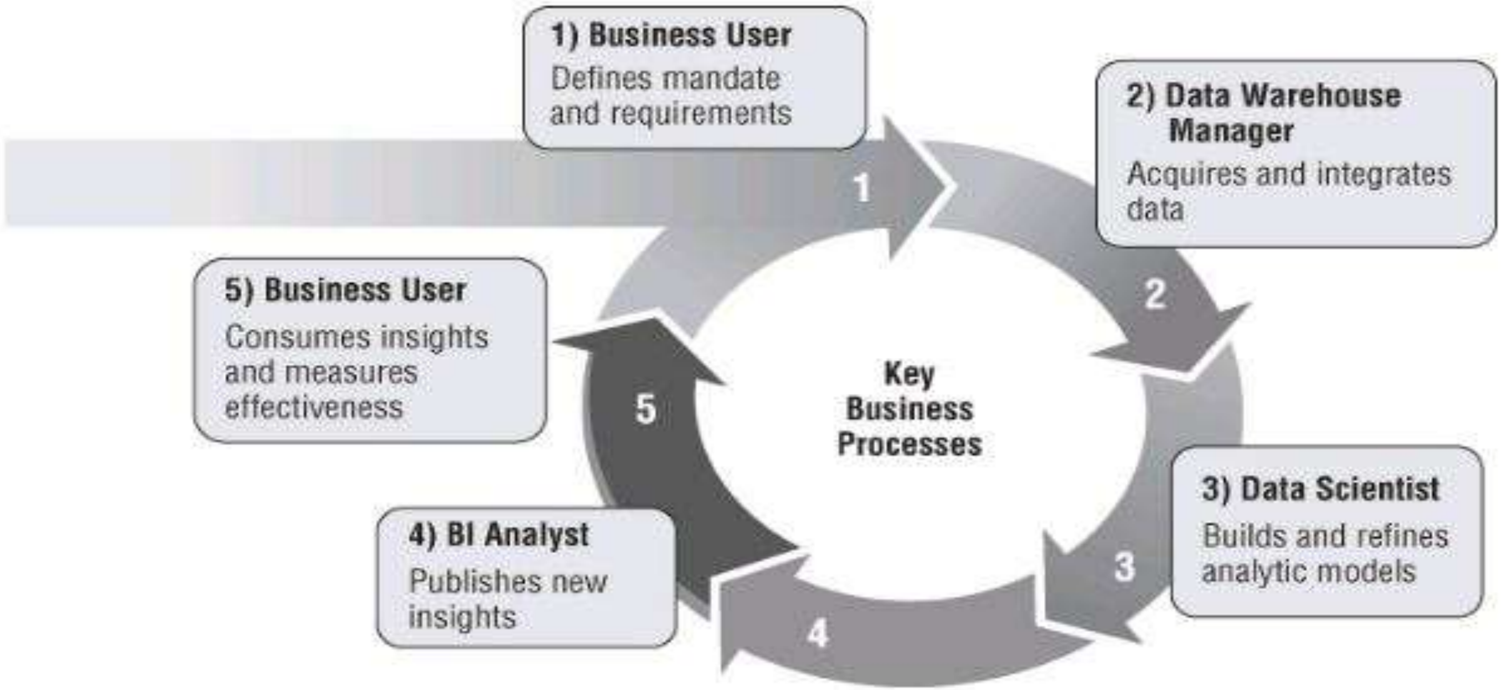
The strategy document breaks the technology requirements into the following components:

- Identify the metrics, measures, and key performance indicators against which the progress and success of each of the business initiatives will be measured.
- Identify the business questions, metrics, and dimensions (using “by” analysis) necessary to support the business initiatives. You should also capture any hierarchical business relationships at this stage.
- Identify the business decisions and the flow of the decisions required to support each key task. Test or prototype these key tasks and decisions to validate that you have captured all the necessary business questions, metrics/facts, and dimensions. Capture the business decisions, decision flow/process, and user experience (UEX) requirements
- Identify the analytic algorithms and data modeling and transformations required to support the predictive components of each of the key tasks. Look for opportunities to insert new predictive “verbs,” such as score, forecast, optimize, recommend, and predict, into the business questions and business decisions.
- Identify the supporting data sources including measures, dimensionality, dimensional attributes, granularity, update frequency, storage location, and access methods.

As a result of this work, you are now in a position to define the technology stack and required data and analytics architecture including the Master Data Management (MDM), ETL/ELT/data enrichment, data warehousing, BI, and advanced analytics requirements necessary to support the Customer Intimacy business strategy.

Data Analytics Lifecycle

- Successful big data organizations continuously uncover and publish new customer, product, operational, and market insights about the business. Consequently, these organizations need to develop a comprehensive process that not only defines how these insights will be uncovered and published, but clearly defines the roles, responsibilities, and expectations of all key stakeholders including the business users, data warehouse managers, BI analysts, and data scientists. Let's use the **Data analytics lifecycle** to gain an understanding of how these different stakeholders collaborate



- **The business user** (which also includes the business analyst) is responsible for defining their key business processes, and identifying the metrics and key performance indicators against which those business processes will be measured. The business users are the ones who understand what questions they are trying to answer and what decisions they are trying to make. The business users are the ones who are trying to leverage the available data and insights to answer those questions and make those decisions.
- **The data warehouse manager** (or DBA in some cases) is responsible for defining, developing, and managing the data platform. The traditional tools of choice for this stakeholder has historically been data warehouses, data marts, and operational data stores. However, new technology innovations are enabling the data warehouse manager to broaden their role by considering new technologies such as Hadoop, in-memory computing, and data federation. These new data platforms support both structured and unstructured data and provide access to data located both inside the organization as well as select data sources that exist outside the four walls of the organization. These modern data platforms also support the ability to ingest and analyze real-time data feeds and enable the "trickle feeding" of data into the data platform.

- **The data scientist** is responsible for mining the organization's data—structured and unstructured data that is both internal and external of the organization—to uncover new insights about the business.
- Data scientists are data hoarders, seeking out new sources of data that can fuel the analytic insights that power the organization's key business processes.
- The data scientist needs a work environment (analytic sandbox) where they are free to store, transform, enrich, integrate, interrogate, and visualize the data in search of valuable relationships and insights buried across the different data sources.
- The data scientist needs an environment that allows them to build, test, and refine data models rapidly—measured in minutes and hours, not days, and weeks—and embraces the "fail enough times" approach that gives the data scientist confidence in the quality of the analytic models.
- "Fail enough times" refers to the point in the analytic model development and testing process where the data scientist has "failed" enough times in testing other variables and algorithms that they feel confident that the resulting model is the best analytic model.

- **The BI analyst** is responsible for identifying, managing, presenting and publishing the key metrics and key performance indicators against which the business users will monitor and measure business success. BI analysts develop the reports and dashboards that the business users use to run the business and provide the “channel” for publishing analytic insights through those reports and dashboards to the business users. This is where the real-time, predictive enterprise vision comes to fruition.

And finally, the analytic process circles back to the business users who use the resulting reports, dashboards, and analytic insights to run their business. It is the business users, and the effectiveness of the decisions that they make, who ultimately determine the effectiveness of the work done by the data warehouse manager, data scientist, and BI analyst. Finally, the results of the decisions that the business users make can be captured and used to fuel the next iteration of the analytic lifecycle.

- The exact nature of the roles, responsibilities, and expectations of these different stakeholders will vary from organization to organization, and even project to project. Some business users may be more comfortable with statistics and predictive analytics, and may seek to do some of the analytic work themselves. Same with the BI analysts who are looking to broaden their skill sets with advanced analytics and data visualization skills.
- It should be noted that the roles and responsibilities for each stakeholder are centered on a targeted key business processes. The roles and responsibilities might very well change for each key business process, depending upon the skills, capabilities, and areas of interest of the different stakeholders. So view this analytics lifecycle more as a framework to provide some level of guidance for organizational collaboration, versus a fixed set of roles and responsibilities that ignores the individual skills and interests of the different stakeholders.

Big Data user experience ramifications

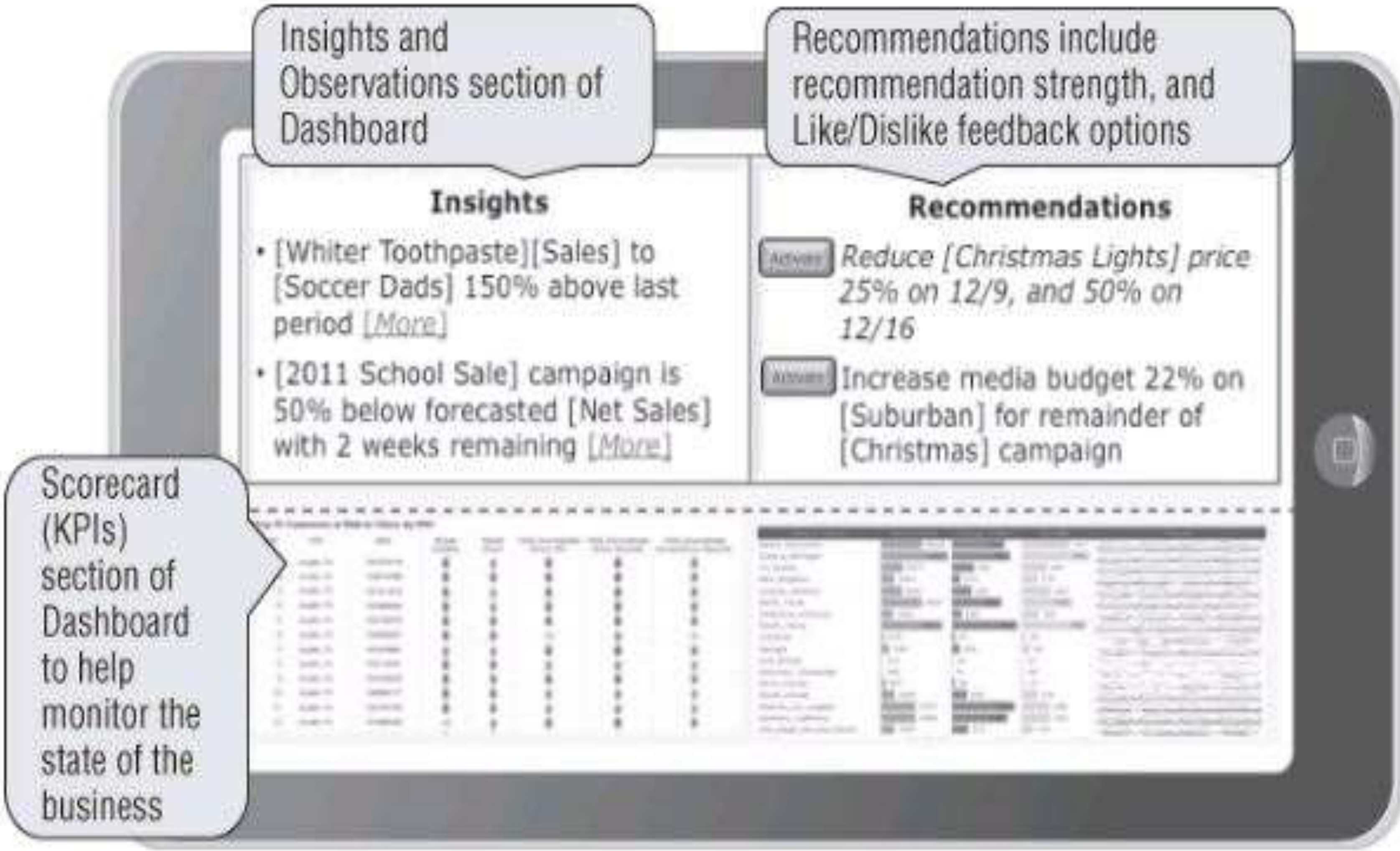
- This new insights-lead analytics process can enable an entirely new, more productive user interface. Traditional dashboard interfaces present the business users with a multitude of seemingly unrelated charts and tables. The data discovery process of trying to find something of interest in the charts and tables (that is, slicing and dicing, drill across, drill down) is left to the user.
- However leveraging the insights-lead analytic process, the user interface can be simplified to present only the information or insights needed to run or optimize the business.
- Think how the iPod revolutionized the established MP3 market by providing a fundamentally simpler user interface—one that presented only those capabilities that enabled anyone to play the songs and the playlists that they wanted to play.
- With the advanced analytics enabled by big data, the user interface could focus on the delivery of two key pieces of **information—insights and recommendations.**

- Insights are unusual behaviors or performance (for example, two standard deviations outside normal performance, 200 percent above or below predicted performance) that might require further investigation by the user. Insights would leverage both simple (time series trends, previous period comparisons, benchmarks) and advanced (predictive analytics, data mining, regression analysis) analytic models to identify performance situations operating outside of normal boundaries. These insights would be starting points for a more detailed investigation by the user.
- A few example insights are:
- Did you know that product A's sales to the [Soccer Dads] customer segment is 150 percent of what it was last period?
- Did you know that marketing campaign [2011 Back to School Sale] is 50 percent below forecasted conversions with only 2 weeks left in the campaign?
- Did you know that the variance level on machine [Charles City Plant Turbine 120] is 20 percent outside the normal control boundaries?

Recommendations are specific actions that are generated based on a detailed analysis of the insights and current state of the business. Recommendations would leverage advanced analytic modeling and real-time feeds to analyze the key business drivers and variables, update or fine-tune analytic models, and make specific recommendations.

- A few examples of recommendations are:We recommend marking down the product category [Christmas Lights] by 25 percent starting December 9, and increasing the markdown to 50 percent on December 16.
- We recommend increasing the media budget by 22 percent on display ad [Chevy Suburban] and decrease media budget 33 percent on display ad [Chevy Volt] for the remainder of the campaign [Holiday Season].
- We recommend repairing your [Maytag Model 3200] washer's drum engine within the next 5 days because there is a 95 percent probability of product failure.
- We recommend that patient A101-23V be admitted into the hospital for an extra day due to the high probability of readmission.

Big Data Analytics-Enabled User Experience



- The analytics that underpin the insights and recommendations can be quite complex, but such analytic complexities are likely not of concern to the business user. The business user wants to have the data tell them what's happening in their business, and wants the technology to make recommendations based on previous learning and best practices. Also, the analytics can be personalized and self-learning so that it is constantly fine-tuning the analytic models based upon the user's feedback on what they like, don't like, and why (think about how Pandora, the online music service, uses “like” and “dislike” feedback to learn more about your music preferences).
- Massive, detailed data sources, coupled with more powerful analytic tools, provides the capabilities to identify significant, material, and actionable insights in the data without forcing users to have the analytic skills or training to quantify why things happened. It enables a completely different user interface—one that is focused on providing greatly simplified insights and recommendations—to help business users optimize their key business processes.

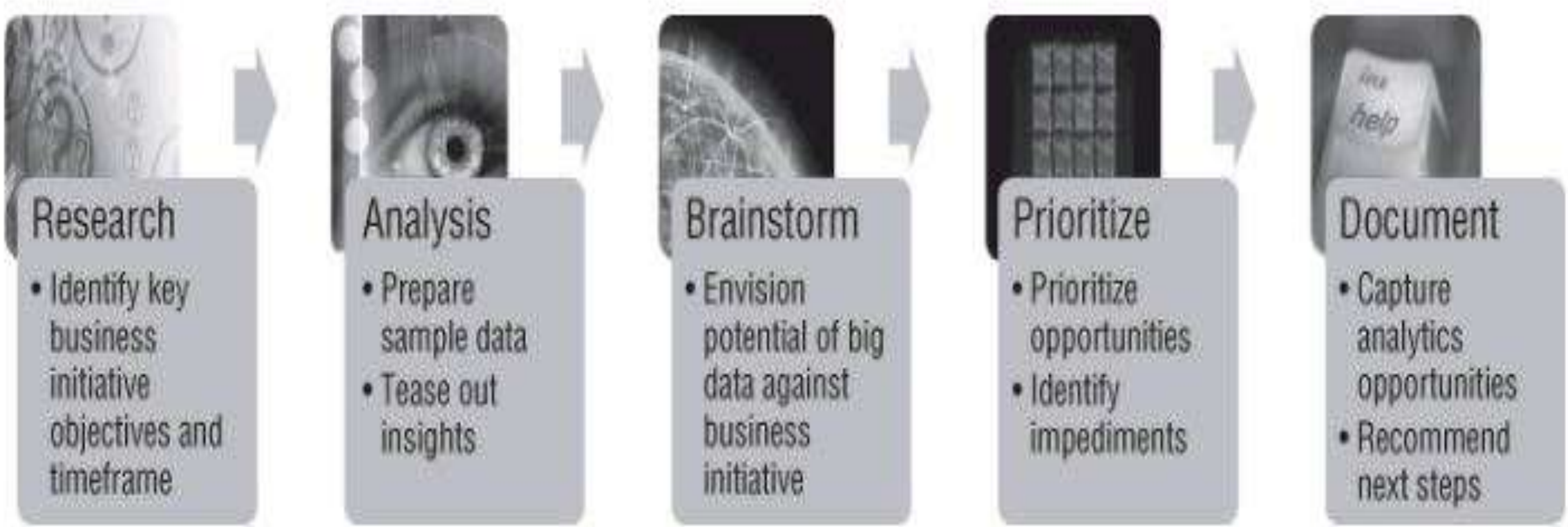
Identifying Big Data Use Cases.

The Big Data Envisioning Process

The envisioning process, which is called the vision workshop, defines where and how big data and advanced analytics can be deployed to transform your business. The vision workshop process is comprised of the following steps..

1. Research and interviews to understand the targeted business initiative or business process.
2. Data preparation and client-specific analytics development.
3. Envisioning exercises to convey the “realm of the possible”.
4. Brainstorm and prioritize big data use cases.
5. Capture implementation risks and business value drivers

The Big Data Envisioning Process

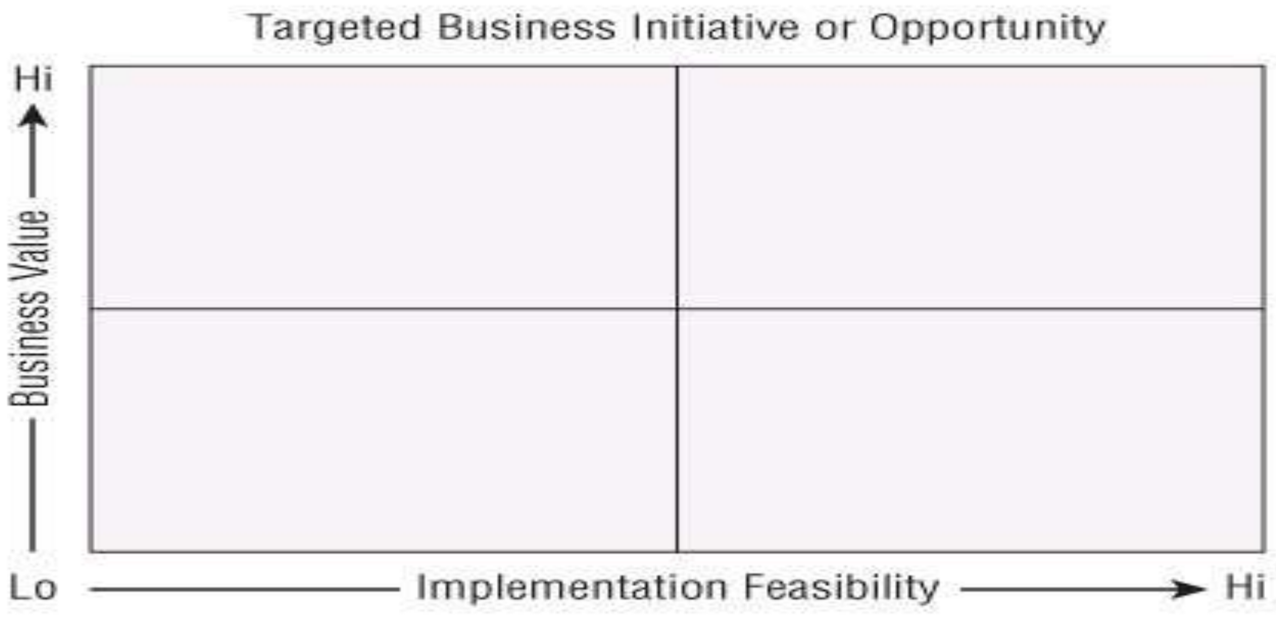


The Prioritization Process

- One key challenge to a successful big data journey is gaining consensus and alignment between the business and IT stakeholders in identifying the initial big data business use cases that deliver sufficient value to the business, while possessing a high probability of success. One can find multiple business use cases where big data and advanced analytics can deliver compelling business value. However, many of these use cases have a low probability of execution success due to:
- Unavailability of timely, accurate data
- Lack of experience with new data sources like social media, mobile, logs, and telemetry data
- Limited data science or advanced analytics resources or skills
- Lack of experience with new technologies like Hadoop, MapReduce, and text mining
- Architectural and technology limitations with managing and analyzing unstructured data, and ingesting and analyzing real-time data feeds
- Weak working relationship between the business and IT teams
- Lack of management fortitude and support

- Prioritization Matrix works to not only prioritize the initial big data use cases, but how to use it to foster an atmosphere of collaboration between the business and IT stakeholders.
- The prioritization process is the single most important step in the envisioning process. While I expect that most readers would think the brainstorming process is the most important, the truth is that many use cases are probably already known ahead of the brainstorming session. The brainstorming session is useful in validating and expanding on those known use cases and helping to fuel the identification of additional use cases.
- The Prioritization Matrix is a 2×2 grid that facilitates the interactive process and debate between the business and IT stakeholders to determine where on the matrix to place each use case in relation to the other use cases. The use cases are placed on the matrix based
- **Business value:** the vertical axis of the matrix. The business stakeholders are typically responsible for the relative positioning of each business use case on the Business Value axis. The Business Value axis reads from low business value at the bottom to high business value at the top as shown in Figure on:

- **Implementation feasibility:** the probability of a successful implementation considering availability, granularity and timeliness of data, skills, tools, organizational readiness, and needed experience. Implementation feasibility is the horizontal axis of the matrix. The IT stakeholders are typically responsible for the relative positioning of each business use case on the Implementation Feasibility axis. The Implementation Feasibility axis reads from low implementation feasibility on the left (higher probability of failure) to high implementation feasibility on the right (higher probability of success).



The Prioritization Matrix

As a reminder, you are not looking for the exact valuation of each use case from a Business Value perspective. Instead, you want to know the relative business value of each use case and some level of justification from the business stakeholders as to the reasoning behind the placement of the use case.

Some more ways are…

1.Identify all possibilities that fall under big data use case categories

There are several big data use case categories, including customer insights, fraud management, and digital advertising, to name a few. Let’s take the example of customer insights. Most of the industries collect a large volume of customer data from various sources, such as social media platforms, interaction with customer care agents, data from wearable devices, and much more. Collecting customer information means it is a B2C company, having goals of attracting new customers, retaining the old ones, and keeping them happy altogether.

2.Select the most relevant big data use cases

Once you select the big data categories, the next step is to identify appropriate use cases. In our earlier example, we have selected ‘customer insights’ as our big data use case category. For companies that might choose ‘customer insights’ as one of their big data categories, should identify how exactly customer data can benefit their organization. Few specific use cases for the category ‘customer insights’ are gauging customer sentiments, attracting new customers, reducing customer rate of attrition, enhancing customer experience, retaining existing customers for long, providing great customer support. Similarly, companies should identify use cases for the selected category.

3.Recognize the complexity of analytics required for the selected use case

The last step is to estimate the complexity of data analytics by analyzing the level of complexity of data needed for a particular use case. The more complex the data is, the more complex is the analysis required. Hence, CAOs should carry out a complete analysis of the level of analytics involved.


Thank you!!! ...