Taming The Data Elephant!

Elm WHITE PAPER SERIES | DATA CONSULTATION GROUP | MAY 2016

AUTHORED BY: WAEL ALALWANI AND ZAHER ALHAJ HUSSEIN

REVIEWED BY: FRANK BUYTENDIJK

GARTNER RESEARCH VP AND DISTINGUISHED ANALYST

DATA MANAGEMENT & INTEGRATION

Elm

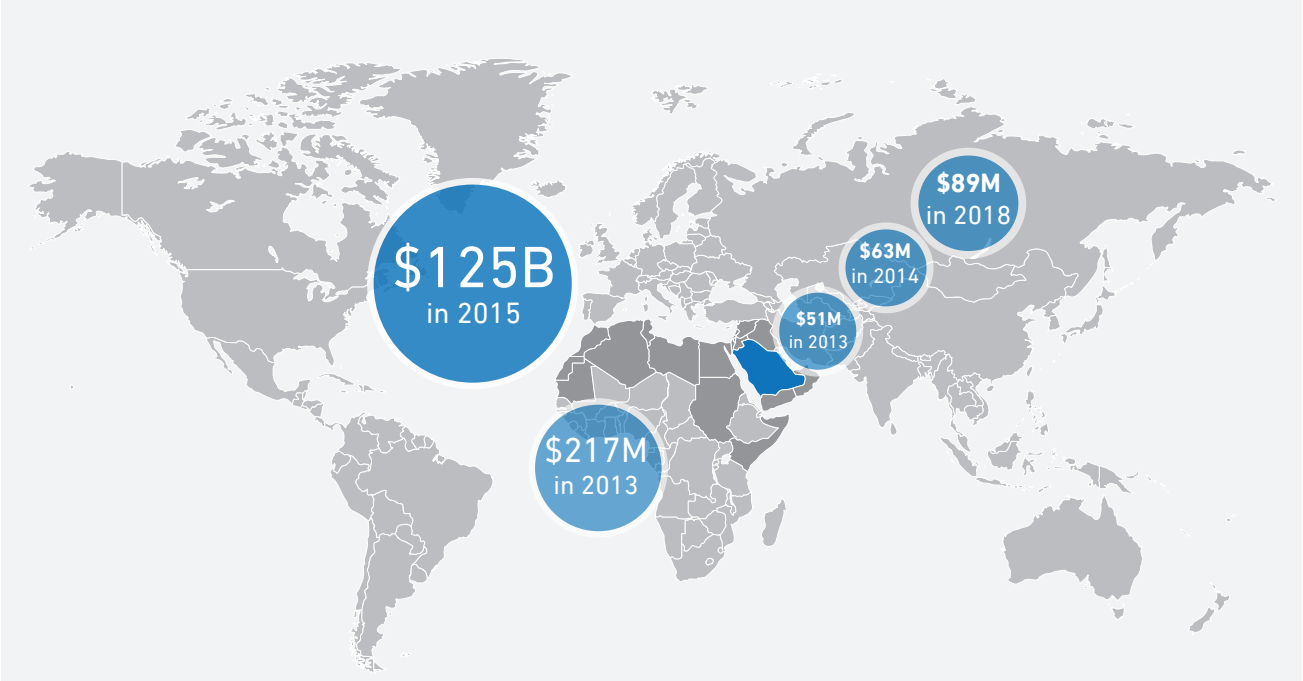
Challenges and Lessons from the Data Frontiers

Discussing the untold story about Business Intelligence and Data Analytics projects

Content

Section 1

Executive Summary



> *In Saudi Arabia, the spending on business analytics software was expected to reach $62.7 million in 2014, compared to $51 million in 2013 . The spending is expected to keep growing and exceed $89 million in 2018.*

Amount of Data We Generate

The world has generated more than 1.8 Zettabyte in 2011 (1 Zettabyte = 1 Trillion Gigabyte). The volume of generated data massively increased in 2013 to reach 4.4 ZB, and the forecasts indicate that it will reach 44 ZB in 2020. If such data were to be stored in 128GB iPads, the stacked tablets would create 6.6 stacks covering the distance from the Earth to the Moon[1].

Spending on BI & Analytics Software

In 2015 the overall big data and analytics market would reach $125 billion worldwide in software, hardware infrastructure, and services[2]. Middle East and North Africa (MENA) totaled $217 million in 2013 in Business Intelligence (BI) and analytics software[3]. In Saudi Arabia, the spending on business analytics software was expected to reach $62.7 million in 2014, compared to $51 million in 2013[4]. The spending is expected to keep growing and exceed $89 million in 2018[5].

This white paper draws attention to a wide range of challenges commonly faced in Business Intelligence and Data Analytics projects. Taking the reader in a journey, the first half of the paper discusses these challenges as a story, portraying a failure that the CEO of a fictitious Company "Food for Us", Mr. Samer, has faced. This occurred while an investment in

BigData initiative at the company was taking place. The failure was the result of many challenges encountered across the strategy, governance, operating model, and data quality levels. In the quest of understanding what went wrong, the second half of the paper analyzes the challenges and ends with Elm recommendations to overcome such challenges.

General Takeaways

The white paper presents a set of general lessons to consider:

1. Investment in technology alone is not a silver bullet. It has to be supported with effective processes, right people, and reliable data.
2. Data culture is not just a fancy concept. According to Gartner, "While 80% of CEOs claim to have operationalized the notion of data as an asset, only 10% say that their company actually treats it that way"[6]. Data culture is an essential and evolving ecosystem that should be ingested in any work environment dealing with data. 81% of respondents representing IT decision makers in Saudi Arabia agreed in a survey conducted by EMC that the biggest barrier to dealing with business priorities is cultural barrier[7]. This is why special emphasis was put on the data culture in this white paper.
3. Data is a corporate Asset like capital, real estates, and human resources. It is not "IT Stuff". Business must own the data, not the IT.
4. Although business should be the main driver, companies should keep an eye on the technological disrupters, which usually change the rules of the whole game and create unimaginable new use cases. The gap between business and technology should be carefully bridges to avoid many challenges.
5. Having huge data does not imply being able to extract reliable insights in a blink of an eye. Data must undergo multiple phases of processing and reasoning (inductive, deductive, and abductive reasoning).
6. Data-oriented projects like BI act as a magnifying glass exposing problematic work cultures like verbal buy-ins (lack of diligence), crippling politics, unsustainable positivity, etc.
7. There is an utmost importance for adopting holistic, comprehensive, and end-to-end framework when tackling data challenges. The framework can be used for multiple purposes (Diagnosis, Planning, Communication, Technical Evaluation, and Solutions Design). Adopting partial and shortsighted solutions might cause sub-optimization.
8. Training your employees is important, but this does not make them instantly qualified. Patience and the right setup are required. For example, establishing the proper mechanisms to foster continuous feedback loops, open communication, and collaborative work. Co-location is a common practice to achieve this.
9. Generic Agile concepts do not directly apply to data analytics projects. Specific tweaks and modified approaches must be introduced. For example, many analytics/BI projects heavily depend on experimentation, making the sprint planning, in its traditional sense, not possible. As a result, it is important to think of ways to integrate such experimentations into the agile methodologies.
10. Traditional hiring practices, widely applied in the local market, should be replaced with better ones to guarantee recruiting qualified employees for business and data analytics projects. Some notable practices are: full-day assessments and case interviews.
11. Administrative and transactional Data alone

will give you good answers, yet insufficient. Collecting data from external data sources whether public, 3rd party, or survey data is essential depending on the questions you want to answer. In global markets, we believe that more than 50% of essential business data will be external in nature.

12. Any company planning to invest in BI and data analytics projects must assess the maturity of its data, processes, people, and technologies. Such projects are highly prone to failures if there was a lack of proper preparation and assessment. Companies are advised to take a bimodal approach. According to Gartner, there are two models: Operation (Keeping things running) and Innovation (Exploring the unknown) so "In order to succeed with a bimodal approach, you must start with culture, and specifically create three subcultures — operator, innovator and guardian."[8]

13. The breadth of data analytics projects stretches between efforts applied to two ends: data integration and analytics/dashboards/reporting. It is frequent that deficiencies in one end create challenges to the other. This usually results in redundant and non-scalable efforts. It is important to find the right balance and solve the challenges where they should be solved.

14. Handling structured and unstructured Arabic data has its own set of challenges, e.g. the lack of proper out-of-the-box data quality knowledge base component.

15. Without clear business needs, BI and BigData projects will not bring any added value. Also, they should be carefully studied in order to know how they fulfill the desired needs. It is not uncommon to hear senior people in the local market mentioning BI, BigData, and other related terms interchangeably without really understanding what each term means.

Goals of the Paper

This paper discusses the first step of the data analytics journey at Food for Us Company, which is the BI project and the challenges associated with it. Elm has been part in number of BI projects and has gathered many insights about the real challenges that exist in the Saudi BI market in both the public and private sectors. The challenges related to the local market are not very different from the global market ones, however, their order and scale are somewhat dissimilar and are shaped by specific realities governing the Saudi market. The white paper follows a non-linear story-telling narrative, where all the characters are fictitious.

The paper has the following goals:

- Discuss the common challenges and obstacles associated with BigData as well as enterprise-level BI/Analytics projects.
- Share insights, lessons-learnt, tried-and-proven solutions, and takeaways for the audience who already started implementing data analytics and BigData projects.
- Engage the audience who is considering investing in BigData projects in a journey where a set of eye-opening points worth consideration will be explained (i.e. getting to know the road ahead).
- Shed lights on the issues that are more relevant to the Saudi market as opposed to other international and regional markets.

The targeted audience for this white paper is:

- Executive IT and Business leaders who are implementing or planning to implement Enterprise BI solutions in their companies (e.g. CIO, CEO, CDO, IT Managers, Enterprise applications managers, etc.)
- Project managers and delivery managers who are in charge of managing sophisticated BI projects.
- IT professionals who would implement BI projects hands-on (business analysts, BI analysts, solution architects, integration architects, lead BI/ETL developers, etc.).
- IT consultants who would get involved as an advisory arm in BI projects.
- BI Vendors and anyone interested to know the challenges of data projects in the Saudi market.

---

[1] "The Digital Universe of Opportunities", IDC-EMC

[2] "IDC Predictions 2015: Accelerating Innovation and Growth", IDC

[3] "Gartner Says MENA BI and Analytics Software Market Grew 11 Percent in 2013", Gartner

[4] "ICT-Based Transformation to Lead KSA Government Spending", IDC

[5] "Spending on Business Intelligence Software Tools in Saudi Arabia will grow rapidly", IDC

[6] "How to Establish a Data-Driven Culture in the Digital Workplace", Gartner

[7] "EMC Survey Reveals Big Data Adoption Trends in Saudi Arabia", EMC. Note that there were very few resources that cover the Saudi Market with respect to such aspects.

[8] "How to Start Implementing the Bimodal Office of the Chief Data Officer", Gartner

Section 2

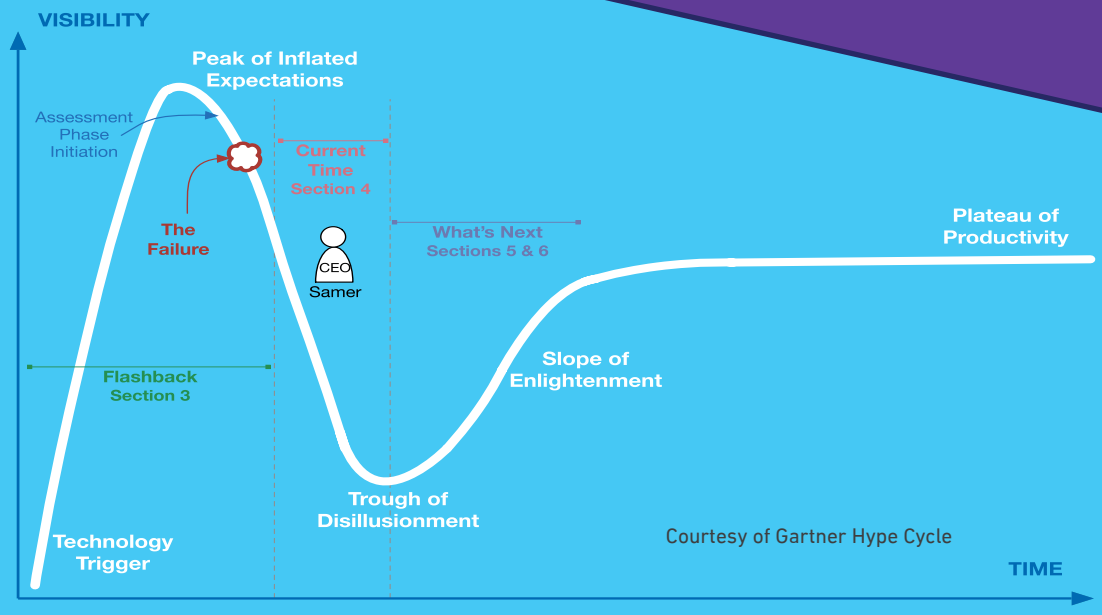
When Bad Things Happen to Good Companies

For the second time in one day, the chief auditor of the joint investigation committee tried to reach Mr. Samer. As the CEO of "Food for Us" company, a major company in the Food industry, Samer found himself in a devastating crisis, as a serious quality issue was discovered in one of his company's products. Alarmingly enough, this incident stirred the social media up against the company. Consequently, a joint committee was formed by the local Industry and Health regulators to start an investigation to resolve this issue. The company was instructed to submit official compliance reports to the committee.

Unfortunately, the submitted reports contained misleading information. As a result, the company was ordered to pay 5 million Riyals as a penalty. Mr. Samer feels very frustrated after this failure and financial loss.

Samer had a sudden flashback remembering the journey his company went through when he decided to build the BI capability in the company as the first step towards a bigger BigData initiative. He realized that Gartner's famous hype cycle really represents the states that he has been into. Starting with the technology trigger, moving up to the inflated expectations phase, and finally being struck with realities he did not foresee leading him towards the trough of disillusionment. The quality failure happened in a very critical time, when Samer was in the middle of assessing the BI investments outcomes. During this assessment, he believed that he has just uncovered some challenges, and the rest are going to be revealed by the end of the assessment. The following figure describes where Samer currently stands, and under which section each experience will be discussed.

Skeleton of the White Paper



Section 3

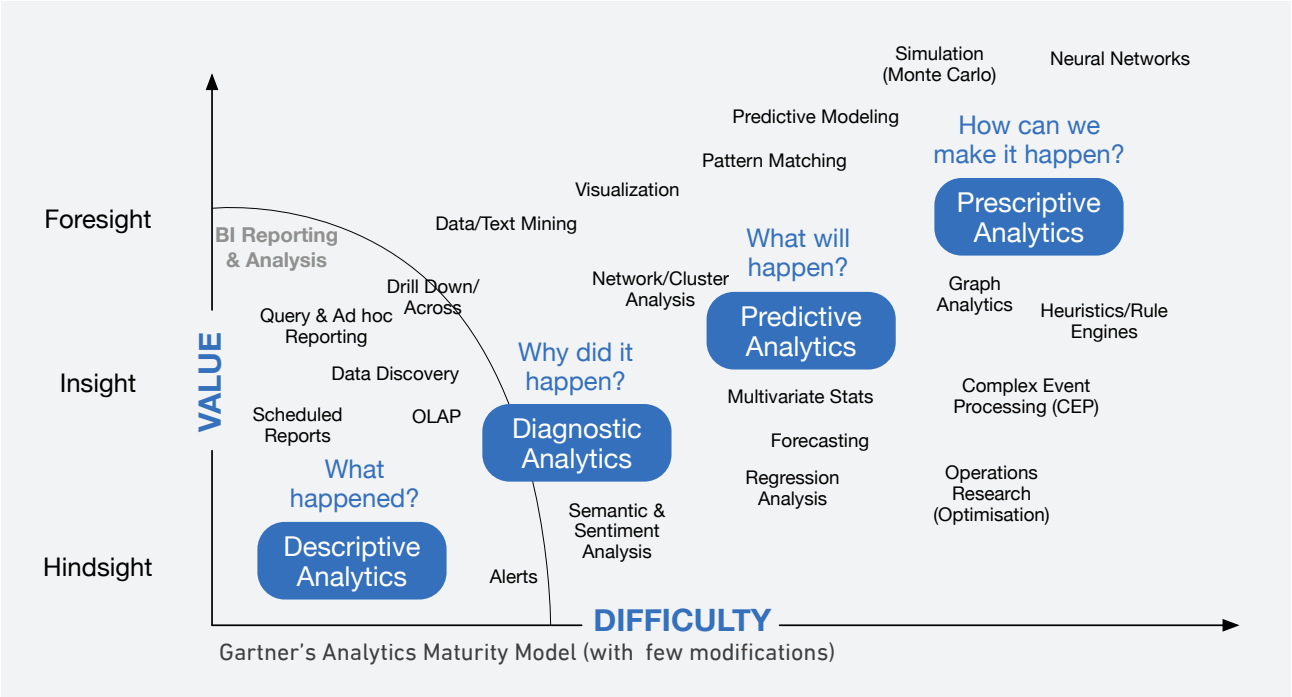
Flashback: The Original Expectations

BI Technology Trigger

One year ago, feeling frustrated from taking all critical decisions based on gut feelings and subjective judgments, Samer was in a meeting with the CIO and Enterprise Architect (EA) discussing the importance of practicing data-driven and fact-based decision making. They realized that "Food for Us" was lagging behind as their competitors have gone far ahead by successfully utilizing BI technologies. So, the EA took the CEO and CIO in a journey around the available technologies and what they can achieve. The EA has been a keen advocate for best utilization of the data as a real value-adding asset.

Afterwards, they started to list the main strategic business drivers to implement new BI system, as the stepping stone towards a comprehensive BigData initiative. As highlighted many times by the EA, the first driver is: current **operational misalignment** between the sales and manufacturing departments that is causing serious problems as the demand-and-supply balance is unattained. The second driver is that the strategic management has no detailed-enough analytical information to take **strategic decisions** like: entering new markets, introducing new products, retiring existing products, etc. The last business driver is that the company must comply with many **stringent regulatory rules**, so having correct and detailed reports for compliance reasons is a high priority.

This BI investment was the first step towards richer analytics roadmap (as demonstrated in the chart below) ending with data analytics including diagnostic, predictive, and prescriptive analytics. Samer was determined to embark this initiative as soon as possible.



Gartner's Analytics Maturity Model (with few modifications)

Rising Expectations

With the executive's full support, the CIO and his team chose the best-of-breed BI tool designated for producing the Key Performance Indicator reports as well as operational performance monitoring, and decision management (e.g. Balance Score Card and Strategy Maps) reports. Samer met all the executives and managers and everyone was very enthusiastic for the project with a wholehearted buy-in.

In a relatively short time, a competent BI vendor started the implementation with huge focus on dashboards design, usability, and the aesthetics of data visualization using samples of real data extracted from the company's internal systems. This took place after identifying the main data sources, which are the ERP Databases, along with secondary sources including social media sites. Everything seemed to work smoothly, and needless to say, Samer's expectations were peaking as the project was reaching its end.

1- The Data Quality Struggle

Believing in the investment he made to have interactive dashboards and periodical reports, Samer started to analyze the reported insights that are directly related to the company's business.

In multiple bar charts, he saw some aggregated values being reported despite that they directly violate essential business rules. Samer asked his CIO about this, and the latter informed him that this is the way data is stored in the products database, and it seems no effort has been made to assess the data against these specific business rules in order to rectify accordingly. Samer then looked at a major trend line chart and was skeptic about how smooth the line was. The data team lead informed him that the data team assumed every outlier as an erroneous value, hence, immediately eliminated. Samer started to become worried after hearing this inaccurate assumption.

"Why can't we see the age groups' distribution of Ha'il on the map visualization? Wait.. and why are there two different 'Jeddah' records in Arabic in the table beneath?" Samer shouted. The Data Team Lead explained that there is missing data for Ha'il, and that the tool they brought doesn't automatically aggregate the slightly-different Arabic city names like (i.e. Jeddah/Jedda). The data team lead also listed many data quality issues they are struggling with, but with little luck. Samer continued to read the recommended actions to take based on the presented information. The actions were suggested by data and business analysts in the data team, but he found them truly misleading as they were based on the information aggregates he just saw and considered doubtful. At this moment, Samer started to lose his faith in the data and realized that with such data incosistency, any decision to be made would be highly unreliable. He was particularly upset because he knew that there was a quality assurance stage before getting such reports and dashboards. Yet, the data quality proved to be problematic.

2- Operating Model Challenges

After a deep retrospective session between Samer, CIO, EA, and data team lead, they realized that the data quality challenge must be the result of other causes, either partially or fully. The operating model, defined as: how the company gets things done; that is the whole processes, activities, rules, and procedures that the company utilize to create, deliver, and capture value, started to become the suspect. Samer delegated the CIO to find out what the causes are (check the next figure to get an idea about Samer's company, some general information, and the organization structure along with the involved teams).

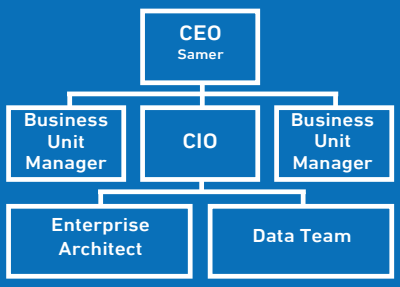
Company Name

Food for Us, Market Leader in Food Manufacturing

Business Strategy

- **Entering new regional markets within 2 years,**
- **Enhancing new product development innovation to come up with 3 new products every 3 months,**
- **Improving the responsiveness to the market by enhancing the operational performance,**
- **Revamping the forecasting capability.**

Organization Structure



General Information

- **Riyadh based (Saudi Arabia)**
- **Lead manufacturing company in Food industry.**
- **5 Billion Riyals revenue**
- **10,000 employees**
- **Year 2015**

With extensive investigation, it became apparent that the issues started at very early stages like how the employees enter the data into the applications (i.e. data capturing) and all the way to the report/ dashboard development stage. Moreover, it was realized that the buy-in collected for such BigData initiative within the company from the Business Unit managers stayed verbal and never turned into fully tangible actions. To make things worse, there was a fierce resistance by middle managers to collaborate, as many were worried that such initiative will expose their departments. This resulted in serious disruption to the overall operations and processes. Although the CEO was fully aware of it, the CIO was not really aware of what was happening because the team used to deliver relatively on time. However after this investigative activity, the CIO understood that the team was operating in an ad-hoc fashion rather than Agile, despite that everyone was saying "we work with agile methodologies". Fire-fighting mode, poor documentation, jeopardized quality, inaccurate analysis, and many more define what adhocism is all about.

When confronted with these findings, the data team (data warehouse and analytics team) confirmed that they are struggling with the operating model in terms of Processes, Technology, Data, and People and an urgent help is required. On the **Process** side, the team did not find generic agile concepts directly suitable for Business Intelligence and Data Integration projects. On the **Technology** level, the team stated that the technical tools were not properly selected to fit their needs and expertise. Moving to the **Data** aspect, the team described how difficult it was to get basic data from other departments, and that everyone tries to make any excuse in order to avoid sharing it like data privacy infringements. Finally, on the **People** side, the team is looking for: support in managing the technical BI vendor that was not properly selected, defining better ways to collect the requirements, and finding the right talents and well-experienced employees. The team ended its debriefing with a very appalling statement that many mangers and deputies in the company do not know what they need, and that their requirements are unclear.

The Failure

Adding insult to the injury, the failure broke out while Samer and the CIO were in the middle of the BI challenges assessment. A serious quality issue was discovered in one of the company's products, which stirred the social media up, and led the local Industry and Health regulators to open an investigation to resolve this issue. Samer's company submitted some official compliance reports to a committee representing the regulators. Unknowingly, the reports contained misleading information beyond the expected level of inaccuracy that Samer had in mind. He knew the compliance reports have some inaccuracy due to the issues he discovered lately, but he thought it was manageable for these specific reports. He did not realize that the numbers severely conflicted with previous compliance reports as well as general statistics published by the Industry regulator.

As a result, the issue escalated very quickly, and the company was ordered to pay 5 million Riyals as a penalty. Despite the fact that Samer felt very frustrated after this failure and the resulting financial loss, he was highly determined to continue the BI assessment process to uncover all the issues in order to resolve them. Samer still believes in the importance of the BigData initiative on his business, regardless of the associated challenges and hassle.

Section 4

Back to "Now": Uncovering More Challenges

In the quest of understanding what went wrong, Samer asked the CIO, EA, and Data Team to continue the investigation the challenges that resulted in the current failure of the BI initiative. Samer realized how ambitious his expectations were and how he is now disillusioned and free-falling towards reality. Nevertheless, Samer and the CIO were finding this journey pretty healthy as it gave them deeper insights about the very company they lead and the existing challenges.

3- Governance Challenges

The CIO talked to the Enterprise Architect on whether the data warehouse and business intelligence team is the sole responsible for all this fiasco. But he realized that many major issues are attributed to a higher layer than the operating model, specifically the governance (briefly defined as how the company keeps things in control and how it takes decisions). In an effort he made to identify the governance issues, he pinpointed four major issues.

During the BI implementation, the vendor and data team were struggling to locate the data as the owners of the data were not clear. There was some data that nobody wants to own, on the other hand, there was data that many departments claimed to have ownership over. Furthermore, there were challenges to understand the data in databases, such as the business activities that generate it, and the reason for this. Hence, the second issue is that the business context was missing. The data team had the worst time of their lives during the data collection meetings, as many managers refused to share the data for privacy reasons. Having no data privacy rules or polices in place rendered the discussions with these managers pointless. The fourth issue is that there was a lack in unified definitions of specific terminologies across the company (like customer and product) along with missing universal taxonomies that could make combining insights from different sources possible.

Such issues affected, directly and indirectly, the communication management across the teams and introduced many deficiencies including the lack of proper conflict management and escalation process. This negative impact did not stop here. The impaired data governance affected work culture and created political zones across the teams that disrupted the overall execution progress. This reinforced the CIO's original belief in the existence of functional silos and an ensuing conflicts between the departments in the company.

4- Strategy Challenges

While discussing the findings so far, the CIO informed Samer that he personally finds himself part of the problem because of his limited short-sighted view about data and information. "Information is a business resource that is used in every aspect of a business: it supports the day-to-day operational tasks and activities; it enables the routine administration and management of the business;

and it supports strategic decision making and future planning". Unfortunately, the CIO did not see that coming at the strategic level. He said that while things seemed to be working and achieving gains at the tactical level, an imminent loss at the strategic level was approaching. This is because data and information was not part of business strategy or strategy imperatives, as everyone was in the rush to start the BI implementation. The implications of such deficiency are: having loosely identified enterprise-level data needs, treating data as a by-product of technical solution not as an asset, and the lack of data-related performance metrics at the corporate level.

> ***Information is a business resource that is used in every aspect of a business: it supports the day-to-day operational tasks and activities; it enables the routine administration and management of the business; and it supports strategic decision making and future planning.***
>
> ***--www.whydatamanagement.org***

Section 5

Reached the Bottom, now What?

Nearly every organization is prone to many of the discussed challenges since they are frequent in any data-oriented solutions like: Data Warehouses, Business Intelligence, Business/Data Analytics, and even Market Research activities. In the story above, Samer gained a holistic awareness about the challenges spectrum that the BI implementation faced. Such challenges are very devastating, and some could be showstoppers.

The reader might have noticed that there are clear interdependence and somewhat causal-effect relationship amongst the categories of the listed challenges. Hence, it is important to understand the interconnectedness nature of these issues. The poor data quality, for instance, is partly attributed to serious issues in the operating model and the overall processes of the project, which failed to address such data quality issues. Many of the operating model issues themselves were also the result of other difficulties at a higher dimension, which is the data governance. Following the chain reaction, governance could not do any better with lack of serious answers and directions at the data/information strategy level.

Table A: The Key Challenges and their Equivalent Categories.

Category	Issue / Challenge
Strategy	Data and information are not part of the business strategy.
	Lack of enterprise-level data definitions for common terms like customer, product, ...
	The lack of data-related performance metrics at the corporate level.
	Lack of having a horizontal view over of the data management landscape.
Governance	Lack of clear data ownership.
	The business context of data is missing.
	Lack of company-wide privacy and security policies.
	Lack of master data management activities to solve many issues like: defining acceptable value domains and resolving conflicts in multiple sources.
	Absence of important company-wide committees like: data quality and data stewardship committee.
	Poor communication management, which make key people oblivious to what is happening.
	Toxic work culture that leads to the formation of political zones/silos across the teams that disrupted the overall execution progress.
	Reporting and Collaboration tools used by different users were poorly selected.

Category	Issue / Challenge
Operating Model	Chaotic and unstructured data gathering process.
	Report and dashboard development activities are detached from proper business analysis.
	Operating in an ad-hoc fashion rather than Agile.
	Technical solutions selection and management is poorly done.
Data Quality	Rely exclusively on the tools' automated data quality analysis while excluding the analysts' important contribution of running extended analysis.
	Lack of coordination between the data quality analyst, the data steward, and the business analyst which leaves many aspects not being checked at the data quality level.
	Improper use of tools to detect any change/update in the data quality of the reacquired data (including full snapshot or deltas).
	Lack of proper measures (KPIs).
	Not being aware of the scope of DQ issues and techniques.

Characteristics of the Executives at This Stage

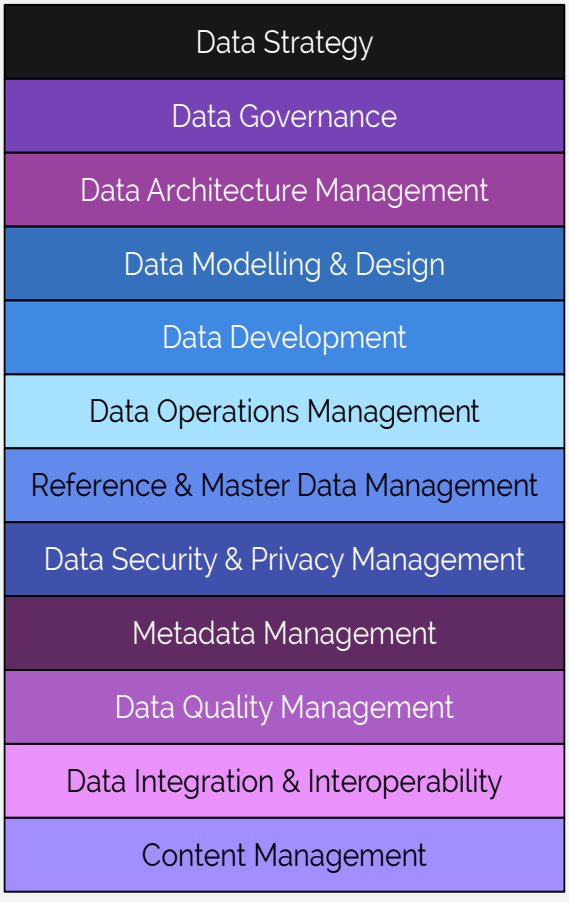
People at this stage, which Gartner calls "trough of disillusionment", like Samer tend to live in denial at the beginning. They seek confirmation to their assumptions from the others surrounding them. Misleading diagnosis to what is happening, makes things much worse. Accepting the fact and moving forward crawls slowly into the picture but with fierce resistance towards investing in additional technologies. Some people might rely on their executives and teams to solve the problems and make the change, whereas others realize that solving such enterprise-wide problems requires well-studied plan, and hence seek external consultants/ teams to provide such plans. Each approach has its own advantages and disadvantages.

Section 6

Hope and Enlightenment (What Elm is Doing, Planning to Do?)

The difficulties that Samer's company faced with its BI implementation and data analytics initiative are very similar to the ones encountered by Elm and many other companies having experience in this kind of projects. The good news is that Elm has dealt with most of these challenges in its diverse projects mainly within the governmental sector. For the few remaining challenges, Elm knows where to start. Generally speaking, Data and Business Analytics projects are quite new in the Saudi market. Here in Elm we are learning, as many other companies in the market are doing. Hence, it is possible to say that most of these challenges presented in the paper demonstrate the realities and difficulties existing in the Saudi market. However, this does not imply that other global markets are immune to such challenges, they do exist there but in different settings and levels.

The importance of this section is that it discusses Elm's findings and general answers to these challenges. This would represent the needed hope for Samer and any analytics troubled-soul to move to the next level, i.e. the transition from the trough of disillusionment towards the slope of enlightenment. It is important to note that the paper is not providing detailed solution to every challenge because we believe that solutions must be tailored to match each company's situation, which is not possible in the paper scope. Rather, the paper provides general solutions and practices following top-down order, i.e. from strategy, governance, operating model, to data quality. While it is always preferred to have top-down sequence when resolving the challenges, we know that it is not always possible. As a result, we decided to keep some challenges that would have been eliminated in response to higher-level resolutions



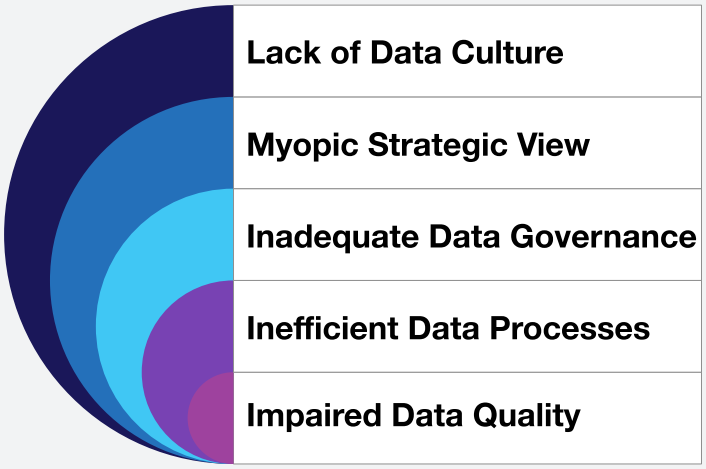
Elm Data Management Functions

visible so that the environments where a bottom-up approach is better would still consider them. We believe also that by just opening the reader's eye to specific challenges, the paper would be fulfilling its objective.

Data Culture: A Desirable Prerequisite

As illustrated in Table A (section 5), the BI and analytics challenges fall into the following categories: Strategy level, Governance level, Operating model level, and Data Quality level. Although the last level could be considered as part of its preceding level, it is better to keep it separate to emphasize its importance. Solutions to these challenges could be implemented in both directions: top-down and bottom-up. However, such enforcement requires a powerful motivation to keep it intact. Elm believes that to have an everlasting motivation, strong data culture must be in place. Change management programs and hiring the right people are highly needed to evangelize this culture in addition to introducing specific changes to the organization structure to support this. Listed below, are number of notable company-wide data culture practices that help in building the motivation needed to drive the top-down and bottom-up approaches to face the data challenges.

- Building the habit of reviewing daily reports and dashboards as well as reading periodical and annual studies related to the company business.
- Nurturing the practice of always making data-driven decisions even for the small tasks.
- Activating Data Democratization: data accessibility for everyone.
- Performing data and numbers standup meeting.
- Integrating dynamicity and resilience in the work culture.
- Preaching for being practical and reasonable by avoiding planning for the happy scenarios always while ignoring the realities and risks.
- Understanding that data analytics projects progress in phases. So, one should not think of them in a big-bang approach.
- Developing an attitude against dichotomous thinking. So, one shouldn't completely ignore what cannot be completely available in hands or totally unambiguous.
- Adjusting the application-centric mentality, most people have, to give further consideration and appreciation to the data entities.
- Developing KPIs that assess data entry, data quality... and many activities applied on the data to measure its health.



Interconnected relationship between key Data Management Aspects

Self Assessment: Why is Data Culture Important?

How beneficial is it to make decisions without being data-informed?

Ask yourself and your team the following questions and compare everyone's answers:

What is the total number of customers the company has? And does it correlate with the company profit in 2014?	How many KPIs the company measures to understand the overall performance?	How many reports does each team member view upon making decisions?
Does he/she think that different versions of the report exist with different numbers with someone else?	Do you trust the numbers you see in the reports?	Are you able to describe the exact problems that harm a specific report's data quality?
Do you think data should be treated as an asset to the company?	Do you feel frustrated every time you ask for a report because it takes so much time to get prepared?	Do you become reluctant when thinking of asking another department to crosscheck your numbers with theirs?

The answers will give you some insights on how key company employees value the data. This is a quick test that just scratches the surface. There are many other data culture health checks and analytics maturity assessments that could be useful to better understand the current situation (check the Additional Reading section at the end of the white paper).

If data culture ever grows to become integral part of the overall company culture, then the company will most likely witness a very flourishing future.

Proposed Solutions: Strategy Level

Recalling Table A (in blue pages), there were four BI and analytics challenges at the strategy level:

Data and information are not part of the business strategy

Placing data and information gathering, generation, and consumption as part of the business strategy would definitely cascade its importance to the lower layers at the execution level, which will enhance all the data activities. Furthermore, the business strategy must reflect the data/information required to achieve the strategy. That is, the business strategy must highlight the information that would enable the strategy. Moreover, this will make data/information monetization very possible and a promisingly profitable new line of business could be created. It is important to note that, as stated before, these challenges are interconnected. Hence, tackling one of them affects the other and they collectively contribute to enhance the overall data strategy. As Data and/or Information are not part of the overall business strategy, it would be very normal not to have enterprise-level data needs. Defining data needs should be done after extensive topdown analysis for the business strategy (drivers, directives, imperatives, goals, etc.) as well as bottom-up analysis for the data entities.

Lacking strategic view about data can be attributed to that traditional mental model about data by treating it as a by-product of technical solution not as an asset. This is due to having an application-centric mentality where data is just a by-product. Instead, data should be treated like the company's capital, equipment/technologies, and people.

Lack of enterprise-level data definitions for common terms like customer, product, …

Having unified definitions for common terminologies across the company is not an easy task. Actually, realizing this fact at a belated stage in the project is tough and expensive. Solving this challenge requires elaborate due-diligence where the key people in the company mitigate any definition conflicts at the strategy level, while maintaining their business function-specific definitions relevant to their units. This also includes defining Business Glossaries or/ and data dictionaries. Eventually, agreements on definitions have to be streamlined all the way to the technical data systems. As a best practice, usually these definitions would be implemented in the metadata repository as business metadata.

The lack of data-related performance metrics at the corporate level

One of the major challenges is the absence of data-related performance metrics at the corporate level to assess the health of data as an asset. These metrics would be defined in the metadata repository as operational metadata, which is usually overlooked by most organizations. The CEO and company executives should always be aware of how their valuable resource, data, is performing across different activities, e.g. data quality Completeness, Timeliness, and Consistency.

Lack of having a horizontal view over of the data management landscape (thus the full-picture is missing)

Trying to solve data issues, quite often most companies fall into the trap of adopting very shortsighted one-dimensional solutions which tend usually to make things worse in the long term. Most companies will try insular solutions, like deploying new technology stack, hiring new staff/ vendors, focusing solely on one data function (e.g. data quality). Although these remedies might be helpful in the short-term, they would rarely yield long-term benefits. Therefore, adopting a holistic, comprehensive, end-to-end framework that reveal the full sophisticated landscape of data management within the organization is of a paramount importance. Such a tool will empower the organization to assess the current maturity of data management practices, plan the target maturity, reveal the gaps, smooth the communication, and objectively evaluate the technical solutions and designs, to mention a few (check the figure in the next page that demonstrates Elm's Data Management Framework).

Addressing these challenges at the strategy level directly enhances the subsequent activities and drastically increases the success chances of any BI and analytics project.

Proposed Solutions: Governance Level

At the governance level, there are eight BI and analytics major challenges:

Lack of clear data ownership

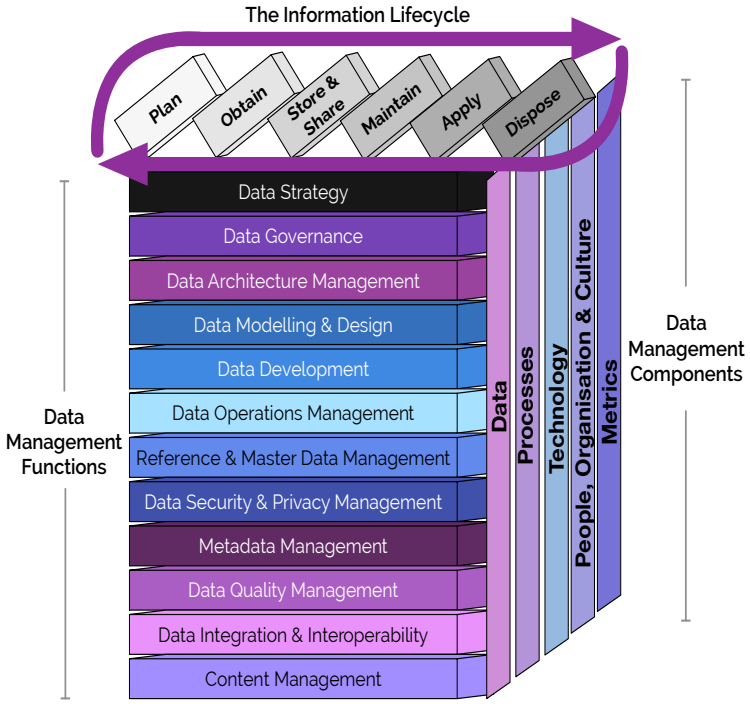
Solving this challenge would require initiating a company-wide program in which data sources/ systems are assigned to owners. The ownership implies managing and being fully charged of maintaining a specific data source on both the technical and business sides (e.g. acceptable range of values entered, modifications on data, business interpretation etc.). Failing to manage the sources would result in being held responsible for any consequences. The information of the data owners and stewards must be published and shared transparently with all relevant parties. This information is subtype of operational metadata. Data ownership program must include conflict resolution and orchestration techniques to assure effectiveness along with data stewardship program to define the actors who can provide support to the data owners.

Having no clear data ownership will naturally result in disparate and shadowy data systems (e.g. Lack of SVOT system Single Version of Truth), which is a clear manifestation of the functional silos problem. The lack of proper as-is assessment before starting such data initiatives is common, unfortunately, and it has severe impact on the operating model. As a best practice and change management tactic, the source of data must be exposed as a technical metadata, using the proper tool (e.g. the metadata repository)

The business context of data is missing

Moving to the second challenge, it is essential that people dealing with data should develop business mindset when working with data to understand the business context around it. Generally speaking, BI and data Analytics projects span the IT and business teams. Teams from both sides have to step closer to each other, rather than constantly waiting the other party to exclusively make the move. Hence, developing the business mindset by data team should operate in conjunction with actions made by

> *There is a serious importance for adopting holistic, comprehensive, and end-to-end framework when tackling data challenges. The framework can be used for multiple purposes like Technical Evaluation, and Solutions Design.*



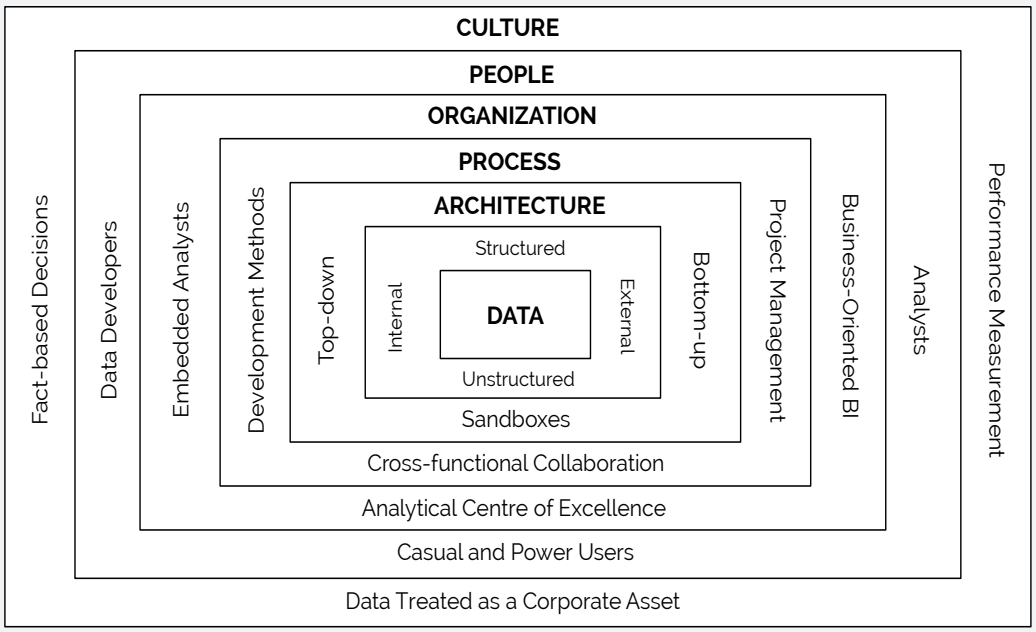
Elm Data Management Framework

the business team to understand the challenges that the data team faces at the first place. Developing the business mindset should also be performed in the lights of an adopted Analytical Framework. The figure below shows the Analytical Framework that Elm recommends. With proper nurturing of the data culture, this process would take place smoothly.

Beside developing the business mindset, it is extremely important to capture and document the business context as business metadata (e.g. business names and definitions of subject and concept areas, entities, and attributes; attribute data types and other attribute properties) and operational metadata (e.g. Business drivers / goals; Data CRUD rules; Data owners; Data sharing rules and agreements / contracts; Data stewards, roles and responsibilities; etc.). This challenge might require capability enhancements through training to individuals equipped with the suitable mental tools and attitudes. Another important aspect to solve such challenge is to recruit the right people for BI and data analytics projects. An important insight we arrived to is to avoid using traditional ways of hiring for the capabilities needed for such projects. Assessment days, competitions, challenging assignments, group activities, and many other recruitment techniques should replace the traditional ones. These techniques focus on measuring the problem solving, analytical thinking, critical thinking, conceptual thinking, and teamwork skills, which are highly needed among the team working on these projects.

Lack of company-wide privacy and security policies

Despite their importance, security and privacy policies are usually overlooked in the Saudi market. Data is used in an unrestrictive way as long as a higher authority in the company takes the responsibility for allowing this. Only when data gets leaked or any blame is received, the policies are brought back into the picture. This is beside a major mental obstacle that is viewing the data as IT stuff, not as an asset to be managed and protected by



The Analytical Framework Recommended by Elm

(Adapted from "Secrets of Analytical Leaders" by Wayne Eckerson "Technics Publications 2012")

the proper polices and processes. It is not an easy job to suggest simple solution for this challenge, since it involves many layers that are specific to the company settings. However, a foundation ground of privacy and security practices covering different states of data (data in motion, data in use, and data at rest) must exist. The practices could be very high level at the beginning but should evolve with time to cover more details. The important thing is that they have to exist at least. Later, enforcing them at the operating model level is as much as important. It is understandable how people usually react to enforcing such policies by considering them as an unnecessary burden. This gets amplified during the busiest day of the projects. But by thinking about the consequences of an accidental data leak, people just hope that such leek does not come, and that they would receive minimum blame if it did. The practical approach to handle such challenge is to roll out the policies in a gradual fashion starting with the easier ones that have large impact.

Lack of master data management activities to solve many issues like: defining acceptable value domains and resolving conflicts in multiple sources

Master Data Management (MDM) is a suit of activities applied on the data in order to reach single point of reference. Such activities could be prioritized in a roadmap to start with the most important ones having larger impact on the data. Usually, master data management is not given much care in BI and analytics projects here in Saudi Arabia because many IT and business professionals are not aware of its role. Such a role is, most of the time, understood when specific consequences start to emerge, i.e. at a late time in the project. Couple of these consequences are: 1) when data cleansing activities commence and they lack proper cleansing rules across the disparate data systems in the company, 2) when reports linking to multiple sources have inconsistent values as they are not properly merged. It is important to note that some of the MDM activities could be done manually especially at the beginning. Afterwards, when a maturity level is reached, MDM technical tools could be considered. Part of MDM activities is to define the acceptable value domains for data objects. This is much more controlled if it stayed within the company. However, as we saw in Samer's story, some reports were delivered to the ministries and it was important to present the information in an understandable and actionable way to them. For example, food categories/subcategories in the reports have to comply with the standard categories followed by the ministry. This includes many other standards like job title classifications, addresses, industrial sectors classification, and more. As a best practice, usually these taxonomies would be implemented in the metadata repository as business metadata

Absence of important company-wide committees like: data quality and data stewardship committee

In Saudi Arabia, steering and executive committees have a negative reputation. Usually, the committees are looked at as the physical manifestation of bureaucracy. But it is important to understand that there are many decisions related to data that have to be taken by such committees, hence, their role has to be activated. For the BI and data analytics projects, data quality and data stewardship committees are among the most important committees to form. Such committees have to have the right people, decision-making methodology, and authority. Similar to data ownership, the information of the data governance bodies and stewards must be published and shared transparently with all relevant parties. This information is subtype of operational metadata.

Poor communication management, which makes key people oblivious to what is happening

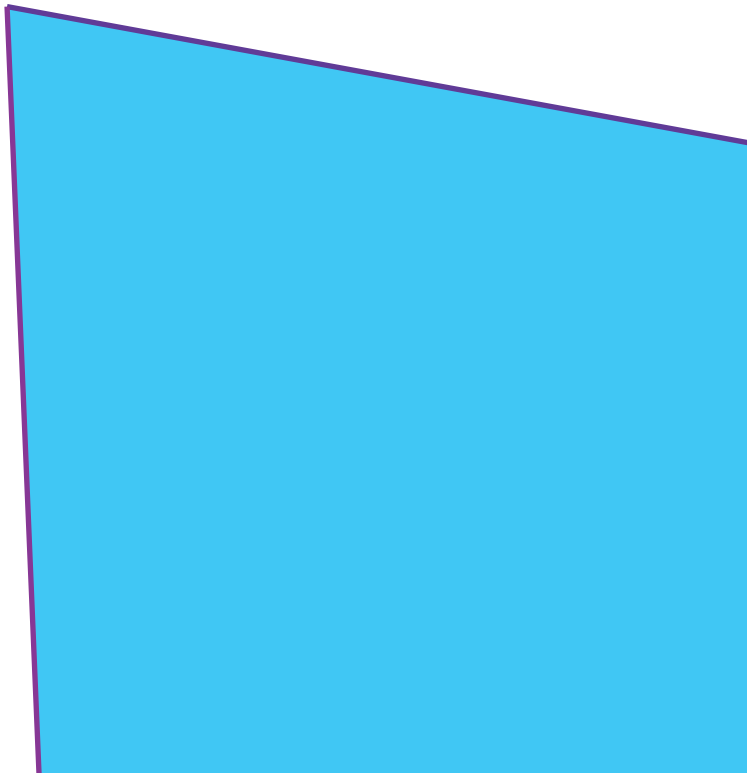
As discussed early, BI and data analytics projects span the IT and business teams. This results in potential deadlocks, bottlenecks, and black holes in terms of communication if not managed properly. Such setup might truly separate the management from what is happening between the teams, and disallow the management to act effectively. While improving communication activities in companies differs from one another, embracing the data culture would help a lot in this regard. Also, there are many techniques to improve communication management in the company, but the recommendation is to introduce specific tweaks to them to gain extra effectiveness for BI and data analytics projects. As a critical part of managing the communication well is to proper conflict management and escalation process in place, due to the nature of such projects where high collaboration between the IT and business teams is needed. These processes must be clear and transparent for all relevant users, thus, they would be defined in the metadata repository as operational metadata, which is usually overlooked by most organizations.

Toxic work culture that leads to the formation of political zones/silos across the teams that disrupted the overall execution progress

The word toxic is relative and might sound a bit harsh, but the nature of BI project magnifies the unfortunate realities of such culture. Neutralization of this challenge would likely to exist as a result of applied solutions at the strategy level including an ideal utilization of data culture. Nevertheless, the strategic solution must never ignore the execution details at the governance level. In such problematic environments, majority of buy-ins are verbal and usually given for political gains. Since BI projects involve many departments, with the existence of such work environment execution becomes a real challenge. One normal symptom of such a culture is to witness excessive resistance by middle managers to collaborate and facing difficulties in getting the data from them. the excessive resistance by middle managers to collaborate is very alarming challenge. The company has to understand the motives that each manager has and gain his support, while offering manageable compromises. The latter statement might seem absurd relative to global work standards, but in the local market, politics and personal relations play major role in driving the work dynamics. So in order to avoid any setback caused by such reality, pragmatic thinking is the best to be offered in order to make the project a big success.

Reporting and Collaboration tools used by different users were poorly selected

Tools selection at this level must be performed while considering different factors and biases that each team has. The traditional trade-offs (usability, cost…) should be balanced in a way that assures optimal utilization of the tools by different users once acquired. All the teams should be open to discuss many tools' advantages/disadvantages and assess how each tool fit to what they want and need.



Proposed Solutions: Operating Model Level

Recalling Table A, the following shows BI and analytics challenges at the operating model level:

Chaotic and unstructured data gathering process

BI and data analytics project taps into a wide range of data sources within the company or outside, e.g. social media and competitors reports. Collecting such data must be done carefully with the goal of making the collection process as structured as possible to ease the data integration/consolidation. It may sound obvious, but data gathering output is what the reports and the project output eventually, to a great extent, rely on. The results of lacking the proper attention at this level gets amplified at the report level. Handling this challenge could be done by developing data sharing agreements supported by data feeds documents or SLA across the company departments. This solution assumes that positive actions have been made at the governance level to facilitate it. Data collection becomes more challenging when external agencies are required to provide their data. With this setup, the project governance must be robust enough to handle the extra flavor of challenges coming from outside.

Report and dashboard development activities are detached from proper business analysis

This challenge would be truly facilitated if specific actions were made at a higher level (governance and strategy). However, at some cases this is not possible. In such cases, it is important to include in the operating model a robust process that continuously mix business analysts with the data team while creating the reports. BI name implies that business is a crucial part in the reports and dashboards. However, it not uncommon to find BI reports that just show data aggregates lacking the business context that will not help the decision makers in taking better business decisions. This is a challenge that must be addressed efficiently. It is not uncommon for some key people in the company not to know what they need, so their requirements would be ambiguous. This challenge could be addressed with proper data culture in place along with being aware of the strategic goals that the BI and data analytics projects seek to achieve. However, in the environments that lack such a setup, the BI project team should guide, evangelize, and help report requesters in understanding the value of this project and what should be expected from their side.

Operating in an ad-hoc fashion rather than Agile

Some employees hide their relative lack of competency by blaming the standard methodologies. Operating under Agile methodology does not imply: less planning, lack of strategic and tactical thinking, over utilizing human resources with redundant and unimportant tasks, doing zero documentation, and working under firefighting mode with no much thoughts on establishing effective processes. This is adhocism at best. Companies that aim to start BI and data analytics projects have to assure how the projects will be developed by the execution arms given they are using the right methodologies. Saying "we do Agile" while practicing adhocism is a hollow slogan and should not be relied on. Even if we want to work using agile methodologies, one has to realize that Agile concepts are not directly suitable for Business Intelligence and Data Integration projects. Some introduced notable modifications like: adding non-standard sprints, additional managing team roles beside the scrum master and product owner, and utilizing pipelined delivery techniques. Moreover, some articles have recently argued that Agility accustomed us to work and think in a task-oriented fashion, but BI and data analytics projects should follow experiment-oriented approach. So, the challenge related to applying generic Agile in BI and data analytics projects should be seriously considered by the companies.

Technical solutions selection and management is poorly done

Understanding the real business need and then choosing the right technological enabler is very crucial. The tool selection criteria have to be rich enough to comprehend many factors in order for the tool to be very suitable as an enabler and make the financial investment worthy. This includes comprehensive understanding about the realities of the market like: tools availability, local success stories, resources availability, direct suitability to business goals, etc. In addition to this, using published studies like Gartner's Magic Quadrant is useful in the selection process, but one has to be careful to avoid complete reliance on such studies when making the decision. Choosing the experienced vendor is not the only important thing. The hosting company, i.e. the client, must be well prepared to embrace the vendor and facilitate its work on the project especially when defining and integrating mutual processes, roles and responsibilities, planning, etc. Also, to gather requirements, the vendor should be as isolated as possible from any considerations not related to the project, e.g. internal politics.

Improper deployment of privacy and security enforcements (including data masking)

No matter how well defined the security and privacy policies are at the governance level, enforcing them at the operating model level is more important. It is understandable how people usually react to enforcing such policies by considering them as an unnecessary burden. This gets amplified during the busiest day of the projects. But by thinking about the consequences of an accidental data leak, people just hope that such leek does not come, and that they would receive minimum blame if it did. The practical approach to handle such challenge is to roll out the policies in a gradual fashion starting with the easier ones that have large impact.

Proposed Solutions: Data Quality Level

Recalling Table A (in blue pages), the following shows the BI and analytics challenges at the data quality levels:

Rely exclusively on the tools' automated data quality analysis while excluding the analysts' important contribution of running extended analysis

Data profiling and quality assessment tools apply many types of analysis and checks in an automated way. However, they have their own limitations. There are many types of data quality analysis that require human-based investigation. This is particularly important when checking whether the data violates any documented or undocumented business rules. The data quality analyst has to be driven by curiosity to squeeze the data and discover any quality issues. Besides, the data quality process must be designed in a way that offers multiple feedback loops coming from different activities and teams related to data.

Lack of coordination between the data quality analyst, the data steward, and the business analyst which leaves many aspects not being checked at the data quality level

The role of data steward is usually overlooked, despite its importance. The lack of having a person who is fully aware of the business terminologies, standards, rules, metadata, logical model, and content of a data source has a negative impact on the work of both the team and the report consumers. For BI and data analytics projects, such a role should exist because of the essential role it will play with the data quality analyst. In case it exists, proper coordination between the interleaving roles should be considered.

Improper use of tools to detect any change/update in the data quality of the reacquired data (including full snapshot or deltas)

With the existence of human factor at the data entry level along with constant changes in some business aspects of a company, data quality would always face challenges. The tools and processes deployed for maintaining data quality must be configured to detect any change/update in the data quality of the periodically acquired data (including full snapshot or deltas). This could be extended also to have dimension and fact tables in the data warehouse to maintain the data quality changes. This is important to keep track of the quality of data.

Lack of proper measures (KPIs)

Data quality must be quantified and measured in order to make proper estimations about the deliverables accuracy, i.e. BI reports accuracy. Moreover, with being able to measure it, the data cleansing and quality enhancement efforts would be evaluated. Almost all data quality tools come with a set of KPIs to measure the data quality score. Depending on the nature of available data and the level of depth required to measure the data quality, sometimes such KPIs are not sufficient. As a result, it is important to define the set of needed KPIs to monitor the progress of the data quality enhancements.

Not being aware of the scope of data quality issues and techniques

Data quality analysis covers many techniques and cases, and it is not as simple as many would assume. It is always advised to understand scope of analysis needed to achieve best results, because this analysis requires resources and time. Prior to launching the BI project, and during the as-is phase, the data quality enhancements scope must be defined. Check the end of the paper for additional resources on this.

Conclusion

BI and data analytics projects generally span the IT and business teams, and this creates a unique set of challenges. These challenges were grouped and discussed in this white paper according to the levels in which they exist, i.e. at the strategy, governance, operating model, and data quality levels. Special emphasis was given to the challenges resulting from the lack of effective data culture and proper data ownership programs in such projects. This white paper discussed many approaches to handle such challenges and demonstrated how essential it is to treat data as an asset while adopting a holistic, comprehensive, and end-to-end data management framework.

To practically actualize a data management practice in your organization, It is recommended to check out the below high-level rough plan (please do not take the milestones literally, as they depend heavily on a myriad of factors. This list is not intended to be an exhaustive one):

- [0 - 30 Days]: Complete a maturity assessment, Develop the data strategy, and Develop high-level program plan.
- [30 - 60 Days]: Identify key initiatives (part of data strategy), Establish the initial Data Governance (including organization roles and responsibilities with stewardship roles), and Develop the high-level Data Architecture.
- [60 to 90-100 Days]: Develop the detailed implementation roadmap (part of data strategy), Identify and recruit team leaders, and Identify information.

Additional Resources:

You may directly contact us if you are interested in getting the following additional resources:

- A detailed list of Data Quality KPIs
- Comprehensive Data Quality Assessment techniques
- Glossary of the key terms in BI and Data Analytics

References

- DAMA International. The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK). February 22, 2010. Technics Publications.
- Method for an Integrated Knowledge Environment MIKE2.0, http://mike2.openmethodology.org
- Danette McGilvray, Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information (1 edition). July 25, 2008. Morgan Kaufmann.
- John Ladley, Making Enterprise Information Management (EIM) Work for Business. July 3, 2010. Morgan Kaufmann.
- CMMI Institute. Data Management Maturity (DMM) Model V1.0. August, 2014. CMMI Institute.
- Frank Buytendijk, Ted Friedman, Mark A. Beyer. Introduction to Gartner's Information Capabilities Framework. 6 June 2014. Gartner.
- Robert Laberge. The Data Warehouse Mentor: Practical Data Warehouse and Business Intelligence Insights. May 12th 2011. McGraw-Hill/Osborne Media.
- Steve Hoberman, Donna Burbank and Chris Bradley. Data Modeling for the Business: A Handbook for Aligning the Business with IT using High-Level Data Models, First Edition. 2009. Technics Publications.
- David Newman, Debra Logan. First 100 Days: Enterprise Information Management Initiatives. December 4th 2009. Gartner.

TAMING THE DATA ELEPHANT!

DATA CONSULTATION GROUP 2016

WAEL ALALWANI
walalwani@elm.sa

ZAHER ALHAJ HUSSEIN
zhussein@elm.sa

8191 شارع التخصصي العليا
الرياض 12333 - 3038
المملكة العربية السعودية

8191 Takhassusi Road
Olaya
Riyadh 12333 - 3038
Saudi Arabia

Tel + 966 (0) 11 288 7444
Fax + 966 (0) 11 288 7555

920029200
www.elm.sa