Palantir

Why Business Intelligence Fails (when it does)

As data grows exponentially, organizations are expecting more from their business intelligence solutions.

Why Business Intelligence Fails (when it does)

Overview –

As the complexity and volume of enterprise data grows exponentially, organizations are (rightfully) expecting more from their business intelligence (BI) solutions. The global BI and analytics market will grow to $22.8 billion by the end of 2020 according to Gartner. Unfortunately for those organizations, 60% of big data projects will fail[1]. Why?

Palantir works with organizations that are at different stages of their data-driven transformations. Through our engagements, we get a chance to observe our customers' historical challenges. Most organizations find out too late how their earlier strategic decisions and platform choices result in an inflexible data landscape. Such decisions result in failure modes that look similar across various industries. These failure modes are tightly interdependent — they cascade from top-level decisions to broken collaboration models, inflexible workflows, intractable data integrity, and ultimately, a data foundation that does not scale with the organization's growth.

The following categories capture areas where business intelligence *may* go wrong:

1. Focusing on the immediate cost of implementation more than the long-term cost of ownership
2. Inability to manage data and business logic in tandem
3. Insufficient rigor in tackling enterprise-wide data integrity
4. Lack of collaboration from the ground up
5. Inability to enrich the data foundation with tribal knowledge
6. Difficulty making a leap from rigid data assets to a scalable ontology
7. Difficulty making a leap from "pretty dashboards" to flexible workflows

[1] Gartner - Newsroom / Gartner - BI Analytics

Failure Mode 1 – Focusing on the immediate cost of implementation more than the long-term cost of ownership

The data transformation journey starts with one or more strategic decisions that set the course for a long-term roadmap. Typical approaches to addressing business intelligence pain include building end-to-end home-grown solutions or asking systems integrators to patch together an assortment of tools with the goal of optimizing costs. However, focusing solely on short-term costs frequently undermines the ability to establish a flexible and well-architected solution for the enterprise in the long-term. Specifically, a crowded architecture with minimal connective tissue amongst its constituents is very likely to result in integration and scalability challenges. Such challenges end up introducing unforeseen future costs (including labor costs, integration costs and / or costs of new platforms) that drive up the total cost of ownership (TCO) and make the original "cost-effective" decision moot.

When considering business intelligence solutions, it is imperative that organizations understand how they can scale the following:

SPEED:	The speed with which they can deliver new and incremental data projects without sacrificing data integrity
BREADTH:	The breadth of the underlying data foundation without turning it into a data swamp
PRO-USERS:	The number of data consumers and analysts in the organization who are able to make data analytics operational without getting help from "technical" counterparts

Ultimately, this boils down to prioritizing long-term value and scale of business intelligence investments over one-off wins that are isolated across the enterprise. It is critical to understand that scale comes from building a healthy data foundation which involves back-end decisions made around how data is managed, as well as front-end decisions around how it is analyzed and visualized. The failure modes mentioned below almost always originate from treating the back end and the front end as architectural silos and not getting the fundamentals right from the ground up. One such fundamental area is managing data and business logic in tandem.

Failure Mode 2 – Inability to manage data and business logic in tandem

A data ecosystem consists of data assets and code that integrates, transforms and models those data assets. The code represents key business logic—the know-how that explains how an enterprise works. As business logic evolves in an organization (as it does very rapidly), new data transformations and integrations are required, resulting in new data assets. It turns out that the ability to manage business logic alongside data at scale is the most difficult challenge organizations face today.

When code management is kept as a separate concern from data management, the value and power of the data in the organization is immensely reduced. Specifically, first-class capabilities such as versioning, lineage tracking, authoring and granular access controls need to be applied to data and code in tandem. Without such capabilities, it is not possible to tackle data integrity across the organization (Failure Mode 3), enable true collaboration across the enterprise (Failure Mode 4) and unlock a growing data foundation (Failure Mode 5).

In practice, most organizations ultimately get stuck making a choice between a rigid data warehouse which gives the data community limited ability to shape the underlying structure of the data or a free-flowing data lake that is open for access but does not retain any transparency into the lineage of the data or the logic. Our customers are increasingly coming to the realization that an ideal world not only allows **flexible access to all relevant data assets but also empowers the data community to understand and manage the business logic that generates them.** This is particularly relevant in the context of data integrity - the key pillar of a healthy data ecosystem.

Failure Mode 3 – Insufficient rigor in tackling enterprise-wide data integrity

In a recent KPMG survey across over 2000 global senior executives, only 35% of participants said they have a high level of trust in the way their organization uses data[2]. Clearly, business intelligence is only as good as the data that it can surface to the business community. Still, many BI projects focus on the "shinier" parts of the BI universe, such as data visualization, and treat back-end data integrity and data quality management as an after-thought.

The reality is that issues with data quality will never cease to exist. What's more important is how these issues are handled on an ongoing basis:

- Can business users flag data quality issues on the analysis platforms as soon as they encounter them?
- Can data engineers address the issues flagged by users in near real time?
- Are users and engineers able to dive into the lineage of the data to understand where quality issues originate?
- Can business users trace back the business logic that generates contentious metrics?
- Can new data transformations be tested in a sandbox without posing a threat to the data integrity in production systems.

[2] KPMG

Business intelligence solutions that do not have *kaizen*, or continuous improvement through small and positive changes to the system, will struggle with data integrity. Based on our experiences with various customers operating at enterprise scale, we have built conviction that such continuous improvements to the data foundation are what make a data ecosystem sustainable in the long term. Such an approach invariably requires a holistic view of the data lifecycle extending from back-end data and logic management to front-end tools and workflows.

Failure Mode 4 – Lack of collaboration from the ground up

The problem with many BI projects is that the business plays a marginal role in the way data-driven workflows are built. While business representatives may be present on joint project teams, they rarely have access to a collaborative platform through which they can truly iterate on both the data assets and the key workflows. When business users are not first-class participants in building a data foundation and the associated workflows, they stop using it. As a result, it is not uncommon to see more IT staff dedicated to building an end-to-end solution than to using it regularly.

Specifically, such lack of collaboration manifests itself as a failure mode when:

- BI platforms do not allow **business users and IT engineers to collaborate on enhancing the workflows they should be jointly building**. For example, projects do not advance in the right direction when there is no explicit feedback loop built into data platforms for business users to flag issues, test alternative hypothesis, or confirm existing business logic.
- Business users are unable to access the same back-end data foundation that is available to IT for purposes of **tracking, exploring, and interrogating the data lineage and building trust in the data**.
- Business users cannot **contribute their subject matter expertise to the data foundation** in the form of new data transformations to accelerate the pace of collaboration. Specifically, this is about subject matter experts having the ability to create new data assets, which takes us to Failure Mode 5.

Failure Mode 5 – Inability to enrich the data foundation with tribal knowledge

In the big data universe, there is a lot of chatter about "data silos" but an equally pervasive challenge is the hidden silos of tribal knowledge. In fact, many companies still operate within a spreadsheet culture in which users extract data from internal systems, load it to spreadsheets and perform their own calculations without sharing them company-wide. According to a recent DataWatch survey[3], 89% of respondents state that they input data into spreadsheets for analysis but 81% are concerned about data quality and consistency when using such methods. The result is the existence of competing frames of reference, or sources of truth, causing confusion and at times even risking data security due to key data assets held locally by individuals on their laptops.

[3] Barchart

The ability to address such fragmentation is largely undermined in legacy BI solutions. Many BI approaches fail because:

- They do not allow subject matter experts across the data universe to write back their personal "calculations" into a common data layer and encode them as transformations
- Because such contributions are not possible on an individual basis, the data community cannot benefit from the "state of the art" data assets and business logic to do their work
- Because numbers are thrown around the business without the business logic behind them, no one can trust the data (e.g. "what is your definition of revenue?")

Solving for this failure mode involves defining key business metrics transparently in a common data foundation. The metrics can then be enriched across the enterprise several times as long as the organization retains the ability to track each revision along the way. In this sense, the abilities to **write back new data transformations into the data foundation and trace their lineage** are critical components of a sustainable business intelligence architecture.

Failure Mode 6 – Difficulty making the leap from rigid data assets to a scalable ontology

Because most organizations treat BI as a series of discrete "fit for purpose" workflows, they tend to focus on producing data assets that service those workflows directly. When new questions emerge in the business, the data preparation steps that lead to the answers are rebuilt from scratch rather than repurposed from existing projects.

The only way to avoid this inefficiency is to focus on a data foundation composed of reusable building blocks, or "business ontology", that represent the foundational concepts for the organization. Such building blocks accelerate the analytic process significantly and increase the lifetime value of every investment. Organizations that have designed a scalable ontology instead of a set of rigid data assets have thought carefully about the following questions:

- Do all projects and all users have easy access to approved data objects that map to intuitive business concepts?
- Can those objects be reused for new questions, analyses, and projects?
- Can new objects be introduced and seamlessly integrated into the existing ontology?

Failure Mode 7 – Difficulty making the leap from pretty dashboards to flexible workflows

An attractive end product for a BI solution is the colorful and dynamic visualization of an organization's data. Unfortunately, “pretty” visualizations do not always translate into effective workflows. Ultimately, all dashboards decay in value because business intelligence is a moving target.

An existing BI workflow can change continuously throughout its tenure to reflect constant changes in data requirements or business expectations. Business analysts may need multiple capabilities based on the problems they are trying to solve. To understand the shape of the data, analysts often require exploratory workflows that are very flexible and ad-hoc in nature. Once they understand the shape of the data, they typically spend majority of their time in hypothesis testing. At times, analysts will also need niche analytical techniques for specific types of data (e.g. analyzing sensor data requires rich time series techniques).

Ultimately, rapid dashboarding capabilities fall short in all these realms and constrain the capability and creativity of business analysts significantly and unnecessarily. We see that our customers benefit immensely from having a repertoire of tools in their toolbox through which they can build and scale their own workflows. Critical questions to consider avoiding this failure mode are:

- Can analysts move beyond the constraints of any individual dashboard?
- Can analysts ask deep questions of the data without depending entirely on data engineers to adjust the shape of the data?
- Can analysts pick from a number of tools that are best suited to the question they're asking?