Section 2: Week 5: Theory to Practice: Business Intelligence

Nate Bachmeier

TIM-7020: Database and Business Intelligence

January 5th, 2019

North Central University

Theory to Practice: Business Intelligence

Business intelligence(BI) refers to the process of transforming data into knowledge for decision processes. As organizations gain insights into their customer experiences and internal processes, they can make more informed adjustments to dynamic market conditions. While it is generally agreed to be a critical aspect, the implemented solution does not always deliver the expectations because of poor planning, resource management, and usability, among other issues. Successful business intelligence systems cannot be ‘another IT project,’ and instead must be treated as an integral extension of the enterprise management system.

# The Story of Intune

Intune offers an online mobile device management platform with many internal processes powered by BI solutions. However, like many other organizations, their journey to a mature set of data driven-decisions experienced significant turbulence along the way.

## Initial Release

Their first release in 2008, stored all service logs on the cluster nodes where they could not be easily accessed. The formation of a separate siloed business analytics unit allowed the company to track usage metrics and project capacity requirements. Due to a lack of domain expertise, the level of granularity available for their forecasts was limited.

## In-Memory Solution

By 2012 the engineering team’s built custom tooling to reduce the friction associated with accessing service logs, though these refinements merely polished a broken system. The business unit recognized the need to centralize the service logs and requested the operations team to provision an Elasticsearch, Kabana, and Logstash (ELK) cluster. ELK is an open-source in-memory data store that supports full-text searches across unstructured data. The system was expensive to operate at scale, so the business could only afford seven days of retention and an insufficient number of read-only replicas. These measures reduced cost at the expense of adoption as the cluster was sluggish and missing longitudinal reporting. A subset of the team learned the technology and became proficient, though skepticism from middle management prevented broader training investments.

## Big Data Solution

In 2014, the senior leadership acknowledged the adoption failures of ELK and rebooted the business intelligence strategy around the Hadoop ecosystem. The data pipeline begins with a log collection process that centralized the service logs into Azure Blob storage. Next open-source tooling schedules SQL-like queries using Apache Hive, then copies the aggregations into a Microsoft SQL Azure database. A custom intranet portal would render the information into hardcoded charts and tables. This solution had many improvements over its predecessor, such as extended retention periods and the SQL-like interface query interface reduced training complexity. However, adding new charts often required changes across the entire data pipeline, and Hive’s batch processing design added significant delays between test iterations. Despite teams understanding the value, they pushed back against even simple changes as it would take weeks of effort and defer customer-facing features.

## Hot, Warm, and Cold

In 2016, the senior leadership took a step back and looked at the holistic problem of providing business intelligence across the organization. Instead of having a single general-purpose solution, they chose to have three built-for-purpose systems to address hard real-time metrics (hot), soft real-time exploration (warm), and offline batch processing (cold). The hot path targeted a service level agreement (SLA) of 1-second latency in exchange for limitations on the dimensionality of data reporting, which was sufficient for aggregate service health status, not special customer status. Microsoft Kusto, a proprietary in-memory analytics engine, was provisioned to hold thirty days of service logs and given access to enormous pools of cloud resources. This warm path targeted an SLA of 5-seconds, which encouraged ad-hoc exploration and integrated into Microsoft PowerBI for extensive visualization and dashboarding solutions. For longitudinal reporting (cold), teams needed to use the Azure Data Lake to execute SQL-like queries that would take 10s of minutes to complete. Unlike the previous Hadoop ecosystem, extensions for the organization’s default Integrated Development Environments (IDE) provided a more familiar experience.

## Evolving Beyond Tooling

Business intelligence projects are not ‘IT projects’ and need to consider the value-add of the organization. The senior leadership, middle management, and business analytics unit come together to build a scorecard solution that focused on key quality metrics. The executive leadership reviewed the scorecard on a regular cadence, which put pressure on lower management to ensure the report was accurate and timely. For instance, the report identified the top ten customer scenarios and measured the end-to-end experience in terms of reliable, latency, and availability. These measurements came from instrumentation across the production service that emits tuples contain client\_id, timestamp, scenario\_id, service\_name, method\_name, and execution\_type values. The logging framework collects millions of these values every hour and places them into the big data cold storage system. Then scheduled automation performs aggregations to reduce the size and derive secondary metrics, such as how many customers received their expected SLA, and pushed the results into the warm storage solution to further promote ad-hoc exploration.

Anderson (2018) describes how the data-driven culture fundamentally reshaped the architecture of the service. For instance, the DevOps team was able to detect quality regressions and pinpoint the culprit component in a matter of minutes. Engineering teams used greater insight into the runtime behavior to test the various hypothesis, such as the influence of different messaging and serialization technologies. The product management team actively reviews feature usage statistics to determine future investments. Through the servicing of the various business intelligence needs, customers receive high-quality products that are cheaper to operate, allowing the organization to be more competitive.

# Analysis of Intune

The journey of Intune is similar to the anonymized organization Acme Manufacturing. Gonzales and Wareham (2019) state that Acme needed to reboot its business intelligence platform three times over a decade. The first release took two times longer than expected, and the cost was 300% over budget. The second release had insufficient hardware that caused significant delays in replicating data across their global presence. Middle management distrusted data accuracy because the platform was not timely. The third and final version centered around learn business statistics that had broad implications. For instance, insight into ordering cycles reduced the inventory holding period from 140 to 47 days. They also focused on supply chain bottlenecks and drove delivery times from 15 to 5 days.

The DeLone and McLean model propose that business intelligence system adoption is directly proportional to user satisfaction (Gonzales & Wareham, 2018). Information, system, and service quality serve as inputs into the model, with an output of individual impact. This model partially explains the poor acceptance found in both case studies, as the system was underpowered could not efficiently meet the needs of the organization. Another challenge in both case studies comes from not gaining the support of management (Rai, 2002). Consider the scenario with Intune, where the scorecard report to executive leadership forced a cultural shift to maintain the quality data feed.