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 Contoso

Contoso 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, teams needed to use the Azure Data lake to execute SQL-like queries. Unlike the previous Hadoop ecosystem, modern tooling linked familiar Integrated Development Environments (IDE) experiences into the new tooling.