Teaching Real-Time Streaming Analytics with Microsoft Azure and Power BI

Loreen M. Powell
lpowell@bloomu.edu
Innovation, Technology, and Supply Chain Management
Bloomsburg University of Pennsylvania
Bloomsburg, PA 17815, USA

Hayden Wimmer
hwimmer@georgiasouthern.edu
Information Technology
Georgia Southern University
Statesboro, GA 30460, USA

Ekene Nwobodo
Information Technology
Georgia Southern University
Statesboro, GA 30460, USA

ABSTRACT

Business and data analytics is a growing field with many career opportunities. As a result, many information system and technology related degree programs are integrating business and data analytics courses into their curriculum. However, this is a limited amount of teaching resources available on this topic. This paper will add the body of knowledge and teaching resources regarding business and data analytics. Specifically, this paper aims to explain real-time digital data streaming, explore business and data analytic tools used for collecting, organization, and visualizing data, as well as explain possible implementation and usage.

Keywords: Real-Time Data Stream, Business and Data Analytics, Data Science, Power BI,

1. INTRODUCTION

Today, there is an increase in the number of data science courses being offered within information systems and technology related degree programs. Additionally, data science is a vast field with numerous application tools that are open sourced and fee based. The fee based tools often create a barrier for educators to use due to several budgetary

restraints within academic institutions (Wimmer & Powell, 2016). As a result, educators often seek teaching resources to supplement the limited text books exploring business and data analytics with Microsoft Excel or R. However, there is a limited

amount of academic literature based teaching resources available for business and data analytics topics (Wimox et al, 2014; March & Pepard, 2013; Tiong & Pei-Chen, 2015). This paper seeks to add to the academic knowledge base of teaching resources for IS and related educators. This paper provides an overview of real-time digital data streaming (RTDDS) as well as business and data analytic tools including but not limited to Microsoft Azure and Power BI. This work has practical implications for higher education institutions, faculty, and computing degree programs. Implications include an overview of the state of the art and knowledge for practitioners and academics as a teaching resource regarding Microsoft Azure and Power BI that can be used within their business and analytics courses. The remainder of this paper is structured as follows: background/review of the literature, business and data analytics tools, and conclusion.

2. BACKGROUND

Business and data analytics is a growing field that focusses on the process that "effectively transforms a data-rich environment into a decision-smart one" pp. 44 (Liu, 2016). Currently, there is an increasing number of employment opportunities in this field (Smigala, 2014; Henry & Venkatraman, 2015). As such many information systems (IS) and technology related degree programs are developing or should be developing business and data analytic courses within their course offerings (Henry & Venkatraman, 2015).

2.1 Teaching Business and Data Analytics

Today, teaching is different than the past. Many courses are delivered via several different way including, face-to-face, blended and online. Additionally, instructors use a variety of instructional strategies including but, not limited to, learning contracts, small group work, discussions, projects, lecture, collaborative learning, self-directed learning, case studies, mentorship, simulations, and role plays to connect with students and ensure student learning (University of Illinois, 2016). As such, many instructors rely on using shared instructional resources for their courses. However, this is not necessarily the case with teaching business and data analytics from and information systems (IS) and other technology related disciplines.

Currently, there is a limited amount of teaching resources regarding business and data analytics (Wimox et al, 2014; March & Pepard, 2013; Tiong & Pei-Chen, 2015).
This shortage in teaching resources may possible be a result of a growing topic. Tiong and Pei-Chen reported that the number of analytics courses has significantly increased since 2010. Regardless, additional teaching resources are needed for sharing and reusing innovative business analytics teaching practices and resources to increase the core body of knowledge (March & Peppard, 2013; Marjanovic, 2013; Tiog & Pei-Chen, 2015).

2.2 What is real-time digital data streaming?

"A digital data stream (DDS) is a continuous digital encoding and transmission of data describing a related class of events" (Piccoli, Rodriguez, & Watson, 2015). DDS is a primary force of value creation within organizations (Pigni, Piccoli & Watson, 2016).

Digital data as expanded in unbelievable amounts since the first big data center created for U.S. government "to store 742 million tax returns and 175 million sets of fingerprints" (Shim, Koh, Fister, & Seo, 2016). With the advancements of information technology, and the development of the web in 1989, data

availability and processing capability have progressed to the point where all assets can recurrently report every state change. Hence, this is RTDDS.

Today, there are several businesses within various industries that are using real-time data stream to leverage a competitive advantage. *Table 1* provides a sample list and overview of usage from various businesses within different industries using RTDDS.

Table 1: Sample List of Business using RTDDS

Business	Overview of usage
San Francisco Municipal Transportation Agency	Installed magnetometers to detect vehicle occupancy (San Francisco Municipal Transportation Agency, 2014).
Macys	Installed iBeacon devices for location-based digital coupon (Halzach, 2014).
Uber	Uses QCon software for driver and rider applications (Tonse, 2016).
NEDBANK	Uses Insight software for real-time data stream transactions (INETCO Systems, 2016).
Nasdaq	Uses software and twitter to produce StockTwits, real-time data streaming of stocks (Corbett, 2009).

One of the first data tools used to analyze big data was Hadoop. Since then several other application tools have been developed. *Table 1* indicates various software programs and devices developed and used by organizations for specific real-time data processing. However, these tools are more customizable to specific disciplines and expensive. Many academic institutions do not have the money to purchase these applications for usage within their specific discipline. Hence, educators look for alternative ways or teaching resources to expose students to business and data analytic tools.

3. BUSINESS AND DATA ANALYTIC TOOLS

Business and data analytics tools are needed for the collection, organization, and manipulation and visualization of data (Lustig, Dietttric, Johnson & Dzickan, 2010; Evans, 2016). Many academic text books focus on using Microsoft Excel and R as tools for teaching business and data analytics because they are free or already in use by other disciplines. However, as indicated in Table 1 many other software tools exists that have additional or better visualization capabilities. Visualization provides a simple way for linking and communicating data within various levels of a business. In fact, many researchers believe that visualization is one of the most useful components of business and data analytics (Evans, 2016). Students need exposure to additional analytic tools beside Microsoft Excel and R. Teaching resources, such as this paper, offers an academic literature exploration of additional business and data analytic tools such as Microsoft Azure and Power BI.

3.1 Power BI and AZURE

Power BI is a Microsoft business intelligence tool for data visualization. Microsoft integrated all excel features including Power Query, Power Pivot, and PowerView. With Excel 2010, Microsoft began offering "power" add-ins to provide better analytical capabilities inside of excel (PowerPivot, PowerView, PowerQuery and PowerMap). In 2013 Microsoft developed a self-service, stand-alone application, Power BI, to consolidate these "power tools" and produce a competing business analytics application of their own. With all these features integrated, it makes it easy for users to transform, visualize and report data in a single application.

In order to utilize the full potential of Power BI, users must have an Office 365 subscription with the Power BI package. Microsoft offers a preview of the software's capabilities by allowing users to download Power BI Desktop, which is a component of Power BI, free of charge. The purpose of Power BI Desktop is to "visually explore your data through a free-form drag-and- drop canvas, a broad range of modern data visualizations, and an easy-to- use report

authoring experience". Power BI Desktop offers a complete analytics life-cycle beginning with data acquisition and ending with publishing intuitive reports for business analytics. The possibilities for this tools are boundless with this live monitoring feature which can be used for IoT (Internet of things) sensor data monitoring, blood glucose level monitoring, asthma monitoring and lots of fields for real time monitoring.

The diagram below shows a tested experiment for real time data visualization using Microsoft Azure, Visual Studios and Power BI. This is an integration of three platforms to realize real-time visualization. The diagram just shows a summary of the steps taken and procedures taken to visualize real-time tweets in Power BI.

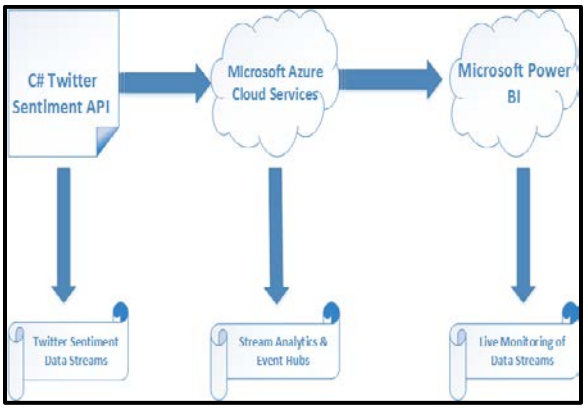

Figure 1 showing the integration of Visual Studios, Azure & Power BI

The twitter API which is written in C# is connected to a twitter account with the use of access tokens which are talked about more in report. Also the API has a section in the C# program where hashtags are put and can be changed whenever the program is stopped. The C# API is connected to Microsoft Azure with the use of a connection String from Azure which is put in the API. The access to the connection string can be seen in the diagram below. This is just one way data can be visualized in real-time using Microsoft Power BI. The great thing about this application is that it's a Microsoft platform which can be connected to any Microsoft platform that can store data. Microsoft has also done a great job of integrating applications like Hadoop, Tableau, Salesforce, SAP HANA etc. which are all integrated in Microsoft Power BI and Azure. Data streams such as log files, sensor data, twitter feeds, transactional data, stock trades etc. can all be visualized with the proper graphs in real-time with Power BI.

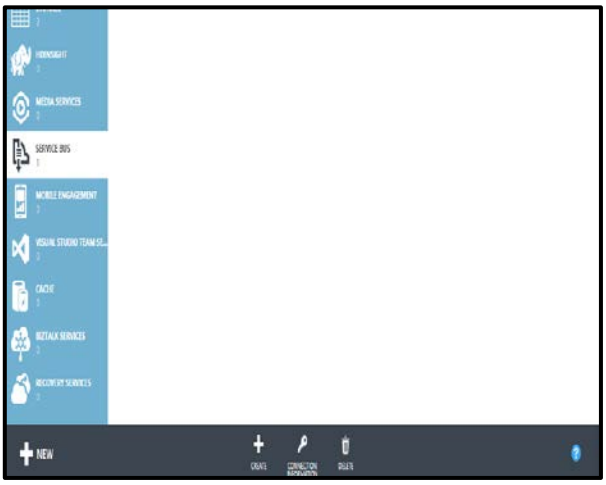

Figure 2: Azure Dashboard

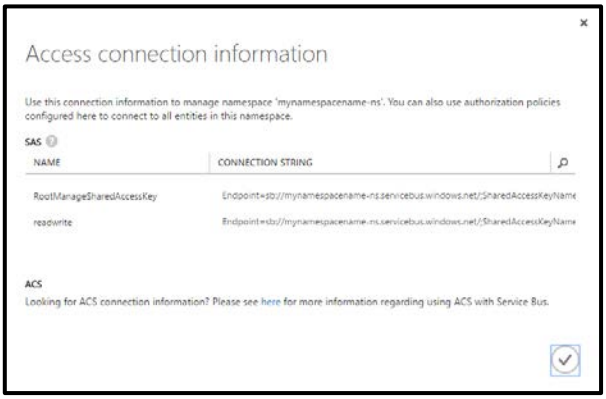

Figure 3 Connection String in AZURE

3.2 Tweet Streams from C# API

This API connects to twitter using twitter's Consumer Key, Consumer Secret Key, Token Key and Token Secret key. Which can be created at https://apps.twitter.com/. It also connected to an Azure account with an Event Hub application which has a connection string and it's put in the API. The API is also queried to shows tweets according to the hashtags added. Whatever hashtag a user

adds to the API, will be the tweets shown in the CMD (command prompt) view. The API shows the timestamp and sentiment score of each tweet but it can be tweaked to show the tweet and user name of the twitter account owner.

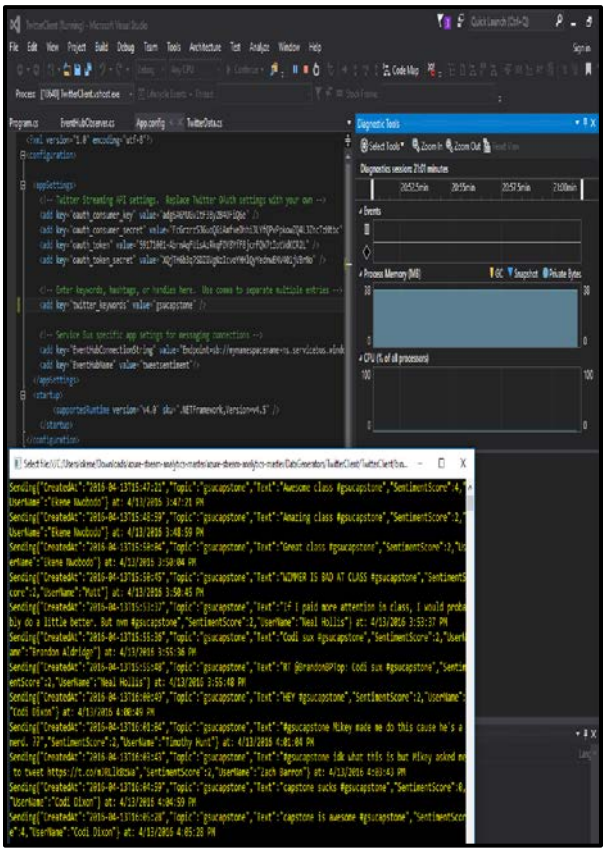

Figure 4: Real-Time Tweets Streaming in C#

3.3 Event Hubs

Microsoft Azure is a cloud computing platform and infrastructure created by Microsoft for building, deploying and managing applications and services through a global network of Microsoft managed datacenters. Azure also has capabilities such as virtual machines, SQL Databases, App Services and Cloud services. The stream analytics application enables data streams compatibility and stores them in an application called Event Hubs. The Event hub takes the real time data and analyzes the amount and size of data that comes in with a live dashboard. In the Event Hub application there's a query section which is used to modify the user's output accordingly. Event hubs do a great job of showing how much data is being stored and shows the flow of data with timestamps on a dashboard. The Event Hub diagram can be seen in detail for further description.

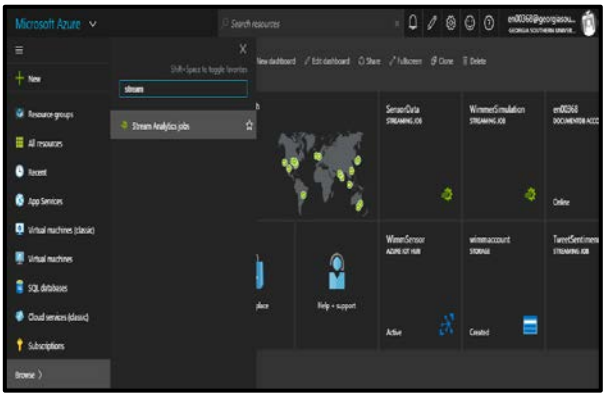

Figure 5: Showing the Stream analytics component in Azure

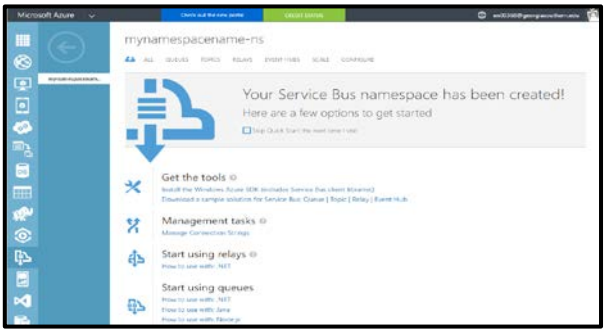

Figure 6 Setting up Event Hubs in AZURE

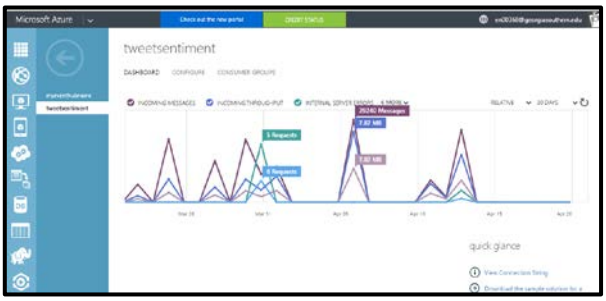

Figure 7 Sample Event Hub for Twitter Sentiment

3.4 Power BI Dashboard

Power BI is a collection of software services, apps and connectors that work together to turn your unrelated sources of data into coherent, visually immersive and interactive insights. This is where visualization is

realized with the help of dashboards, graphs, charts and custom visuals. With the help of Microsoft integrating Power View, Power Pivot, Power Query, Power BI which is cloud based is provides users to access this platform from any given site.

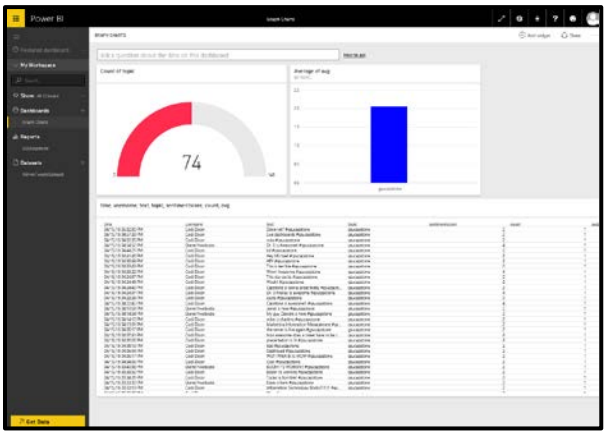

Figure 8 Power BI Dashboard

3.5 Power BI Features

It has dataset section where columns and rows can be dragged and dropped and with the use of graphs, charts and custom visuals. Then the reports are pinned to a dashboard with the "Pin Live" feature, which takes the diagram from the dataset section and pins the diagram to the dashboard. Here are couple of features the Power BI platform offers:

Real Time Analytics: This feature is used for real time monitoring of data as it comes in. Alerts can be configured using parameters to monitor the data.

Pin Live Page Feature (Auto refresh feature): This feature is used in the report section to upload charts and graphs to the dashboard, but the special feature is the auto refresh capability on data streams.

Report Section: This is for making charts and graphs from a dataset. The user configures how they want their visualizations to appear.

Dashboard Section: This is where reports are posted. Here, users can analyze charts and graphs to better understand the data.

Data Stream Capabilities: Power BI can collect both dynamic and static data. Log files, tweets, sensor data can be collected and visualized using Power BI.

Connect to Azure Stream Analytics, Hadoop, Salesforce, Tableau etc. Data Streams: The following applications can connect to Power BI as a data source (this is not an exhaustive list of applications)

Share Dashboard: Dashboards from Power BI can be shared to any individual who has access to a Power BI account. For example, businesses can share a central dashboard and drop reports into it allowing employees across the business to view them.

Q&A Feature: Questions can be typed into the search bar with column name, row values or values relevant to the dataset, and the returned answer is in a visualization format.

Analyze in Excel Feature: The dashboards can be exported to Microsoft Excel for further analysis if needed by user.

Create a Group Feature: Groups can be created for users to facilitate multiple users on a single dataset, reports or dashboards.

Power BI Gateways: This is an admin feature for users to sign in before they can get access to a Power BI user interface.

Create and View Content Pack: Interfaces can be packaged and sent to other users which can be viewed but not edited.

Development Tool Feature: This is a platform where custom visuals are created and can be edited by normal users.

Creation of Embedded Code: Here, embedded code for blogs and social media are made, which eventually are copied and pasted to web pages.

Workspace Feature: This is a feature where different users can have a multiple user interfaces to work with, all in one Power BI window.

Power BI Community: This is a direct access to the Power BI user forums for questions, discussions, and updates of new technology being added to the software.

4. CONCLUSION

Today, data science careers are in demand (Wimmer & Powell, 2016; Evans, 2016). As such, many IS and technology related programs are incorporating business and data analytic courses into their curriculum. Wimmer and Powell (2016) state that teaching students business and data analytics is challenging within the current budgetary constraints of academic institutions. Furthermore, teaching resources for business and data analytics are limited (Wimox et al, 2014; March & Pepard, 2013; Tiong & Pei-Chen, 2015). This paper highlighted business and data science tools that can proved educators with an academic literature resources to expose the student to additional business and data analytic tools and their features.

Future research will include further exploring implementations of additional business and data analytic tools, comparison of business and data analytic tools, as well as providing additional teaching resources regarding business and data analytics.

5. REFERENCES

Corbett, P. (2009). NASDAQ to Stream Real-Time Data to StockTwits and Twitter. Available at https://isl.co/2009/08/nasdaq-to-stream-real-time-data-to-stocktwits-and-twitter/

Evans, J. R. (2016). *Business Analytics: Methods, Models, and Decisions*. Boston, Pearson.

Halzach, S. (2014). Is the New Technology at Macy's Our First Glimpse of the Future of Retail? *The Washington Post*, Available at http://www.washingtonpost.com/news/business/wp/2014/09/25/is-the-new-technology-at-macys-our-first-glimpse-of-the-future-of-retail/

Henry, R., & Venkatraman, S. (2015). Big data analytics the next big learning opportunity. *Academy Of Information & Management Sciences Journal*, 18(2), 17.

INETCO Systems (2016). INETCO Insight® transaction monitoring software for omni-channel banking and payment processing environments. Available at https://www.inetco.com/products/inetco-insight/#sthash.FnKRoB3a.dpuf

Liu, K. (2016). What your industrial data analytics courses need. *Industrial Engineer*: IE, 48(4), 43-46.

Lustig, I., Diettric, B., Johnson, C. & Dzickan, C. (2010). The Analytics Journey, *Analytics*, 11/12(2016)

Marchand, D. A. & Peppard, J. (2013). Why IT Fumbles Analytics. *Harvard Business Review*, 91(1-2), 104-112.

Marjanovic, O. (2013). Sharing and Reuse of Innovative Teaching Practices in Emerging Business Analytics Discipline. Paper presented at 46th Hawaii International Conference on System Sciences (HICSS), 2013.

McDonald, M.P. & Rowsell-Jones, A. (2014), *The Digital Edge: Exploiting Information & Technology for Business Advantage* (Stamford, CT: Gartner, 2012); San Francisco Municipal Transportation Agency, "SFpark: Pilot Project Evaluation," Available at http://direct.sfpark.org/wp-content/uploads/eval/SFpark_Pilot_Project_Evaluation.pdf

Pigni, F., Piccoli, G., & Watson, R. (2016). Digital Data Streams: Creating Value From The Real-Time Flow Of Big Data. *California Management Review*, 58(3), 5-25.

Piccoli, G, Rodriguez, J. & Watson, R.T. (2015), Leveraging Digital Data Streams: The Development and Validation of a Business Confidence

Index. Paper presented at 46th Hawaii International Conference on System Sciences (HICSS), Kauai, HI, 2015.

San Francisco Municipal Transportation Agency, (2014). SFpark: Pilot Project Evaluation. Available at http://direct.sfpark.org/wp-content/uploads/eval/SFpark_Pilot_Project_Evaluation.pdf

SHIM, J. P., KOH, J., FISTER, S., & SEO, H. Y. (2016). Phonetic Analytics Technology and Big Data: Real-World Cases. *Communications Of The ACM*, 59(2), 84-90.

Smigala, H. (2013), IBM partners with 28 business schools and universities, to help train tomorrow's data scientist. Retrieved from http://www.03.ibm.com/press/us/en/pressrelease/44031.wss

Tiong T., G., & Pei-Chen, S. (2015). Teaching Social Media Analytics: An Assessment Based on Natural Disaster Postings. *Journal of Information Systems Education*, *26*(1), 27-36.

Wixom, B., Ariyachandra, T., Douglas, D., Goul, M., Gupta, B., Iyer, L., & Turetken, O. (2014). The Current State of Business Intelligence in Academia: The Arrival of Big Data. *Communications of the Association for Information Systems*, 34(1).

University of Illinois (2016). Online Teaching Activity Index. Available at http://www.ion.uillinois.edu/resources/otai/

Wimmer, H. & Powell, L. (2016). A Comparison of Open Source Tools for Data Science. *Journal of Information Systems and Applied Research*, 9(2), 4-12.