BI Endgame – When BI meets AI

THE LINUX FOUNDATION

ODPi
odpi.org

ODPi BI & AI

Table of Contents

Introduction

Recently, while speaking at conference in Washington DC organized by one of the biggest BI vendors in the industry, I proclaimed that "**BI is dead…**" As expected, saying this resulted in a moment of silence as many people began reconsidering their careers. But then I continued…

"BI is dead, unless you incorporate AI in it"

Like it or not, **BI Endgame** has come. We can either fight back or wait to be killed. In this whitepaper, six BI leaders will share their strategy for this battle. In particular, we explore how they leverage AI in the legacy BI platform to make it more competitive. Interestingly, after working with them for the past few months to understand their secret weapons, I found out that they fall into the Infinity Stone categories well, with a little twist:

Power – Microsoft Power BI. Besides sharing the word "Power", it allows users to extend already familiar products like Excel with a dose of AI steroid to empower the BI offering.

Space – MicroStrategy. Because of the unified architecture design, it allows users to "teleport" into multiple data sources efficiently. Addition of the AI Bifrost shortens the analytics journey.

Time – Qlik. Even though it cannot help in time travel, its distinct Associative Engine makes analytics runs very quickly including AI model which is typically a time-consuming process.

Soul – SAS. It puts user experience, the soul of the analytics, right in the middle of the BI platform. Combining that with the AI spirit, this allows users focusing on business.

Reality – Tableau. It has superior visualization capability allowing people to see and understand the reality hidden behind data, which may not be always obvious.

Mind – ThoughtSpot. It leverages AI in the data insight engine within the BI platform. This results in reading your mind to know what you want, before you even ask for it.

Associating each BI tool to one Infinity Stone does not mean they are good at only one aspect. In fact, all vendors are capable of different areas to a certain level. However, just like everything else, there will be something relatively more outstanding than others. Hence this analogy is to highlight the unique area where individual shows more prominent performance.

Moreover, a common theme across all platforms is AI can be incorporated in BI in 2 different aspects: (1) AI is applied to the tool itself to make it more intelligent (2) Expose AI to the end users so that they can obtain more insights on the data set which may not be possible with only the traditional BI functionality.

Back in November 2017 at the Conference on Health IT and Analytics (CHITA) organized by University of Maryland, I first proposed a concept: **AI + BI = CI** This equation suggests that true cognitive intelligence (CI) can only be achieved when machines accelerate (AI) the direction intuited by human insight (BI) Now, this concept seems to be pervasive. Let's now take a deep dive and see how each vendor is handling this Endgame in more detail in the following sections!

– Cupid Chan

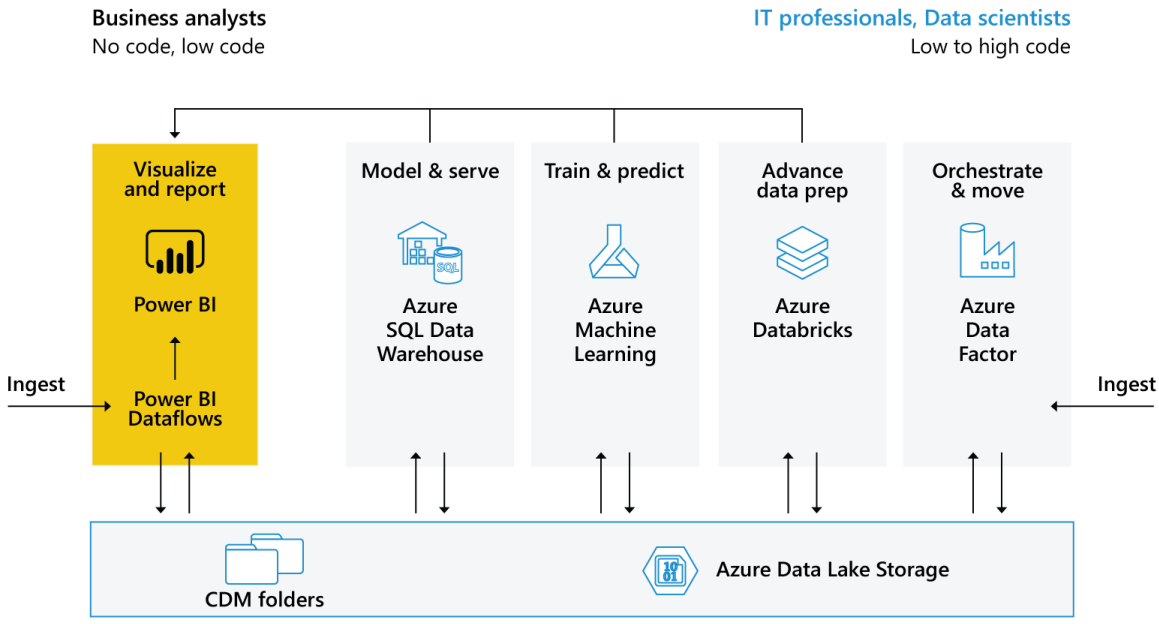
Microsoft PowerBI

Bring the full power of Microsoft cloud technology into the hands of the everyday use

According to the new IDC Spending guide[1], released September 2019, it is estimated that almost $100 billion will be spent on Artificial Intelligence in 2023. There is no question that companies will continue to invest heavily in AI technology. The outstanding question, however, will be "*how* will they choose to invest in AI?" Currently the most common expenditure for an AI project is currently on IT services (outside expertise.) The companies that implement AI technology that is intuitive/easy to learn/deploy will set themselves apart by saving large amounts of money and gaining even more insight.

By making artificial intelligence intuitive and easy to implement, BI empowers business users to analyze data in ways that were previously only reserved for data scientists with advanced degrees. AI can aid in data exploration, automatically find patterns in data, and predict future outcomes to help business drive results in a fraction of the time/cost.

Microsoft is directing great amount of development resources[2] and large investments toward making Artificial Intelligence as intuitive as possible[3]. Building on this framework, these AI capabilities can be easily adopted by analysts in a low-code/no-code environment. The ease of use is realized by analysts that have been traditionally working with Excel and many other products, which makes it easy to adopt. Artificial Intelligence can be integrated at every stage of the report authoring experience to accelerate report development and uncover a wide range of new possibilities for data discovery.



[1] https://www.idc.com/getdoc.jsp?containerId=prUS45481219&mkt_tok=eyJpIjoiT1dNMVkySmxaRE5pWTJJNSIsInQiOiJPNWgzREIwZzZUdVRLZ3NqRXpBZlNvQzMxTlpZdGl0XC9ZSldLUDF6SFZ0WUtEOGxRaG1oSkhpbjRpbkNkbmppUnFLaHJhOGRxclV1ZisxNG1MeXlvXC9sU3JwT05uV2lLMDBjKzc3b0dBdGwzRUFoMXhBd1owT3dIMld2VUlRXC9keSJ9

[2] https://www.geekwire.com/2019/microsoft-invests-1-billion-openai-vows-build-ai-technology-platform-unprecedented-scale/
https://www.cnbc.com/2018/06/20/microsoft-buys-ai-start-up-bonsai.html

[3] https://techcrunch.com/2018/09/13/microsoft-acquires-lobe-a-drag-and-drop-ai-tool/ ;
https://www.forbes.com/sites/louiscolumbus/2019/01/06/microsoft-leads-the-ai-patent-race-going-into-2019/#32fa75f944de ;
https://www.fool.com/investing/2018/06/20/heres-how-microsoft-is-investing-in-ai.aspx

Extending Artificial Intelligence models BI

Some of the largest cloud vendors in our industry are rapidly developing AI capabilities that can be extended to BI to uncover new frontiers of insight. We integrate with these AI capabilities in our data ingestion process (https://docs.microsoft.com/en-us/power-bi/service-dataflows-overview) which can be configured following these links: https://aka.ms/enableaiworkload & https://aka.ms/azuremlpbi

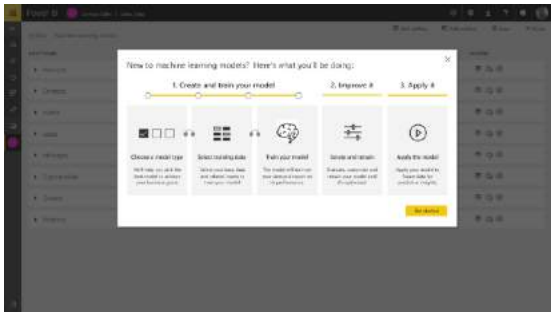
Image recognition and sentiment analysis without coding

Users looking to gain insight ***from unstructured data*** can now use the AI from Azure Services to get capabilities such as image recognition and sentiment analysis. And the best part is that no code is required so BI users to discover hidden, actionable insights in their data and drive better business outcomes with easy-to-use AI without having to write custom code.

These capabilities are enabled through integration with Azure Cognitive Services (https://azure.microsoft.com/en-us/services/cognitive-services/) to classify images, detect language sentiment, etc. Once this is configured, cognitive services are accessible through the AI Insights browser, which is an editor for dataflows. From this editor, a user can rate the sentiment analysis of a customer's review or even detect the language the text is in.

Introduce predictive analytics in BI

Users looking to **identify patterns in large, noisy structured data** sets can create machine learning models directly using Automated Machine Learning (https://studio.azureml.net/), which is a cloud predictive analytics service that makes it possible to quickly create and deploy predictive models as analytics solutions. Models built by data scientists can now be easily shared with business analysts. This makes collaboration among business analysts and data scientists easier and faster than ever before.



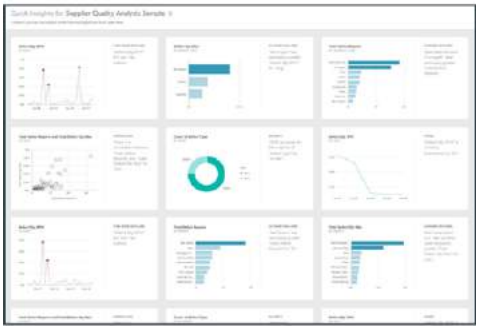
The AI Insights browser displays all ML models that have been shared. These models can be introduced to preexisting queries and will return the appropriate score that the ML model calculates. Users can access existing ML models via AI Insights in the editor or they can navigate to any dataflow and build a new ML model through AutoML: https://powerbi.microsoft.com/en-us/blog/creating-machine-learning-models-in-power-bi/

As an example, a business analyst could leverage the automated machine learning technology to quickly and easily build a model to predict how likely an open sales opportunity is to be won. This could help a sales manager prioritize which high value opportunities to focus on and how likely they are to meet their target.

Artificial Intelligence within BI

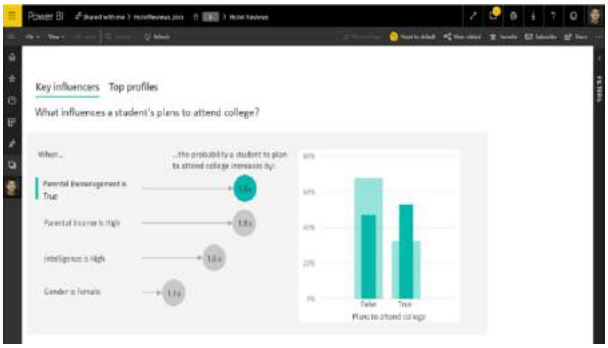
New insights can also be generated on top of the data that is published to the cloud service. The following are features that anyone accessing a BI report can take advantage of from their web browser.

Generate visualizations with AI



Another AI integration with BI to is to allow users to find insights in their data in new and intuitive ways by a set of advanced analytical algorithms. With a click, user can search different subsets of the data set while applying a set of sophisticated algorithms to discover potentially interesting insights. It scans as much of a data set as possible in an allotted amount of time: https://docs.microsoft.com/en-us/power-bi/service-insights

Uncover key relationships with AI



BI user should be able to visualize the AI result so that analysts can understand what drives an outcome. It uses ML.NET, an open-source machine learning toolkit, to run logistic and numeric regressions over the data set to surface anything that seems to be a key driver in effecting an outcome. Results that are calculated to have a significant influence (p=0.05) are automatically ranked and visualized for the analyst. For example, consider the likelihood that someone will repay a loan and the various factors that contribute to this. Key driver analysis automatically surfaces those things that matter most, like interest rate, loan program, or even location.

This can be configured by simply dragging an independent variable into the "Analyze" field in the visualization banner and whatever remaining relevant data fields to the "Explain by" field in the visualization banner: https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-influencers

Natural Language Processing in BI

End users can also take advantage of natural language processing to get an answer to a question they have or quickly build a visual that they want to embed in a dashboard. This is particularly insightful because it leverages AI to navigate through all of the data in a particular report or workspace and builds a visual on-the-fly[4]

Artificial Intelligence in BI with no cost

With a desktop app you can import data and use many of the same features discussed above. Key Influencer, Q&A in Desktop[5], and Insights[6] are configured very similarly. Additionally, a report author can take advantage of Add Column by Example[7] and Automatic relationship mapping[8] to accelerate their report development work.

All in all, by introducing AI to BI, we extend the outer limits of what was previously possible for data scientists and the intuitive nature of the product expands the audience of potential adopters of the technology.

[4] https://docs.microsoft.com/en-us/power-bi/consumer/end-user-q-and-a
[5] https://docs.microsoft.com/en-us/power-bi/natural-language/q-and-a-best-practices
[6] https://docs.microsoft.com/en-us/power-bi/desktop-insights
[7] https://docs.microsoft.com/en-us/power-bi/desktop-add-column-from-example
[8] https://docs.microsoft.com/en-us/power-bi/desktop-create-and-manage-relationships

MicroStrategy

Viewing AI as a tool for breaking down complex business problems and BI as a means for distributing or embedding the insights

If you examine popular consumer-facing applications, spanning retail, social networking, media, and other industries, you're likely to find examples of artificial intelligence (AI) at work. Key to these product's success is the fact that AI-powered functionality is simple to use and augments end-users in some significant way. Take, for example, Kayak.com. With each search, the site generates price predictions that enable future travelers to evaluate if now is a good time (or not) to book a flight. Another example is Spotify. With the tap of a finger, Spotify can generate unique playlists that are personalized to each user's unique taste in music. In both cases, AI is used for complex, multi-variate decision making, but operates in the context of a much broader system.

Compared to the role of AI in consumer-facing applications, in business intelligence (BI), AI is most often relegated to augmenting historically-oriented reports, dashboards, and visualizations with basic predictive insights. In other words, in BI, the types of problems AI is applied to is often very different than in other domains. This could be because organizations are still discovering how to apply AI to their business and there hasn't been a need for BI yet, or that AI and BI practitioners sit in different parts of the organization and their job responsibilities don't intersect. Unknown to both AI and BI practitioners is the fact that many of the data engineering, systems integration, and security challenges faced by data scientists are problems that BI systems are historically well-adapted to solving. AI and BI teams have much to gain if the gaps, both technical and organizational, were reduced or eliminated. To achieve this, we need simpler integration points between specialized AI technologies and BI platforms.

When accessing data directly, data scientists have to deal with data that is spread out over numerous databases and they spend vast amounts of time trying to make sense of it all. Fortunately, the BI system's job is to centralize cross-departmental data, provide a common definition, and secure mechanisms for data access. What if data scientists could leverage these assets? How would this improve their productivity and the quality of their solutions? If data scientists could easily tap into these assets, it would make the BI system more attractive as a data source and a solution deployment vehicle. Through APIs, the data catalog, data source connectivity profiles, and security frameworks that are managed by the BI system could be exposed through integrations with data scientists' preferred tools such as Python, R, Spark, or cloud-based auto-ML platforms. Doing so would enable data scientists to consume high-quality business data and, in turn, contribute sophisticated predictive insights to the BI system for other users to take advantage of. To do this, the BI platform must provide APIs, extensive database connectivity, and in-platform integrations that give data scientists a path to production.

If fully integrated within the BI system, the operators of these models would need administrative tools for architecting and controlling model-related operational processes. These include the ability to connect to new databases, define queries for extracting data, create and manage refresh schedules, detect data anomalies that occur during refresh processes, and provide dashboards for observing the performance of the system over time. Because model refresh processes can be time and resource-intensive, these solutions should run on horizontally-scalable, cloud-friendly infrastructure and isolate model-related workflows from unrelated analytics workflows occurring elsewhere in the BI system. Finally, while monitoring the quality of model decisions is a separate activity typically performed by model developers, model operators need to monitor the performance of scoring and inference

processes. They need to know how quickly predictions are generated to ensure that their solution continues to satisfy latency requirements and that predictions are generated in a stable way.

When considering how to seamlessly integrate AI and BI, it is clear that evolution within BI is needed. First, BI systems must evolve to support technologies that enable simple, scalable, and secure integration with tools and techniques preferred by AI practitioners. Doing this will make it more likely for AI solutions to be adopted. On the other hand, BI practitioners need to learn how to apply AI to new types of problems and recognize that doing so may challenge the form factors used traditionally to convey analytical information within BI. This paradigm shift – viewing AI as a tool for breaking down complex business problems and BI as a means for distributing or embedding the insights – is inspired by examples of AI in the consumer space where AI is more prevalent and trusted by users for decision making. At MicroStrategy, we're focused on enabling data scientists by surfacing MicroStrategy platform capabilities within the tools data scientists are experts in. By building solutions on a unified BI platform, data scientists can focus on hard problems that affect the entire workforce and leverage extremely deep data source connectivity, API, and SDK functionality to embed predictions in reports, dashboards, custom web portals, and productivity applications that can be scaled across the enterprise.

Qlik

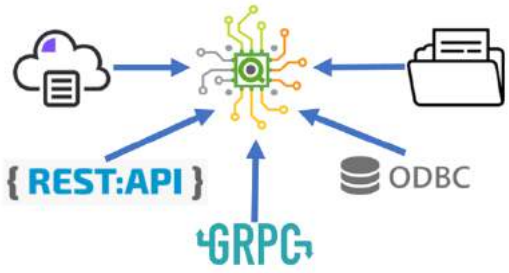
Once the data is in the engine, it can be elegantly joined up with other entities in the model and made "associative" with the rest of the entities in the application.

Machine Learning Approaches

BI platform can offer many options when it comes to integrating with machine learning algorithms and platforms and includes many native advanced analytic functions and machine learning driven capabilities. The various integration methods consist of consuming output from other systems, Server-Side Extensions (SSE) for scoring and unsupervised algorithms, asynchronous workflows where the engine pushes the data to an external system and later consumes the enriched results and use the BI platform as a source for data scientists. There are also 3rd party integration capability through extensions which allow interacting with elements such as Natural Language Processing (NLP) engines, D3js charting, or any programmatic system with an API.

Consumption of Output

Enriching an existing dashboard that analysts already use is a powerful way to include valuable insights generated by data scientists. Users can consume data from REST endpoints, file extracts, ODBC, or gRPC based microservices. Once the data is in the engine, it can be elegantly joined up with other entities in the model and made "associative" with the rest of the entities in the application.



ODBC Connectivity

It's quite common to have machine learning pipelines output results to a data lake or database that has connectivity via ODBC, this is helpful for loading outputs or leveraging SQL-based machine learning platforms without having to worry about long running batch queries.

REST

Using a "no-SQL" analytical engine approach to load XML, JSON, or CSV data presented via HTTP is a natural and easy method of ingestion. Ingesting via REST opens a lot of options for outputs presented in most machine learning platforms without having to write out to a database first. Pagination, many authorization options, and all HTTP parameters are supported.

File Extract

If you have a storage system where the ML pipeline writes out to, BI Platform should naturally consume delimited text, JSON, and XML. It can either be sourced locally on the server, file share, or it can be read from a cloud storage service.

Server-Side Extensions (SSE), gRPC Connectivity

SSE is a form of ingestion directly from Machine Learning algorithms and platforms. These extensions implement a function that can be called in an expression for a measure; or can be called in the Load Script for ingesting data. This means a function written in Python, R, Java, Go, C++, or anything gRPC, can

be exposed directly to the engine by simply aliasing the columns returned, as it automatically joins up with the rest of the data model. You can send in parameters, including dynamically with variables, and can even load a previously loaded table as a parameter. This means you can interactively push data into the machine learning models, have an algorithm run on it, and then ingest the results back into the analytic data model.

SSE Integration

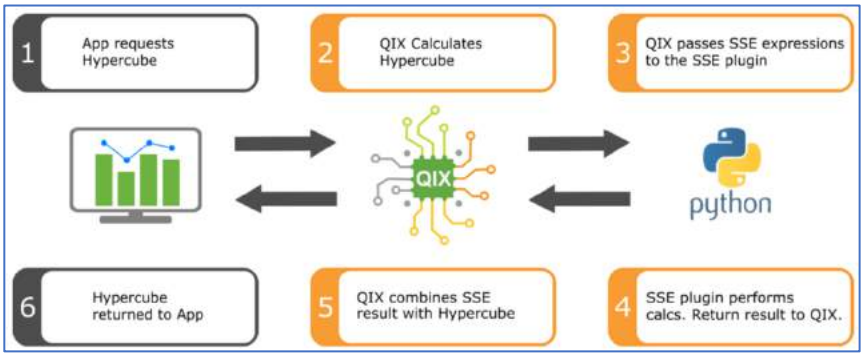


Figure 1: SSE Architecture

This use case shows the SSE plugin called PyTools. It will score and geospatially visualize a data submitted to a Python clustering algorithm based on fatal accidents by region in Australia

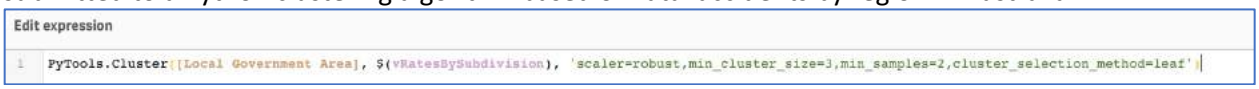


Figure 2: Python Script

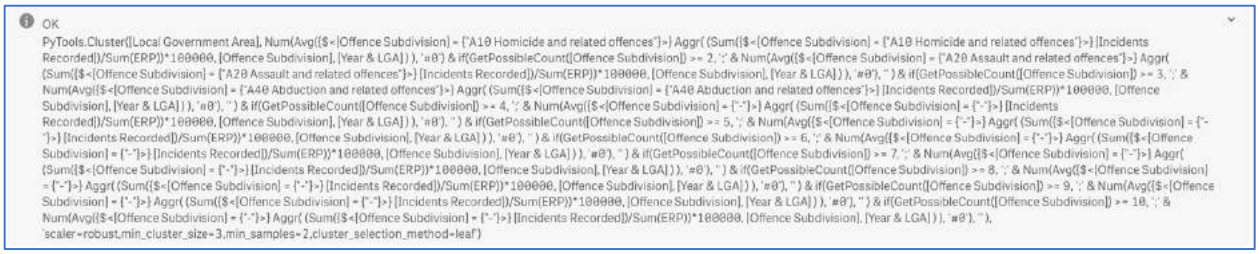


Figure 3: Submits to Python

This function returns clustered information which is used to plot in a multilayer map. The Python SSE plugin is running on the server, instead of desktop to execute this computation. The power of the SSE engine means that any user via just a browser on any device can harness the power of data science.

Qlik as a Source (QaaS)

Besides a consumption mechanism of data science output, a BI platform can also be used as a data science data input. The API is websocket/JSON-RPC based which external applications can access to perform multiple functions such as:

- List available apps pre-created by the business intelligence team members or other self-service users.

- Attach to an app and get a list of columns and other metadata about the data model loaded into the app.
- Define and create a “hypercube”, a virtual representation of that a particular schema, (for example, requesting 3 dimensions and 5 measures)
- Request the data extract. Using a python module, extract a defined data set into a dataframe which is then sent to machine learning system such as CDSW, Python, etc and then visualized in matplotlib or plotly.
- When the result is properly defined, this data can be sent back into the source application or a new app can be created for other users to later consume.

Built-in Machine Learning Features and Functions

Another aspect of AI in BI is to analyze data and present chart suggestions which enhance the self-service user experience with capabilities such as Insight Advisor. There are also data specific advanced analytic functions such as correlation, linear regression (non-model based), and standard deviation which use native C++ libraries and are very fast and powerful. On the other hand, a recent addition of “Insight Bot” enables a conversational AI bot providing insights on demand through a natural language interface.

Visualizations

There are several visualizations that are helpful and common in data science results presentation/analysis: The histogram, scatter plot, distribution chart, and of course the box plot. In addition to these visualizations there are many more available through the Extension API and can be downloaded through the open source community. “Visualization bundle” is provided with advanced visualizations developed through the Extension API. All visualizations can also be embedded into third party applications / websites, so democratization of data science insights is in many cases just a line of code.

Summary

Qlik democratizes access to insights driven by AI/ML platforms. BI can now serve as a flexible and powerful data science AI/ML platform via its integration with all programmatic systems with a range of options including ODBC, gRPC, and REST. The ability to provide data as either outputs or inputs to data science systems provides powerful capabilities for the data science and advanced knowledge worker community.

SAS

Empowers customers to transform a world of data into a world of intelligence by converging the power of data, analytics, AI, ML, and BI to drive transparent, trusted and reliable insights across our industry solutions.

The lines are starting to blur between the once different offerings of Business Intelligence (BI) and Artificial Intelligence (AI). Traditionally BI told the story of the past and AI told the story of the future. As BI solutions start to embrace AI into their offerings, the past and the future are coming together.

Why are BI and AI Merging?

AI is the science of training systems to emulate human tasks through learning and automation. With AI, machines can learn from experience, adjust to new inputs and accomplish tasks without manual intervention.

The explosion in market hype around the term AI is closely tied to advances in deep learning and cognitive science, but AI spans a variety of algorithms and methods. It doesn't require the flashiest new technologies to be considered an AI application.

As a topic of interest for years – from science fiction plots to futurists' prophecies – the promise of AI has always been at the forefront of our minds. But what was once a distant vision is becoming reality as organizations embrace the value of AI now:

- By 2025, the AI market will surpass $100 billion. (Source: Constellation Research)
- 72% of business leaders believe AI will be fundamental in the future. (Source: PwC)
- In the immediate future, executives are looking for AI to alleviate repetitive, menial tasks such as paperwork (82 percent), scheduling (79 percent) and timesheets (78 percent). (Source: PwC)

AI must have a way to illustrate to the business community its findings and present the results of its algorithmic learning to the decision makers. This is where BI comes into play by translating the results of the AI process into something that is consumable to the business community. Based on the market predictions, it makes sense BI solutions want to get into the AI game and AI solutions are looking at ways to provide a presentation layer to the end users. The beauty of this merger is that AI will automate BI activities, proactively decide on the right presentation (e.g., chart, graph) to use based on the data and past BI design preference learnings and explain the findings in natural language.

Why AI is a hot topic?

C-level executives are taking a close look at AI for a host of good reasons:

AI automates repetitive learning and discovery through data. Unlike robotics, which automates manual tasks, AI automates high-volume computing tasks such as search and classification – reliably and tirelessly. The emergence of Natural Language Processing (NLP) will expand the use of AI because now human language can be used to ask AI questions.

AI adds intelligence to existing products. Think about how Siri added new value to Apple products. Or how AI can make online chat with a bot feel like talking with a human. AI enhances technologies in the home or workplace, from consumer marketing and security intelligence to investment analysis.

AI adapts through progressive learning. The data does the programming - an algorithm can teach itself how to play chess or what product to recommend next online. Through back-propagation, AI models adapt to new data they are given or what they learned from experience.

AI analyzes more and deeper data using neural networks that have many hidden layers. For example, building a fraud detection system with five hidden layers was almost impossible a few years ago. It's achievable now, thanks to incredible computer power and big data.

AI achieves accuracy that was previously impossible. Our interactions with Alexa, Google Search and Google Photos (all based on deep learning) keep getting more accurate the more we use them. In the medical field, AI techniques can find cancer on MRIs as well as highly trained radiologists.

AI gets the most out of data. When algorithms are self-learning, the data itself can become intellectual property and a competitive differentiator. The answers are in the data; applying AI can ferret them out.

For the first time, companies have access to a full set of building blocks to begin embedding machine intelligence in their business processes. Almost every industry is already seeing the effects, from agriculture to transportation, health care to financial services. Machine learning (ML) empowers people to be more productive with tools that exist today.

While you may think your company is way behind regarding AI, Gartner says only 4 percent of all companies are currently using it. The vast majority are researching, taking a wait-and-see approach and determining how to be successful when they are ready to act.

Why data management matters

Here's the reality: ML and AI systems don't just extract insights from the data they are fed and do predictions like traditional analytics do. They change the underlying algorithm based on what they see in the data. The more data they are fed, the more tightly they define the algorithm and the more confidently they make classifications or predictions.

So, the "garbage in, garbage out" truism that applies to all analytic pursuits is truer than ever. If the data that feeds ML algorithms is not well managed, the results could be like the result of the whisper game – wrong statements where errors have multiplied upon themselves. The dangers in that are obvious: inconsistency, inaccurate insights, loss of trust and questionable AI results.

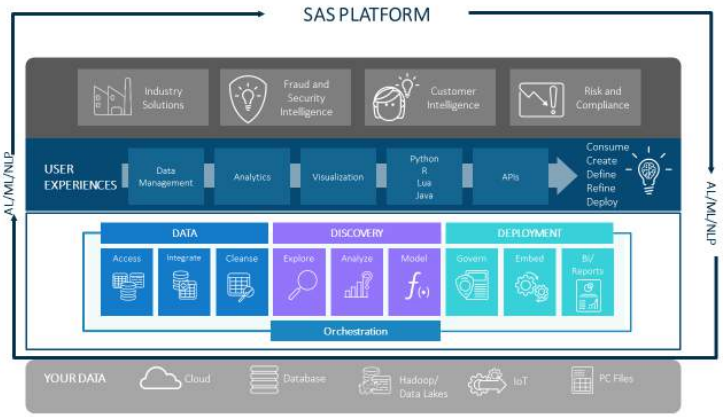
On the other hand, since data carries more weight than ever before, data management can become a real competitive advantage. Even if everyone is applying similar techniques in a competitive industry, the one with the best data management program will win. So, it's no surprise that 95 percent of C-level executives believe data is an integral part of forming their business strategy. (Data Management Benchmark Report, Experian, 2018. http://www.experian.com.vn/wp-content/uploads/2018/02/2018-global-data-management-benchmark-report.pdf). This has always been true, but ML and AI magnify the possibilities.

There is a general assumption that AI's automation handles much of the data management legwork. This is an incorrect assumption - data management is very much needed at the outset. While AI can support data management processes (such as determining what data to keep or discard or classifying data for optimal storage) it doesn't by default manage data for its own consumption or integrity.

AI and BI calls for chief data officers to take a holistic approach to data integration, data quality and governance. Underlying all these areas is cloud, where we're likely to see the most growth in the analytics data management market in general. In short, machine learning, AI, and BI is only as good as the data that goes in it – and it provides the best return when it is supported by a well-governed data management program.

As data continues to grow and becomes more diverse, AI and BI are the keys to understanding which data has value and how it should be presented to decision makers. Big data will continue to become bigger and the complexity will increase. AI and ML techniques will apply intelligence to big data based on what it has learned from past user preferences and data management procedures. Thus, data management practices will also evolve as the use and preferences of data will ultimately influence AI and ML processes. In turn, BI will progress because the algorithms being applied to the data will also dictate which visualizations will be presented to the end users. It is an exciting time in the world of data!

"*At SAS, we now look to the future with the same spirit of curiosity that has always led us to innovation*", said Kim. If data without analytics is value not yet realized, what does that mean in a world where data is growing exponentially? It's fair to say the world is at a turning point with data AND analytics. Soon everything and everyone will be generating data around the clock, and successful organizations will transform this new abundance of potential intelligence into value with analytics. We know that our customers will need guidance through the increasing complexity of data and analytics, technologies to stay ahead of the curve, and expertise to help guide them on their data and analytics journey. SAS' vision is to turn a world of data into a world of intelligence. We fully embrace our vision across our entire suite of technologies that make up the platform by weaving analytics across our comprehensive offerings.



We have heard the phrase- 80% of the time is spent preparing the data for analytics, AI, and BI while only 20% is spent using the data for business value. This is not a good way to run a business. The better way is to give our clients a way to seamless access, integrate, cleanse and govern data in an automated, augmented and cognitive fashion. So that users are empowered to spend less time on data preparation and more time on AI and BI functions to drive business value. Some examples weaving assortment of ML and AI capabilities into data preparation solutions include:

- Apply MI and ML modeling techniques that can be employed to generate a list of suggestions to improve data. Over time, actions on these suggestions (accept or reject) will help to improve the results of the system.
- Data Profiling will automatically find and flag personal information that can influence user or system behavior. A table tagged as "personal information" may prevent data from being viewable by some users, for example.
- A discovery engine that analyzes data and metadata to suggest data transformations, and leverages machine learning to provide suggestion improvement over time.
- Recommendations to users and suggestions for "next-best "actions during the data preparation process.

Combining our robust heritage in data management with SAS' expertise in analytics, AI, ML, and Computer Vision, BI users can now dramatically reduce the time they spend with the data and increase the time they spend deriving business value from BI and analytics.

Tableau

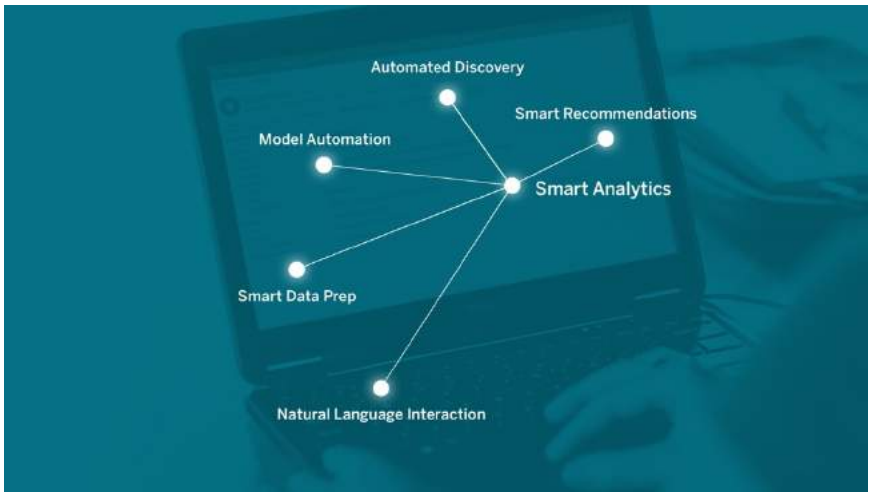
Help people see and understand their data

How BI can integrate with AI seamlessly?

Tableau is on a mission to help people see and understand their data. To achieve this mission, Tableau promotes the democratization of data with the fundamental belief that the people who know the data the best should be empowered to ask questions of their data.

Tableau leverages artificial intelligence (AI) to elevate its analytic platform to enhance the user experience through automation, ease of use, and speed.

Today BI integrates AI to augment analytics: automated discovery (Explain Data), data catalog, smart data prep (Prep Builder), model automation (Data Model), smart recommendations and search, and natural language interaction (Ask Data).



ClearGraph enables smart data discovery and data analysis through Natural Language Processing (NLP). It makes it easy to analyze data using natural language. It brings a consumer-like experience to users by connecting disparate data sources and making them accessible and intelligible through simple conversational style search. This unique natural language query technology stores semantic data in knowledge graphs that can expand and learn over time. Accessing and analyzing data then requires no technical training, as the system can infer users' intent through natural language. It is the key to Ask Data experience, by which users can ask questions of any published data source and get answers in the form of a data visualization. It enables users with the ability to explore data at the speed of thought and make better data-driven decisions.

Artificial intelligence (AI) and machine learning technologies can help people answer richer and deeper questions faster from their data. Empirical Systems specializes in automated statistics which originated at the MIT Probabilistic Computing Project. A unique analytics engine was developed and applied sophisticated statistical techniques to automatically uncover insights hidden in data. Without having to examine every dimension and variation in the data manually, people are automatically guided to relevant insights and alerted to data points that are worth exploring. The engine automates the analysis

and data modeling that would typically require a trained statistician and makes that analysis available in real-time so everyone can more easily spot correlations, outliers, and patterns in their data.

Combing with this AI capability, people can leverage the power of statistical analysis and get to insights faster, enabling more people to do sophisticated analysis without requiring deep statistical or data science expertise. These advancements present an opportunity to further simplify the analytics process, helping people more than ever before to see and understand data. This is key to "Explain Data" experience, which automatically provides AI-driven explanations for the value of a data point within a data visualization with a single click.

In summary, a key role that AI plays in BI platform is efficiency by taking what might be otherwise labor-intensive or manual processes and automating them or making them more efficient. AI is a force multiplier for BI.

To summarize, these are ways that AI is integral to the BI platform:

- Ask Data – Simplifies querying data through natural language as users can use English like statements to query their data. This helps with creating a data culture by fueling speed to insight and engaging with immediacy versus deterrence with latency.
- Data Catalog – through the creation of a semantic layer that ingests data from different sources and harmonizes it, the data catalog manages a single view of the truth and resolves potentially problematic data challenges like duplicate data because of data granularity issues in data joins that users may not naturally understand at first glance of the data.
- Explain Data – an AI-powered analysis of data that offers up "the why" behind unexpected values in data.
- Prep Builder – through the use of fuzzy matching logic, several algorithms can be applied, including pronunciation, spelling, and common character, to consolidate/group like or similar data values that would otherwise require manual analysis to achieve the same.

Tableau integrates AI capabilities to enhance automation, cognition, and efficiency of the analytic user experience and streamline the decision-making process.

ThoughtSpot

A Search and AI platform that can uncover insights for you automatically thru the data insights engine which generates trusted advice from billions of rows of data with just a single click.

IDC forecasts the amount of data created annually at a staggering 175 Zettabytes (or 175 trillion gigabytes) in 2025[1]. Compare that to 33 Zettabytes of data created in 2018. Connected people – 75 percent of the world's population – will interact with data every 18 seconds on average in 2025, according to IDC.

Unfortunately, while data volume is rapidly growing, the volume of insights we're able to extract from it has been fundamentally limited. That's because in today's analytics paradigm there's a huge gap between data supply and data demand. On one end, there are many data consumers across every line of business who crave new insights. On the other end, there are a few data producers – the data experts – who are required to extract value from data. As more and more data is collected, this small group of trained experts is under more pressure.

That's why AI and machine learning present such a significant opportunity in the world of analytics. By infusing AI into analytics workflows, you can transform your organization and bridge the supplier-consumer divide by giving everyone access to the tools they need to make data-driven decisions. The good news is that AI has already arrived and is changing the way business people – such as marketers and sales people – interact with the data they have at their disposal. Today, the uses of AI in analytics can be boiled down to three categories of technology: automated data discovery; search and text-based analytics, and; intelligent modeling and recommendations.

Automated data discovery encompasses a class of technologies that automates the process of data analysis and exploration in real-time. This includes everything from selecting data sets to explore, running queries automatically, combing through results for insights, and choosing a best-fit visualization paired with a natural language description of each insight.

The number of possible questions to ask of data is often too much for any human. With automated data discovery technologies, business people can rely on machine-driven smarts to explore complex datasets with a few clicks and get insights explained to them in natural language. They don't need a trained analyst it would otherwise take to explore the data manually.

AI-powered data modeling and recommendations can reduce time spent on this kind of work by automatically generating statistics about data sets, inferring data types, identifying hierarchies and relationships within data sets, and dynamically aggregating data at query-time.

This covers just a few examples of how AI is fundamentally changing the world of analytics. Applied correctly, artificial intelligence has the potential to substantially impact or predict business outcomes, exponentially improve employee productivity and decision-making, and even create new jobs within the data team by increasing data literacy.

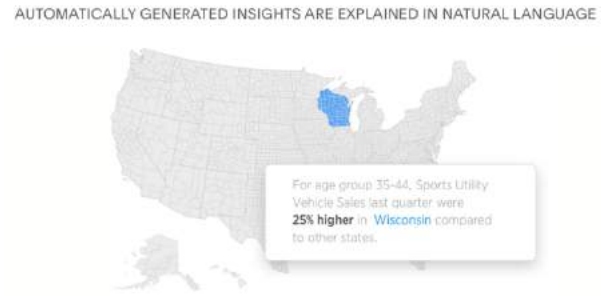
Trust-based AI

Despite the promise of these offerings, there's a problem lurking at the core of AI in today's world. Many understand what innovative AI technologies can accomplish but few know how they work, creating a general feeling of distrust. Many organizations count on analytics technologies to drive their most critical decisions, so they may find it difficult to adopt revolutionary technologies such as AI without understanding what's happening under the hood.

And that's why trust is the key to adoption of AI-infused analytics. When it comes to analytics, trust is created by delivering accurate, relevant, and transparent results. To do this, machines should not rely solely on their own built-in learning algorithms but must work together with humans to ensure every result meets these standards of trust.

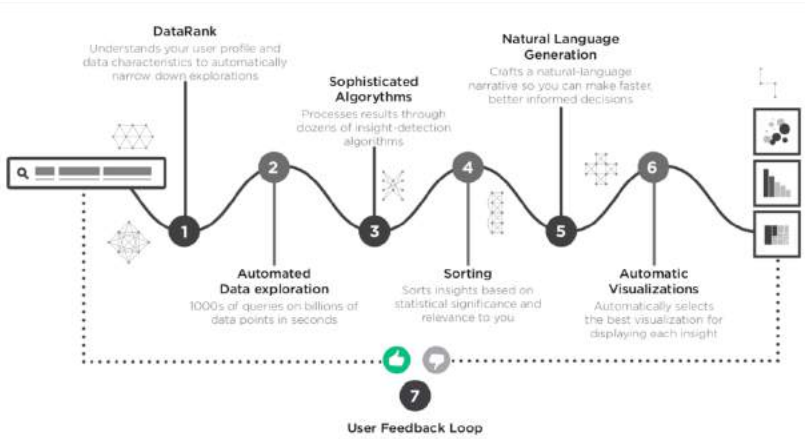
This philosophy underlies ThoughtSpot's SpotIQ – an automated data insights engine that makes it easy for any business person to automatically generate trusted insights with a single click.

Combining search & AI into a single analytics platform puts the power of a thousand analysts in the hands of every business person. Users can search to easily analyze data or automatically get trusted insights pushed to you rather than having to dig thru data to find the information. There is a relational search engine to curate deep and relevant insights for users that they may not have thought to look for on their own. AI applied to data can then be used to automatically ask thousands of questions about billions of data points and bring back dozens of insights in seconds. In a traditional BI paradigm, you would have to hire a thousand analysts, tell them exactly what questions you want them to answer, and then wait for them to deliver relevant reports and dashboards.



With AI, there is a huge opportunity to enable millions of people to make smarter decisions fueled by automated AI-driven automated insights. In this example, we use AI to analyze retail data to share with the user that SUV Sales were 25% higher in Wisconsin.

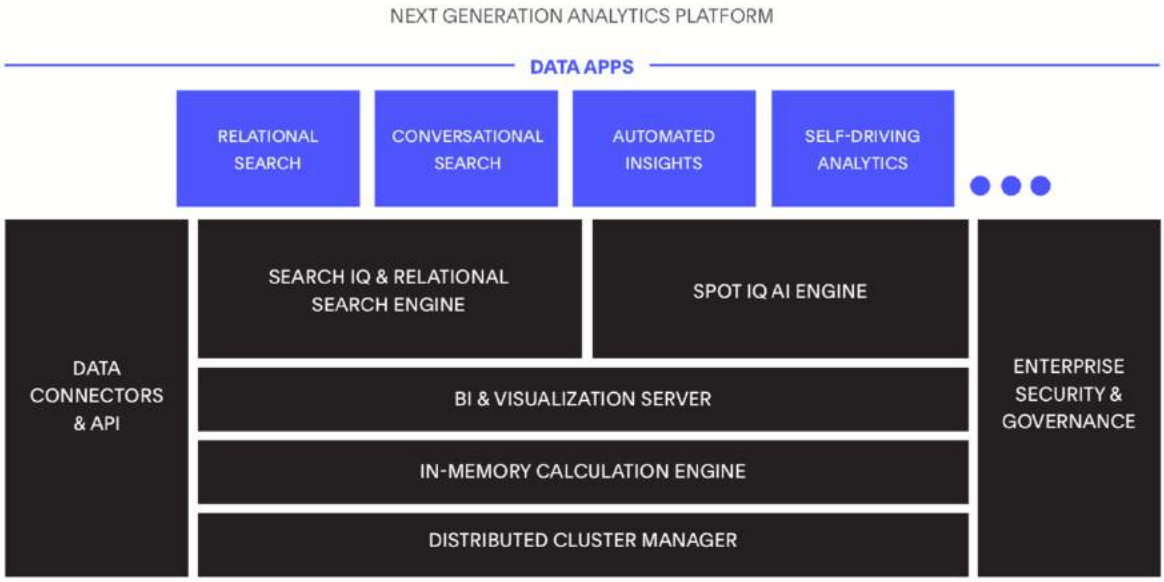
By integrated AI directly into the analytics experience, a user types in a question, the relational search engine calculates an answer in the form of a best fit visualization. Furthermore, users can auto analyze data and answers, and custom analyze to focus on specific aspects of the data. Selecting either of these options will again kick off an AI workflow, which executes the following steps to augment the analysis with automated insights on the data.



Effective AI Requires a Platform-up Approach

AI requires the ability to perform complex calculations on massive amounts of data at extremely high speeds to deliver on its promise. But how can you provide a simple experience that can handle the user and data scale along with the complexities of the enterprise to make AI accessible to everyone? Traditional disk-based solutions, or even hybrid in-memory and disk-based solutions, are not adequate to meet the performance and scale demands of AI.

Rather, AI-driven analytics require a distributed, massively parallel, in-memory execution engine to provide processing speeds at scale orders of magnitude faster than traditional architectures. Combine that with enterprise-grade requirements like security, governance, high availability, and manageability and the only possible solution is a vertically-aware analytics stack built from the ground-up for AI-driven analytics.



ThoughtSpot has delivered this next generation search & AI-driven analytics platform that combines the precision of the world's first relational search engine with the smarts of a robust AI engine and the scale of a massively parallel in-memory data cache and calculation engine. As data is cached, the entire data model is indexed, including the raw data, metadata, and relationships. This makes it easy and fast for anyone to perform database joins, drill anywhere in their data model, and change aggregations on the fly; which in turn makes the application smarter and more relevant for any user in the organization, regardless of their technical aptitude. This approach eliminates the need for rigid summary structures like cubes and data marts that require hours or months for technical resources to build, in turn delivering an Enterprise grade solution faster, allowing you to meet today's requirement of doing more with less.

Authors

Cupid Chan

Cupid Chan is a seasoned professional who is well-established in the industry. His journey started out as one of the key players in building a world-class BI platform. He has been a consultant for years providing solutions to various Fortune 500 companies as well as the Public Sector. He is the Lead Architect of a contract in a government agency leading a BI and analytics program on top of both Big Data and traditional DB platforms. He is the Board of Directors, Technical Steering Committee (TSC) and the Chairperson of BI & AI Project in Linux Foundation ODPi.

David Freriks

David Freriks is a Technology Evangelist in the Office of Strategy Management team at Qlik. Dave's mission is to help spread the word on the power of the Qlik Platform with Big Data systems and integration with partner technologies related to that ecosystem. Currently focused on data integration with Cloud data systems and associated technologies like AI/ML and change data capture, he has spent 20+ years in the business intelligence space working at Qlik, SAP, IBM, and Cognos helping launch new products to market. Dave has a background in data warehousing and a Mechanical Engineering degree from Texas Tech University. He is married with two kids, and two Australian Shepards.

David Harsh

David works for Microsoft specializing in the Power Platform. While he has been a complete data junkie for the last 10 years, David's Big Data journey began in 2014, when he was the first consultant to successfully stand up a BI connection to Presto at Netflix. Ever since he has spent his time helping some of the largest companies in the world optimize their ability to build BI applications on top of their Big Data clusters including Facebook, eBay, NBC Universal, and American Express. Before Microsoft, he was a consultant and product manager for MicroStrategy.

Kim Kaluba

Kim Kaluba is a Senior Manager for Data Management solutions at SAS. For the past 25 years she has helped organizations derive the most value out of their most valuable asset-data. A data enthusiast and thought leader in data management best practices, Kim believe that data is the heart of any corporate initiative and without the right data at the right time, an organization can't function properly

Monica McEwen

Monica McEwen leads ThoughtSpot's Public Sector team for North America. She has more than 20 years of experience in the analytics and data warehousing space. Prior to joining ThoughtSpot, Monica started Qlik's Federal practice, where she developed a successful go-to-market strategy and built a high performing team that consistently doubled annual revenue.
Monica has spent her entire career in the data management and data analytics space in a variety of technical, sales, and sales leadership roles. Monica is passionate about helping organizations leverage their data to drive organizational efficiency. Monica holds a BA in Business Administration from Colby College where she was President of her class.

Scott Rigney

Scott joined MicroStrategy in 2017 and is the Principal Product Manager for machine learning, data science, APIs, and SDKs. Before MicroStrategy, he worked in risk management and built machine learning systems for predicting business application outages, simulation and resource optimization, and IT system dependency discovery using network graph models. Scott has a master's degree in Data Science from Northwestern University and a bachelor's of science in Finance from Virginia Tech. Scott works at MicroStrategy HQ in Mclean, Virginia and lives in nearby Arlington, Virginia.

Gerard Valerio

With more than 6-years at Tableau, Gerard Valerio currently leads a team of sales consultants and solutions architect in evangelizing Tableau to the U.S. Federal Government and helping customers see and understand their data using Tableau. Mr. Valerio is also an adjunct professor teaching Tableau at Montgomery College in Maryland and is the chief organizer of the Tableau Meet-Up for DC, Maryland, and Virginia. He has built a 20+ year career on data spanning mainframe and mid-range-based reporting systems (referred to as decision support systems and executive information systems) to first generation sub-Terabyte data warehouses in Oracle, Informix, and Sybase front-ended by business intelligence tools like SAP Business Objects, IBM Cognos, and MicroStrategy. Mr. Valerio also worked in the data integration space as a customer and employee of Informatica. His Big Data experience spans working with Terabyte and Petabyte-sized data volumes staged on in-memory columnar databases like Vertica, Teradata, and others to structured/unstructured data residing in Hadoop-based data lakes to log data captured in Splunk. Mr. Valerio holds an Electrical Engineering degree from University of Illinois.