

Adobe® Insight, powered by Omniture®

Accelerating government intelligence to the speed of thought

- 1 Challenges that analysts face
- 2 Analysis tools and functionality
- 3 Adobe Insight
- 4 Summary

Never before has data collection and analysis needed to be performed as quickly as today. More and more data is being collected continuously, but data is only valuable for a limited amount of time—the window of opportunity to glean value from that data is slim. Government agencies and organizations need close to real-time access to data as it's collected to respond effectively to issues that analysis reveals.

Dedicated government agency analysts, including those in law enforcement and military organizations, perform regular intelligence analyses that are critical to the day-to-day operations and stability of the country. These analysts assemble and process data from myriad sources that include open sources like the Internet and closed sources such as intelligence records in government databases.

These analysts must assemble, categorize, analyze, and report on vast amounts of disparate and rapidly changing data in a wide variety of formats. Yet they don't have the luxury of time to perform analysis because they are frequently under pressure to deliver results before the data turns stale. Intelligence analysts analysis because to "make sense of it all" very quickly and convert volumes of raw data into actionable knowledge vital to the nation's well-being.

In addition, intelligence analysts must identify patterns and persons of interest from billions of raw data points. They need to quickly combine cell phone, Global Positioning System (GPS), email, and call detail records to produce valuable intelligence, and they need to explore vast volumes of data without the help of a large IT team creating a new view or cube for every new question that arises.

Challenges that analysts face

Analysts are faced with three fundamental challenges when analyzing data. First, the sheer magnitude of the data that needs to be combed is staggering. In today's digital world, individuals, corporations, organizations, and governments all generate huge volumes of data. Every phone call, credit card transaction, and visit to a website creates a data point that is collected and stored in a different system. This translates to massive datasets for analysts to access and manipulate. Lesser analytic platforms are limited by scalability and fail at addressing these huge volumes.

The second issue is that the data is in a multitude of disparate data types. Transactional data that provides information about events, such as credit card purchases, the details of a cell phone call, or an airline flight booking. There is also streaming data, or a continuous flow of log-type information that follows movement, such as a web log that streams information on what a user is doing—every page visited, every click, every item viewed, and so forth. Lastly, unstructured data is constantly being generated, such as blogs, comments on articles, discussion forums, email, and other socially driven communication and commentary. Combining these disparate data types into one common environment for analysis is a daunting task.

Thirdly, the analyst is faced with the challenge of keeping up with rapidly changing datasets that evolve within minutes or seconds from when the last dataset was captured. Analytic tools need to be able to adapt to the constantly changing data. Analysis tools must be flexible enough to adapt to modifications to the underlying data structure and transition to quickly evolving queries without having to rebuild datasets.

Analysis tools and functionality

A wide spectrum of tools and functionality is available for analysts. However, not all these analytic technologies are created equal—each has its own strengths and limitations and is better suited for certain types of analysis than others. Central to all these tools is the traditional functionality of query, reporting, and analysis. Data quality and data integration tools help to accurately and consistently consolidate data from multiple sources. At the front end, dashboards and other visualization tools help users quickly understand analytic results. Lastly, predictive analytics helps to discover hidden patterns and enable what-if analysis.

Query and reporting

A simple query might access a database to determine how many widgets a company has in its inventory or to calculate total sales and revenue for a defined time frame. Most query tools also provide simple reporting functionality to generate a simple report listing, such as the accrued vacation of all employees, sorted and totaled by department.

Enterprise reporting usually involves high-volume reports that are run on a regular basis. A sales report that shows monthly sales and associated commissions sorted by salesperson and customer is an example of an enterprise report. Report distribution is usually controlled to ensure that each sales manager only sees entries for his or her salespeople.

Query and reporting tools are usually based on relational database technologies that were designed for online transaction processing (OLTP). As such, these tools can read and write data to individual records very quickly but are limited in the types of questions that can be asked of the data. These tools typically have much longer response times to complex queries that involve multiple joins from different data sources.

Multidimensional analysis

With more advanced analysis functionality, users can view data across multiple classifications or dimensions (for example, product, customer, location, or time period). Users can slice and dice the data to look at various combinations. Multidimensional analysis functionality also allows the definition of hierarchies. For example, an analyst can look at sales for different regions and drill down to view the sales in each country for that region. Filtering allows analysts to include or exclude certain dimensions.

Many of these advanced functionalities are available only by preprocessing and pre-aggregating the raw data into a data warehouse or in specialized online analytical processing (OLAP) products that use proprietary databases. These tools pre-aggregate the underlying data and store the results, providing fast response times to even very complex queries involving drill-down, slice-and-dice, and filtering functionality. However, the downside of pre-aggregation is that it takes a long time to preprocess very large datasets into a format that the user can then query. Additionally, if the user has queries of the data that were not predetermined in the design phase, the system has no way of answering the questions without going back to the raw data in the source. It is also difficult and time consuming to implement any changes to the data structure. For example, changing a hierarchy to add another geographical layer between a state and region to adapt to changing environmental requirements is an onerous task.

Visualization

Graphical and visualization techniques complement other analytic technologies because they add visual context to data-heavy reports. With graphical gauges and color-coded symbols, analysts can quickly identify relationships, trends, and exception conditions. Users can typically drill down from visual images to the underlying detail.

Data mining and predictive analytics

Query, reporting, and multidimensional analysis view or analyze what has already occurred, while data mining and predictive analytics allow users to predict what might occur. These technologies use sophisticated mathematical and statistical techniques to find relationships that are not obvious. They help determine which factors are closely related to other factors or outcomes, such as consumer behavior based on income, education, age, and so forth.

"Adobe Insight combines the ease of use and graphical sophistication made popular by the interactive entertainment industry with the raw data processing power of parallelized real-time computing. [This] breakthrough combination of interactive visualization and real-time data analysis capabilities is disruptive technology that can redraw the capabilities of government and private sector alike."

Gilman Louie, former president, In-Q-Tel



Adobe Insight

Adobe Insight is a different solution that employs unique technological innovations to answer ad hoc queries in real time. By using randomized statistical techniques, all system queries are progressively refined until completion. Queries are instantly answered with an approximation, but the results become increasingly more accurate or near exact as the numbers are continuously updated over the course of a few seconds or minutes.

Often, analysts don't need exact answers—to discover an important trend or pattern and, therefore, shouldn't have to slow down to wait for them. This unique feature provides an unmatched ability to explore data and draw conclusions while the task at hand is still fresh in the analyst's mind.

Unlike other query and analytic tools, Adobe Insight does not require pre-aggregations or preprocessing of data. As a result, there are no predefined questions, predefined filters, nor limitations on how deep analysts can drill down into data while still benefiting from real-time responses.

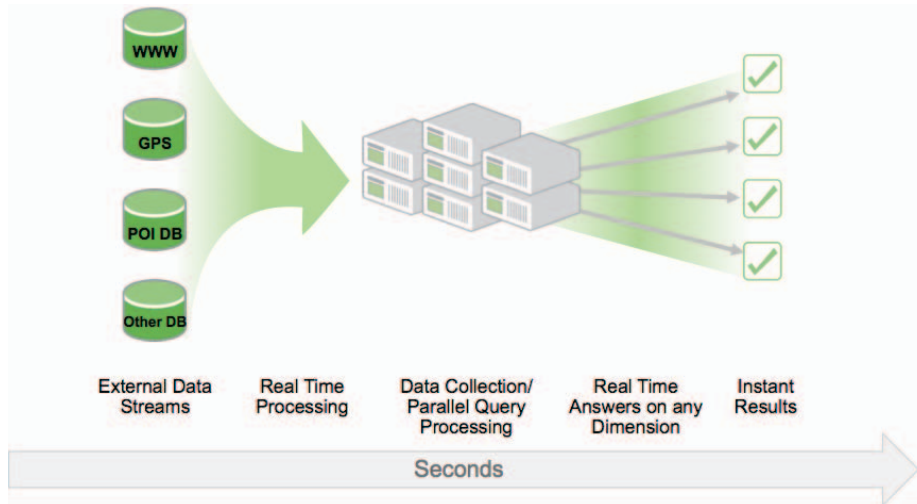
Powerful and flexible data extraction and transformation

Adobe Insight provides a visual mapping of all data and the transformations at every step. Backed by a high-performance scalable platform, terabytes of data can be brought into the system in near real-time data streams or via batch processes. Adobe Insight can load and process datasets with billions of rows of data, and a flexible set of transformations allows analysts to build an array of custom processes to synthesize practically any structured data type. For example, analysts can filter out extraneous data, correlate events at different times, or incorporate secondary sources of data.

High-performance data processing

While Adobe Insight excels in utilizing advanced analysis capabilities to identify patterns and trends in datasets with long-term historical data, it also has powerful capabilities to analyze data in real time. Unlike other systems that provide answers nightly, Adobe Insight is always up to date—a set of graphs on an analyst's screen will update continually as new data streams in. This continuous flow of information provides the ability to react quickly and eliminates the need to rerun a time-consuming query again to get up-to-date numbers.

Adobe Insight uses an innovative data storage schema to ensure that every query includes all data that is currently stored in the system. Because the schema is not a relational database and does not use SQL, the results are always correct and are not based on a sampling or aggregation of the data.



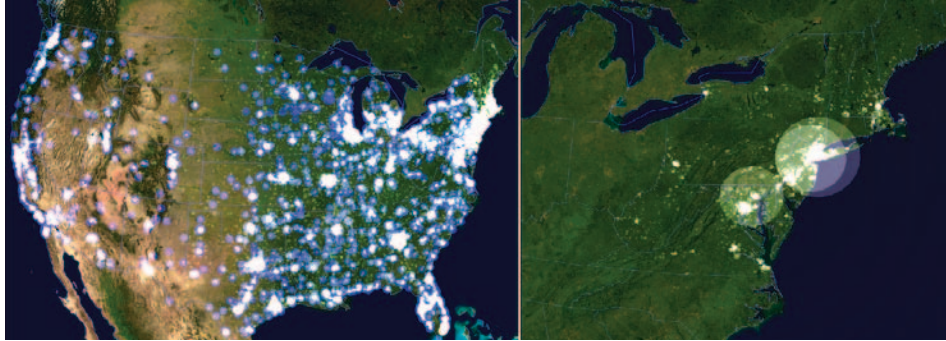
Rapid data discovery for very large datasets

Adobe Insight is designed to hold query time relatively constant, regardless of data size. Whether the dataset is a few hundred gigabytes or many terabytes, the analyst receives an instant approximation response for every query and a 100% completed query in minutes, if not seconds.

State-of-the art visualization

Powerful back-end technology is delivered via a visually stunning and graphical interface that encourages exploration and makes it easy to present complex queries graphically. A variety of visualizations, including 2D and 3D process maps, geospatial analysis, and cluster diagrams, provide analysts a rapid visual representation of data associations and links.

Adobe Insight includes a number of tools to rapidly identify time-dependent events, in combination with geospatial analysis to provide targeting solutions. One of these is Visual Geography—a geographical imagery and analysis visualization tool that shows location data in a topographical display. Data can be layered by city, longitude and latitude coordinates, country, define DMA, and IP domain. Additional layers can be created via lookup tables.



Geospatial analysis with heat maps

Adobe Insight also provides the capability to annotate these visualizations to produce high-quality publishable reports that can be distributed to others via email, the reports that can be distributed to others via email, the Web or hard-copy..

Summary

Faced with the daunting task of having to bring together volumes of disparate data for swift analysis, intelligence analysts have a number of analytic technologies to choose from. These different technologies possess both strengths and limitations, depending on the analysis needs and types of data involved. However, for government intelligence analysts who require near real-time results, rapid data integration, and sophisticated analysis, many of the analytic technologies fall short in meeting their rapidly evolving needs. Adobe Insight is a solution that enables intelligence analysts to glean valuable information from vast volumes of constantly changing data and accelerates intelligence to the speed of thought.

Near real-time results

- Instant query approximation response with 100% query completion in seconds or minutes
- Responses to changes in filters, segments, or dimensions in seconds, not in hours or days
- Constant query response time, regardless of data size

Rapid data integration

- Virtually no limitation to dataset size
- Flexible data schema that allows the integration of new data sources for analysis
- Processing of incremental data into the system from log files and some applications
- Complete retransform for new data sources in 12 hours

Easy extraction and transformation

- Easy extension of the data schema for analysis
- Visibility into processing steps for data validation

- Automated system checkpoints for faster data reprocessing and recovery
- Extensive data transformation options , including parsing, appending, merging, and categorizing of data for processing and extraction

Geospatial analysis

- Analysis of patterns and events within topographic visualization
- Ability to create multiple layers to analyze dependencies between activities, events, groups, or other dimensions

Time-series event analysis

- Latency analysis to rapidly identify patterns and correlations between events and times
- Ability to link latency analysis to geospatial views and layers

N-dimensional analysis

- Automatic correlation of all data
- Shared dimensions, metrics, and analysis workspaces to collaborate with colleagues
- Annotation callouts for data clarification and explanation
- Spreadsheets to build scenarios
- Navigation from high-level trends to contributing data at its most granular level
- Data correlations and confidence metrics to guide user analysis and validate conclusions

Lowest level analysis

- Ability to spot high-level trends and then drill down to the lowest level of granularity
- No data summarization: every query run against every data point in the dataset

For more information

For more details about Adobe Insight, powered by Omniture, visit www.omniture.com.

