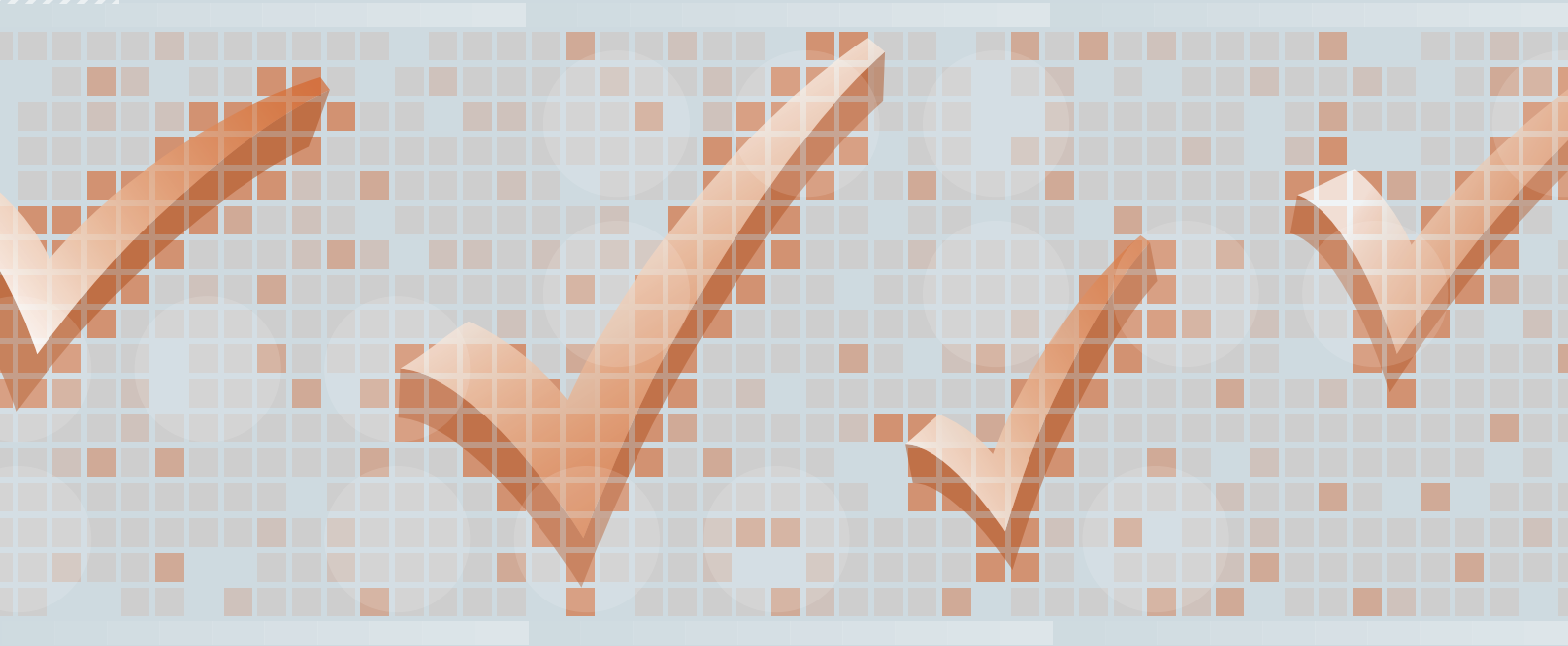
TDWI RESEARCH

TDWI CHECKLIST REPORT

OPERATIONAL INTELLIGENCE

Real-Time Business Analytics from Big Data

By Philip Russom



Sponsored by

splunk>

tdwi.org

tdwi

AUGUST 2013

TDWI CHECKLIST REPORT

OPERATIONAL INTELLIGENCE

Real-Time Business Analytics from Big Data

By Philip Russom

1201 Monster Road SW, Suite 250
Renton, WA 98057

T 425.277.9126
F 425.687.2842
E info@tdwi.org

tdwi.org

TABLE OF CONTENTS

INTRODUCTION

As a result of the big data phenomenon, data volumes and the diversity of new data sources are exploding around us. Yet, most organizations are missing the analytic opportunities of big data because they are still focused on gleaning insights from structured data via traditional tools for business intelligence (BI) and data warehousing (DW). A brave new world of insight awaits organizations, and the path to it is through the exponentially growing volumes of unstructured and semi-structured data, especially from new sources such as machines, sensors, logs, and social media. One of the challenges is that traditional BI/DW tools were not designed for these new data sources and data types. BI/DW tools are certainly not going away, but there's a need to complement them with new technologies for the new sources of big data, and operational intelligence supports this growing need.

Defining Operational Intelligence

Operational intelligence (OI) is an emerging class of analytics that provides visibility into business processes, events, and operations as they are happening. The practice of OI is enabled by special technologies that can handle machine data, sensor data, event streams, and other forms of streaming data and big data. OI solutions can also correlate and analyze data collected from multiple sources in various latencies (from batch to real time) to reveal actionable information. Organizations can act on the information by immediately sending an alert to the appropriate manager, updating a management dashboard, offering an incentive to a churning customer, adjusting machinery, or preventing fraud.

Use Cases for Operational Intelligence

The point of operational intelligence is to gain insight into new data sources so that business opportunities, organizational threats, and performance issues are detected and addressed as soon as possible, thereby enabling reactions that leverage or correct a given situation. Real-world implementations of operational intelligence monitor and analyze business activities to give a wide range of users the real-time visibility they need to see a problem or opportunity, make a fully informed and fast decision, and then act accordingly:

- See a product recurring in abandoned shopping carts. Run a promotion to close more sales of that product.
- Perform capacity planning for mobile networks as new high-bandwidth services are introduced. Improve customer experience.

- See a social media sentiment or pattern. Direct it or correct it as it evolves.
- See potentially fraudulent activity while it's in process. Prevent action and proactively mitigate impact.
- See that your utility grid has excess capacity. Sell that capacity while it's available.
- Understand customer behavior in real time across channels—Web, mobile, and social.

OI's Fit in the Analytics Landscape

Time-sensitive business processes. The pace of business continues to accelerate such that speedy decisions based on fresh information have become a competitive advantage. This is true of many modern business practices, from operational business intelligence to business performance management, and from just-in-time inventory to facility monitoring. OI now joins these technologies, but with support to process semi-structured and unstructured data in true real time.

Greater speed and agility. Business growth is sustained on a daily basis by fast decisions made from fresh and complete information, with insights based on rapidly growing volumes of data from new sources. This is true whether you're preventing a customer from churning or identifying online fraud. OI provides insight into new sources so managers can make informed business decisions rapidly and with more complete information.

Complementing business intelligence and data warehousing. BI/DW originated to support business decision making from structured data sources and to provide analytics from a historical perspective. OI complements BI/DW by providing insight into new unstructured and semi-structured data in real time. It handles big data in ways BI/DW cannot.

Complete views of business entities and situations. Data of different latencies tells managers different things about a business entity, such as a customer, transaction, or business process. OI can correlate the real-time analysis of streaming big data and machine data with historical data (typically managed in a data warehouse or similar database) to present a complete view.

The Four Primary Capabilities of Operational Intelligence

One way to think of operational intelligence is that it's the combination of multiple leading-edge technologies that provide unparalleled visibility into a business, as summarized in Figure 1.

1. **Real-time data handling.** Capture and process data in seconds or milliseconds from multiple sources (both traditional and new), including streaming data, event streams, and message queues.
2. **Advanced analytics.** Link and correlate related events, regardless of their origins or latency, to discover problems or opportunities that merit immediate attention.
3. **Big data and machine data.** Ingest and analyze multi-terabyte data volumes daily and tens of terabytes to petabytes of historical data, ranging from relational data to human language text, with an emphasis on real-time machine data.
4. **Business visibility.** Provide complete views of business entities and situations based on both real-time and latent data, presented in terms that are business friendly and actionable.

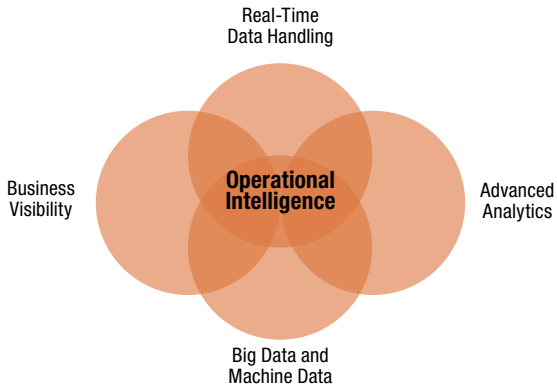


Figure 1. *The four primary capabilities of operational intelligence*

This TDWI Checklist Report drills into the many technologies and capabilities needed to make operational intelligence possible for a technology team and successful for a business.

NUMBER ONE

EXECUTE ANALYTICS IN TRUE REAL TIME.

OI accelerates insight into seconds and milliseconds. Originally the term *real time* literally meant that all processing (from event reception to system response) executes within seconds, maybe milliseconds. We have become somewhat sloppy in our use of the term, in that we sometimes say "real time" when the fetching and delivery of data takes minutes or is executed every few hours. However, note that OI operates in *true* real time by receiving event messages and other incoming data, processing them, making correlations, and assessing correlations within seconds or milliseconds.

True real-time analytics are critical for many analytic applications. For example, true real time is required for the continuous monitoring of a process, network, or facility. Consider that a few minutes on a rail or truck schedule can affect customer satisfaction, and they can add up, amounting to major delays. Additional time to access content on a mobile device can lead to serious customer satisfaction issues and unwanted churn.

OI makes correlations across real-time data, plus data sets of other latencies. A confluence of events may include events that just happened in one department plus those that happened weeks or months ago in other organizations. For example, a potentially fraudulent insurance claim is revealed when the same person, location, or vehicle is linked across multiple claims, occurring at different times.

Business rules can automatically take action based on analytic outcomes. Though OI can return a result in milliseconds or faster, no human can respond that quickly. For maximal response, many OI solutions support business rules and alerts that can make decisions and spawn a software job that takes action immediately. Obviously, business rules are equally useful in situations that are not as time sensitive.

OI complements BI/DW by operating in true real time. The popular practice of operational BI involves querying data that is a few hours or minutes old, representing recent business events and performance, as seen from structured data. Standard reports are generated from data that's even older. Although the use of historical data has indisputable value, it does not represent the complete, up-to-date picture. Fully informed business decisions made in brief time frames require that real-time data be presented alongside historical data, if you're to get full value from exponentially growing new data sources. Hence, BI and OI complement one another by supporting different time frames, as well as different types of data.

NUMBER TWO

ADOPT ADVANCED ANALYTICS AND CORRELATE NEW DIVERSE DATA SOURCES.

Operational intelligence depends on advanced technologies to enable real-time analytics.

Correlations across multiple data sets. OI excels with multi-data-source correlations, even when the sources are an eclectic mix of new ones (streaming data, machine data, social media data, NoSQL data platforms such as Hadoop) and traditional enterprise sources (enterprise applications, relational databases). Unlike solutions exclusively for structured data, OI also deals with data sets that are schema free, metadata free, and evolving structurally. That's a long list of old and new data sources—each with its own requirements—and a mature OI platform must support them all.

Streaming data. As more organizations move deeper into monitoring operations in real time, there's a growing need to quickly capture and process events expressed as messages or events in a stream that generates and delivers data almost continuously. At the same time, the number of data streams is increasing because many new forms of big data are communicated via streams (especially sensor data and machine data). In many ways, analytic correlation across multiple data streams is the epitome of operational intelligence.

Event processing. Technologies for event processing have been around for years, but most are designed to monitor only one stream of events at a time. Even if users monitor multiple streams, they end up with multiple siloed views into real-time business operations. Operational intelligence creates a more unified view that correlates events from multiple streams and other information sources, arming businesses with better insights.

Actionable analytic outcomes. OI's combination of leading-edge analytic technologies helps users and their organizations act immediately on the results of real-time analyses:

- Understand customer behavior so you can improve the customer experience
- Monitor and maintain the availability, performance, and capacity of interconnected infrastructures, such as utility grids, computer networks, and manufacturing facilities
- Identify compliance and security breaches, then take action to prevent future ones
- Spot and stop fraudulent activity, even as fraud is being perpetrated
- Evaluate sales performance in real time and take measures to achieve quotas

NUMBER THREE

DEMAND PLATFORM SPEED AND SCALABILITY FOR HANDLING BIG DATA.

Operational intelligence (OI) handles data in extreme environments, where big data volumes are counted in terabytes and real-time data is generated in continuous streams. Therefore, platform speed and scalability are critical success factors for OI, despite the many challenges that big data's size and diversity present:

Big data is mostly defined as very large data sets. A telco may process terabytes of data daily via operational intelligence, coming from their content delivery networks (CDNs), set-top boxes, call detail records, and broadcast operations, among other sources.

Big data gets big because it comes from many sources, in many data types and formats. This includes new frontiers, such as sensor data and machine data, plus other frontiers such as unstructured data (human language text) and multi-structured data (XML, JSON, CSV). Traditional data types are still with us, too, in the form of structured data, relational data, and record-oriented flat files.

Some forms of big data stream in real time. OI's real-time analytics and business monitoring depend on correlations across many sources of big data that are inherently streaming, typically clickstreams from Web servers, machine data, data from devices, CSV, events, transactions, and customer interactions.

High counts of small messages or events can add up to big data. A successful solution for OI must do several things with streaming data. OI must capture each event from a stream, separate events of interest from noise, make correlations with other streams and databases (including data warehouses), react to some events in real time, and store events for offline analytics. Technology for OI excels with high performance for each of these atomic units of work, so that the aggregate performance of the overall OI system is fast and scalable.

NUMBER FOUR

SEIZE THE MANY BUSINESS OPPORTUNITIES OF MACHINE DATA.

Nowadays, just about any machine, device, living organism, product, or building can be fitted with sensors for data collection via some kind of network. Obviously, many machines are networked by nature, such as computers, mobile devices, manufacturing robots, medical devices, point-of-sale systems, ATMs, and slot machines. All of these machines and sensors periodically or continuously generate and transmit so-called "machine data" about their current condition and recent events. This makes machine data a large and rapidly growing segment within big data, and the business opportunities for leveraging it are equally large.

Machine data is inherently real time. Machine data is usually generated one small message at a time, as a record of an event or condition that just occurred. Depending on the machine involved and how it's configured, the message may be transmitted immediately, so software, users, and other machines are informed and up to date. OI solutions can capture and analyze these messages to enable real-time business visibility and reaction.

Machine data should also be stored. Though broadcast in real time, most machine data today is captured in log files, as seen in the logs of Web servers and other enterprise applications. As a historic record, machine data is rarely updated or deleted; this characteristic makes large stores of unchanging machine data ideal for offline analytics. Furthermore, stored machine data provides a historic context for the most recent machine data generated in real time. Ideally, an OI solution should make correlations across data of various vintages or latencies.

There are *de jure* data and protocol standards for common types of machine data. An OI platform should support most of these out of the box. Examples include clickstream data, server logs, CSV, Extensible Markup Language (XML), JSON, Java Messaging Service (JMS), supervisory control and data acquisition (SCADA), call detail records (CDRs), and other log formats. Furthermore, many popular vendor products have similar but proprietary formats.

Machine data can be unique. For example, most robots that assemble products in manufacturing are one of a kind because they're designed to install a single component in a specific product. Hence, data collected via sensors on that robot typically has a proprietary format that's unique. OI solutions must flexibly enable developers to create models and analytic processing for any machine data, no matter how proprietary and unique it is.

Some machine data is generated intermittently, not continuously. For example, most railroads nowadays have multiple sensors on each railcar, monitoring each for extreme temperature, vibration, vertical orientation, and so on. However, data from the sensors is only collected when the car passes through a rail yard or station to avoid the expense of deploying radio-frequency receivers all along a rail route. When data from a railcar is received, an OI solution can determine within moments whether the car needs maintenance.

GPS is an important form of machine data. It's obvious that all physical assets and resources have a location; yet relatively few organizations have developed geographic dimensions for their models of customers, products, partners, mobile assets, and offices. Key to populating geographic dimensions with actionable data are the many devices that can transmit or record GPS coordinates, including sensors within cell phones, vehicles, and product packages. When leveraged by OI technology, GPS machine data reveals where customers are when they make certain purchases, which of your trucks is closest to the location where one is needed, and what route products took from your manufacturing facility to retail shelves.

Machine data contributes to 360-degree views for a more accurate picture. For example, the profitability of most customers is high when calculated from sales data, but many customers turn out to be unprofitable when you correlate data from non-sales customer touch points and channels. The money your customer spends is easily burned up when they return products, demand excessive phone support, require field service for broken products, and are located in an isolated area that drives up shipping and service costs. OI can make these correlations by tapping Web logs, application logs, business process management logs, call detail records, shipping manifests, GPS coordinates, mobile device logs, and social channels. Hence, tapping diverse machine data sources via OI improves 360-degree views so that no matter what you need to do with your customers, you have the complete and up-to-date information you need to make that effort successful.

NUMBER FIVE

EASILY EXPLORE AND STUDY NEW DATA FOR ACTIONABLE INSIGHTS.

Exploration is an important path to business value from big data and streaming machine data. After all, most big data and machine data today come from sources that are new to an organization, and therefore have not been studied much. Exploring and studying new data leads to the discovery of new facts about the business, which in turn leads to actionable insights.

For example, in recent years, a number of trucking companies in the U.S. have added various types of sensors to their fleet vehicles. The machine data that now streams from the trucks has led to more efficient route designs, safer driving habits, and substantial cost reductions via reduced fuel consumption.

Data exploration is also a prelude to developing new applications for operational intelligence. Retailers that manage their own truck fleets have correlated inventory data with truck manifest data and truck location data. That way, when a store suddenly has low stock for a profitable product, merchandisers can restock the store "just in time" from a nearby truck, not just from a regional warehouse, which was the older practice.

Given the size and diversity of today's big data, you need OI solutions that are built for exploring a wide range of enterprise data:

Search technology for exploring diverse data types. Data exploration should be as easy as Google, it should parse data of many formats and structures, and it should allow any open-ended question, not just those confined to a predefined data model. Search technology satisfies all of these requirements.

High ease of use for user productivity. This is critical because some users are business people who need to see the data for themselves but through a business friendly view. Ease of use accelerates technical developers' productivity, too.

Short time to use and business value. A data exploration capability with high ease of use enables a wide range of uses to get acquainted with data quickly, yet keep digging deeper over time for new business facts and the opportunities they lead to.

Query capabilities in support of data exploration. Technical users, in particular, depend on query capabilities to find just the right data and structure the result set of a query so it is ready for immediate use in analytic applications.

NUMBER SIX

COMBINE STREAMING DATA WITH STRUCTURED DATA.

Enrich streaming data with structured data. Although visibility into streaming data by itself is extremely valuable to the business, its value can be further enhanced by combining it with data that already exists in structured databases and data warehouses. For example, you could combine an insurance claim ID in streaming data from the applications that support claims processing with additional profile data from a customer master. This helps you understand real-time claims processing analytics in the context of specific customer attributes and profile information.

Present streaming data and historic data side by side. In the user interface of an OI application, the latest value of a parameter culled from streaming data should be compared to previous values of that parameter, at meaningful time periods (say, the same time of day yesterday, last week, and last month). Likewise, the latest value should be compared to the average, adjusted for seasonality. This way, the end user is fully informed about the tracked entity's performance and is therefore well equipped to make a good decision.

Develop thresholds for all entities tracked via streaming data. Nowadays, most chemical manufacturing plants are monitored online via OI technologies and most adjustments to the manufacturing process are made via software. If the temperature reading from a sensor is outside a prescribed threshold, the software automatically executes a script that adjusts the machinery. If, say, vibration readings are high on a device, the software alerts a maintenance engineer to examine the device in person or via a surveillance camera.

Develop business rules for interpreting streaming data. When a streaming event says that a customer deactivated service a moment ago, a business rule should automatically look up that customer's profile, which includes metrics for profitability, lifetime spend, loyalty, etc. Based on that information, the business rule can calculate whether to make an incentive offer, asking the customer to reinstate service. For such practices to work, looking for a customer ID in the stream is key for combining stream data with other enterprise data.

NUMBER SEVEN

LOOK FOR SOLUTIONS CONDUCIVE TO OPERATIONAL INTELLIGENCE.

As we've seen throughout this report, the technical requirements of operational intelligence are fairly extreme, such that very few user organizations could build their own. Hence, TDWI recommends that users turn to vendor tools and platforms for OI, instead of attempting a homegrown solution. Yet, not just any vendor tool can achieve the demanding requirements of operational intelligence. Based on the discussion of this report, here is a concise list of critical features and functions users should look for when evaluating technologies for operational intelligence:

True real-time operation, in seconds and milliseconds. After all, this differentiates OI from similar technologies and it's the leading value proposition for OI.

Analytics that correlate events from multiple events and other data sources. Analytic correlation across multiple real-time streams and latent data sets is the epitome of OI.

Fluency for machine data and other forms of streaming data. In many ways, machine data is the killer app for OI, but only when an OI technology can ingest data from multiple sources and combine that data with relational data and other enterprise data.

Flexible data acquisition. By nature, an OI tool integrates with multiple systems of diverse types, so it can quickly bring on board and acquire traditional structured data, as well as new forms of big data, streaming data, unstructured data, schema-free data, and data with an evolving structure.

Proven scalability and high performance. Most OI tools, being fairly new, were built for extreme environments, dominated by big data and streaming data. Be sure to check a vendor's references to confirm that a tool scales and performs as advertised.

Capabilities for data exploration. Getting to know big data and machine data is a critical first step to developing effective solutions for them. Look for business-friendly data exploration capabilities that support both search and query access to data.

Complement BI/DW infrastructure with OI. OI won't replace a mature BI/DW implementation, and it's unlikely you can stretch BI/DW to the extreme real-time performance of OI. However, the two work well together because they serve different purposes.

To gain competitive advantage in today's environment, organizations need to expand their data analytics capabilities beyond structured data to new data sources. BI/DW professionals seeking analytic support for big data and machine data should consider extending their software portfolios to include technology for OI. OI users should integrate their solutions with a data warehouse as an additional analytic store for machine data.

ABOUT OUR SPONSOR

splunk>

www.splunk.com

Splunk Inc. (NASDAQ: SPLK) provides the engine for machine data™. Splunk® software collects, indexes, and harnesses the machine-generated big data coming from the websites, applications, servers, networks, sensors, and mobile devices that power business. Splunk software enables organizations to monitor, search, analyze, visualize, and act on massive streams of real-time and historical machine data. 5,600 enterprises, universities, government agencies, and service providers in over 90 countries use Splunk Enterprise to gain Operational Intelligence that deepens business and customer understanding, improves service and uptime, reduces cost, and mitigates cybersecurity risk. Splunk Storm®, a cloud-based subscription service, is used by organizations developing and running applications in the cloud.

To learn more, please visit www.splunk.com/company.

ABOUT THE AUTHOR

Philip Russom is director of TDWI Research for data management and oversees many of TDWI's research-oriented publications, services, and events. He is a well-known figure in data warehousing and business intelligence, having published over 500 research reports, magazine articles, opinion columns, speeches, Webinars, and more. Before joining TDWI in 2005, Russom was an industry analyst covering BI at Forrester Research and Giga Information Group. He also ran his own business as an independent industry analyst and BI consultant and was a contributing editor with leading IT magazines. Before that, Russom worked in technical and marketing positions for various database vendors. You can reach him at prussom@tdwi.org, @prussom on Twitter, and on LinkedIn at linkedin.com/in/philiprussom.

ABOUT TDWI RESEARCH

TDWI Research provides research and advice for business intelligence and data warehousing professionals worldwide. TDWI Research focuses exclusively on BI/DW issues and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of the business and technical challenges surrounding the deployment and use of business intelligence and data warehousing solutions. TDWI Research offers in-depth research reports, commentary, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

ABOUT THE TDWI CHECKLIST REPORT SERIES

TDWI Checklist Reports provide an overview of success factors for a specific project in business intelligence, data warehousing, or a related data management discipline. Companies may use this overview to get organized before beginning a project or to identify goals and areas of improvement for current projects.