

SOLUTION BRIEF

Data Center: Big Data and Analytics
Improving Products/Services/Efficiencies



Create Real-time Business Value with Advanced Analytics

Derive more accurate and contextually relevant big data results than ever before with Intel® architecture-optimized real-time analytics

This solution brief describes how to solve business challenges through investment in innovative technologies.

If you are responsible for...

- **Business strategy:**

You will better understand how a next-generation analytics solution will enable you to successfully meet your business outcomes.

- **Technology decisions:**

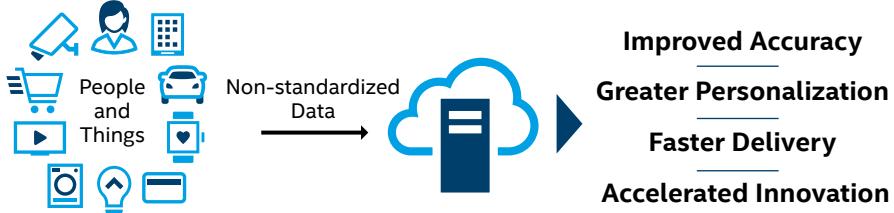
You will learn how a next-generation analytics solution works to deliver IT and business value.

Executive Summary

Industry research has characterized technology investments in big data and data science as strategic. Some organizations seeking tangible benefits from big data analytics have found it hard to achieve real-time relevance from the use of a broad array of data sources. By placing Intel's new breed of big-data-tuned components at the foundation of servers while implementing a more optimized mix of applications and analytics frameworks in the software stack, organizations can gain unprecedented results from their analytics efforts. Users can realize greater levels of solution accuracy, utility, speed, and innovation—providing more stability, security, and profitability.

Across nearly every field, both public and commercial, analytics can provide the intelligence needed to help IT decision makers and executives glean critical business insights faster than ever before. Such decisions can mean the difference between stability and economic collapse, security and exploitation, or health and sickness. Moreover, the speed at which data is ingested and processed often matters as much as the number of data sources and the amount of data being crunched. Legacy analytics platforms often don't provide an ability to ingest and analyze information at Internet speed from disparate data sources. Intel® architecture-optimized real-time analytics offers users the ability to embrace disparate data sets in a nearly instantaneous manner, enabling a quicker path to more intelligent, actionable results.

Intel® Architecture-optimized, Real-time Analytics



Author

Melvin Greer

Principal Engineer and Director
Data Science and Analytics
Intel Federal

Figure 1. Intel® technology-driven, next-generation analytics have the processing performance needed to combine multiple data sets and yield mission-critical business benefits in near real time.

Solution Benefits

- **Faster ingestion of big data.** Task-optimized hardware and high-performance analytics software enable quicker results.
- **Higher ingestion bandwidth and capacity.** Incorporates more varied data sets, leading to deeper analysis.
- **Higher-performing analytics platform.** Yields more individualized, granular results so that data outcomes can better service specific user needs.
- **Potential savings in the billions of dollars.** The deeper assessment and faster performance of next-generation analytics can help recoup tremendous amounts in annual losses—savings that can be passed on to individuals.

Business Challenges: Too Much Raw Data, Not Enough Insight

IDC estimates that by 2020, every person in the world will generate 1.7 MB of data per second.¹ That's enough to fill the average hard drive in less than 10 days.² Whether a government is looking to shape policy using big data to improve citizens' lives or an insurance company is looking to minimize waste and fraud, the amount of data that must be gathered and fed into analytics systems grows at a staggering pace. Meanwhile, organizations must be able to simultaneously ingest these massive data sets from potentially multiple sources in the shortest amount of time.

Consider that there are more than 16 different U.S. intelligence agencies (in addition to hundreds of foreign counterparts) that must function separately yet share data for millions of people. The speed at which collaboration occurs could dramatically affect world events. There are nearly 800,000 healthcare companies in the United States alone;³ imagine if all of them could seamlessly and securely share data in ways that allowed for faster, more accurate diagnoses (without requiring a new set of forms to be filled out at every new office). Or ponder the possibilities of public utilities joining data sets with personal activity trackers and weather services, home monitoring companies, and other similar groups to devise the smartest possible models for reducing home energy consumption and carbon emissions.

Today, most large organizations are challenged with mass-scale data ingestion and are seeking the platform horsepower to keep up with the constant influx of new data. Even after conquering that challenge, the issue remains of how to combine multiple data sets in ways that quickly yield useful conclusions and do so while meeting privacy regulations. Solutions are emerging—with the need for those solutions underscored by a wide array of industry facts:

- By 2020, 40 percent of new investment in business intelligence and analytics will stem from predictive and

prescriptive analytics, which, in turn, hinge on more effective ingestion and processing.⁴ Organizations increasingly feel the need for improving their big data efforts, but the systems purchased to shoulder those loads must keep pace.

- Cross-organizational analytics can play a significant role in alleviating social inequality and creating new opportunities. For example, according to Whitehouse.gov and the Consumer Financial Protection Bureau, 30 percent of consumers in low-income neighborhoods are "credit invisible" and another 15 percent are "unscorable."⁵ Both conditions stem from an inability to compute a relevant credit history for these groups. This blocks certain demographic groups from access to opportunities for financial growth. However, by coordinating disparate databases, from public records to cell phone use, better big data analytics can open the door to credit help for these underserved groups.
- Blue Cross Blue Shield notes that fraud costs roughly USD 68 billion annually in the United States.⁶ Insurance companies coordinate with the FBI and the Inspector General for Health and Human Services, as well as state and local police. With such a massive target to tackle, these groups need all the cross-database analytics help they can get.
- Genomics (DNA sequencing and analysis) stands as one of the highest growth (and highest need) areas for big data analytics. Some estimates state that data growth in this space will increase by 1 million times by 2020, and others have pegged the daily data accumulation rate as doubling every seven months.⁷ All this data should lead to some phenomenal advances, such as the ability for pharmaceutical providers to create medications customized to react optimally with a patient's unique DNA. Though the amount of analysis required for such precision appears staggering, with highly scalable analytics systems with which to ingest and process data, bioinformatics scientists could perform this kind of breakthrough research.



Figure 2. Due to bandwidth and capacity limitations, legacy analytics solutions often either ignore large swaths of valuable source data or deliver results too slowly to be practical.

In every case, the problems boil down to the process of putting raw data in and pulling insights out. The challenge lies in constructing better tools for both sides of that analytics equation so that organizations can realize better results more quickly.

Bigger Data, More Powerful Applications

Applications for analytics already pervade our world. From e-tailer recommendation systems to speech recognition to delivery route planning, big data-driven analytics already touch most aspects of our daily lives. However, as the following use cases indicate, analytics platform advances are leading to incredible breakthrough capabilities and models.

Broader Sourcing and Cyber Intelligence

Cyber intelligence focuses on risk assessment and the identification of internal and external vulnerabilities and threats. Legacy solutions tend to be rule-based and deterministic: if this event happens, then trigger that action. The approach is very reactive and largely based on event data gathered from sources within an organization.

In contrast, next-generation analytics can employ a far broader set of information inputs. Cyber intelligence should be able to pull from social media accounts, financial histories, employment records, and all manner of other data pertinent to identifying individual risks. Clearly, this involves ingesting and processing a greater magnitude of data—but, in combination with the right frameworks and algorithms, it also allows for predictive intelligence that can provide actionable risk signals far in advance of what legacy systems are able deliver.

Less Fraud in Healthcare

Imagine a patient who visits the doctor for a broken wrist. The doctor prescribes an opiate painkiller, which the patient obtains at Pharmacy A. The patient might also have the prescription filled at Pharmacy B. If he or she has multiple insurance policies, the patient could bill two different insurers for the same prescription. Alternatively, the person might go to another doctor for a different condition and be prescribed a medication that could cause a negative reaction with the painkiller. All this fraud, waste, and potential tragedy could be avoided with the help of a secure analytics platform able to combine these siloed data sets and deliver real-time alerts (or block transactions) whenever transgressions were attempted.

Trusted Analytics Platform (TAP) is a software toolkit for building analytics applications; it establishes a software framework for analysis, modeling, and algorithmic processing of raw data. Recently, a solution integrator (SI) allied with Intel and the open source TAP software community to receive training on how to use TAP software tools and optimize

applications for making the most of Intel® technology-based hardware features. The SI created intellectual property around algorithms specifically designed for clinical trial data sets and wove this into the TAP software framework, ultimately delivering successful, fraud-reducing proof-of-concept demonstrations to government organizations including the Veteran's Administration, Centers for Disease Control, U.S. Army, and National Institutes of Health. Interestingly, SIs now find themselves with multiple revenue models for the insight gained from analytics work.

Solution Value: Deeper Insight, Faster Results

Companies deploying a next-generation big data analytics solution can expect benefits that fall into four primary categories:

- **Improved accuracy.** This might take the form of more customized, accurate drugs or preventing fraud in financial markets. Accuracy means making sure that data falls within acceptable parameters and flagging instances when it does not. Potential benefits for organizations abound here, but it's important to note that individuals also stand to gain in ways from shorter travel times to lower hospital costs.
- **Greater personalization.** With the ability to ingest more data from more sources comes the realization of greater context. Providing context for hyper-relevant data against a backdrop of highly correlated yet disparate data sets can paint a richer picture and allow systems to make much more accurate and meaningful recommendations. This value applies equally to both government and commercial environments.
- **Faster delivery.** Using an integrated, application-optimized platform to handle larger, multi-source analytics loads can approach real-time performance, thus speeding time to value.
- **Accelerated innovation.** When an organization innovates, it either improves existing operations or devises new ones, all in the pursuit of better achieving organizational goals, such as customer satisfaction, profitability, public health, preservation of resources, and so on. By processing larger, more diverse data sets more quickly, enterprises and agencies can innovate intelligence offerings and aim for even higher outcomes.
- **Forward-looking insight.** As organizations gather more data and enhance their ability to analyze it quickly using techniques like artificial intelligence and machine learning, they can move beyond 'rear view mirror' or even real time insight to predict and act upon what will happen next. In competitive retail environments or time-sensitive clinical settings, this ability to pre-empt issues and prepare for opportunities can be critical.

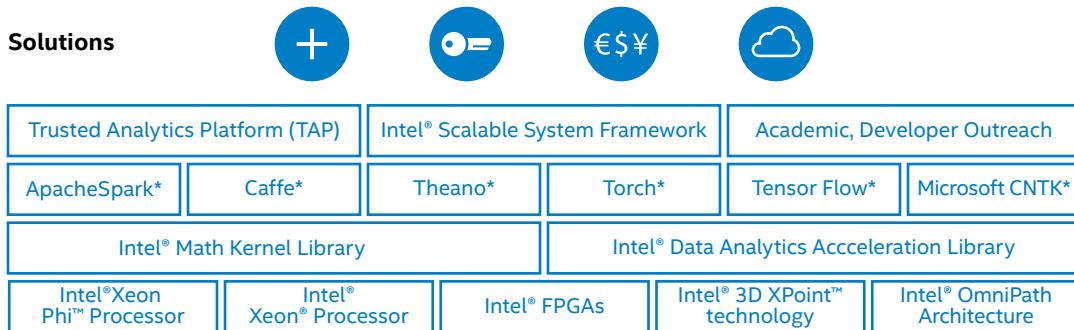


Figure 3. Intel® technology forms the hardware foundation of next-generation analytics platforms while also helping with advances in analytics software tools.

Solution Architecture: Optimizing High and Low

In essence, big data ingestion flows through a four-level solution stack (see Figure 3). At the top, applications serve as the front end through which data sets enter. Below this, analytics software does the hard work of crunching all that data. Next, data platforms—operating systems, virtualization platforms, and the like—provide the solution's software foundations. And finally, a hardware platform specifically geared to the needs of high data throughput and analytics workloads provides the physical means to power and accelerate all software operations. Every analytics solution must work through these four layers, and factors such as scalability, efficiency, and total cost value will depend on how well this stack is optimized for analytics tasks in general and even the specific workloads in question.

Currently, most of the innovation improving analytics performance is happening at the top and bottom of this stack. Advances in processor capabilities, memory architectures, and networking bandwidth all continue to open the throttle on analytics throughput. At the application level, one of the goals must be to find better ways to weave together the dozen or more software tools that collectively comprise an analytics solution. Intel met this problem head-on by spearheading development of TAP, which became an open source software effort in August 2015. As a key contributor to the TAP effort, Intel continues to work on optimizing the toolkit to make the most effective use possible of Intel® technology-based hardware resources while keeping software open and agnostic for industry-wide compatibility.

At the stack's foundation stands a reliable, massively scalable, high-performance infrastructure able to handle vast quantities of data. Based on Intel testing, a reference server platform built with the design elements shown below can ingest terabytes of data in seconds.

While infrastructure specifics should derive from applications and workloads, these design principles incorporate a common set of functional building blocks:

- **Scalability and speed.** Intel® Xeon® processors embed a range of enhancements that target performance specifically for the types of workloads common with big data. This includes accelerators at the silicon level as well as architectural paths for operations like in-memory computing. Combined with right-sized Intel® technology-based storage and networking resources, Intel® architecture-based analytics servers have the flexibility to grow as data demands continue to climb.
- **Security.** Particularly in regulated industries that stress compliance, keeping data private and secure is paramount. Intel® processor technologies such as Intel® Cloud Integrity Technology, Intel® Trusted Execution Technology with One-Touch Activation, and Intel® Platform Trust Technology help to thwart data intrusions and help protect information even after it leaves the server.
- **Optimized data ingestion.** Parallel data ingestion from a wide variety of data sources enables high-speed data aggregation and faster time to conclusion. The solution architecture layers are optimized for Intel®-based hardware for maximum efficiency and total value.
- **Faster solution development.** By using code libraries from Intel and solutions from Intel's many third-party software collaborations, developers will have a shorter route to a more robust final analytics solution. Intel and its associates may also be able to assist in validating solutions before they reach the market.

Conclusion

Nearly every organization and every individual in the world stands to gain from the innovations that improved analytics delivers. Big data can thwart crime, improve lives, and transform economies. The scale of benefits is only bounded by the scale of solutions. For this reason, smart organizations need to adopt technologies that help them to ingest data faster and from a broader variety of sources. With high-performance components from Intel at the heart of analytics servers, and tools optimized for Intel technology throughout the solution stack, next-generation analytics and accelerated innovation are now within reach.

Better Results from Big Data with Intel® Technology

Intel® Xeon® processor-based platforms are the dominant choice in data analytics solutions, but some customers can further accelerate their specific ingestion and processing loads by configuring analytics servers with task-optimized hardware components:

- **Intel® Xeon® Scalable Processor.** The latest Intel Xeon processor-based platform generation reflects Intel's dedication to delivering unmatched enterprise-ready platforms to support advanced analytics workloads. These may range from in-memory computing to highly distributed workloads like Hadoop*, or a synthesis of both models, such as streaming analytics. Significantly increased cores, memory bandwidth and I/O enable the Intel Xeon Scalable processor to deliver improved performance and faster results across a range of databases and applications.
- **Intel® Optane™ Solid State Drives.** Built on pioneering Intel® 3D XPoint™ technology, Intel Optane SSDs help overcome the capacity and performance limitations of traditional storage devices. The combination of high I/O at low queue depths, quality of service and low latency under load enable Intel Optane drives to break through bottlenecks in a fast cache or storage usage. This enables greater server scaling and reduces transaction costs.
- **High-speed Integrated Intel® Ethernet (up to 4x10GbE).** Next-gen analytics conquers data silos, but insufficient network bandwidth will all but turn an analytics server into its own island. High-speed Integrated Intel Ethernet keeps data flow between network nodes running at peak levels. This can help reduce total system cost, lower power consumption, and improve transfer latency of large storage blocks and virtual machine migration.
- **Intel® Omni-Path Architecture.** Intel Omni-Path Architecture (Intel® OPA) delivers the performance for tomorrow's high performance computing (HPC) analytics workloads and the ability to scale to tens of thousands of nodes—and eventually more—at a price competitive with today's fabrics.

Learn More

You may also find the following resources useful:

- [Intel® Xeon® Scalable processor family](#)
- [TEDx: Big data requires big visions for big change](#)
- [White Paper: Tame the Data Deluge](#)
- [White Paper: Streaming Analytics in the Real-time Organization](#)
- [Solution Brief: Drive Up Performance for Analytics in Risk Management Workloads](#)

Find the solution that is right for your organization. Contact your Intel representative or visit intel.com/analytics.



¹ IDC Digital Universe Study: Big Data, Bigger Digital Shadows and Biggest Growth in the Far East, 2012, whizpr.be/upload/medialab/21/company/Media_Presentation_2012_DigiUniverseFINAL1.pdf

² Average size in May 2016: 1.43TB. HDD Units Down, Average Capacities Up; forbes.com/sites/tomcoughlin/2016/05/03/hdd-units-down-average-capacities-up/#7cc8cca25e39 com.

³ Big Data Trends for 2017, statisticbrain.com/health-care-industry-statistics

⁴ Roundup Of Analytics, Big Data & BI Forecasts And Market Estimates, 2016; forbes.com/sites/louiscolumbus/2016/08/20/roundup-of-analytics-big-data-bi-forecasts-and-market-estimates-2016/#dff06649c5f

⁵ Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights; whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

⁶ Blue Cross Blue Shield Statistics, bcbsm.com/health-care-fraud/fraud-statistics.html

⁷ "Government of Canada invests in new genomics "big data" research projects aimed at real-world challenges," [canadianinsider.com/government-of-canada-invests-in-new-genomics-big-data-research-projects-aimed-at-real-world-challenges](http://GenomeCanada.ca, canadianinsider.com/government-of-canada-invests-in-new-genomics-big-data-research-projects-aimed-at-real-world-challenges).

⁸ Nature, Is the \$1,000 genome for real? nature.com/news/is-the-1-000-genome-for-real-1.14530

⁹ GenomeNext, Powered by Amazon Web Services and Intel, Achieves Unprecedented Throughput of 1,000 Genomes Analyzed Per Day, Enabling Population-Scale Genomics, genomenext.com/single-post/2016/03/17/GenomeNext-Powered-by-Amazon-Web-Services-and-Intel-Achieves-Unprecedented-Throughput-of-1000-Genomes-Analyzed-Per-Day-Enabling-PopulationScale-Genomics

All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software, or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer, or learn more at intel.com.

System configurations, SSD configurations and performance tests conducted are discussed in detail within the body of this paper. For more information go to intel.com/performance.

Cost reduction scenarios described are intended as examples of how a given Intel- based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction. No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Copyright © 2018 Intel Corporation. All rights reserved. Intel, Xeon, Xeon Phi, Optane, 3D XPoint, and the Intel logo are trademarks of Intel Corporation in the U.S. and/or other countries.

^{*}Other names and brands may be claimed as the property of others.

0818/CAT/JS/PDF

© Please Recycle 338006-001EN