(intel)

Improve Employee Wellness through Predictive Analytics

With the right analytics infrastructure and software, organizations can improve population health and lower costs by tapping into the power of biometric data and other data sources

This reference architecture provides a fully functional technical picture for developing a cohesive business solution.

If you are responsible for:

- **Investment decisions and business strategy…**
  You'll learn how predictive analytics of biometric data and other data sources can help solve the pressing challenges facing insurance companies today.

- **Figuring out how to implement predictive analytics…**
  You'll learn about the architecture components and how they work together to create a cohesive business solution.

Executive Summary

Many organizations and their employees face continuing cost increases in their health plans. An estimated USD 2 trillion per year in healthcare costs globally are attributable to poor health habits.[1] Digitization—collecting data from wearable devices and other sources—plus advanced predictive analytics on that data can help improve population health and reduce these costs by putting behavioral data in the hands of healthcare and benefits professionals.

In this paper, we first explore how biometric data and predictive analytics can enable organizations to securely use data to engage employee populations with event-triggered personalized messaging, contesting, and rewards to encourage better health. We then look deeper into the solution architecture of the COVALENCE™ Health Analytics Platform from Big Cloud Analytics, powered by Intel® technology. Using this information, benefits managers and population health managers can take the next step toward unleashing the power of biometric data and predictive analytics.

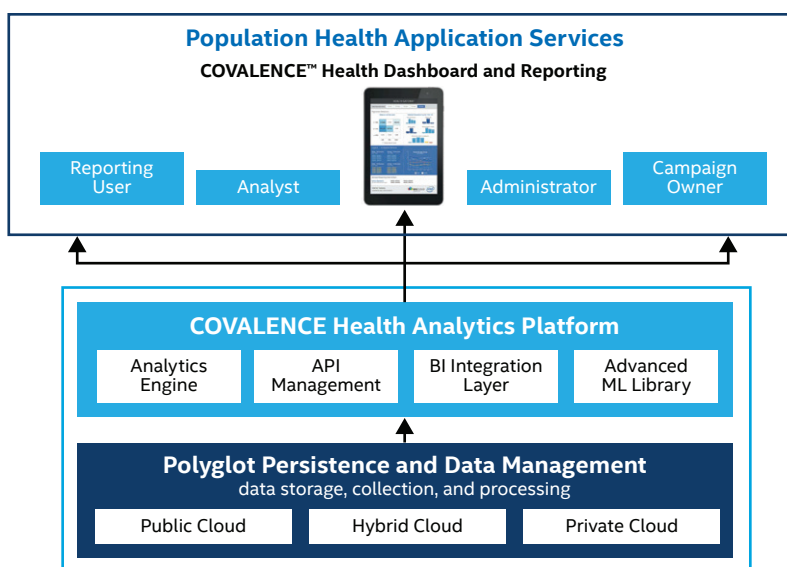[1] Source: www.mckinsey.com/insights/mgi/in_the_news/the_obesity_crisis

Figure 1. The COVALENCE™ Health Analytics Platform solution architecture supports innovative, personalized services and can enhance and optimize multiple business application services.

Table of Contents

Introduction

With the digital era and the arrival of the Internet of Things (IoT), organizations can use this new channel to substantially increase population engagement, resulting in improved health, increased productivity, reduced sick days, and cost. As personal technology becomes more mainstream, employers are adopting wearable technology due to its ease, availability, and benefits. Additionally, wearable technology gives benefits managers the ability to gather and apply data to healthcare models in new and innovative ways. The improvement in care and treatment can be significant—by identifying early warning signs, population health managers can proactively address the onset of chronic diseases, enhance lives, and generate long-lasting savings.

Population Health Transformation and the Impact of Wearable Technology

Many organizations are already using advanced data-mining techniques to improve decision making and to provide insights into improved business process management. In healthcare, the focus on health analytics can lead to improvements in population health and decreases in healthcare costs that benefit employers and their employees. However, data mining in population health management is still evolving. To thrive, organizations that are intent on improving the well-being of their members need to invest in exploring innovation, advanced data collection, and analysis techniques to achieve the same improvements in healthcare that are currently being realized in business process management.

Wearable technology presents important opportunities for benefits managers to embrace digitization. Consumer wearable devices allow employers to proactively engage their employees in an attempt to positively change their health habits, which provides value to the employees, to their employers, and to insurance companies.

Data collected from members' wearable devices, as well as other sources, can help benefits managers reduce risks through increased data accuracy and lower costs associated with employees' healthier and longer lives. Employers and their employees both reap rewards. Employers benefit because healthy employees may be more productive and less expensive to insure; employees benefit through the use of their data to help improve their quality of life and increase their life expectancy. Additionally, as their risk profiles improve, it is possible that the cost of their insurance premiums may drop.

By becoming a pioneer in the wearable device field and focusing on improving population health, employers strengthen their brands as innovators who make a difference in their employees' health. Proactive engagement can also improve trust between employees and their employers. In addition, collecting large quantities of biometric health data points for enrolled populations provides an opportunity for data analysis and predictive modeling. Data modeling can lead to insights that inform many aspects of population health management including best practices for wellness motivation, benefit personalization, targeted care plans, and fraud detection during claims processing.

Predictive Analytics Use Cases in Population Health Management

Biometric data combined with other data sources can fuel predictive analytics that can transform benefit management for population members. Predictive analytics using contextual data from fitness and sleep trackers provides cohort grouping by sleep, activity level, heart rate, and other factors to show each person how their activity compares to others with similar profiles. The automatic, continuous, and accurate log of activity and biometrics motivates participants to achieve fitness goals by showing the impact of improved fitness activity over time. Making even small changes in activity levels can reduce mortality and morbidity rates that result in significant savings on healthcare costs. While the application of biometric data and predictive analytics in population health management can manifest health improvements in many areas, the following four primary use cases stand out for employers.

Increased Morale and Productivity

Wellness programs can take advantage of the fun of fitness tracking by promoting healthy competitions and sharing of goal achievement. Leaderboards, digital awards, and progress gauges create a sense of shared community. Group activities develop involvement that builds corporate enthusiasm and contributes to improved population health and productivity. An additional benefit of a successful wellness program is the potential to foster a culture of well-being and pride in the community achievement of wellness.

Early Intervention

Gathering biometric information transforms data into business intelligence, predictive analytics, and actions. The COVALENCE™ Health Analytics Platform generates wellness scores, bio-identity scores (see sidebar, "Bio-Identity Helps Prevent Fraud"), and many other scores that can be used to engage with the participants. Insights based on analysis of the data points and trends can provide an early indication of potential health issues or lack of progress toward health goals. Population health managers use this data to support customizable actions and scalable engagement with wearable device users. Coaching can be provided to members who are identified with risk factors. Proactive management of members showing such risk factors can lead to better management of those members' health, which can help delay or even eliminate the debilitating onset of chronic conditions (see Figure 2).

Bio-Identity Helps Prevent Fraud

Biometrics is defined as the measurement and analysis of unique physical or behavioral characteristics (such as fingerprint or voice patterns) especially as a means of verifying personal identity. "Bio-identity" is a shorthand term used to describe an individual's unique biometric characteristics.

Predictive analytics involves many statistical techniques including machine learning, predictive modeling, and data mining. These techniques are used to analyze historical and current data to make forecasts. Predictive models leverage patterns found in data to identify risks and opportunities. Models capture relationships and correlations within the data. The combination of these analytic techniques defines a predictive score or probability. A bio-identity score is the probability that a collected data pattern is associated with a particular individual. Since individuals' biometric characteristics are so unique, the chance of two individuals having the same bio-identity score is less than 1 in 7 billion.[2]

Big Cloud Analytics has developed a patent-pending bio-identity algorithm that uses biometric sensor information to help analysts identify when a device is changing users or ownership, so that health and wellness programs can battle fraud. Due to the personal nature of the data collected, it is recommended that employers notify employees of the purpose of data collection and obtain the employees' consent.

---

[2] Based on Dr. V. Kumar's patent-pending algorithm.

Data Sources

External	Internal
STANDARD DATA	**STANDARD DATA**
Medical Information Bureau	Employee Records
Electronic Rx Profile	Application Data
Personal Medical Record	Benefit Plan Data
Health Exchange Data	
NON-STANDARD DATA	**NON-STANDARD DATA**
Medical Device Data	Email
Social Media	Voice
Event Registration Data	Click Stream
Survey Data	
Census Data	
Transaction Data	

Analytics Engine Algorithm

Chronic Condition Rules
- Chronic Condition Thresholds
- Risk Factors for Chronic Disease

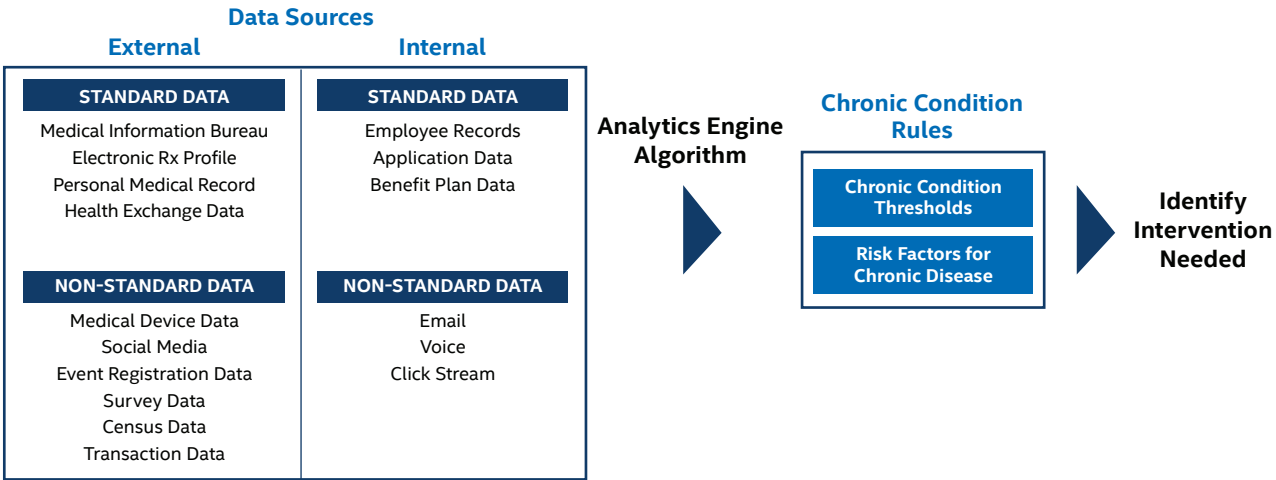**Identify Intervention Needed**

Figure 2. Advanced analytics supports the detection of chronic conditions. Automated detection of the onset of chronic conditions driven by data and augmented by clinical counseling can lead to population members better managing their health.

Reduced Insurance Plan Churn

Incorporating IoT data into predictive modeling for measuring employee benefit patterns can focus on areas such as insurance plan churn. Benefit plan enrollment turnover can be reduced using models for predicting the likelihood of group members defecting from an insurance plan. Using predictive analytics, these models can consume member data from group plan historical records to help identify defectors in advance and allow benefits managers to mitigate benefit plan churn and reduce coverage costs. Analysis of plan usage is also leveraged to recommend enrollment in insurance plans that will be more cost effective for both the employer and the employee.

New Healthcare Insights

Predictive analytics applied to population health management can facilitate the exploration of cause and effect. Health data gathered from trackers can greatly increase the number of data points that can inform healthcare professionals and can lead to new insights. Low activity levels might correlate to a rise in glucose levels, increase in sleep time may be attributed to a rise in activity levels, or improvements in sleep quality could depend on bedtime. Exploring these associations with population members managing chronic disease can lead them to make better health decisions.

Solution Architecture: COVALENCE™ Health Analytics Platform

The COVALENCE Health Analytics Platform provides an all-encompassing software-as-a-service (SaaS) platform with secure collection, storage, processing, and data analysis.

Solution Architecture Overview and Benefits

As Figure 3 shows, data may come from a variety of sources, ranging from biometric data collected by wearables and other sources to historical health data, self-reported health information, and more—all stored and analyzed in an encrypted standards-compliant cloud solution or on-premises data lake. To deliver value to the employer and employee, the COVALENCE Platform relies on the interaction of the following key components:

- **Wearable devices, such as fitness and sleep trackers.** These edge devices monitor and report biometric data (heart rate, sleep patterns, activity and exercise, skin temperature, etc.). The COVALENCE Platform features robust support for a variety of these devices, including scales, glucometers, and blood pressure cuffs.
- **Smart phones or Intel® processor-powered Internet of Things (IoT) gateways.** These edge devices gather biometric data and send it over security-enabled channels to data stores with the capacity for managing and analyzing large amounts of data.
- **A security-enabled cloud or on-premises data lake.** This can be a private, public, or hybrid cloud or an on-premises data lake (behind the enterprise firewall) capable of storing and analyzing large amounts of customer biometric and health data gathered from wearable devices and other data streams.
- **An analytics engine.** This engine is capable of analyzing longitudinal data generated by wearable devices and from other data sources, leveraging Big Cloud Analytics' patent-pending scoring and bio-identity algorithms, as well as providing real-time visualization tools and insurance workbenches.

The first three components are exterior to the COVALENCE Platform and may differ across implementations. The fourth component, the analytics engine, is the core of the COVALENCE Platform.
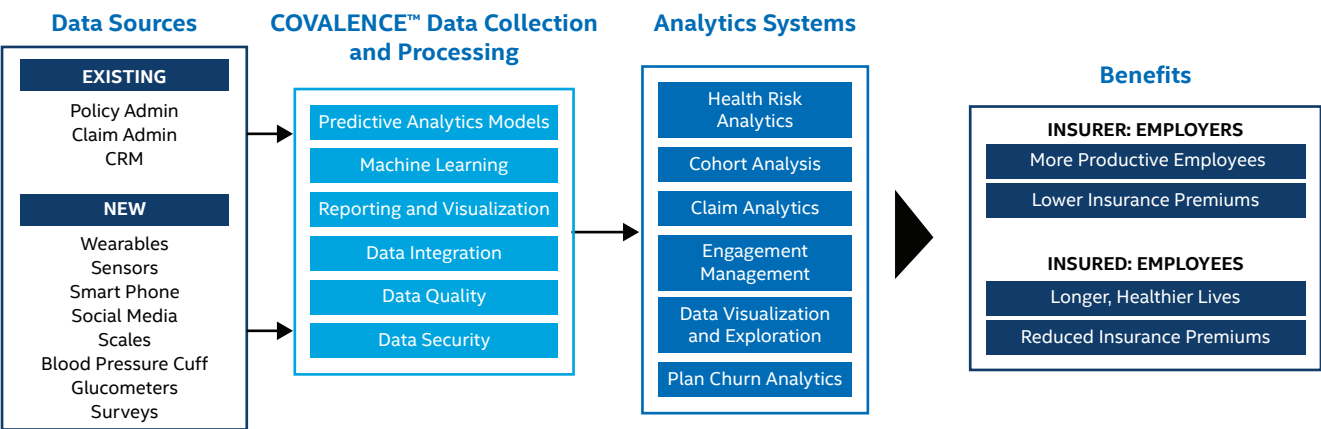

Figure 3. The COVALENCE™ Health Analytics Platform accesses data sources, processes the data, and makes the results available to a wide range of use cases and applications.

As shown in Figure 4, The COVALENCE Platform provides comprehensive, enterprise-level dashboard functionality for visualization of analysis, scoring, and trending of synthesized data streams from multiple sources regardless of format. This enables benefits managers or administrators to use insights to transform wellness campaigns and the health of the population.

Dashboards available in the COVALENCE Platform offer insight for both administrators and enrolled program participants, with a combination of historical and real-time reporting, analysis, benchmarking, and predictive analytics. At a glance, benefits managers can understand member behavior and risk level, which employee segments are at risk, which cohorts are exhibiting factors that can lead to chronic disease, and other insights into population health management. Powerful and flexible visualization tools allow administrators and other user roles to visualize employee data sets by demographic, behavior, or segmentation attributes.

Additionally, powerful user role management features enable administrators to define alerts and triggers for various conditions (such as device abandonment, changes in activity or sleep levels, recovery heart rate, elevated resting heart rate, body mass index, and elevated stress levels) to notify administrators and launch personalized, event-triggered messaging that can help guide participants to better health practices.

Population health managers can also see how their program is currently performing and obtain predictions of future performance and timing.

Solution Architecture Details

With the prospect of innovating population health management, organizations need a robust, predictive analytics platform and end-to-end program management. The COVALENCE Health Analytics Platform from Big Cloud Analytics is a modular, comprehensive solution that can help employers provide innovative, personalized services, as well as use the advanced analytics output in multiple business functions to enhance and optimize existing processes.

The COVALENCE Platform is versatile and modular, supporting multiple deployment strategies based on use cases, workload management scenarios, and regulatory compliance. If the organization desires, device data can be stored in a public cloud while the COVALENCE Platform runs on-premises—a hybrid cloud usage model. Alternatively, if the use case requires it, the entire application infrastructure including the data store can be deployed in a private or public cloud.

The COVALENCE Platform (see Figure 1) is based on an extensible open architecture that uses models built using both proprietary and open source technologies such as SAS Software*, IBM SPSS Statistics*, KNIME (Konstanz Information Miner*), and R*. The COVALENCE Platform is designed to display information in real time using the dashboard and a combination of existing and new web-based applications.
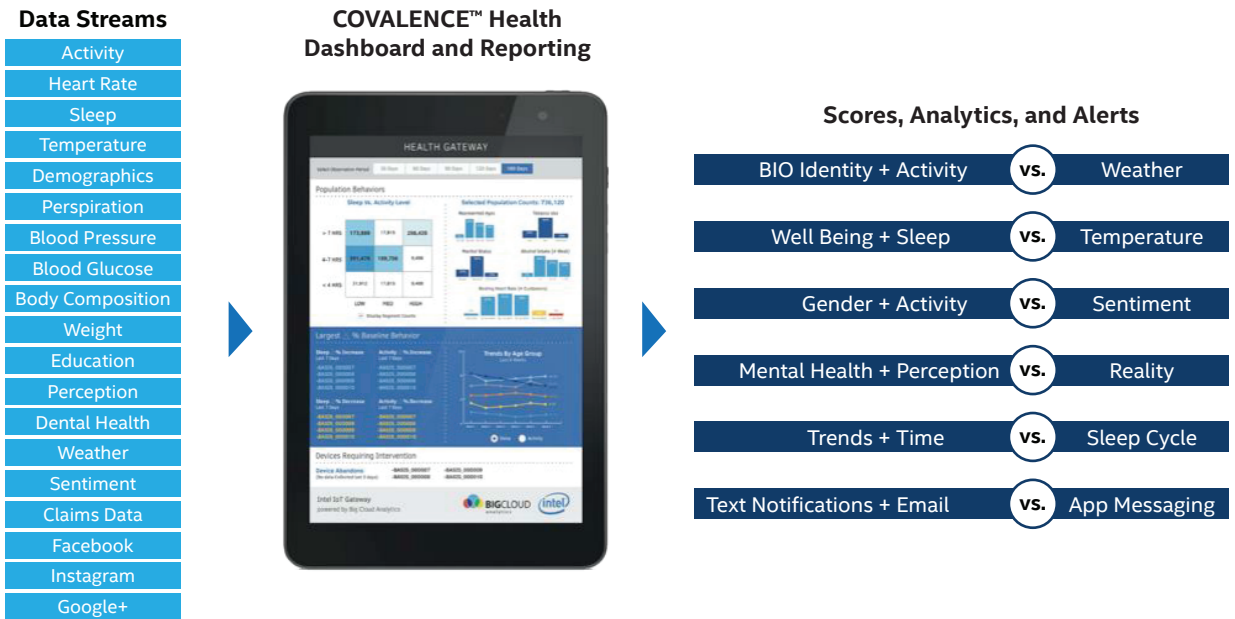

Figure 4. Most solutions available in the market only count steps as their major data point. The COVALENCE™ Health Analytics Platform goes far beyond counting steps, using up to 72,000 daily data points, plus external data sources. The easy-to-use, rich interface empowers employers to get the most out of all their data streams.

Architecture Components

The following components are included in the COVALENCE Health Analytics Platform:

- The **analytics engine** combined with a comprehensive engagement dashboard provides insight into health data and behavior patterns, including the use of predictive modeling. The advanced analytics engine combines the services of a Bayesian Computation Engine, Weighted Heuristics, and a Composite Score Computation Engine. The resulting predictive analytics and analytics insights can proactively identify potential behavioral issues that drive outcomes and lead to proactive care delivery rather than simply reactive care. This provides an opportunity to proactively engage population health members through automated health alerts, which can encourage better health habits, improve customers' quality of life, and reduce their risk profiles and claims.

- The **API management layer** supports connectivity to multiple devices and applications including the Apple Health Kit as well as devices produced by Withings, Garmin, Pivotal Living, and more.

- The **business intelligence (BI) integration layer** provides dashboards delivered to browsers on laptops, tablets, and smart phones as well as connectivity to SAS, Crystal Reports*, Pentaho platforms, MicroStrategy platform, and other custom BI applications.

- The **polyglot persistence (data storage) layer**[3] supports many computing languages and protocols, and can run in a public, private, or hybrid cloud environment. This layer can use new or existing storage infrastructure such as Amazon S3/4*, Microsoft Azure* Blob, Apache Hadoop* and Hadoop Distributed File System* (HDFS), and other file systems.

- In the **data management layer,** various applications are supported, including various Oracle products, Microsoft SQL*, MySQL*, SAP HANA*, Hadoop, and NOSQL*.

Additionally, the platform features fulfillment, enrollment, and customer support portal capabilities, as well as integration to rewards platform technology.

Data Formats

For employers to effectively take advantage of all the available data collected from wearable and other IoT-enabled devices, they should not have to worry about data format. As shown in Figure 5, the COVALENCE Platform can access, store, integrate, and analyze many kinds of data. Some data is structured, such as telemetric data coming from wearable and other devices. Other data is semi-structured or unstructured, such as COVALENCE self-reported data, surveys, social, and other data sources. The Platform's flexibility and adaptability enable it to read many data formats such as CSV, XLS, and JSON.

---

[3] Polyglot persistence is the process of storing data using multiple data storage technologies. The goal is, for each type of data, to use a data store that enables an application or a component of an application to optimally access that data. For more information, visit jamesserra.com/archive/2015/07/what-is-polyglot-persistence and mapr.com/products/polyglot-persistence.

Inputs

| Ad Hoc Batch |
| Automated Batch |
| Real-Time |

COVALENCE™ Health Analytics Platform

Scores

| Ad Hoc Batch |
| Automated Batch |
| Real-Time |

Figure 5. The COVALENCE™ Health Analytics Platform is designed to support real-time, batch, or ad hoc data ingestion, analytics, and visualization.

One Solution for Many Applications

The COVALENCE™ Health Analytics Platform from Big Cloud Analytics, powered by Intel® technology, can provide significant benefits to population health managers as detailed in this paper. Other industries, such as healthcare providers and life and health insurance companies can also use this same solution architecture to revolutionize services and operations. Imagine this:

A **healthcare provider** integrates biometric data from a smart watch, diet logs from an online portal, social media activity, and personal medical records to craft a personalized exercise plan for each patient. Based on subsequent data analysis, patients are rewarded with incentives for smart health choices, like a gift card or a discount on their next office visit.

A **trucking company** uses wearable devices along with cab cameras to monitor alertness and stress levels of long-haul truckers to assure their safety and the safety of the loads they carry, as well as the people sharing the road.

A **life and health insurance company** can reduce underwriting costs. Algorithms and analytics on longitudinal behavioral and other data can help underwriters process customer-submitted applications faster. Additionally, with access to many sources of data, including biometric data, insurance companies are able to identify unusual shifts in behavior patterns to more easily detect fraudulent claims.

A **pharmaceutical company** can create a more efficient clinical trial process by recruiting patients and collecting data that is not subject to reporting bias. There are demonstrable benefits to using wearables in clinical trial settings as they increase the volume and speed of data collection. Using data science techniques to identify trends and outliers in the clinical data helps researchers to visualize the data, apply the insights, and to design better subsequent trials.

Deployment Recommendations

The COVALENCE Health Analytics Platform can be deployed in a private, public, or hybrid cloud infrastructure behind a client firewall or as a hosted service. The method of deployment depends on specific business requirements. Careful selection of infrastructure technologies can help ensure a high level of scalability, high availability, security, and compliance, such as using a relational database management system (RDBMS) with built-in data encryption.

The following two sections discuss deployment considerations based on the number of users: 100,000 or fewer and more than 100,000 users. Both deployment scenarios support all deployment options. The primary difference between the two deployment scenarios involves data volume. For 100,000 or fewer users, an RDBMS is generally sufficient.[4] For more than 100,000 users, a scalable infrastructure such as Apache Hadoop is required to cope with massive data volume and data processing requirements.

Deployment Scenario #1: 100,000 or Fewer Users

Figure 6 depicts a deployment scenario for 100,000 users and fewer. In general, an RDMS is sufficient to handle the data volume and data processing requirements for this size of deployment.

The exact data storage and processing requirements should be calculated based on individual organizations' data collection points and the types of wearable devices generating the data. Fitness trackers collect readings on multiple biometric data points surrounding activity, sleep, and heart rate patterns. The measurements are gathered continuously and data collected is synchronized securely via smart phone or IoT gateway. The volume of data collection and the number of data sources for each deployment can be determined by an organization's business model and the type of services provided to members that utilize wearable or other IoT devices. If the employer offers comprehensive, personalized services, additional data sources would be required such as social media, weather, and demographics like location and level of education.

---

[4] Enterprises may choose to use Apache Hadoop* for smaller deployments, even though not strictly required, to future-proof the deployment (or perhaps an existing data lake is already available). In this type of deployment, as the number of participants increases, the implementation can easily scale without requiring new supporting infrastructure.
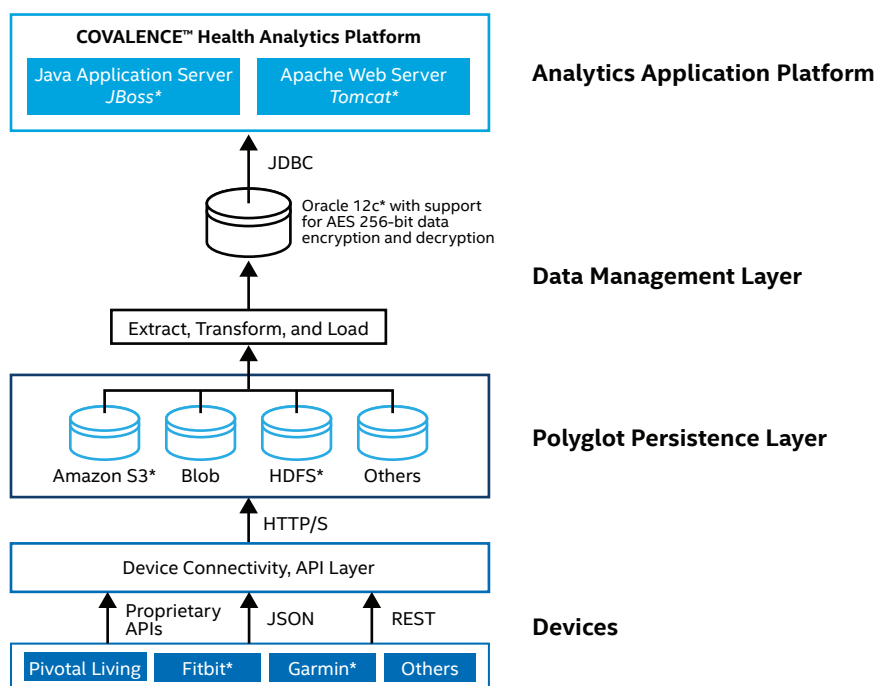

Figure 6. The COVALENCE™ Health Analytics Platform deployment scenario for 100,000 or fewer users is based on modular, decoupled architecture. This solution enables IT architects to choose various existing or new applications for data storage, ETL processing, and data management.

Data Flow

In this scenario, data from wearable devices, other IoT-enabled devices, or exogenous data sources is typically collected (using JSON and RESTful APIs and other custom or supplier-provided APIs). The choice of API depends heavily on the type of device and data storage architecture. After the data is collected, it is extracted, transformed, and loaded (ETL processing) into the database. The COVALENCE Platform can access the data through a Java Application Server and the result of the analysis can be published using a web server to internal and external applications to transform data into actionable information.

Security Features

Oracle Database 12c* helps provide high levels of data security and compliance with regulatory requirements such as the Health Insurance Portability and Accountability Act (HIPAA), the Payment Card Industry Data Security Standard (PCI DSS), and so on. Oracle 12c supports 256-bit encryption and decryption using Intel® Advanced Encryption Standard-New Instructions (Intel® AES-NI).[5]

Scalability

Oracle 12c provides both scale-up and scale-out configurations using clusters of Intel® Xeon® processor E5 family-based servers or using high-end 4- to 32-CPU Intel® Xeon® processor E7 family-based servers. To increase the performance and throughput, we recommend using high-endurance Intel® Solid-State Drives (Intel® SSDs) in conjunction with Intel® 10 Gigabit Server Adapters.

High Availability: Layered Architecture

To achieve the highest level of infrastructure availability for the deployment configuration with 100,000 or fewer users, we recommend a layered architecture. Using traditional load balancers and clustering can provide adequate high availability for web, application, and database servers. For the device connection layer, high availability is more device-specific and requires further analysis based on supported data transfer mechanisms. If the solution is deployed in a public cloud or as a hosted service, high availability and disaster recovery are part of the service-level agreement within the contract.

Deployment Scenario #2: More than 100,000 Users

In the deployment with 100,000 or fewer users, an RDBMS can sufficiently handle data volume and processing; however, for large-scale deployments with more than 100,000 users, we recommend using a scalable data storage and processing infrastructure, such as Hadoop. Such an infrastructure can handle the large number of data sources, data flows, and volumes generated by this number of users.

As shown in Figure 7, a converged infrastructure is a cost-effective and scalable solution based on open source technology that not only provides data storage but data processing and analysis as well. Scalability and high availability are critical for the secure collection, processing, and analysis of data. The functional requirements for a scalable infrastructure are as follows:

- The ability to communicate with large number and variety of heterogeneous devices, plus support for multiple communications protocols
- The ability to consume and aggregate large volumes of data at varying velocities and timeframes (be sure to account for any limits)
- Support for multi-application and multi-tenant environments so data can be used for diverse business purposes
- The ability to scale on demand and withstand failure by being cloud-enabled
- Support for modular and open architecture philosophy, including the use of open source solutions such as Hadoop where appropriate
- The ability to ingest a large volume of multi-structured data, store it, and expose it to other applications
- The ability to integrate with the existing infrastructure and applications and support for batch, ad hoc, and real-time data processing and analysis
- The ability to provide built-in data security at rest and in motion
- Compliance with regulatory requirements

While these functional requirements also apply to smaller deployments, they can be more challenging in a large-scale deployment.

---

[5] For more information about Oracle 12c* security features and compliance visit oracle.com/technetwork/database/security/security-compliance-wp-12c-1896112.pdf.

As shown in Figure 7, a managed, open source big data management and analytics platform is necessary. While there are several such platforms from which to choose, Cloudera Enterprise* is an excellent candidate for a COVALENCE Platform implementation that supports more than 100,000 users. Cloudera's fast, easy, and secure Hadoop ecosystem enables employers of all types and sizes to overcome data silos. This platform helps benefits managers, data scientists, and other key business stakeholders to access rich data sources, blend and analyze data from any source in any amount, detect patterns, model risk, and gain valuable real-time insights that deliver results.

Cloudera Enterprise enables population health managers to get more insights for more users at a lower total cost of ownership. With powerful open source access frameworks like Apache Spark* (machine learning), Apache Impala* (high-performance SQL), and Apache Kudu* (fast analytics on fast data), users can leverage various modeling capabilities that enable them to get results faster. Cloudera Enterprise also includes the only native Hadoop search engine and provides active data optimization to enable continuous performance improvements.

Cloudera Enterprise can be deployed on-premises or in the cloud. The unified big data management and analytics hub enables organizations to secure sensitive data while also meeting compliance requirements. Cloudera works with Intel and the open source community to deliver security without compromising flexibility or performance. (For more information on security in this deployment scenario, see the "Security" section and the "Cloudera Navigator* Security Features" sidebar.)

Data Flow

In Figure 7, ETL data processing is performed by Cloudera Enterprise; in the meantime, the COVALENCE Analytics Engine can access the data through HiveQL* or SQL—or in case of near-real-time analytics, through Spark SQL.[6] For real-time analytics it is highly recommended to use Cloudera Impala* and Kudu*.

As a big data platform, Cloudera Enterprise can provide data processing and analysis using a variety of open source and commercially available tools such as Spark, Mahout*, SAS Software, and SAP HANA. The big data platform analyzes millions of data points, identifies patterns, and makes the insights available as services that are transformed by visualization tools. Dashboards, portals, and mobile applications are built

---

[6] For more information on supported Cloudera Spark SQL, Impala, and Kudu visit cloudera.com/documentation/enterprise/latest/topics/spark_sparksql.html and blog.cloudera.com/blog/2015/09/kudu-new-apache-hadoop-storage-for-fast-analytics-on-fast-data.
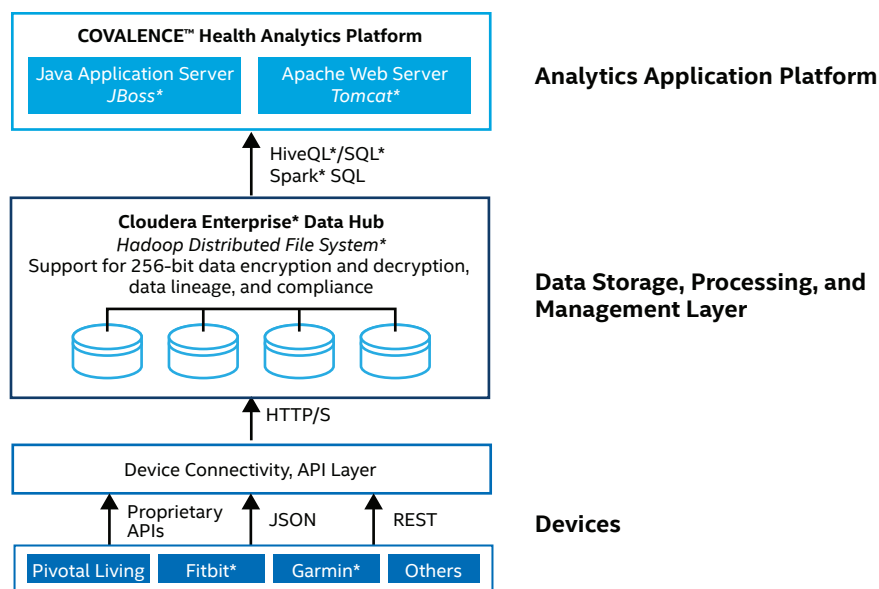

Figure 7. The COVALENCE™ Health Analytics Platform deployment scenario for more than 100,000 users is based on a managed open source data platform solution—Cloudera Enterprise* is one option. This solution provides a scalable infrastructure that can store and process large volumes of multi-structured data while providing data security using high-performance, file-level encryption and decryption. It also supports a wide variety of open source data processing and analysis tools.

at the top of the solution stack and make the data actionable. Advanced analytics provide additional insight that can lead to meaningful interaction with the participant to improve outcomes. Predictive analytics can anticipate outcomes and lead to proactive involvement by the benefits manager or the participant.

Optional Intelligent Gateway Component

An additional option for device connectivity in a large scalable deployment is Intel® Intelligent Gateway-based solutions[7], as shown in Figure 8, which enable multiple tracking devices to communicate back to the central organization—whether in a private, public, or hybrid cloud.

The gateway also provides the following capabilities:
- Encrypting data from fitness trackers and smart watches
- Updating controls
- Secure device pairing
- 3G management
- Whitelisting
- Secured communications
- Enforcement of application signing
- Encrypted entries
- Cloud connectivity

Security

Data encryption within Hadoop can occur at the network, OS volume, and HDFS folder/file levels:
- At the network level, industry-standard protocols such as SSL/TLS encrypt data just before it travels across a network and decrypt it when it arrives at the other end.
- At the OS volume level, Cloudera Navigator Encrypt*, in conjunction with Navigator Key Trustee, can encrypt an entire Linux* volume of cluster data inside and outside HDFS, such as temp/spill files, configuration files, and metadata databases. See the sidebar, "Cloudera Navigator* Security Features" for more information.
- At the HDFS folder level, HDFS data-at-rest encryption, also in conjunction with Navigator Key Trustee, applies encryption to specific HDFS folders. (It cannot encrypt data outside of HDFS.)

---

[7] For more information about Intel® Intelligent Gateways, visit intel.com/content/www/us/en/internet-of-things/gateway-solutions.html
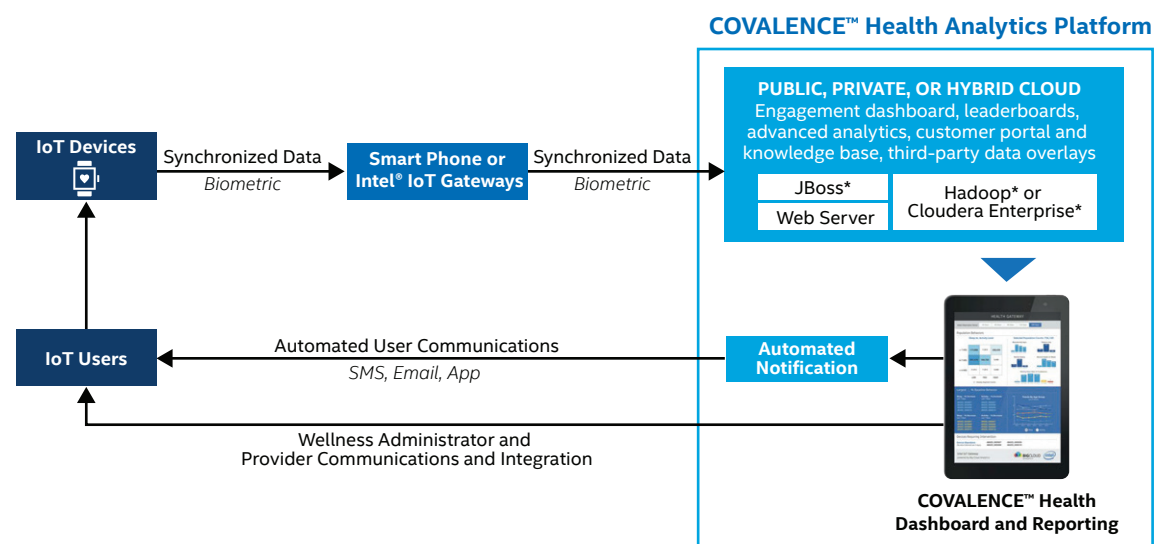

Figure 8. Intel® Intelligent Gateway-based solutions can be added to a large scalable deployment of the COVALENCE™ Health Analytics Platform. While not strictly required, such solutions can provide additional capabilities such as extended security features, data filtering, and remote management.

High Availability: Built-In Architecture

For large-scale deployments, we recommend using built-in high-availability capabilities such the following:

- Hadoop clustering and replication in case of node failure and data loss due to faulty storage medium

- Traditional load balancing and clustering for web and application servers (as for the smaller deployment scenario)

This architecture can be further enhanced by deploying a disaster recovery site (a second data center) as either passive/active (hot stand-by) or active/active (both sites are serving population members but one can continue services in case the other site fails). If the solution is deployed in a public cloud or as a hosted service, high availability and disaster recovery are part of the service-level agreement within the contract.

Conclusion

Population health managers need to delight member participants across their entire benefits journey. Improving the population health and well-being solution combines fitness tracking with a robust analytics platform that transforms various data streams into predictions, insights, and actions that improve participants' well-being. This proactive management of population health provides health and cost saving benefits to members as well as to population health managers. Predictive analytics and intelligent wearable devices, combined with other internal and external data sources, are poised to transform population health management.

Using predictive analytics, organizations can translate large quantities of structured and unstructured data into actionable insights about participant preferences and behaviors. These insights can drive development of new services and products that can enhance a company's competitive edge. They can also help reduce costs and improve the well-being of their member populations. Wearable technology presents important opportunities for benefits managers to embrace digitization. Consumer wearable devices allow employers to proactively engage their employees in an attempt to change their health habits, which provides value to the employees, to their employers, and to the benefits managers.

Intel® Xeon® Processor E5-2600 v3 Product Family: The Server Platform of Choice for a Scalable and More Secure Data Platform

Intel optimizes each new generation of processor to perform Intel® AES-NI faster, implementing intensive sub-steps of the AES algorithm into the hardware. The Intel® Xeon® processor E5-2600 v3 product family has reduced the latency of AES instructions from eight to seven cycles, with optimized throughput achieved by reducing the number of micro-operations. The reduction in latency helps improve serial AES encryption operating modes, such as cipher-block chaining (CBC) Encrypt. The increase in throughput aids parallel modes of operation, such as CBC Decrypt and multi-buffer.

The Intel Xeon processor E5-2600 v3 product family adds 50 percent more cores and cache over the previous generation[8] and includes numerous other hardware enhancements, such as Intel® Advanced Vector Extensions 2 (Intel® AVX2) and Intel® Quick Path Interconnect (Intel® QPI).

These innovations deliver up to 2.2x the performance over the previous generation,[9] significantly boosting output across a broad set of workloads. The Intel Xeon processor E5-2600 v3 product family also delivers an increase in virtualization density of up to 1.6x compared to the previous generation,[10] ever more important in the modern data center. With up to 18 cores per socket, 45 MB of last-level cache (LLC), and next-generation DDR4 memory support, the Intel Xeon processor E5-2600 v3 product family delivers significant performance improvements in workloads across all industries, from small businesses to large corporations in enterprise and technical computing, communications, storage, and private clouds.

Intel® Data Protection Technology (Intel® DPT) with Intel AES-NI accelerates data encryption and decryption up to twice as fast as previous-generation Intel® Xeon® processor families.[11] With many workloads, Intel AES-NI encryption/decryption is practically transparent to system resources. Combined with Intel DPT with Intel® Secure Key random number generation, the Intel Xeon processor E5-2600 v3 product family provides even stronger data protection. Intel® Platform Protection Technology, with Intel® Trusted Execution Technology (Intel® TXT), helps protect platform firmware and the OS kernel from pre-boot attacks. Intel TXT now supports trusted platform module (TPM) 2.0, with stronger cryptographic capabilities.

---

[8] Intel® Xeon® Processor E5-2699 v3 (18 cores, 45 MB cache) compared to Intel® Xeon® Processor E5-2697 v2 (12 cores, 30 MB cache).

[9] 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 www.intel.com. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations, and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information go to intel.com/performance.

[10] Relative performance is calculated by assigning a baseline value of 1.0 to one benchmark result, and then dividing the actual benchmark result for the baseline platform into each of the specific benchmark results of each of the other platforms, and assigning them a relative performance number that correlates with the performance improvements reported. Baseline Configuration and relative performance on Server Virtualization Benchmark: Intel® Server Board S2600CP with two Intel® Xeon® Processor E5-2697 v2, 16x32GB DDR3-1600 QR LR-DIMM, VMware* ESXi* 5.5 GA, relative performance: 1.0. New configuration: Intel® Server Board S2600GZ with two Intel® Xeon ® Processor E5-2699 v3, 16x32GB DDR4-2133 QR LR-DIMM, VMware* ESXi* 5.5 GA, relative performance: 1.6. Source: Intel internal measurements TR#1411 as of August 2014.

[11] Source as of June 2014 on AES-128-GCM Encryption algorithm: Intel internal measurements using Intel® Server Board S2600CW2S with two Intel® Xeon® processor E5-2658 v3, DDR4-2133, CentOS* v3.8.4, Open SSL v1.0.2-beta1. Baseline Configuration: Intel internal measurements with two E5-2658 v2, DDR3-1866, CentOS* v3.8.4, Open SSL v1.0.2-beta1.

Cloudera Navigator* Security Features

Cloudera Enterprise*—a big data management and analytics hub—includes Cloudera Navigator*, a native, end-to-end governance solution for Apache Hadoop*-based systems.[12] Cloudera Enterprise is the only platform available on the market that offers native encryption for data-in-motion between processes and systems, as well as for data-at-rest as it persists on disk or other storage media. Navigator is the first fully integrated data security and governance application for Hadoop-based systems. Of equal importance, Cloudera Enterprise enables the most complete data security protocols.

Cloudera Navigator Encrypt* (integrated with Cloudera Navigator) provides massively scalable, high-performance encryption for critical Hadoop data. It utilizes the industry-standard 256-bit Advanced Encryption Standard (AES) and provides a transparent layer between the application and file system. Navigator Encrypt also includes process-based access controls, allowing authorized Hadoop processes to access encrypted data while simultaneously preventing admins or super-users (such as root) from accessing data that's not necessary.

Navigator Encrypt includes the following features:
- Advanced key management. Stores keys separately from the encrypted data to prevent data breaches from resulting in the loss of the cryptographic key
- Transparent data encryption. Protects data at rest with minimal performance impact and doesn't require complex changes to databases, files, applications, or storage
- Process-based access controls. Restricts access to specific processes rather than by OS user
- Encrypting and decrypting unstructured data. Helps protect personally identifiable information, intellectual property, log files, and any other sensitive data that could be considered damaging if exposed outside the business
- Performance. Supports the Intel® AES-NI cryptographic accelerator for enhanced performance in the encryption and decryption process
- Compliance. Helps enable compliance with HIPAA-HITECH, PCI-DSS, FISMA, EU Data Protection Directive, and other data security regulations
- Multi-distribution support. Supported on Debian, Ubuntu, CentOS, Red Hat, and SUSE distributions of Linux*
- Simple installation. Distributed in RPM and DEB packages, as well as SUSE KMPs
- Multiple mount-points. Support for multiple encrypted mount-points managed with individual keys

The security of cloud data depends on the safety of cryptographic keys, SSL certificates, database tokens, and other opaque objects. Navigator Encrypt also helps to protect these critical IT secrets from unauthorized access and helps meet strict data security regulations.

Cloudera Enterprise* Data Hub Security Architecture

---

[12] For more information about Cloudera data governance and compliance, visit cloudera.com/developers/get-started-with-hadoop-tutorial/data-governance-and-compliance.html.

Solutions Proven by Your Peers

The COVALENCE™ Health Analytics Platform, powered by Intel® technology, provides population health managers with advanced predictive analytics. This and other solutions are based on real-world experience gathered from customers who have successfully tested, piloted, and/or deployed these solutions in specific business use cases. Solution architects and technology experts for this solution reference architecture include:

- **Wendy Bohner**
  Solution Architect, Health & Life Sciences Solutions, Intel Corporation
- **Parviz Peiravi**
  Chief Technical Officer, Financial Services Industry Solutions, Intel Corporation
- **Robert E. Cabral**
  VP of Services, Big Cloud Analytics
- **Robert J. Gentry**
  Chief Operating Officer & Chief Financial Officer, Big Cloud Analytics

Intel Solution Architects are technology experts who work with the world's largest and most successful companies to design business solutions that solve pressing business challenges. Big Cloud Analytics, a market leader in real-time predictive analytics technology for the Internet of Things, has worked closely with Intel Solution Architects to develop the solution described in this doument.

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

Learn More

This solution reference architecture complements product documentation and is part of an entire solution kit of content that is full of key insights and learnings:

- **Solution Brief:** "Improve Population Health with Predictive Analytics"

You may also find the following resources useful:

- **Big Cloud Analytics:** bigcloudanalytics.com
- **Cloudera Enterprise Data Hub:** cloudera.com/content/dam/cloudera/Resources/PDF/solution-briefs/Cloudera-EDH-ExecutiveBrief.pdf
- **Intel® IoT Gateways:** intel.com/content/www/us/en/internet-of-things/gateway-solutions.html

Solution Provided By: