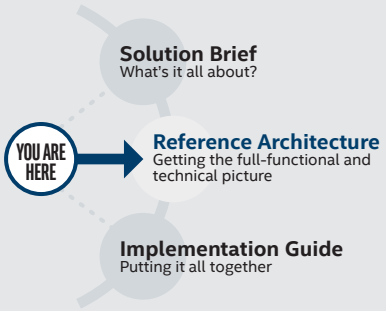
REFERENCE ARCHITECTURE

Financial Services
Data Management and Analytics

Deliver Innovative Services through Predictive Analytics

With the right analytics infrastructure, life and health insurance companies can improve service, increase revenue, and lower costs by tapping into the power of biometric data and other data sources.



What You'll Find in This Solution Reference Architecture

This solution provides a starting point for developing a predictive analytics infrastructure.

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

Life and health insurance companies face two seemingly opposing market pressures: attracting and retaining more customers through innovation and customized offerings, and cutting costs by increasing operational efficiency. Digitization—collecting data from wearable devices and other sources—plus advanced predictive analytics on that data can help meet these challenges and increase an insurance company's competitive edge.

In this paper, we first explore how biometric data and predictive analytics can change the nature of risk assessment, operational management, and the consumer engagement and services delivery model. 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, insurance companies can take the next step toward unleashing the power of biometric data and predictive analytics.

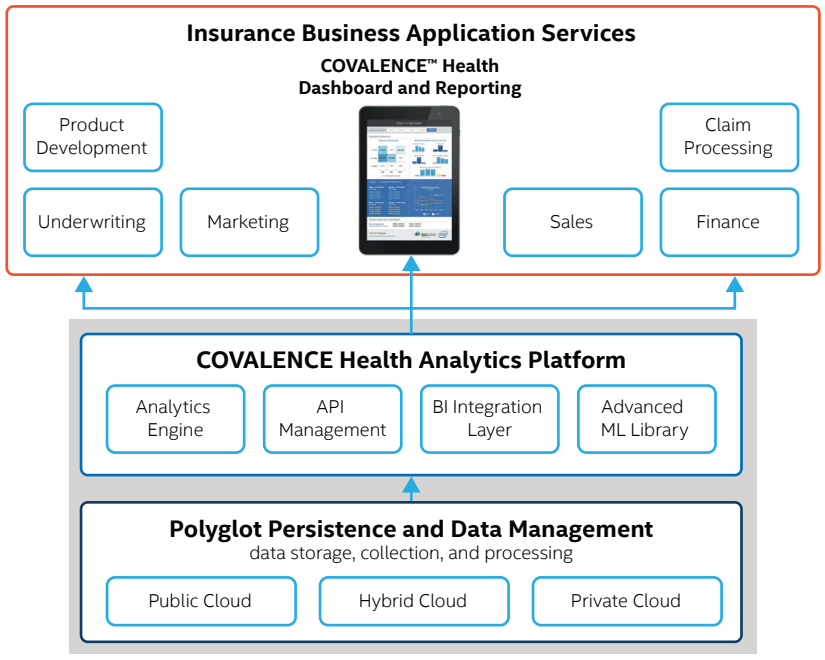


Figure 1. The COVALENCE™ Health Analytics Platform solution architecture supports innovative, personalized services and can enhance and optimize multiple business application services.

Table of Contents

Introduction

With the digital era and the arrival of the Internet of Things (IoT), insurance companies can use this new channel to substantially increase customer engagement. As personal technology becomes more mainstream, customers are adopting wearable technology due to its ease, availability, and benefits. Additionally, wearable technology gives insurance companies the ability to gather and apply data to actuarial models in new and industry-changing ways. The return on investment can be significant as use cases span an insurance company's entire business from awareness to customer acquisition to claims and customer retention.

Industry Transformation and the Impact of Wearable Technology

Many industries such as property and casualty insurance, financial institutions, and healthcare providers are already using advanced data-mining techniques to improve decision making. This has resulted in lower risk and better customer relations. However, data-mining in the life and health insurance industries is still evolving. To thrive, life and health insurance companies need to invest in exploring innovation, advanced data collection, and analysis techniques to achieve the same improvements to risk and customer relations that the other industries have benefitted from.

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

Data collected from customers' wearable devices can help insurance companies reduce actuarial risk as a result of increased data accuracy and lower costs associated with customers' healthier and longer lives (see Figure 3 for additional benefits). Employers and customers also benefit. Employers benefit because healthy employees may be more productive and less expensive to insure; insurance customers can use their data to help improve their quality of life and increase their life expectancy. As their risk profiles improve, it's possible that the cost of their insurance premiums may drop.

By becoming a pioneer in the wearable device field and focusing on improving customer health, insurers strengthen their brands as innovators who make a difference in the health of communities where they have a presence. Proactive engagement also improves trust between customers and insurance providers. In addition, collecting large quantities of biometric health data points for enrolled populations provides an opportunity for data analysis and predictive modeling, which can lead to insights that inform many aspects of the insurance industry including underwriting processes, product personalization, targeted marketing, and fraud detection during claims processing.

Predictive Analytics Use Cases in the Insurance Industry

Whether a company is offering health or life insurance, biometric data combined with other data sources can fuel predictive analytics that can transform the business. Predictive analytics enables life and health insurance companies to create a comprehensive roadmap for increasing engagement with their customers' journeys. It provides an enterprise-wide view of a customer by gathering insights and identifying opportunities across all business lines while improving internal business processes. While the application of biometric data and predictive analytics in the insurance industry is broad, four primary use cases stand out for life and health insurance companies.

Underwriting

Incorporating IoT data into predictive modeling for underwriting is quickly becoming a common use case that can empower health and life insurance companies to segment and underwrite risks through a more consistent and less expensive process. A life insurer will typically spend approximately one month and several hundred dollars underwriting each applicant. The opportunity for annual savings from reduced and streamlined requirements and processing time are in the millions.[1] For example, algorithms and analytics on longitudinal behavioral and other data can help underwriters process customer-submitted applications. As shown in Figure 2, an analytics engine algorithm can analyze many sources of data. The outcome of this analysis, which could take as little as 20 minutes, may enable a life insurance underwriter to waive medical requirements for certain customers. These customers benefit from an easier, faster time-to-issue and personalized premium pricing, while the insurer benefits from reduced underwriting requirements and costs.

Personalized Offerings

Using biometric data from wearable devices, along with other data sources, helps insurers gather insights and identify new opportunities across business lines. Insurers can use data to make better decisions and adapt customer acquisition and channel strategies. For example, health and life insurers can offer usage-based policies, with variable premiums based on real-time data collected from the insured. (This is already commonplace in the car insurance industry, where premiums are often based on a customer's driving habits.)

Market Segmentation

Marketing expenses make up a large portion of an insurance company's budget. Therefore, utilizing marketing budgets efficiently is a key operational strategy. Through analysis and segmentation of health data from wearable devices, plus other sources of information such as social media, insurers are able to target insurance products to specific market segments. With a "360-degree" view of the customer, predictive analytics can identify potential customers who are more likely to qualify for life insurance products, or who would be interested in a specific health insurance product or other portfolio products.

Fraud Detection

Fraudulent claims cost insurers—and indirectly, consumers. According to the Federal Bureau of Investigation, insurance fraud is the second most costly white-collar crime in America, and accounts for more than $30 billion in losses every year.[2] 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.

[1] "Predictive Modeling for Life Insurance: Ways Life Insurers Can Participate in the Business Analytics Revolution," 2010, Deloitte Development LLC

[2] "Fraud Detection in Insurance Claims," Bob Biermann. Predictive Analytics World, April 2013.

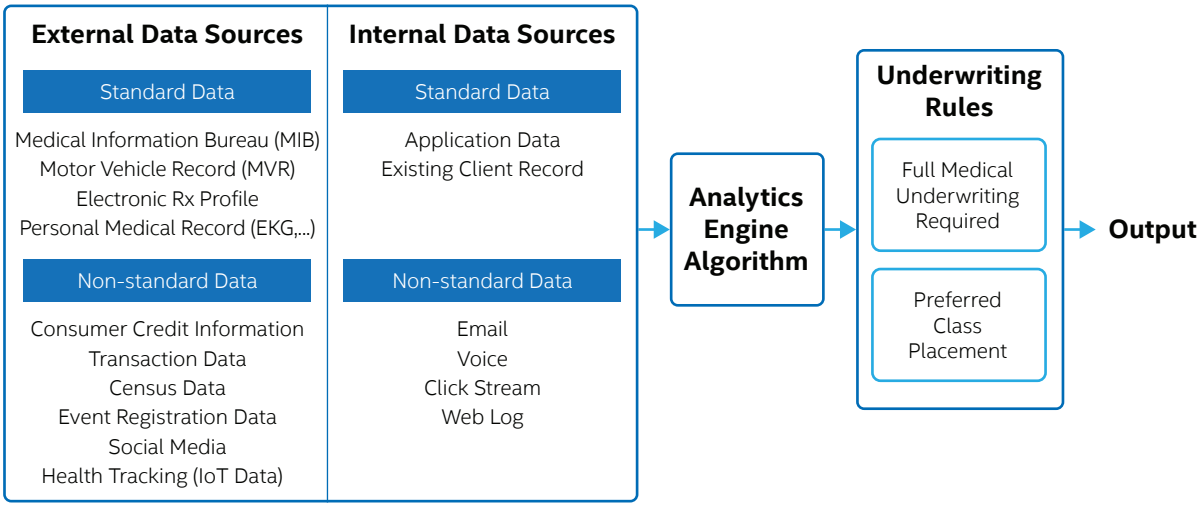


Figure 2. Advanced analytics supports the underwriting process. Automated decisions driven by data and augmented by expert advice can significantly streamline underwriting and cut operational costs.

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 medical devices to historical health data, self-reported health information, and more—all stored and analyzed in an encrypted standards-compliant cloud solution.

To deliver value to the insurer and the customer, 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 secure channels to data stores with the capacity for managing and analyzing large amounts of data.
- **A secure 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.

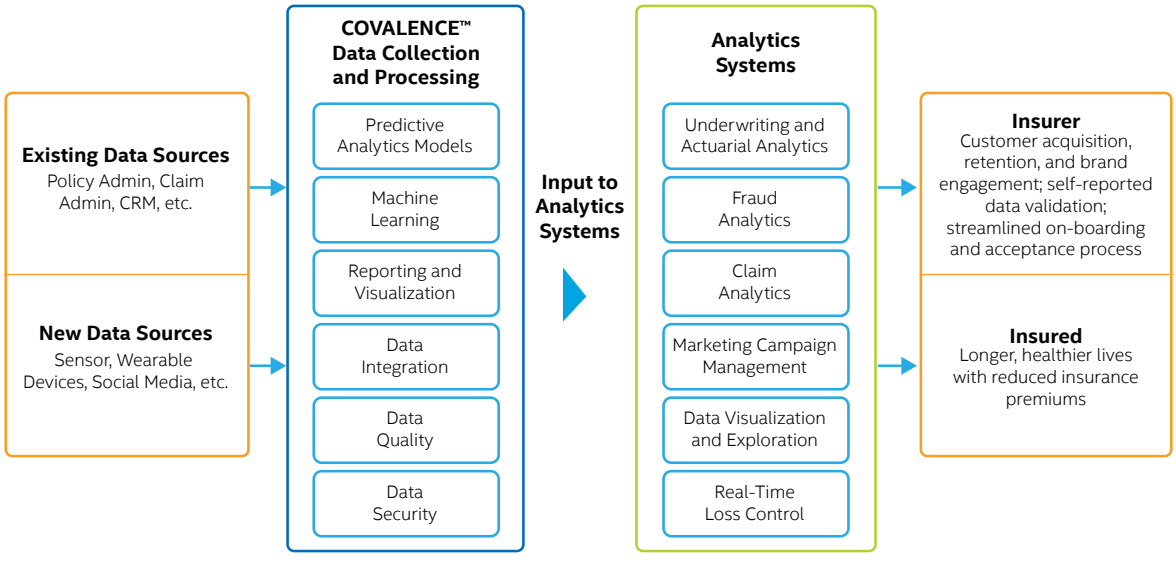


Figure 3. The COVALENCE™ Health Analytics Platform accesses data sources, processes the data, and makes the results available to a wide range of use cases and applications.

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 insurers to use insights to transform processes and campaigns.

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, insurers can understand customer behavior and risk level, which customer segments are at risk, which regions and markets have the most potential for expansion, and other insights into population health management. Powerful and flexible visualization tools allow administrators and other user roles to visualize customer 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, elevated resting heart rate, and elevated stress levels) to notify administrators and launch personalized, event-triggered messaging that can help guide users to better health practices.

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

Solution Architecture Details

With the prospect of revolutionizing both consumer relations and operations, insurance companies 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 insurers provide innovative, personalized services, as well as use the advanced analytics output in multiple business functions to enhance and optimize the 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 insurance company 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.

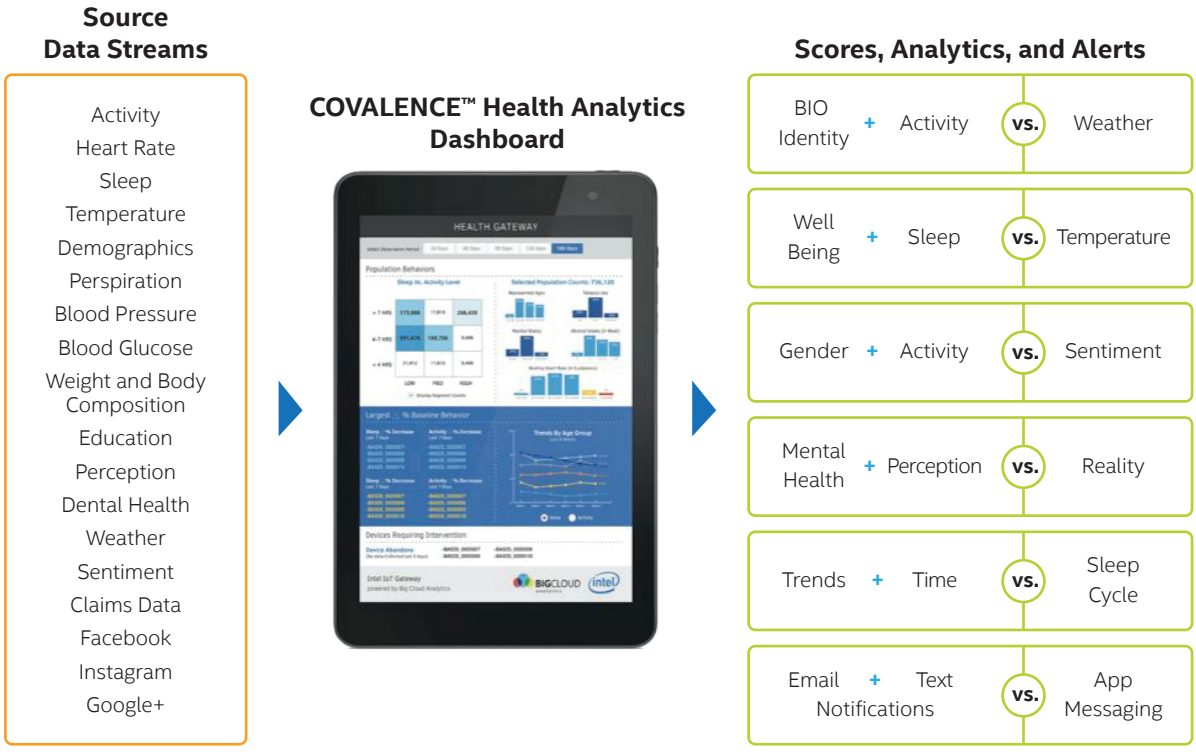


Figure 4. Most solutions available in the market only count steps as their major data point. The COVALENCE™ Health Analytics Platform goes far beyond counting steps, using up to 72,000 daily data points, plus external data sources. The easy-to-use, rich interface empowers insurers to get the most out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

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.

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 to identify ways to mitigate potential risk factors. The engine calculates risk scores dynamically, in real time, based on established industry normative values. This provides an opportunity to proactively engage the customer 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 Fitbit* and 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*, 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 insurance companies 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, XML, 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.

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 the life and health insurance industries, as detailed in this paper. Others, such as healthcare providers, trucking companies, and employers in general 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.

An employer can help employees increase activity and sleep for better general wellness. For example, an employee could be alerted to high levels of stress, dehydration, and high resting heart rate. The alerts can suggest actions and provide programmatic content or contests for improvement. (The results could be used either in a double-blind, aggregated sense or to reward personalized effort.)

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 insurance customers' data collection points and the types of wearable devices generating the data. In general, most fitness trackers collect readings on up to 50 biometric data points and synchronize the data 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 insurance company's business model and the type of services provided to customers that utilize wearable or other IoT devices. If the insurer 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 users increases, the implementation can easily scale without requiring new supporting infrastructure.

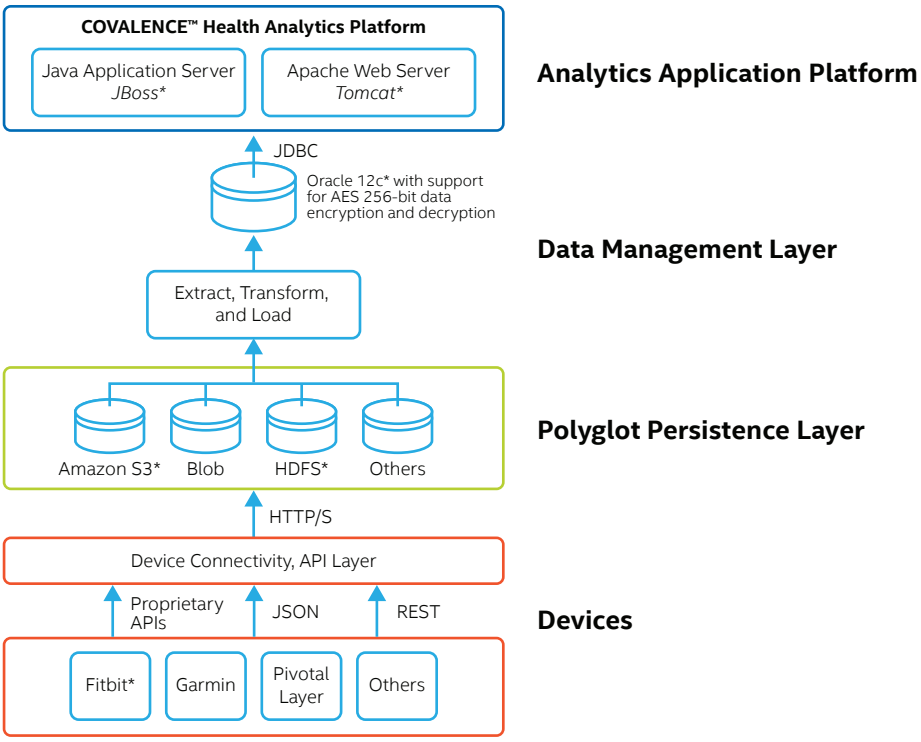


Figure 6. The COVALENCE™ Health Analytics Platform deployment scenario for 100,000 or fewer users is based on modular, decoupled architecture. This solution enables IT architects to choose various existing or new applications for data storage, ETL processing, and data management.

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

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.

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

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 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, 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, 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. 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.

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.

A managed open source data platform solution such as Cloudera Enterprise*—a big data management and analytics platform—is a good candidate for a COVALENCE Platform implementation that supports more than 100,000 users (see Figure 7). Cloudera's fast, easy, and secure Hadoop ecosystem enables insurers of all types and sizes to overcome data silos. This platform helps actuaries, underwriters, 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 insurers 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.

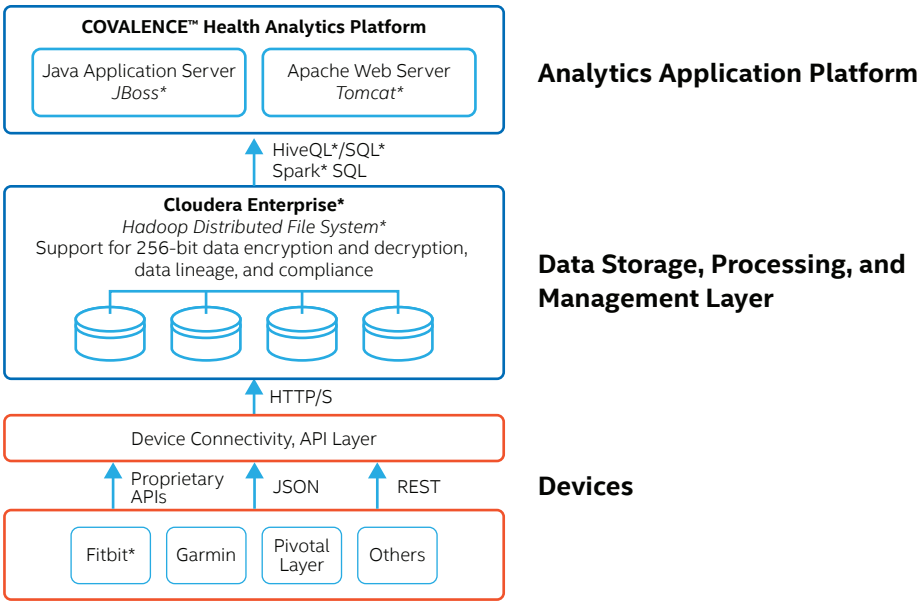


Figure 7. The COVALENCE™ Health Analytics Platform deployment scenario for more than 100,000 users is based on a managed open source data platform solution. This solution provides a scalable infrastructure that can store and process large volumes of multi-structured data while providing data security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

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

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 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 customer to improve outcomes. Predictive analytics can anticipate outcomes and lead to proactive action by the insurer or the customer.

Optional Intelligent Gateway Component

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

[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.

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

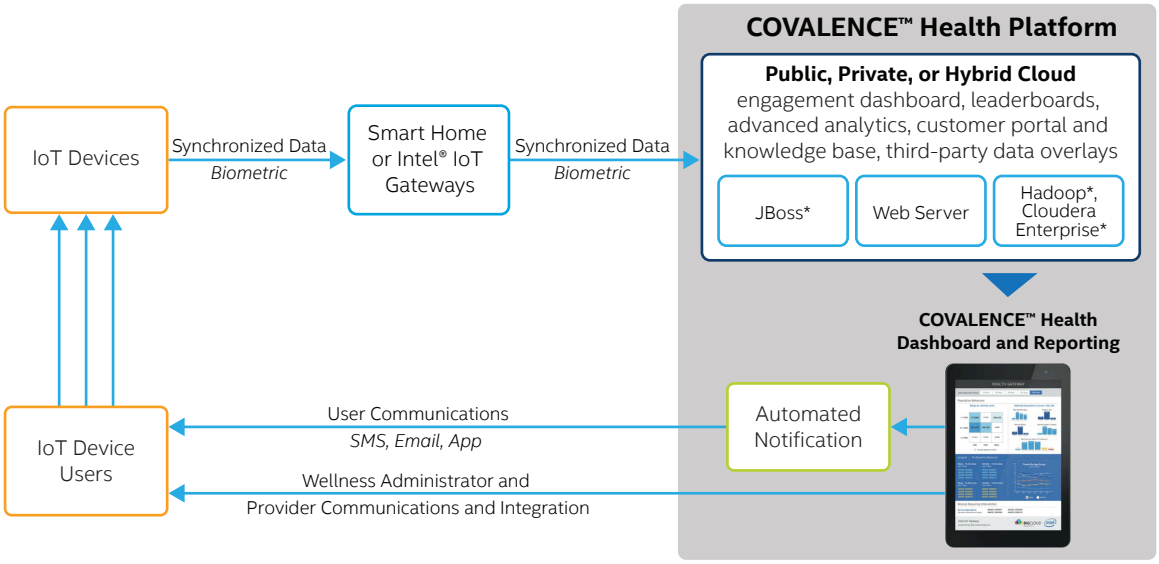


Figure 8. Intel® Intelligent Gateway-based solutions can be added to a large scalable deployment of the COVALENCE™ Health Analytics Platform. While not strictly required, such solutions can provide additional capabilities such as extended security features, data filtering, and remote management.

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.)

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 customers 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

Life and health insurers need to delight customers across their entire customer journey—from brand awareness, first contact with the company, customer acquisition, selecting and using a product, on-going customer engagement, service and retention. To this end, the industry must develop innovative services and products that are based on a thorough understanding of consumer needs and that are easy to access or purchase at competitive rates. Predictive analytics and intelligent wearable devices, combined with other internal and external data sources, are poised to transform the insurance industry by providing a powerful means to do just that.

Using predictive analytics, insurers can translate large quantities of structured and unstructured data into actionable insights about customer 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 enhance efficiency by transforming operational activities such as underwriting processes and reducing fraudulent claims.

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.[8] 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.

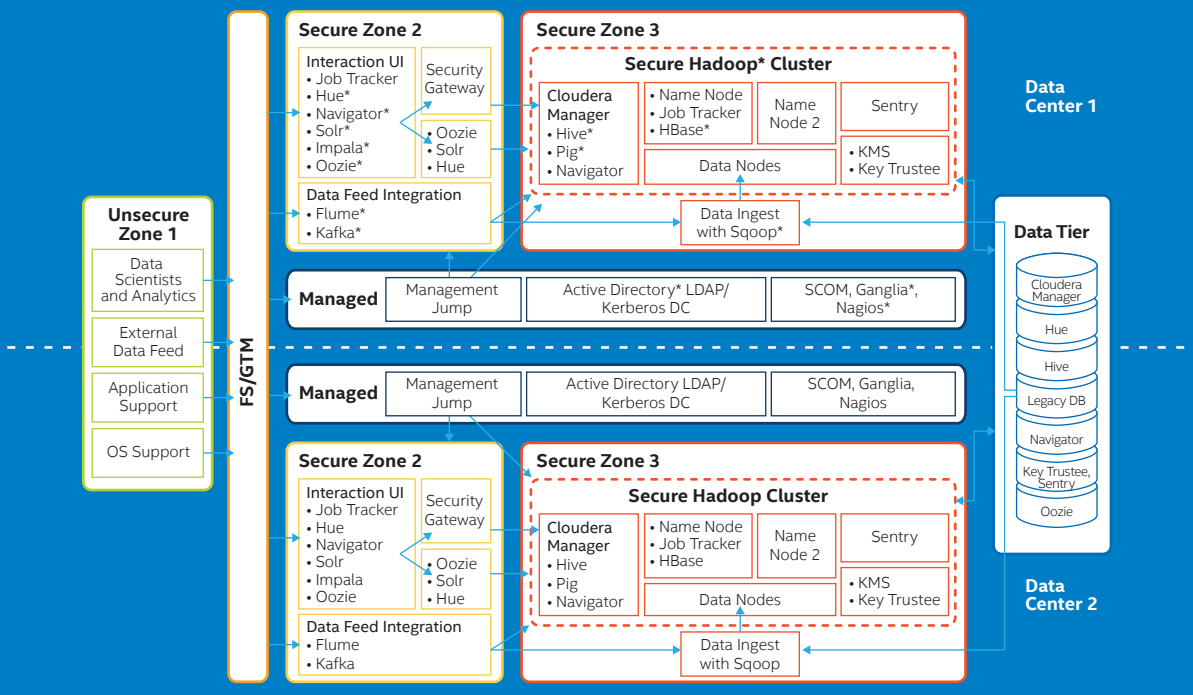
Cloudera Navigator Encrypt* (integrated with Cloudera Navigator) provides massively scalable, high-performance encryption for critical Hadoop data. It utilizes the industry-standard 256-bit Advanced Encryption Standard (AES) and provides a transparent layer between the application and file system. Navigator 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



[8] 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 insurers 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:

- **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 and 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's right for your organization.
Contact your Intel representative or visit **intel.com/FSI**.

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: "**Reinventing the Insurance Customer Relationship through Wearable Technology**"**

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 Financial Services Solutions:** intel.com/FSI
- **Intel® IoT Gateways:** intel.com/content/www/us/en/internet-of-things/gateway-solutions.html

Solution Provided By:

BIGCLOUD analytics

intel®

Printed in USA 1116/CPET/KC/PDF ♻ Please Recycle 334068-001