



# Cloud Computing and Machine Learning Uses in the Actuarial Profession





# Cloud Computing and Machine Learning Uses in the Actuarial Profession

<b>AUTHOR</b>			
	Van Beach, FSA, MAAA Principal Milliman, Inc.	Alexandre Boumezoued Senior Consultant Milliman, Inc.	Josh Dobiac, LL.M., JD, MS, CAIA Consultant Milliman, Inc.
	Jonathan B. Glowacki, FSA, MAAA, CERA, CSPA Principal and Consulting Actuary Milliman, Inc.	Joe Long Assistant Actuary, Data Scientist Milliman, Inc.	Antoine Ly, FIA Senior Consultant Milliman, Inc.
	Makho Mashoba Data Scientist Milliman, Inc.	Zohair Motiwalla, FSA, MAAA Principal and Consulting Actuary Milliman, Inc.	David South Actuarial Analyst Milliman, Inc.
	Benjamin Buttin, ASA Associate Actuary Milliman, Inc.		

## Caveat and Disclaimer

The opinions expressed and conclusions reached by the authors are their own and do not represent any official position or opinion of the Society of Actuaries or its members. The Society of Actuaries makes no representation or warranty to the accuracy of the information.

Copyright © 2019 by the Society of Actuaries. All rights reserved.

# CONTENTS

<b>Acknowledgments .....</b>	<b>5</b>
<b>Section 1: Executive Summary .....</b>	<b>6</b>
<b>Section 2: Introduction and Background .....</b>	<b>8</b>
SECTION 2.1 WHAT IS THE CLOUD? .....	8
SECTION 2.1.1 SAAS .....	10
SECTION 2.1.2 IAAS .....	10
SECTION 2.1.3 PAAS .....	11
SECTION 2.2 IMPACT ON ACTUARIES .....	14
<b>Section 3: Financial Modeling, Actuarial Process, and the Cloud .....</b>	<b>17</b>
SECTION 3.1 FINANCIAL MODELING .....	18
SECTION 3.2 DATA COLLECTION .....	19
SECTION 3.2.1 MODELING .....	20
SECTION 3.2.2 REPORTING .....	22
SECTION 3.2.3 GOVERNANCE .....	23
SECTION 3.3 INSURANCE INNOVATIONS USING THE CLOUD .....	24
SECTION 3.4 TO THE CLOUD .....	25
<b>Section 4: Machine Learning / AI and Cloud Data Impact on Actuarial Science .....</b>	<b>27</b>
SECTION 4.1 INCREASED PROCESSING CAPACITY THROUGH DISTRIBUTED SYSTEMS .....	27
SECTION 4.1.1 CHOOSE THE RIGHT TECHNOLOGY AND CLOUD SERVICE .....	27
SECTION 4.2 CURRENT LEVEL OF USE OF MACHINE LEARNING AND AI BY ACTUARIES .....	29
SECTION 4.2.1 TRENDS AND EXPECTED USE OVER NEXT FIVE YEARS .....	30
SECTION 4.3 DATA CONSIDERATIONS FOR USING MACHINE LEARNING / AI .....	30
SECTION 4.3.1 LEVEL OF GRANULARITY, SCALABILITY, AND AUTOMATION .....	30
SECTION 4.3.2 USE AND AVAILABILITY OF DATA / BIG DATA / DATA ARCHITECTURE .....	31
SECTION 4.4 CASE STUDIES OF MACHINE LEARNING FOR ACTUARIES .....	32
SECTION 4.4.1 CASE STUDY .....	32
SECTION 4.5 COMMON MACHINE LEARNING ALGORITHMS AND TOOLS .....	34
SECTION 4.5.1 MACHINE-LEARNING ALGORITHMS (CART, RANDOM FORESTS, K-MEANS, ETC.) .....	34
SECTION 4.5.2 COMMON PACKAGES .....	36
SECTION 4.5.3 CONSIDERATIONS FOR USING OPEN-SOURCE SOFTWARE .....	37
SECTION 4.5.4 MAINTENANCE .....	37
SECTION 4.6 MODEL SELECTION AND INTERPRETATION OF RESULTS .....	37
SECTION 4.6.1 MODEL SELECTION .....	37
SECTION 4.6.2 INTERPRETATION AND FAIRNESS OF ALGORITHMS .....	38
SECTION 4.7 INSIGHTS / LESSONS LEARNED / RISKS .....	38
SECTION 4.7.1 DO YOUR HOMEWORK .....	38
SECTION 4.7.2 CHECK RELEVANCE OF THE IMPLEMENTATION / SELECTION OF ALGORITHM .....	38
SECTION 4.7.3 ENSURE THE FAIRNESS OF THE RESULTS .....	39
SECTION 4.7.4 CHECK STABILITY OF THE ANALYSIS OVER TIME .....	39
SECTION 4.7.5 INCREASE THE COMPLEXITY OF MODELS PROGRESSIVELY AND DO NOT HESITATE TO COMPARE THEM WITH TRADITIONAL APPROACHES .....	39
<b>Section 5: Future Developments and Insights .....</b>	<b>40</b>
SECTION 5.1 CHANGES TO ACTUARIAL PROFESSION .....	40
SECTION 5.1.1 PRICING ACTUARY .....	40
SECTION 5.1.2 VALUATION AND RESERVING .....	41
SECTION 5.1.3 ENTERPRISE RISK MANAGEMENT .....	41
SECTION 5.1.4 EXPERIENCE ANALYSIS AND ASSUMPTIONS .....	41
SECTION 5.2 OPPORTUNITIES AND EDUCATION REQUIREMENTS / SKILLS REQUIRED .....	42
SECTION 5.2.1 OPPORTUNITIES .....	42

<b>Glossary of Terms .....</b>	<b>44</b>
<b>Appendix A: References .....</b>	<b>46</b>
<b>Appendix B: Cloud Computing Survey Questions .....</b>	<b>50</b>
<b>About The Society of Actuaries .....</b>	<b>63</b>

## Acknowledgments

The authors wish to thank the Society of Actuaries and the Actuarial Innovation & Technology Program Steering Committee for sponsoring this paper, as well as the members of the Project Oversight Group for their assistance in preparing this paper. The Project Oversight Group is comprised of Katie Barron, Robert Barg, Nick Leimer, Min Mercer, Kathleen Wang, Trevor Howes, Justin Hanson and Erik Lie. In addition, the authors gratefully acknowledge the assistance of Korrel Crawford, Mervyn Kopinsky, and Dale Hall from the Society of Actuaries, who provided much appreciated leadership and coordination. We also thank Phil Borba and other members of the global Milliman community for their thoughtful feedback and peer review.

Lastly, the authors wish to thank the survey participants from companies around the world for providing valuable insights that helped to shape and enrich this report.

# Cloud Computing and Machine Learning uses in the Actuarial Profession

## Section 1: Executive Summary

Actuarial modeling has gone through several distinct phases largely defined by the underlying technology platform. Mainframe computing gave way to personal computers and DOS-based applications. The emergence of the Windows platform necessitated new architectures and enabled new features. Today, cloud computing is again redefining the technology foundations of actuarial modeling and related processes.

The “cloud” enables a fundamental shift in how models are developed, deployed, and maintained. The cloud includes a wide array of offerings—software as a service, platform as a service, infrastructure as a service, etc.—that provide different benefits to support the modeling function, which can be deployed through public, private, community, or hybrid cloud models.

Migrating to the cloud allows for integrated model governance, centralized data, collaboration and version control, and other benefits to the modeling process. In addition, cloud storage is relatively cheap and the amount of data collected and stored is growing exponentially. The types of analytics applied to this data are also expanding, with increased applications of machine learning (ML) and artificial intelligence (AI).

Actuaries are currently and will continue to be impacted by the growth of cloud storage and cloud computing. The cloud has the potential to reshape the actuarial function. This will create opportunities for actuaries to improve existing methods (for governance, scalability, automation, etc.), deepen the analyses in existing areas (e.g., more granularity means more closely refined assumptions), and expand to new and different analyses enabled by the cloud (e.g., machine learning, predictive analytics).

With these advances, actuaries are in a position to leverage their current skills to new areas within and outside of traditional roles at insurance companies. However, there is also the potential to misuse data and advanced analytical techniques. Actuaries must be aware and proficient in their understanding of the appropriate analytic techniques, data, and applications.

This research is presented in the following four sections:

1. **Introduction and Background:** This section provides an introduction to the cloud service models and their impact on the actuarial profession. The section also provides some insights into a survey conducted for this research, polling insurance companies from around the world on their current and future uses of the cloud. The survey results indicate actuaries believe the cloud will be a benefit to the profession.

Industry publications indicate the enterprise cost of data storage and computation will decrease by migrating to a cloud platform. However, in the authors' experience, if left unchecked, the technology costs allocated to actuarial / analytics departments could increase as a result of increased computational capabilities and associated fees, software license fees, and staffing additions with sufficient experience developing cloud-based analytic frameworks.

2. **Financial Modeling, Actuarial Processes, and the Cloud:** This section discusses the use of the cloud in terms of financial modeling and actuarial processes. In this section, we discuss the different types of cloud environments,

how actuaries can or are currently leveraging the cloud to improve the model life cycle, and deployment options for leveraging the cloud in a production environment.

Actuaries are leveraging the cloud to optimize model life cycles, enhance reporting, and improve model governance. In a cloud framework, data and programs can be seamlessly integrated across an organization, which reduces the number of steps required to update and run complex model processes. This can lead to an increased frequency of model updates, decreased run-time and/or increased capacity to perform complex stochastic modeling, and frequency of enterprise-wide capital analysis. The results of any analysis can be integrated into a reporting framework that is delivered through browser-based tools, providing senior management more frequent and timely information on the financial status of the enterprise. Further, the entire process can be integrated with model governance tools to ensure data accuracy, reduce human error, and increase auditability of the process.

3. **Machine Learning / AI and Cloud Data Impact on Actuarial Science:** This section discusses the use of the cloud in terms of the increased ability to collect more data (and more granular forms of data) to perform advanced analytics. We discuss common uses of the cloud today for these types of advanced analyses, including case studies across actuarial practice areas and considerations for actuaries in deploying machine learning and AI models. This section also provides an introduction into common machine learning techniques used by insurance companies.

Insurance companies are using the cloud and machine learning algorithms to enhance pricing algorithms, analyze and segment data into actionable insights, and process granular data in a more efficient methodology. Companies may be collecting individual health data from consumer wearables, telematics data for auto insurance, or other types of large and granular datasets. Machine-learning algorithms are efficient for analyzing this type of data.

4. **Future Development and Insights:** This section provides an outlook for the potential ways expanded use of cloud technology combined with machine learning / AI could impact the actuarial profession as the technology evolves. We discuss ways for actuaries to leverage the cloud and machine-learning techniques in their work, and we provide some considerations for actuaries as the demand for newer technologies and advanced analysis increases.



## Section 2: Introduction and Background

### SECTION 2.1 WHAT IS THE CLOUD?

Cloud computing is now a ubiquitous term alongside big data and data science. As of 2018, over 70% of insurers reported using some type of cloud-based service, a proportion that was only a third that size just a few years ago<sup>1</sup>. As with all new technology, the terminology tends to proliferate faster than the understanding of what it is and what, if any, advantages it may offer. For the purposes of this report, **cloud computing** is the internet-based outsourcing of digital storage along with access to a collection of resources, including software, databases, data warehouses, and computational processing. This includes both public cloud options (i.e. third-party providers such as Amazon Web Services and Microsoft's Azure), as well as private clouds (i.e. internal clusters of shared computing resources). This compares to an "on-premise" computing architecture where all hardware and software are located on the premises. Data can be accessed locally through a network by many users, but software is installed locally, and individually, for users. When there are software updates, each user must install the updates, or rely on their IT department to push updates. For actuaries, this can create version control issues if they are using two different versions of the same software, where there may be slight differences in the underlying algorithms or sampling methodologies. In a cloud architecture, this can be overcome by using shared computing resources that are maintained to ensure there is a single version of the software in production for all users.

Cloud computing allows organizations to save costs, enable collaboration, scale resources, and/or securely protect their data like never before. This is because the maintenance, server cost, and back-up are often provided for by the cloud service provider, who distributes the cost among many customers. Think of the cloud as the ability to rent versus buy a property (without a mortgage). The cost of renting is far less than the cost of buying a given property with cash at the time of the transaction. In addition, the landlord may provide additional services such as utilities, snow removal, and lawn care that one would otherwise assume in a purchase. Cloud providers purchase the required hardware and "rent" them to organizations through economies of scale. Note, in this context we are referring to the enterprise-wide cost of hardware and software licenses. Later in the paper, we discuss the actuarial cost of a cloud environment. Depending on the use and service model, the cost to actuarial or analytic departments may increase through the use of a cloud-based analytical framework.

Although cloud computing in its current form is a recent phenomenon, its historical threads are linked back to the 1950s, when "dumb terminals" were connected to massive and archaic mainframe computers<sup>2</sup>. Subsequently, in the 1960s, the Defense Advanced Research Projects Agency (DARPA)<sup>3</sup> collaborated with the Massachusetts Institute of Technology (MIT) to further expand the earlier technological capabilities for multiple-user access<sup>4</sup>. The term "virtualization" began to take shape as the early predecessor of virtual machines, an important technological construct for cloud services that did not quite hold the same meaning then as it does today. Due to the high cost of purchasing the computing equipment, the extremely limited computational capacity, and the maintenance required, these rudimentary versions of cloud computing were siloed, meaning they were limited to intra-organizational

---

<sup>1</sup> Wojciechowicz, Rebecca. (2018). "70% of Insurers Now Using Cloud Computing, With Additional Growth Planned, Says Novarica." *NOVARICA*. Available at: <https://novarica.com/70-insurers-now-using-cloud-computing-additional-growth-planned-says-novarica/>.

<sup>2</sup> Neto, Maximillano Destefani. (2014). "A brief history of cloud computing." *IBM*. Available at: <https://www.ibm.com/blogs/cloud-computing/2014/03/18/a-brief-history-of-cloud-computing-3/>.

<sup>3</sup> DARPA. (2019). "About DARPA." Available at: <https://www.darpa.mil/about-us/about-darpa>.

<sup>4</sup> Foote, Keith D. (2017). "A brief history of cloud computing." *Dataversity*. Available at: <https://www.dataversity.net/brief-history-cloud-computing/>.



systems and processes. Additionally, there was a crucial component missing: a way to link a broader array of computers to a centralized processing system.

The initial seed of what we know today as the internet was planted by J.C.R. Licklider, a computer scientist who—in 1963—envisioned an “Intergalactic Computer Network” (ICG)<sup>5</sup> where an interlinking homogenous “language” would allow the speedy sharing of data between individual computers<sup>6</sup>. At the close of the 1960s, Licklider’s ICG began to materialize through the Advanced Research Projects Agency Network (ARPANET), which produced the technological foundations for “Resource Sharing Computer Networks”<sup>7</sup>. While a section of the scientific community labored to create more efficient and broader connections between computer systems, IBM’s experts worked to improve their earlier versions of virtualization via various iterations of their mainframe systems<sup>8,9</sup>.

Prior to IBM’s advances in virtualization, batch processing was the main method of operation<sup>10</sup>, but during the early years of the technology it could only run a single program at any given time. As the amount of data began to grow, and an increasing number of enterprises were adding computerized functions to their daily work, a way to partition the workload was needed where computational resources could be shared without disrupting the other users<sup>11</sup>. Thus, by the 1970s, virtual machine technology was well on its developmental path.

The 1980s and 1990s were the eras of personal computers and the expansion of telecommunication services, which fueled the cloud-computing evolution and set the stage for massive data collection via the World Wide Web. Indeed, cloud computing has been a fundamental factor in the development of the web, transitioning from static websites (known as “Web 1.0”) to the proliferation of technologies being applied to provide individual experiences based on specific profiles (referred to as “Web 3.0”) <sup>12</sup>, underscoring a continual feedback loop increasing the amount of data that can be collected, while also deploying more complex analytical models for faster and more accurate diagnostic, predictive, and prescriptive insights.

Today, many organizations are evaluating shifting their computing infrastructure from on-premise servers to cloud-hosted solutions. There are many considerations for this evaluation including the security of the data, the short- and long-term costs of maintenance, updates, and fees, computational demands and capabilities, and scalability. The sections below discuss different service models for the cloud, deployment models for the cloud, and potential benefits of cloud computing.

### 1. Service models (e.g., SaaS, IaaS, PaaS)

Cloud-based services as we know them today, or cloud services, is a meta-category term that contains three main service models: software as a service (SaaS), infrastructure as a service (IaaS), and platform as a service (PaaS). As

<sup>5</sup> Licklider, J.C.R. (1963). “Memorandum for members and affiliates of the Intergalactic Computer Network.” Available at: <http://www.kurzweilai.net/memorandum-for-members-and-affiliates-of-the-intergalactic-computer-network>.

<sup>6</sup> Ibid.

<sup>7</sup> Hauben, Michael. (1995). “Behind the net: The untold story of the ARPANET and computer science.” Available at: <http://www.columbia.edu/~rh120/ch106.x07>.

<sup>8</sup> Conroy, Sean. (2018). “History of virtualization.” Available at: <https://www.idkrtn.com/history-of-virtualization/>.

<sup>9</sup> Oracle. (2012). “Brief history of virtualization.” Available at: [https://docs.oracle.com/cd/E26996\\_01/E18549/html/VMUSG1010.html](https://docs.oracle.com/cd/E26996_01/E18549/html/VMUSG1010.html).

<sup>10</sup> IBM. (2010). “What is batch processing?” Available at: [https://www.ibm.com/support/knowledgecenter/zosbasics/com.ibm.zos.zconcepts/zconc\\_whatbatch.htm](https://www.ibm.com/support/knowledgecenter/zosbasics/com.ibm.zos.zconcepts/zconc_whatbatch.htm).

<sup>11</sup> Kohlbrenner, Eric; Morris, Dana; Morris, Brett. (n.d.). “The history of virtual machines.” Available at: <http://denninginstitute.com/itcore/virtualmachine/history.htm>.

<sup>12</sup> Expert Systems. (n.d.). “What is web 3.0?” Available at: <https://www.expertsystem.com/what-exactly-is-web-30/>.

happens frequently within competitive markets, other niche cloud-based services also known as anything as a service (XaaS)<sup>13</sup>—have surfaced, including managed software as a service, data as a service, AI as a service, payments as a service, and so on<sup>14</sup>. At this time, however, the primary service models are the dominant use cases across industries.

### SECTION 2.1.1 SAAS

The SaaS model, also frequently referred to as “on-demand software,” provides immediate access to software and applications that have been placed on cloud servers, usually by third-party vendors. SaaS is the most popular option and is also the most comprehensive option. Vendors manage all systemic items, e.g., applications, data, run time, middleware, etc. Information technology (IT) staff do not have to download updates or maintain systems. For example, consider Gmail. Google manages all the hardware, all the software, and the operating systems. Users easily log onto the web to see their email. There is minimal configuration or updating involved.

Most software in use today is deployed using the SaaS model: Office 365, Slack, Google Docs, and Adobe Creative Cloud are a few examples of commonly known (and used) SaaS packages. As such, SaaS holds the largest market share amongst the three service models. Approximately 66% of the global public cloud revenue generation stems from SaaS, and its total worldwide revenue is expected to reach \$346 billion by the year 2027<sup>15</sup>. Gone are the days of purchasing hard copies of software that would need to be repurchased whenever the newest version was released—or holding on as long as possible to the prior version until the most current operating system no longer supported it.

### SECTION 2.1.2 IAAS

IaaS providers offer on-demand infrastructure such as servers, storage, virtual machines, and networking capabilities. Essentially, an enterprise is renting the infrastructure for use via web-based access while still being responsible for managing applications, data, run time, middleware, and operating systems<sup>16</sup>. Because hardware is such an expensive part of large-scale computing, it often makes sense to turn that into a continuously variable cost. IaaS makes it easy to switch gears. Companies can switch what kind of software runs on the computers. Companies can switch how much hardware they use. For the most part, managers can make the same decisions they could make if the systems were housed in the local office.

The IaaS vendor is responsible for the management and maintenance of the aforementioned components, which can significantly reduce the costs associated with purchasing, implementing, and maintaining both the hardware and software on-premise. In IDG’s 2018 Cloud Computing study, 73% of the respondents stated that they had “at least a portion of their computing infrastructure already in the cloud”<sup>17</sup>. Gartner forecasts that the IaaS portion of

---

<sup>13</sup> McLellan, Charles. (2017). “XaaS: Why ‘everything’ is now a service.” ZDnet.com. Available at: <https://www.zdnet.com/article/xaas-why-everything-is-now-a-service/>.

<sup>14</sup> Tribe, Jennifer. (2016). “The big -aaS list of As-a-Service offerings.” Available at: <https://www.auvik.com/franklymsp/blog/aas-as-a-service-list/>.

<sup>15</sup> Statista. (2019). “Public cloud Software as a Service (SaaS) revenue worldwide from 2016 to 2027 (in billion U.S. dollars).” Available at: <https://www.statista.com/statistics/477742/public-cloud-software-revenue-forecast/>.

<sup>16</sup> Harvey, Cynthia. (2018). “IaaS vs PaaS vs SaaS: Which should you choose?” *Datamation*. Available at: <https://www.datamation.com/cloud-computing/iaas-vs-paas-vs-saas-which-should-you-choose.html>.

<sup>17</sup> IDG. (2018). 2018 Cloud Computing Survey. Available at: <https://www.idg.com/tools-for-marketers/2018-cloud-computing-survey/>.

the cloud services market will increase from \$31 billion in 2018 to \$63 billion by 2021<sup>18</sup>. Amazon AWS, Microsoft Azure, Google Compute Engine, Rackspace Cloud, and BM SmartCloud Enterprise are the major competitors within the IaaS space.

### SECTION 2.1.3 PAAS

Under PaaS, a cloud service provides the platform that a company can use to make and distribute software. The cloud service provides a basic framework for building applications and programs. It supplies the hardware and delivers the software a company needs to build its own applications. For example, a cloud provider would supply the operating systems (e.g., Windows 10 or Mac OS). In addition, it would provide other applications that might make it easier to run programs. It handles the software updates, the system upgrades, and the infrastructure. PaaS retains both flexibility and ease of use. Using these platforms means companies do not have to start from scratch building programs they want to deploy. This saves a lot of time and money. They are easy to run without much system administration expertise.

The PaaS model expands upon the vendor-managed functions of the IaaS model, enabling the company to focus its IT resources on application development and deployment, maintenance of applications, and data management (but not necessarily storage). All other resources (e.g. servers, storage, networking, etc.) remain the responsibility of the PaaS vendor. With regard to market share, PaaS currently lags behind SaaS and IaaS, but Gartner projects a steady revenue increase, from \$18.8 billion in 2019 to \$27.7 billion by 2021<sup>19</sup>. Salesforce, Heroku, and Google App Engine are a handful of well-known PaaS providers.

Most companies today use all of the service models discussed above. For the purposes of this report, we will focus mostly on IaaS. Our discussions in the report, and their application to actuaries, generally focus on increased storage, computing capacity, and governance that can be provided through the use of the cloud.

## 2. Deployment models (e.g., public, community, private, hybrid)

At this time, enterprise use of public clouds outpaces other deployment models such as private, hybrid, or community. The RightScale-Flexera’s 2019 State of the Cloud Report said that 91% of the organizations surveyed have adopted a public cloud, and 31% considered the public cloud as their “top priority”<sup>20</sup>. Notably, within the same report, 84% are deploying a multi-cloud strategy, using a combination of public and private clouds, with the public clouds distributed across different cloud service providers<sup>21</sup>. Overall, enterprises and organizations throughout all industries are adopting one or more of the following deployment models:

- **Public cloud:** IT resources are shared amongst vendor customers, and the cloud environment is owned, as well as maintained, by the third-party vendor (e.g., Amazon, Microsoft, Oracle, IBM).
- **Private cloud:** IT resources may either be on-premise or deployed via a third-party vendor, but access to the services and/or infrastructure is limited to a single organization. Private clouds are a common

---

<sup>18</sup> Gartner. (2018). “Gartner forecasts worldwide public cloud revenue to grow 17.3 percent in 2019.” Available at: <https://www.gartner.com/en/newsroom/press-releases/2018-09-12-gartner-forecasts-worldwide-public-cloud-revenue-to-grow-17-percent-in-2019>.

<sup>19</sup> Ibid.

<sup>20</sup> RightScale-Flexera. (2019). “State of the Cloud Report.” Available at: <https://info.flexerasoftware.com/SLO-WP-State-of-the-Cloud-2019>

<sup>21</sup> Ibid.

deployment model for enterprises and institutions that require adherence to regulatory requirements or those that need greater cloud environment control.

- **Hybrid cloud:** IT resources are distributed across different cloud-based systems, such as utilizing a public cloud for data that has been identified as non-sensitive, while employing a private cloud for internal business systems and processes that demand the safeguarding of intellectual property and/or sensitive data.
- **Community cloud:** IT resources are used within the context of several organizations having permissioned access to a select set of applications. Generally, the organizations have a shared industry or objective, such as insurers or financial institutions (e.g., compliance, privacy, security)<sup>22</sup>.

There are additional variations of cloud deployment that have not yet fully emerged as broadly discussed use cases. These include federated cloud, micro cloud, cloudlet, ad hoc cloud, and heterogeneous cloud<sup>23</sup>. It cannot be overstated that all technologies, including cloud services, are always changing. New service and deployment models will come to light as the cloud services industry matures.

When deciding which deployment model is right for a given organization, general considerations are the sensitivity of the data (e.g. does it contain personally identifiable information or protected health information, PII and PHI respectively), level of customization required (i.e. does your organization need a specific type of dedicated environment), and scalability. Public clouds typically have a lower cost given the shared resources of the cloud and are more scalable (i.e. more resources can be easily added to your cloud environment on demand). Private clouds can offer an additional level of security (i.e. they have dedicated hardware to one organization) and customization. However, private clouds will have a slightly higher cost relative to a public cloud, all else being equal.

### 3. Main benefits to using the cloud

Overwhelmingly, the primary benefit of cloud utilization to a given organization is the decrease in enterprise costs<sup>24,25,26,27,28,29</sup>. McKinsey estimates that, by using the cloud, enterprises can “reduce IT overhead costs by 30 to 40 percent”<sup>30</sup>. Regardless of the service or deployment models, the offloading of hardware and software purchases,

<sup>22</sup> U.S. National Institute of Standards and Technology. (2011). “The NIST Definition of Cloud Computing.” Available at: <https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-145.pdf>.

<sup>23</sup> Vaghese, Blesson; Buyya, Rajkumar. (2017). “Next generation cloud computing: New trends and directions.” *Future Generation Computer Systems*, ISSN: 0167-739X, Elsevier Press, Amsterdam, The Netherlands, 2017 (in press). Available at: <https://arxiv.org/pdf/1707.07452.pdf>.

<sup>24</sup> Microsoft Azure. (2019). “What are public, private, and hybrid clouds?” Available at: <https://azure.microsoft.com/en-us/overview/what-are-private-public-hybrid-clouds/>.

<sup>25</sup> Salesforce. (2015). “Why move to the cloud? 10 benefits of cloud computing.” Available at: <https://www.salesforce.com/uk/blog/2015/11/why-move-to-the-cloud-10-benefits-of-cloud-computing.html>.

<sup>26</sup> Webroot. (n.d.). “5 financial benefits of moving to the cloud.” Available at: <https://www.webroot.com/au/en/resources/tips-articles/five-financial-benefits-of-moving-to-the-cloud>.

<sup>27</sup> Amazon Web Services. (n.d.). “Six advantages of cloud computing.” Available at: <https://docs.aws.amazon.com/aws-technical-content/latest/aws-overview/six-advantages-of-cloud-computing.html>.

<sup>28</sup> Bozicevic, Vedran. (2018). “Cloud computing benefits: 7 key advantages for your business.” *GlobalDots*. Available at: <https://www.globaldots.com/cloud-computing-benefits/>.

<sup>29</sup> Goasduff, Laurence. (2015). “The financial case for moving to the cloud.” *Gartner*. Available at: <https://www.gartner.com/smarterwithgartner/the-financial-case-for-moving-to-the-cloud/>.

<sup>30</sup> McKinsey Digital. (2018). “Cloud adoption to accelerate IT modernization.” Available at: <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/cloud-adoption-to-accelerate-it-modernization>.

as well as maintenance—and, by extension, personnel resources—is a decisive cost reduction factor. Granted, a hybrid cloud model may mitigate cost reduction due to a portion of the cloud functions being sourced from on-premise components (e.g., data center and servers). Still, the on-demand cost calculations for public cloud services, along with other agile cost offerings via the cloud service provider<sup>31</sup>, lend themselves to higher probabilities of lowering IT costs.

For actuaries, scalability is another potential significant benefit. With on-premise IT, scalability is limited by hardware and software considerations. Adding servers to the existing infrastructure is costly, and there is a high risk of downtime throughout the process. Additional capital expenditures require large, up-front investments of time and money. For actuaries needing to perform complex simulations or estimate complex models, the ability to quickly increase computing capacity and/or storage is an attractive feature of using cloud computing versus on-premise computing. Cloud services offer immediate upscaling or downscaling of resources. Consequently, as long as usage costs are monitored<sup>32</sup>, enterprises have greater agility to meet increasing resource demands. It is worth noting that, without formal data and use governance, the costs for cloud computing could exceed on-premise costs or budget for actuarial modeling (it is important to point out the overall enterprise cost may decrease, but the costs allocated to modeling and analytics might increase). This would occur in situations where various teams within an actuarial department run several large and complex models using many nodes in a cloud environment.

Governance and data security are important discussion areas when evaluating the use of the cloud. With increased scrutiny on data privacy, it is important to evaluate how secure data is on external (i.e. cloud) systems. Cloud providers are aware of these concerns and have been spending significant resources ensuring data security. Private clouds offer a solution for companies to increase data security while leveraging the benefits of cloud computing and storage. Cloud providers and users have also developed tools to improve data and model governance, such as Git, which facilitate version control for data processes and complex statistical models.

Due to the extensive server network that cloud service providers possess, their systemic protocols are highly resistant to incidences of failure. Certainly, cases of failure do occur. However, there is less risk of a complete system failure as cloud service providers can redirect resource needs to other servers while they identify and solve the cause of issues. Furthermore, cloud service providers will mirror “the data and applications across at least two” of their data centers<sup>33</sup>. As such, the likelihood of data loss and the previously mentioned downtime are also greatly reduced.

Regardless of the industry, enterprises and institutions exist within a global marketplace. Whether they are sending data to and receiving data from consumer apps or their workforces and clientele dispersed throughout the world, 24/7 access to computational resources, business documents, and other systems and processes is now an essential feature. Both public and private cloud services easily accommodate access via any type of device, from anywhere in the world, at any time of the day or night. The organization can still maintain control over accessibility while also providing the benefit of mobile access and increased collaboration<sup>34</sup>.

---

<sup>31</sup> Harvey, Cynthia. (2017). “Cloud computing cost calculators.” Available at: <https://www.datamation.com/cloud-computing/cloud-costs-calculators.html>.

<sup>32</sup> RightScale-Flexera. (2018). “Cloud cost optimization.” Available at: <https://www.rightscale.com/learn/cloud-optimization/cloud-cost-optimization-best-practices>.

<sup>33</sup> Webroot. (n.d.). “5 financial benefits of moving to the cloud.” Available at: <https://www.webroot.com/au/en/resources/tips-articles/five-financial-benefits-of-moving-to-the-cloud>.

<sup>34</sup> Salesforce. (2015). “Why move to the cloud? 10 benefits of cloud computing.” Available at: <https://www.salesforce.com/uk/blog/2015/11/why-move-to-the-cloud-10-benefits-of-cloud-computing.html>.

## SECTION 2.2 IMPACT ON ACTUARIES

### 1. The cloud: Services, applications, and data applicability to actuaries

The impact of migration to cloud services from an actuarial perspective involves several interrelated internal and external forces, which include ever-shifting consumer and technological demands (e.g., digital natives, dynamic risk management, machine learning / AI), as well as changes in regulatory frameworks. While it is difficult to pinpoint the quantitative effects of these influences, they will assuredly lead to further changes in the actuarial operating model.

At the external marketplace level, consumers and technological advances have converged to reveal a largely digital demographic whose dependence on their mobile devices is steadily increasing<sup>35</sup>. In a study conducted by Applied Systems, 74% of the Millennial and Gen Z respondents selected mobile app access as an important feature in selecting an insurance provider<sup>36</sup>. Accordingly, Millennials and Gen Z are generational cohorts who want their initial contact with insurers to be digital, where they can receive immediate and individualized quotes for insurance services<sup>37</sup>. Additional insurer preferences include Omni channel and 24/7 customer service availability, a humanistic brand they can trust, and evidence of a brand's social consciousness<sup>38</sup>. These preferences indicate the requirement of a strong online presence that does not merely promise an individualized approach<sup>39</sup>. Considering that these digital natives are very quickly becoming the dominant consumer class, and that Millennials alone will soon surpass Baby Boomers as the largest generation<sup>40</sup>, it is imperative for insurers to continually find ways to actively engage this generation (as well as Gen Z).

Technological and consumer transformation breed market competitors who have the agility for faster adaptation. Insurtech is a prime example of where these new, tech savvy, and generationally aware enterprises are emerging. Lemonade is a home and renters insurance provider<sup>41</sup> that currently has a first iteration with all of the qualities younger consumers demand:

- A highly responsive website written in a conversational tone
- A swift return on a quote, which leads the customer to a user-friendly dashboard for choosing a coverage amount and policy options, with clear statements on what is not covered—including a pop-up detailing information such as which dog breeds are considered high risk, etc.

---

<sup>35</sup> Pew Research Center. (2018). "Mobile fact sheet." Available at: <https://www.pewinternet.org/fact-sheet/mobile/>.

<sup>36</sup> Applied Systems. (2018). "The next generations and insurance: Understanding buying behaviours and preferences of younger adults." Available at: [https://www1.appliedsystems.com/media/3203/understanding-millennial-gen-z-buying-behaviors\\_en-uk.pdf](https://www1.appliedsystems.com/media/3203/understanding-millennial-gen-z-buying-behaviors_en-uk.pdf).

<sup>37</sup> Gallup. (2015). "Insurance companies have a big problem with Millennials." Available at: <https://news.gallup.com/businessjournal/181829/insurance-companies-big-problem-millennials.aspx>.

<sup>38</sup> Applied Systems. (2018). "The next generations and insurance: Understanding buying behaviours and preferences of younger adults." Available at: [https://www1.appliedsystems.com/media/3203/understanding-millennial-gen-z-buying-behaviors\\_en-uk.pdf](https://www1.appliedsystems.com/media/3203/understanding-millennial-gen-z-buying-behaviors_en-uk.pdf).

<sup>39</sup> Turner, Heather. (2018). "What do millennials want from their insurance agency?" *Insurance Business America*. Available at: <https://www.insurancebusinessmag.com/us/news/technology/what-do-millennials-want-from-their-insurance-agency-115007.aspx>.

<sup>40</sup> Fry, Richard. (2018). "Millennials projected to overtake Baby Boomers as America's largest generation." *Pew Research*. Available at: <https://www.pewresearch.org/fact-tank/2018/03/01/millennials-overtake-baby-boomers/>.

<sup>41</sup> Lemonade. (2019). Available at: <https://www.lemonade.com/>.

- Near-instant comparison quotes
- Easy access to quotes, payment processing, and policy documents through web and mobile channels

On the vendor side of insurance, App Orchid provides enterprise-level machine learning and AI services to “solve business problems that have historically remained untouched in the insurance industry”<sup>42,43</sup>. Actuaries now have more than internal data to factor into their models. Almost all generations have converged onto social media. The increase in industrial Internet of Things (IoT) and the widespread use of personal health tracking apps and other data-intensive technologies are enormous data sources that can now be tapped for building more complex risk models. Unstructured data streaming from a variety of sources also poses data collection, storage, and feature engineering challenges. Insurtech enterprises such as App Orchid are leveraging innovations in cloud services and advanced machine learning or AI capabilities.

With regard to the impact on insurers, and ultimately on actuaries, the business will require ongoing technological adjustments (e.g., mobile apps that can deliver initial real-time quotes and data gathering from various devices) and advanced analytical capabilities for dynamic pricing<sup>44</sup>. The data actuaries use to construct risk models will, in itself, be dynamic.

The cloud is also having an impact on the way actuaries conduct work. Actuaries are finding that they can slash time on large data projects and the construction of machine-learning models. Others are building solutions into the cloud, allowing actuaries to work together on projects that are both scalable and available at all times. The cloud opens up the possibilities for work actuaries have always done using data that is more granular compared to aggregated historical information. With the cloud, actuaries can compute in parallel, which allows them to finish projects in a matter of hours that previously could have taken months. Actuaries can have more time to review and validate their work. Presenting their work to regulators becomes much easier when organizations can share results online effortlessly.

In recent years, actuaries have seen more extensive data collecting, more need for stress testing, and more complicated demands for modeling. Along with the increasing demands of the Solvency II requirements and the institution of International Financial Reporting Standards (IFRS), there will be a need to ramp up capacity at certain times of the year. The cloud is helping actuaries complete the vast financial modeling necessary for reporting in the modern era. With the move from deterministic modeling to stochastic modeling, many actuaries are discovering that cloud computing offers the benefits of being able to use computing power tailored to the task. Actuaries can use what they want when they want.

## 2. More data and enhanced access: impact on actuarial work today and in the future

Already we are seeing actuaries adopt cloud technologies in significant numbers. Milliman conducted a survey for this research report, polling insurance companies on their current and anticipated use of the cloud (see the Appendix for the results of the survey). According to this survey, 29% of respondents are using the cloud on a daily basis, and 52% are using the cloud at least once a month. The use of the cloud was generally consistent between practice areas (Life, Health, Property & Casualty, and other). In this context, using the cloud is inclusive of data storage, processing,

---

<sup>42</sup> CIO Applications. (n.d.). “App Orchid: Harnessing the power of AI.” Available at: <https://insurtech.cioapplications.com/vendor/app-orchid-harnessing-the-power-of-ai-cid-1713-mid-96.html>.

<sup>43</sup> App Orchid Claims 360. (2019). Available at: <https://www.apporchid.com/claims-360/>.

<sup>44</sup> Deloitte. (2018). “Dynamic pricing is not just about profitability.” Available at: <https://www2.deloitte.com/nl/nl/pages/financial-services/articles/dutch-insurance-outlook-2017-dynamic-pricing.html>.



and collaboration. The most common use for actuaries is leveraging the cloud for faster computation (that is, for distributed computing) and model development. Specifically, at least 75% of the respondents reported they are either currently using or plan to use some type of distribution calculation engine within the next year, and at least 70% of respondents are either currently using or have plans to use the cloud for model development. The use of distributed computing is most common in the life insurance sector.

The view of the impact of the cloud on actuarial work is positive. A full 57% of respondents see the cloud as having a large positive impact on their work, and 83% see at least a small positive impact. In addition, many of the respondents believe the use of the cloud will grow over the next five years, with 77% indicating the cloud is likely to become more prevalent within their actuarial departments over the next five years.

### **3. Faster and more advanced analytical capabilities: impact of actuarial work today and in the future**

Respondents to our survey also expect to see a positive qualitative impact from the cloud. Specifically, 73% of respondents indicated using cloud services would have a positive impact of their workflow, with 71% indicating the cloud is already important, or would be in the future, to performing their job functions. The positive impact can come from increased abilities to perform advanced analytics, faster run time, enhanced collaboration, and other factors.

The following three bullet points are results from the survey indicating what actuaries view as the most positive impact on the cloud:

- 70% of our respondents believe that the cloud delivers faster and more advanced analytical capabilities;
- 47% of the survey respondents believe the cloud brings more centralized processes and increases collaboration; and
- Actuaries see a cost saving component to using the cloud, with 20% of the survey respondents indicating there is a benefit in the cost savings of the cloud.

### **4. Professional impact**

The cloud and the data that comes with it are changing the shape of the actuarial profession. Consider, for example, the addition of the data scientist to the insurance company team. In recent years, insurance companies have been adding data from external sources that may be structured or unstructured. Managing this data requires programming experience that exceeds the capabilities and expertise of many actuaries (or at least the capabilities required through the examination process).

Historically, data scientists started out as statisticians, but as computers became more powerful and data storage cheaper, some of these statisticians started making calls for bringing together the fields of computer science, information science, statistics, and mathematics. The idea was to bring the scientific method together with statistics to analyze the large amount of data that organizations could now store. Today, the term “data science” has become watered down, with little consensus as to what it means.

Still, modeling actuaries are beginning to take on data science techniques. Using predictive analytics and unstructured data, actuaries are adapting to gain insights into their work. Some companies are training actuaries to use data science methods in-house. Modeling actuaries are also getting in the mix by combining these statistical modeling techniques along with their specialized training in insurance, statistics, economics, and probability. These actuaries have found that they are able to get the best of both worlds when they combine their specialized training with the broad spontaneity of data science.

## Section 3: Financial Modeling, Actuarial Process, and the Cloud

It is truly an exciting time in insurance. The accelerating pace of technological innovation presents a rare and valuable opportunity for insurers to strategically reimagine the way they do business. AXA uses blockchain and smart contracts on the Ethereum network to offer parametric insurance against delayed flights with an entirely automated back end connected to flight databases to immediately pay policyholders upon a delayed flight<sup>45</sup>. Farmers Insurance uses virtual reality to simulate hundreds of home damage scenarios to better train their claims representatives while saving on significant travel expenses<sup>46</sup>. MetLife embraced Docker (according to Amazon Web Services, “Docker is a software platform that allows you to build, test, and deploy applications quickly. Docker packages software into standardized units called containers that have everything the software needs to run, including libraries, system tools, code, and runtime”), projecting tens of millions in expense savings<sup>47</sup> by reducing their number of virtual machines by 70%<sup>48</sup> and removing the headache of maintaining apps across a variety of environments.

Along with building capabilities in-house, insurers are investing in private technology companies, with a total of 120 private placements in 2017, according to Willis Towers Watson Securities and Willis Re<sup>49</sup>. Maximizing these partnerships and leveraging the innovation pouring into the industry depends on an agile, flexible, and scalable technological platform that rigid legacy systems running on-premise were not originally built for. Many are looking to the cloud for this required agility.

As mentioned previously, the majority of insurers plan to, or have already made, a move to cloud solutions, with life insurers and pension administrators moving at the same speed as property and casualty providers. In Figure 1, Accenture provides the breakout by development phase<sup>50</sup>.

---

<sup>45</sup> AXA. (2017). "AXA goes blockchain with fizzy." Available at: <https://group.axa.com/en/newsroom/news/axa-goes-blockchain-with-fizzy>.

<sup>46</sup> Farmers Insurance. (2017). "First-of-its-kind training module's immersive, virtual environments deliver dynamic learning experience for claims representatives." Available at: <https://www.farmers.com/news/2017/farmers-insurance-pioneers-innovative-virtual-reality-training-program-for-claims-representatives/>.

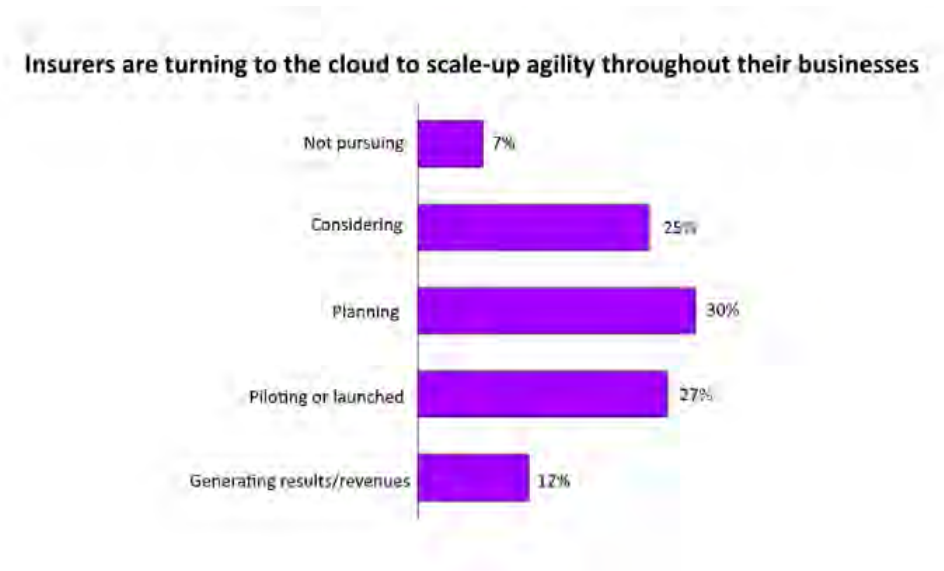
<sup>47</sup> Datamation. (2017). "MetLife navigates DevOps with Docker containers." Available at: <https://www.datamation.com/cloud-computing/metlife-navigates-devops-with-docker-containers.html>.

<sup>48</sup> Junod, Betty. (2017). "MetLife uses Docker Enterprise Edition to self fund containerization." Docker. Available at: <https://blog.docker.com/2017/10/metlife-docker-enterprise-edition-self-funded-containerization/>.

<sup>49</sup> Gangcuangco, Terry. (2018). "Willis Towers Watson reveals the extent of the insurtech boom." Insurance Business. Available at: <https://www.insurancebusinessmag.com/uk/news/technology/willis-towers-watson-reveals-the-extent-of-the-insurtech-boom-91200.aspx>.

<sup>50</sup> Lyman, Michael. (2018). "Insurers strive for greater agility to secure big digital pay-off." Accenture. Available at: <https://insuranceblog.accenture.com/insurers-strive-for-greater-agility-to-secure-big-digital-pay-off>.

**Figure 1**  
BREAKOUT BY DEVELOPMENT PHASE



Despite the massive interest, Accenture in early 2017 reported that some companies are still grappling with the cost of the cloud, with 35%<sup>51</sup> of insurers believing the cloud imposes a higher total cost of ownership (TCO) than existing IT infrastructure. This is in contrast with prior studies indicating a key consideration in migrating to the cloud is cost savings. These results indicate the cost-benefit analysis in evaluating a cloud-based framework is a complex process. Whether the cloud makes dollars and sense depends on the company’s existing tech infrastructure and chosen go-to-cloud strategy. The cost-benefit analysis involves many predictable and obvious expenses like housing, hardware, software, and networking, as well as countless less predictable ones like utilization rates, frequency of software updates, future server performance requirements, employee receptiveness (both in and out of the actuarial function), IT support, and management costs. However, focusing too much on the monetary estimates made today of simply lifting-and-shifting existing infrastructure entirely misses the point: the cloud positions enterprises to share in the most cutting-edge technological advancements made in the cloud ecosystem. Companies properly equipped with a dynamic technological foundation capable of capitalizing on these innovations will gain a real competitive advantage.

**SECTION 3.1 FINANCIAL MODELING**

The last several decades have witnessed tremendous advancements in computer software and hardware solutions. Perhaps as a result, the demand in the insurance sector for increasingly complex actuarial financial reporting has more than kept pace with these technological advancements. Reporting paradigms such as principle-based reserving (PBR), Actuarial Guideline 43 and C-3 Phase II, Solvency II, IFRS, long-duration targeted improvements under U.S. GAAP, and economic capital all require stochastic modeling over a range of economic scenarios. These high expectations are set by both insurance companies and the regulatory authorities whose mandate it is to ensure that insurance companies are able to stay solvent and meet their obligations to policyholders.

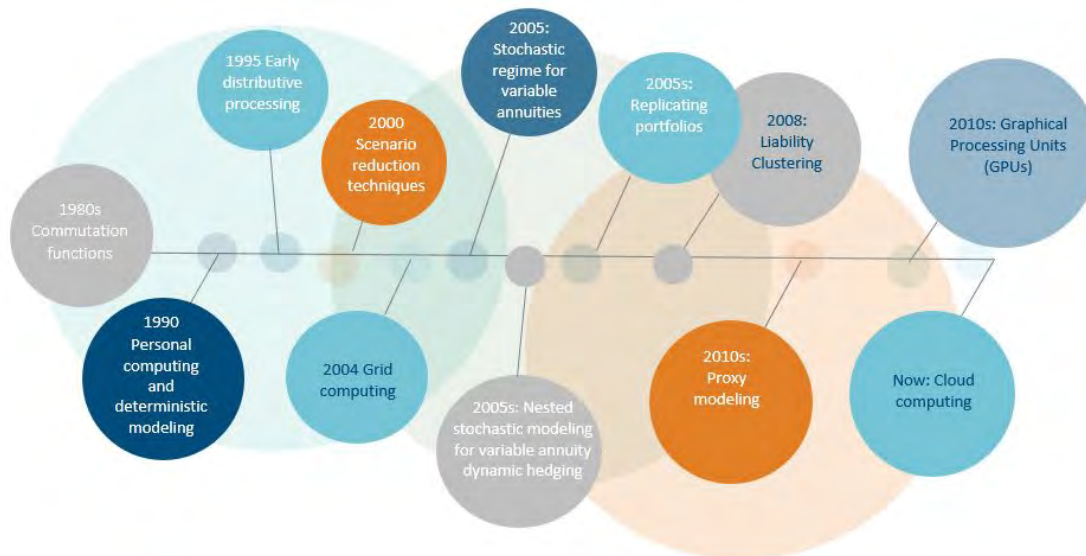
---

<sup>51</sup> Presutti, Daniele. (2017). "Cloud computing offers insurers big TCO savings." Accenture. Available at: <https://insuranceblog.accenture.com/cloud-computing-offers-insurers-big-tco-savings>.

The timeline below presents some important historical milestones for the insurance industry at large. Cloud computing is at the current forefront of technological opportunities for insurance companies and represents the next step in the evolution of actuarial financial reporting.

**Figure 2**

USE OF TECHNOLOGY IN THE INSURANCE SECTOR, HISTORICAL MILESTONES



The actuarial financial modeling process at insurance companies typically involves the following:

- Collecting liability and asset data,
- Analyzing this data and setting assumptions,
- Building, maintaining, and running a representative model that reflects the current insurance company portfolio, and
- Reporting financial results that are based in large part on the output from that underlying representative model.

The cloud will play a different role across the various components of this process as some pieces are less critical and not subject to as much regulation, while others require more internal controls, governance, and security.

**SECTION 3.2 DATA COLLECTION**

According to research conducted by the Boston Consulting Group, data-driven decision making poses a 6% to 9% upside in an insurer’s combined ratio: a key profitability measure that is the ratio of net claims over net premiums<sup>52</sup>. The first step in unlocking this potential is to collect data. Traditional data collection for financial modeling depends on the availability of structured data from a limited number of internal sources and information handoffs from one system to another using hard-coded business logic with tools constrained by storage and computing capacity. These

<sup>52</sup> Gandhi, Rash. (2018). "Building new data engines for insurers." Boston Consulting Group. Available at: <https://www.bcg.com/en-us/publications/2018/building-new-data-engines-insurers.aspx>.

batch jobs often take time, resulting in delays downstream for dependent processes, and are difficult to adjust in real-time (an advantage when there must be controls in place to prevent inadvertent errors and ensure data quality, but a disadvantage when trying to quickly test an idea). In contrast, the cloud has practically no limit on storage as it can scale arbitrarily large, with additional capacity at-the-ready. Different pricing tiers charge consumers depending on the required availability of data: frequently accessed data such as those critical to production work will cost the most to store, but relatively unused data like archives of prior month-end valuation work—which make up the bulk of an insurer’s data from an actuarial financial reporting perspective—can be kept in “colder” storage, which is slower to access but far cheaper. For example, Amazon’s Simple Storage Service (S3) starts at \$0.023 per GB but goes all the way down to \$0.00099 per GB<sup>53</sup>. Furthermore, cloud data can be replicated in multiple, separate servers to ensure that it is never lost and also quicker to access in different geographic regions.

Aside from capacity, the cloud is also packed with application programming interfaces (APIs) to ease connectivity to data of heterogeneous formats from third-party vendors and public records, making it easier to enrich internal policyholder information with external data sets. While comma-separated value (CSV) files are a straightforward and reliable data format, newer file types like Apache Parquet and Avro increase compression and optimize query performance. Investing on-premise runs the risk of locking into a particular set of tools, whereas partnering with cloud providers allows insurers to accept new features and capabilities on a rolling basis so that efficiencies, such as the Parquet format, can be quickly integrated. The API-based cloud ecosystem also makes it easy for internal processes to talk to one another so data can be quickly handed off from one process to the next, making automation easier and the development and deployment of new workflows quicker. Again, the pricing for these jobs can vary depending on the urgency. Interruptible, spot-cloud instances provide attractively priced options for low-urgency jobs so the cost of running the autonomous data collection, for example, can happen during off-peak hours for a fraction of the price.

More important than the massive capacity and connectivity of the cloud is the ability to effectively leverage the collected data. It is in the best interest of cloud providers to continuously improve and push out new analytics capabilities to customers through their platforms. For example, Amazon’s Timestream boasts the ability to store and apply machine-learning algorithms on trillions of daily transactions posted by the IoT<sup>54</sup>, a burgeoning field as wearables track physical fitness for life and health insurers and motor telematics relay driving behavior for auto insurers. Offerings such as Timestream require niche expertise and, rather than developing the talent in-house to capitalize on these new fields, insurers can utilize tools developed by cloud providers with the appropriate specialties. Somewhat less tangible, yet still valuable, cloud providers offer tools that act as a sort of social media layer on top of data repositories, where database stewards can define field names, list caveats of particular data sets, and respond to questions from users, providing greater taxonomy and, ultimately, more usable data.

It is important to note that, while most of this can be accomplished with an on-premise solution, a capital investment in bare metal leaves the insurer facing risks of missing out on future innovations that appear to be coming apace.

### SECTION 3.2.1 MODELING

Models play a foundational role in the actuarial financial reporting process. The actuary is involved with developing, validating, maintaining, running, and, ultimately, retiring the models that are used in this process. From a cloud perspective, there are two aspects to such models:

1. Migrating the entire model to the cloud.

---

<sup>53</sup> Amazon. "Amazon S3 Pricing." Available at: <https://aws.amazon.com/s3/pricing/?nc=sn&loc=4>.

<sup>54</sup> Amazon. "Amazon Timestream." Available at: [https://aws.amazon.com/timestream/?nc2=h\\_m1](https://aws.amazon.com/timestream/?nc2=h_m1).

2. Keeping models local (i.e., on-premise), but leveraging the scalable computing power available in the cloud to aid in distributed processing to run such models faster

There is likely to be some inertia to fully migrating actuarial models to the cloud because it is a bespoke process specific to insurance. Moreover, the responsiveness to change at insurance companies tends to vary from company to company—some companies have a culture that embraces such changes, whereas some are wary of new ways of doing things. Nevertheless, the prevailing sentiment is that this is the direction the industry is going and, accordingly, vendors are working on adding such functionality to their actuarial platforms in anticipation of the demand for such services. It will be interesting to see whether insurance companies fully utilize this option for all modeling tasks, particularly for models outside of financial reporting, such as those in the pricing and enterprise risk management functions where the modeling process tends to be different from valuation and financial reporting (the latter is performed on a regular basis while pricing, for example, tends to be ad hoc). Tech companies are focused on broader use cases so insurance applications are still in their infancy, although there is growing attention. By way of comparison, the banking sector generally does not have any intention of migrating their core processes and calculations off of old systems because they work well, provide total control over security, and are highly regulated. Some insurance companies may opt for a dual approach of preserving certain legacy systems while migrating noncore processes to the cloud. This is discussed in more depth in the “To the Cloud” section of this report below.

Perhaps more likely in the short term is the notion that companies will keep models local, but use the scalable distributed computing available in the cloud to run their models faster. Over the last several decades, the evolution of the actuarial models used by insurance companies has changed from single deterministic economic scenarios to stochastic sets of economic scenarios. It is now often necessary to perform either stochastic-on-stochastic (commonly referred to as nested stochastic) or deterministic-on-stochastic runs in order to see a distribution of emerging earnings or surplus for entire in-force portfolios across many economic scenarios, but with future reserves and/or capital requirements themselves calculated at certain preset nodes using stochastic approaches in accordance with the relevant accounting principles at hand. For purposes of nomenclature, we refer to the economic scenarios as “outer loop” scenarios and the scenarios used to calculate the future reserve and capital requirements as “inner loop” paths.

All models are, of course, only approximations, but the goal of building a useful model is to accurately capture as many of the key underlying risks as possible, while also carefully treading the line between parsimony and efficiency. This precept is particularly true for nested stochastic or deterministic-on-stochastic models, where a full-blown approach can easily result in geometric increases in run time due to the modeling of some combination of a large number of outer loop scenarios or inner loop calculations at which future reserves and/or capital requirements are calculated (either a large number of inner loop paths or a large number of nodes or both).

Generally, the more complicated the model the more time is needed for that model to be executed from start to finish. There are really only a few underlying drivers that can reduce model run time:

- Faster hardware and/or software
- Accepting simplifications in approach
- Access to more engines for distributed processing

The first bullet is, of course, a moving target—remarkably, both hardware and software continue to improve in accordance with Moore’s Law, which estimates that computing resources and/or capacity should double every two years. Certain technological advancements such as graphical processing units (GPUs) also have widespread use in noninsurance sectors and have started to be used by insurance companies as well.

The second bullet refers to the actuary adopting certain reasonable simplifications in approach and methodology in order to strike an appropriate balance between run time and accuracy. For example, the actuary could choose to model fewer outer loop scenarios, fewer nodes, and/or fewer inner loop paths for determining reserve and/or capital requirements, clustering of liabilities and/or assets, replicating portfolios or least squares Monte Carlo regressions, or possibly any combination of these. The litmus test here is to ensure that the simplified version of the model faithfully reproduces the results of the full-blown model, at least to within an acceptable margin of error. This test may need to be performed periodically<sup>55</sup>.

The third and last bullet is where the cloud can have a meaningful impact. The usual actuarial financial reporting process for production purposes tends to be performed on a regular basis, such as a quarterly reporting cycle. However, there may also be a need for the actuaries in the reporting function to perform non-production runs in a sandbox environment or for model change management purposes. In addition, as noted above, the pricing and risk management functions may have more of an uneven need for running models. For example, due to the complexity of healthcare systems, when trying to quantify the potential impact that health policy changes may have on a health plan or the system in whole, an actuary may be required to conduct computationally intensive simulations in a short amount of time<sup>56</sup>. Taken as a whole, there are times when an on-premise high-performance computing (HPC) grid that has a fixed number of engines may be completely utilized (to the point of having queued runs that are waiting to be submitted) and other times where the HPC grid may be underutilized or even unused.

The key takeaway here is that models do not have a constant demand for processing power. An on-premise HPC grid has an associated fixed maintenance cost and, given that the demand for its use may be uneven, this may not be the most optimal solution. Access to the cloud allows for on-demand scalability of computing power, when and where it is needed the most—and no more than is needed—with a variable cost to the insurance company<sup>57</sup>.

One exciting feature that actuarial platform vendors are working with cloud providers to implement is the ability to efficiently distribute nested stochastics or deterministic-on-stochastic runs. This will be a tremendous advantage in reducing run time and would apply not only to the actuarial financial reporting function, but also to the pricing and risk management functions, which also commonly utilize such runs. However, this ability is also associated with new challenges—namely, the cost of accessing and using these computing resources, which still requires significant computing capacity that comes at a cost. Organizations employing this type of computation in the cloud must carefully optimize the trade-off between run time and cost.

### SECTION 3.2.2 REPORTING

The reporting of financial results that are based on the output from the underlying model will likely be easier to migrate to the cloud, particularly if the back-end (or post-model) processing involved in adjusting the raw model output (if applicable) has been fully automated to remove any manual steps that involve possible human intervention. In particular, migration to the cloud can mean that:

- Adjusted model output can be fed straight into visualizations and reporting frameworks using robust industry business analytic tools such as Tableau Online, Microsoft Power BI, Looker, and QlikView.

---

<sup>55</sup> Freedman, Avi & Reynolds, Craig W. (2008). "Cluster analysis: A spatial approach to actuarial modeling." Milliman. Available at: <http://www.milliman.com/insight/research/insurance/Cluster-analysis-A-spatial-approach-to-actuarial-modeling/>.

<sup>56</sup> Mills, Alan. (2013). "Simulating health behavior." Society of Actuaries. Available at: <https://www.soa.org/Files/Research/Projects/research-2013-05-mills-a-sim-health-beh-report-final-02.pdf>.

<sup>57</sup> Digital Insurer. "Actuarial models meet the cloud – a perfect marriage?" Available at: <https://www.the-digital-insurer.com/actuarial-models-meet-the-cloud/>.



- Without the cloud, there are often manual steps required in shifting model output from one location or output to a reporting database. This is particularly true if the model is stored on a local machine. Because data, code, and computation can be integrated within a cloud environment, the entire modeling process to reporting can be seamlessly integrated. Not only does this make the process more efficient, it can also reduce the potential for human error.
- Existing static report templates that are currently delivered via Microsoft Office tools such as PowerPoint, Excel, and Word can be replaced with dynamic web-based dashboards accessible by any authorized user at any time instead of waiting for an email notification that files are ready.
  - Without the cloud, the process to update templates and reports can be cumbersome and prone to human error in setting the correct assumptions or documenting the specific model run.

Automating the reporting pipeline in this way potentially allows for quicker financial close times and faster refreshes. Perhaps most importantly, the visualization element can help end users and senior management better understand and make use of the financial results, thereby allowing for more informed strategic decision-making.

### SECTION 3.2.3 GOVERNANCE

With respect to governance, there are two considerations for insurers: data governance and model governance. The cloud poses pros and cons for each.

As the cloud is a rather new development for insurers and may pose unexpected privacy concerns, insurers will need to update their data governance frameworks to include roles responsible for interfacing with cloud providers. These individuals must represent all stakeholders and all the key business functions, both technical and nontechnical—a dedicated cloud model governance committee that spans the whole company would be ideal.

There is also the concern of data in transit to, and at rest in, the cloud: using a cloud provider requires a lot of trust in their security protocols. However, the trend in the information security industry is shifting to recognize that the cloud may offer more security around data than on-premise solutions. Hector Rodriguez, Microsoft’s Worldwide Health chief information security officer, says, “Cloud-based solutions have matured to the point where they are more secure than local solutions. The reality is that these solutions, when properly integrated, should and do strengthen an enterprise’s overall cybersecurity posture by adding additional layers of security and monitoring.”

But aside from the risks associated with depending on an external partner to safeguard data, the model governance standards related to operating models in the cloud (regardless of whether it is the entire model or simply the leveraging of engines in the cloud for distributed processing) should be relatively consistent with (and complementary to) the model governance standards already adopted by a company as part of the formal corporate model governance framework.

Regardless of where the model lives (in the cloud or on-premise), there must be a well-defined change management process for updating models. While following Excel checklists and manually documenting changes are sufficient and necessary in the current state, cloud providers offer monitoring tools and built-in version control that track any changes that are made, when they were made, and who they were made by. They also have the ability to restore prior versions. Of course, there are IT tools that already exist to perform similar functions, but the cloud democratizes this oversight through user-friendly interfaces, allowing individual teams to control access to their data within the company’s overarching security position. When implemented properly, a cloud-based model governance structure can reduce potential model error and significantly improve the model process audit trail, facilitating easy replication, validation, and external audit.

### SECTION 3.3 INSURANCE INNOVATIONS USING THE CLOUD

The agility and capacity offered by the cloud has enabled new forms of insurance to be spearheaded by Insurtech companies, many of which are partnering up with larger established insurers. Trov delivers micro-duration contracts through a mobile app to cover individual personal items like laptops and watches and can be toggled on and off directly within the cloud-powered app. Trov's credibility has been affirmed via partnerships with Munich Re, AXA, and Suncorp in 2016<sup>58</sup>.

Ant Financial, owned by China's Alibaba, is taking healthcare back to a centuries-old model of collective protection, leveraging cloud-stored customer information generated from Alipay's online payment platform to effectively market to Alibaba users. This collective health insurance product, called Xiang Hu Bao (which literally means mutual protection), has already attracted 50 million users as of April 10, 2019 since its inception in October 2018 and aims to cover 20% of China's population within two years. It is all based on cloud technology, attesting to the power of big data<sup>59</sup>. There are no premiums: all users share the cost of claims payouts on around 100 illnesses, up to \$44,700.

Travelers partnered with Amazon in late 2018 to offer smart-home kits at discounts on Amazon's online platform. The products include water sensors, security cameras, motion detectors, and an Amazon Echo<sup>60</sup>. Travelers has even developed new Amazon Alexa skills to help policyholders with questions regarding bills and payments and grant them access to home safety and property maintenance tips. These devices all interact with one another, generating data pushed to the cloud for analysis and monitoring.

In 2015, Ageas became the first insurer to adopt end-to-end enterprise insurance software based in the cloud to provide benefits, well summarized by Hans Van Wuijckhuijse, a regional director of business development in Asia:

"With this comprehensive set of next-generation enterprise software from SAP, Ageas gets to focus on its customers and the core of its business, instead of being in the business of also running a midscale technology services firm internally. Previously we would have struggled to reconcile numerous independent systems from multiple providers involving multiple systems integrators. This integrated set of SAP solutions, coupled with SAP taking end-to-end responsibility for this software-as-a-service offering, will give Ageas the much-needed agility to help us establish and penetrate into new markets"<sup>61</sup>.

Everything from underwriting, quotes, policy administration, claims management, and financial reporting is handled on this platform. Ageas has also partnered with Tractable to deploy artificial intelligence. Tractable's AI Approval connects directly to an insurer's back-end systems and will analyze photos of auto accidents to generate repair estimates and flag any suspicious activity, reducing fraud. Deploying AI has allowed Ageas to reduce the time it takes for customers to process claims and receive their cars back. In an era where consumers expect quick service with minimal delays and intuitive user interfaces, the value of having a streamlined data engine with modern applications and interfaces cannot be overstated.

Even more traditional insurance companies have been able to leverage the cloud to manage risk. Many companies, for example, especially reinsurance companies, are investing in credit risk transfer deals, but participating in these transactions requires models based on billions of transactions. By using the cloud, an analytics company has been

<sup>58</sup> Wikipedia. "Trov." Available at: <https://en.wikipedia.org/wiki/Trov>.

<sup>59</sup> Jingli, Song. (2019). "Alipay's mutual aid platform Xiang Hu Bao aims to cover 1/5 of China's population in two years." KrAsia. Available at: <https://kr-asia.com/alipays-patient-payout-platform-xiang-hu-bao-aims-to-cover-1-5-of-chinas-population-in-two-years>.

<sup>60</sup> Moorcraft, Bethan. (2018). "Travelers announces major Amazon partnership." Insurance Business. Available at: <https://www.insurancebusinessmag.com/us/news/breaking-news/travelers-announces-major-amazon-partnership-113456.aspx>.

<sup>61</sup> SAP News. (2015). "Global insurance group Ageas adopts industry's first end-to-end enterprise insurance software in the cloud thanks to SAP." Available at: <https://news.sap.com/2015/10/global-insurance-group-ageas-adopts-industrys-first-end-end-enterprise-insurance-software-cloud-thanks-sap/>.

able to build models on this data and share access with those participating reinsurers. These companies can access the data and manipulate model conditions as end users.

### SECTION 3.4 TO THE CLOUD

As mentioned previously, there are four main cloud deployment models in the market: public, private, hybrid, and community. If we wish to be even more reductive, these four categories can be distilled into two: public and private. Both hybrid and community clouds leverage aspects of both, so they really represent variations on those two, and do not necessarily represent unique structures in and of themselves. With these deployment models, insurers can often choose different methods of leveraging public clouds versus retained traditional storage and processing, and versus community or private clouds. Noncore processes can be pushed to the public clouds, while more critical or sensitive processes may be retained in a private cloud, or other internal systems for the extra security.

Noncore processes are far more likely to use SaaS. For those processes where security is a priority, however, IaaS will enhance core processes by allowing on-demand provisioning of extra computing power.

Wasabi is a startup seeking to disrupt the cloud computing marketplace. Founded in 2015 by David Friend and Jeff Flowers, their goal is to provide flexible, low-cost storage without the traditional egress charges associated with large cloud service providers. It is not our goal to endorse or suggest any provider, but it is clear that the industry is not a settled one. As time goes on, and more companies use the cloud, the variety of solutions will continue to expand, allowing more bespoke structures that will cater to the specific needs of a client or industry. So, while Amazon, Microsoft, Google, and IBM are the largest current public cloud supporters, their high margins are inducing many new companies to enter the space.

In terms of security, there is a debate as to whether public clouds are sufficiently secure for core or critical business processes. While cloud data is almost always encrypted, these encryptions can be broken if someone has the keys. This is true regardless of which type of cloud is being used, but private clouds are managed by your IT staff, and only your IT staff, which arguably reduces the risk. That being said, public cloud vendors are in the business of cloud computing, so while there is broader access to the cloud, the security on average may be substantially more advanced than your own.

Independent of the security of the storage, another major concern is how secure the data transferring process is with a public cloud. How private the transmission of data from a local network to the cloud is going to be depends on several factors, not all of which are under the control of the cloud vendor, but it will also depend strongly on your internal security processes for handling the data that will be transmitted.

While these concerns may militate against using a public cloud for core processes in production, that does not mean there may not be value in them in general. An attractive solution is to build proofs of concept on the public cloud with test data. This is a low-cost option where you only need to pay for what you use. As the proof of concept transitions to production, it can be migrated over to a private cloud with real data. This utilizes the flexibility of the public clouds (and the attendant ease of quicker implementation) while benefiting from the security and control of a private cloud.

Actuaries, when thinking about the type of cloud structure to adopt, need to be careful about the exact purpose and needs the cloud is meant to satisfy. Key considerations are:<sup>62,63</sup> budget; security and compliance requirements; hardware and virtual server control; failover control; service level agreements (SLAs); cloud resource utilization and consistency; what data will be used in the cloud environment; internal IT resources to support the services; how many teams or groups will be utilizing the cloud and how similar the processes are; and how much automation can be achieved if a private cloud is utilized.

Innovation is always more successful if it is thoughtfully implemented rather than blindly incorporated. Cloud computing often has the opportunity to greatly improve many processes of a company, but the cost savings can turn into a nightmare if the process is not thought out thoroughly in advance. The trade-offs between different structures and vendors should be carefully considered, and sufficient internal support and training should be provided for whatever solution is selected.

---

<sup>62</sup> Myerson, Judith. (2014). "Public or private cloud? 5 criteria to help you choose." TechRepublic. Available at: <https://www.techrepublic.com/article/public-or-private-cloud-5-criteria-to-help-you-choose/>.

<sup>63</sup> Jasso, David. (2017). "Private vs public cloud costs: Surprising 451 research results." VMware Blogs. Available at: <https://blogs.vmware.com/management/2017/07/private-vs-public-cloud-costs-surprising-451-research-results.html>.

## Section 4: Machine Learning / AI and Cloud Data Impact on Actuarial Science

In addition to the above impacts of the cloud on actuarial processes, the cloud also allows for users to “scale up” computational resources on demand by using more servers. This allows users to leverage more computationally-intensive algorithms that would not be possible for a local machine or even a local server without consuming significant resources. These types of algorithms are commonly called “machine learning” (ML) and more recently “artificial intelligence” (AI).

Artificial intelligence can be defined as any attempt to make machines learn from experience, and to perform human tasks. Currently, AI is becoming increasingly important thanks to big data, ever-improving algorithms, and the greater capacities of storage and computing.

Machine learning is a subfield of artificial intelligence that allows machines, programs, or algorithms to learn and improve from data. Machine learning is used in a variety of different fields, from finance to governance and transportation. It is based on statistical theory and uses different features to make conclusions. This section of the research discusses machine learning and AI within the field of actuarial science. It is included in this report as advances in cloud computing have made it more feasible for insurance companies to use and leverage these techniques.

### SECTION 4.1 INCREASED PROCESSING CAPACITY THROUGH DISTRIBUTED SYSTEMS

Although academic papers on AI have been in circulation for more than 50 years, their uses have been limited in commercial application until recently due to advancements in cloud computing and its relatively low cost. Prior to advancements in cloud computing, computer science development had been ruled by Moore’s law. Therefore, although there had been research performed on AI, the application of complex algorithms in the real world would take time to implement given computing limitations. Cloud computing, through the use of distributed computing, has vastly increased computing capacity and has made implementing these algorithms feasible.

A distributed system allows speeding up computations and improving efficiency and performance using different networked computers. Distributed computing consists of multiple instances of software on multiple computers but run as a single system. As many computers work on the same “project,” users can add new computers to the system or replace defective computers without interrupting the work flow. If a computer stops running, the work can continue.

Thanks to the cloud and associated AI algorithms, actuaries can now explore new methodologies in claims management (e.g. by using picture analysis or machine learning algorithms to identify high-risk exposures) and pricing (by using machine learning algorithms to consider many complex relationships within data), or even propose new services based on the Internet of Things (e.g. connected cars, e-consultation, health monitoring, etc.).

Exploring these new tools requires that actuaries have an understanding of cloud services and the associated algorithms, including the risks of using these algorithms.

#### SECTION 4.1.1 CHOOSE THE RIGHT TECHNOLOGY AND CLOUD SERVICE

In order to meet the computational requirements an actuary requires in a cloud environment, insurers have to work on specific IT environments that can be easily deployed onto the cloud. This section provides background and discusses some of the IT considerations for a cloud environment that is required for actuarial analysis.

### *Multiprocessing on CPU/GPU*

The use of multiprocessing is the most common approach to increasing the computation of algorithms. Multiprocessing refers to a coordinated processing of programs by more than one computer processor. Multiprocessing is the technique of performing multiple tasks at the same time.

Multiprocessing can be performed both via a central processing unit (CPU) or a graphical processing unit (GPU). The latter is commonly used for training deep-learning models. Deep learning is a subfield of machine learning and is currently popular for extracting information from unstructured data-like images, sounds or text. A GPU is essential for deep learning because those algorithms require a computationally intensive amount of matrix multiplication that can be executed on a GPU. More generally, a CPU is the brain of any computing device. There are different kinds of CPUs. Single-core CPUs can perform one operation at a time. Multicore CPUs are faster and can perform multiple operations at a time. The “multi” in multiprocessing refers to the multiple cores in a computer’s CPU. Using a multicore CPU can dramatically improve processing time. On top of CPUs, there are also GPUs, which are smaller in size but have many cores. Basically, GPUs are built to perform graphical calculations, which are often used in gaming applications. Recently, however, GPUs have been utilized to perform non-graphical calculations.

For multiprocessing, users often choose between CPUs or GPUs for the hardware, and the ultimate choice depends on the task one wants to accomplish. CPUs are generally thought to be better at single-complex calculations while GPUs are generally thought to be better at handling multiple and simpler calculations at a time. The choice is a trade-off between speed, reliability, and cost.

### *Parallel computing*

Another technology used to improve computing calculation time is parallel computing. Without parallel computing a single machine sends commands to an individual computer, and the commands are executed sequentially on a single processor, which means calculations are performed in order, one at a time. The goal of parallel computing is to simultaneously use multiple-system resources to solve a computational problem. Instructions are executed simultaneously on different processors, and the system acts as a single computer on which the instructions of a program are executed simultaneously, using additional cores of the processor or by connecting several computers via a network or in the cloud. In the second case, networks connect multiple stand-alone computers (nodes) to make a “cluster.” Parallel computing is often used for statistical modeling, simulation, and understanding complex, real-world phenomena.

### *Distributed computing on cluster*

When distributed computing on one device is not enough to meet the computational demands required, organizations can connect several devices together to increase computing capacity. Distributed computing is a network of computers working together as one system. Either the machines can be located in close proximity or they can be dispersed geographically, but most of the time the machines are close and connected through a network. It can be used for connecting hardware (for example, a printer) in a company; each time you make a new Wi-Fi connection, you are entering a new computer network. Another example would be a company that runs a web application. Every day, the company attracts more and more users. Thus, the company needs more and more computing resources to continue running the site. Distributed computing is a solution to this common issue: the company adds more computers to the system to accommodate the increased workload. On top of that, each processor or computer can be dedicated to a single repetitive task.

Distributed computing has strengths and weaknesses. Among the strengths, distributed computing has a fault tolerance, which means that if a machine on the network goes down, the entire system does not crash and can

continue what it was doing. It also offers better efficiency compared to centralized systems. This approach is very scalable because it is easy to add a computer when more power is needed. The counterargument is that this system is relatively complex compared to a centralized system. It also requires more security because each device on the network needs to be protected and secure. Finally, this system is more expensive than a centralized system, as it generally requires more hardware.

Cloud providers offer services for both parallel computing and distributed computing.

## SECTION 4.2 CURRENT LEVEL OF USE OF MACHINE LEARNING AND AI BY ACTUARIES

This section provides a brief introduction into current uses of machine learning and AI by actuaries, which is made possible by the type of expanded computing capacity described above. In this section, we discuss considerations for using machine learning, provide a brief introduction into common algorithms, and provide examples of how insurance companies are using machine learning and leveraging the cloud.

In late 2018, the SOA commissioned a survey to understand the use of machine learning in actuarial science. A total of 143 individuals (predominately from U.S. insurance companies) responded to the 46 questions of the survey<sup>64</sup>. According to this survey, the use of machine learning in actuarial science is relatively new. Most of the actuaries who use machine learning have been using predictive algorithms in their work for less than five years.

The most common uses of machine learning in actuarial science included: pricing, claims, in-force management, risk, underwriting, valuation, and disease management. Most of the people who answered the survey think that machine learning in actuarial science improves underwriting (e.g., by being able to detect fraud), evaluates lapse, surrender, and mortality rates, or improves pricing. According to the survey, predictive models are also used to better analyze businesses and customers, gain competitive insights, identify potential risks, and improve the efficiency of different models. Actuaries who work on data sets that exceed 200 GB are rare, which means that most of the time actuaries do not need distributed computing. Regarding large data sets, most of the respondents use statistical sampling and a few use new techniques such as Spark or Hadoop (that is, systems that allow users to work with large amounts of data).

According to the results of this survey, traditional models such as generalized linear models (GLM) and generalized additive models (GAM) are more common than machine-learning models. Concerning machine-learning models, the ones that are the most used are classification models, especially classification and regression trees (CART) and random forests. They attach great importance to the hypothesis of normality, independence, and model form when developing models, but they also allow for the hypothesis of normality, independence, and constant variability to be violated for the sake of practicality. To account for missing data, the survey respondents prefer to use imputation or remove incomplete observations; for outliers, most survey respondents opt for capping or removing data. Only 50% of actuaries use cross-validation to check the performance of their models.

When it comes to feature selection, they use several methods, such as backward, forward, or stepwise selection, Lasso, Ridge, or algorithms that determine the importance of features (such as random forest). The selection criteria are principally mean square error (MSE), R-squared, adjusted R-squared, log likelihood, area under curve (AUC), and Bayesian information criterion (BIC) or Akaike information criterion (AIC).

---

<sup>64</sup> SOA (2019). "Guidelines for Predictive Modeling in Insurance Applications." Available at: <https://www.soa.org/resources/research-reports/2019/considerations-predictions-insurance-applications/>.



From this poll, we can conclude that there are different kinds of actuaries who use machine learning. It is difficult to extract a potential trend because of the diversity of practice.

#### SECTION 4.2.1 TRENDS AND EXPECTED USE OVER NEXT FIVE YEARS

According to a 2019 Gartner survey of chief information officers (CIOs), "AI adoption in organizations has tripled in the past year"<sup>65</sup>. As data becomes increasingly available and employees become more efficient at processing and analyzing data, it is likely this trend will continue.

According to the Gartner survey, AI is going to become more important in the process of decision-making. Respondents predicted AI will be one of the top tools to "drive infrastructure decisions" in the near future, but to do so, companies must follow this trend and adapt to this important evolution. As machine learning and deep learning are complex, it is important that IT and business teams work in collaboration in order to support the transition. If companies want this transition to be efficient, it must be done step by step, which means they should not rush into it. Companies that are new to AI may want to skip steps; for example, by directly applying deep learning without having performed the initial step of exploratory data analysis in advance of applying these complex algorithms. It is crucial to understand the basics before applying such complex techniques to solve problems. Gartner predicts that organizations will be using complex deep-learning models for tasks where simpler techniques would be sufficient, stating, "classical machine learning techniques are extremely underrated"<sup>65</sup>. Thus, even though deep learning and neural networks seem to be the future of data science, it is important to understand that applying simple models first is a crucial step in the transition to AI.

More and more, AI and similar technological tools will become commonplace in the near future. They are going to be part of our daily lives as part of the process of continuous innovation. Autonomous technologies are going to fill an increasingly important role in our lives and these technologies, such as robots or self-driving cars, are all based on AI. These tools are commonly known as the Internet of Things (IoT) and, not only do they rely on data to work, but they also generate data that can be used to reinforce the algorithms and provide information on trends and use, offering insights into consumers.

For actuaries, this type of data can be valuable in providing customized rates and products to consumers. Examples of these types of data being incorporated into insurance products are miles-based car insurance, adjustments to life and/or health insurance premiums as a function of activity levels, and retirement tools that incorporate current and projected spending patterns in order to provide guidance on how much consumers should save today to meet their retirement goals. All of these products were developed by actuaries and data scientists who incorporate outside data and information using advanced analysis techniques for product innovation.

### SECTION 4.3 DATA CONSIDERATIONS FOR USING MACHINE LEARNING / AI

#### SECTION 4.3.1 LEVEL OF GRANULARITY, SCALABILITY, AND AUTOMATION

AI, machine learning, and big data are topics that may intimidate a large part of the population, with consumers asking themselves whether they are permanently monitored or under surveillance. Clearly, there are plenty of data that are collected every day, especially on the internet, but data must be analyzed before they can be turned into actionable insights. Used alone and individually, these data are not really useful. They make sense when they are

---

<sup>65</sup> Costello, Katie. (2019). "I&O leaders need to strategically leverage AI as a core accelerant to digital business initiatives." Gartner. Available at: <https://www.gartner.com/smarterwithgartner/gartner-predicts-the-future-of-ai-technologies/>.

aggregated, analyzed, and integrated into processes that can provide individualized recommendations, pricing, or optimization.

Each context for the use of data is different and, according to the context, data are used in various ways. For an actuary, an important item to know when collecting the required data for a project is the type of data available and how to store data records. For example, are we working on spatial, temporal, or social data? If we are working on temporal data, are the data collected every second, minute, hour, or day? If we are working on spatial data, are they collected by street, city, or country? Are the data individually collected or do they represent aggregated statistics (mean, median, etc.)? These questions refer to the granularity of the data.

For actuaries in the process of migrating to cloud storage, there are opportunities going forward to expand the data collected and stored. It is important to be involved in discussions within the company as data capabilities are built out, so actuaries can ensure they have the appropriate and correct types of data collected for consideration in future applications. For example, for life and health insurance companies, it is important to build the data architecture today for potential pricing schemes in the future. It may be prohibitive within an existing data architecture to collect, store, and analyze data from wellness applications, which can provide real-time information on individual activity levels. Therefore, adding data storage from the cloud could be beneficial to allow the actuary to build up the data necessary to potentially analyze and incorporate such information into their future products. Once the data are collected, the actuary will also need ways to analyze the data. If the data are too large to store on internal hardware, it is likely the data are too large to analyze on internal hardware.

#### SECTION 4.3.2 USE AND AVAILABILITY OF DATA / BIG DATA / DATA ARCHITECTURE

The cloud has enabled the collection of big data and has modified both the use of data and the way users access, share, and store data. Companies use data to produce insights, and these insights help determine a company's strategy.

Given the rapid emergence of new technologies such as IoT, data are increasingly collected as so-called "unstructured data" from a variety of sources, including emails, images, sounds, PDF files, and others. Unstructured data can be obtained from sources such as social media platforms, mobile applications, location services, and IoT. Since the data are diversified without a common organization structure, two common issues are encountered with unstructured data: the data require resources to transform the data into usable formats, and the datasets are larger and require more storage, which makes them more expensive (in terms of investment). Most companies choose to collect unstructured data and store them in a central location known as a "data lake."

Gathering unstructured data and designing a data lake can be expensive. In response to this problem, some systems allow users to efficiently store and access data. For example, Hadoop, Spark, or Elastic Search are distributed systems that are available on the market to manage data volume (i.e. the quantity of data collected), variety (e.g. text data, time data, location data, event data, etc.) and velocity (i.e. the speed at which new data are created, such as every second, minute, hour, day). Most of the technologies are provided through cloud platforms, and they can be easily activated by knowledgeable programmers. These systems are open-source technologies specifically created to handle large amounts of data and provide an efficient and low-cost approach to establishing a coherent and organized architecture of data.

All companies that want to analyze unstructured data ultimately have to convert the data into an organized format, known as "structured data." This kind of data is then stored in a databases and can include pre-defined fields such as customer information (name, address, and phone number), transaction details (financial data, orders), product information, etc. One of the core value-adds of structured data is that it is easily understandable, searchable, and analyzable, not only by humans but also by machines.

When considering cloud storage and what data to add into a data lake, actuaries should be mindful of how that data will ultimately be used, the process to transform the data from unstructured to structured, and how the structured data will be analyzed in the future. Considerations for these items include processing time, frequency and velocity of data, and internal resources available to design and maintain the architecture.

#### SECTION 4.4 CASE STUDIES OF MACHINE LEARNING FOR ACTUARIES

According to the Be An Actuary website<sup>66</sup>, actuaries are experts in:

- Evaluating the likelihood of future events—using numbers, not crystal balls
- Designing creative ways to reduce the likelihood of undesirable events
- Decreasing the impact of undesirable events that do occur

For the majority of pricing, reserving, and modeling work, actuaries traditionally relied upon development triangles and their derivatives (common in property and casualty [P&C] exposures), mortality tables, morbidity tables, and other aggregate statistics to estimate risks associated with future exposure. As data became increasingly available, industry data and statistics were merged with internal experience using statistics such as credibility theory to develop experience analysis and blend internal company experience with industry data. This approach is also used in lines of business with low frequency or volatile results. Careful attention is spent to ensure internal data is not overly relied upon if the data is thin or not robust.

In today's environment, actuaries are increasingly using machine-learning techniques to maximize the value of internal and external data. The concepts are similar to traditional actuarial work, but they are being applied at more granular levels with more data. The same notion of not over-fitting the historical experience, but still capturing enough variation such that the resulting assumption generalizes well to the future experience, is also an important subject in the study of machine learning. However, what is known in machine learning as the bias-variance trade-off has many similarities to traditional actuarial credibility theory. Assuming knowledge of these similarities and understanding the differences, an actuary can get comfortable with using machine learning to help automate the assumption-setting process. This can lead to uncovering new relationships in data that may not yet have been detected with traditional methods.

Actuaries in a variety of industries are adopting the use of machine learning and have been using it to enhance traditional actuarial analyses. Below are some case studies of machine learning in practice.

##### SECTION 4.4.1 CASE STUDY

###### *Underwriting and pricing*

Machine learning is a common tool for developing predictive models. Building predictive models is a regular activity for an actuary, and it is particularly interesting to use machine learning to develop advanced rating systems. Machine learning can be a way to incorporate complicated interactions within the data efficiently (being careful not to over-fit the data, of course!). Thus, pricing optimization is a natural application of machine learning within insurance markets.

One common-use case for this type of analysis is leveraging telematics data for car insurance products. Telematics is a method of monitoring a vehicle by using its GPS and onboard information. In insurance, it can be used in pricing

---

<sup>66</sup> Be An Actuary. Available at: <http://www.beanactuary.org/>.

to segment the cost of insurance between lower-risk and higher-risk drivers, based on the behavior of an insured. The data is often used as a way to reward low-risk drivers.

Telematics data provides information on vehicle technologies and driver behavior. The data are often large (imagine a database that captures your every turn, speed, and acceleration when you drive!). Therefore, cloud storage provides a great solution to being able to store and analyze the data. Machine-learning algorithms provide an efficient methodology for car insurance companies looking to analyze the many complex interactions within the data. The data provided by such a tool allow the insurer to better evaluate risks associated with driving, and thus to better value prices. Telematics provides a large number of new variables (distance traveled, accelerations, turns) that an insurance company can use in different applications, such as scoring, pricing, or underwriting. Using standard actuarial methods, the models required to price such data would be cumbersome and difficult to use to develop granular prices. The data could quickly become thin and lose credibility. Even with GLM, it would be difficult to properly capture and control for the various relationships in the data. However, with machine learning, these roadblocks can be solved and reliable pricing algorithms can be developed.

#### *Claims management*

Due to the rapidly increasing cost of healthcare in the United States, machine learning is being applied in a variety of ways throughout the healthcare industry, with the goal of reducing the severity of healthcare claims. For example, care management teams are using population stratification tools that rely on multiple data sources and machine learning to identify individuals who are likely to incur large claims in the near future that may be avoided or mitigated. Once identified, the care management team works with these individuals to lower their risks of incurring these claims. Machine-learning techniques are also being used beyond just identifying these high-risk individuals to recommending care paths based on previous clinicians' recommendations of similar cases. All of this has great potential for reducing the overall increasing healthcare expenditures.

These examples are again possible only through the collection of granular data on a large number of policyholders, and the ability to analyze the data. An efficient way to store and analyze this data is through the cloud.

#### *Reserving*

Loss reserving for general insurance is traditionally based on aggregate triangle-based models, which use a single runoff triangle as a representation of the activity of several cohorts of claims over their respective lifetimes. These classical methods may generate material cumulative errors in the reserve estimates when portfolio characteristics of claims cohorts inherent in the triangle data change over time in unanticipated ways.

The combination of individual claim models (ICMs) with machine-learning techniques is an emerging area of research and practice that uses individual claim-level data to estimate loss reserves. Evolution in technology with respect to efficient data collection, storage, and analysis has made ICMs more accessible. To date, there is convergence neither with respect to an ICM analysis framework nor to the universe of model parameter assessment and validation techniques. Further, the amount of expert judgement required in an ICM analysis can be substantial. Nevertheless, the use of valuable information embedded in individual claim data is a promising feature of the approach that should lead to more reliable loss-reserve estimates. The use of machine learning and AI techniques in this field constitutes a promising area.

#### *Mortality forecasting*

For life insurance companies, an important activity is to accurately forecast mortality based on their experience and trends in the underlying insured population. There are common models used by actuaries to estimate future mortality rates, such as the Lee-Carter model. There have been some extensions of this model, but the extended

models are difficult to calibrate. Recently, there has been discussion around the use of neural networks to extend this model to select an optimal model structure. The Lee-Carter model is applied to a single population, and the goal of this extension is to use multiple population forecasting, which would be more robust. The goal is to capture some features or trends among different groups, in order to apply similar factors to different groups. On top of that, when using single-population forecasting, there could be incoherence between different countries, while using a multiple-population model could reinforce the coherence. This neural networks approach, based on multiple-population forecasting, requires less manual engineering than usual and it can be fit to many populations simultaneously.

#### *Nontraditional roles (e.g., marketing or optimization)*

Machine learning can be particularly interesting and useful for marketing, especially in e-commerce. For this application, user data is required to target specific individuals for marketing purposes. This data can be structured or unstructured with both a high velocity of new data and size of data that can be collected. The cloud provides an efficient way to store and analyze this data.

Machine learning allows companies to adapt their business strategies to current situations. Consider the example of e-commerce to illustrate the role of machine learning. Machine learning can be used to recommend items through a recommendation system: for example, after purchasing item X, the recommendation system would suggest purchasing item Y, as the two were commonly purchased together in the past. It can also be used to detect fraud or optimize a voucher campaign, for example. Some machine-learning algorithms can determine different clients' profiles in order to better understand demand and, thus, better adapt to the market.

With respect to marketing, machine learning can also be used in recommendation systems through cross-selling or association rules, in particular to optimize customer selection. The goal of cross-selling is to propose to clients a basket that fits their expectations. In finance, it can be a basket of financial products that optimize the client's financial strategy. Care must, of course, be taken to ensure that the variety of financial products being offered are appropriate for the customer (so as to avoid suitability concerns).

## **SECTION 4.5 COMMON MACHINE LEARNING ALGORITHMS AND TOOLS**

This section discusses common machine-learning algorithms and tools used both in the insurance industry and data science.

### SECTION 4.5.1 MACHINE-LEARNING ALGORITHMS (CART, RANDOM FORESTS, K-MEANS, ETC.)

Thanks to all the tools being provided through the internet, it is relatively easy to estimate a machine learning model. However, understanding how the algorithms work is the first and most important step in data science. Plenty of algorithms are available and many can be adapted to a single problem.

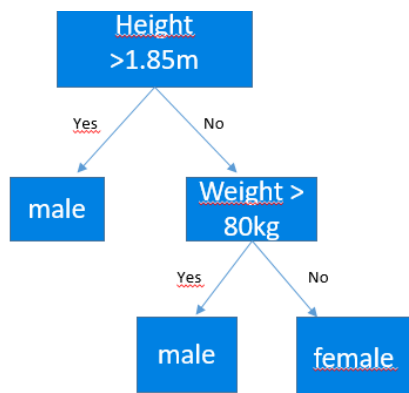
In machine learning, there are two kinds of tasks: supervised learning and unsupervised learning. When we talk about supervised learning, it means that given a set of inputs, we want to predict an output. To do so, the input and output we know are used to train the database in order to approximate a mapping function. Once the mapping function is approximated, it is used to predict new outputs from new inputs. When the mapping function is trained, we already know the answer (the output), but we want to determine the model that best fits the data. Supervised learning can be divided into two types of algorithms: regression and classification. In regression, the output variable is a continuous value, while in classification, the output variable represents a label: it can be an integer or a category.

In unsupervised learning, only input variables are available. The goal of unsupervised learning is to use the characteristics of the data to draw inferences that are not explicit. The most common form of unsupervised learning is cluster analysis. Cluster analysis uses characteristics of individuals in order to identify some potential cluster of

similar individuals. For example, given a set of 100 individuals and their associated features, can we create an algorithm that segments the population into four “like” groups? The examples below provide a more thorough description of supervised and unsupervised learning, respectively.

One of the most famous supervised learning algorithms is the classification and regression tree (CART) algorithm. Decision-tree algorithms can be used both for regression and for classification. A CART represents a binary tree, where each node is summarized by a yes/no question, such as “is this man taller than 1.85 meters?” or “is the individual blue-eyed?” Each node represents a variable and a split value of this variable (the variable can be numeric or categorical). The leaf of the tree represents the output variable and is used to make a prediction. Making a prediction through a decision tree is easy. Given a set of inputs, the data flows through the tree according to the value of the input, up to arriving at a leaf, which represents the predicted value of the output.

Figure 3: Decision Tree



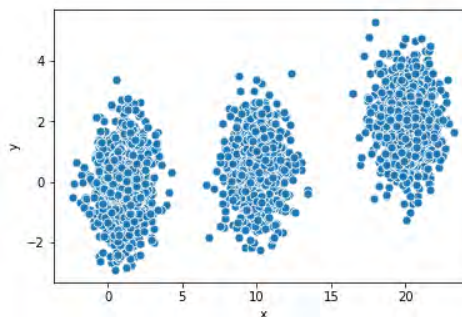
Suppose one wants to predict someone’s sex, given that person's height and weight. A decision tree could look like the graph in Figure 3.

Another algorithm from supervised learning is the random forest, which also works for both regression and classification. It is called a forest because it is composed of an ensemble of decision trees. Each tree is built “randomly” using a subset of the total potential variables in the data; the resulting trees are aggregated together to create a final prediction. As it is composed of several trees, the input data goes through all of the trees. In the case of a classification problem, the output value of the algorithm is the most predicted class from all the trees combined. In a regression problem, the output of the random forest is the mean of all predicted values by the different trees. An advantage of the random forest over a decision tree is that it limits over-fitting.

CART and random forests are commonly used in pricing analysis, creating reserving algorithms, and evaluating risks with complex interactions.

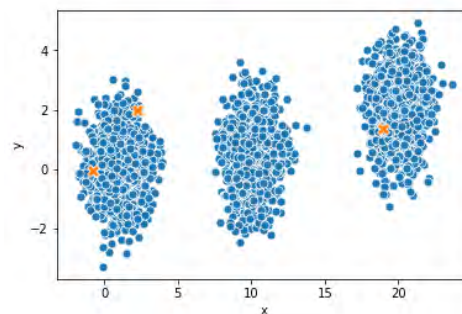
A common unsupervised learning algorithm is the k-means algorithm. The goal of this algorithm is to group individual data into clusters according to their characteristics. Thus, similar individuals will be in the same cluster. One of the important points of k-means is to determine the best number of clusters (as it is unsupervised learning, we do not have a quantity of classes or labels on which to base the quantity of clusters). To illustrate the k-means algorithm, we have generated three groups of points through three random normal laws. In the graph in Figure 4, we can distinguish three groups. Intuitively, we are going to choose a k-means with three clusters.

Figure 4: Visualization of Data



To do that, we have to initialize the k-means, selecting three random points of data – known as centroids (as shown in Figure 5). Using this algorithm, we have decided to initialize the k-means with existing random points in the data, but there are other k-means algorithms that initialize with nonexistent points. The k-means algorithm works iteratively and on each iteration there are two steps: assigning its cluster to each point and moving centroids. In the first step of the iteration, the algorithm associates each point with the cluster it is closest to. In the second step, the algorithm moves the centroids, computing the average of all the points in each cluster. These two steps are repeated until a stopping criteria is met (e.g., the clusters remain the same, the gain of inertia reached a threshold, or a maximum number of iterations was reached).

Figure 5: Initialization of Centroids



After the first iteration, the k-means looks like the graph in Figure 6. And after two iterations, it looks like the graph in Figure 7. At the third iteration, the graph is the same as in Figure 7, which means the k-means procedure is done.

Figure 6: First Step of the K-Means

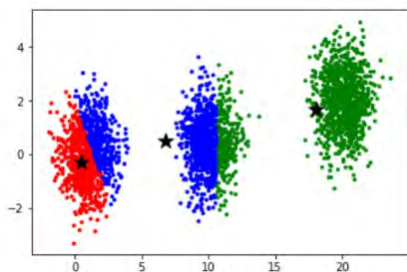
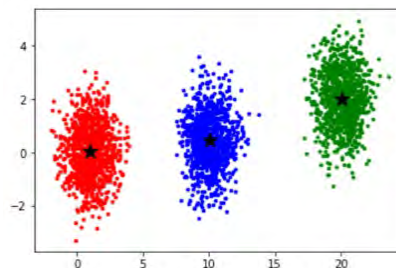


Figure 7: Second Step of the K-Means



K-means clustering is commonly used in marketing campaigns to identify similar exposures for claims management and process optimization.

#### SECTION 4.5.2 COMMON PACKAGES

R and Python are common programming languages used to perform machine-learning analysis. Different packages in R and Python allow a community to do machine learning quite easily. A package is a set of modules (a module is a file created to perform a task) that keep all linked modules in the same location. Packages allow a hierarchical structure of the module name space. Thus, a package could be compared to a directory, with modules as files composing the directory.

In Python and R, packages are open source, which means that the code is freely accessible and anyone can modify it, or add new modules, in order to benefit a community of users. Most R and Python packages have their own descriptions, in order to be understandable and to describe the parameters to use in the package. For example, this is the case for scikit-learn, which is the principal machine-learning package in Python. It allows running many different types of algorithms, such as random forest or regression trees.

Most of the code for the main packages are online and can be accessed through GitHub. GitHub allows these packages to be open source and accessible to all.



### SECTION 4.5.3 CONSIDERATIONS FOR USING OPEN-SOURCE SOFTWARE

As its name suggests, open-source software is open source, which means that any IT professional or any member of the community can access the code in order to alter, improve, and distribute it. Open-source software is a perfect example of a successful communal collaboration. It can be used for any type of software, such as word processing or cloud computing.

Open-source software provides many advantages. It has strong durability because it is maintained by a community of users who discuss their problems and find solutions together. Since the software is free, it allows many companies to avoid both license and maintenance fees linked to software. The open-source characteristics also make it easy to learn because the community of users is large and help is always available. Maintenance of open-source software is generally much faster than maintenance of proprietary software. When there is an issue, users do not have to wait for the IT team of a software company to solve the bug; the community of users is usually quick to resolve identified issues.

However, even though it provides many advantages, open-source software also has its disadvantages. There is no legal warranty in case of bugs because there is no contract with a supplier. On top of that, in case of a technical issue, there is no certified technician that could come and help in order to solve the problem. Some claim that open-source software is not secure relative to proprietary software.

### SECTION 4.5.4 MAINTENANCE

In general, all computer code needs maintenance, because code is modified over time for enhancements. This means that a set of code, or a package, never has a “final” version; there are always new “last” versions. Even though the last version is better than the previous one, it is important to keep track of all modifications on a project, particularly for audit purposes and replicability. This is known as versioning. Versioning means handling new versions of the code progressively with modifications applied on the prior code.

There are different version control systems like Git that allow users to go back not only to a previous version of a file, but also to an entire project, to compare changes over time or to see who last modified a file. If a file is lost, it can be recovered easily. Using versioning software is a common developer’s practice, both for keeping track of the evolution of code, but also to share one’s project with the community. As most packages are open source, it is important to keep track of the different changes in the code in order to quickly trace potential bugs. Indeed, when modifying code, some bugs can arise and become difficult to track down. With a version control system, each time a file is added or deleted, a .txt file is modified and saved in a commit. A commit is a version of the code at a given time. Aggregating the historical commit files provides an auditable record of the code development process.

## SECTION 4.6 MODEL SELECTION AND INTERPRETATION OF RESULTS

This section discusses considerations for model selection and interpreting the results of machine-learning algorithms.

### SECTION 4.6.1 MODEL SELECTION

With machine learning, in particular supervised learning, an important step is to divide the set of data into a training set and a test set. The training set of data is used for training the algorithm. This means that the algorithm is built based on the training set and it learns how to adjust to best fit the training set. Then, to validate the efficiency of the algorithm, the algorithm is run on the test set. The algorithm may be perfectly adapted on the training set, but not on the test set. That is what we call over-fitting. In order to ensure the coherence of the algorithm, there are resampling methods, such as cross-validation. The goal of cross-validation is to train and test the algorithm multiple times in order to be sure to avoid over-fitting. When testing the algorithm, it is also important to choose a good

evaluation criterion, and to compare it to a threshold, in order to check the relevance of the algorithm. The training/test validation is one of the most important steps when using machine-learning algorithms.

The method of resampling can also be used when applying more traditional methods, such as linear models (GLM, regression). However, in general, the metric to evaluate an algorithm is different from traditional methods. In general, there is a tendency to compare R-squared, likelihood function, or AIC and BIC to check the results of a model.

#### SECTION 4.6.2 INTERPRETATION AND FAIRNESS OF ALGORITHMS

The benefit of traditional methods over a machine-learning algorithm is that traditional methods are easier to interpret. They generally provide a coefficient where a causal relationship can be quantified. Traditional models, such as linear regression, logistic regression, or GLM, are not complex in theory and are thus easier to understand. On the other hand, machine learning and deep learning models are intrinsically more complex because they require multiple calculations and iterations. Understanding a model is very important in machine learning, not only to interpret results, but also to detect potential mistakes. That is why the transparency of a model remains particularly important. Once a model is trained, it can be difficult to understand the origin of the results. This can be a disadvantage in certain situations.

That is what we generally call the trade-off of complexity and interpretability. It is linked to the fact that the smarter an algorithm is, the more incomprehensible it is. Consider the example of CART and random forest. A CART has the advantage of being easily understood because the model is relatively simple. A random forest, however, is an ensemble of multiple trees—it is more accurate and efficient than a single tree, but it is also much more difficult to interpret. In a CART, it is easy to distinguish which variables have the most influence on the result, whereas using a random forest is more difficult because the random forest uses the result of many trees to predict the output. Thus, when a simple model and a more complex one are similarly efficient, it is better to use the simplest model, all else being equal, because it is easier to interpret. As Cedric Villani says, with the emergence of big data, we have to be careful about the interpretability of different models. It is as important to analyze results closely to understand what is hidden behind models.

### SECTION 4.7 INSIGHTS / LESSONS LEARNED / RISKS

#### SECTION 4.7.1 DO YOUR HOMEWORK

Thanks to open-source software and packages in Python and R, machine-learning algorithms are easy to use, but also hard to understand. When implementing a new machine-learning algorithm, it is important to understand the theory behind the algorithm in order to understand how the algorithm works, in what instances the algorithm is appropriate, and what range of parameters is appropriate for a given situation. Most of these algorithms need many parameters that can change the results. This implies that knowing the role of these parameters and their impact on the model results are very important, both to execute the algorithms and also to interpret the results. While output can generally be obtained from a given data set and algorithm, a misused algorithm can lead to misleading results. Being aware of what happens within a theory or a given algorithm is the first step to understanding the appropriate uses of any individual algorithm.

#### SECTION 4.7.2 CHECK RELEVANCE OF THE IMPLEMENTATION / SELECTION OF ALGORITHM

Choosing a good algorithm to answer a problem is particularly important. Plenty of algorithms can solve a single question, but which one is the most efficient and relevant? It is important to be able to answer this question by looking at some evaluation criteria. Most of the time, the criterion that is used is the mean squared error (MSE), but with an unbalanced set of data, using the MSE is not meaningful. For example, in a problem of classification in which

there is 95% of zero and 5% of one, using the MSE as a selecting criterion is absurd because the MSE would probably be around 95%, which is close to always predicting zero. Instead of MSE, it is often better in such instances to look at other evaluation criteria that evaluate true negative rates (known as specificity) and true positive rates (known as sensitivity) such as the Gini Coefficient.

To evaluate an algorithm, we have to fix a threshold according to data and compare the value of the evaluation criterion to this threshold.

The relevance of an algorithm also relies on the choice of variable. It is important to use hypothesis testing to know the importance of each variable. For example, in a random forest, allowing the different trees to determine which feature is particularly important in the final result.

#### SECTION 4.7.3 ENSURE THE FAIRNESS OF THE RESULTS

It is important to check the fairness of the results of an algorithm. However, the notion of fairness can have different meanings. For example, the fairness of an algorithm can rely on the original data. If the data comes from a survey, the answers of the poll may be biased and the result of the algorithm similarly biased, but the fairness of the results can also mean checking whether the results are ethical or not. As machine learning models are used more and more in insurance and banking, it is important to implement strategies to ensure the fairness of those models and to detect potential discriminative behaviors during the estimation process. We can take the example of an insurance company that uses a predictive model in order to decide whether to insure a client or not. The inputs could be the client's attributes such as age, race, income, education, gender, etc. The results of such a model could have some legal implications if the model discriminates unfairly. In order to prevent this from occurring, ensuring the fairness of the results remains a step that should not be skipped.

#### SECTION 4.7.4 CHECK STABILITY OF THE ANALYSIS OVER TIME

On top of checking the relevance and the fairness of the results, it is necessary to check whether the analyses are stable over time. Indeed, an analysis can be true at a time  $t$ , but false at a time  $t + 1$ . This is even truer when we work on long-term analyses or predictions because the short-term is generally easier to predict than the long-term. Unexpected events such as natural disasters and terrorist attacks can influence the results of an analysis. That is why checking the results of an analysis and of a prediction regularly remains important in machine learning.

#### SECTION 4.7.5 INCREASE THE COMPLEXITY OF MODELS PROGRESSIVELY AND DO NOT HESITATE TO COMPARE THEM WITH TRADITIONAL APPROACHES

With machine learning, patience and simplicity are virtues. Sometimes complex models can be relevant and sometimes they are not. That is why it is interesting to progressively increase the complexity of models. For example, before using a random forest, a CART should be used in order to see which variables may be important. If a CART is not coherent, then a random forest is not coherent either. It is important not to rush into anything and to be patient to understand the problems.

Comparing machine-learning approaches and traditional approaches can also be interesting and provide useful information. For example, considering a classification problem, one could compare the result of a logistic regression and a CART or a random forest in order to see which individuals are classified similarly. This is useful to ensure the relevance of the algorithms and the results.

## Section 5: Future Developments and Insights

This section of the report discusses insights from the authors on possible future developments of cloud technology and its potential impacts on the actuarial profession. It also provides resources for actuaries to reference when learning about the topics discussed in this report in order to leverage cloud-enabled technological advances.

### SECTION 5.1 CHANGES TO ACTUARIAL PROFESSION

Cloud technology has the potential to impact many practice areas of actuaries. This section will summarize current and potential future impacts to the actuarial profession. This section refers to actuaries in general and does not differentiate between practice areas such as life, health, and property and casualty.

#### SECTION 5.1.1 PRICING ACTUARY

Cloud computing offers material increases in storage, data, and computing capacity compared to the most common data architecture available to an actuary. In the current environment, many actuaries are able to read data from centralized databases and data extracts are often summarized or aggregated. This potentially limits the complexity of pricing models in terms of what data can be analyzed for granular risk-based pricing.

With cloud-based data storage and computational capacity, these restrictions are reduced and actuaries can leverage more data to develop more granular pricing models at the policy level, instead of aggregate pricing models. As more data becomes available and more companies look to differentiate pricing to the individual, new techniques are required to develop actuarially sound pricing schemes.

It is possible that actuaries will need to expand their current skill set or develop innovative methodologies to be able to leverage additional data to create more granular pricing. Not only will actuaries themselves need to understand and get comfortable with the analysis required to develop individual pricing, but they will also need to be able to effectively communicate pricing changes and relationships both internally and with regulators. This is where machine learning may fall short of such requirements. While machine learning may result in more accurate predictions by being able to more efficiently identify and capture complex interactions within the data, it can be difficult, particularly for complex models such as gradient boosting, for the actuary to clearly articulate the resulting price indications to regulators to ensure fair and adequate insurance rates. Other nontechnical aspects, such as overall company strategy, suitability, legal, and regulatory risk that may not be straightforward to codify in a machine-learning algorithm, also need to be considered when developing such price indications.

To address this issue, the National Association of Insurance Commissioners (NAIC) recently formed an Innovation and Technology (EX) Task Force to explore the technological developments in the insurance sector. The Task Force will provide a forum "for the discussion of innovation and technology developments ... in order to educate state insurance regulators on how these developments will affect consumer protection, insurer and producer oversight ... and the state-based insurance regulatory framework"<sup>67</sup>. The task force will also look at emerging issues related to insurers or licensees leveraging new technologies, including artificial intelligence. One important initiative for the task force is to develop and discuss ways to share data used for pricing insurance using big data and to propose methodologies that will provide regulators with the ability to analyze such data.

---

<sup>67</sup> NAIC. "Innovation and Technology (EX) Task Force." Available at: [https://www.naic.org/cmt\\_e\\_ex\\_itf.htm](https://www.naic.org/cmt_e_ex_itf.htm).

A separate trend that could impact the pricing actuary is the movement from manual underwriting to automatic underwriting<sup>68</sup>. As actuaries incorporate both internal and external data (such as lifestyle information or geospatial data) into pricing algorithms, the underwriting process will be streamlined to provide personalized, on-demand quotes to policyholders. Pricing actuaries will have to consider available data sources, their reliability, and their robustness when developing such pricing algorithms in the future.

#### SECTION 5.1.2 VALUATION AND RESERVING

Similar to trends for pricing actuaries, reserving and valuation models are becoming more and more granular, with additional models employed to track policies through the life cycle of an insured. For example, mortgage insurance claims reserves historically were estimated as a function of how many payments a borrower missed. The more missed payments, the greater the likelihood of a claim. Increasingly, models are being deployed that take into consideration the number of missed payments, the equity position of the borrower, geographic-specific differences, and other features of the borrower or mortgage. In the future, data will be connected directly with servicers to understand if a borrower is in contact with the servicer to further segment between borrowers that are likely to result in a claim or cure the mortgage through a loan modification or other type of workout.

#### SECTION 5.1.3 ENTERPRISE RISK MANAGEMENT

Because cloud data is stored in remote data centers and can be efficiently shared with users across the globe, actuaries in enterprise risk management (ERM) can model enterprise-wide risk simultaneously to seamlessly integrate risks and models from various business units. Current practice is typically for each business unit to separately run their models under specific scenarios or confidence levels, then aggregate results at the holding-company level. With a fully integrated cloud platform, the holding company could have the ability to run company-wide analysis on-demand using the same shared production models within the underlying business units.

This type of functionality will become increasingly helpful to insurance companies as regulatory demands, such as Solvency II, require robust risk models across a given enterprise. It is reasonable to expect that these types of demands will become more common and require more granular analysis and reporting in the future.

In addition to sharing data and models, an integrated cloud platform across an enterprise could also be designed to facilitate reporting, as results could be directly fed into a business intelligence tool or dashboard that is shared with key stakeholders to enable drilling down into the data in real-time. An important function of enterprise risk management is to not only quantify risks but also to explain and manage them as they emerge. Applying business intelligence tools on top of detailed model results can facilitate real-time monitoring of risks and efficient communication. This communication can be either used for internal purposes or shared externally with regulators and/or rating agencies in the future.

#### SECTION 5.1.4 EXPERIENCE ANALYSIS AND ASSUMPTIONS

Another critical area that might be affected by deploying cloud-based platforms for actuaries is the development of assumptions and evaluating models and forecasts against actual experience in real-time. With an integrated cloud solution, data can be collected with respect to expenses, claims, premiums, and other factors and compared to model estimates. If the model assumptions (which can be used for pricing, valuation, or risk) are not consistent with model expectations, the model assumptions can be adjusted on-the-fly to produce revised forecasts or alerts to key

---

<sup>68</sup> McKinsey & Company. (2018). "Insurance 2030 – the impact of AI on the future of insurance." Available at: <https://www.mckinsey.com/industries/financial-services/our-insights/insurance-2030-the-impact-of-ai-on-the-future-of-insurance>.

stakeholders.

This is important for the actuary because it would provide more frequent and updated information that allows the actuary to understand and explain deviations between reporting periods. Additionally, it could help actuaries better segment differences between actual results and forecasts as there are typically many moving parts that impact model forecasts from period to period, such as assumptions, economic changes, and changes in the underlying composition of an insured population. Having access to real-time data and trends can facilitate better understanding of causes for deviations.

## SECTION 5.2 OPPORTUNITIES AND EDUCATION REQUIREMENTS / SKILLS REQUIRED

### SECTION 5.2.1 OPPORTUNITIES

The advances in data storage and computing, as enabled by the cloud, offer great opportunities to today's actuaries and actuaries of the future. Actuaries are in a unique position with respect to applying these advances to:

- Streamline processes and data, allowing actuaries to spend more time analyzing and communicating results
- Develop and implement granular, risk-based pricing tools
- Assist with the development of algorithms to target specific potential policyholders
- Optimize pricing to maximize the risk / reward trade-off of products

Actuaries are best equipped to take advantage of these opportunities because of their domain expertise in insurance, as well as their general analytical capabilities. However, in order to benefit from the opportunities presented by cloud computing and granular analysis, actuaries will need to either become data and technology experts or become familiar enough with these topics to effectively provide the required solutions and skill sets to employers.

The list below provides some introductory material actuaries can reference to expand their current skill sets to include proficiency in the topics discussed in this report.

#### *The 8 Best Places to Learn Cloud Computing*

This article provides a nice summary of places to learn cloud computing. While it is not an exhaustive list, it is a good start for those interested in furthering their knowledge of cloud computing.

Link: <https://www.akraya.com/blog/the-8-best-places-to-learn-cloud-computing>

#### *Cloud Computing: Concepts, Technology, & Architecture*

This book discusses the basic concepts of the cloud. It will dive into clusters, the difference between parallel and distributed computing, and much more. Details on architecture and mechanisms can also be found if the reader is interested.

Link: <https://books.google.com/books?id=czCij6sbhpAC>

### *Best Practices for Scientific Computing*

This report suggests best practices for projects that deal with leveraging the use of computing resources. It discusses important practices that are aimed at improving the productivity and reliability of results when using computing resources.

Link: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001745>

### *An Introduction to Statistical Learning*

This book introduces the important concepts of machine learning while being focused on the application of its techniques. It does this by providing a number of R labs that give real-world use cases of various techniques. It is also on the list of required reading on the SOA's syllabus for the Statistics for Risk Modeling Exam.

Link: <http://www-bcf.usc.edu/~gareth/ISL/>

### *Kaggle*

Kaggle provides a variety of ways to increase data science skills. It hosts competitions where users can practice their skills and provides educational courses for those looking to learn new skills. It also has useful blogs and kernels where people share and discuss their code and solutions.

Link: <https://www.kaggle.com/>

### *Coursera / EdX*

There are a variety of online educational resources providing access to course material from universities such as MIT, Harvard, and others, as well as classes led by industry thought leaders and companies. Two examples of these providers are Coursera and EdX. Both have material specific to cloud developers, machine learning, and additional topics discussed in this research report.

Link: <https://www.coursera.org/>

Link: <https://www.edx.org/>



## Glossary of Terms

**Artificial Intelligence (AI)** – Any attempt to make machines learn from experience and perform human tasks.

**Central Processing Unit (CPU)** –The “brain” of a computer that carries out program instructions and calculations.

**Classification and Regression Tree (CART)** – A decision-tree algorithm that can be used both for regression and for classification.

**Cloud Computing** – The internet-based outsourcing of digital storage along with access to a collection of resources, including software, databases, data warehouses, and computational processing.

**Community Cloud** – The cloud-deployment model where IT resources are used within the context of several organizations having permissioned access to a select set of applications.

**Distributed Computing** – A computational method made up of a network of computers working together as one system.

**Graphical Processing Unit (GPU)** –The graphics processor of a computer, comparable to a CPU, commonly used for training deep-learning models.

**Hybrid Cloud** – The cloud-deployment model where IT resources are distributed across different cloud-based systems, such as utilizing a public cloud for data that has been identified as non-sensitive while employing a private cloud for internal business systems and processes that demand the safeguarding of intellectual property and/or sensitive data.

**Infrastructure as a Service (IaaS)** –The cloud-service model where providers offer on-demand infrastructure such as servers, storage, virtual machines, and networking capabilities.

**Insurtech** – The products and companies that specialize in the combination of insurance and emerging technologies.

**Internet of Things (IoT)** – The connection of ordinary devices and objects to the internet.

**K-means** – An algorithm that groups individual data into clusters according to their characteristics.

**Machine Learning (ML)** –A subfield of artificial intelligence concerned with allowing machines, programs, or algorithms to learn and improve from data.

**Parallel Computing** – A coordinated computational method that simultaneously uses multiple-system resources to solve a computational problem.

**Platform as a Service (PaaS)** – The cloud-service model where a provider offers the platform that a company can use to make and distribute software. The cloud service provides a basic framework for building applications and programs.

**Private Cloud** – The cloud-deployment model where IT resources may either be on-premise or deployed via a third-party vendor, but access to the services and/or infrastructure is limited to a single organization.

**Public Cloud** – The cloud-deployment model where IT resources are shared amongst vendor customers, and the cloud environment is owned, as well as maintained, by the third-party vendor (e.g., Amazon, Microsoft, Oracle, IBM).

**Software as a Service (SaaS)** – The cloud-service model that provides immediate access to software and applications that have been placed on cloud servers, usually by third-party vendors.

## Appendix A: References

- Amazon Web Services. (n.d.). "Six advantages of cloud computing." Available at: <https://docs.aws.amazon.com/aws-technical-content/latest/aws-overview/six-advantages-of-cloud-computing.html>.
- Amazon Web Services. (n.d.). "What is a data lake?" Available at: <https://aws.amazon.com/fr/big-data/datalakes-and-analytics/what-is-a-data-lake/>.
- Applied Systems. (2018). "The next generations and insurance: Understanding buying behaviours and preferences of younger adults." Available at: [https://www1.appliedsystems.com/media/3203/understanding-millennial-gen-z-buying-behaviors\\_en-uk.pdf](https://www1.appliedsystems.com/media/3203/understanding-millennial-gen-z-buying-behaviors_en-uk.pdf).
- App Orchid. (2019). <https://www.apporchid.com/claims-360/>.
- AXA. (2017). "AXA goes blockchain with fizzy." Available at: <https://group.axa.com/en/newsroom/news/axa-goes-blockchain-with-fizzy>.
- Badr, Will. (n.d.). "Evaluating machine learning models' fairness and bias." Available at: <https://towardsdatascience.com/evaluating-machine-learning-models-fairness-and-bias-4ec82512f7c3>.
- Baltazar, Gino. (2018). "CPU vs GPU in machine learning." Oracle. Available at: <https://www.datascience.com/blog/cpu-gpu-machine-learning>.
- Batty, Marc & Morel, Médéric. (n.d.). *Big Data et Machine Learning - 2e éd.* De Pirmin Lemberger.
- Bozicevic, Vedran. (2018). "Cloud computing benefits: 7 key advantages for your business." GlobalDots. Available at: <https://www.globaldots.com/cloud-computing-benefits/>.
- Braun, Mikio. (2014). "What is scalable machine learning?" Available at: <http://blog.mikiobraun.de/2014/07/what-is-scalable-machine-learning.html>.
- BrightPlanet. (n.d.). "Structured vs. unstructured data." Available at: <https://brightplanet.com/2012/06/structured-vs-unstructured-data/>.
- Brownlee, Jason. (2016). "Classification and regression trees for machine learning." Available at: <https://machinelearningmastery.com/classification-and-regression-trees-for-machine-learning/>.
- Brownlee, Jason. (2016). "Supervised and unsupervised machine learning algorithms." Available at: <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>.
- CIO Applications. (n.d.). "App Orchid: Harnessing the power of AI." Available at: <https://insurtech.cioapplications.com/vendor/app-orchid-harnessing-the-power-of-ai-cid-1713-mid-96.html>.
- Conroy, Sean. (2018). "History of virtualization." Available at: <https://www.idkrtm.com/history-of-virtualization/>.
- Costello, Katie. (2019). "I&O leaders need to strategically leverage AI as a core accelerant to digital business initiatives." Gartner. Available at: <https://www.gartner.com/smarterwithgartner/gartner-predicts-the-future-of-ai-technologies/>.
- DARPA. (2019). "About DARPA." Available at: <https://www.darpa.mil/about-us/about-darpa>.

- Deloitte. (2018). "Dynamic pricing is not just about profitability." Available at: <https://www2.deloitte.com/nl/nl/pages/financial-services/articles/dutch-insurance-outlook-2017-dynamic-pricing.html>.
- Expert Systems. (n.d.). "What is web 3.0?" Available at: <https://www.expertsystem.com/what-exactly-is-web-30/>.
- Farmers Insurance. (2017). "First-of-its-kind training module's immersive, virtual environments deliver dynamic learning experience for claims representatives." Available at: <https://www.farmers.com/news/2017/farmers-insurance-pioneers-innovative-virtual-reality-training-program-for-claims-representatives/>.
- Elastic. (n.d.). "Au cœur de la Suite Elastic." Available at: <https://www.elastic.co/fr/products/elasticsearch>.
- Foote, Keith D. (2017). "A brief history of cloud computing." Dataversity. Available at: <https://www.dataversity.net/brief-history-cloud-computing/>.
- Fry, Richard. (2018). "Millennials projected to overtake Baby Boomers as America's largest generation." Pew Research. Available at: <https://www.pewresearch.org/fact-tank/2018/03/01/millennials-overtake-baby-boomers/>.
- Gallup. (2015). "Insurance companies have a big problem with Millennials." Available at: <https://news.gallup.com/businessjournal/181829/insurance-companies-big-problem-millennials.aspx>.
- Gangcuangco, Terry. (2018). "Willis Towers Watson reveals the extent of the insurtech boom." Insurance Business. Available at: <https://www.insurancebusinessmag.com/uk/news/technology/willis-towers-watson-reveals-the-extent-of-the-insurtech-boom-91200.aspx>.
- Gartner. (2018). "Gartner forecasts worldwide public cloud revenue to grow 17.3 percent in 2019." Available at: <https://www.gartner.com/en/newsroom/press-releases/2018-09-12-gartner-forecasts-worldwide-public-cloud-revenue-to-grow-17-percent-in-2019>.
- Git. (n.d.). "Getting started – About version control." Available at: <https://git-scm.com/book/en/v2/Getting-Started-About-Version-Control>.
- Goasduff, Laurence. (2015). "The financial case for moving to the cloud." Gartner. Available at: <https://www.gartner.com/smarterwithgartner/the-financial-case-for-moving-to-the-cloud/>.
- Harvey, Cynthia. (2017). "Cloud computing cost calculators." Available at: <https://www.datamation.com/cloud-computing/cloud-costs-calculators.html>.
- Hauben, Michael. (1995). "Behind the net: The untold story of the ARPANET and computer science." Available at: <http://www.columbia.edu/~rh120/ch106.x07>.
- IBM. (2010). "What is batch processing?" Available at: [https://www.ibm.com/support/knowledgecenter/zosbasics/com.ibm.zos.zconcepts/zconc\\_whatbatch.htm](https://www.ibm.com/support/knowledgecenter/zosbasics/com.ibm.zos.zconcepts/zconc_whatbatch.htm).
- IBM. (n.d.). "What is distributed computing?" Available at: [https://www.ibm.com/support/knowledgecenter/en/SSAL2T\\_8.1.0/com.ibm.cics.tx.doc/concepts/c\\_wht\\_is\\_distd\\_comptg.html](https://www.ibm.com/support/knowledgecenter/en/SSAL2T_8.1.0/com.ibm.cics.tx.doc/concepts/c_wht_is_distd_comptg.html).
- IDG. (2018). 2018 Cloud Computing Survey. Available at: <https://www.idg.com/tools-for-marketers/2018-cloud-computing-survey/>.

- Junod, Betty. (2017). "MetLife uses Docker Enterprise Edition to self fund containerization." Docker. Available at: <https://blog.docker.com/2017/10/metlife-docker-enterprise-edition-self-funded-containerization/>.
- Kerner, S.M. (2017). "MetLife navigates DevOps with Docker containers." Datamation. Available at: <https://www.datamation.com/cloud-computing/metlife-navigates-devops-with-docker-containers.html>.
- Kohlbrenner, Eric; Morris, Dana; Morris, Brett. (n.d.). "The history of virtual machines." Available at: <http://denninginstitute.com/itcore/virtualmachine/history.htm>.
- Lemonade. (2019). Available at: <https://www.lemonade.com/>.
- Licklider, J.C.R. (1963). "Memorandum for members and affiliates of the Intergalactic Computer Network." Available at: <http://www.kurzweilai.net/memorandum-for-members-and-affiliates-of-the-intergalactic-computer-network>.
- Lyman, Michael. (2018). "Insurers strive for greater agility to secure big digital pay-off." Accenture. Available at: <https://insuranceblog.accenture.com/insurers-strive-for-greater-agility-to-secure-big-digital-pay-off>.
- McKinsey Digital. (2018). "Cloud adoption to accelerate IT modernization." Available at: <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/cloud-adoption-to-accelerate-it-modernization>.
- McLellan, Charles. (2017). "XaaS: Why 'everything' is now a service." ZDnet.com. Available at: <https://www.zdnet.com/article/xaas-why-everything-is-now-a-service/>.
- Microsoft Azure. (2019). "What are public, private, and hybrid clouds?" Available at: <https://azure.microsoft.com/en-us/overview/what-are-private-public-hybrid-clouds/>.
- Neto, Maximillano Destefani. (2014). "A brief history of cloud computing." IBM. Available at: <https://www.ibm.com/blogs/cloud-computing/2014/03/18/a-brief-history-of-cloud-computing-3/>.
- Oracle. (2012). "Brief history of virtualization." Available at: [https://docs.oracle.com/cd/E26996\\_01/E18549/html/VMUSG1010.html](https://docs.oracle.com/cd/E26996_01/E18549/html/VMUSG1010.html).
- Panetta, Kasey. (2018). "Gartner top 10 strategic technology trends for 2019." Gartner. Available at: <https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2019/>.
- Pew Research Center. (2018). "Mobile fact sheet." Available at: <https://www.pewinternet.org/fact-sheet/mobile/>.
- Presutti, Daniele. (2017). "Cloud computing offers insurers big TCO savings." Accenture. Available at: <https://insuranceblog.accenture.com/cloud-computing-offers-insurers-big-tco-savings>.
- RightScale-Flexera. (2018). "Cloud cost optimization." Available at: <https://www.rightscale.com/learn/cloud-optimization/cloud-cost-optimization-best-practices>.
- RightScale-Flexera. (2019). "State of the Cloud Report." Available at: <https://info.flexerasoftware.com/SLO-WP-State-of-the-Cloud-2019>
- Rothrie, Sarah. (2019). "What is distributed computing? Examining blockchain's backbone." Coincentral. Available at: <https://coincentral.com/what-is-distributed-computing/>.

Salesforce. (2015). "Why move to the cloud? 10 benefits of cloud computing." Available at: <https://www.salesforce.com/uk/blog/2015/11/why-move-to-the-cloud-10-benefits-of-cloud-computing.html>.

Society of Actuaries. (2019). "Guidelines for Predictive Modeling in Insurance Applications." Available at: <https://www.soa.org/resources/research-reports/2019/considerations-predictions-insurance-applications/>.

Statista. (2019). "Public cloud Software as a Service (SaaS) revenue worldwide from 2016 to 2027 (in billion U.S. dollars)." Available at: <https://www.statista.com/statistics/477742/public-cloud-software-revenue-forecast/>.

Tribe, Jennifer. (2016). "The big -aaS list of As-a-Service offerings." Available at: <https://www.auvik.com/franklymsp/blog/aas-as-a-service-list/>.

Turner, Heather. (2018). "What do millennials want from their insurance agency?" Insurance Business America. Available at: <https://www.insurancebusinessmag.com/us/news/technology/what-do-millennials-want-from-their-insurance-agency-115007.aspx>.

U.S. National Institute of Standards and Technology. (2011). "The NIST Definition of Cloud Computing." Available at: <https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-145.pdf>.

Vaghese, Blesson; Buyya, Rajkumar. (2017). "Next generation cloud computing: New trends and directions." Future Generation Computer Systems, ISSN: 0167-739X, Elsevier Press, Amsterdam, The Netherlands, 2017 (in press). Available at: <https://arxiv.org/pdf/1707.07452.pdf>.

Villani, Cedric. (2017). "Entretien avec Cedric Villani." Available at: <http://www.accroche-press.fr/7448-2/>.

Webroot. (n.d.). "5 financial benefits of moving to the cloud." Available at: <https://www.webroot.com/au/en/resources/tips-articles/five-financial-benefits-of-moving-to-the-cloud>.

Wojciechowicz, Rebecca. (2018). "70% of Insurers Now Using Cloud Computing, With Additional Growth Planned, Says Novarica." NOVARICA. Available at: <https://novarica.com/70-insurers-now-using-cloud-computing-additional-growth-planned-says-novarica/>.

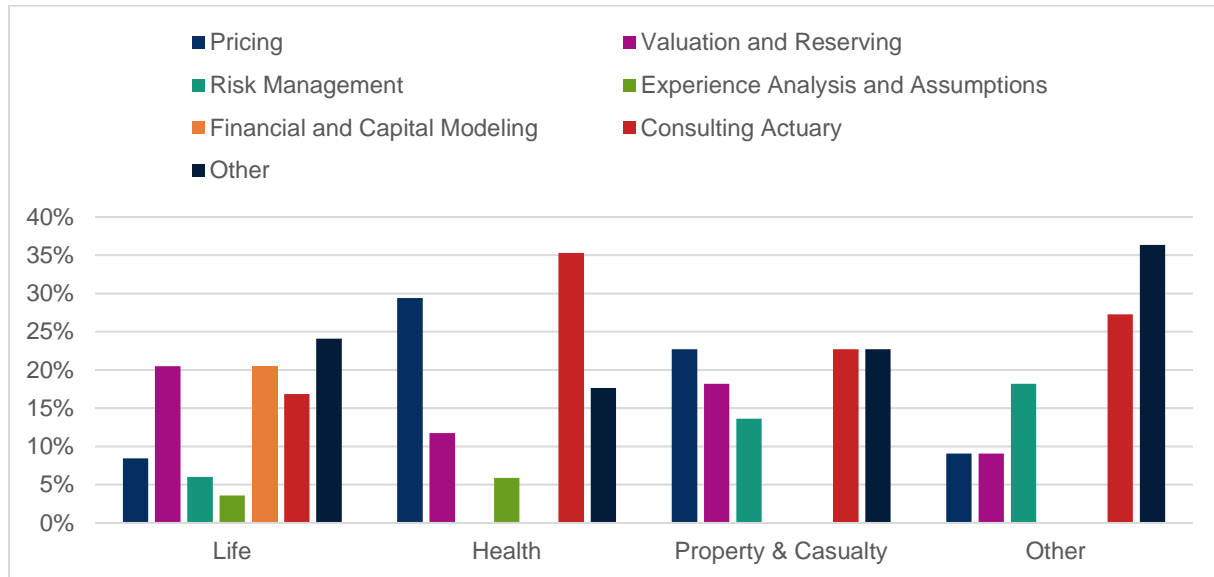
Wüthrich, Mario V. (n.d.). "A neural network extension of the Lee-Carter Model to multiple populations."

Wüthrich, Mario V. (n.d.). "Machine learning in individual claims reserving."

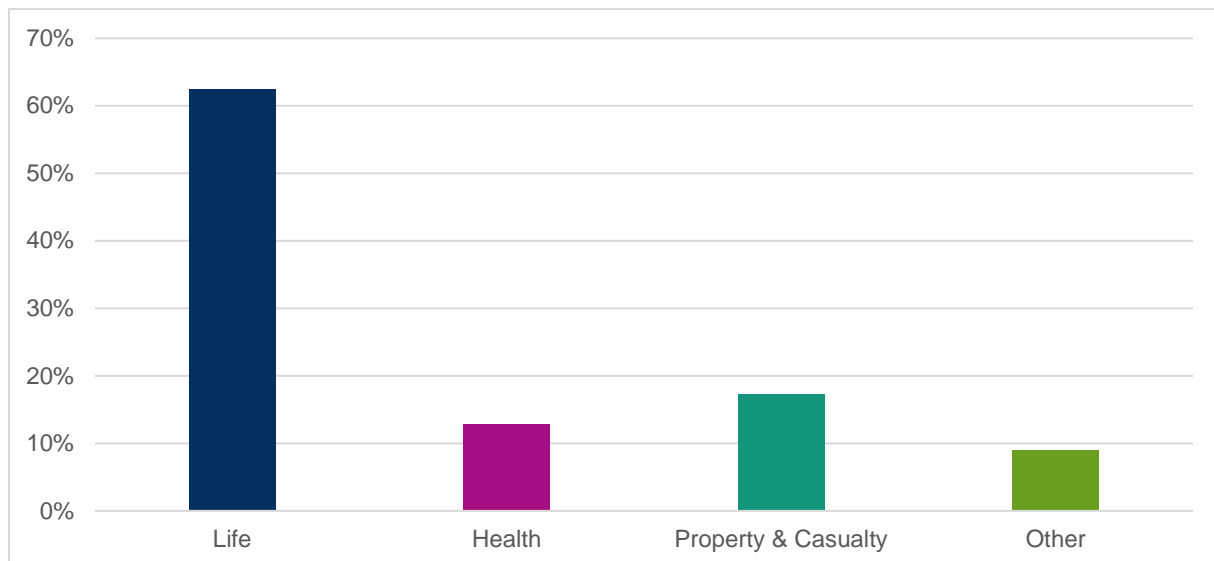
Wüthrich, Mario V. (n.d.). "Neural networks applied to chain-ladder reserving."

## Appendix B: Cloud Computing Survey Questions

Q1. What is your current actuarial role?

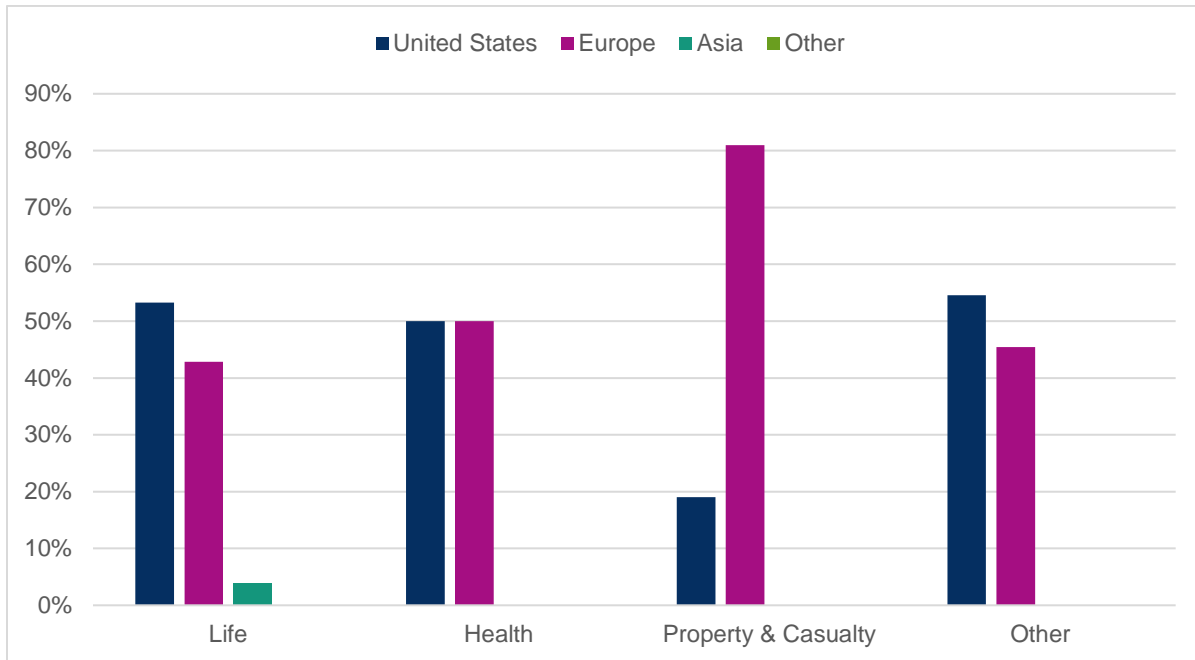


Q2. What is your primary area of practice?

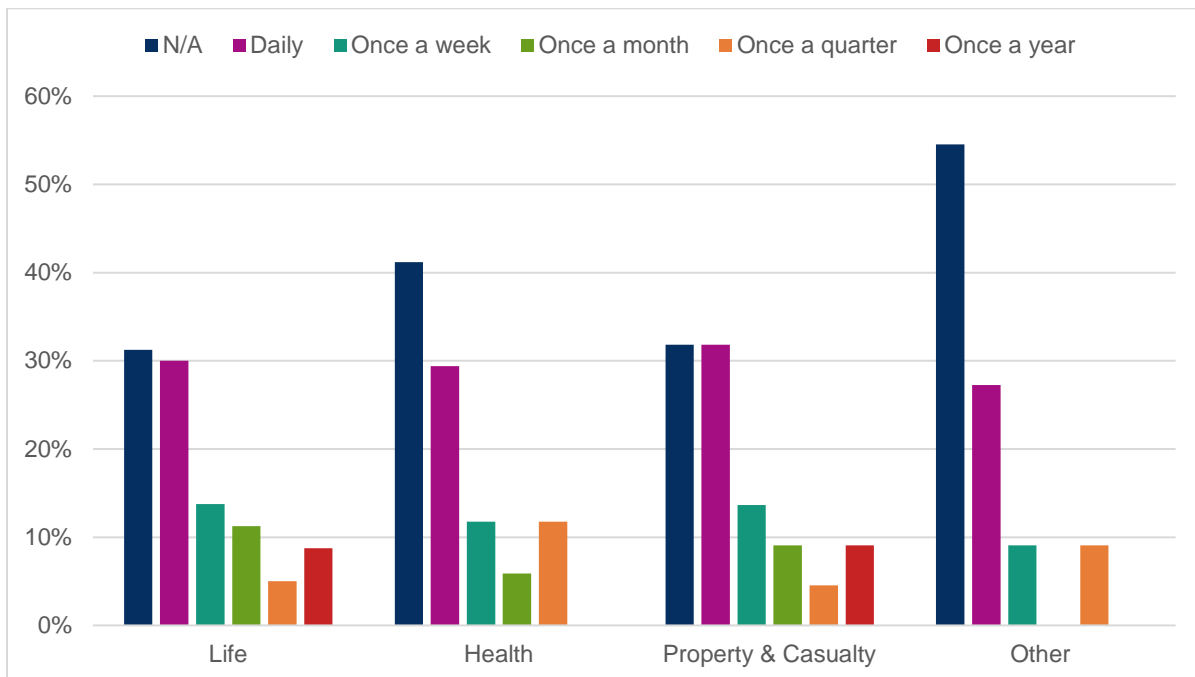




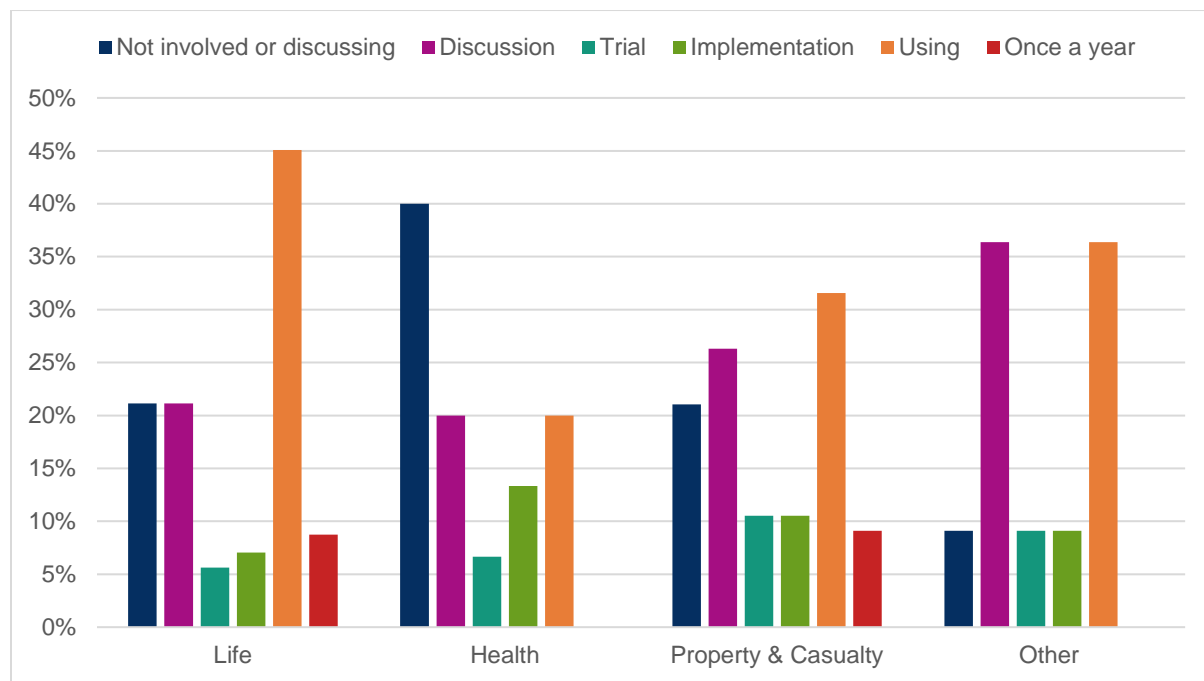
Q3. What is the primary geographic region of your office?



Q4. How frequently do you use cloud computing in your current role as an actuary?

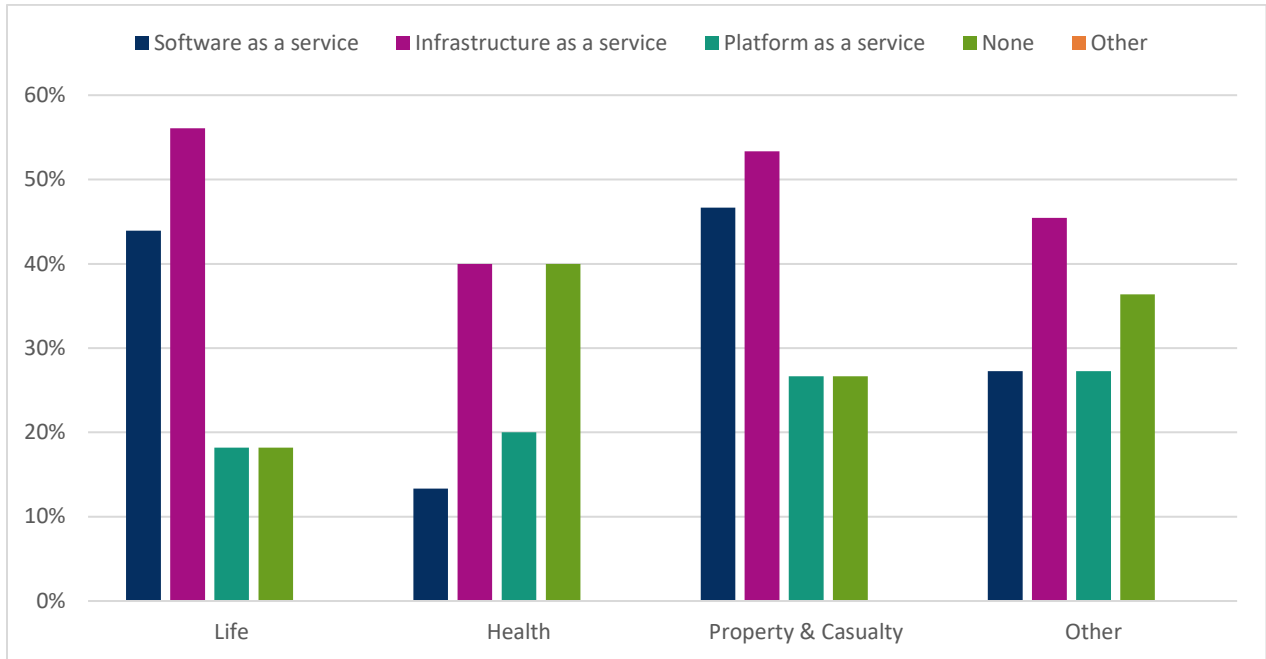


Q5. At what stage is the actuarial department within your organization at with regards to leveraging either cloud storage or computing for financial modeling and/or actuarial processes for your current role?

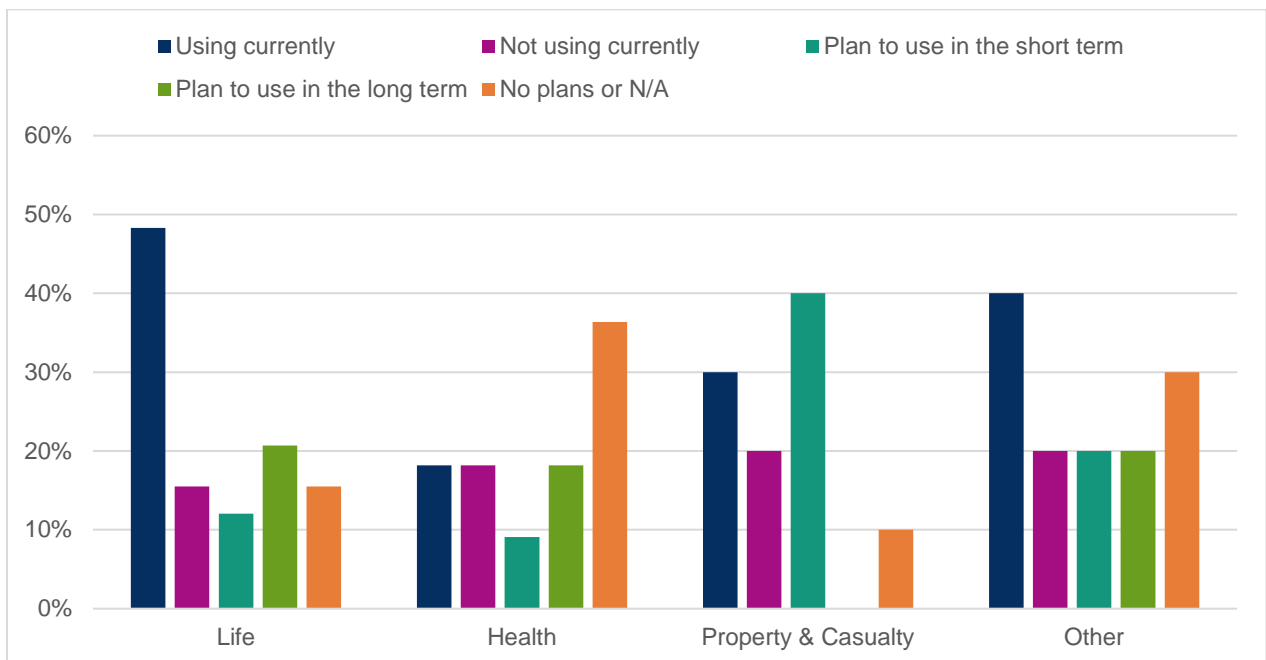


Q6. Which of these types of cloud computing is your organization currently using for financial modeling and/or actuarial processes (select all that apply)?

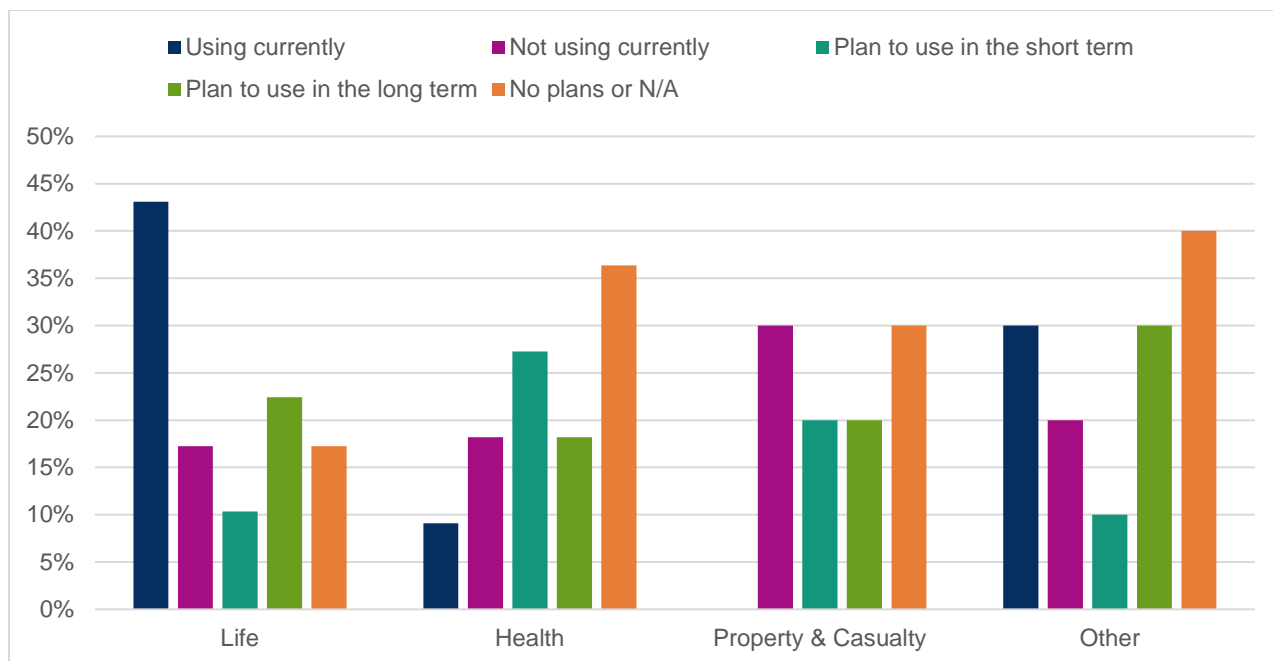
A) Service Model



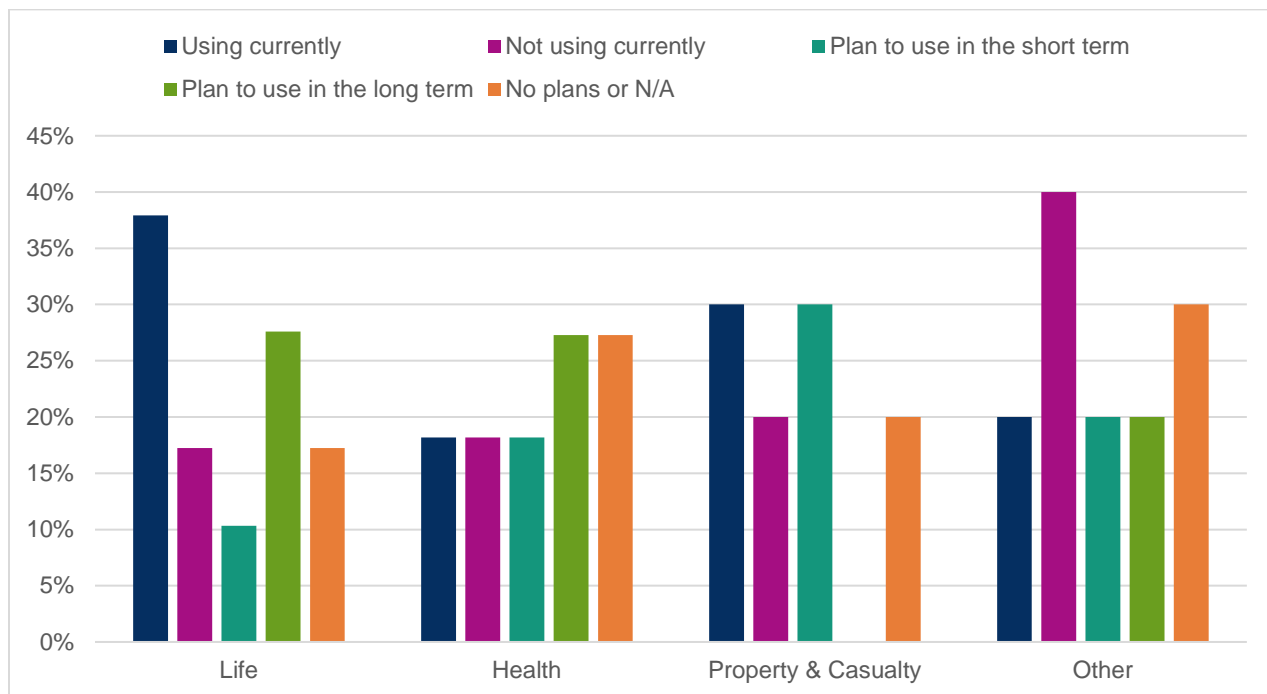
B) Distribution Calculation Engine (on an ad-hoc basis)



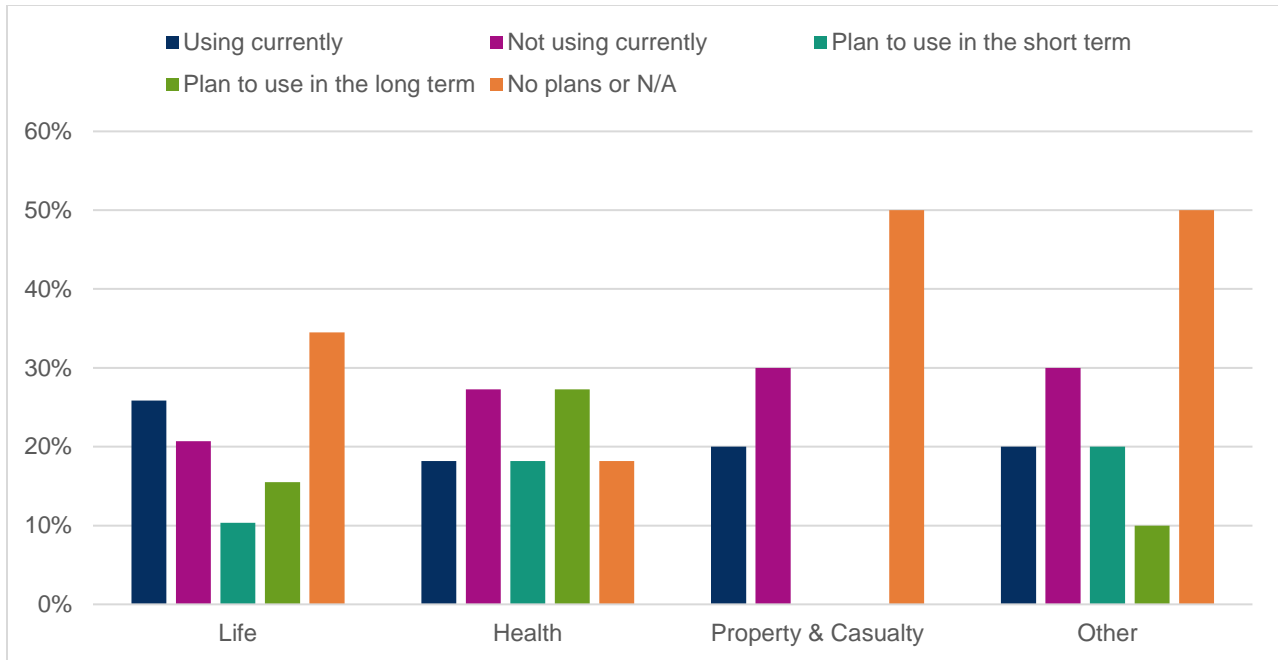
C) Distributed Calculation Engine (for production)



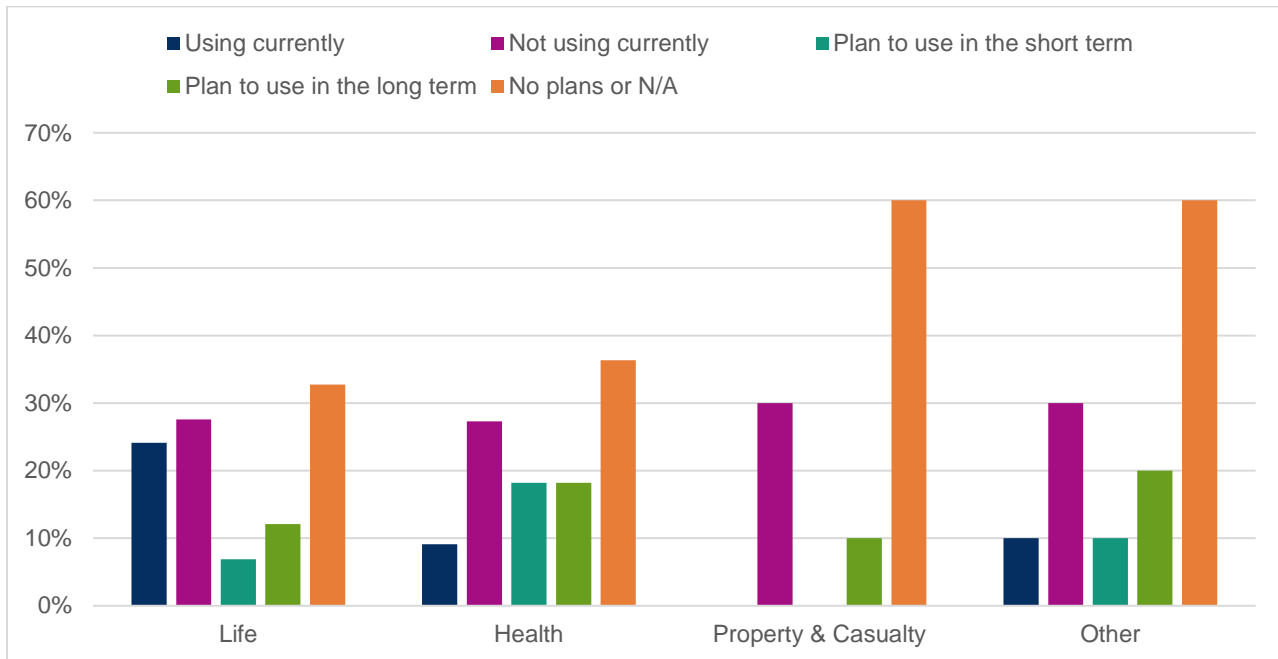
D) Model Development



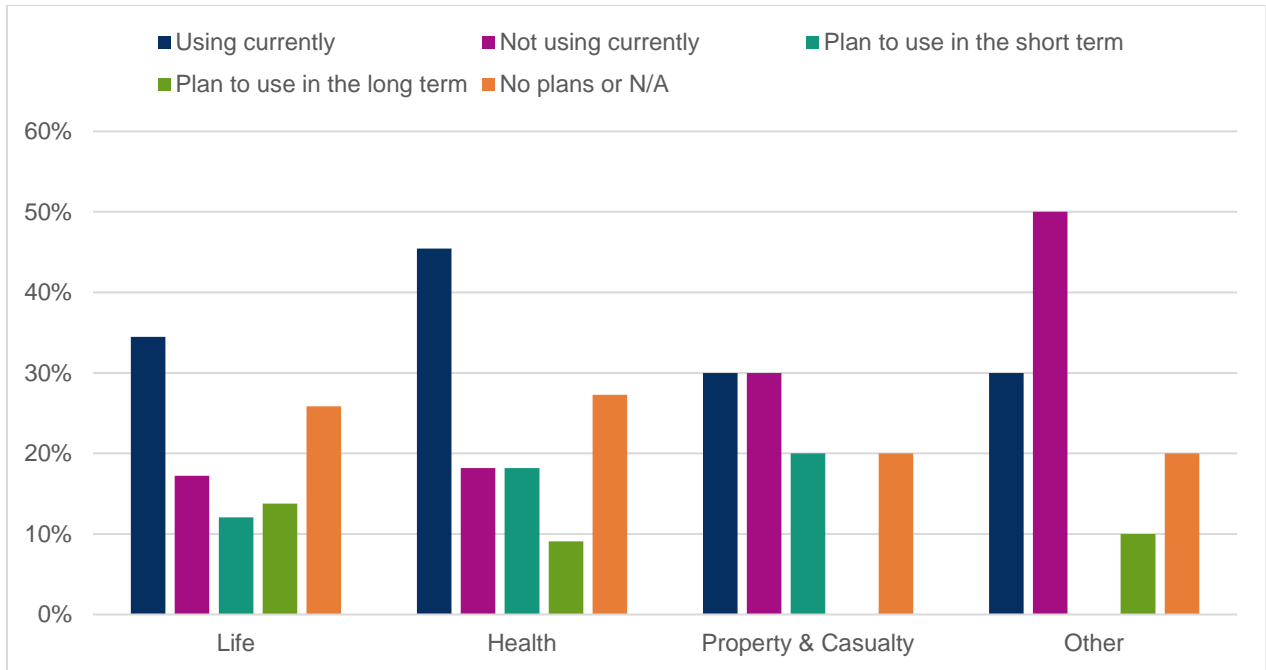
E) Model Governance



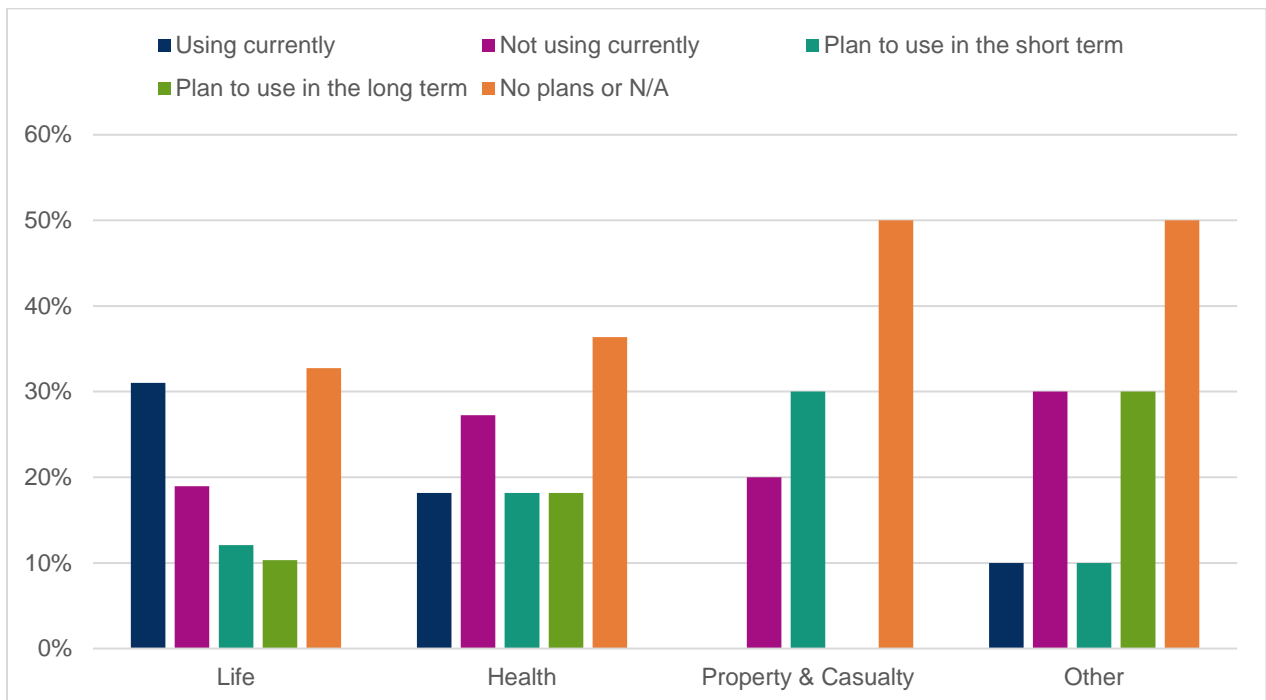
F) Assumption Management



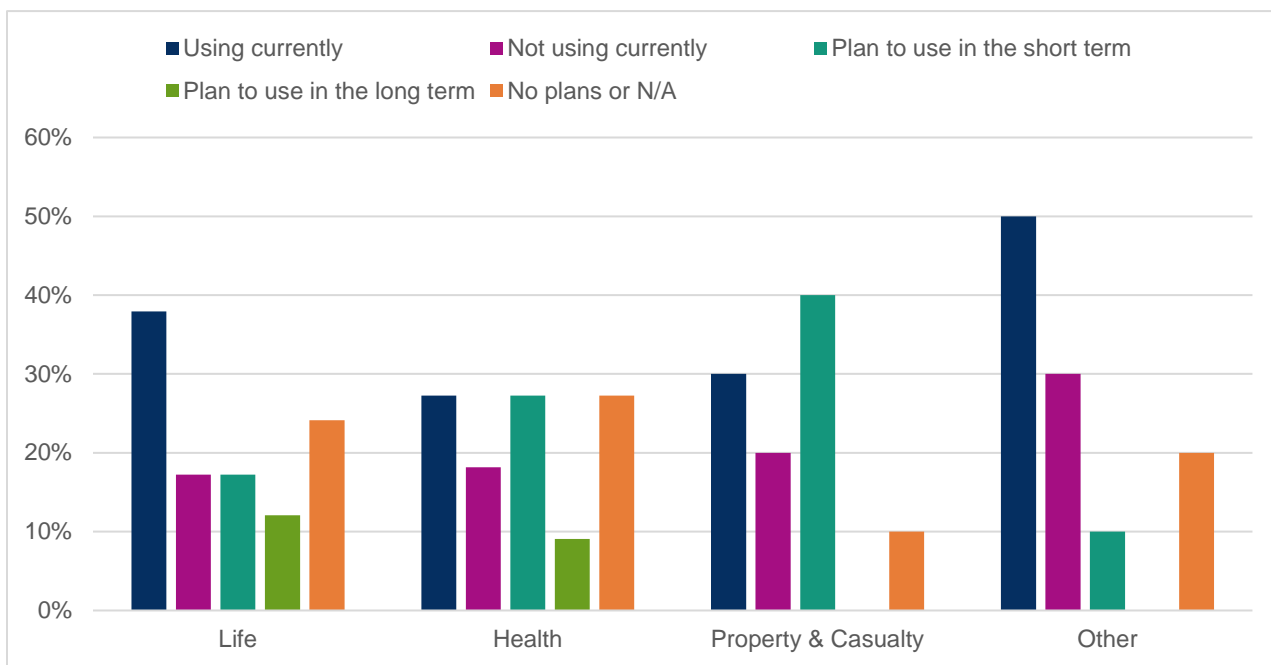
G) Data Input Processing



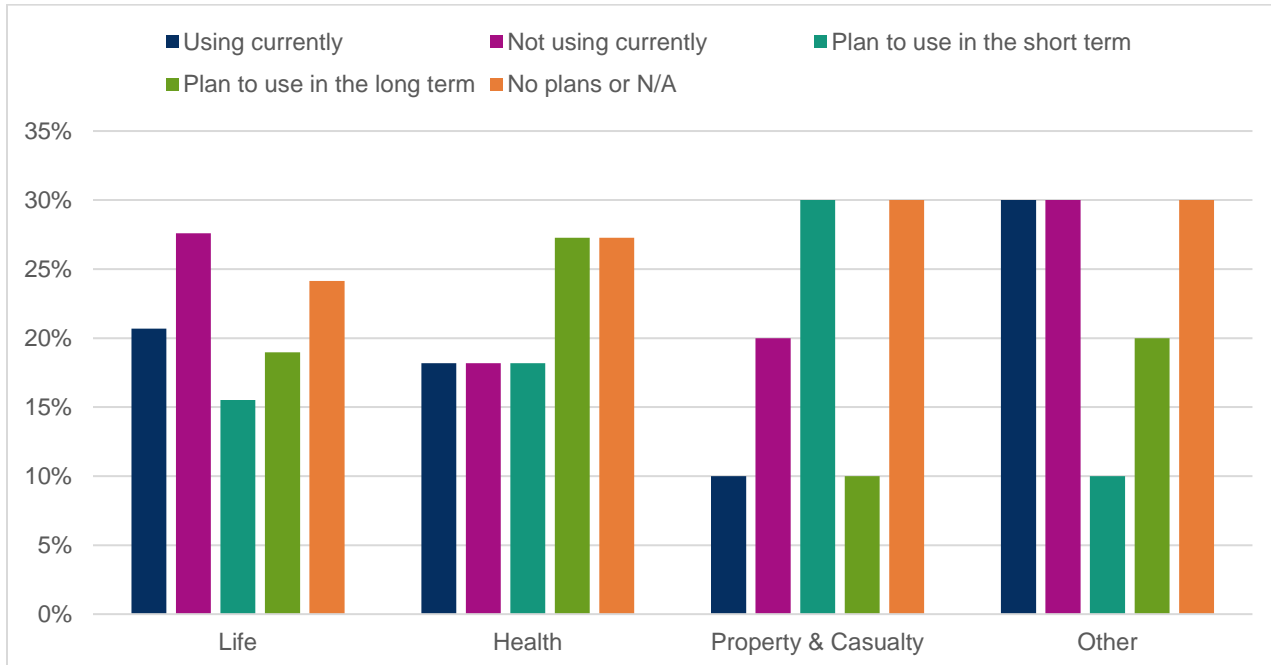
H) Model Post-Processing



I) Data Visualization\Reporting

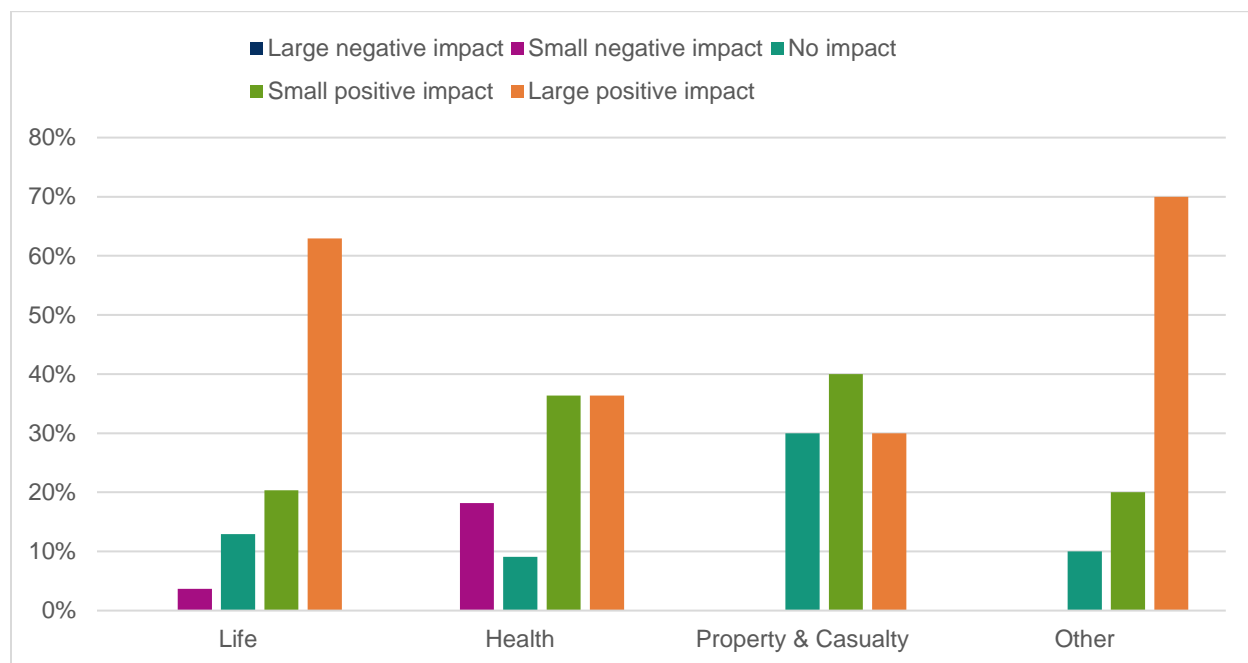


J) Predictive Analytics

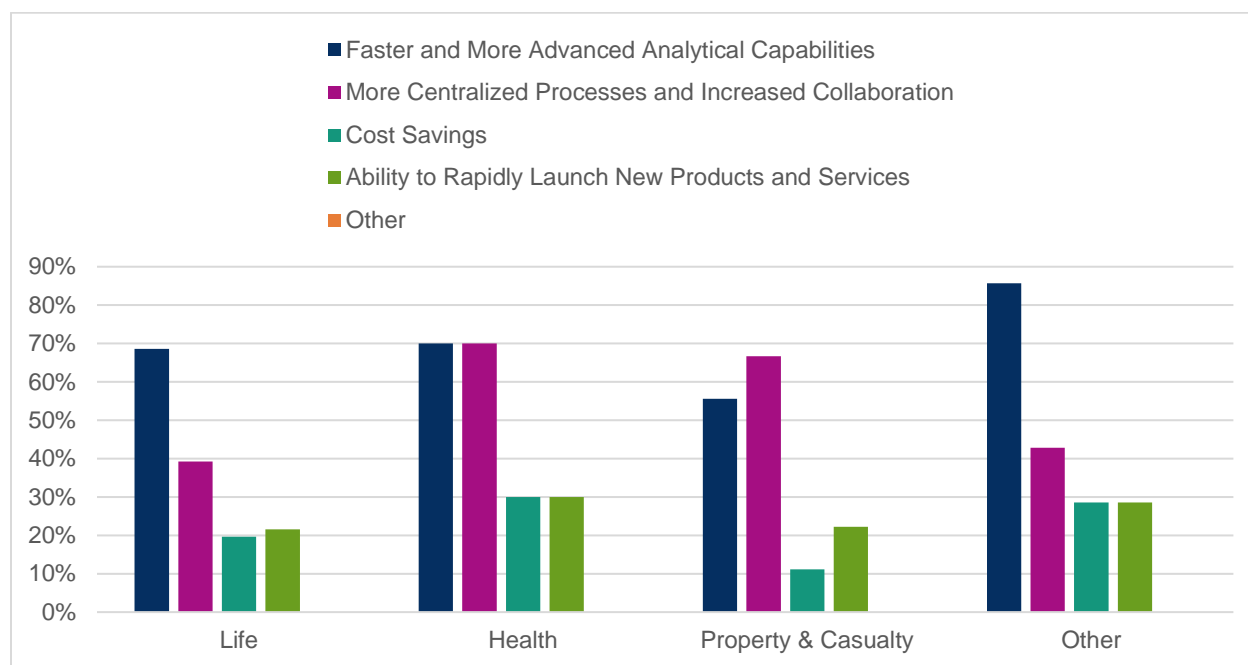




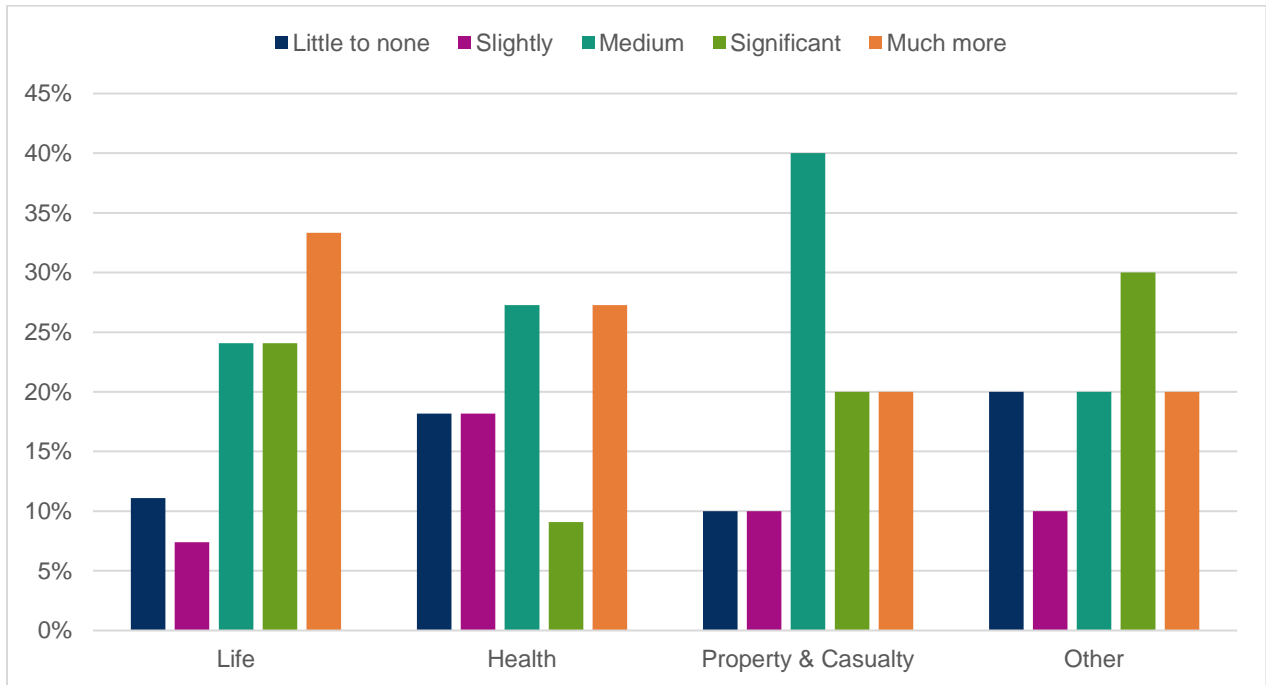
Q8. How large of an impact to actuarial work does access to the cloud offer? For this question we are referring to both cloud storage (e.g. larger data sets) and cloud computing (e.g. faster computation times). (1 being large negative impact, 5 being high positive impact)



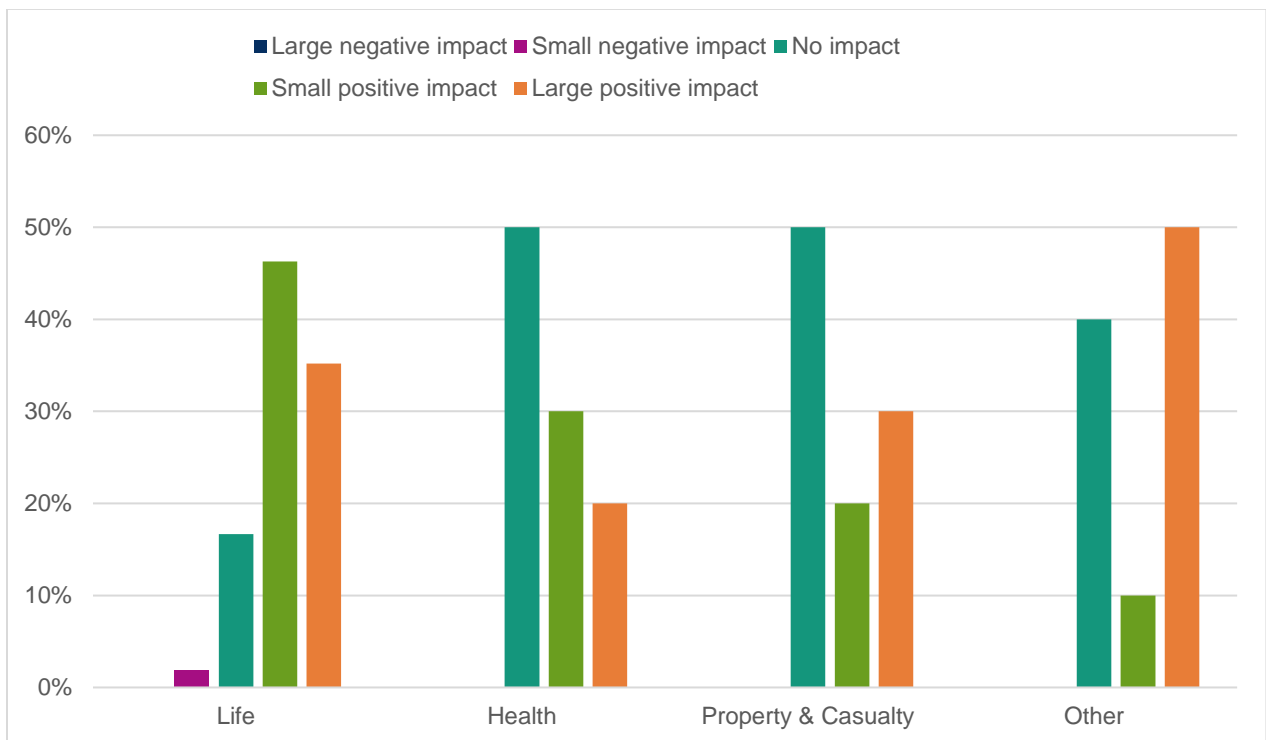
Q9. What do you see as being the most useful benefits of the cloud for the actuarial department of your organization (select one)?



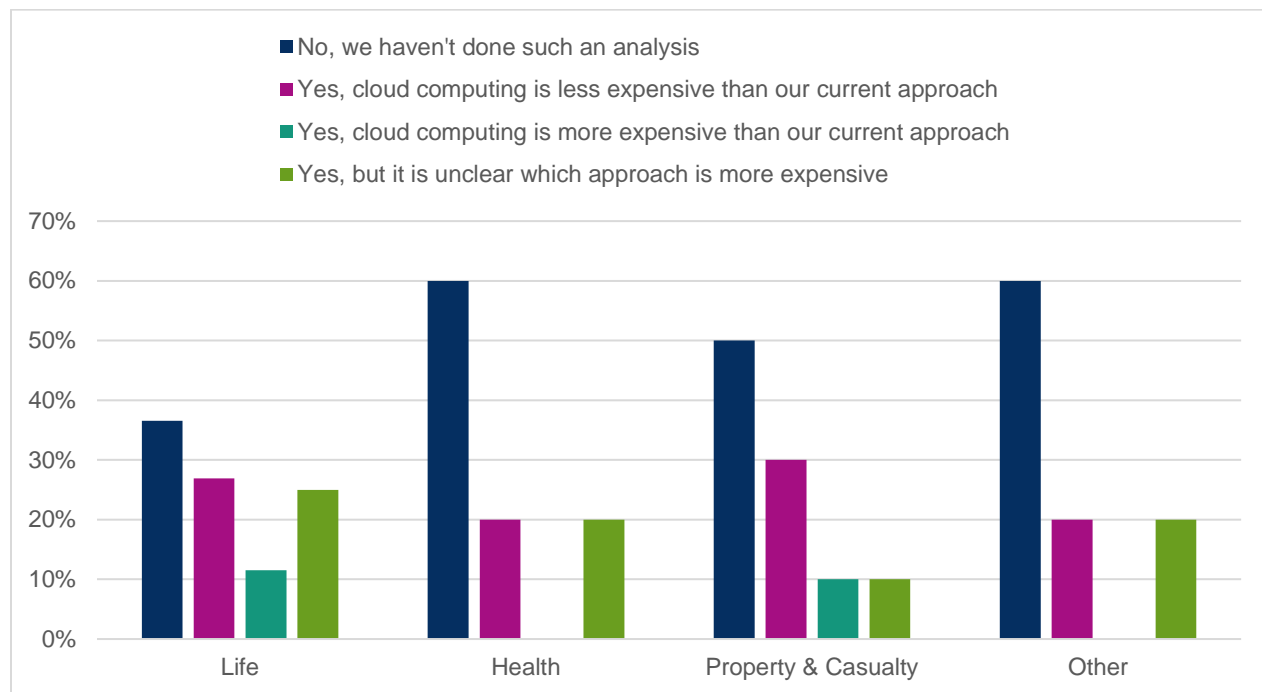
Q10. Within the next 5 years, how much more prevalent do you see the cloud becoming in the actuarial department of your organization? (1 being little to none and 5 being much more)



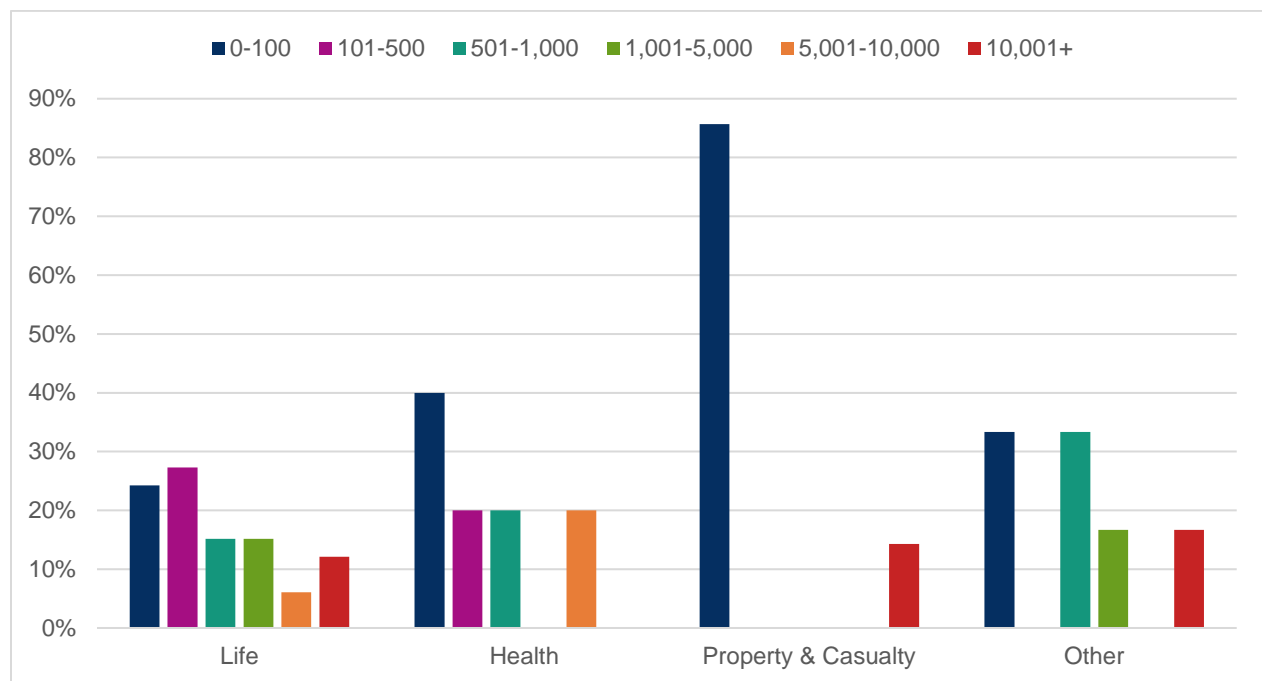
Q11. How much do you expect cloud computing to change either the work flow or analysis demanded from your current job function over the next five years? (1 being large negative impact, 5 being large positive impact)



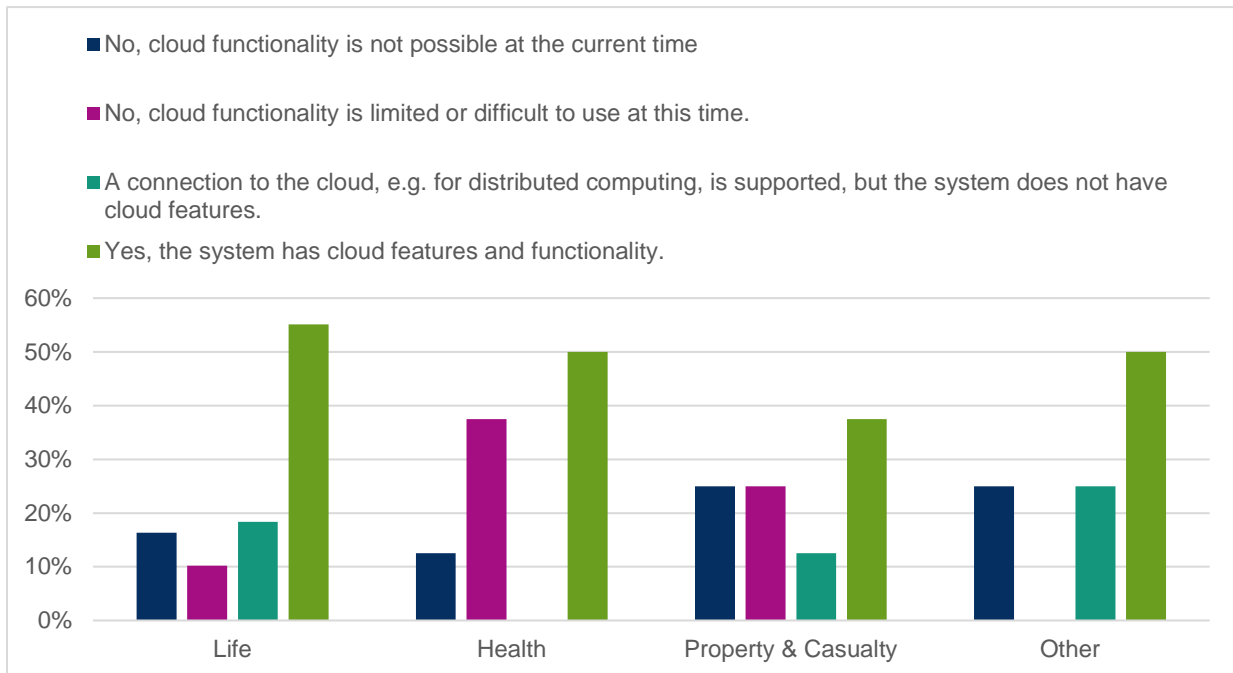
Q12. Has your organization performed a cost-benefit analysis of using the cloud for financial modeling and/or actuarial processes?



Q13. If you are currently using the cloud for financial modeling distributed computing, how many cores do you typically use for production in your current role?



Q14. In your view, is your actuarial modeling platform well equipped to take advantage of the cloud?

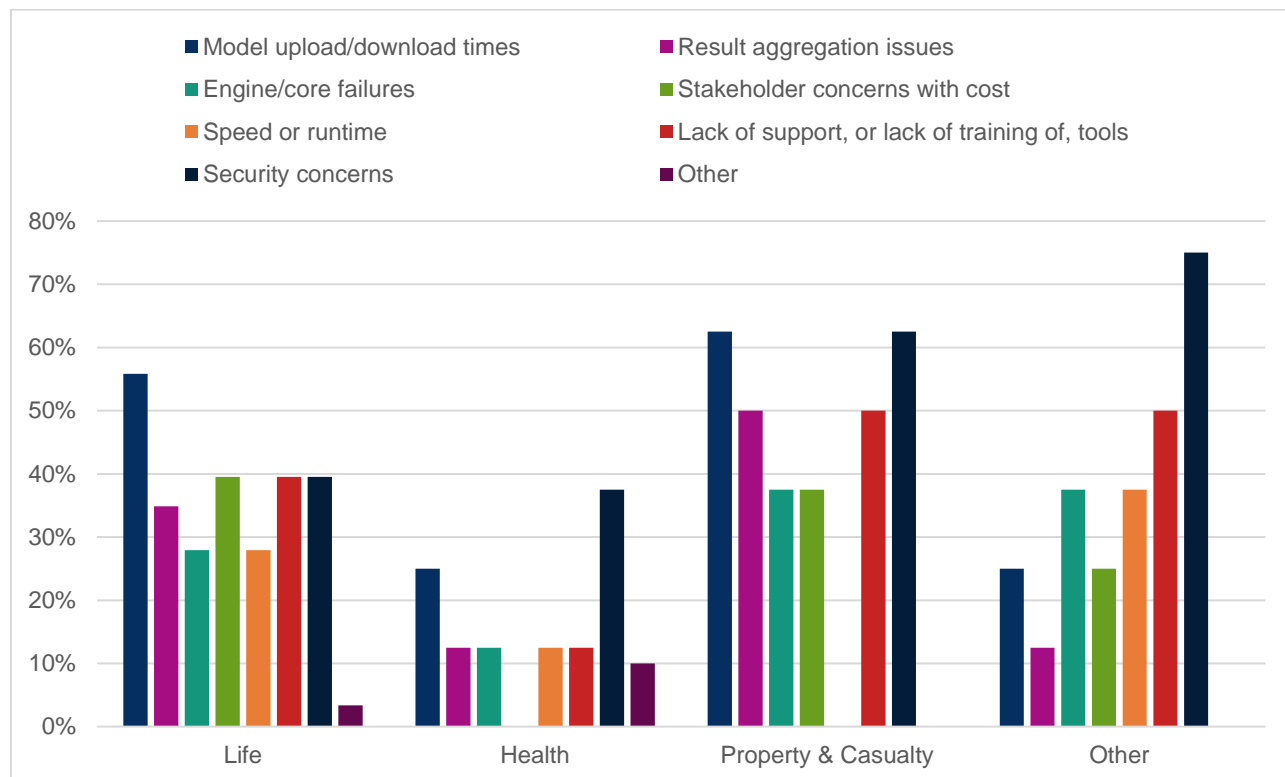


Q15. When considering future actuarial modeling platforms, how important is the ability of the platform to have cloud capabilities?



Q16. Are there any bottlenecks you have faced with using the cloud for financial modeling and/or actuarial processes when using distributed processing?

Please select all that apply. (If you do not currently use the cloud, please base your response on which items are factors in your decision)



## About The Society of Actuaries

The Society of Actuaries (SOA), formed in 1949, is one of the largest actuarial professional organizations in the world dedicated to serving more than 32,000 actuarial members and the public in the United States, Canada and worldwide. In line with the SOA Vision Statement, actuaries act as business leaders who develop and use mathematical models to measure and manage risk in support of financial security for individuals, organizations and the public.

The SOA supports actuaries and advances knowledge through research and education. As part of its work, the SOA seeks to inform public policy development and public understanding through research. The SOA aspires to be a trusted source of objective, data-driven research and analysis with an actuarial perspective for its members, industry, policymakers and the public. This distinct perspective comes from the SOA as an association of actuaries, who have a rigorous formal education and direct experience as practitioners as they perform applied research. The SOA also welcomes the opportunity to partner with other organizations in our work where appropriate.

The SOA has a history of working with public policymakers and regulators in developing historical experience studies and projection techniques as well as individual reports on health care, retirement and other topics. The SOA's research is intended to aid the work of policymakers and regulators and follow certain core principles:

**Objectivity:** The SOA's research informs and provides analysis that can be relied upon by other individuals or organizations involved in public policy discussions. The SOA does not take advocacy positions or lobby specific policy proposals.

**Quality:** The SOA aspires to the highest ethical and quality standards in all of its research and analysis. Our research process is overseen by experienced actuaries and nonactuaries from a range of industry sectors and organizations. A rigorous peer-review process ensures the quality and integrity of our work.

**Relevance:** The SOA provides timely research on public policy issues. Our research advances actuarial knowledge while providing critical insights on key policy issues, and thereby provides value to stakeholders and decision makers.

**Quantification:** The SOA leverages the diverse skill sets of actuaries to provide research and findings that are driven by the best available data and methods. Actuaries use detailed modeling to analyze financial risk and provide distinct insight and quantification. Further, actuarial standards require transparency and the disclosure of the assumptions and analytic approach underlying the work.

Society of Actuaries  
475 N. Martingale Road, Suite 600  
Schaumburg, Illinois 60173  
[www.SOA.org](http://www.SOA.org)