PROJECT NO.

SENG7R16/4836

Leveraging Other Industries – Big Data Management (Phase I)

Leveraging Other Industries – Big Data Management (Phase I)

Prepared by:

Raja Kadiyala, Ph.D.
CH2M

Christopher Macintosh
CH2M

2018

THE Water Research FOUNDATION

The Water Research Foundation (WRF) is a nonprofit (501c3) organization which provides a unified source for One Water research and a strong presence in relationships with partner organizations, government and regulatory agencies, and Congress. The foundation conducts research in all areas of drinking water, wastewater, stormwater, and water reuse. The Water Research Foundation's research portfolio is valued at over $700 million.

The Foundation plays an important role in the translation and dissemination of applied research, technology demonstration, and education, through creation of research-based educational tools and technology exchange opportunities. WRF serves as a leader and model for collaboration across the water industry and its materials are used to inform policymakers and the public on the science, economic value, and environmental benefits of using and recovering resources found in water, as well as the feasibility of implementing new technologies.

For more information, contact:
The Water Research Foundation

Alexandria, VA Office
1199 North Fairfax Street, Suite 900
Alexandria, VA 22314-1445
Tel: 571.384.2100
www.werf.org
werf@werf.org

Denver, CO Office
6666 West Quincy Avenue
Denver, Colorado 80235-3098
Tel: 303.347.6100
www.waterrf.org
Info@WaterRF.org

WRF ISBN: 978-1-60573-354-8
WRF Project Number: SENG7R16/4836

Smart Water Networks (SWAN) Forum, River to Tap (R2T)

Cover image: Courtesy of CH2M, 2016.

Acknowledgments

Research Team

Principal Investigators:
Raja Kadiyala, Ph.D.
Jacobs

Project Team:
Chris Macintosh, M.B.A., B. Eng.
Jacobs

WRF Issue Area Team

John B. Barber, Ph.D.
Eastman Chemical Company

Christopher Chow, Ph.D.
University of South Australia

John McMahon
Metro Vancouver

Daniel Murray
U.S. EPA

Mary M. Smith
Water Research Foundation

Helen Stratton, Ph.D.
Griffith University

Lisa C. Thompson, Ph.D.
Sacramento Regional County Sanitation District

Ting Lu, Ph.D.
Clean Water Services

Research Advisory Council Liaison

John Willis, P.E., BCEE
Brown and Caldwell

Water Research Foundation Staff

John Albert, MPA
Chief Research Officer

Walter L. Graf, Jr.
Program Director

Abstract and Benefits

Abstract:

This research project examined the current state of knowledge and application of Big Data analytics and Internet of Things (IoT) technology in the water/wastewater industry using an online survey of water/wastewater utilities. A selection of companies from other industries outside the water/wastewater industry were surveyed for comparison. A literature search was then performed to locate guidance and applications from water/wastewater industry organizations and information technology research firms that can assist with better implementation of Big Data analytics and IoT technology in the water/wastewater industry. Finally, research was undertaken to determine trends and future technology paths for Big Data analytics and IoT technology. Research into published material suggests the guidance about implementation is similar across all industries. The results of the surveys show that the water industry is currently lagging in terms of implementation of Big Data analytics and IoT technology. As the guidance for implementation is not industry specific, the water industry can benefit from knowledge gained by other industries.

Benefits:

- Provides an understanding of the benefits of implementation of Big Data analytics and IoT technology.
- Identifies opportunities to implement Big Data analytics and IoT technology in water/wastewater utilities to increase operational efficiency and enhance capital planning.
- Identifies approaches to mitigate the typical impediments for water/wastewater utilities to implement Big Data analytics and IoT technology.
- Discusses future trends and technologies within the Big Data analytics and IoT technology space and their relevance to the water sector.
- Provides information that utilities can use to compare their Big Data analytics and IoT technology capabilities and maturity level to other utilities.
- Enhances regulatory compliance.

Keywords: Big Data, IoT, analytics, technology, wastewater, water.

Contents

Acronyms and Abbreviations

AI	Artificial intelligence
AMI	Advanced metering infrastructure
AWWA	American Water Works Association
CIO	Chief Information Officer
CIS	Customer information system
CMMS	Computerized maintenance management system
CSO	Combined Sewer Overflow
DCS	Distributed Control System
DNN	Deep Neural Network
EAMS	Enterprise asset management system
ETL	Extract, transform, load
GAN	Generative Adversarial Network
GIS	Geographic information system
GPU	Graphical processing unit
HBR	Harvard Business Review
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IMT	International Mobile Telecommunication
IoT	Internet of Things
ITU	International Telecommunications Union
LIMS	Laboratory information management system
LTE	Long-Term Evolution
LTE-A	LTE Advanced
N/A	Not applicable
NIST	National Institute of Standards and Technology
PLC	Programmable Logic Controller
R2T	River to Tap
SCADA	Supervisory control and data acquisition
SRS	Surveillance and response system
SSO	Sanitary Sewer Overflow
SWAN	Smart Water Networks Forum
SWMM	Stormwater Management Model
WEF	Water Environment Federation
WRF	The Water Research Foundation

Executive Summary

Internet of Things (IoT) technology and the need to process great amounts of data from multiple, often disparate sources, are becoming very important in all industries, including the water/wastewater industry. Data alone provides minimal value until it is processed to derive information and knowledge that leads to insight. With IoT and increasingly large, dissimilar data sets, the use of Big Data analytics is essential for cities and utilities to rapidly process these data sets to maximize gained value while minimizing impact to daily staff operations. Because data are so critical to business operations, collection, processing, and management must be performed in a cohesive system to be actionable and enable better decision making.

Utilities are becoming overwhelmed with data. CH2M recently queried Chief Information Officers (CIOs) from 50 of the largest water and wastewater utilities in the United States on the amount of data they use of the data they generate. A surprisingly low number – only 10% – is actually used. This poor result is primarily driven by the lack of Big Data analytics tools and staff resources. The inherent value in the other 90% of data not used (often referred to as dark data) is its significant potential to create new insights to improve operational efficiency and enhance regulatory compliance. Understanding how to harness Big Data using real-time analytics can provide utilities with a variety of benefits.

The overall goal of this project was to determine the current capabilities and state of knowledge of IoT and Big Data processing within the water industry and certain non-water sectors. The following tasks were undertaken to meet this goal:

- Task 1, Survey Water and Wastewater Utilities, included the development and administration of a survey to determine the current state of the use of IoT technology and Big Data analytics by water and wastewater utilities. Thirty utilities responded to this survey.
- Task 2, Survey Water Industry Organizations and IT Research Firms, was a literature search to identify approaches for the use of Big Data analytics and IoT technology specific to the water/wastewater industry and generally compare them to approaches in other industries.
- Task 3, Survey Large Firms in Non-Water Sectors, included the development and administration of a survey with questions similar to those in Task 1 to provide a comparison to the data gathered in Task 1. A small number of firms outside the water/wastewater industry responded.
- Task 4, Determine Trends and Future Technology Paths, was a literature search to identify trends for Big Data analytics and IoT technology using near-term product, early adopters, and research and development timeframes.
- Task 5, Prepare Final Report, was the impetus for this document.
- Task 6, Project Management, included completion of administrative and project-specific tasks, such as cost tracking and quality assurance.

Overall conclusions obtained from completion of the tasks described above were:

- Generally, the water/wastewater industry has not embraced Big Data analytics and IoT as rapidly as other industries.
- There is still a “wait and see” element within the water sector that is impacting adoption of these concepts and technologies. This concern will ease as more lighthouse examples are implemented.

- While utilities are collecting significant amounts of data, data quality is an issue that hampers data utilization.
- Generally, the technology solutions available address only single water industry uses. This leaves the onus on the utility to provide the integration capabilities when multiple applications are implemented.
- Future technologies are becoming available that will ease implementation of data analytics.

This final report includes a description of the approach for the project, a summary of the findings, and technical memorandums developed for each of the associated tasks (Appendices A through D).

CHAPTER 1

Task 1 – Survey Water and Wastewater Utilities

The Project Team designed a survey to determine the current use of Big Data analytics and Internet of Things (IoT) technology by water and wastewater utilities. The survey was administered online to 30 water and wastewater utilities, and results were collected and analyzed by the Project Team.

The survey was divided into the following sections:

- Introductory information – concepts and definitions to familiarize the respondent with Big Data technologies.
- Utility information (four questions).
- Status and future willingness to implement Big Data analytics (three questions).
- Understanding and adoption of Big Data analytics (11 questions).
- Impediments to implementation (one question).

Summary conclusions based on the survey responses are as follows:

- Challenges that utilities see of greatest concern, in order of importance, are:
  - Aging utility infrastructure.
  - The aging workforce.
  - Management of operational costs.
  - Data management.
  - Management of capital costs.
  - Justification of capital improvements and any associated rate changes that may be required.
  - Resiliency and reliability.
- Of the 30 respondents, 19 have implemented components to store and process Big Data.
- 80% of respondents plan to invest in Big Data in the near term.
- Most respondents consider Big Data analysis to be important, but that it is only one of many things to be addressed. Only one-third consider it to be a significant issue that needs to be addressed as a priority.
- Data currently produced are primarily used for operational decision making, with short-term operations and maintenance planning being next.
- The benefits of Big Data that are seen as important include:
  - Optimize the operation of treatment plants and networks.
  - Predict system and equipment failure.
  - Decrease expenses through operational cost efficiencies.
  - Improve workforce management.
  - Extract greater value from existing analytical tools.
  - Mitigate knowledge loss from an aging workforce.
  - Accelerate the speed with which new capabilities and service are deployed.
- Utilities generally consider that they have advanced skills for management of Big Data.

- Utilities feel they lack skills in real-time analysis of Big Data.
- Utilities believe that Big Data will have a dramatic impact on customer relationships, the way operations are organized, and making the business more data-focused.
- Impediments to the use of Big Data include:
  - Data quality.
  - Lack of talent to run Big Data processing and analytics on an ongoing basis.
  - Data security.
  - Lack of talent to implement Big Data.

The administered survey and detailed analysis of the results appear in Appendix A.

CHAPTER 2

Task 2 – Survey Water Industry Organizations and IT Research Firms

The Project Team reviewed papers and presentations from multiple sources covering the following:

- Guidance on Big Data analytics and IoT technology published by water industry organizations in the form of technical white papers, committees, presentations, and so forth.
- The use of Big Data analytics and IoT technology in non-water sectors as published by information technology (IT) research firms, periodicals, and other organizations.

Due to the evolving nature of this topic, very few peer-reviewed papers were available. The majority of the references were, therefore, reports, articles from periodicals, and presentation slides available on the websites of industry organizations and separately through research organizations.

Specific examples of applicable lessons identified in the literature include:

- A common theme in the interviews and in the literature, is that existing infrastructure should be monitored and assessed, focusing efforts on rectifying poor-quality data from data sources. The literature generally concludes that it is more important to have good quality data and the right data rather than having more data. One solution to this is to have some level of data quality analytics looking at the sensor data itself. Often times sensor issues are occurring, but go unnoticed. These analytics can be as simple as threshold analysis to as complex as noise and drift analysis.
- It is critical to ensure there is only one integrated platform managing the data. “Separate silos of data only create separate silos of insight.” This is in a slight juxtaposition to the current state of solutions of having to rely on Big Data and IoT applications from different vendors – there is no fully integrated solution available. This implies that an organization undertaking a Big Data and IoT project will need to accept the integration effort that will be required.
- A shift in organization, processes, and culture may be required to successfully make the most out of becoming a digitally focused organization. Business processes that fit in a setting with minimal digitization may become more of a roadblock in the execution of a workflow. Staff education and management may require different organizational alignment to make sure that the entity can properly maintain systems going forward.
- Establishing proper key performance indicators early on to be able to quantify value generation is important. Many IoT projects are technology driven, resulting in insufficient business end-user acceptance and an inability to measure return on investment properly with reference to business goals.

Water utilities interested in pursuing Big Data management to address top challenges can learn from not only the non-water sector but also innovators within the water sector. Overall, this work will be beneficial to all water utilities as technologies and platforms are constantly evolving and IoT is growing exponentially.

A detailed summary of references and findings appears in Appendix B.

CHAPTER 3

Task 3 – Survey Large Firms in Non-Water Sector

The Project Team developed a survey to determine how firms outside the water sector use Big Data analytics and IoT technology. The survey sought to understand whether investments in Big Data analytics and IoT technology are considered important and the primary applications for which they are used. The survey was based on the questions in the Task 1 survey to enable comparisons to be made. Firms known to the researchers were identified, and the survey was administered by personal contact to increase the response rate and completeness of the responses. During discussions, firms confirmed the value of the lessons gleaned from the literature, but were reluctant to add anything further, as they considered their processes to provide them a competitive edge.

The survey was divided into the following sections:

- Organizational information (three questions).
- Status and future willingness to implement Big Data analytics (four questions).
- Adoption of Big Data analytics and IoT technology (five questions).
- Results achieved by using Big Data analytics and IoT technology (three questions).
- Impediments to implementation (one question).

The survey had nine respondents covering multiple industries. Summary conclusions include the following:

- Components to store and process Big Data and IoT data were either put in place more than three years ago or not implemented, with IoT implemented in fewer cases than Big Data.
- Two-thirds of respondents plan to invest in Big Data analytics in the near term.
- None of the respondents considered Big Data analysis to be unimportant. Only one of the respondents considered it the single most important way to maintain operations and reduce costs. Four considered it a top-five issue, while three considered it important but only one of many things to be addressed.
- Two-thirds of respondents use Big Data analytics and IoT technology in the production and research and development departments.
- Monitoring and sensing physical quantity and quality, and targeting customers are the primary functions for which Big Data analytics and IoT technology are used.
- Respondents considered that they had advanced skills in data engineering, statistical and quantitative analysis, and stream processing, with few skills in R programming and Scala programming.
- Organizations have most often seen measurable results from the use of Big Data and IoT technology when establishing a data-driven culture.
- Impediments to the use of Big Data include data security and data quality.

The administered survey and detailed analysis of the results appear in Appendix C.

CHAPTER 4

Task 4 – Determine Trends and Future Technology Paths

The project team reviewed research papers from firms with large research budgets, IT research organizations, and university research groups to identify trends for IoT technology and Big Data analytics.

This research applied the three horizon views as a means of identifying the different stages of maturation of technologies that are considered important for the water industry. The technologies identified are as follows:

- Horizon 1: Near-term products of interest include 5G cellular networking, automated machine learning, and Apache Spark.
- Horizon 2: Technologies in the early adoption phase include deep learning and machine learning for cybersecurity.
- Horizon 3: Technologies in the research and development phase include generative adversarial networks and quantum computing.

It is expected that these technologies will be used in the future by the water industry, and preparations for their implementation should be considered.

CHAPTER 5

Conclusions

Task 1 surveyed the water and wastewater industry, and Task 3 similarly surveyed organizations in other industries. Both survey sample sets were small, with 30 respondents for Task 1, and nine respondents for Task 3. When comparing the results of the two surveys, the following observations were made:

- Have not implemented components to store and process Big Data:
  - Water industry – 11 of the 30 respondents (36%).
  - Other industries – two of the nine respondents (22%).
- Plan to have, or increase, Big Data investments in the near term:
  - Water industry – 24 of the 30 respondents (80%).
  - Other industries – six of the nine respondents (66%).
- Importance of Big Data analysis (seen as most important or a top five issue):
  - Water industry – 10 of 30 respondents (33%).
  - Other industries – five of the nine respondents (55%).
- Most influential impediments to implementing Big Data (in order):
  - Water industry – data quality, lack of talent to run Big Data processing and analytics on an ongoing basis, data security, and lack of talent to implement Big Data.
  - Other industries – data security and data quality.

Although it is recognized that the size of both sample sets is small, these results indicate that the water industry is lagging when it comes to investing in Big Data analytics when compared to other industries, with a lower percentage of the water industry having currently made a significant investment, and with the expectation of investments in the near future. The water industry recognizes similar challenges and impediments to other industries, although, based on the survey responses, the water industry does not yet see Big Data analysis as important as do other industries.

Lessons learned from the use of Big Data analytics and IoT technology by a variety of industries can be applied to the water industry. Key fundamentals, foundations, and processes can be leveraged to enhance the success of this effort. Water utilities interested in pursuing Big Data management to address top challenges can learn from not only the non-water sector, but also innovators within the water sector. Overall, this work will be beneficial to all water utilities as technologies and platforms are constantly evolving and the use of IoT-related technology is projected to grow exponentially.

APPENDIX A

Task 1 – Survey Water and Wastewater Utilities

Leveraging Other Industries – Big Data Management

Task 1 – Survey Water and Wastewater Utilities

PREPARED FOR:	WE&RF
PREPARED BY:	CH2M
DATE:	November 19, 2017
PROJECT NUMBER:	692034 (SENG7R16)
REVISION NO.:	Draft

As part of the Big Data Management project effort, the initial task involved establishing a baseline for the water sector. *Task 1 – Survey Water and Wastewater Utilities* included the design of a survey to determine the current state of the use of IoT and big data related technology. The survey focused on providing quantifiable measures of the level of use of Big Data related technologies within the water sector. The survey was administered by the SWAN Forum using the Survey Gizmo online survey service. This Technical Memorandum (TM) provides a summary of the process and results of the survey.

Approach

A set of survey questions was developed to gather the information about the current state of the use of IoT and big data related technologies in water and wastewater utilities. These questions are presented in Attachment A.

The survey was divided into the following sections:

- Introductory information – concepts and definitions to familiarize the respondent with the Big Data technologies
- Utility information (4 questions)
- Status and future willingness of Big Data implementation (3 questions)
- Big data understanding and adoption (11 questions)
- Impediment factors (1 question)

The details and analysis of the survey results are presented in this document under the sections of the survey.

A list of potential water and wastewater utilities was generated to provide guidance for who would be contacted to undertake the survey. This list of targeted respondents was initially generated from contacts of CH2M, WE&RF, and SWAN.

The survey content and potential list of respondents were reviewed by the WE&RF Project Steering Committee, who provided comments that were incorporated prior to administering the survey.

Survey Results

Discussion of the survey results is presented under the topics used for the survey.

Introductory Information

General information was sought from the utilities participating in the survey including name, location, city, and country. Of the 30 respondents, 23 were US based, with three respondents from Europe, one from South America, two from South-East Asia, and one from Australia.

Is your utility a public or a private entity?

Utilities were asked whether they were public or private entities. The results presented in Figure 1 show that 25 of the 30 respondents where public entities, with five private entities.

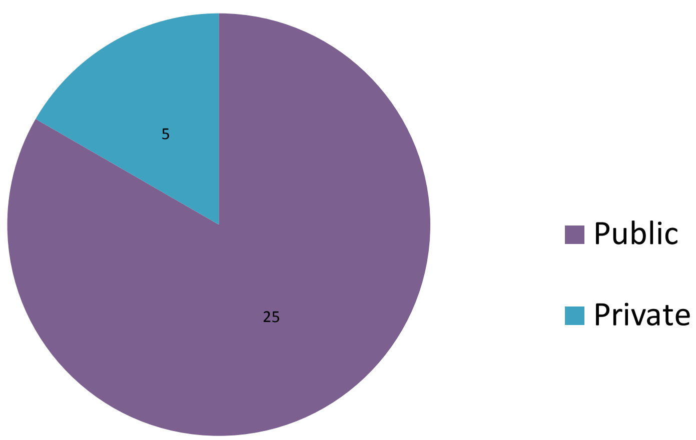

Figure 1. Public vs Private Utilities

What types of facilities does your utility currently manage?

This question asked for all the types of facilities that the utility managed, with the results presented in Figure 2. Most utilities managed more than one type of facility, the detailed breakdown is:

- One utility is drinking water only
- Six utilities are wastewater only
- Four utilities include all four types of facilities
- All utilities with stormwater or water reuse facilities also have wastewater facilities

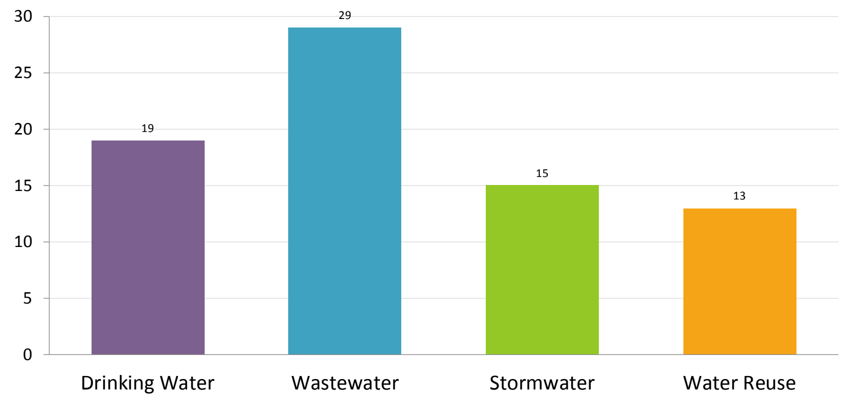

Figure 2. Facilities Managed

What population does your utility serve?

The population served for the responding utilities varied, with a bias towards large utilities as shown in Figure 3.

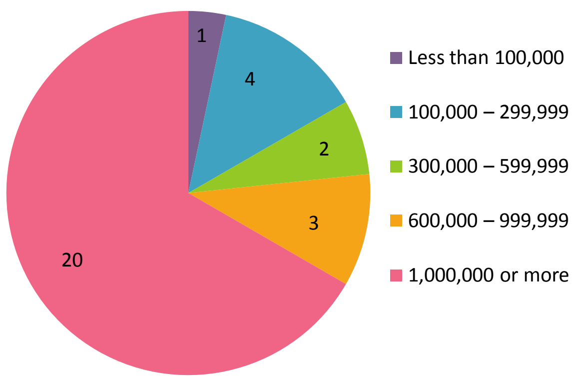

Figure 3. Population Served

Please rate current situation of each of the following challenges to your utility

A rating of 1-5 was used, with 1 = not a challenge to our utility; 5 = challenge that requires immediate attention/solution. The challenges to be rated were:

- Aging of utility infrastructure
- Managing capital costs
- Managing operational costs
- Justifying improvements/rate requirements
- Resilience/Reliability
- IT infrastructure (servers, network, storage)
- Data management (databases, visualization and analysis tools)
- Industrial control systems (SCADA, PLCs, DCS)
- Aging workforce
- Treatment technology
- Cybersecurity threats
- Recruiting employees for specialized fields
- Water conservation

- Political will to establish sustainable rates
- Availability of funding
- Water scarcity or availability
- Water loss (Non-revenue water)
- Cross-connections or redundancy
- Meeting treated water discharge regulations
- SSO and/or CSO occurrences within the system
- Customer satisfaction and raising awareness
- Managing stormwater runoff with green stormwater infrastructure

Results obtained using the 1-5 scale range created somewhat dispersed results that were difficult to cluster. The ratings were therefore adjusted to a Low/Medium/High scale where 1 and 2 were combined to be "Low", 3 became "Medium", and 4 and 5 were combined to be "High".

Results using the Low/Medium/High rating scale are presented in Table 1. The cells highlighted in yellow are those representing greater than half of the survey respondents in each group (i.e., >=15). The results indicate the following:

- Aging of utility infrastructure is the greatest concern
- The aging workforce is the second greatest concern
- Other issues of concern are:
    - Managing operational costs
    - Data management
    - Managing capital costs
    - Justifying improvement and rate requirements
    - Resiliency and reliability
- Of least concern are:
    - Water conservation
    - Water scarcity or availability
    - Cross connections or redundancy

Table 1. Current Situation Summary

	Low	Med	High
Aging of utility infrastructure	0	6	24
Managing capital costs	2	12	16
Managing operational costs	3	10	17
Justifying improvements/rate requirements	3	11	16
Resilience/Reliability	5	9	15
IT infrastructure (servers, network, storage)	5	12	13
Data management (databases, visualization and analysis tools)	7	6	17
Industrial control systems (SCADA, PLCs, DCS)	6	11	13
Aging workforce	3	5	22

Treatment technology	12	11	7
Water conservation	15	5	10
Political will to establish sustainable rates	6	14	10
Availability of funding	9	11	10
Water scarcity or availability	16	5	9
Water loss (Non-revenue water)	11	11	7
Cross-connections or redundancy	15	13	1
Meeting treated water discharge regulations	13	9	8
SSO and/or CSO occurrences within the system	8	10	12
Customer satisfaction and raising awareness	5	11	14
Managing stormwater runoff with green stormwater infrastructure	13	12	5

Status and future willingness of Big Data implementation

This section of the survey focused on what components have been implemented to store and process big data, and what near-term plans the utility may have. Results by utility size are also included to show trends for smaller and midsize utilities versus the larger utilities. These results are displayed as percentages on the right chart for each figure.

Has your utility implemented components to store and process Big Data?

If YES, when did you implement it?

The results are presented in Figure 4. Eleven respondents haven't implemented components to store and process Big Data. Of the 19 who have implemented these components, nine were in the last year. One third of respondents have had systems to store and process Big Data for more than one year.

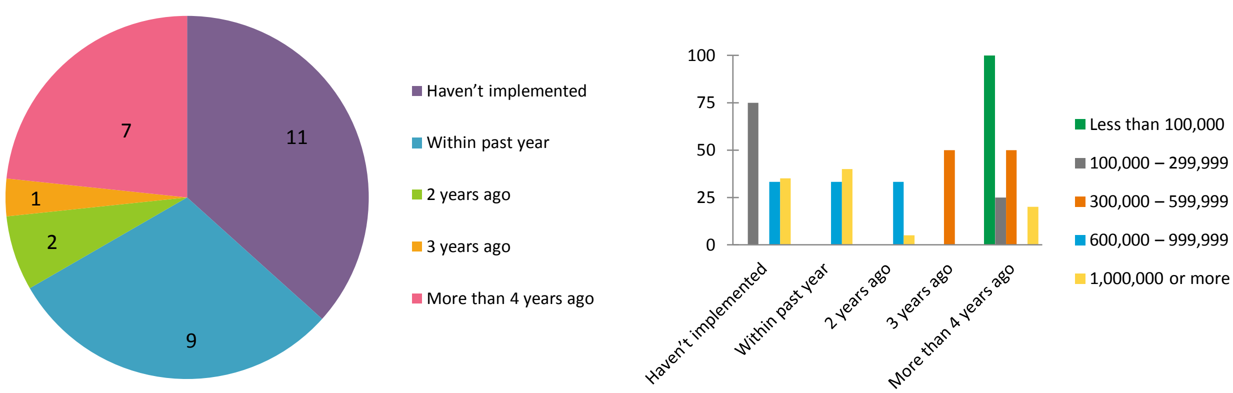

Figure 4. Implementation of Components to Store and Process Big Data

Does your utility plan to have or increase Big Data investments in the near term?

Figure 5 presents the results which indicate that 80% of respondent intend to invest in Big Data in the near term.

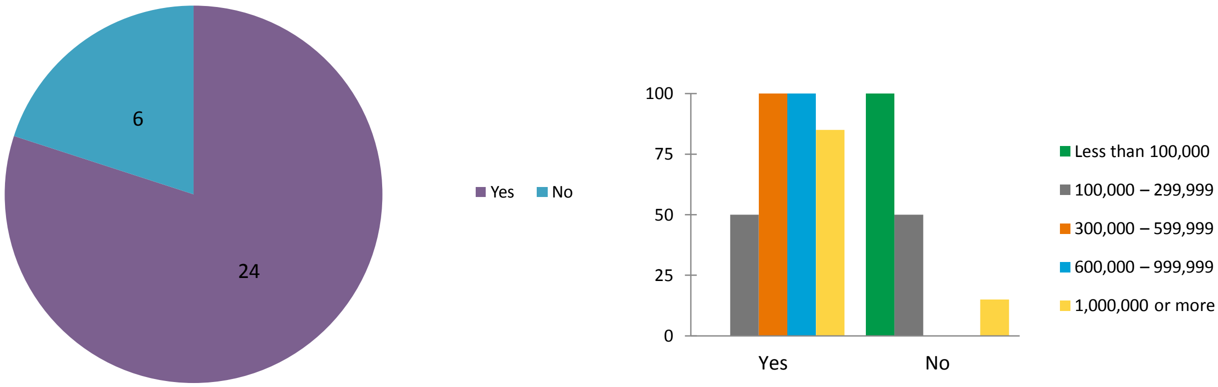

Figure 5. Plans for Big Data Investments

If YES, when do you see your utility will have or increase investment on Big Data?

Of the 24 respondents, half are expecting to invest in the next year, while 11 are planning to invest in 2-3 years. The one respondent planning to invest in 4-5 years from now has invested in the past year.

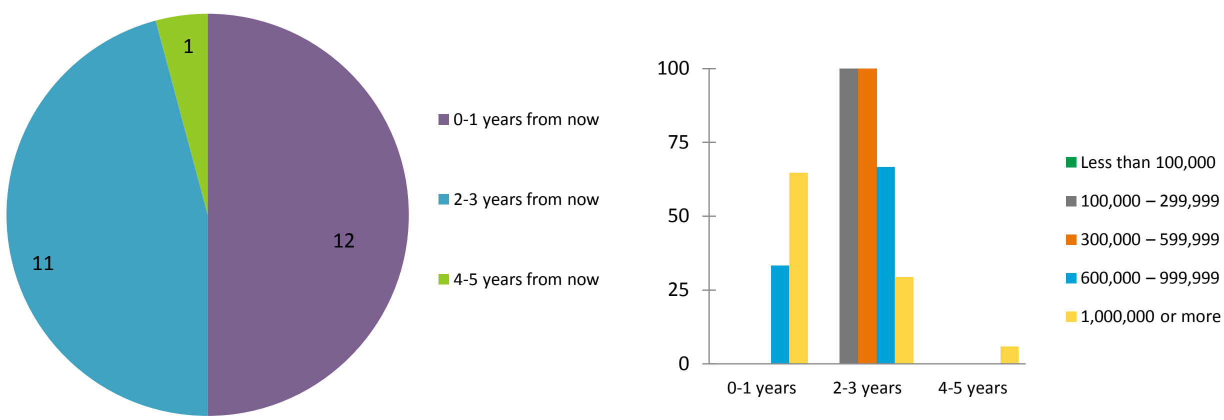

Figure 6. When Investment is Expected

If NO, what's the reason for not investing in Big Data?

Figure 7 indicates that the greatest reason for not investing in Big Data is that there is no need identified to justify the investment. The second greatest reason is availability of budget.

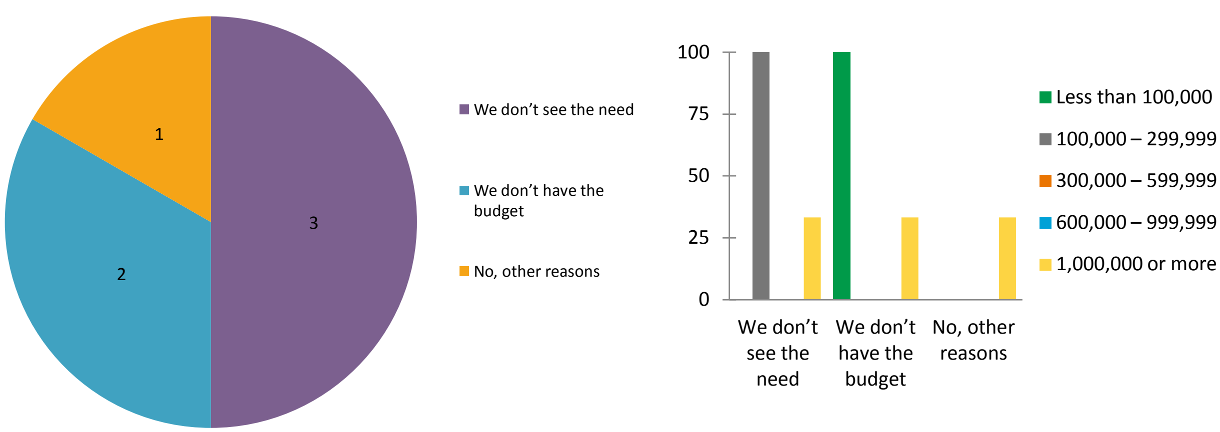

Figure 7. Reasons for Not Investing

How important is Big Data analysis to your utility?

Options were:

- It is the single most important way for us to maintain operations and reduce costs
- Is a top 5 issue that gets significant time and attention from top leadership
- Is important but is only one of many other challenges/opportunities that we need to address
- Not very important
- Don't know/NA

The results in Figure 8 show that the majority of respondents consider Big Data analysis to be important, but that it is only one of many things to be addressed. Only one third consider it to be a significant issue that needs addressing as a priority.

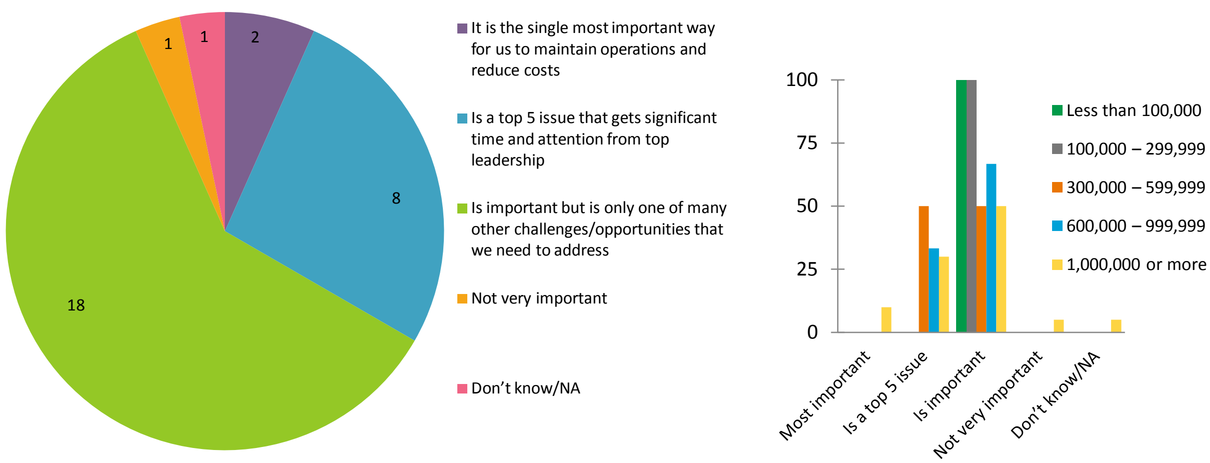

Figure 8. Importance of Big Data Analysis

Big data understanding and adoption

The questions for this section of the survey sought to understand the sources of Big Data currently existing within the utility and the use of the data.

Which of the following do you consider as part of Big Data?

This question asked to select all that apply from the following list:

- Large data files (20 terabytes or larger)
- Advanced analytics or analysis
- Advanced visualization tools
- Data from social networks
- Unstructured data (e.g., video, open text, voice)
- Geospatial/location information
- Sensor data
- Telematics/vehicular data

Results are presented in Figure 9.

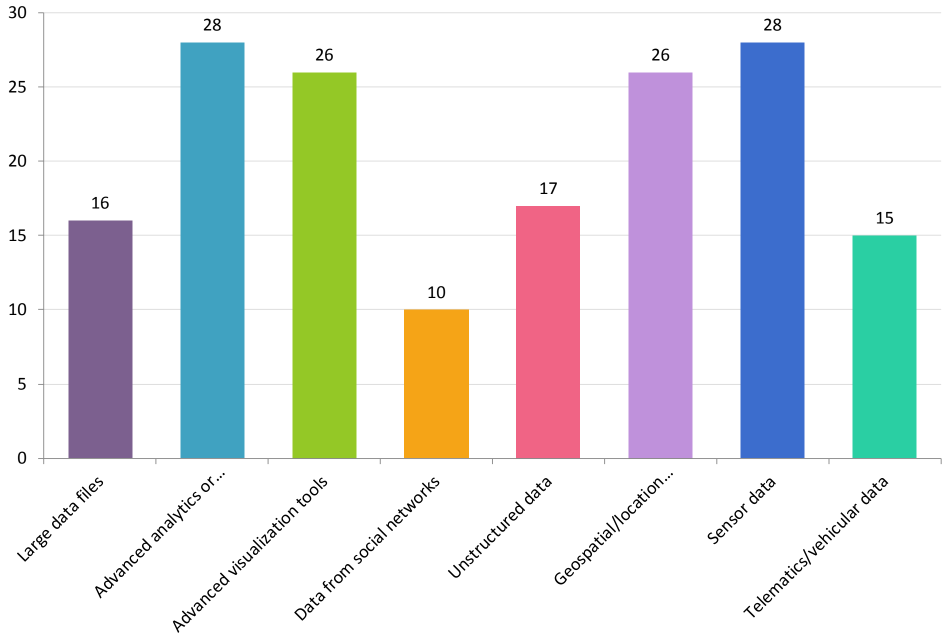

Figure 9. Topics Considered Part of Big Data

What system(s) below have you already implemented in your utility?

This question asked to select all that apply from the following list:

- Advanced metering infrastructure (AMI)
- Customer information system (CIS)
- Computerized maintenance management systems (CMMS)
- Laboratory information management systems (LIMS)
- Supervisory control and data acquisition (SCADA) system
- Enterprise asset management system
- Surveillance and Reponses System (SRS)

Results are presented in Figure 10. Some of these systems didn’t pertain to the particular utility, for example a wastewater only utility would not have an AMI or SRS.

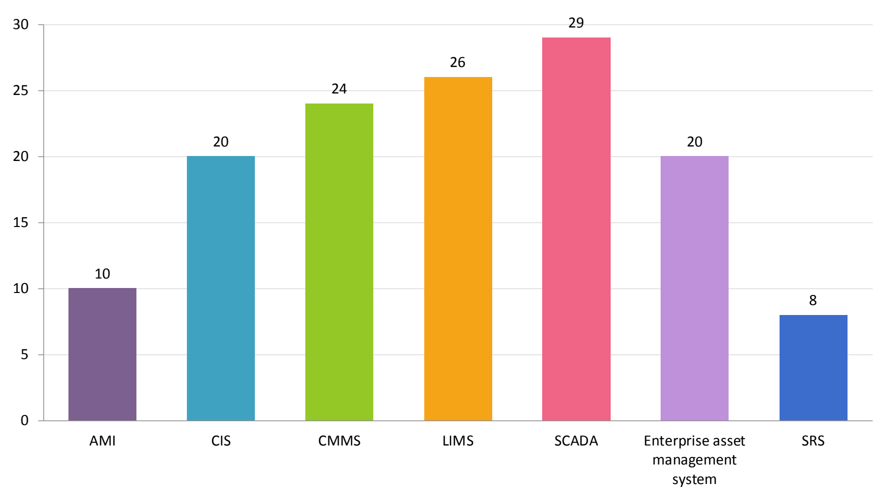

Figure 10. Systems Already Implemented

Which system(s) does your utility use on a daily basis?

This question asked to select all that apply from the following list:

- Advanced metering infrastructure (AMI)
- Customer information system (CIS)
- Computerized maintenance management systems (CMMS)
- Laboratory information management systems (LIMS)
- Supervisory control and data acquisition (SCADA) system
- Enterprise asset management system
- Surveillance and Reponses System (SRS)

Results are presented in Figure 11.

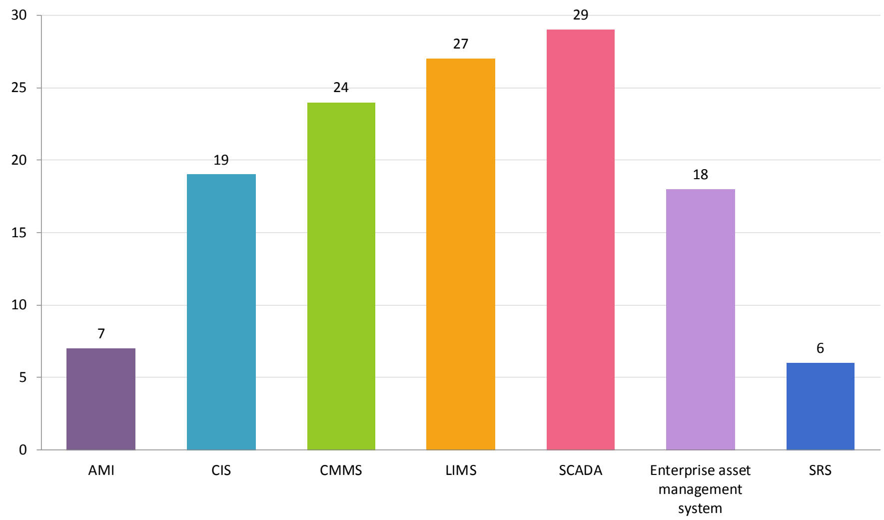

Figure 11. Systems Used on a Daily Basis

How important are the below systems to your utility?

Ratings were 1-5, where 1 = not important; 5 = very important. The systems to be rated were:

- Advanced metering infrastructure (AMI)
- Customer information system (CIS)
- Computerized maintenance management systems (CMMS)
- Laboratory information management systems (LIMS)
- Supervisory control and data acquisition (SCADA) system
- Enterprise asset management system
- Surveillance and Reponses System (SRS)

Results obtained using the 1-5 scale ratings didn't provide significant clustering of patterns. The ratings were therefore adjusted to a Low/Medium/High scale where 1 and 2 were combined to be "Low", 3 was made "Medium", and 4 and 5 were combined to be "High".

Results using the Low/Medium/High rating scale are presented in Table 2. The cells highlighted in yellow are those representing greater than half of the survey respondents in each group (i.e., >=15). This question also provided the option to indicate "N/A" which was used by most respondents where they also indicated that they currently didn't have such a system.

The results indicate the following systems are considered important:

- Supervisory control and data acquisition (SCADA) system
- Computerized maintenance management systems (CMMS)
- Laboratory information management systems (LIMS)
- Enterprise asset management system
- Customer information system (CIS)

Table 2. Importance of Certain Utility Systems

	N/A	Low	Med	High
Advanced metering infrastructure (AMI)	12	2	3	12
Customer information system (CIS)	5	4	0	20
Computerized maintenance management systems (CMMS)	1	1	1	27
Laboratory information management systems (LIMS)	2	0	2	26
Supervisory control and data acquisition (SCADA) system	0	0	0	30
Enterprise asset management system	3	2	4	21
Surveillance and Reponses System (SRS)	11	3	7	8

What system(s) below is your utility most likely to implement, upgrade or extend in the near future?

This question asked to select all that apply from the following list:

- Advanced metering infrastructure (AMI)
- Customer information system (CIS)
- Computerized maintenance management systems (CMMS)
- Laboratory information management systems (LIMS)
- Supervisory control and data acquisition (SCADA) system
- Enterprise asset management system
- Surveillance and Reponses System (SRS)

The results are presented in Figure 12. Twenty-seven of the respondents identified systems that they might implement, upgrade, or extend in the near future with the primary focus on asset management systems.

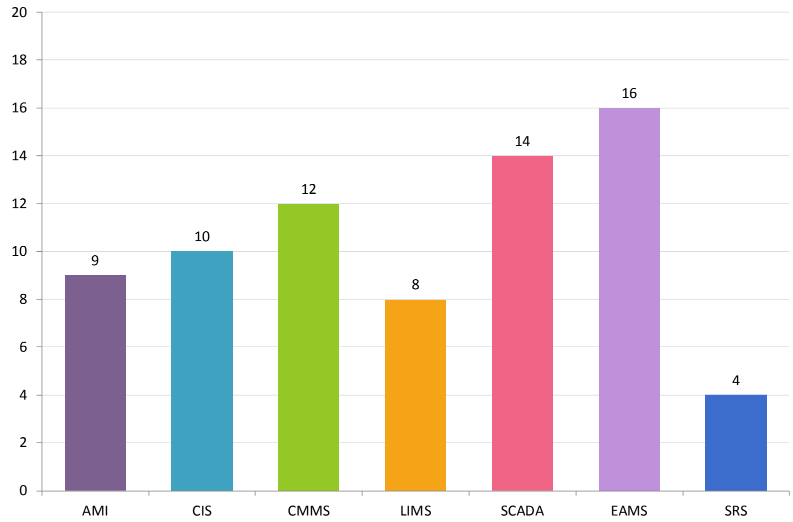

Figure 12. Systems Most Likely to be Implemented or Upgraded

What does the information from these systems get used for?

This question asked to select all that apply from the following list.

- Operational decision making
- Short-term operations planning
- Short-term maintenance planning
- Long-term operations planning
- Long-term maintenance planning
- Capital expenditure planning
- Long-term asset management
- Other

The results presented in Figure 13 indicate that all respondents use the data for operational decision making, with short-term operations and maintenance planning being next. The category "other" included regulatory compliance and billing.

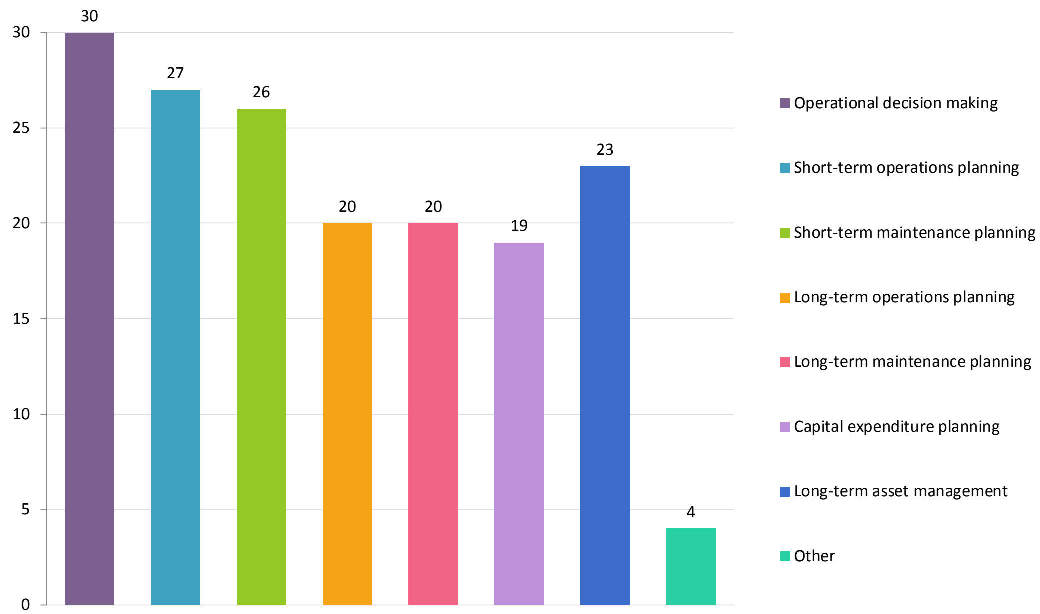

Figure 13. Information Use

Please rate the importance of each of the following benefits that Big Data analysis will bring to your utility?

Ratings were 1-5, where 1 = not important; 5 = very important. The benefits to be rated were:

- Optimal operation of treatment plants and networks
- Predict system and equipment failure
- Accelerate the speed with which new capabilities and service are deployed
- Decrease expenses through operational cost efficiencies
- Mitigate knowledge loss from aging workforce
- Improve workforce management
- Extract greater value from existing analytical tools
- Reduce non-revenue water to minimize water and revenue losses
- Reduce pollution events

As in the previous ratings based analysis, ratings were adjusted to a Low/Medium/High scale where 1 and 2 were combined to be "Low", 3 was made "Medium", and 4 and 5 were combined to be "High".

Results using the Low/Medium/High rating scale are presented in Table 3. The cells highlighted in yellow are those representing greater than half of the survey respondents in each group (i.e., >=15). The results indicate the following:

- The most important benefit identified was optimal operation of treatment plants and networks
- Next most important included the following three benefits:
  - Predict system and equipment failure
  - Decrease expenses through operational cost efficiencies
  - Improve workforce management
- Other benefits considered important include

- Extract greater value from existing analytical tools
- Mitigate knowledge loss from aging workforce
- Accelerate the speed with which new capabilities and service are deployed

Table 3. Importance of the Benefits of Big Data Analysis

	Low	Med	High
Optimal operation of treatment plants and networks	3	2	25
Predict system and equipment failure	1	5	24
Accelerate the speed with which new capabilities and service are deployed	5	10	15
Decrease expenses through operational cost efficiencies	1	5	24
Mitigate knowledge loss from aging workforce	1	9	20
Improve workforce management	2	4	24
Extract greater value from existing analytical tools	3	5	22
Reduce non-revenue water to minimize water and revenue losses	12	3	14
Reduce pollution events	8	9	13

Please rate the level of skills available within your utility for each of the following

Ratings were 1-5, where 1 = no skills available; 5 = advanced skills available. The skills to be rated were:

- Management (storage, indexing and retrieval) of Big Data
- Off-line analysis of Big Data
- Real-time analysis of Big Data
- Maintenance of systems to manage and analyze Big Data

The ratings were adjusted to a Low/Medium/High scale where 1 and 2 were combined to be “Low”, 3 was made “Medium”, and 4 and 5 were combined to be “High”.

Results using the Low/Medium/High rating scale are presented in Table 4. The cells highlighted in yellow are those representing greater than half of the survey respondents in each group (i.e., >=15). The results indicate the following:

- Utilities generally consider that they have advanced skills for management of Big Data
- Utilities feel they lack skills in real-time analysis of Big Data

Table 4. Skills Available

	Low	Med	High
Management (storage, indexing and retrieval) of Big Data	5	10	15
Off-line analysis of Big Data	13	8	9
Real-time analysis of Big Data	16	10	4
Maintenance of systems to manage and analyze Big Data	8	11	11

Please rate the level of Big Data's impact on each of the following areas of your utility in the next five years

Ratings were 1-5, where 1 = no impact; 5 = will have a dramatic impact. The areas to be rated were:

- Impacting customer relationships
- Changing the way we organize operations
- Making the business more data-focused

The ratings were adjusted to a Low/Medium/High scale where 1 and 2 were combined to be "Low", 3 was made "Medium", and 4 and 5 were combined to be "High".

Results using the Low/Medium/High rating scale are presented in Table 5. The cells highlighted in yellow are those representing greater than half of the survey respondents in each group (i.e., >=15). The results indicate that utilities believe that Big Data will have a dramatic impact on customer relationships, the way operations are organized, and making the business more data-focused.

Table 5. Impact of Big Data

	Low	Med	High
Impacting customer relationships	5	6	19
Changing the way we organize operations	4	3	23
Making the business more-data-focused	3	4	23

What do you see as the water utility's role in developing "smart cities"?

This question asked to select from the following list:

- Integral part of the process from planning to capital funding and spending
- Advisory/consultative
- Observe and report back to stakeholders/board
- Don't know

The results presented in Figure 14 indicate that utilities believe they have an integral role in the process of developing smart cities.

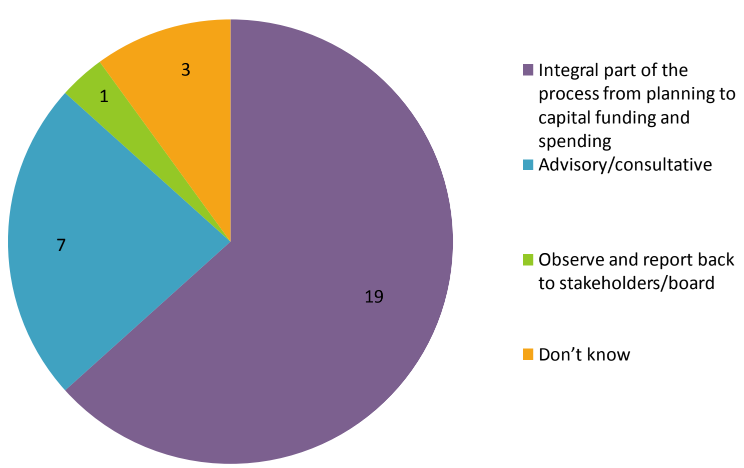

Figure 14. Water Utility's role in Developing Smart Cities

If you have implemented Big Data analysis, can you share one of your Big Data projects as an example?

Below is a list of examples provided by the respondents. Most of the examples included the analysis of time series data, with some respondents doing the analysis in an automated workflow.

1. Energy efficiency project - partnership with Cascade Energy and Envirosim. Data sources were the SCADA (Proficy iHistorian), dataset for electricity submetering data, and process data and LIMS database for lab data. Analysis done with Hach WIMS to integrate iHistorian and LIMS data and for some data analysis/visualization. Biowin for wastewater process modeling and Excel for data analysis and visualization.
2. ILPM - a product for looking at burst and leakage data to focus resource deployment. Comes from Schneider Electric.
3. Implemented Maximo, GIS, and NCIS
4. Predicting SR Bacti failure using R for analysis and Tableau for presentation. Pretty manual extraction process from data sources.
5. Reporting of LIMS and OSI-PI data for operational and regulatory reporting using up to 5 years of data at 11 treatment plants.
6. Using SWMM LIVE, InfoMaster, stormwater realtime control, Telog, and Lucity.
7. The City of Atlanta implemented a Smart Water Analytics Platform.
8. Water Consumption Portal project. Data extracted to Oracle-based data cubes, visualized using Jasper analytics.
9. We're implementing a water integration management system project that aims to use Big Data in order to support decisions in operations.

Impediment factors

This section had one question about impediments to adoption of Big Data analysis within a utility. The survey question is provided in Attachment A, with the results presented and discussed below.

Please rate the level of influence of each of the following impediment factors have on Big Data analysis adoption in your utility

Ratings were 1-5, where 1 = not influential; 5 = most influential. The factors to be rated were:

- Data security
- Data quality
- Lack of budget
- Lack of talent to implement Big Data
- Lack of talent to run Big Data processing and analytics on an ongoing basis
- Resistance to integrate existing systems
- Procurement limitations on Big Data vendors
- Lack of middle management adoption and understanding
- Lack of data governance policies and practices

The ratings were adjusted to a Low/Medium/High scale where 1 and 2 were combined to be “Low”, 3 was made “Medium”, and 4 and 5 were combined to be “High”.

Results using the Low/Medium/High rating scale are presented in Table 6. The cells highlighted in yellow are those representing greater than half of the survey respondents in each group (i.e., >=15). The results indicate the following:

- Impediments in order of most influential are:
    - Data quality
    - Lack of talent to run Big Data processing and analytics on an ongoing basis
    - Data security
    - Lack of talent to implement Big Data
- Procurement limitations on Big Data vendors are not seen as influential
- Lack of middle management adoption and understanding may have medium influence

Table 6. Impediment Factors to Big Data Adoption

	Low	Med	High
Data security	5	8	17
Data quality	0	8	22
Lack of budget	7	12	11
Lack of talent to implement Big Data	5	8	17
Lack of talent to run Big Data processing and analytics on an ongoing basis	5	4	21
Resistance to integrate existing systems	6	12	12
Procurement limitations on Big Data vendors	15	12	3
Lack of middle management adoption and understanding	5	15	10
Lack of data governance policies and practices	7	14	9

Summary/Conclusions

The survey had 30 respondents across the water and wastewater sector. Summary conclusions include:

- Challenges that utilities see of greatest concern are:
    - Aging of utility infrastructure is the greatest concern
    - The aging workforce is the second greatest concern
    - Managing operational costs
    - Data management
    - Managing capital costs
    - Justifying improvement and rate requirements
    - Resiliency and reliability
- Of the 30 respondents, 19 have implemented components to store and process Big Data
- 80% of respondents plan to invest in Big Data in the near term
- The majority of respondents consider Big Data analysis to be important, but that it is only one of many things to be addressed. Only one third consider it to be a significant issue that needs addressing as a priority.

- Data currently produced are primarily used for operational decision making, with short-term operations and maintenance planning being next.
- The benefits of Big Data seen as important include:
    - Optimal operation of treatment plants and networks
    - Predict system and equipment failure
    - Decrease expenses through operational cost efficiencies
    - Improve workforce management
    - Extract greater value from existing analytical tools
    - Mitigate knowledge loss from aging workforce
    - Accelerate the speed with which new capabilities and service are deployed
- Utilities generally consider that they have advanced skills for management of Big Data
- Utilities feel they lack skills in real-time analysis of Big Data
- Utilities believe that Big Data will have a dramatic impact on customer relationships, the way operations are organized, and making the business more data-focused.
- Impediments to the use of Big Data include:
    - Data quality
    - Lack of talent to run Big Data processing and analytics on an ongoing basis
    - Data security
    - Lack of talent to implement Big Data

Attachments:

A. Survey Questions for Wastewater Utilities

Big Data Management Survey - Utilities

Water Environment and Reuse Foundation (WE&RF)/CH2M Study

Nearly every discipline from sports and advertising to public health and science rely on data-driven analysis for decision-making. Tag-lined the “Age of Big Data,” we are becoming more and more reliant on data-driven evidence and analysis for nearly every decision we make. Data is not only becoming more available to the general public, but also more understandable, thanks to increased compute resources and advanced algorithms for analytics.

Even utilities are finding more advanced, efficient methods of sensing and managing data. These range from sensor networks and analytics to aid in system resiliency and tackle wet weather challenges to helping reduce non-revenue water issues on the distribution side. By leveraging all the data generated at a utility (a truly Big Data set), the utility is not only able to provide rapid detection and response to anomalous events, but also valuable insight into the asset performance and both the collection system and distribution network.

As you progress through the survey, it is useful to frame your responses to the challenges facing a utility and the core meaning of Big Data (versus *a lot of data*). As defined by the National Institute of Standards and Technology in NIST SP 1500-1, Big Data refers to “the inability of traditional data architectures to efficiently handle the new datasets. Characteristics of Big Data that force new architectures are:

- Volume (i.e., the size of the dataset);
- Variety (i.e., data from multiple repositories, domains, or types);
- Velocity (i.e., rate of flow); and
- Variability (i.e., the change in other characteristics).”

In summary, Big Data consists of:

extensive datasets -- primarily in the characteristics of volume, variety, velocity, and/or variability -- that require a scalable architecture for efficient storage, manipulation, and analysis.

Introductory Information

All responses will be kept confidential and presented only on an aggregated basis. * Denotes a mandatory field.

Your Name

Your Title

Utility Name *

Utility City, State (if applicable) *

Utility Country *

1. **Is your utility a public or a private entity?** *

- Public
- Private

2. **What types of facilities does your utility currently manage? Please select all that apply.** *

- ☐ Drinking Water
- ☐ Wastewater
- ☐ Stormwater
- ☐ Water Reuse

3. **What population does your utility serve?** *

- ○ Less than 100,000
- ○ 100,000 – 299,999
- ○ 300,000 – 599,999
- ○ 600,000 – 999,999
- ○ 1,000,000 or more

4. **Please rate current situation of each of the following challenges to your utility (1-5, 1 = not a challenge to our utility; 5 = challenge that requires immediate attention/solution).** *

	1	2	3	4	5
Aging of utility infrastructure	○	○	○	○	○
Managing capital costs	○	○	○	○	○
Managing operational costs	○	○	○	○	○
Justifying improvements/rate requirements	○	○	○	○	○
Resilience/Reliability	○	○	○	○	○
IT infrastructure (servers, network, storage)	○	○	○	○	○
Data management (databases, visualization and analysis tools)	○	○	○	○	○
Industrial control systems (SCADA, PLCs, DCS)	○	○	○	○	○
Aging workforce	○	○	○	○	○
Treatment technology	○	○	○	○	○
Cybersecurity threats	○	○	○	○	○
Recruiting employees for specialized fields	○	○	○	○	○
Water conservation	○	○	○	○	○
Political will to establish sustainable rates	○	○	○	○	○
Availability of funding	○	○	○	○	○
Water scarcity or availability	○	○	○	○	○
Water loss (Non-revenue water)	○	○	○	○	○
Cross-connections or redundancy	○	○	○	○	○
Meeting treated water discharge regulations	○	○	○	○	○
SSO and/or CSO occurrences within the system	○	○	○	○	○
Customer satisfaction and raising awareness	○	○	○	○	○
Managing stormwater runoff with green stormwater infrastructure	○	○	○	○	○

Status & Future Willingness of Big Data Implementation

5. **Has your utility implemented components to store and process Big Data? If YES, when did you implement it?** *

- Haven't implemented
- Within past year
- 2 years ago
- 3 years ago
- More than 4 years ago

6. **Does your utility plan to have or increase Big Data investments in the near term?**

- Yes
- No

If YES, when do you see your utility will have or increase investment on Big Data?

- 0-1 years from now
- 2-3 years from now
- 4-5 years from now
- More than 5 years from now

If NO, what's the reason for not investing in Big Data?

- We don't see the need
- We don't have the budget
- Insufficient internal capabilities
- No, other reasons

7. **How important is Big Data analysis to your utility?**

- It is the single most important way for us to maintain operations and reduce costs
- Is a top 5 issue that gets significant time and attention from top leadership
- Is important but is only one of many other challenges/opportunities that we need to address
- Not very important
- Don't know/NA

Big Data Understanding & Adoption

8. **Which of the following do you consider as part of Big Data? Please select all that apply.**

- ☐ Large data files (20 terabytes or larger)
- ☐ Advanced analytics or analysis
- ☐ Advanced visualization tools
- ☐ Data from social networks
- ☐ Unstructured data (e.g., video, open text, voice)
- ☐ Geospatial/location information
- ☐ Sensor data
- ☐ Telematics/vehicular data

9. **What system(s) below have you already implemented in your utility? Please select all that apply.**
(SRS definition link: EPA SRS Introduction)

- ☐ Advanced metering infrastructure (AMI)
- ☐ Customer information system (CIS)
- ☐ Computerized maintenance management systems (CMMS)
- ☐ Laboratory information management systems (LIMS)
- ☐ Supervisory control and data acquisition (SCADA) system
- ☐ Enterprise asset management system
- ☐ Surveillance and Reponses System (SRS)

10. **Which system(s) does your utility use on a daily basis? Please select all that apply.**

- ☐ Advanced metering infrastructure (AMI)
- ☐ Customer information system (CIS)
- ☐ Computerized maintenance management systems (CMMS)
- ☐ Laboratory information management systems (LIMS)
- ☐ Supervisory control and data acquisition (SCADA) system
- ☐ Enterprise asset management system
- ☐ Surveillance and Reponses System (SRS)

11. **How important are the below systems to your utility? (1-5, 1 = not important; 5 = very important)** *

	N/A	1	2	3	4	5
Advanced metering infrastructure (AMI)	○	○	○	○	○	○
Customer information system (CIS)	○	○	○	○	○	○
Computerized maintenance management systems (CMMS)	○	○	○	○	○	○
Laboratory information management systems (LIMS)	○	○	○	○	○	○
Supervisory control and data acquisition (SCADA) system	○	○	○	○	○	○
Enterprise asset management system	○	○	○	○	○	○
Surveillance and Reponses System (SRS)	○	○	○	○	○	○

12. **What system(s) below is your utility most likely to implement, upgrade or extend in the near future? Please select all that apply.**

- ☐ Advanced metering infrastructure (AMI)
- ☐ Customer information system (CIS)
- ☐ Computerized maintenance management systems (CMMS)
- ☐ Laboratory information management systems (LIMS)
- ☐ Supervisory control and data acquisition (SCADA) system
- ☐ Enterprise asset management system
- ☐ Surveillance and Reponses System (SRS)

13. **What does the information from these systems get used for? Please select all that apply.** *

- ☐ Operational decision making
- ☐ Short-term operations planning
- ☐ Short-term maintenance planning
- ☐ Long-term operations planning
- ☐ Long-term maintenance planning
- ☐ Capital expenditure planning
- ☐ Long-term asset management
- ☐ Other - Write In (Required)

*

14. **Please rate the importance of each of the following benefits of Big Data analysis will bring to your utility? (1-5, 1 as not important; 5 as very important)** *

	1	2	3	4	5
Optimal operation of treatment plants and networks,	○	○	○	○	○
Predict system and equipment failure	○	○	○	○	○
Accelerate the speed with which new capabilities and service are deployed	○	○	○	○	○
Decrease expenses through operational cost efficiencies	○	○	○	○	○
Mitigate knowledge loss from aging workforce	○	○	○	○	○
Improve workforce management	○	○	○	○	○
Extract greater value from existing analytical tools	○	○	○	○	○
Reduce non-revenue water to minimize water and revenue losses	○	○	○	○	○
Reduce pollution events	○	○	○	○	○

15. **Please rate the level of skills available within your utility for each of the following (1-5, 1 as no skills available; 5 as advanced skills available)?** *

	1	2	3	4	5
Management (storage, indexing and retrieval) of Big Data	○	○	○	○	○
Off-line analysis of Big Data	○	○	○	○	○
Real-time analysis of Big Data	○	○	○	○	○
Maintenance of systems to manage and analyze Big Data	○	○	○	○	○

16. **Please rate the level of Big Data's impact on each of the following areas of your utility in the next five years (1-5, 1 as no impact; 5 as will have a dramatic impact)?** *

	1	2	3	4	5
Impacting customer relationships	○	○	○	○	○
Changing the way we organize operations	○	○	○	○	○
Making the business more-data-focused	○	○	○	○	○

17. **What do you see as the water utility's role in developing "smart cities"?** *

- ○ Integral part of the process from planning to capital funding and spending
- ○ Advisory/consultative
- ○ Observe and report back to stakeholders/board
- ○ Don't know

18. **If you have implemented Big Data analysis, can you share one of your Big Data projects as an example? (project information, partnership, data sources, data analysis, tools and skills)**

Impediment Factors

19. **Please rate the level of influence of each of the following impediment factors have on Big Data analysis adoption in your utility (1-5, 1 as not influential; 5 as most influential)** *

	1	2	3	4	5
Data security	○	○	○	○	○
Data quality	○	○	○	○	○
Lack of budget	○	○	○	○	○
Lack of talent to implement big data	○	○	○	○	○
Lack of talent to run big data processing and analytics on an ongoing basis	○	○	○	○	○
Resistance to integrate existing systems	○	○	○	○	○
Procurement limitations on big data vendors	○	○	○	○	○
Lack of middle management adoption and understanding	○	○	○	○	○
Lack of data governance policies and practices	○	○	○	○	○

Thank You!

Thank you for taking our survey. Your response is very important to us.

APPENDIX B

Task 2 – Survey Water Industry Organizations and IT Research Firms

Leveraging Other Industries – Big Data Management

Task 2 – Survey Water Industry Organizations and IT Research Firms

PREPARED FOR:	Water Environment & Reuse Foundation
PREPARED BY:	CH2M
DATE:	March 14, 2018
PROJECT NUMBER:	SENG7R16

Task 2 – Survey Water Industry Organizations and IT Research Firms involved two aspects:

- A review of research from firms (such as Gartner, Forrester, and the ARC Advisory Council) on IoT and Big Data use in non-water sectors
- Research of water industry organizations (including the Water Environment Federation Knowledge Development Forum Team, American Water Works Association Information Management and Research Committee, SWAN Forum, Smart Cities Council, International Water Association, and others) to identify technical white papers, committees, and guidance on internet of things (IoT) and Big Data

IoT and Big Data applications in the water industry include examples such as real-time water quality monitoring, asset management, customer billing, and customer demand. Some utilities find these applications to be cost prohibitive and in conflict with their current culture, and they raise security concerns and issues. Experience from the non-water sector can be leveraged to assist with resolving these issues.

Approach

Papers and presentations were reviewed from as many sources as possible. Due to the evolving nature of this topic there were very few peer reviewed papers. The majority of the references were articles from periodicals, or presentation slides available on the web sites of industry organizations. This information was generated by industry experts, subject matter experts, and professionals within these organizations.

Gartner is a research and advisory company that specializes in the IT industry. Forrester is a research and advisory company working with businesses. ARC Advisory Group is a technology research and advisory firm for industry, infrastructure, and cities.

Introduction to IoT, Big Data, and Analytics

IoT, Big Data, and analytics have been used in non-water industries to transform data into information and knowledge. Principles of use from other industries can be applied to the water industry. This Technical Memorandum provides an overview of the research, including industry definitions for IoT and Big Data analytics, ideas for the application of these technologies from the non-water sector, and concludes with suggestions for how these can be applied in the water industry.

The Gartner IT Glossary (Gartner, 2018) defines The Internet of Things (IoT) as the network of physical objects that contain embedded technology to communicate and interact with their internal states or the external environment. This is further explained by McLellan as “a fast-growing constellation of internet-connected sensors attached to a wide variety of 'things'. Sensors can take a multitude of possible measurements, internet connections can be wired or wireless, while 'things' can literally be any object (living or inanimate) to which you can attach or embed a sensor. (McLellan, 2015)”

Kushmaro notes that Big Data and IoT are loosely intertwined, making it difficult to talk about one without the other (Kushamaro, 2017). While Big Data has existed to perform analytics long before the IoT, information from IoT devices resides in Big Data. Big Data analytics creates the insights from data delivered from IoT.

From 1960 through 2000, the amount of data generated grew exponentially. This data growth happened to overlap with growth in processing capabilities (Moore’s Law) and storage densities. During the early 2000s, the variety of data that needed to be processed, along with the sheer volume and velocity with which it was being collected, began to overwhelm traditional systems. New techniques needed to be applied.

Initial Big Data techniques typically involved processing in batches rather than in real-time. These technologies were derived from solutions developed by internet search companies as they struggled to manage their indexing and classification efforts on the internet. Though first coined in 2005, the NIST Big Data interoperability framework (NIST, 2015) defines Big Data as data that can no longer be efficiently handled by traditional data architectures such as static databases. Big Data is defined by four characteristics:

- Volume – size of the dataset
- Variety – data is from multiple sources
- Velocity – rate of flow of the data
- Variability – changes in other characteristics

Other authors (McLellan, 2015 and Kushmaro, 2017) include veracity (the uncertainty in the data) instead of variability. No explanation is provided by either author for this difference.

The IoT can provide data at a volume, variety, and velocity that creates Big Data and analytics can be performed on the data to derive information and knowledge that leads to insight.

Several basic types of analytics can be performed depending on the data source and particular application. The advantages analytics can offer are embedded in the four primary types (Conner, 2016; WEF, 2016), which are foundational to construction of an analytical program to fit a specific need as described in Table 1.

Table 1. Analytic Approaches

Analytic Type	**Result**
Descriptive	Backward looking analysis to tell us what happened
Diagnostic	Backward looking analysis to explain what and why an event occurred
Predictive	Forward looking analysis of what might happen in the future
Prescriptive	Forward looking analysis, recommendation on what should be done

Figure 1 provides a detailed breakdown of the value added by various advanced analytics to illustrate the concepts described above. Forward-looking analytics provide the most value to a business by predicting what may happen, and providing guidance on how to deal with the outcome, or better still, how to influence the outcome.

Leveraging Other Industries – Big Data Management

Task 3 – Survey Large Firms in the Non-Water Sector

PREPARED FOR:	Water Environment and Reuse Foundation
PREPARED BY:	CH2M
DATE:	January 16, 2018
PROJECT NUMBER:	SENG7R16

Task 3 – Survey Large Firms in the Non-Water Sector sought to determine how non-water sector organizations use Internet of Things (IoT) technology and Big Data analytics. The purpose of the survey was also to understand whether investments in IoT technology and Big Data analytics are considered important and the primary applications used. Firms were identified that were known to the researchers, and the survey was administered by personal contact to increase the response rate. This Technical Memorandum provides a summary of the survey development process and results of the survey.

Approach

A set of survey questions was developed to gather the information about how the non-water sector is using IoT technology and Big Data analytics. These questions are presented in the Attachment.

The survey was divided into the following sections:

- Organizational information (three questions)
- Status and future willingness of Big Data and IoT implementation (four questions)
- Big Data and IoT adoption (five questions)
- Big Data and IoT results achieved (three questions)
- Impediment factors (one question)

Survey Results

The survey had nine respondents. Details and analysis of the survey results are presented in this document under the sections of the survey.

Organizational Information

Respondents to the survey were asked what industry they operated in, the size of the organization, and the role of the person completing the survey.

Organization Size

Organization size was identified within the following groups:

- Less than 49 employees
- 50 to 199 employees
- 200 to 499 employees
- Greater than 500 employees

The results presented in Figure 1 show that five of the responding organizations were organizations of over 500 employees and four were of under 500 employees.

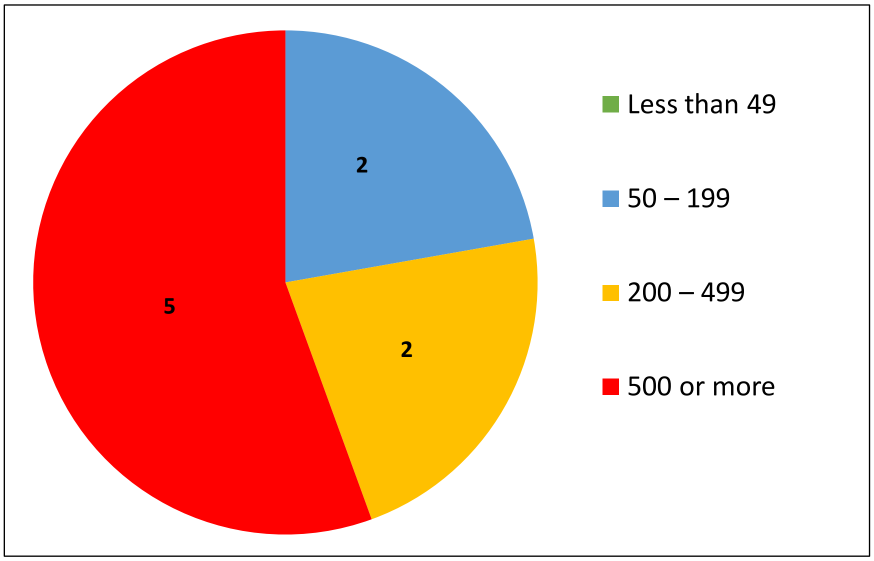

Figure 1. Organization Size

Industry

Respondents were asked in which of the following industries they worked. Some examples were provided; however, as it is not possible to identify all possible industries, the option for "Other" was included. The list presented the following options:

- Consulting
- Financial services
- Healthcare/life sciences
- Media/entertainment
- Retail
- Telecommunications
- Other

Survey responses are shown in Figure 2. Industries represented included consulting, financial services, retail, telecommunications, and three identifying as "Other." Other included instrumentation design, food processing technology, and logistics.

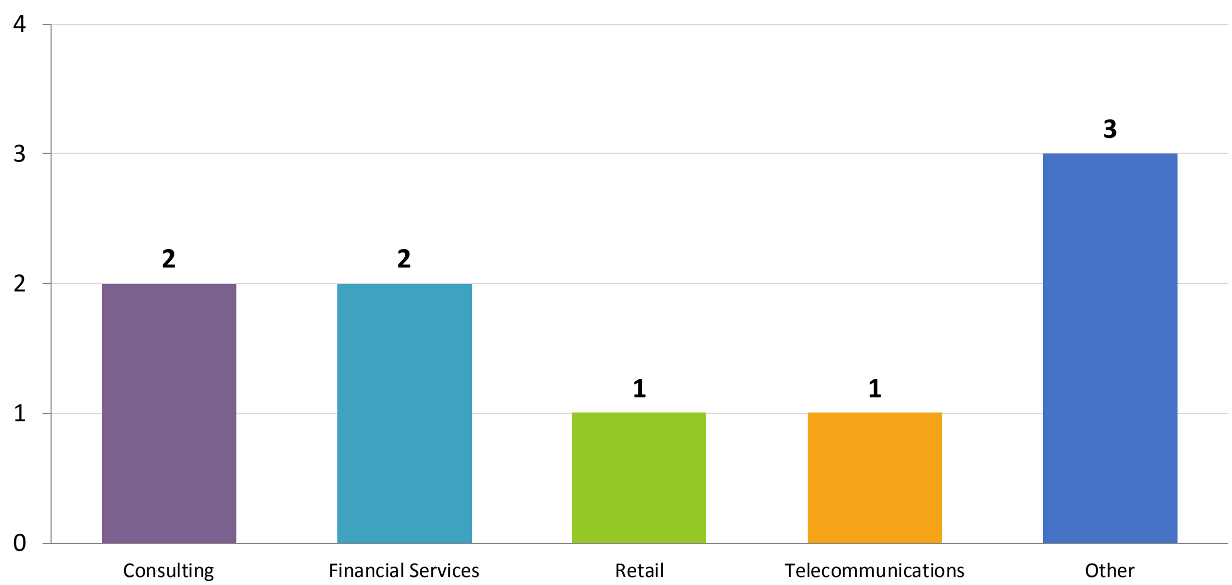

Figure 2. Industry

Organizational Role

This question was intended to identify the organizational level of the person completing the survey. Responses included Logistic Specialist, Consultant, Engineer, Data Analyst, Head of Big Data, Head of Product Management and Engineering Development, and Assistant Vice President Investment/Research.

Status and Future Willingness of Big Data/IoT Implementation

This section of the survey focused on the components that have been implemented to store and process Big Data and IoT data, and the near-term plans the company may have.

If your company has implemented components to store and process Big Data, when was it implemented?

The following five options were provided for this question:

- Within the past year
- Two years ago
- Three years ago
- More than 4 years ago
- Not implemented

The results are presented in Figure 3. Two respondents have not implemented components to store and process Big Data. Of the seven who have implemented these components, one was more than 3 years ago, and the other six were more than 4 years ago.

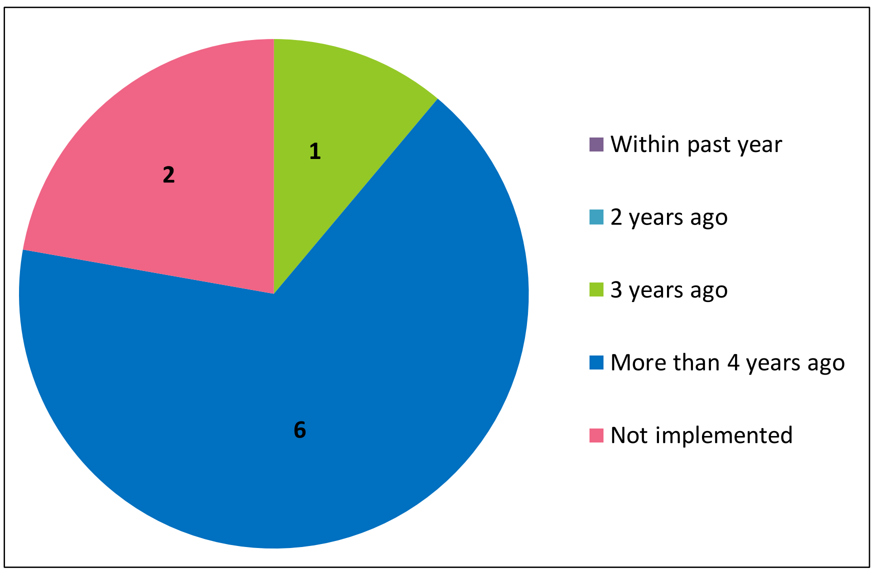

Figure 3. Implementation of Components to Store and Process Big Data

If your company has implemented components to store and process data produced by Internet of Things (IoT), when was this implemented?

The following five options were provided for this question:

- Within the past year
- Two years ago
- Three years ago
- More than 4 years ago
- Not implemented

The results are presented in Figure 4. Four respondents have not implemented components to store and process IoT data. Of the five who have implemented these components, one was more than 3 years ago, and four were more than 4 years ago.

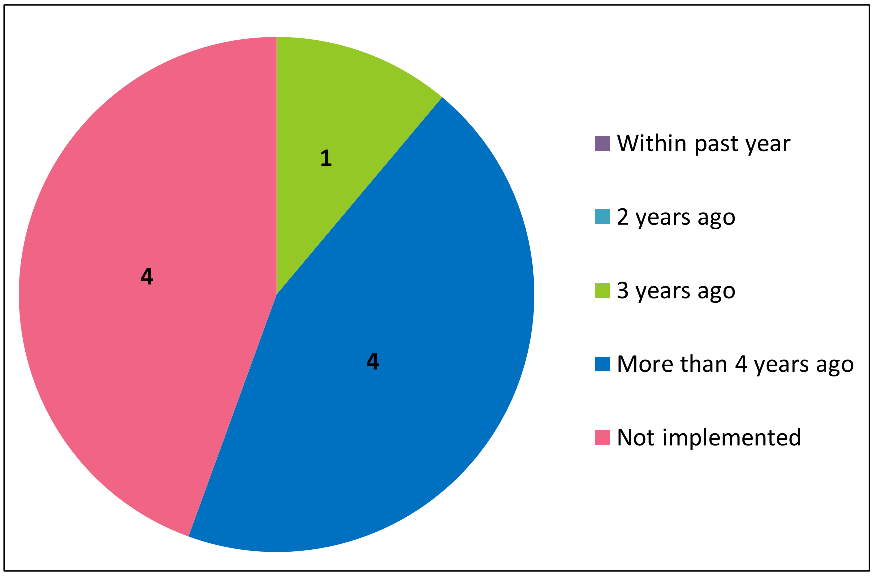

Figure 4. Implementation of Components to Store and Process IoT Data

Does your company plan to have or increase investment on Big Data/IoT in the near term?

Figure 5 presents results that indicate two-thirds of respondents intend to invest in Big Data/IoT in the near term.

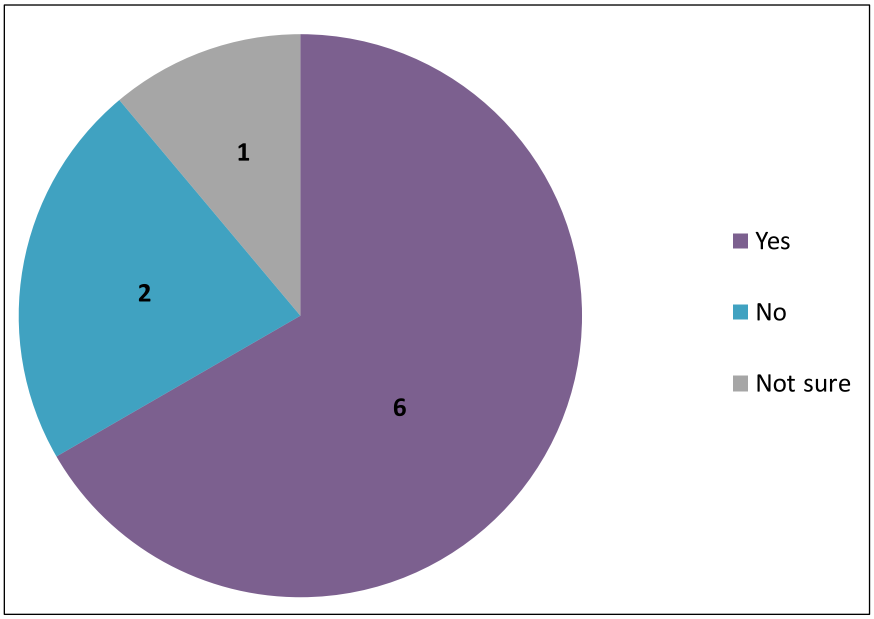

Figure 5. Plans for Big Data Investments

If YES, when do you see your company will have or increase investment on Big Data/IoT?

Of the seven respondents that answered "Yes" or "Not Sure" to the previous question, Figure 6 shows that six are expecting to invest in the next year, while one is planning to invest in 2 to 3 years (this is the organization that responded with "Not Sure" in the previous question).

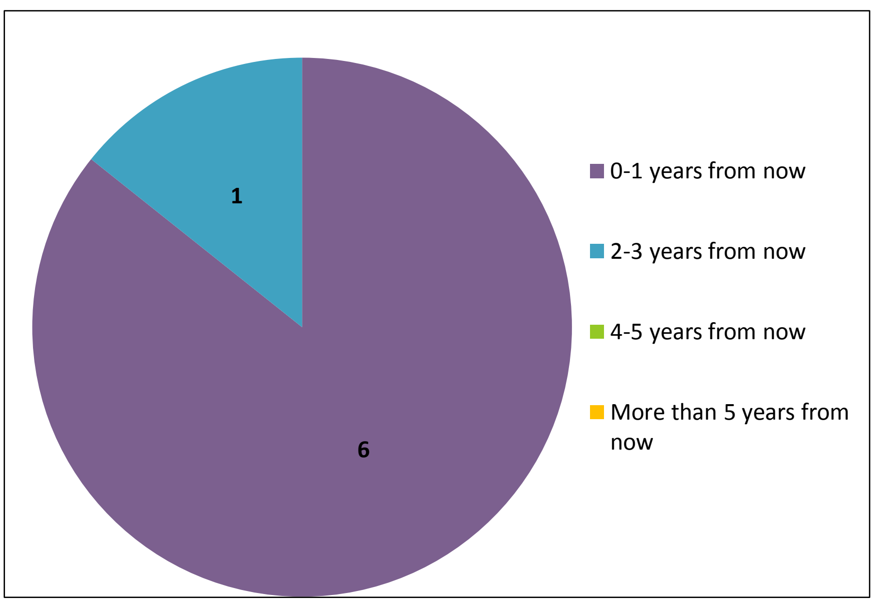

Figure 6. When Investment is Expected

If NO, what is the reason for not investing in Big Data/IoT?

Options were:

- We don't see the need
- We don't have the budget
- Insufficient internal capabilities
- Other reasons

Figure 7 indicates that, of the three respondents that answered "No" or "Not Sure" to the previous question, two do not have the budget (including the "Not Sure") and one (a consulting firm) replied that there is no need identified to justify the investment.

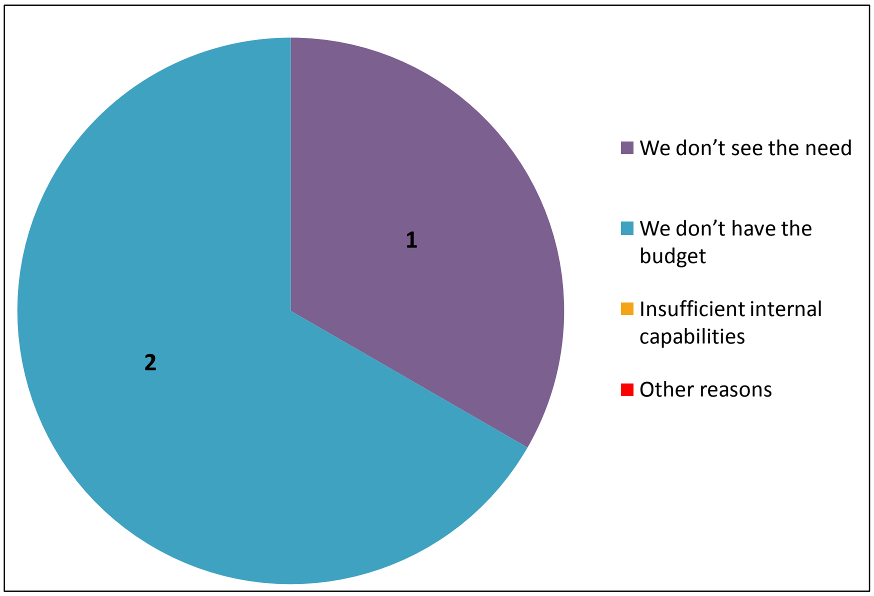

Figure 7. Reasons for Not Investing

How important is Big Data/IoT to your company?

Options were:

- It is the single most important way for us to maintain operations and reduce costs
- Is a top 5 issue that gets significant time and attention from top leadership
- Is important but is only one of many other challenges/opportunities that we need to address
- Not very important
- Don't know/NA

The results in Figure 8 show that only one of the respondents considers Big Data analysis to be the most important way to maintain operations and reduce costs. Four of the respondents identify it as a top five issue, and three identify it as important, but that it is only one of many things to be addressed.

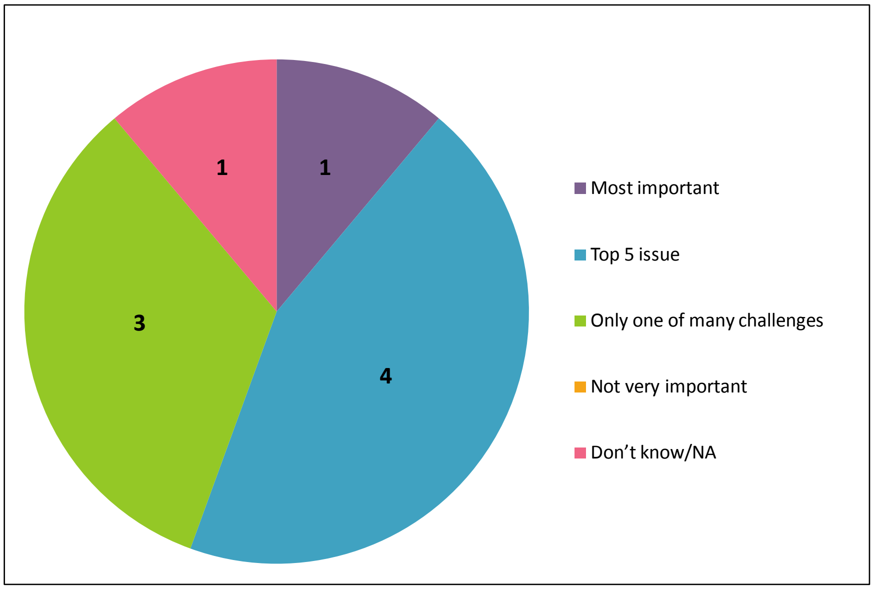

Figure 8. Importance of Big Data Analysis

Big Data/IoT Adoption

The questions for this section of the survey sought to understand the uses of Big Data and IoT within the company.

Which departments in your company have adopted Big Data/IoT?

This question asked the respondent to select all that apply from the following list:

- Accounting and finance
- Customer service
- Energy
- Finance
- Human resource management
- Logistics
- Marketing
- Production
- Research and development
- Supply chain
- Other
- We use outsourcing for Big Data/IoT analysis

Results are presented in Figure 9. Production and research and development were identified as the main users of Big Data and IoT analysis. Results for “Other” included “software development” and “don’t know”.

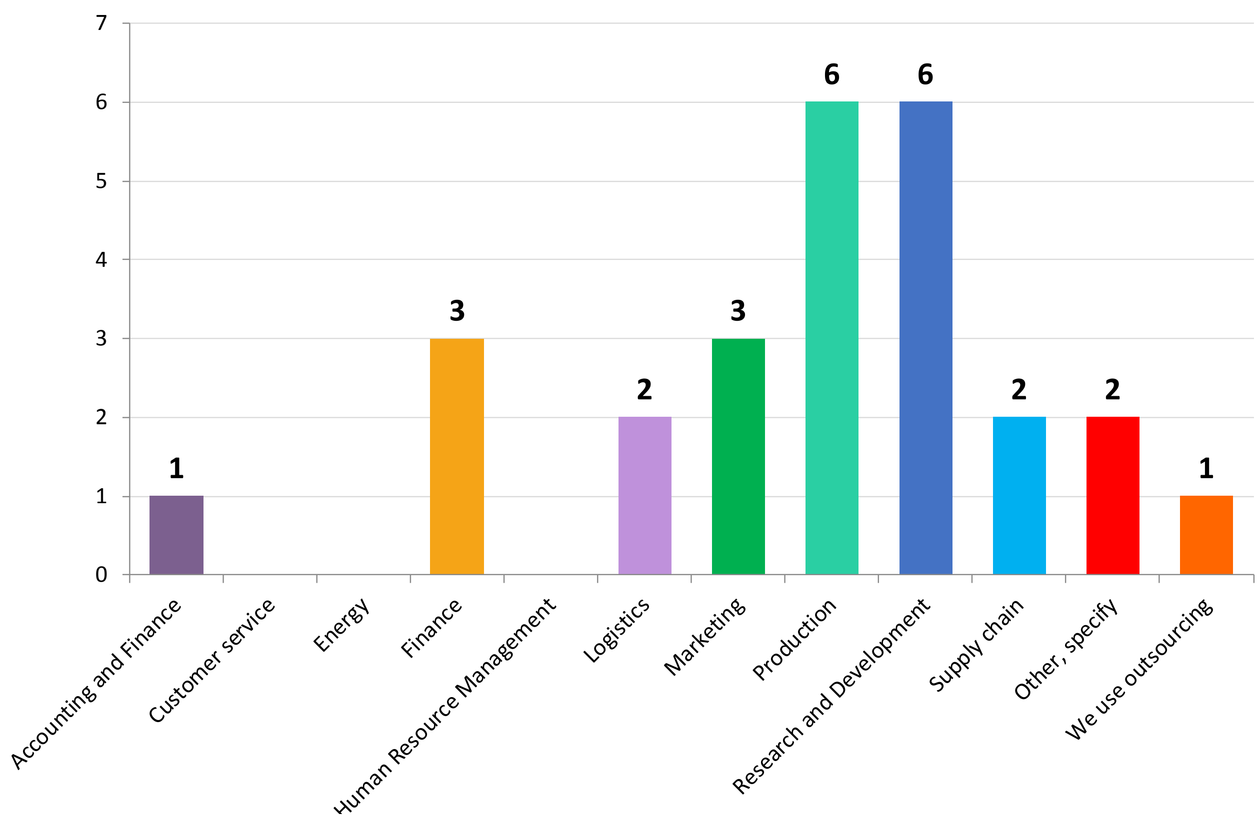

Figure 9. Departments that have Adopted Big Data/IoT

Which departments in your company do you think will adopt Big Data/IoT in the near future?

This question asked the respondent to select all that apply from the following list:

- Accounting and finance
- Customer service
- Energy
- Finance
- Human resource management
- Logistics
- Marketing
- Production
- Research and development
- Supply chain
- Other
- We will keep using outsourcing for Big Data/IoT analysis

Figure 10 provides a comparison of the results in Figure 9: the departments that have adopted Big Data/IoT (the solid color bars) and those that are expected to adopt Big Data/IoT in the near future (the patterned bars). Results for “Other” included “software development” and “don’t know”. These results indicate that each company identified different future pathways for the use of Big Data and IoT. There were no common themes across the respondents.

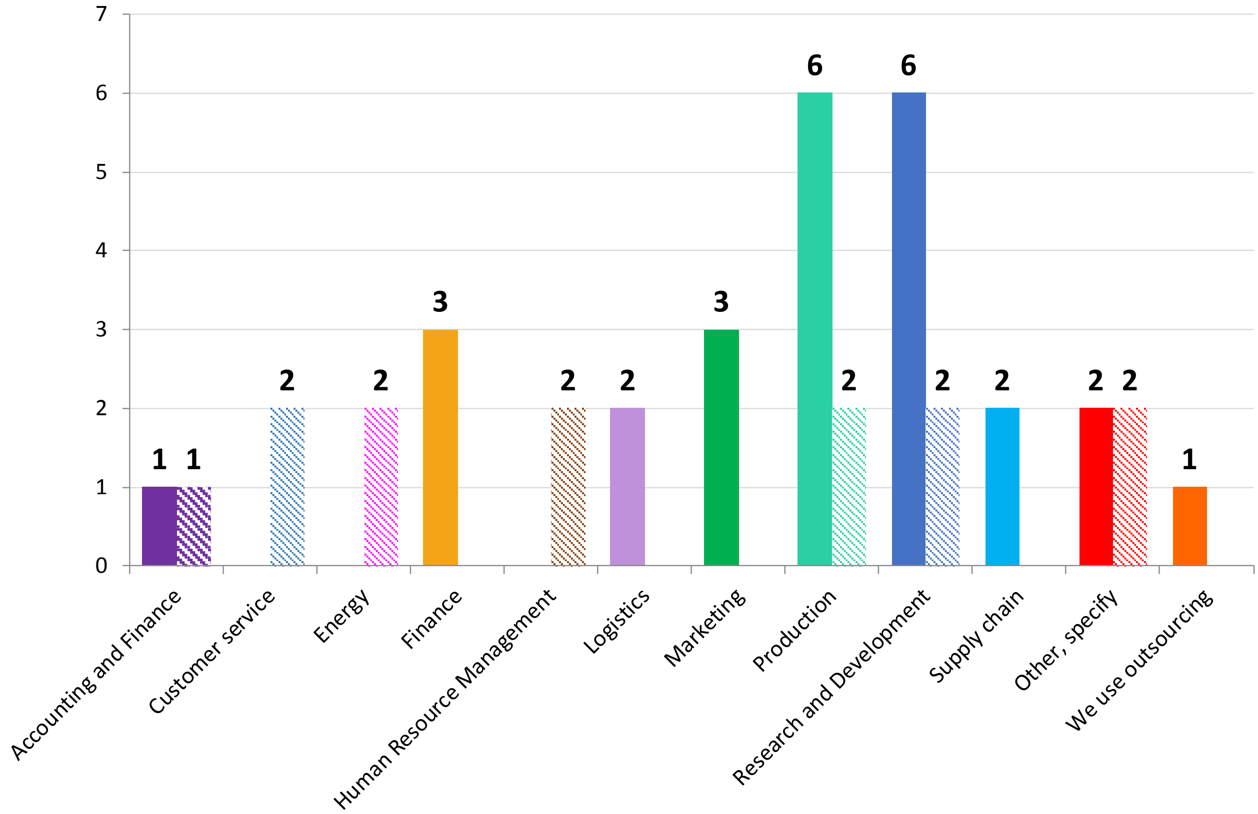

Figure 10. Departments That Have Adopted and Those That Are Likely to Adopt Big Data/IoT in the Near Future

How much has your company invested in Big Data/IoT?

This question asked the respondent to select from the following:

- Greater than $50 million
- $5 to $50 million
- $1 to $5 million
- Under $1 million
- Not available
- No investment

Results are presented in Figure 11.

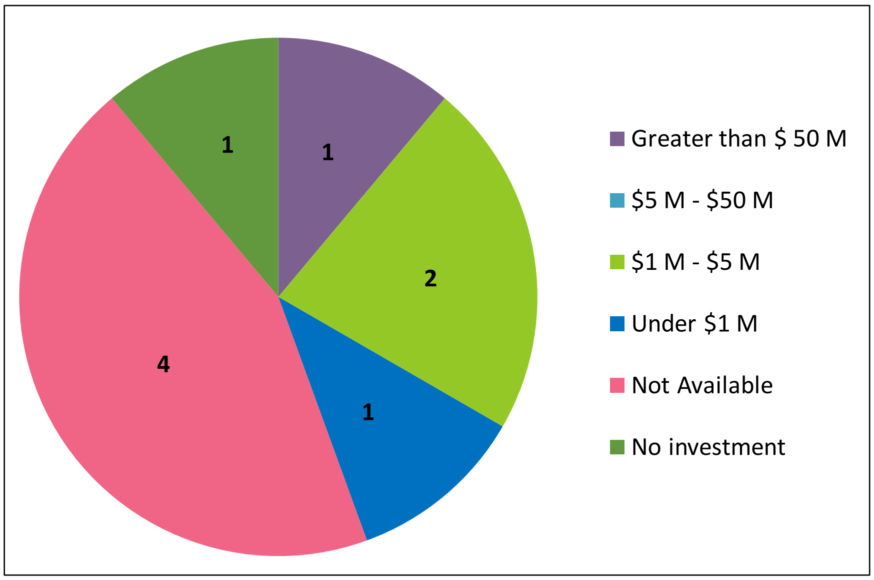

Figure 11. Investments in Big Data

Which functions of Big Data/IoT are used in your company?

This question asked the respondent to select all that apply from the following list:

- Monitoring and sensing physical quantity and quality
- Tracking physical movement
- Targeting customers (for advertising by marketers)
- Dynamic response to product demands
- Real-time optimization of manufacturing production and supply-chain networks
- Monitoring and controlling operations of infrastructure
- Market projection
- Energy and resource management
- Dynamic interaction between different components (i.e., the vehicle, infrastructure, driver or user)
- Other, specify
- Not applicable

The results indicate that monitoring and sensing physical quantity and quality, and targeting customers are the most common uses of Big Data and IoT by the respondents

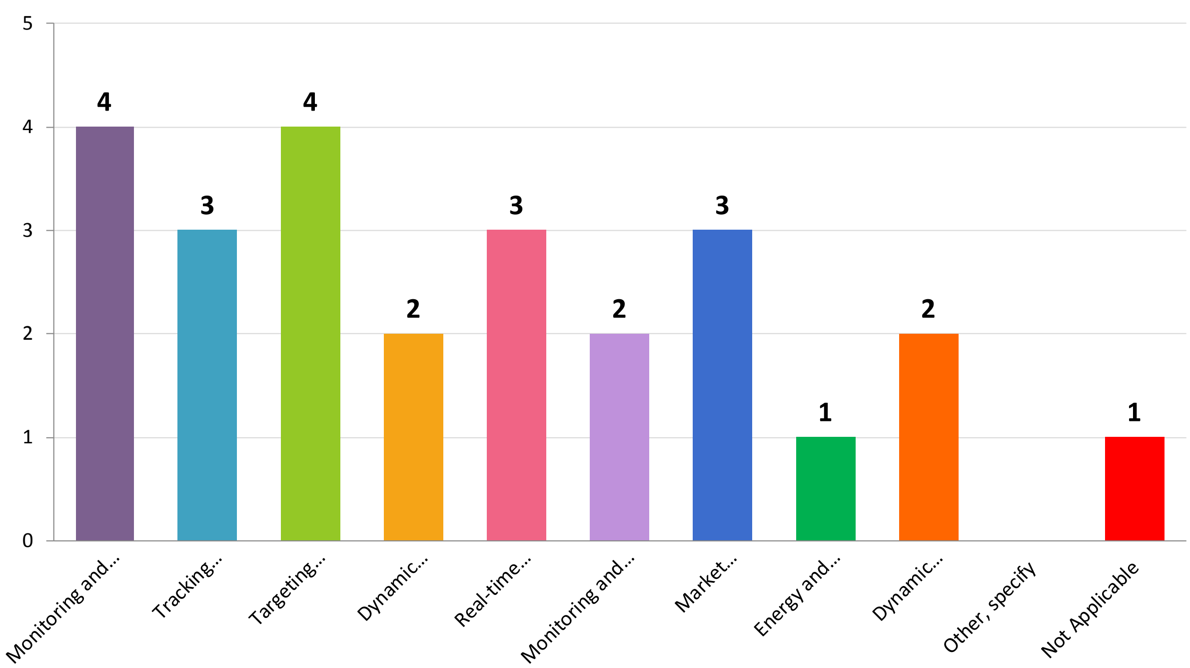

Figure 12. Functions of Big Data/IoT used Currently

Please rate the level of skills available within your company for each of the following

This question asked the respondent to rate the skills in the following list from 1 to 5 (1 as no skills available; 5 as advanced skills available):

- Data engineering (infrastructure/architecture to support data scientists)
- NoSQL for operational data storage
- Statistical and quantitative analysis
- Machine learning
- Stream processing
- Data visualization
- R programming
- Python programming
- Scala programming

The ratings were adjusted to a Low/Medium/High scale where 1 and 2 were combined to be "Low," 3 was made "Medium," and 4 and 5 were combined to be "High."

Results using the Low/Medium/High rating scale are presented in Table 1. Multiple respondents did not rate all skills in the table with no explanation. Respondents considered that they had advanced skills in data engineering, statistical and quantitative analysis, and stream processing, with low skills in R programming and Scala programming.

Table 1. Level of Skills Currently Available

	Low	Medium	High
Data engineering (infrastructure/architecture to support data scientists)	3		5
NoSQL for operational data storage	2	1	4
Statistical and quantitative analysis	1	3	5
Machine learning	2	1	4
Stream processing	2		5
Data visualization	1	3	3
R programming	5		1
Python programming	4		3
Scala programming	5		1

Big Data/IoT Results

This section sought to identify the positive results achieved by the respondents' implementation of Big Data and IoT analysis.

Can your company quantify results from Big Data/IoT?

These answers were categorized as follows:

- Yes, estimated value is $______
- No
- Not Applicable

The results presented in Figure 13 indicate that four of the respondents marked "Not Applicable." Of these four companies, three of them have invested in Big Data and IoT, and intend to invest further (based on responses to previous questions). Of the three "No" responses, two of these companies have invested in in the technologies in the past.

The two "Yes" responses were accompanied with the values $20M (from a consulting company) and $100M+ (from a telecommunications company).

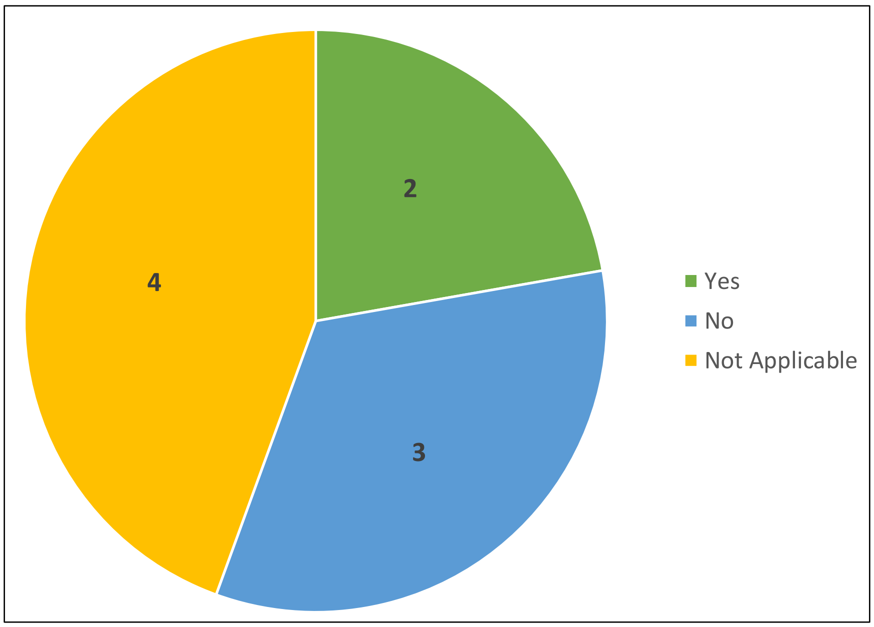

Figure 13. Quantification of Results from Big Data/IoT

In which areas have your company received measurable results from Big Data/IoT?

This question asked the respondent to select all that apply from the following list:

- Decrease expenses
- Establish a data-driven culture
- Accelerate the speed with which new capabilities and services are deployed
- Improve customer service
- Better understand customer trends
- Identification of failures
- Launch new product and service offerings
- Expand market share
- Other, specify
- Not applicable

Results are presented in Figure 14. Establishing a data-driven culture was selected most often, with decreasing expenses and improving customer service close behind.

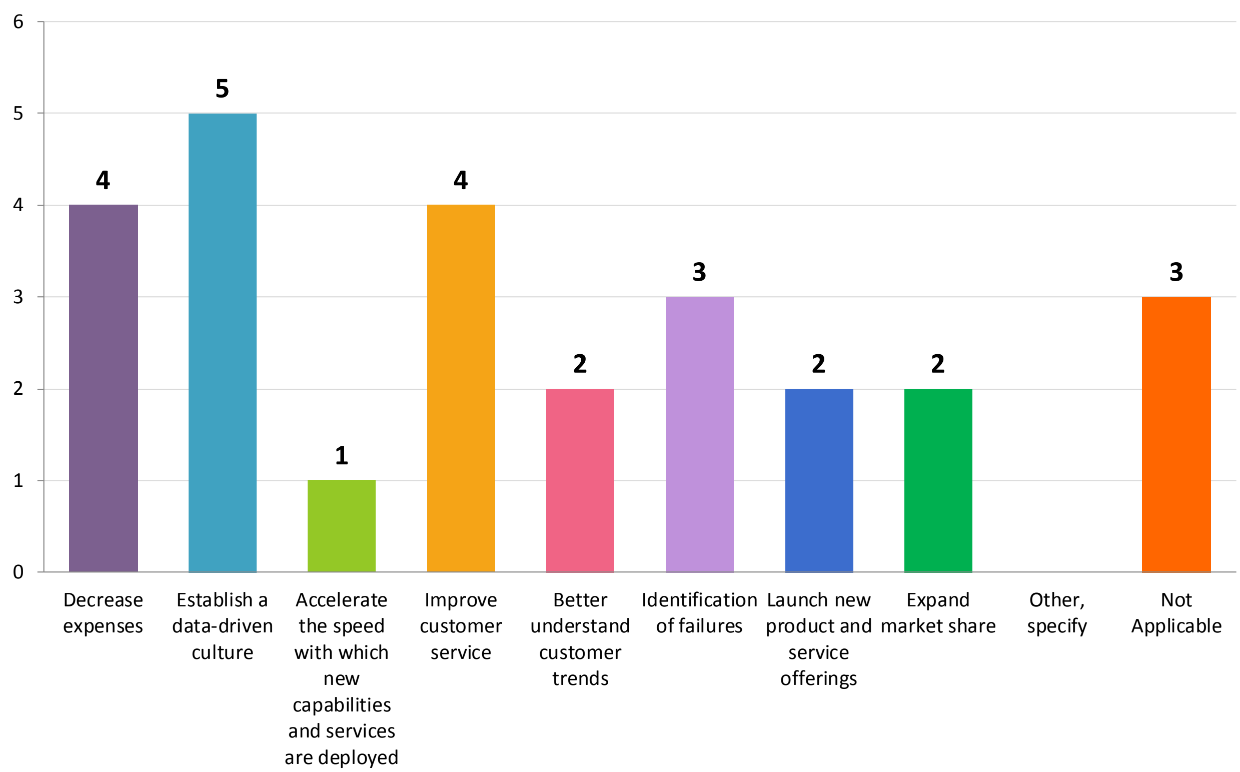

Figure 14. Areas that have Received Measurable Results from Big Data/IoT

Please rate the level of Big Data's impact on each of the following areas of your company in the next five years

Ratings were 1 to 5, where 1 = no impact; 5 = will have a dramatic impact. The areas to be rated were as follows:

- Impacting customer relationships
- Changing the way we organize operations
- Making the business more data-focused

The ratings were adjusted to a Low/Medium/High scale where 1 and 2 were combined to be "Low," 3 became "Medium," and 4 and 5 were combined to be "High."

Results using the Low/Medium/High rating scale presented in Table 2 indicate that making the business more data focused ranks the highest of the three areas.

Table 2. Impact of Big Data

	Low	Medium	High
Impacting customer relationships	2	3	3
Changing the way we organize operations	1	4	4
Making the business more data-focused	1	1	6

Impediment Factors

This section had one question about impediments to adoption of Big Data analysis within a utility.

Rate the level of influence that each of the following impediment factors have on Big Data analysis adoption in your company

Ratings were 1 to 5, where 1 = not influential and 5 = most influential. The factors to be rated were as follows:

- Data security
- Data quality
- Lack of budget
- Lack of talent to implement Big Data
- Lack of talent to run Big Data processing and analytics on an ongoing basis
- Resistance to integrate existing systems
- Procurement limitations on Big Data vendors
- Lack of middle management adoption and understanding
- Lack of data governance policies and practices

The ratings were adjusted to a Low/Medium/High scale where 1 and 2 were combined to be "Low," 3 became "Medium," and 4 and 5 were combined to be "High."

Results using the Low/Medium/High rating scale are presented in Table 3. The results indicate the following:

- Impediments considered most influential are data security and data quality.
- Lack of budget and lack of talent to implement Big Data are not seen as influential.

Table 3. Impediment Factors to Big Data Adoption

	Low	Medium	High
Data security	3		5
Data quality	2	1	5
Lack of budget	6		2
Lack of talent to implement Big Data	6		1
Lack of talent to run Big Data and analytics on an ongoing basis	4	1	2
Integration resistance with existing systems	5		2
Procurement limitations on Big Data vendors	5	2	
Lack of middle management adoption and understanding	4	2	2
Lack of data governance policies and practices	5		2

Summary/Conclusions

The survey had nine respondents covering multiple industries. Summary conclusions include the following:

- Components to store and process Big Data and IoT data were either put in place more than 3 years ago or not implemented, with IoT implemented in fewer cases than Big Data.
- Two-thirds of respondents plan to invest in Big Data in the near term.
- None of the respondents considered Big Data analysis to be unimportant. Only one of the respondents considered it the single most important way to maintain operations and reduce costs. Four considered it a top five issue, while three consider it important, but only one of many things to be addressed.
- Two-thirds of respondents use Big Data/IoT in the production and research and development departments.
- Monitoring and sensing physical quantity and quality, and targeting customers are the primary functions for which Big Data and IoT are used.
- Respondents considered that they had advanced skills in data engineering, statistical and quantitative analysis, and stream processing, with few skills in R programming and Scala programming.
- Measurable results were only seen for establishing a data-driven culture.
- Impediments to the use of Big Data include data security and data quality.

Attachment--
Survey Non-Water Sector

- Industry:
  - ☐ Consulting
  - ☐ Financial Services
  - ☐ Healthcare /Life Sciences
  - ☐ Media/Entertainment
  - ☐ Retail
  - ☐ Telecommunications
  - ☐ Other (transportation, shipping, maintenance monitoring (e.g., GE aircraft engines), etc.)

- Organizational Role:
  - ☐ Chief Data Officer
  - ☐ Chief Analytics Officer
  - ☐ Chief Information Officer
  - ☐ Head of Big Data
  - ☐ CEO/President
  - ☐ Chief Marketing Officer
  - ☐ Other, specify

Status & Future Willingness of Big Data/IoT Implementation

- If your company has implemented components to store and process Big Data, when was it implemented?
  - ☐ Within past year
  - ☐ 2 years ago
  - ☐ 3 years ago
  - ☐ More than 4 years ago
  - ☐ Not implemented.

- If your company has implemented components to store and process data produced by the Internet of Things (IoT), when was this implemented?
  - ☐ Within past year
  - ☐ 2 years ago
  - ☐ 3 years ago
  - ☐ More than 4 years ago
  - ☐ Not implemented.

- Does your company plan to have or increase investment on Big Data/IoT in the near term?
  - ☐ Yes
  - ☐ No

  If YES, when do you see your company will have or increase investment on Big Data/IoT?
  - ☐ 0-1 years from now
  - ☐ 2-3 years from now
  - ☐ 4-5 years from now
  - ☐ More than 5 years from now

  If NO, what's the reason of not investing on Big Data/IoT?
  - ☐ We don't see the need
  - ☐ We don't have the budget
  - ☐ Insufficient internal capabilities
  - ☐ Other reasons

- How important is Big Data/IoT to your company?
  - ☐ Is single most important way for us to gain competitive advantage
  - ☐ Is a top 5 issue that gets significant time and attention from top leadership
  - ☐ Is important but is only one of many other challenges/opportunities that we need to address
  - ☐ Not very important
  - ☐ Don't know/NA

Big Data /IoT Adoption

- Which departments in your company have adopted Big Data/IoT (select all that apply)?
  - ☐ Accounting and Finance
  - ☐ Customer service
  - ☐ Energy
  - ☐ Finance
  - ☐ Human Resource Management
  - ☐ Logistics
  - ☐ Marketing
  - ☐ Production
  - ☐ Research and development
  - ☐ Supply chain
  - ☐ Other, specify
  - ☐ We use outsourcing for Big Data /IoT analysis, specify

- Which departments in your company do you think will adopt Big Data/IoT in the near future (select all that apply)?
  - ☐ Accounting and Finance
  - ☐ Customer service
  - ☐ Energy
  - ☐ Finance
  - ☐ Human Resource Management
  - ☐ Logistics
  - ☐ Marketing
  - ☐ Production
  - ☐ Research and development
  - ☐ Supply chain
  - ☐ Other, specify
  - ☐ We will keep using outsourcing for Big Data /IoT analysis, specify

- How much has your company invested in Big Data/IoT?
  - ☐ Greater than $ 50 M
  - ☐ $5 M - $50 M
  - ☐ $1 M - $5 M
  - ☐ Under $1 M
  - ☐ Not Available
  - ☐ No investment

- Which functions of Big Data/IoT are used in your company (select all that apply)?
  - ☐ Monitoring and sensing physical quantity and quality
  - ☐ Tracking physical movement
  - ☐ Targeting customers (for advertising by marketers)
  - ☐ Dynamic response to product demands
  - ☐ Real-time optimization of manufacturing production and supply chain networks
  - ☐ Monitoring and controlling operations of infrastructure
  - ☐ Market projection
  - ☐ Energy and resource management
  - ☐ Dynamic interaction between different components (i.e. the vehicle, infrastructure, driver or user)
  - ☐ Other, specify
  - ☐ Not Applicable

- Please rate the level of skills available within your company for each of the following (1-5, 1 as no skills available; 5 as advanced skills available)?
    - Data Engineering (infrastructure/architecture to support data scientists) __
    - NoSQL for operational data storage __
    - Statistical and quantitative analysis __
    - Machine learning __
    - Stream processing __
    - Data visualization __
    - R programming __
    - Python programming __
    - Scala programming __

Big Data/IoT Results

- Can your company quantify results from Big Data/IoT?
    - ☐ Yes, estimated value is $______
    - ☐ No
    - ☐ Not Applicable

- In which areas have your company received measurable results from Big Data/IoT (select all that apply)?
    - ☐ Decrease expenses
    - ☐ Establish a data-driven culture
    - ☐ Accelerate the speed with which new capabilities and services are deployed
    - ☐ Improve customer service
    - ☐ Better understand customer trends
    - ☐ Identification of failures
    - ☐ Launch new product and service offerings
    - ☐ Expand market share
    - ☐ Other, specify
    - ☐ Not Applicable

- Please rate the level of Big Data's impact on each of the following areas of your company in the next five years (1-5, 1 as no impact; 5 as will have a dramatic impact)? *
    - Impacting customer relationships __
    - Changing the way we organize operations __
    - Making the business more data-focused __

Impediment Factors

- Rate the level of influence that each of the following impediment factors have on Big Data analysis adoption in your company (1-5, 1 as not influential; 5 as most influential)
    - Data security __
    - Data quality __
    - Lack of budget __
    - Lack of talent to implement big data __
    - Lack of talent to run big data and analytics on an ongoing basis __
    - Integration resistance with existing systems __
    - Procurement limitations on big data vendors __
    - Lack of middle management adoption and understanding __
    - Lack of data governance policies and practices __

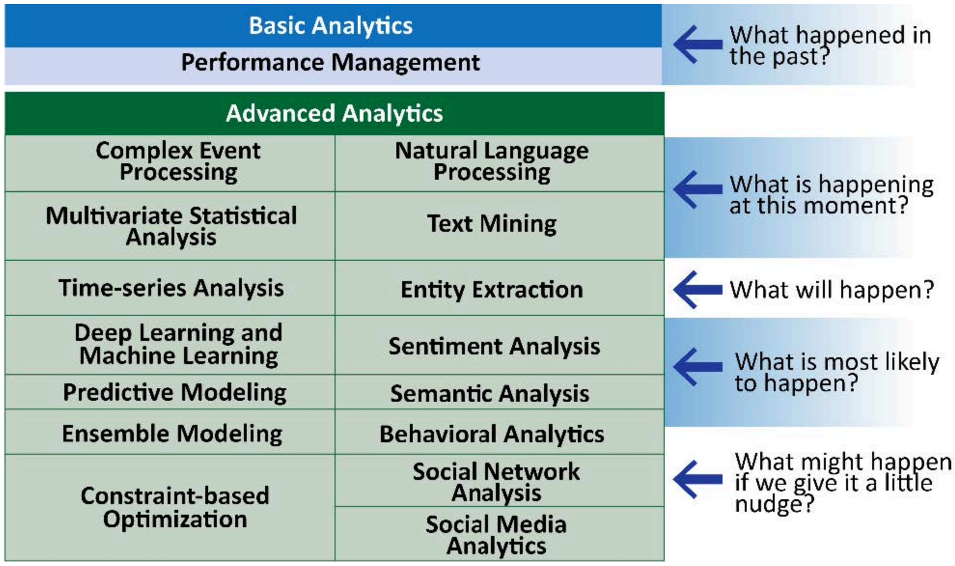

Figure 1. Analytics Overview

Application of Big Data Management in Non-Water Industries

Big Data management includes data sources, software platforms and integration, and Big Data analytics. These have been leveraged by non-water industries to create information and knowledge.

Data Sources

Big Data doesn't mean simply collecting large quantities of data from all kinds of different places. It is about the information produced being useful and accessible by the people in your organization (HBR, 2014).

A common theme found in the literature is that existing infrastructure should be monitored and assessed focusing efforts on rectifying poor-quality data from data sources. Quality is more important than quantity and therefore maintenance, calibration and proper use of the sensors, with validation is important to provide good primary measurements (HBR, 2014).

It is critical to process existing data efficiently and prove value from existing information before investing in further instrumentation (HBR, 2014). Innovative data analysis processes are reducing the need to "over-sensorize" a system so it is recommended to measure only what is useful. Over measuring will lead to higher maintenance and the instruments to get ignored (HBR, 2014).

A common downfall of some organizations is to be too tolerant of poor-quality data. Solving the data problem often isn't hard, but it may involve a change to the status quo or organizational culture (HBR, 2014). Also important, is consideration of how to handle outliers and extraordinary results and how to avoid these poor-quality data points having an impact on decisions. This can be handled through analytics (HBR, 2014).

The literature generally concludes that it is more important to have good quality data and the right data rather than more data.

Software Platforms and Integration

No vendor today has a complete platform solution for data collection, aggregation, and analysis (Gartner, 2016). Enterprises are therefore forced to implement solutions from multiple vendors, and take responsibility for their integration into central enterprise resources. This can be especially challenging with compatibility issues often occurring between software platforms and enterprise resources.

Platform capabilities may be distributed between cloud services and on-premises gateways or devices with the center of data processing and aggregation distributed as appropriate across these environments. It is expected that 50 percent of enterprises will move to the use of public-cloud for data, big data, and analytics, in 2018 as they look to control costs (Forrester, 2018).

Most IoT platforms do not represent the full set of services needed to assist an enterprise to build an IoT business solution, much less to become a digital business. Any IoT platform will need to be configured, modified and integrated with back-end systems and data to best meet the requirements of the specific IoT project or deployment (see "IoT Drives New Integration Challenges and Best Practices") (Gartner, 2016).

The literature describes all implementations of Big Data and IoT having the need to integrate applications from different vendors, as there is no fully integrated solution available. This implies that an organization undertaking a Big Data and IoT project will need to accept the development organization that will be required.

Big Data Analytics

Big Data analytics is about taking all the data an organization has and turning it into knowledge that can enable better operation. The right data, analytics, and decision framework can drive optimal performance (Shaw, 2017).

Key takeaways noted by ARC (Abel, 2015) include:

- Data needs to be turned into operational intelligence to be useful
- A combination of technology in the cloud and in datacenters is used to achieve the outcomes
- The technology is used to develop innovative business solutions that extract business value from data

Artificial intelligence (AI) is expected to be used in 20 percent of enterprises in 2018 to make decisions and provide real-time instructions such as suggesting what to offer customers, recommend terms to give suppliers, and instruct employees on what to say and do (Forrester, 2018).

Six key aspects for laying the foundation for Big Data analysis identified in (Liner and Kenel, 2016) are:

- **Integration:** It is critical to ensure there is only one platform managing the data. “Separate silos of data only create separate silos of insight”. An integrated solution also should include more than one technology.
- **Analytics tools:** These create more sophisticated, accurate, and actionable information.
- **Visualization tools:** These present the information in a form that is understandable by decision makers.
- **Development tools:** These provide the means to enhance the analytical and visualization tools as well as supporting the overall platform.
- **Workload optimization:** This focuses on efficient processing and storage of the data.
- **Security and governance:** These provide the means for maintaining the sensitive data that must be protected.

Gartner notes that moving to a digital business is not an incremental change because it is dramatic and all-encompassing. Incremental approaches tend accommodate silos which need to be broken down as part of moving to a digital business (Gartner, 2016a). Likewise, implementing Big Data in an organization should be thought of as a process focused on how to get to new insights and turn them into action, resulting in business value (Gartner, 2016b). HBR (2014) notes that data visualization is important when selling a complex answer to a complex problem.

Key ideas from other industries which can be helpful when considering Big Data management and implementation as a process include (HBR, 2014):

- **Focus on the top journeys.** Select three to five journeys that matter most to customers and the bottom line and prioritize these.
- **Don't wait for the data to be perfect.** The challenge is pulling the data together.
- **Focus on analytics, not reporting.** Much greater value comes from analyzing data to pinpoint cause and effect and make predictions rather than just reporting.

The literature recommends a number of approaches to developing and implementing Big Data management. These approaches have been summarized in Table 2.

Table 2. Approaches to Developing and Implementing Big Data Management

Approach	Description	Reference
A	• A deductive approach starts with business issues. Identify a business issue that could benefit from better insight, the use relevant data sources to help provide that insight. • Alternatively, the inductive approach starts with data. Analyze as much raw data as possible in a certain area from a wide variety of sources, using different styles of analytics, to find anomalies, outliers, new clusters, and patterns that can lead to new insights and ideas, and new relevant business questions.	Gartner, 2016b
B	• Start by developing a list of interesting ideas that should be further explored, and prioritize these ideas based on expected feasibility, impact, or broader applicability. • Perform many experiments in a short timeframe to test the ideas, so that failures can be determined quickly. • Perform validation to test the value of any new insight to confirm the business case and return on investment. • Finally, implement the solution by turning the analytic into a product-level solution that is operated and maintained in the most efficient and repeatable way.	Gartner, 2016b
C	• Approach data and analytics as a core strategic capability within a larger digital business strategy, not simply as a set of related initiatives. • Leverage data and analytics for intelligent business processes, which identify and enable digital business moments. • Start small and build momentum by alleviating operational pain and deriving business value. • Build an agile, repeatable methodology, positioning data and analytics capabilities as embedded within the infrastructure of digital business.	Gartner, 2016a

D	• Consider an approach for any IoT projects initiated or supported in an enterprise based on short, medium, and long term strategic actions.	Gartner, 2016

In addition to the approaches described in Table 2, it is essential to identify key staff. To build a more analytical organization, the two cadres of employees needed are analytics professionals and analytical businesspeople (HBR, 2014). Analytics professionals mine and prepare data, perform statistical operations, build models, and program the surrounding business applications, while analytical businesspeople, are those able to use the information and analyses in their work.

One of the implications of data and analytics as part of a digital business is that all of the organization and processes are a target. Organizations should start small but also start anywhere there is operational pain and an opportunity to get traction and deliver business value. Data and analytics can be used to both report on the operation of each element of a business, as well as be used to operate that element (Gartner, 2016a).

The general literature overall suggests a major shift in organization and culture may be required to successfully make the most out of becoming a digitally focused organization.

Application of Big Data Management in the Water Industry

While the water industry has been gathering various types of data, such as SCADA data, for many years, it is often managed and used in isolated systems. For data to be useful it must be converted to information using analytics (Shaw, 2017), and be available throughout an organization. But analytics is only successful when it's being used to improve business performance (HBR, 2014).

Data Sources and Quality

This section considers the sources of data used, whether they are individual sensors (such as smart meters), or data from manual entry (such as customer details or complaints), or data created from other applications (such as mapping tools).

Instrumentation now is better than ever, and maintenance and calibration is well understood, however it is still important to verify that the data produced is of good quality (Shaw, 2017).

When there is an assumption that if the answer is to be found in the data, then there is a perception that more data from more sensors will provide better answers. However, solving real operational and management problems with Big Data includes the integration of all operational data, such as customer billing, GIS, hydraulic modeling databases, SCADA, and computerized maintenance management software (CMMS), which is often stored in disassociated platforms. If leveraging existing data for specific management and operational goals is difficult, then adding more data may prove especially challenging (Conner, 2016).

Despite the current focus for many on improving analytics and/or decisions, there's also a lot to be said for making sure our foundations are sound (Shaw, 2017).

Below are five keys to making Big Data work and avoiding the pitfalls of poor-quality data (Shaw, 2017):

- **Focus on data quality rather than quantity.** Not even the most sophisticated analytics can overcome measurement errors.
- **Measure only what's useful to you.** It's best to have a handful of good instruments, positioned in locations that are meaningful, and keep those sensors running well.
- **Think dynamics, not steady state.** Due to daily variations in measured parameters, there is a need to measure and analyze the dynamics.

- **Recognize different timescales.** Hand in hand with dynamics is the need to think about different timescales: diurnal (daily) variations, weekly trends (especially weekend versus weekday differences), and seasonal shifts.
- **Consider how to handle outliers and extraordinary events.** Outliers need to be identified, and dealt with appropriately rather than just ignore them.

Software Platforms and Integration

The software available to the water industry can be very application specific (Theo, 2017). This requires integration at the system level to provide a holistic, utility-wide view of the system. Some examples of software products that target specific applications in the water industry are provided in Liner and Kenel, 2016. Utilities typically integrate these products as required.

These type of software products often utilize more than one data source to reduce errors. As an example, sensor data might to be correlated with maintenance activities from the CMM to remove false positives that might be identified by analysis of just sensor data when maintenance activities are being undertaken.

Big Data Analytics for the Water Industry

Potential applications of Big Data analytics to the water industry were discussed in various papers and presentations. The 2015 AWWA State of the Water Industry survey (AWWA, 2015) reported that 48 percent of respondents said their utilities are in various stages of exploration, implementation, or operation of Big Data systems.

Examples of applications that Big Data analytics and IoT can help with have been identified in multiple references, with most of these applications appearing in more than one of them (Conner, 2016; Water Online, 2016; Westerling, 2017; AWWA, 2014; Cohen, 2017; Li, 2015; WEF, 2016). A summary of these applications is presented in Table 3.

The applications listed above were likely developed to target the top challenges facing water utilities. The annual 2017 AWWA State of the Water Industry Report (AWWA, 2017) presented the top ten issues for the water industry for the previous 4 years, according to its annual surveys. These are presented in Table 3, in the order for 2017, with an indication of the big data and IoT applications available to address them.

Table 3. Water Industry Issues and Poential Big Data/IoT Applications

Water Industry Top Ten Issues	Big Data and IoT Applications for the Water Industry
Renewal and replacement of aging water and wastewater infrastructure	• Management of aging water infrastructure • Support for asset management • Support for increasing asset and operational efficiency
Financing for capital improvements	• Prioritization of capital expenditure • Identification of maintenance options
Long-term water availability	• Support for managing water scarcity
Public understanding of the value of water systems and services	• Provision of real-time insights for demand management • Management of energy consumption • Support for asset management • Support for increasing asset and operational efficiency • Communication with consumers and support for actively enlisting their participation

	• Monitoring of water quality in real-time • Support for addressing challenges of an aging workforce
Public understanding of the value of water resources	• Provision of real-time insights for demand management • Communication with consumers and support for actively enlisting their participation
Watershed/source water protection	
Emergency preparedness	
Cost recovery (pricing water to accurately reflect its true cost)	• Management of billing and the customer interface • Support for increasing asset and operational efficiency • Identification of maintenance options • Communication with consumers and support for actively enlisting their participation • Greater efficiency of regulatory compliance
Public acceptance of future water and wastewater rate increases	• Support for increasing asset and operational efficiency • Communication with consumers and support for actively enlisting their participation
Water conservation/efficiency	• Identification of leakage and non-revenue water • Provision of real-time insights for demand management

Additionally, examples of smart technologies (Westerling, 2017) that can help water utilities to become more proactive than reactive when addressing challenges include:

- Using video analytics to automate inspection of corroding pipes, enabling predictive maintenance
- Analyzing data in real time to identify leaks that would otherwise go unnoticed
- Leveraging software to help utilities and consumers track home water usage

Common Barriers for Water Utilities

Several common themes on the barriers for Big Data and IoT technologies in water utilities are mentioned in the literature (Liner and Kenel, 2016; Westerling, 2017; Liner, 2016; Grady, 2016; Grilo, 2016; Cohen, 2017), which include:

- Siloed communication within the utility
- Lack of resources and expertise
- Culture tends to be not open to change
- Tight budgets
- Justifying return on investment
- The challenge of managing the scope, scale, and security of smart systems
- The long RFP cycle where technologies can become out of date
- Water is too cheap, prohibiting a lot of investment in the area

Most water utilities are very small, with the EPA estimating about 8 percent of the 52,000 water utilities across the United States serve 82 percent of the population (Grady, 2016). Larger systems can more readily employ advanced metering infrastructure, data management software, and leak detection using the latest sensor technology because they have larger revenue.

These barriers have been overcome by many of the larger utilities, although others have found the culture to be the greatest barrier.

Data Security and Governance Concerns

The collection and analysis of data by water utilities raises legal and ethical questions. The main questions relate to who owns the data, who can profit from it, and what rights consumers have to the data (Noonan, 2014; Westerling, 2017; Grilo, 2016). One approach proposed is to share data collected with customers to foster transparency and empowerment (Noonan, 2014).

Suggestions from other industries (HBR, 2014) include clarifying who owns the data by the use of written contracts to clearly allocate ownership of data and derivative works.

Summary

Lessons learned from use of IoT and Big Data analytics by a variety of industries can be applied to the water industry. Key fundamentals, foundations, and processes can be leveraged to enhance the success of this effort. Water utilities interested in pursuing Big Data management to address top challenges can learn from not only the non-water sector but also innovators within the water sector. Overall, this work will be beneficial to all water utilities as technologies and platforms are constantly evolving and IoT is growing exponentially.

Many commonalities about the need for data quality can be found among industries. The goal is to evaluate the best quality data from reliable sources to create information and knowledge. The basis for the approaches to do so suggested in the water industry literature and general literature are similar, with both recommending a common approach of trying to build small solutions in an experimental manner before transitioning to production.

References

The general industry references and water sector references have been separated below. This separation will help the reader find industry specific references.

General Industry References

Abel, J. 2015. Transforming the Enterprise with Connected Data and Operational Intelligence. ARC view

https://www.arcweb.com/blog/transforming-enterprise-connected-data-and-operational-intelligence

Forrester. 2018. Predictions 2018: The Honeymoon For AI Is Over

https://www.forrester.com/report/Predictions+2018+The+Honeymoon+For+AI+Is+Over/-/E-RES139744?utm_source=forbes&utm_medium=pr&utm_campaign=2018_predictions&utm_content=predictions_ai

Gartner. IT Glossary. Definition for "Internet of Things". Accessed February 2018. https://www.gartner.com/it-glossary/internet-of-things/

Gartner. 2016b. Market Guide for IoT Platforms. https://www.gartner.com/doc/3380746?ref=SiteSearch&sthkw=big%20data&fnl=search&srcId=1-3478922254

Gartner. 2016a. Mastering Data and Analytics: 'Table Stakes' for Digital Business. https://www.gartner.com/doc/3343918?ref=SiteSearch&sthkw=big%20data&fnl=search&srcId=1-3478922254

Gartner. 2016. Organizing for Big Data Through Better Process and Governance. https://www.gartner.com/doc/3002918?ref=SiteSearch&sthkw=big%20data&fnl=search&srcId=1-3478922254

HBR. 2014. From the HBR.org Insight Center "From Data to Action". Harvard Business Publishing. https://kupdf.com/download/hbr-from-data-to-action-107218_58ac23dc6454a7c34fb1e8ee_pdf

Kushmaro, P. 2017. The IoT and Big Data: Making The Connection. Huffington Post. https://www.huffingtonpost.com/philip-kushmaro/the-iot-and-big-data-maki_b_12116608.html

McLellan, C. 2015. The internet of things and Big Data: Unlocking the power. ZDNet. http://www.zdnet.com/article/the-internet-of-things-and-big-data-unlocking-the-power/

NIST. 2015. NIST Special Publication 1500-1. NIST Big Data Interoperability Framework: Volume 1, Definitions. http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf

Water Industry References

AWWA. 2014. AWWA Webinar Program: Smarter Water Management. https://www.awwa.org/portals/0/files/education/webcast/smarterwatermanagementhandouts.pdf

AWWA. 2015. 2015 AWWA State of the Water Industry Report. https://www.awwa.org/Portals/0/files/resources/water%20utility%20management/sotwi/2015-AWWA-State-of-the-Water-Industry-Report.pdf

AWWA. 2017. 2017 State of the Water Industry Report. https://www.awwa.org/resources-tools/water-and-wastewater-utility-management/state-of-the-water-industry.aspx

Cohen, S. 2017. SWAN Blog: Lessons from the SWAN 2017 Conference. SWAN Forum. https://www.swan-forum.com/blog/

Conner, J.N. 2016. Water Reuse – The Need for Predictive Analytics. AWWA. http://www.awwa.org/portals/0/files/education/conferences/membrane/mtc16/membrane%20papers/wed08%20papers/03-conner%20paper.pdfGrady, B. 2016. Can Big Data save our water infrastructure? GreenBiz. https://www.greenbiz.com/article/can-big-data-save-our-water-infrastructure

Grilo, J. 2016. The Impact of ICT in the Water Sector. International Water Association. http://www.iwa-network.org/the-impact-of-ict-in-the-water-sector/

Li, H. 2015. Wise Use of Water With Smart Technologies. International Water Association. http://www.iwa-network.org/wise-use-of-water-with-smart-technologies/

Liner, B. 2016. Big Data Making Waves in Utility and Investment Communities. http://www.ncsafewater.org/news/298363/

Liner, B. and Kenel, P. 2016. Exciting future for Big Data solutions. Water Environment Federation. https://www.wef.org/globalassets/assets-wef/3---resources/topics/a-n/innovation/technical-resources/worldwaterbigdataapril2016.pdf

Noonan, C. 2014. Breaking through Big Data, analytics and the Internet of Things. Smart Cities Council. https://smartcitiescouncil.com/article/breaking-through-big-data-analytics-and-internet-things

Shaw, A. 2017. Understanding Big Data in the Water Industry. Water Online. https://www.wateronline.com/doc/understanding-big-data-in-the-water-industry-0002

Theo, G. 2017. Tapping into data: The analytics revolution for the business and customer. SWAN Forum. https://www.swan-forum.com/wp-content/uploads/sites/218/2017/05/SWAN-2017_Keynote-Unitywater_George-Theo.pdf

Water Online 2016. Surging 28%, US Municipal Water Infrastructure Forecast Reaches $532B. Water Online. https://www.wateronline.com/doc/surging-us-municipal-water-infrastructure-forecast-reaches-b-0001

WEF. 2016. Intelligent Water Systems: A Smart Start. Water Environment Federation. https://wef.org/globalassets/assets-wef/3---resources/online-education/webcasts/presentation-handouts/iws-webcast---presentation-handouts.pdf

Westerling, K. 2017. Smart Water Questions Answered. Water Online. https://www.wateronline.com/doc/smart-water-questions-answered-0001

APPENDIX C

Task 3 – Survey Large Firms in Non-Water Sector

Leveraging Other Industries – Big Data Management

Task 3 – Survey Large Firms in the Non-Water Sector

PREPARED FOR:	Water Environment and Reuse Foundation
PREPARED BY:	CH2M
DATE:	January 16, 2018
PROJECT NUMBER:	SENG7R16

Task 3 – Survey Large Firms in the Non-Water Sector sought to determine how non-water sector organizations use Internet of Things (IoT) technology and Big Data analytics. The purpose of the survey was also to understand whether investments in IoT technology and Big Data analytics are considered important and the primary applications used. Firms were identified that were known to the researchers, and the survey was administered by personal contact to increase the response rate. This Technical Memorandum provides a summary of the survey development process and results of the survey.

Approach

A set of survey questions was developed to gather the information about how the non-water sector is using IoT technology and Big Data analytics. These questions are presented in the Attachment.

The survey was divided into the following sections:

- Organizational information (three questions)
- Status and future willingness of Big Data and IoT implementation (four questions)
- Big Data and IoT adoption (five questions)
- Big Data and IoT results achieved (three questions)
- Impediment factors (one question)

Survey Results

The survey had nine respondents. Details and analysis of the survey results are presented in this document under the sections of the survey.

Organizational Information

Respondents to the survey were asked what industry they operated in, the size of the organization, and the role of the person completing the survey.

Organization Size

Organization size was identified within the following groups:

- Less than 49 employees
- 50 to 199 employees
- 200 to 499 employees
- Greater than 500 employees

The results presented in Figure 1 show that five of the responding organizations were organizations of over 500 employees and four were of under 500 employees.

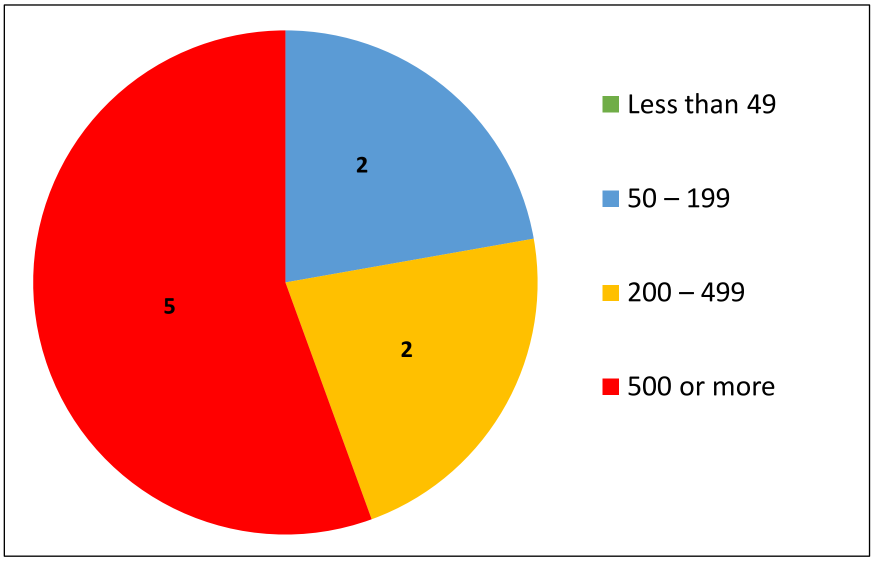

Figure 1. Organization Size

Industry

Respondents were asked in which of the following industries they worked. Some examples were provided; however, as it is not possible to identify all possible industries, the option for "Other" was included. The list presented the following options:

- Consulting
- Financial services
- Healthcare/life sciences
- Media/entertainment
- Retail
- Telecommunications
- Other

Survey responses are shown in Figure 2. Industries represented included consulting, financial services, retail, telecommunications, and three identifying as "Other." Other included instrumentation design, food processing technology, and logistics.

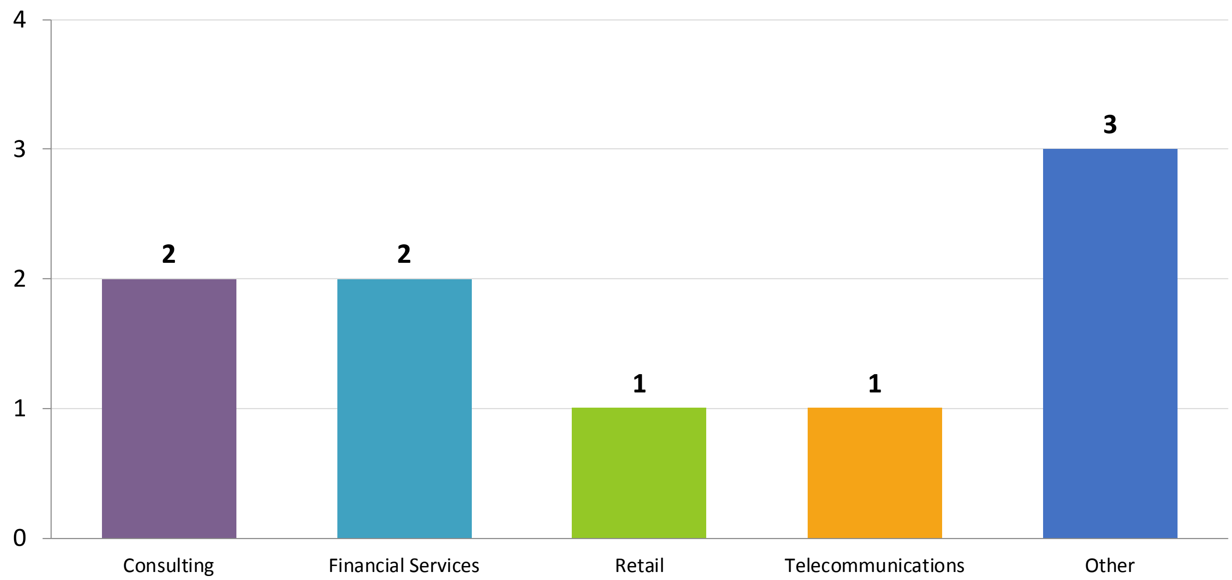

Figure 2. Industry

Organizational Role

This question was intended to identify the organizational level of the person completing the survey. Responses included Logistic Specialist, Consultant, Engineer, Data Analyst, Head of Big Data, Head of Product Management and Engineering Development, and Assistant Vice President Investment/Research.

Status and Future Willingness of Big Data/IoT Implementation

This section of the survey focused on the components that have been implemented to store and process Big Data and IoT data, and the near-term plans the company may have.

If your company has implemented components to store and process Big Data, when was it implemented?

The following five options were provided for this question:

- Within the past year
- Two years ago
- Three years ago
- More than 4 years ago
- Not implemented

The results are presented in Figure 3. Two respondents have not implemented components to store and process Big Data. Of the seven who have implemented these components, one was more than 3 years ago, and the other six were more than 4 years ago.

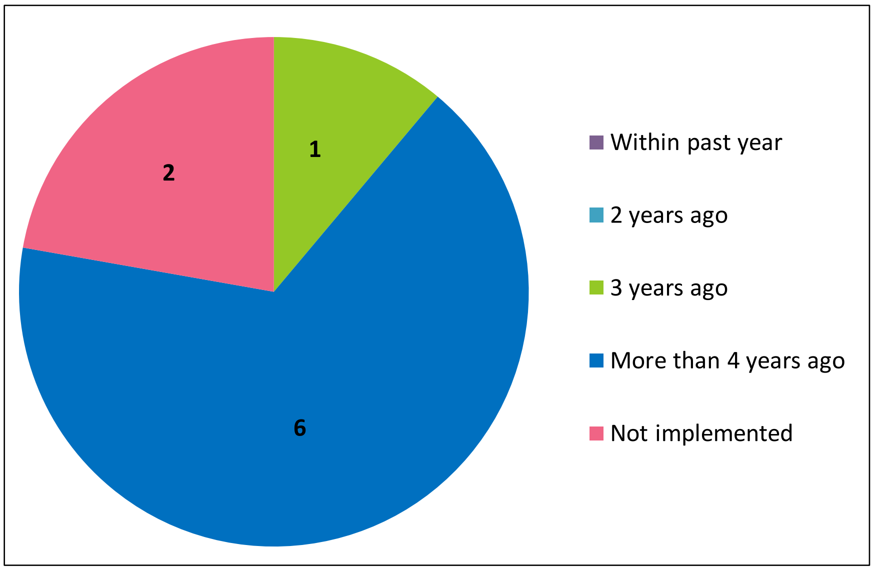

Figure 3. Implementation of Components to Store and Process Big Data

If your company has implemented components to store and process data produced by Internet of Things (IoT), when was this implemented?

The following five options were provided for this question:

- Within the past year
- Two years ago
- Three years ago
- More than 4 years ago
- Not implemented

The results are presented in Figure 4. Four respondents have not implemented components to store and process IoT data. Of the five who have implemented these components, one was more than 3 years ago, and four were more than 4 years ago.

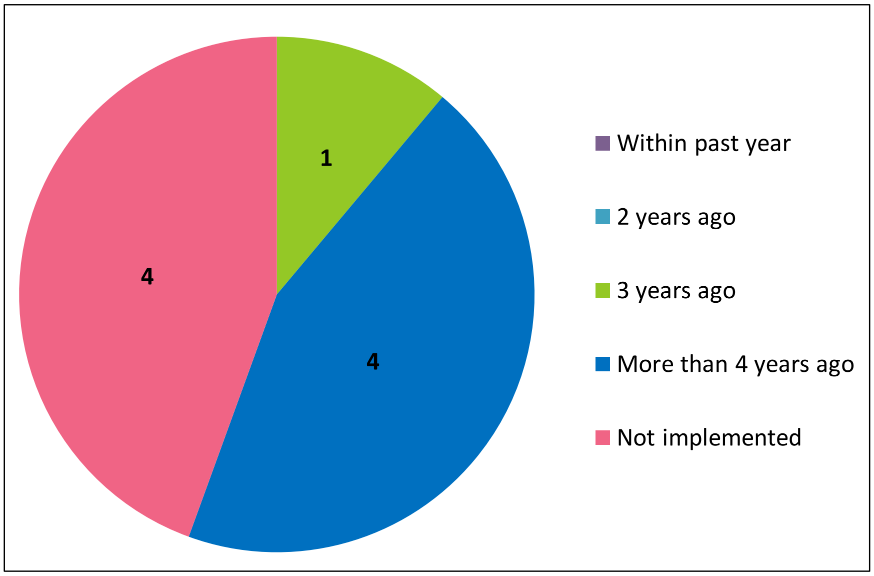

Figure 4. Implementation of Components to Store and Process IoT Data

Does your company plan to have or increase investment on Big Data/IoT in the near term?

Figure 5 presents results that indicate two-thirds of respondents intend to invest in Big Data/IoT in the near term.

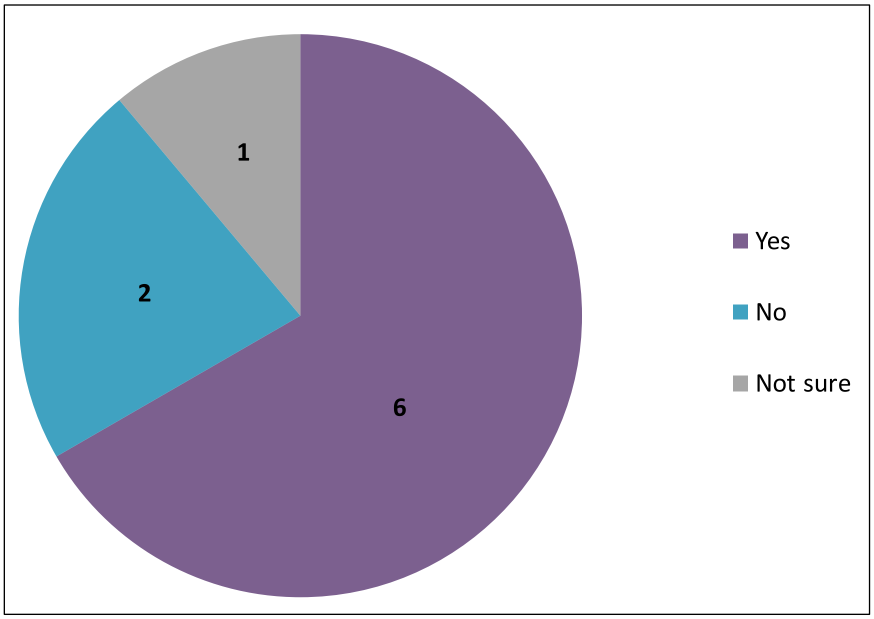

Figure 5. Plans for Big Data Investments

If YES, when do you see your company will have or increase investment on Big Data/IoT?

Of the seven respondents that answered "Yes" or "Not Sure" to the previous question, Figure 6 shows that six are expecting to invest in the next year, while one is planning to invest in 2 to 3 years (this is the organization that responded with "Not Sure" in the previous question).

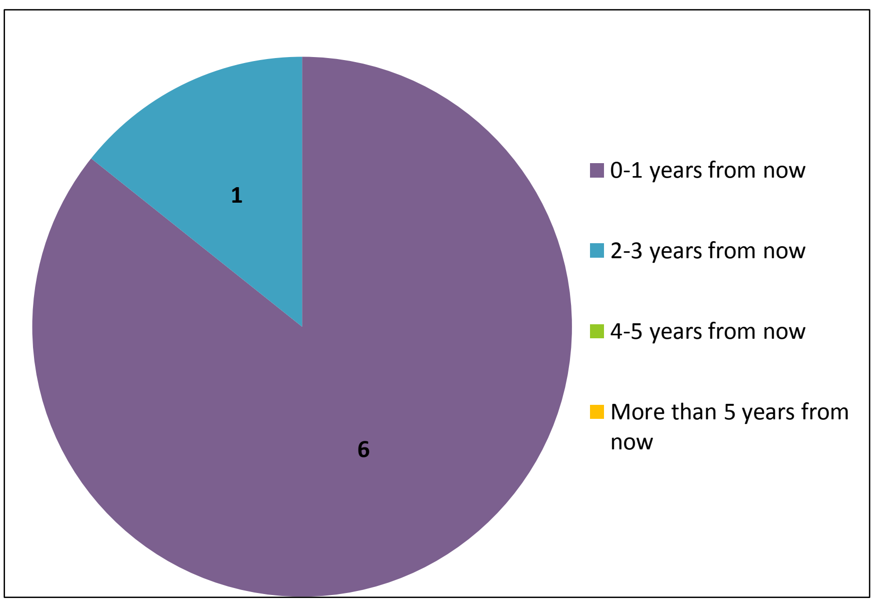

Figure 6. When Investment is Expected

If NO, what is the reason for not investing in Big Data/IoT?

Options were:

- We don’t see the need
- We don’t have the budget
- Insufficient internal capabilities
- Other reasons

Figure 7 indicates that, of the three respondents that answered “No” or “Not Sure” to the previous question, two do not have the budget (including the “Not Sure”) and one (a consulting firm) replied that there is no need identified to justify the investment.

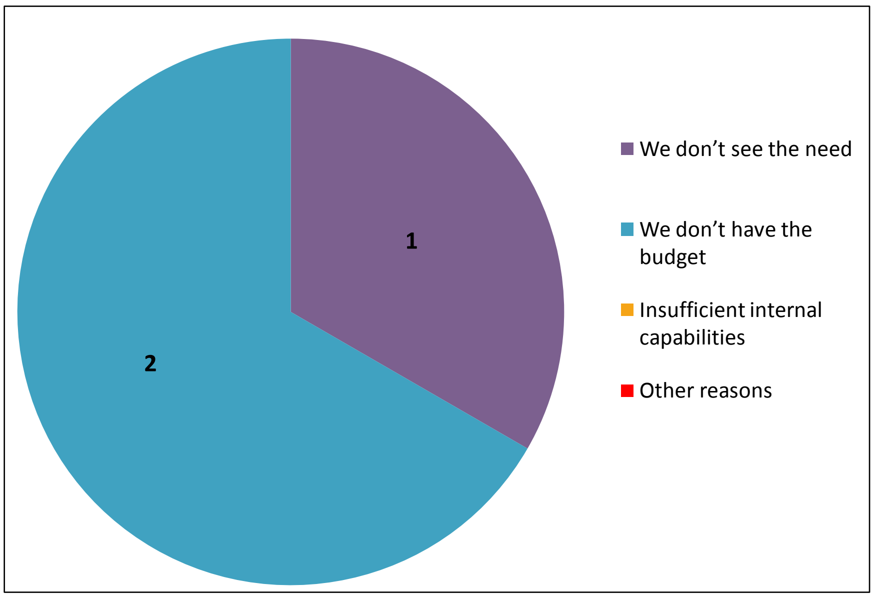

Figure 7. Reasons for Not Investing

How important is Big Data/IoT to your company?

Options were:

- It is the single most important way for us to maintain operations and reduce costs
- Is a top 5 issue that gets significant time and attention from top leadership
- Is important but is only one of many other challenges/opportunities that we need to address
- Not very important
- Don't know/NA

The results in Figure 8 show that only one of the respondents considers Big Data analysis to be the most important way to maintain operations and reduce costs. Four of the respondents identify it as a top five issue, and three identify it as important, but that it is only one of many things to be addressed.

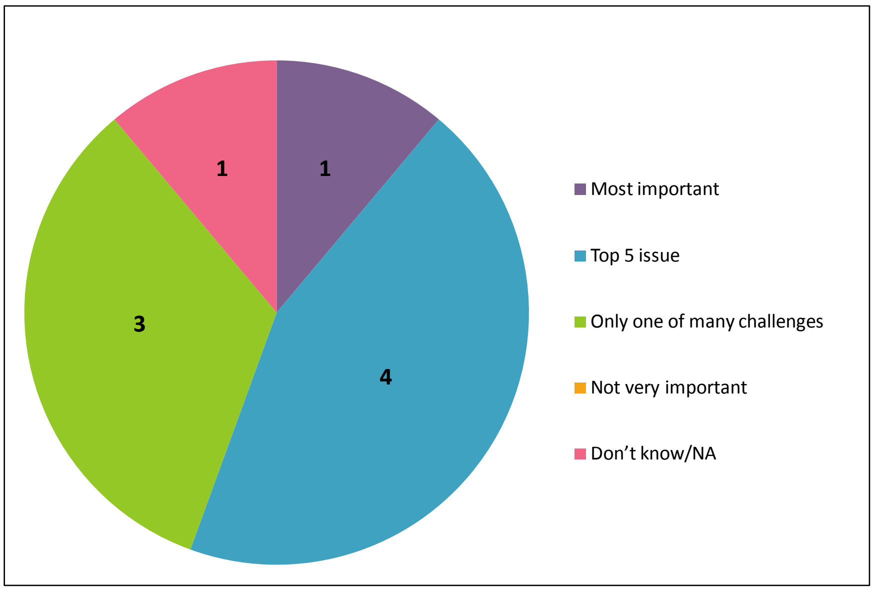

Figure 8. Importance of Big Data Analysis

Big Data/IoT Adoption

The questions for this section of the survey sought to understand the uses of Big Data and IoT within the company.

Which departments in your company have adopted Big Data/IoT?

This question asked the respondent to select all that apply from the following list:

- Accounting and finance
- Customer service
- Energy
- Finance
- Human resource management
- Logistics
- Marketing
- Production
- Research and development
- Supply chain
- Other
- We use outsourcing for Big Data/IoT analysis

Results are presented in Figure 9. Production and research and development were identified as the main users of Big Data and IoT analysis. Results for "Other" included "software development" and "don't know".

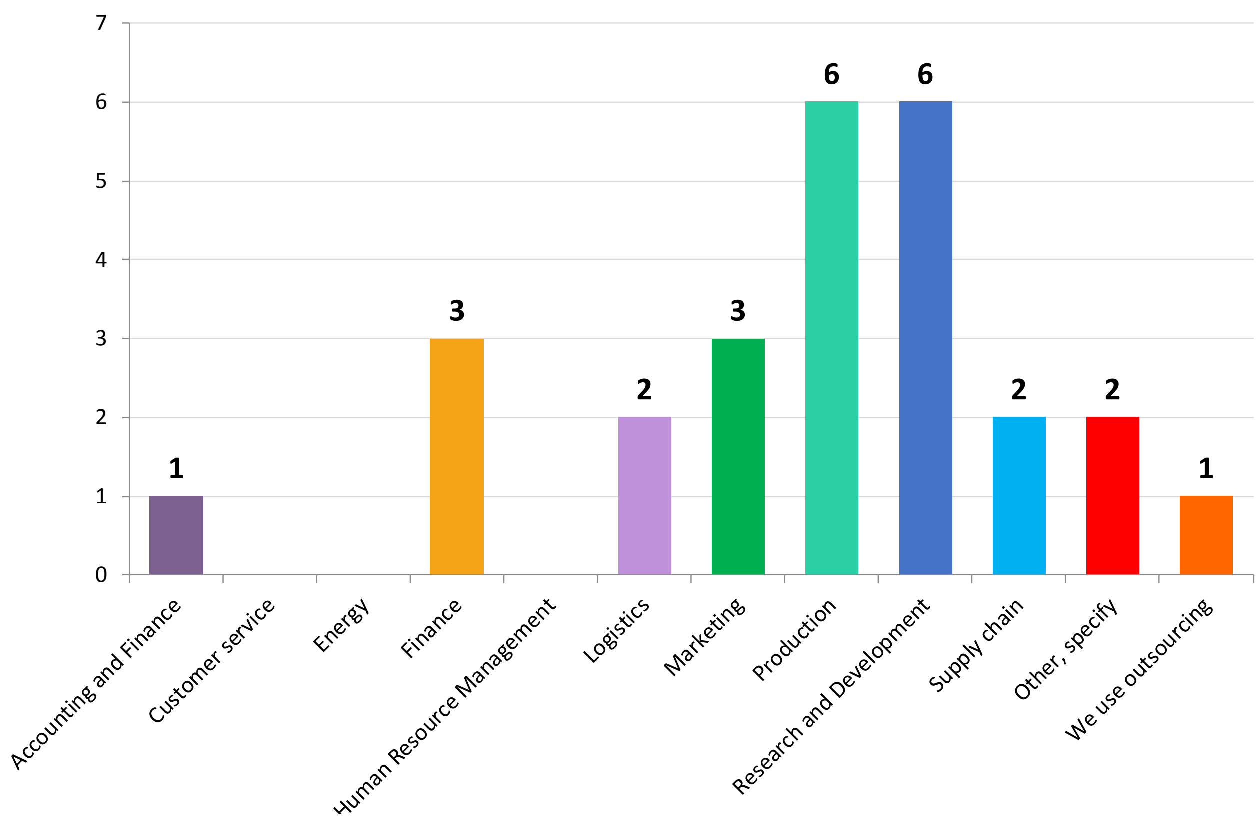

Figure 9. Departments that have Adopted Big Data/IoT

Which departments in your company do you think will adopt Big Data/IoT in the near future?

This question asked the respondent to select all that apply from the following list:

- Accounting and finance
- Customer service
- Energy
- Finance
- Human resource management
- Logistics
- Marketing
- Production
- Research and development
- Supply chain
- Other
- We will keep using outsourcing for Big Data/IoT analysis

Figure 10 provides a comparison of the results in Figure 9: the departments that have adopted Big Data/IoT (the solid color bars) and those that are expected to adopt Big Data/IoT in the near future (the patterned bars). Results for “Other” included “software development” and “don’t know”. These results indicate that each company identified different future pathways for the use of Big Data and IoT. There were no common themes across the respondents.

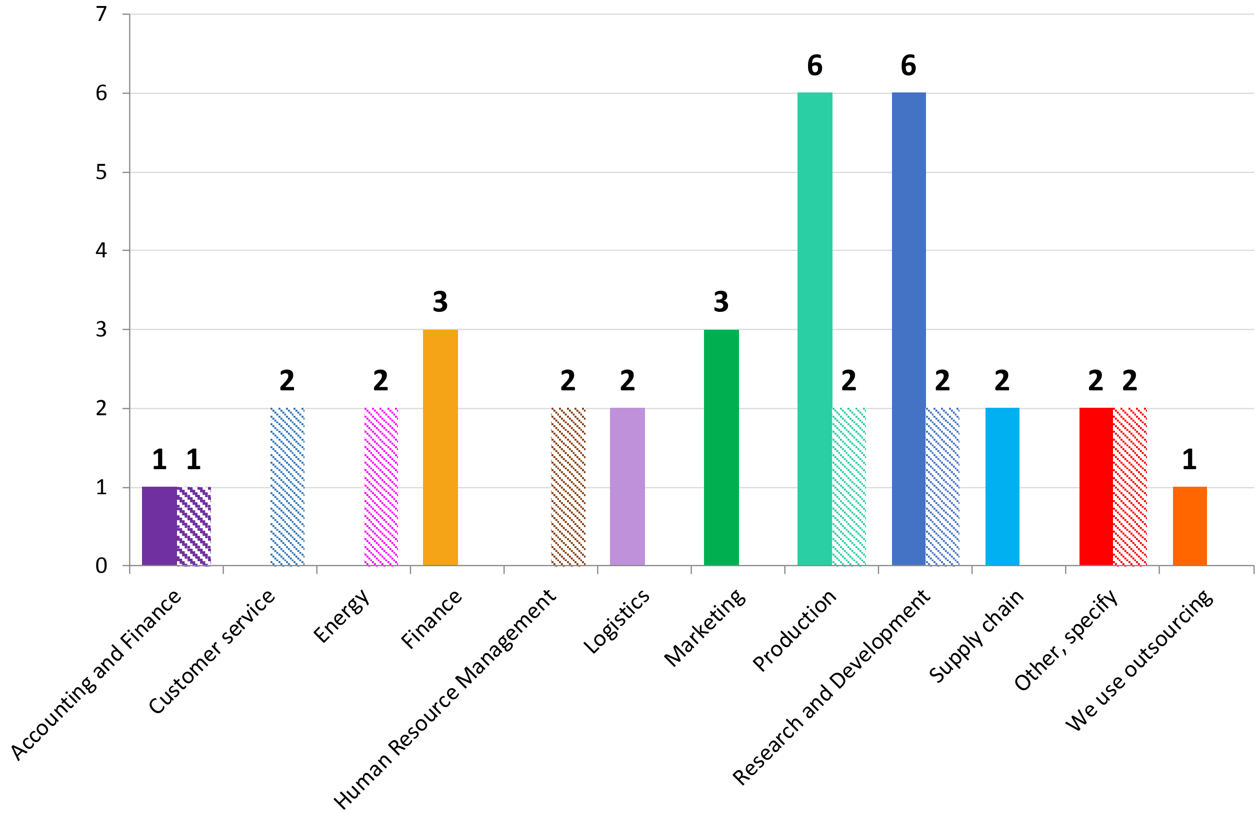

Figure 10. Departments That Have Adopted and Those That Are Likely to Adopt Big Data/IoT in the Near Future

How much has your company invested in Big Data/IoT?

This question asked the respondent to select from the following:

- Greater than $50 million
- $5 to $50 million
- $1 to $5 million
- Under $1 million
- Not available
- No investment

Results are presented in Figure 11.

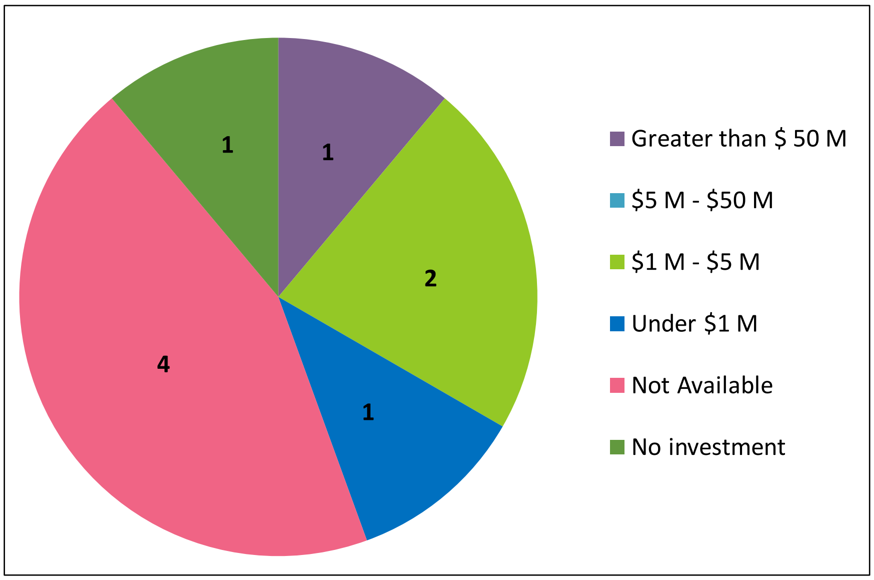

Figure 11. Investments in Big Data

Which functions of Big Data/IoT are used in your company?

This question asked the respondent to select all that apply from the following list:

- Monitoring and sensing physical quantity and quality
- Tracking physical movement
- Targeting customers (for advertising by marketers)
- Dynamic response to product demands
- Real-time optimization of manufacturing production and supply-chain networks
- Monitoring and controlling operations of infrastructure
- Market projection
- Energy and resource management
- Dynamic interaction between different components (i.e., the vehicle, infrastructure, driver or user)
- Other, specify
- Not applicable

The results indicate that monitoring and sensing physical quantity and quality, and targeting customers are the most common uses of Big Data and IoT by the respondents

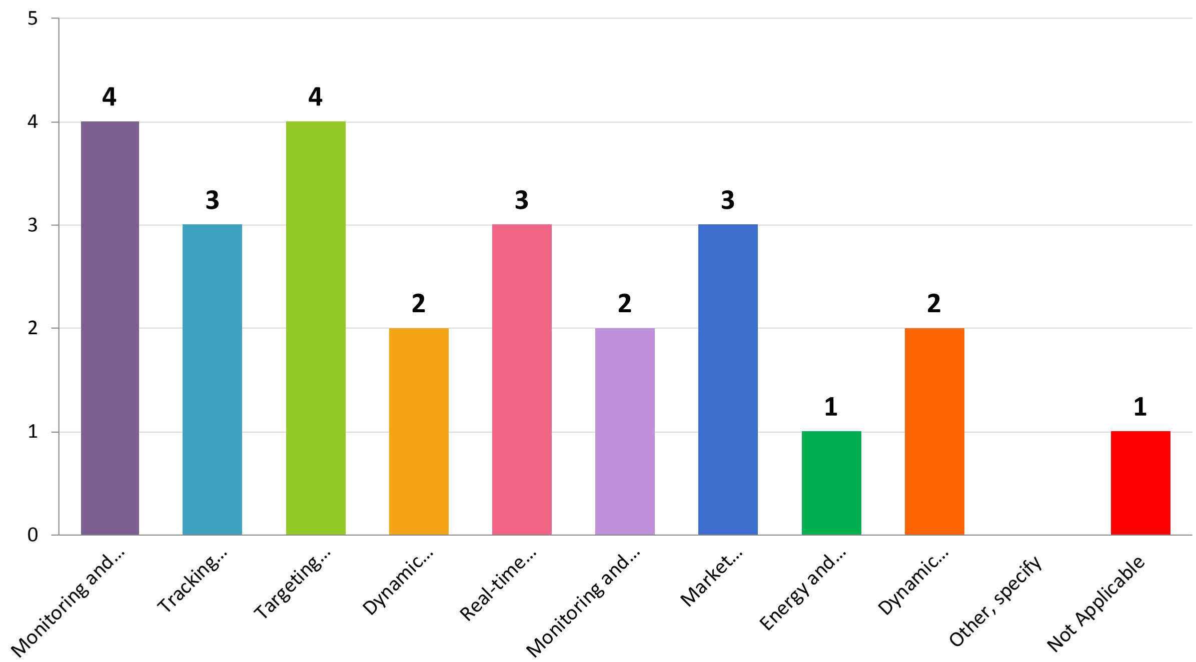

Figure 12. Functions of Big Data/IoT used Currently

Please rate the level of skills available within your company for each of the following

This question asked the respondent to rate the skills in the following list from 1 to 5 (1 as no skills available; 5 as advanced skills available):

- Data engineering (infrastructure/architecture to support data scientists)
- NoSQL for operational data storage
- Statistical and quantitative analysis
- Machine learning
- Stream processing
- Data visualization
- R programming
- Python programming
- Scala programming

The ratings were adjusted to a Low/Medium/High scale where 1 and 2 were combined to be "Low," 3 was made "Medium," and 4 and 5 were combined to be "High."

Results using the Low/Medium/High rating scale are presented in Table 1. Multiple respondents did not rate all skills in the table with no explanation. Respondents considered that they had advanced skills in data engineering, statistical and quantitative analysis, and stream processing, with low skills in R programming and Scala programming.

Table 1. Level of Skills Currently Available

	Low	Medium	High
Data engineering (infrastructure/architecture to support data scientists)	3		5
NoSQL for operational data storage	2	1	4
Statistical and quantitative analysis	1	3	5
Machine learning	2	1	4
Stream processing	2		5
Data visualization	1	3	3
R programming	5		1
Python programming	4		3
Scala programming	5		1

Big Data/IoT Results

This section sought to identify the positive results achieved by the respondents' implementation of Big Data and IoT analysis.

Can your company quantify results from Big Data/IoT?

These answers were categorized as follows:

- Yes, estimated value is $______
- No
- Not Applicable

The results presented in Figure 13 indicate that four of the respondents marked "Not Applicable." Of these four companies, three of them have invested in Big Data and IoT, and intend to invest further (based on responses to previous questions). Of the three "No" responses, two of these companies have invested in in the technologies in the past.

The two "Yes" responses were accompanied with the values $20M (from a consulting company) and $100M+ (from a telecommunications company).

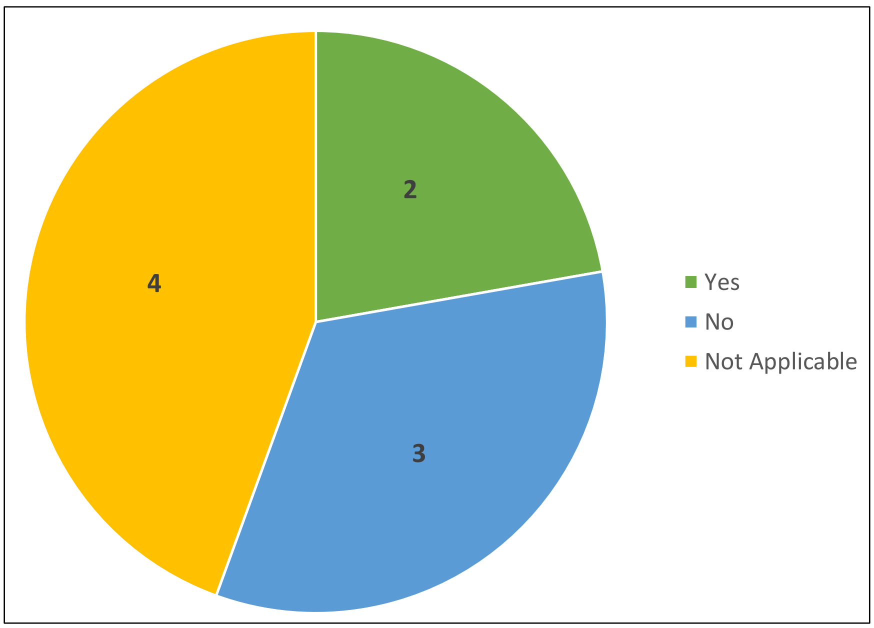

Figure 13. Quantification of Results from Big Data/IoT

In which areas have your company received measurable results from Big Data/IoT?

This question asked the respondent to select all that apply from the following list:

- Decrease expenses
- Establish a data-driven culture
- Accelerate the speed with which new capabilities and services are deployed
- Improve customer service
- Better understand customer trends
- Identification of failures
- Launch new product and service offerings
- Expand market share
- Other, specify
- Not applicable

Results are presented in Figure 14. Establishing a data-driven culture was selected most often, with decreasing expenses and improving customer service close behind.

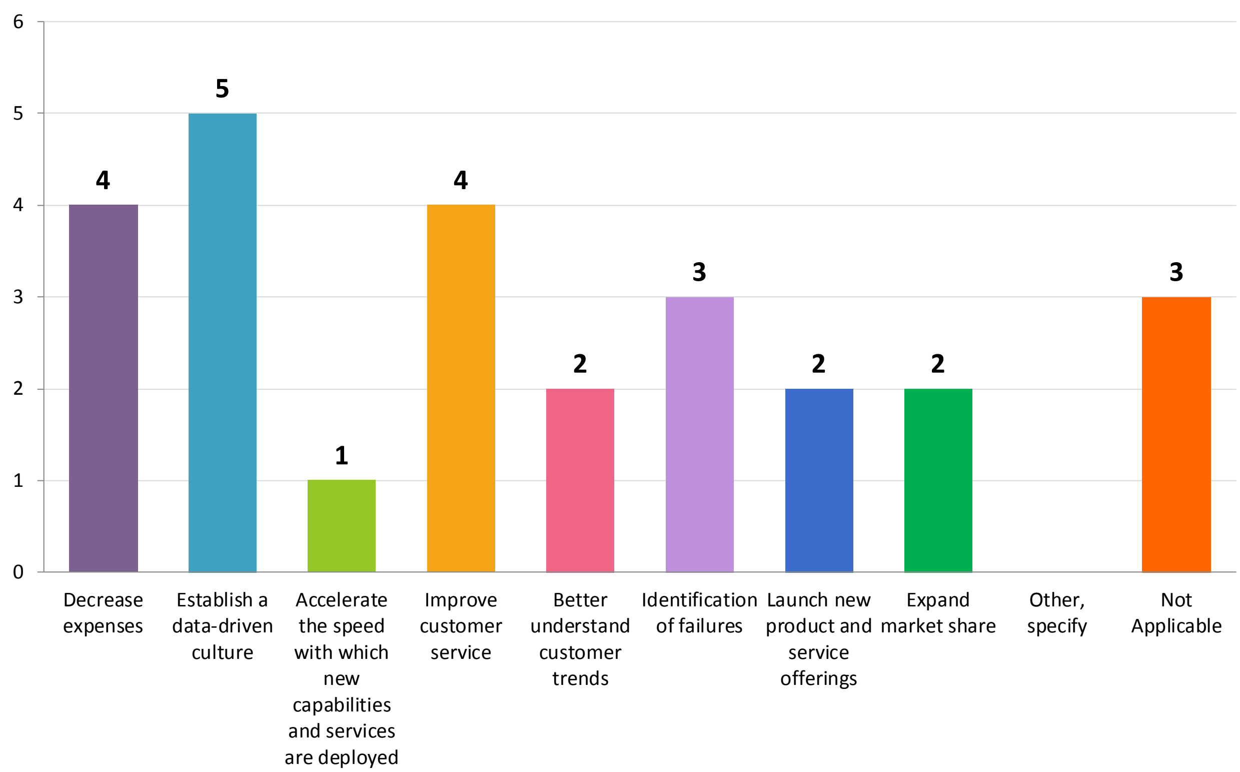

Figure 14. Areas that have Received Measurable Results from Big Data/IoT

Please rate the level of Big Data's impact on each of the following areas of your company in the next five years

Ratings were 1 to 5, where 1 = no impact; 5 = will have a dramatic impact. The areas to be rated were as follows:

- Impacting customer relationships
- Changing the way we organize operations
- Making the business more data-focused

The ratings were adjusted to a Low/Medium/High scale where 1 and 2 were combined to be "Low," 3 became "Medium," and 4 and 5 were combined to be "High."

Results using the Low/Medium/High rating scale presented in Table 2 indicate that making the business more data focused ranks the highest of the three areas.

Table 2. Impact of Big Data

	Low	Medium	High
Impacting customer relationships	2	3	3
Changing the way we organize operations	1	4	4
Making the business more data-focused	1	1	6

Impediment Factors

This section had one question about impediments to adoption of Big Data analysis within a utility.

Rate the level of influence that each of the following impediment factors have on Big Data analysis adoption in your company

Ratings were 1 to 5, where 1 = not influential and 5 = most influential. The factors to be rated were as follows:

- Data security
- Data quality
- Lack of budget
- Lack of talent to implement Big Data
- Lack of talent to run Big Data processing and analytics on an ongoing basis
- Resistance to integrate existing systems
- Procurement limitations on Big Data vendors
- Lack of middle management adoption and understanding
- Lack of data governance policies and practices

The ratings were adjusted to a Low/Medium/High scale where 1 and 2 were combined to be "Low," 3 became "Medium," and 4 and 5 were combined to be "High."

Results using the Low/Medium/High rating scale are presented in Table 3. The results indicate the following:

- Impediments considered most influential are data security and data quality.
- Lack of budget and lack of talent to implement Big Data are not seen as influential.

Table 3. Impediment Factors to Big Data Adoption

	Low	Medium	High
Data security	3		5
Data quality	2	1	5
Lack of budget	6		2
Lack of talent to implement Big Data	6		1
Lack of talent to run Big Data and analytics on an ongoing basis	4	1	2
Integration resistance with existing systems	5		2
Procurement limitations on Big Data vendors	5	2	
Lack of middle management adoption and understanding	4	2	2
Lack of data governance policies and practices	5		2

Summary/Conclusions

The survey had nine respondents covering multiple industries. Summary conclusions include the following:

- Components to store and process Big Data and IoT data were either put in place more than 3 years ago or not implemented, with IoT implemented in fewer cases than Big Data.
- Two-thirds of respondents plan to invest in Big Data in the near term.
- None of the respondents considered Big Data analysis to be unimportant. Only one of the respondents considered it the single most important way to maintain operations and reduce costs. Four considered it a top five issue, while three consider it important, but only one of many things to be addressed.
- Two-thirds of respondents use Big Data/IoT in the production and research and development departments.
- Monitoring and sensing physical quantity and quality, and targeting customers are the primary functions for which Big Data and IoT are used.
- Respondents considered that they had advanced skills in data engineering, statistical and quantitative analysis, and stream processing, with few skills in R programming and Scala programming.
- Measurable results were only seen for establishing a data-driven culture.
- Impediments to the use of Big Data include data security and data quality.

Attachment--
Survey Non-Water Sector

- Industry:
  - ☐ Consulting
  - ☐ Financial Services
  - ☐ Healthcare /Life Sciences
  - ☐ Media/Entertainment
  - ☐ Retail
  - ☐ Telecommunications
  - ☐ Other (transportation, shipping, maintenance monitoring (e.g., GE aircraft engines), etc.)

- Organizational Role:
  - ☐ Chief Data Officer
  - ☐ Chief Analytics Officer
  - ☐ Chief Information Officer
  - ☐ Head of Big Data
  - ☐ CEO/President
  - ☐ Chief Marketing Officer
  - ☐ Other, specify

Status & Future Willingness of Big Data/IoT Implementation

- If your company has implemented components to store and process Big Data, when was it implemented?
  - ☐ Within past year
  - ☐ 2 years ago
  - ☐ 3 years ago
  - ☐ More than 4 years ago
  - ☐ Not implemented.

- If your company has implemented components to store and process data produced by the Internet of Things (IoT), when was this implemented?
  - ☐ Within past year
  - ☐ 2 years ago
  - ☐ 3 years ago
  - ☐ More than 4 years ago
  - ☐ Not implemented.

- Does your company plan to have or increase investment on Big Data/IoT in the near term?
  - ☐ Yes
  - ☐ No

  If YES, when do you see your company will have or increase investment on Big Data/IoT?
  - ☐ 0-1 years from now
  - ☐ 2-3 years from now
  - ☐ 4-5 years from now
  - ☐ More than 5 years from now

  If NO, what's the reason of not investing on Big Data/IoT?
  - ☐ We don't see the need
  - ☐ We don't have the budget
  - ☐ Insufficient internal capabilities
  - ☐ Other reasons

- How important is Big Data/IoT to your company?
  - ☐ Is single most important way for us to gain competitive advantage
  - ☐ Is a top 5 issue that gets significant time and attention from top leadership
  - ☐ Is important but is only one of many other challenges/opportunities that we need to address
  - ☐ Not very important
  - ☐ Don't know/NA

Big Data /IoT Adoption

- Which departments in your company have adopted Big Data/IoT (select all that apply)?
  - ☐ Accounting and Finance
  - ☐ Customer service
  - ☐ Energy
  - ☐ Finance
  - ☐ Human Resource Management
  - ☐ Logistics
  - ☐ Marketing
  - ☐ Production
  - ☐ Research and development
  - ☐ Supply chain
  - ☐ Other, specify
  - ☐ We use outsourcing for Big Data /IoT analysis, specify

- Which departments in your company do you think will adopt Big Data/IoT in the near future (select all that apply)?
  - ☐ Accounting and Finance
  - ☐ Customer service
  - ☐ Energy
  - ☐ Finance
  - ☐ Human Resource Management
  - ☐ Logistics
  - ☐ Marketing
  - ☐ Production
  - ☐ Research and development
  - ☐ Supply chain
  - ☐ Other, specify
  - ☐ We will keep using outsourcing for Big Data /IoT analysis, specify

- How much has your company invested in Big Data/IoT?
  - ☐ Greater than $ 50 M
  - ☐ $5 M - $50 M
  - ☐ $1 M - $5 M
  - ☐ Under $1 M
  - ☐ Not Available
  - ☐ No investment

- Which functions of Big Data/IoT are used in your company (select all that apply)?
  - ☐ Monitoring and sensing physical quantity and quality
  - ☐ Tracking physical movement
  - ☐ Targeting customers (for advertising by marketers)
  - ☐ Dynamic response to product demands
  - ☐ Real-time optimization of manufacturing production and supply chain networks
  - ☐ Monitoring and controlling operations of infrastructure
  - ☐ Market projection
  - ☐ Energy and resource management
  - ☐ Dynamic interaction between different components (i.e. the vehicle, infrastructure, driver or user)
  - ☐ Other, specify
  - ☐ Not Applicable

- Please rate the level of skills available within your company for each of the following (1-5, 1 as no skills available; 5 as advanced skills available)?
  - Data Engineering (infrastructure/architecture to support data scientists) __
  - NoSQL for operational data storage __
  - Statistical and quantitative analysis __
  - Machine learning __
  - Stream processing __
  - Data visualization __
  - R programming __
  - Python programming __
  - Scala programming __

Big Data/IoT Results

- Can your company quantify results from Big Data/IoT?
  - ☐ Yes, estimated value is $______
  - ☐ No
  - ☐ Not Applicable

- In which areas have your company received measurable results from Big Data/IoT (select all that apply)?
  - ☐ Decrease expenses
  - ☐ Establish a data-driven culture
  - ☐ Accelerate the speed with which new capabilities and services are deployed
  - ☐ Improve customer service
  - ☐ Better understand customer trends
  - ☐ Identification of failures
  - ☐ Launch new product and service offerings
  - ☐ Expand market share
  - ☐ Other, specify
  - ☐ Not Applicable

- Please rate the level of Big Data's impact on each of the following areas of your company in the next five years (1-5, 1 as no impact; 5 as will have a dramatic impact)? *
  - Impacting customer relationships __
  - Changing the way we organize operations __
  - Making the business more data-focused __

Impediment Factors

- Rate the level of influence that each of the following impediment factors have on Big Data analysis adoption in your company (1-5, 1 as not influential; 5 as most influential)
    - Data security __
    - Data quality __
    - Lack of budget __
    - Lack of talent to implement big data __
    - Lack of talent to run big data and analytics on an ongoing basis __
    - Integration resistance with existing systems __
    - Procurement limitations on big data vendors __
    - Lack of middle management adoption and understanding __
    - Lack of data governance policies and practices __

APPENDIX D

Task 4 – Determine Trends and Future Technology Paths

Leveraging Other Industries – Big Data Management

Task 4 – Determine Trends and Future Technology Paths

PREPARED FOR:	Water Environment & Reuse Foundation
PREPARED BY:	CH2M
DATE:	April 4, 2018
PROJECT NUMBER:	SENG7R16

Task 4 – Determine Trends and Future Technology Paths proposed interviewing large Internet of Things (IoT) firms (such as Cisco Systems, Qualcomm, Intel, and Microsoft), reviewing research from IT research organizations (obtained during Task 2), and tracking up-and-coming open-source tools from university research groups to identify trends for IoT technology and Big Data analytics.

It became apparent early in this task that the information that could be gained from interviewing the firms proposed was already available in the public domain. The focus of this task was therefore re-aligned to the review of current literature addressing the topics, with this Technical Memorandum being the summary of this task.

Approach

A three-horizon review approach was used to determine the trends and future technology paths for IoT and data analytics. The three-horizon methodology was originally applied to the development of a growth and innovation strategy within companies. This methodology helped define how investments would be made within the company as its products and services matured, but also helped ensure that the company was not left with stagnant products. Inevitably, competitors would innovate and, without continued investment for innovation and growth, a company that didn't develop a growth and innovation strategy would be left behind in the market.

Another closely related concept used to identify future trends is the technology hype cycle promoted by Gartner (Gartner, 2018), a leading research, analysis, and advisory company to the Information Technology (IT) industry. The technology hype cycle follows a given set of related technologies from initial introduction through development, implementation, and refinement. The five phases in the hype cycle, as defined by Gartner, are shown in Figure 1:

1. Innovation Trigger. The cycle begins with a potential technology breakthrough, possibly including early proof-of-concept.
2. Peak of Inflated Expectations. Expectations are raised by early success stories.
3. Trough of Disillusionment. As early implementations fail to deliver, investment is reduced to only those technologies that can satisfy early adopters.
4. Slope of Enlightenment. Second- and third-generation products appear, and users, becoming more familiar with the technology, identify more applications.
5. Plateau of Productivity. Mainstream adoption starts to take off.

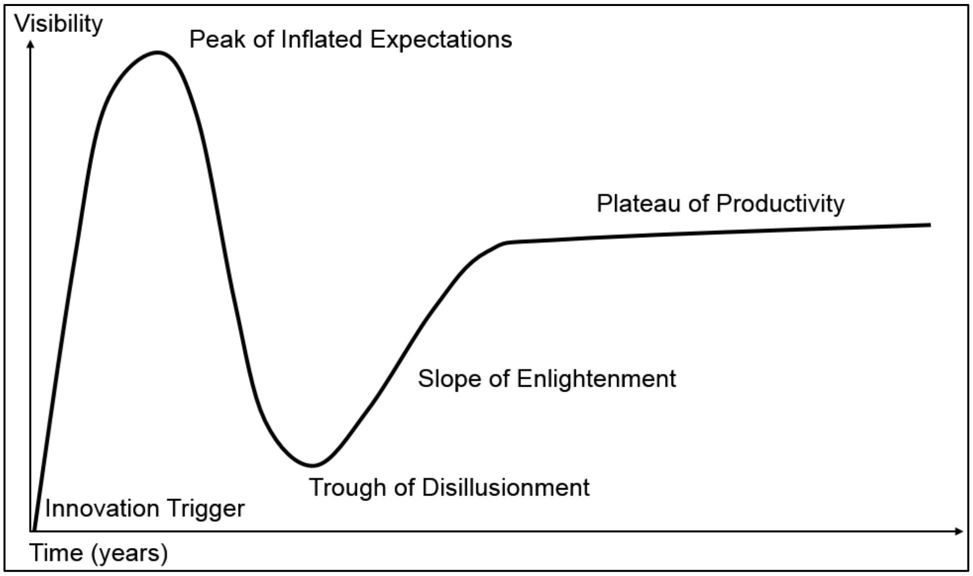

Figure 1: Technology Hype Cycle
(Adapted from Gartner, 2018)

While the hype cycle is forward looking on the maturation of given technologies, the three-horizon view (Baghai, Coley, and White, 1999) is somewhat of a reverse view of the hype cycle in that it provides insight into organizational priorities and investments. Horizon 1 (near-term products) focuses on the product roadmap along with research and development efforts to further the product applications. Horizon 2 (early adoption) builds upon the items outlined in Horizon 1 by augmenting them with IoT and data analytic trends. The longer-term research and development view (Horizon 3) applies research undertaken by key educational institutions and technology firms to determine the focus and growth areas for the future, long-term life of the product.

In the subsequent sections, content is aligned to these three horizons with two to three examples provided for each horizon that define the technology or solution and discuss the relevance of the technology.

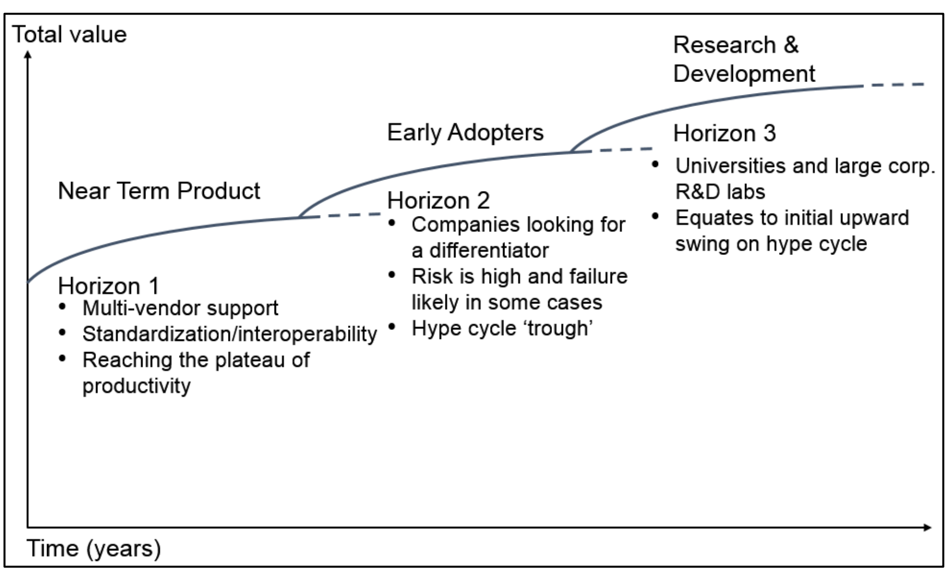

Figure 2: Three-Horizon Technology View
(Adapted from Baghai, Coley, and White, 1999)

Horizon 1 – Near-Term Products

The near term is seeing the adoption of 5G technology on the IoT front and the utilization of AutoML on the data analytics side. Apache Spark links IoT and analytics for real-time data stream processing from IoT devices.

5G

5G is the first generation of cellular networks with a focus on IoT devices, with IoT use cases included in the design of standards, chipsets, and carrier technology. The emphasis on IoT has led to the need for improvements in latency and energy efficiency over previous generation networks.

Definition

Gartner (Ray and Ramamoorthy, 2017) defines 5G as the next-generation cellular standard after 4G (including its variants Long-Term Evolution [LTE], LTE Advanced [LTE-A] and LTE-A Pro). The details of the 5G standard are currently being defined across several global standards bodies, with the official International Telecommunications Union (ITU) specification, International Mobile Telecommunications-2020 (IMT-2020), specifying a target maximum downlink and uplink throughputs of 20 gigabits per second and 1 gigabit per second, respectively, latency below 5 milliseconds, and massive scalability.

Relevance

3G and 4G cellular networks currently deployed limit high bandwidth IoT applications due to network constraints and transmission delays. Where low-power applications are implemented that require the IoT device to initiate a call, transmit data, and go to sleep, 3G and 4G networks require significant time and power to initiate the call. The proposed 5G network standards will reduce the time and power required to function.

The other area in which 5G provides improvement is that of energy efficiency. Many edge IoT solutions within the water sector are battery powered. Increasing the battery life in these devices also reduces the maintenance effort as battery replacement can be a timely (and, in some cases, impossible) endeavor on in-situ and infrastructure-based sensors. Utilization of the RF (radio frequency) section of an IoT device is often the most power-intensive operation. Improvements in RF efficiency, radio sleep modes, and data connection/transmission cycles allow efficiency gains with 5G. One of the design use cases for 5G efficiency was for a 10-year operational battery life on IoT devices.

All major cellular network providers worldwide have announced 5G rollouts, and some already have working pilots.

Conclusion

5G is a near-term technology intended to provide improved support for IoT devices. Technology providers of 5G include Cisco, Ericsson, Huawei Technologies, Intel, NEC, Nokia, and Qualcomm.

AutoML

One of the larger challenges facing data scientists is the challenge of what model to use when developing analytics. AutoML (automated machine learning) eases this effort by recommending the underlying model to use and providing a tuned model based on the data feed into the learning process.

Definition

AutoML supports algorithm selection and parameter tuning by interactively searching for the best machine learning model based on user-provided data and allowing iterative modifications.

Relevance

Data analytics continue to be used within the water sector. While there is typically good domain knowledge of the water and wastewater aspects of a given problem, often there is not a good understanding of the machine-learning algorithms and models that can be applied to help solve the given use case. Knowing when to use a gradient boosting versus a random forest or deep learning model is not a trivial exercise. AutoML assists in this selection.

AutoML is not a magical single “press here” button that solves the overall data analytics problem. A human must still be involved in gathering the data, cleansing the data, labelling and developing features within the data to properly feed the AutoML process, so AutoML will not replace data scientists.

Conclusion

AutoML is a near-term technology that will support the work of data scientists in the selection and tuning of the appropriate algorithms for use with a given dataset. AutoML is currently available from vendors such as Alpine Data, DataRobot, Google (cloud), H2O.ai, IBM, Microsoft (cloud), and Skytree, with open source examples including Auto-sklearn, Auto-Weka, Machine-JS, and TPOT.

Apache Spark

Apache Spark is different than the combination of Hadoop and MapReduce because it is able to process large volumes of data significantly faster. With over 200 contributors, Apache Spark’s development ecosystem is one of the largest within the Big Data open-source community.

Definition

Apache Spark is an open-source data processing framework. By utilizing in-memory processing, Apache Spark can more readily handle real-time data streams as part of an overall machine-learning workflow.

Relevance

One of the main drivers for using Apache Spark is to better process streaming or real-time data. With so much real-time data being produced in the water sector by IoT sensors and SCADA systems, it has become necessary for utilities to be able to process it all in real-time. Some believe that Apache Spark could be the de facto platform for stream processing because Spark is able to handle disparate data in its workflows, allowing developers to use a single framework without regard to data source.

Apache Spark is also useful in streaming ETL (extract, transform, load) workflows. Non-streaming ETL workflows use batch processing, which would store the transformed data in a data warehouse. This processing involved reading the data, converting it into a compatible format, and then writing it to the back-end database. With streaming ETL, the workflow becomes continuous instead of batched to decrease latency.

Conclusions

Apache Spark is a near-term technology for processing Big Data in real-time that is built on an open-source data processing framework. Commercial technology providers that support Apache Spark include Amazon Web Services, Cloudera, Databricks, DataStax, Hortonworks, IBM, MapR Technologies, and SAP.

Horizon 2 – Early Adoption

Horizon 2 solutions are ones that are commercially available, but may not be widely implemented within a given sector, such as the water industry. Two data analytic and machine learning trends will be important in the water sector: deep learning and machine learning for cybersecurity.

Deep Learning

The annual ImageNet Large Scale Visual Recognition Challenge (ILSVRC) has taken place for nearly a decade. This challenge allows research teams to test algorithms in a competitive environment against various visual recognition tasks. Beginning in 2010, computer algorithms have exponentially improved image recognition within the ILSVRC, primarily by using deep learning. Deep learning, which uses many multiple layers of neural networks rather than a shallow set of neural network layers, allows for a more complete classification or recognition model to be generated. Initial implementations for ILSVRC were shallow in nature with very few layers within the neural network. As the number of layers increased, the error rate for image recognition decreased (as shown in Figure 3).

In 2015, the number of layers used by the winning algorithm increased dramatically to 152 layers from less than eight in 2010 (He, et al., 2015). At this point, the winning algorithm had an error rate of less than 5 percent, which is a point where the algorithm exceeds human ability. The use and training of these highly layered neural networks is referred to as Deep Learning.

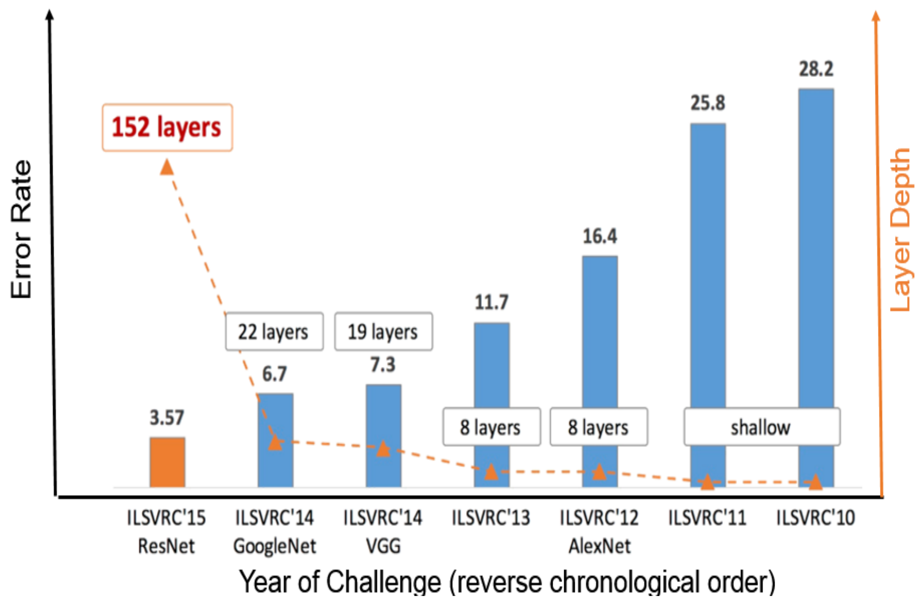

Figure 3: ImageNet Large Scale Visual Recognition Challenge Results
(Adapted from He, et al., 2015)

Definition

Deep learning is a type of machine learning that utilizes highly layered neural networks to provide a more accurate model that is capable of learning in an unsupervised manner from data that may be unstructured or unlabeled. Deep learning is also referred to as a Deep Neural Network (DNN).

Relevance

Three elements are driving the increased use of deep learning:

- Improvements and increased availability of high-end hardware and GPUs (graphical processing units) on premise and in the cloud
- Increased availability of training data
- Improved learning algorithms

Use of deep learning has increased in other industries (financial and insurance sectors as an example) to enable new capabilities from language recognition and near real-time, machine-driven language translation to video analytics and object classification.

Deep-learning capabilities are becoming more accessible in data science platforms, and some debate that they are expected to be a standard component in 80 percent of data scientists' tool boxes in 2018 (Vorhies, 2017).

DNNs are just now being applied within the water sector for use cases such as demand forecasting, water quality prediction, and root cause analysis (such as identification of operational conditions that lead to filter breakthrough).

Conclusions

Deep Learning is in the early adopter phase in certain sectors, with the technology being available by providers such as Amazon, Arimo, deepsense.io, Google, H2O.ai, IBM and Microsoft. Open source versions of the technology include Caffe, Deeplearning4j, TensorFlow, and Theano.

Machine Learning for Cybersecurity

The concepts of machine learning and deep learning are being applied to many areas, such as cybersecurity, due to the myriad of threat challenges to organizations daily. As cyber countermeasures are implemented within an organization, the threats change in response. Having learning capabilities integrated with cybersecurity solutions decreases the likelihood of cyber issues going undetected for an extended period of time.

Definition

Use of artificial intelligence systems to identify and respond to in-progress cyber-threats. These systems mimic the human immune system in being able to adapt to changing and varying threat vectors.

Relevance

Maintaining cybersecurity within an organization is a challenging process, and current technology is not winning the battle. CNN (Doran and Britton, 2018) estimates that cyber-attacks will cost the world $8 trillion by 2022. Much of an enterprise's cyber effort is focused on keeping the perimeter of its network from being penetrated by those on the outside. This strategy ignores the potential of an employee inadvertently opening an infected attachment in an email or bringing in a virus on a thumb drive. The inside threat is not addressed by perimeter network defense.

An example of a virus that has infiltrated many organizations are cryptocurrency mining bots, which typically are introduced into the internal network and rapidly install themselves on other computers connected to that network. These particular viruses do not cause any data loss or grab any sensitive information, but they wait until the computer is idle and begin using CPU cycles to mine cryptocurrency (a computation-intensive process that utilizes lengthy calculations to generate or mine cryptocurrency). Because the virus is not data destructive, it often goes undetected. Check Point, a leading cyber technology firm, estimates as many as 55 percent of organizations globally have been infected (Check Point, 2018). Security researchers at Wandera (Wandera, 2017), a mobile security technology firm, claim that cryptocurrency mining instances on mobile devices increased by 287 percent between October and November 2017.

A cybersecurity system that learns internal network behaviors in real-time and creates a learned model that is unique and specific to your network would be able to detect the abnormal traffic patterns generated by mining bots and other anomalous activity.

As water utilities becomes more and more connected to meet service and operational needs, the cyber threat increases. With the reporting of increased interest by hostile foreign entities attempting to infiltrate water utilities, water utilities should not only anticipate that their networks will be penetrated, but assume that it has happened. Machine-learning cybersecurity provides the ability to have dynamic protection that learns from the traffic it sees.

Conclusions

Machine learning for cybersecurity is in the early adoption phase with the technology provided commercially by Darktrace, deepinstinct, and Splunk.

Horizon 3 – Research and Development

The third horizon focuses on research related to data analytics and processing being undertaken by universities and research groups within large firms. The most relevant topics are Generative Adversarial Networks (GANs) on the data analytics front and Quantum Computing on the processing side. While these technologies are early in their development cycle, they are already being utilized and extended within the university space. As both areas are advancing at a rapid pace, these technologies may be hitting the market sooner than later. As these topics are still in the research stage, commercially available solutions are not listed.

Generative Adversarial Networks

One of the challenges within data analytics and machine learning is having enough data to drive the learning process. Generative Adversarial Networks (GANs) provide a unique way to reduce the time required for the learning process. The concept of a GAN was first introduced in research work out of the University of Montreal and has been dubbed as “the most interesting idea in the last 10 years in ML” by Facebook’s artificial intelligence research director Yann LeCun (Forbes, 2016).

Definition

A GAN is a composition of two neural networks—a generative network and a discriminative network—that work in tandem to provide artificial data that drives more complete learning within the discriminative network. In this composition of networks, the generative network is attempting to “fool” the discriminative network, while the discriminative network is attempting to build a more complete model.

Relevance

Within the water sector, analytics are often linked with physical assets and processes that operate in the real-world. While many utilities are collecting data through their SCADA systems or IoT sensors, the occurrence of operational events does not always happen on a regular basis. For instance, nitrification events may not occur on a regular basis within a water distribution network. This lack of specific event data, which is often a key event that a utility would like to train an algorithm to detect, makes it difficult to train a machine-learning algorithm.

As an example of attempting to create sufficient training data in the physical world, consider the challenge of training self-driving cars. Waymo, Google’s autonomous car development company, has captured data from over 5 million miles of autonomous driving (with human assistance) and data capture. This is a rather large effort that a company the size of Google is able to undertake. Even Google realized that this data set was not sufficient and has augmented the data set with 2.5 billion miles driven in a simulated environment. The company is essentially creating artificial or synthetic data to augment its learning effort to develop autonomous driving algorithms.

Adversarial networks, sometimes referred to as dueling networks, pits a data generator against a data discriminator to dynamically generate data in a closed loop fashion that aids in the machine-learning effort in the absence of a sufficient number of events. A high-level overview of a GAN is shown in Figure 4 (Hitawala, 2018).

Within the water sector, a GAN could potentially be used to help train a model to predict overflow situations when there is not enough historical data to train the neural network. A generator could be used to create simulated event data.

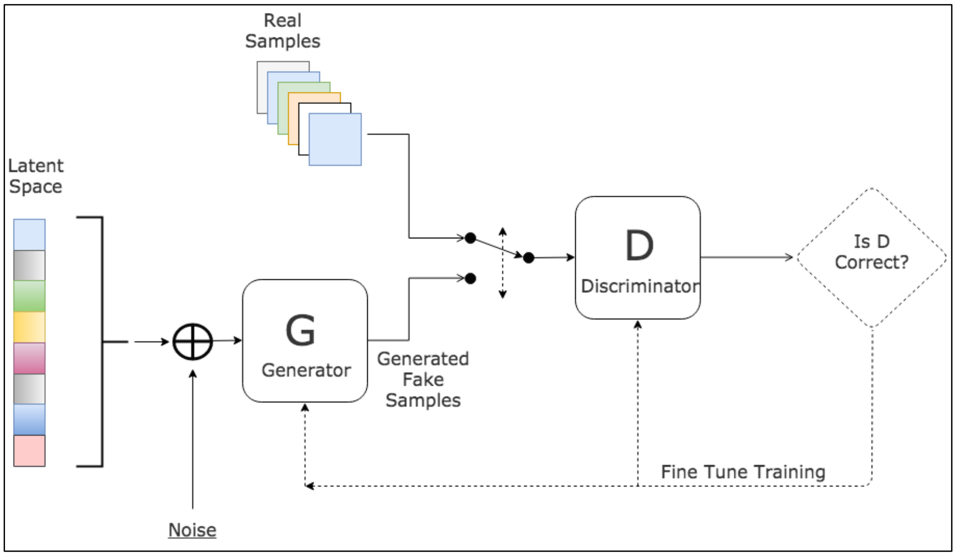

Figure 4: Generative Adversarial Network Composition

Quantum Computing

Current computing technology has been following Moore's law for nearly six decades. Moore's law states that the number of transistors on an integrated circuit doubles approximately every 2 years. While this increase in transistor count has led to increased computational power, one component of classical computing cannot be circumvented: a single computer core can only compute one thing at a time. Quantum computing breaks this constraint.

Definition

From Wired Magazine (Beal and Reynolds, 2018):

> Quantum computing takes advantage of the strange ability of subatomic particles to exist in more than one state at any time. Due to the way the tiniest of particles behave, operations can be done much more quickly and use less energy than classical computers. In classical computing, a bit is a single piece of information that can exist in two states – 1 or 0. Quantum computing uses quantum bits, or 'qubits' instead. These are quantum systems with two states. However, unlike a usual bit, they can store much more information than just 1 or 0, because they can exist in any superposition of these values.

Relevance

While much of the focus currently in quantum computing is centered in cryptography, quantum computing will have long-term potential in the ability to simulate complex physical systems at speed.

Computer scientists are looking at using quantum computing to generate much more timely and accurate weather forecasts. The equations governing weather simulations and forecasting are fairly complex and tax even the latest hardware; in some cases, the analysis runs may take longer than the actual weather event to occur within a classical computation context. Researchers at the Massachusetts Institute of Technology have discovered that the equations governing the weather possess a hidden wave nature, with wave mechanics amenable to solution by a quantum computer.

Within the water sector, water reservoirs are under more stress due to extreme weather events. Not only can quantum computing help utilities in watershed and reservoir management through increased fidelity and timeliness of weather forecast information, quantum computing can also be utilized to better understand the dynamics within the actual reservoir and help drive aeration methodology.

Conclusions

Adapting the three-horizon view for technology as a means of identifying technologies in different stages of maturation, the following technologies are important for the water industry to consider in the three horizons as follows:

- Horizon 1: Near-term products of interest include 5G cellular networking, automated machine learning, and Apache Spark
- Horizon 2: Technologies in the early adoption phase include deep learning and machine learning for cybersecurity
- Horizon 3: Technologies in the research and development phase include generative adversarial networks and quantum computing

It is expected that these technologies will be used in the future by the water industry, and preparations for their implementation should be considered.

References

References used in this technical memorandum are provided in this section.

Baghai, M.; Coley, S.; White, D. *The Alchemy of Growth;* Basic Books, 1999.

Beal, A.; Reynolds, M. What are quantum computers and how do they work? WIRED explains. *WIRED* [Online] February 16, 2018. http://www.wired.co.uk/article/quantum-computing-explained.

Check Point Research Team. December's Most Wanted Malware: Crypto-Miners Affect 55% of Businesses Worldwide. January 15, 2018. https://blog.checkpoint.com/2018/01/15/decembers-wanted-malware-crypto-miners-affect-55-businesses-worldwide/.

Doran, T.; Britton, B. How a company founded by ex-spies uses AI to fight hackers. *CNN.* February 6, 2018. https://www.cnn.com/2018/02/05/tech/darktrace-cybersecurity-immune-system/index.html.

Gartner. Gartner Hype Cycle. https://www.gartner.com/technology/research/methodologies/hype-cycle.jsp (accessed March 27, 2018).

Hackers move to mobile as cryptojacking threat evolves. *Wandera Blog.* December 4, 2017. https://www.wandera.com/blog/cryptojacking-mobile-threat/.

He, K.; Zhang, X.; Ren, S.; Sun, J. *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification* [Online], 2015. http://openaccess.thecvf.com/content_iccv_2015/papers/He_Delving_Deep_into_ICCV_2015_paper.pdf (accessed April 2, 2018).

Hitawala, S. Comparative Study on Generative Adversarial Networks. arXiv:1801.04271 [cs.LG]. January 12, 2018. https://arxiv.org/pdf/1801.04271.pdf.

Ray, W.; Ramamoorthy, G. *Hype Cycle for IoT Standards and Protocols* [Online]; Gartner: 2017. https://www.gartner.com/doc/3762285/hype-cycle-iot-standards-protocols (accessed March 27, 2018).

This is the Cutting Edge of Deep Learning Research. *Forbes.* August 5, 2016. https://www.forbes.com/sites/quora/2016/08/05/this-is-the-cutting-edge-of-deep-learning-research.

Vorhies, W. 6 Predictions about Data Science, Machine Learning, and AI for 2018. *Data Science Central.* December 12, 2017. https://www.datasciencecentral.com/profiles/blogs/6-predictions-about-data-science-machine-learning-and-ai-for-2018.

References

AWWA. 2014. AWWA Webinar Program: Smarter Water Management. https://www.awwa.org/portals/0/files/education/webcast/smarterwatermanagementhandouts.pdf.

AWWA. 2015. 2015 AWWA State of the Water Industry Report. https://www.awwa.org/Portals/0/files/resources/water%20utility%20management/sotwi/2015-AWWA-State-of-the-Water-Industry-Report.pdf.

AWWA. 2017. 2017 State of the Water Industry Report. https://www.awwa.org/resources-tools/water-and-wastewater-utility-management/state-of-the-water-industry.aspx.

Baghai, M.; Coley, S.; White, D. *The Alchemy of Growth;* Basic Books, 1999.

Beal, A.; Reynolds, M. What are quantum computers and how do they work? WIRED explains. *WIRED* [Online] February 16, 2018. http://www.wired.co.uk/article/quantum-computing-explained.

Black & Veatch. 2016. 2016 Strategic Directions: Smart City/Smart Utility Report. https://www.bv.com/reports/2016/smart-cities.

Check Point Research Team. December's Most Wanted Malware: Crypto-Miners Affect 55% of Businesses Worldwide. January 15, 2018. https://blog.checkpoint.com/2018/01/15/decembers-wanted-malware-crypto-miners-affect-55-businesses-worldwide/.

Cohen, S. 2017. SWAN Blog: Lessons from the SWAN 2017 Conference. SWAN Forum. https://www.swan-forum.com/blog/.

Conner, J.N. 2016. Water Reuse – The Need for Predictive Analytics. AWWA. http://www.awwa.org/portals/0/files/education/conferences/membrane/mtc16/membrane%20papers/wed08%20papers/03-conner%20paper.pdfGrady, B. 2016. Can Big Data save our water infrastructure? GreenBiz. https://www.greenbiz.com/article/can-big-data-save-our-water-infrastructure.

Doran, T.; Britton, B. How a company founded by ex-spies uses AI to fight hackers. *CNN.* February 6, 2018. https://www.cnn.com/2018/02/05/tech/darktrace-cybersecurity-immune-system/index.html.

Gartner. IT Glossary. Definition for "Internet of Things". Accessed February 2018. https://www.gartner.com/it-glossary/internet-of-things/.

Gartner. 2016b. Market Guide for IoT Platforms. https://www.gartner.com/doc/3380746?ref=SiteSearch&sthkw=big%20data&fnl=search&srcId=1-3478922254.

Gartner. 2016a. Mastering Data and Analytics: 'Table Stakes' for Digital Business. https://www.gartner.com/doc/3343918?ref=SiteSearch&sthkw=big%20data&fnl=search&srcId=1-3478922254.

Gartner. 2016. Organizing for Big Data Through Better Process and Governance. https://www.gartner.com/doc/3002918?ref=SiteSearch&sthkw=big%20data&fnl=search&srcId=1-3478922254.

Gartner. Gartner Hype Cycle. https://www.gartner.com/technology/research/methodologies/hype-cycle.jsp (accessed March 27, 2018).

Grilo, J. 2016. The Impact of ICT in the Water Sector. International Water Association. http://www.iwa-network.org/the-impact-of-ict-in-the-water-sector/.

Hackers move to mobile as cryptojacking threat evolves. *Wandera Blog.* December 4, 2017. https://www.wandera.com/blog/cryptojacking-mobile-threat/.

HBR. 2014. From the HBR.org Insight Center "From Data to Action". Harvard Business Publishing. https://kupdf.com/download/hbr-from-data-to-action-107218_58ac23dc6454a7c34fb1e8ee_pdf.

He, K.; Zhang, X.; Ren, S.; Sun, J. *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification* [Online], 2015. http://openaccess.thecvf.com/content_iccv_2015/papers/He_Delving_Deep_into_ICCV_2015_paper.pdf (accessed April 2, 2018).

Hitawala, S. Comparative Study on Generative Adversarial Networks. arXiv:1801.04271 [cs.LG]. January 12, 2018. https://arxiv.org/pdf/1801.04271.pdf.

Kushmaro, P. 2017. The IoT and Big Data: Making The Connection. Huffington Post. https://www.huffingtonpost.com/philip-kushmaro/the-iot-and-big-data-maki_b_12116608.html.

Li, H. 2015. Wise Use of Water With Smart Technologies. International Water Association. http://www.iwa-network.org/wise-use-of-water-with-smart-technologies/.

Liner, B. 2016. Big Data Making Waves in Utility and Investment Communities. http://www.ncsafewater.org/news/298363/.

Liner, B. and Kenel, P. 2016. Exciting future for Big Data solutions. Water Environment Federation. https://www.wef.org/globalassets/assets-wef/3---resources/topics/a-n/innovation/technical-resources/worldwaterbigdataapril2016.pdf.

McLellan, C. 2015. The internet of things and Big Data: Unlocking the power. ZDNet. http://www.zdnet.com/article/the-internet-of-things-and-big-data-unlocking-the-power/.

NIST. 2015. NIST Special Publication 1500-1. NIST Big Data Interoperability Framework: Volume 1, Definitions. http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf.

Noonan, C. 2014. Breaking through Big Data, analytics and the Internet of Things. Smart Cities Council. https://smartcitiescouncil.com/article/breaking-through-big-data-analytics-and-internet-things.

Ray, W.; Ramamoorthy, G. *Hype Cycle for IoT Standards and Protocols* [Online]; Gartner: 2017. https://www.gartner.com/doc/3762285/hype-cycle-iot-standards-protocols (accessed March 27, 2018).

Shaw, A. 2017. Understanding Big Data in the Water Industry. Water Online. https://www.wateronline.com/doc/understanding-big-data-in-the-water-industry-0002.

Theo, G. 2017. Tapping into data: The analytics revolution for the business and customer. SWAN Forum. https://www.swan-forum.com/wp-content/uploads/sites/218/2017/05/SWAN-2017_Keynote-Unitywater_George-Theo.pdf.

This is the Cutting Edge of Deep Learning Research. *Forbes.* August 5, 2016. https://www.forbes.com/sites/quora/2016/08/05/this-is-the-cutting-edge-of-deep-learning-research.

Vorhies, W. 6 Predictions about Data Science, Machine Learning, and AI for 2018. *Data Science Central.* December 12, 2017. https://www.datasciencecentral.com/profiles/blogs/6-predictions-about-data-science-machine-learning-and-ai-for-2018.

Water Online 2016. Surging 28%, US Municipal Water Infrastructure Forecast Reaches $532B. Water Online. https://www.wateronline.com/doc/surging-us-municipal-water-infrastructure-forecast-reaches-b-0001.

WEF. 2016. Intelligent Water Systems: A Smart Start. Water Environment Federation. https://wef.org/globalassets/assets-wef/3---resources/online-education/webcasts/presentation-handouts/iws-webcast---presentation-handouts.pdf.

Westerling, K. 2017. Smart Water Questions Answered. Water Online. https://www.wateronline.com/doc/smart-water-questions-answered-0001.

THE
Water
Research
FOUNDATION

1199 North Fairfax Street, Suite 900
Alexandria, VA 22314-1445
www.werf.org | werf@werf.org

6666 West Quincy Avenue
Denver, CO 80235-3098
www.waterrf.org | info@waterrf.org