confluent

2017
APACHE KAFKA™
REPORT

CONTENTS

INTRODUCTION

Over the past several years, organizations across many industries have discovered, and are filling, an increasingly important gap in their data infrastructure. It sits at the nexus of big data, data integration, and all of their data stores and applications – a gap that is being filled by streaming platforms like Apache Kafka™. Confluent has enjoyed a front row view as companies adopt streaming platforms to create new products, become more responsive to customers and make business decisions in real time.

This survey focuses on why and how companies are using Apache Kafka and streaming data and the impact it has on their business.

CONFLUENT SURVEYED OVER 350 ORGANIZATIONS FROM 47 COUNTRIES AND A WIDE VARIETY OF INDUSTRIES TO UNDERSTAND THE EVOLVING APACHE KAFKA USER BASE, USE CASES AND DEPLOYMENTS.

Those surveyed range in position from developers and architects to technical management and engineers. 1 in 4 respondents (26%) work for organizations with more than $1 billion in annual sales, illustrating how quickly this open source technology has gained traction across large enterprises.

Executive Summary

The problem of managing data in enterprises is an exercise in complexity. As companies add new applications, integrate and modernize existing systems, change to a microservices architecture, and embrace a factor increase in data-producing endpoints, their infrastructure, and the engineers who support it, face an existential question:

Should we pursue each new data project as a unique discrete effort, or find an approach with greater scalability?

For many, Kafka has become the answer to this question. It transforms an ever-increasing number of new data producers and consumers into a simple, unified streaming platform at the center of their organization. It allows any team to join the platform, allows a central team to manage the service, and scales to trillions of messages per day while delivering messages in real time.

Kafka takes on legacy technology across many different areas, including ETL, data warehouses, Hadoop, messaging middleware and data integration technologies, to massively simplify an organization's infrastructure. In many cases, Kafka can replace or augment an existing system to make data more consistently available, faster and less costly to deliver.

MORE THAN

15%

RESPONDENTS

PROCESS A BILLION MESSAGES A DAY
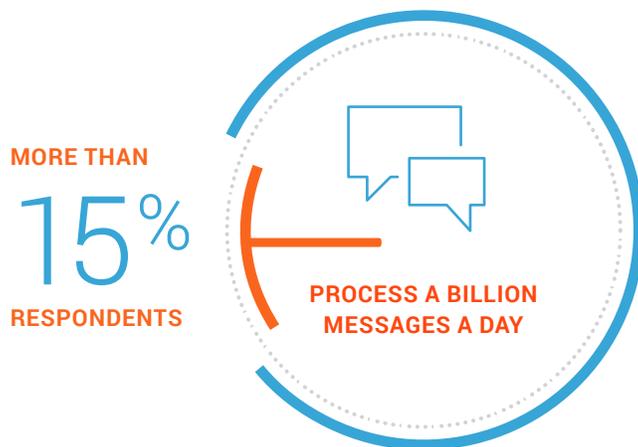
SOME PARTICULARLY REVEALING RESULTS FROM THIS YEAR'S SURVEY INCLUDE:

Kafka use is experiencing a surge:
86% of respondents reported that the number of their systems that use Kafka is increasing and a fifth (20%) reported that the number is "growing a lot!" A majority (52%) of organizations have at least 6 systems running Kafka with over a fifth (21%) having more than 20. According to last year's report, only 41% of organizations had at least 6 systems running Kafka and only one-tenth (10%) had more than 20.

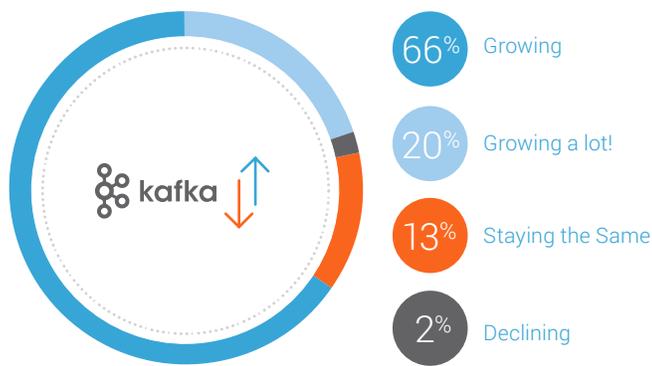**86%** of respondents reported that the number of their systems that use Kafka is increasing

**20%** reported that the number is "growing a lot!"

52% of organizations have at least 6 systems running Kafka.

IS THE NUMBER OF SYSTEMS USING KAFKA GROWING, DECLINING OR STAYING THE SAME?

66% Growing

20% Growing a lot!

13% Staying the Same

2% Declining

Note: Only a single option could be selected

HOW MANY SYSTEMS OR APPLICATIONS USE KAFKA TODAY?

4% 0

6% 1

10% 2

27% 3-5

20% 6-10

11% 11-20

13% 21-50

8% 51+
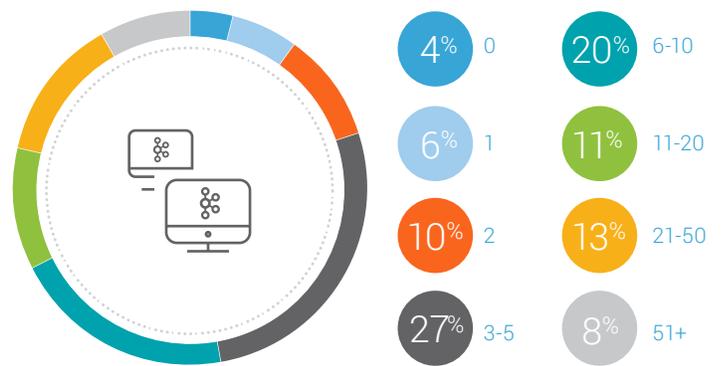
Note: Only a single option could be selected

Kafka is broadly used in the cloud:
Kafka is used by organizations in some combination of virtual private clouds (34%), public clouds (52%), and on premises (57%). Nearly one-third (32%) of respondents who use Kafka in the cloud have at least 6 Kafka applications in the cloud.

60%
RESPONDENTS

amazon web services
USE AWS AS THEIR PUBLIC CLOUD
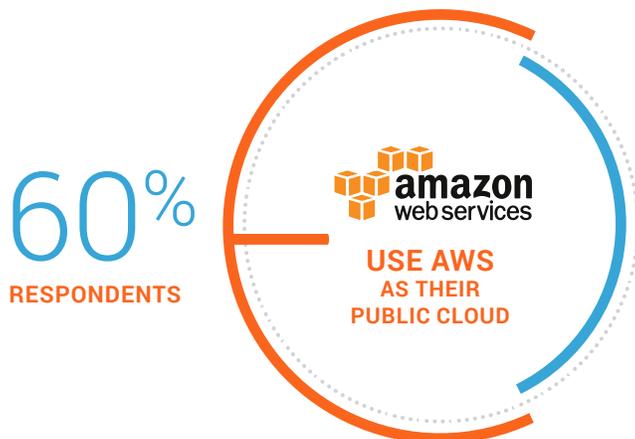
IN WHICH OF THESE ENVIRONMENTS ARE YOU USING KAFKA TODAY?

On premises
48%

A public cloud
43%

My own virtual private cloud
25%

All of the above
9%
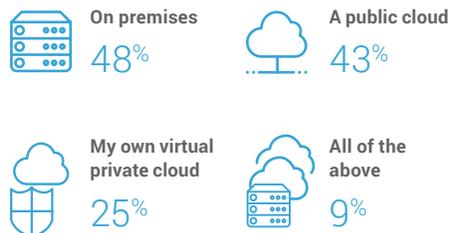
Note: Respondents could select multiple options

KAFKA CREATES NEW BUSINESS OPPORTUNITIES AND IMPROVES THE OLD ONES, TOO

While we've seen organic growth of Apache Kafka over the years and the downloads to prove it, this survey quantifies what we see from customers and adopters: streaming data enables organizations to do what they couldn't before:

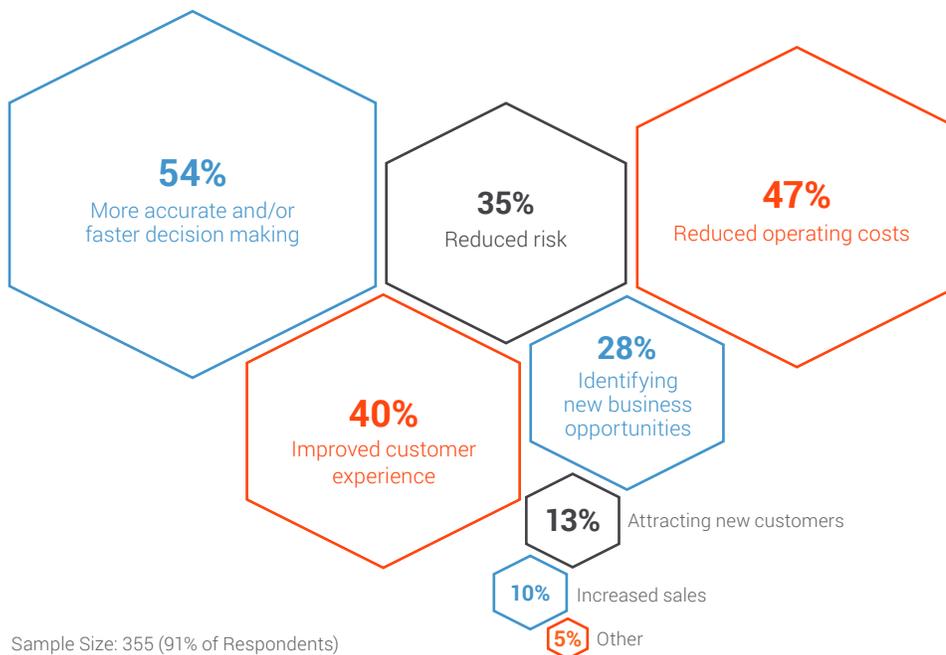- Kafka enables companies to create new market opportunities. Because data is available, shared and immediate, companies can create new products and significantly transform existing ones. As Kafka is deployed in more mission-critical infrastructures, a majority (54%) of surveyed organizations say that their business can make more accurate and/or quicker decisions thanks to Kafka.

- IIn addition to creating new opportunities, companies leverage Kafka to be more efficient and transform existing processes. Of those surveyed, other business benefits worth noting are reduced operating costs (47%) and improved customer experience (40%).

WHAT VALUE DOES YOUR BUSINESS GET FROM USING KAFKA AND STREAMING PLATFORMS? (SELECT ALL THAT APPLY.)

54%
More accurate and/or faster decision making

35%
Reduced risk

47%
Reduced operating costs

40%
Improved customer experience

28%
Identifying new business opportunities

13% Attracting new customers

10% Increased sales

5% Other

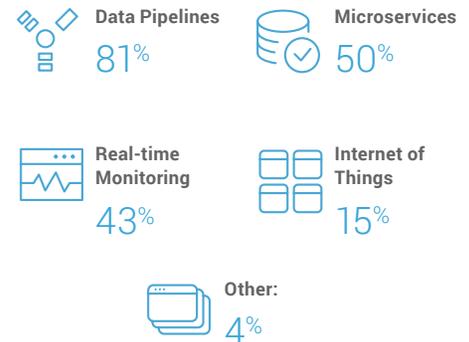Sample Size: 355 (91% of Respondents)

PUTTING APACHE KAFKA TO USE: FAR BEYOND PUB-SUB

Kafka can be used in a variety of different ways for many different use cases. Last year we found a surge of companies adopting streaming platforms. With this approach, companies build mission critical, real-time applications that power their core business — all the way from small to large-scale use cases that handle millions of events per second.

Now, companies are solving a new problem with Kafka: microservices. While microservices involve many independent services, the goal is broader than simply running them across different machines. It's about facing up to a world that is, itself, inherently distributed. Not in some narrow technical sense, but rather as a broad ecosystem composed from many people, many teams and many programs, all of which need the agility that microservices affords them.

This year, we're seeing organizations use Apache Kafka in many ways: two-thirds (66%) use it for stream processing and three out of five (60%) use it for data integration. The most common use case for Kafka is data pipelines (81%), while half (50%) are already using it for microservices.

WHAT ARE YOUR CURRENT USE CASES FOR KAFKA?

Data Pipelines
81%

Microservices
50%

Real-time Monitoring
43%

Internet of Things
15%

Other:
4%

Note: Respondents could select multiple options

WHAT WAYS ARE YOU USING APACHE KAFKA TODAY? (SELECT ALL THAT APPLY.)

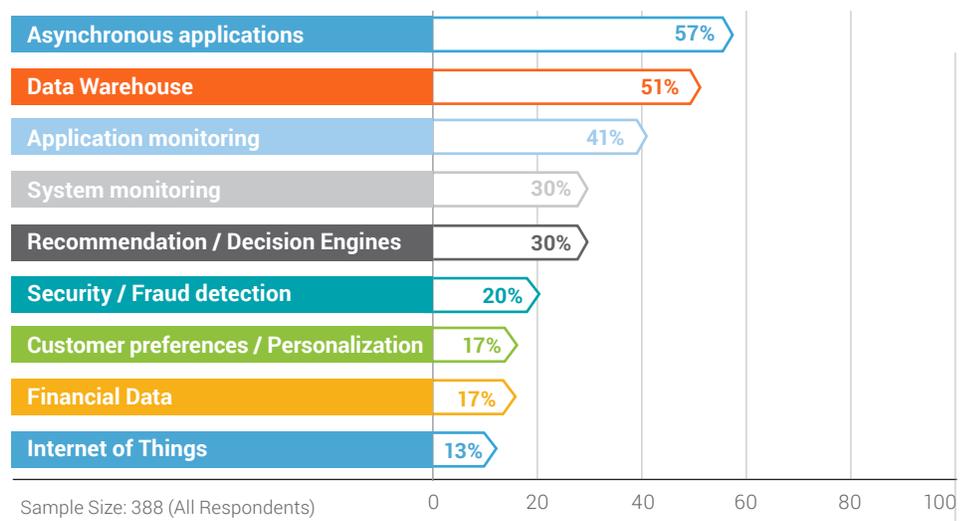| 66%	60%	50%	30%	3%
Stream Processing	Data Integration	Messaging	Log Aggregation	Other
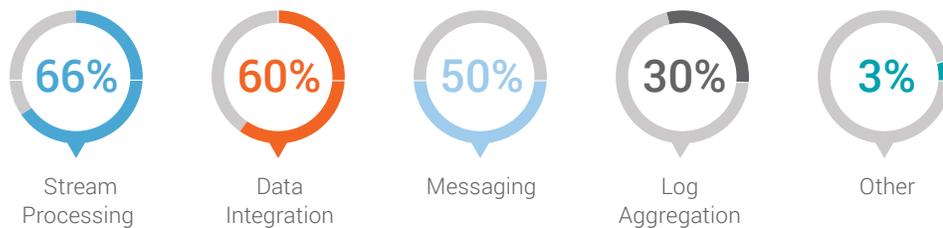
Sample Size: 168 (43% of Respondents)

Similar to last year's report, three-fourths (75%) of organizations have applications connected to their Kafka systems. These applications process data from sensors, websites, analytics and monitoring tools, etc. and share the information so the right teams can process the data they need to make decisions.

The types of applications that a majority of respondents connect to with Kafka are asynchronous applications (57%) and data warehouse (51%). Organizations also have application monitoring (41%), system monitoring (30%), or recommendation/decision engines (30%) powered by Kafka.

WHAT TYPE OF APPLICATIONS ARE POWERED BY KAFKA WITHIN YOUR ORGANIZATION? (SELECT ALL THAT APPLY.)

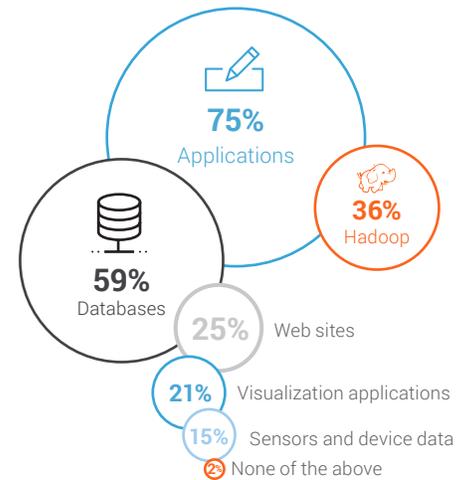| Application	%
Asynchronous applications	57%
Data Warehouse	51%
Application monitoring	41%
System monitoring	30%
Recommendation / Decision Engines	30%
Security / Fraud detection	20%
Customer preferences / Personalization	17%
Financial Data	17%
Internet of Things	13%

Sample Size: 388 (All Respondents)

THE IMPACT OF KAFKA'S GROWING FEATURE LIST

The popularity of the Kafka Connect API has grown significantly over the past year. The Kafka Connect API, included in Kafka, makes it easy to add new datastores to your data pipelines without having to write the interfaces from scratch.

There was a 15-point increase in organizations using the Kafka Connect API over last year (37% in 2017 vs 12% in 2016). While a majority (59%) of respondents have databases connected to their Kafka clusters, only 36% use the Kafka Connect API with Hadoop/HDFS, which is a 4-point drop from last year.
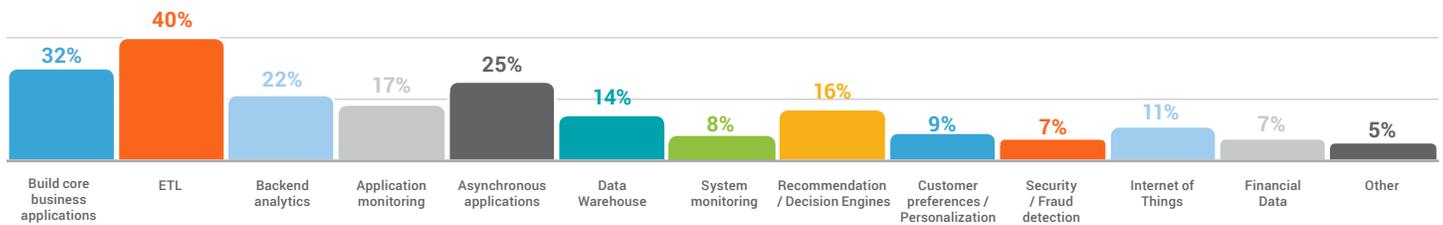
Introduced last year, the Kafka Streams API integrates into your streaming application to build and execute powerful stream processing functions, without needing to set up and run separate infrastructure. Despite its recency, most (89%) organizations are already familiar with the Kafka Streams API. The most common uses for the Kafka Streams API are ETL (40%), building core business applications (32%), and asynchronous applications (25%).

WHAT SOURCES OR SINKS ARE CONNECTED TO YOUR KAFKA CLUSTERS? (SELECT ALL THAT APPLY.)

- **75%** Applications
- **36%** Hadoop
- **59%** Databases
- **25%** Web sites
- **21%** Visualization applications
- **15%** Sensors and device data
- **2%** None of the above

Note: Respondents could select multiple options

WHAT ARE YOU USING THE KAFKA STREAMS API FOR? (SELECT ALL THAT APPLY.)

Build core business applications	ETL	Backend analytics	Application monitoring	Asynchronous applications	Data Warehouse	System monitoring	Recommendation / Decision Engines	Customer preferences / Personalization	Security / Fraud detection	Internet of Things	Financial Data	Other
32%	40%	22%	17%	25%	14%	8%	16%	9%	7%	11%	7%	5%

Note: Respondents could select multiple options

As we mentioned earlier, microservices is the new way to build custom applications. This is made much easier with the Kafka Streams API as developers have the ability to work with data in flight. Of the organizations that have microservices, 28% use the Kafka Streams API to manage them. The Kafka Streams API is most frequently utilized by developers (85%), but architects (48%) and application teams (43%) also use it.

THERE IS A SHORTAGE OF SKILLED KAFKA ENGINEERS

According to a Dice report, people with Kafka skills receive one of the highest salaries in the technology market. However, despite the salary and the growth of Kafka within organizations, three-quarters (75%) of respondents find it difficult to find the right talent with Kafka skills.

75%
RESPONDENTS

HAVE DIFFICULTY
FINDING GOOD TALENT
WITH KAFKA SKILLS

HAVE YOU FOUND IT DIFFICULT TO FIND THE RIGHT TALENT WITH KAFKA SKILLS?

Yes, extremely difficult
15%

Yes, somewhat difficult
60%

No, not very difficult
22%

No, not difficult at all
3%

Note: Only a single option could be selected

RESPONDENTS ARE VERY SATISFIED WITH KAFKA

An factor in the growing popularity of Kafka is that it meets the expectations of its users. Overall satisfaction with Kafka remains high with 70% of organizations saying that they are either very or completely satisfied with it.

HOW SATISFIED ARE YOU WITH KAFKA?

Not at all satisfied
<1%

Slightly satisfied
4%

Moderately satisfied
26%

Very satisfied
61%

Completely satisfied
9%

Note: Only a single option could be selected

CONCLUSIONS

The results of this survey demonstrate the increasing use of Apache Kafka for a broadening set of business and technical objectives. Many companies are implementing the distributed streaming platform for more accurate and faster decision making, reduced operating costs, improved customer experiences and reduced risk.

With demands from customers to react and respond in real time and legacy technologies holding companies back, it's no wonder companies have turned to Apache Kafka to embrace the power of streaming data for competitive advantage.

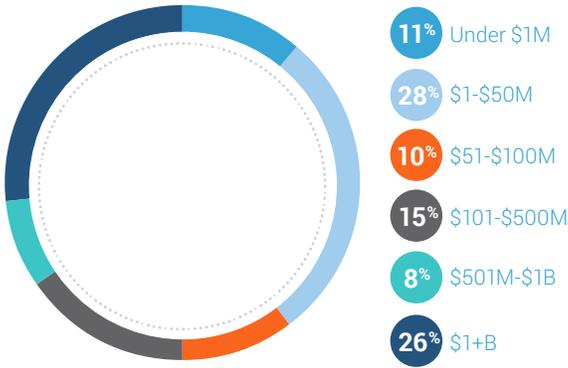## APACHE KAFKA IS A DOMINANT STREAMING DATA SOLUTION FOR ALL ENVIRONMENTS

With the rise of microservices and the cloud, Kafka is best-suited to meet the needs of these distributed environments. This survey demonstrates companies are not only actively deploying these architectures, but they are looking and implementing technologies that help them scale.

KAFKA SKILLS ARE IN DEMAND

Companies are increasing their use of Kafka and are looking for people with the right skillset. With this demand, it's hard to imagine a world without streaming platforms at the heart of every company in the very near future.
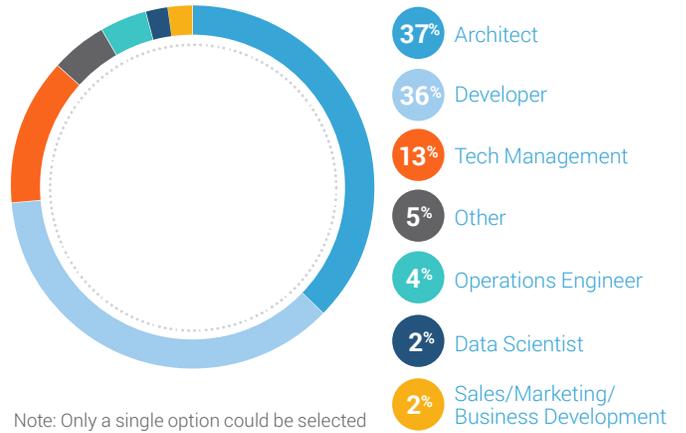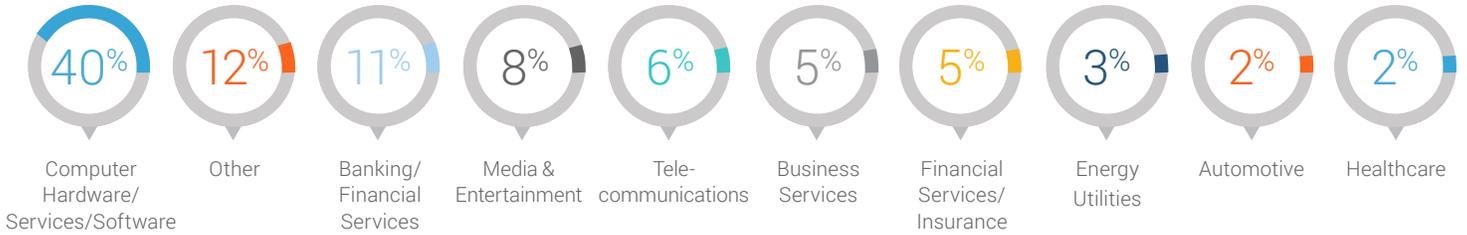
WHO TOOK THE SURVEY?

COMPANY SIZE

- **11%** Under $1M
- **28%** $1-$50M
- **10%** $51-$100M
- **15%** $101-$500M
- **8%** $501M-$1B
- **26%** $1+B

Note: Only a single option could be selected

RESPONDENT ROLE

- **37%** Architect
- **36%** Developer
- **13%** Tech Management
- **5%** Other
- **4%** Operations Engineer
- **2%** Data Scientist
- **2%** Sales/Marketing/ Business Development

Note: Only a single option could be selected

INDUSTRIES

- **40%** Computer Hardware/ Services/Software
- **12%** Other
- **11%** Banking/ Financial Services
- **8%** Media & Entertainment
- **6%** Tele-communications
- **5%** Business Services
- **5%** Financial Services/ Insurance
- **3%** Energy Utilities
- **2%** Automotive
- **2%** Healthcare

LOCATIONS

- Sweden 2%
- UK 7%
- Canada 2%
- Netherlands 2%
- Germany 6%
- France 5%
- Italy 3%
- Spain 4%
- US 45%
- India 4%

Confluent, founded by the creators of open source Apache Kafka™, provides the only streaming platform that enables enterprises to maximize the value of data. Confluent Enterprise empowers leaders in industries such as retail, logistics, manufacturing, financial services, technology and media, move data from isolated systems into a real-time data pipeline where they can act on it immediately. Backed by Benchmark, Index Ventures, and Sequoia, Confluent is based in Palo Alto, California.

To learn more about how Confluent Platform and Apache Kafka can help your organization, visit: www.confluent.io