

2018 Apache Kafka[®] Report

Table of Contents

Introduction	3
Executive Summary	6
Demographics	8
Key Findings	9
Apache Kafka Powers the Real-Time, Event-Driven Enterprise	9
Kafka's Feature Set Enables the Central Nervous System	14
Kafka Users Remain Satisfied; Top Benefit is Increased Agility	15
More than 90% of Companies will Implement New Kafka Applications	17
Replacing Technologies and Recruiting Staff as Kafka Matures	19
Kafka Users Deploy the Streaming Platform Across a Variety of Environments	20
Conclusion	23
Who Took the Survey	24
The Apache Kafka Community	26
About Confluent	31

Introduction

Apache Kafka® is a streaming platform that spans to the far edges of an organization, capturing the very impulses of the business in a **central nervous system** that can unify everything that happens into streams of events. A platform that allows the business to respond in real time to every customer experience, transaction, sale and market movement.



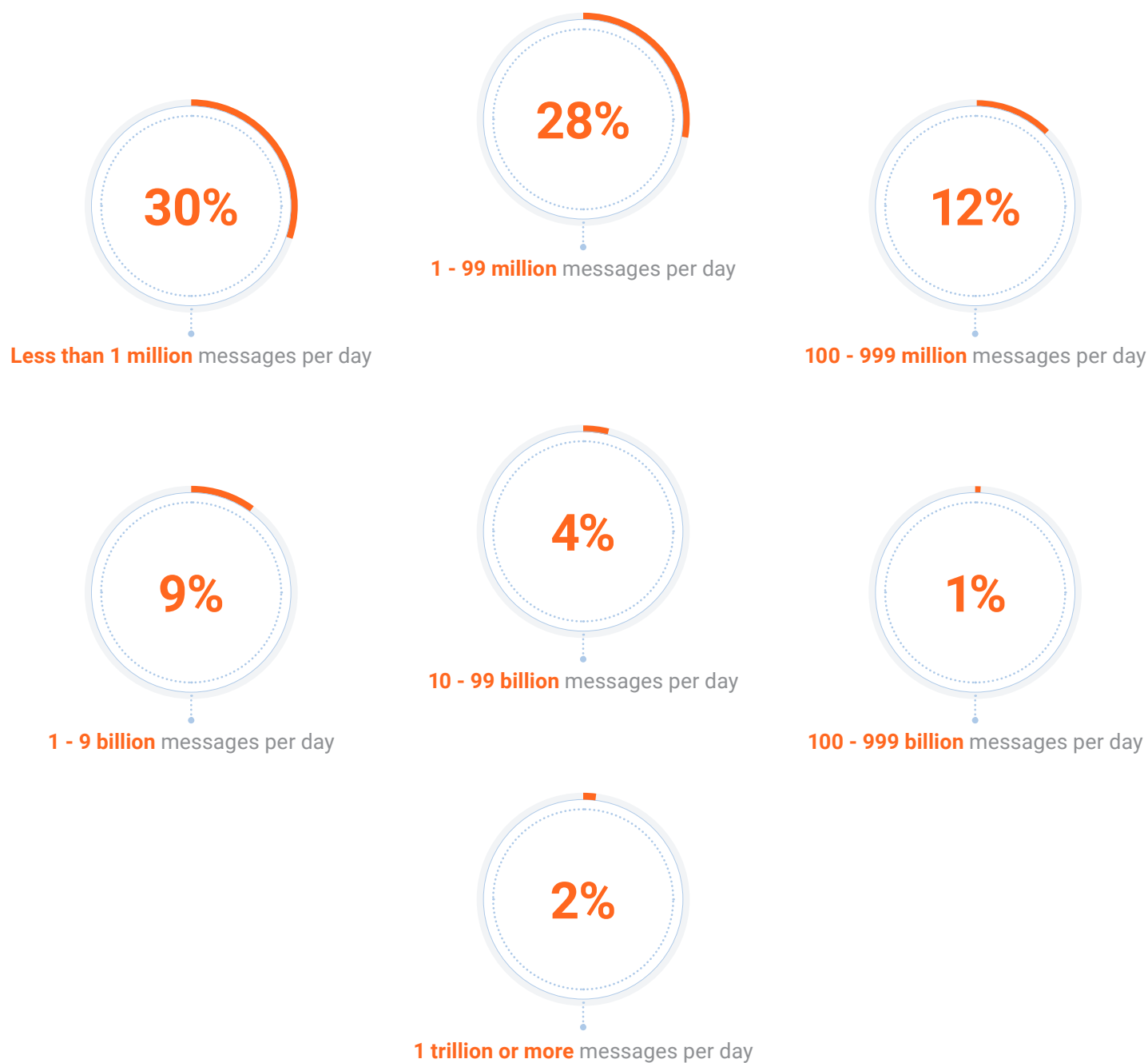
Now in its third year, the survey of the **Apache Kafka** community shows how and why companies are adopting streaming platforms to build event-driven architectures, and the bottom line impact they have on the business.

Nearly six out of ten companies (57%) process 1 million or more messages daily through Kafka: 28% process between 1 and 99 million, and 2% process 1 trillion or more.



How many messages per day does your company process through Kafka?

Sample Size : 564 (93% of Respondents)



Executive Summary

Since Kafka was created in 2012, the growth has been astonishing. Particularly notable in this year's report was the breadth and depth of companies using Kafka as a strategic platform for their business, as well as the scope and centrality of the applications developed.

Companies are reimagining retail, health-care, cars, travel, banking, and virtually every other industry imaginable, around streams of real-time events, creating a new technology pattern.

An event-driven architecture is the idea that one should design software starting with **events**, things that happen in the real world and have real business meaning, and that these events should be the central concept in the architecture.

Top findings from the report include:

Kafka adoption is growing. This year, the vast majority of organizations (94%) plan to deploy new applications or systems using Kafka. Two thirds (67%) plan to deploy between 1 and 10 new applications or systems.

Building distributed systems is challenging. Apache Kafka provides a powerful toolkit for putting together big services as a set of scalable, decoupled components. For six out of ten organizations, their microservices architecture (63%) relies on Kafka to preserve and communicate state, more than twice as many did last year (29%).

Truly data-driven organizations keep events at the center of the business. In fact, 90 percent of respondents said Apache Kafka is mission critical to power this central nervous system across many different applications and use cases.



Will Kafka be a mission-critical part of your data infrastructure/application architecture in 2018?

Sample Size : 537 (89% of Respondents)

90%

Yes

Nine out of ten organizations (90%) say Kafka will be a mission-critical part of their data infrastructure / application architecture in 2018.



Demographics



Confluent surveyed over 600 Apache Kafka users from 59 countries in order to better understand the behavior, usage and attitudes of Kafka's growing user base.

Respondents were most often data architects (**36%**), developers (**26%**), or data engineers (**13%**). They came from a variety of industries – most frequently computer systems (**20%**), financial services (**12%**) and media and entertainment (**10%**). Companies ranged in size: while **30%** grossed between \$1 and \$49 million annually, nearly as many (**25%**) grossed \$1 billion or more per year.

Key Findings

APACHE KAFKA POWERS THE REAL-TIME, EVENT-DRIVEN ENTERPRISE:

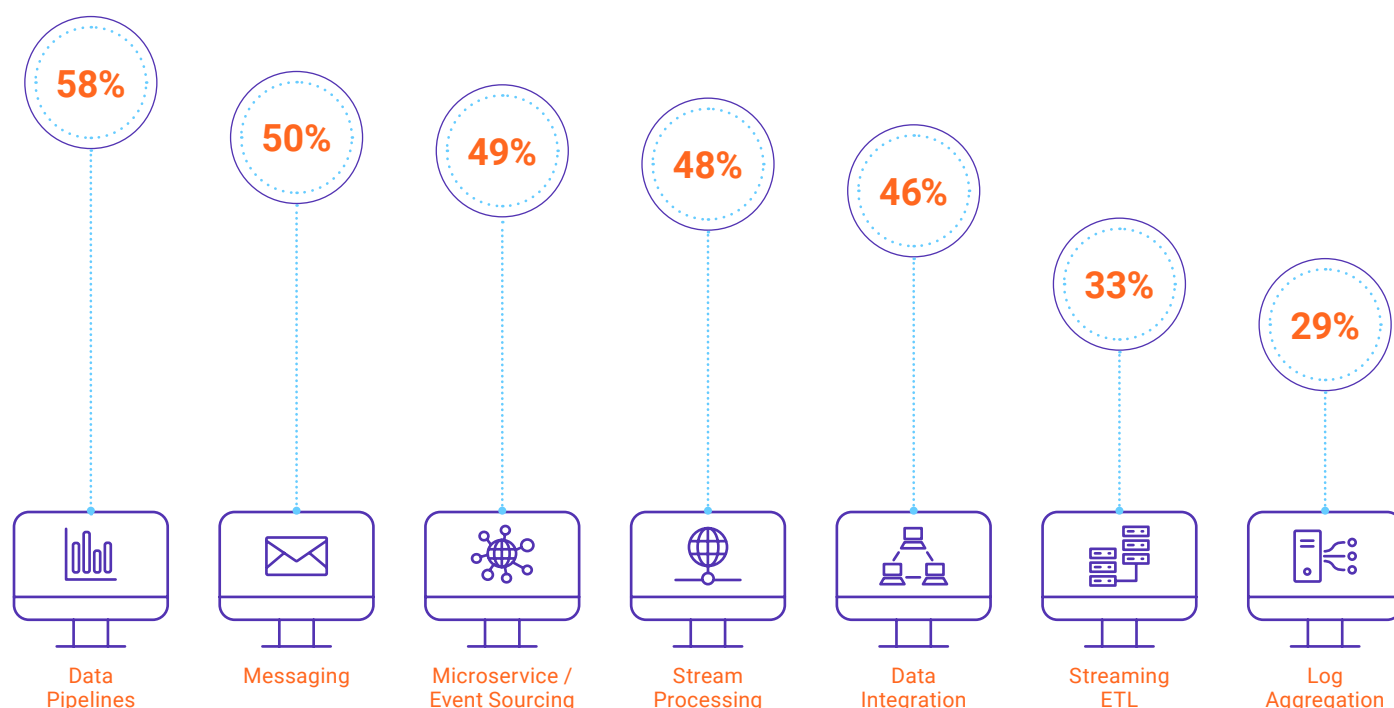
While Kafka powers a wide variety of applications, many industries dealing with similar challenges adopt the platform to impact their bottom line. E-commerce (49%) and media and entertainment companies (42%) use Kafka more than any other industries to power their recommendation and decision making applications. Almost two thirds of financial services (61%) and banking and financial (66%) companies use Kafka with financial data applications, while computer hardware, services and software companies are more likely than others to use it with security and fraud detection (22%) and Internet of Things (35%) applications.

In order to respond more effectively to the increasing demands of current and future business, technology leaders must support **real-time business decisions**. Apache Kafka meets this need by acting as the **central nervous system for the modern digital enterprise**. This streaming platform helps businesses improve customer experiences, reduce risk, and evolve faster to deliver on the promise of digital business transformation.



In what ways are you using Apache Kafka today?

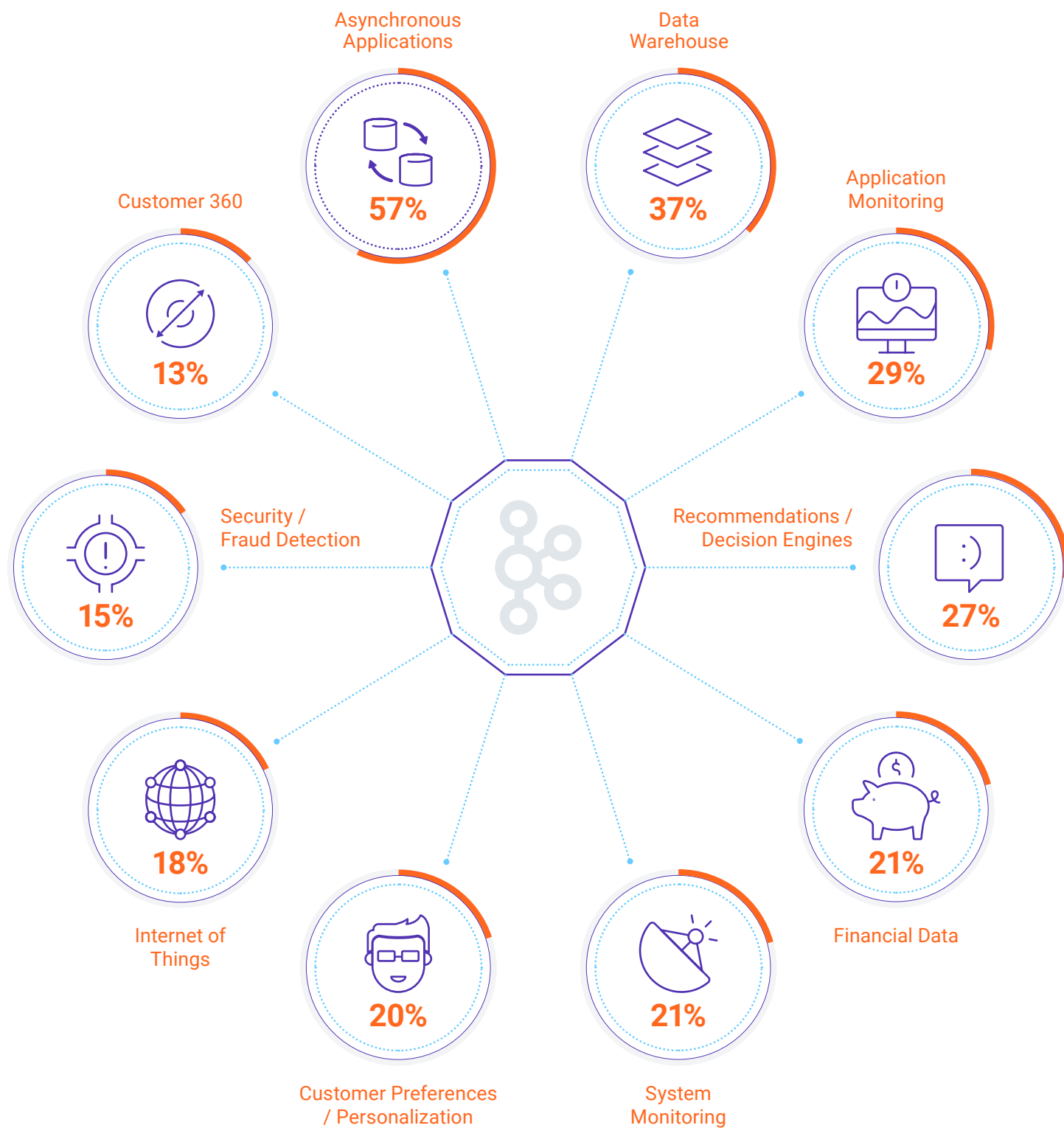
Sample Size : 604 (All Respondents)





What type of applications are powered by Kafka within your organization? - Top 10

Sample Size : 602 (All Respondents)

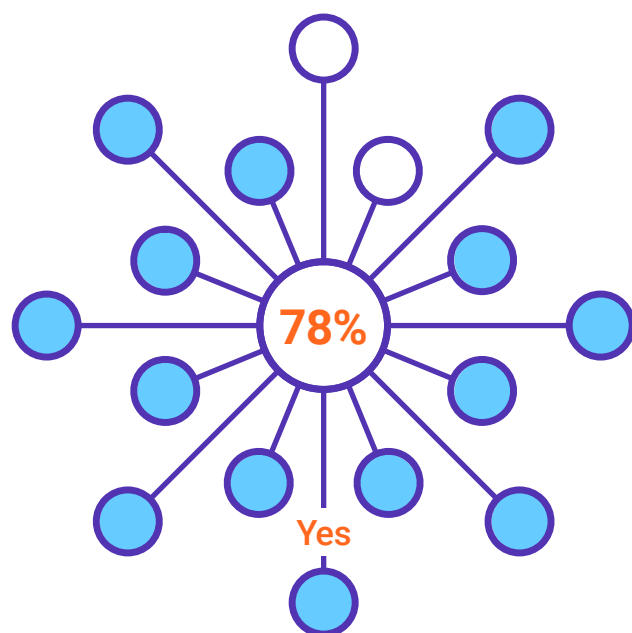


Building distributed systems is challenging. Luckily, Apache Kafka provides a powerful toolkit for putting together big services as a set of scalable, decoupled components. For six out of ten organizations, their **microservices architecture** (63%) relies on Kafka to preserve and communicate state, **more than twice as many did last year** (29%).



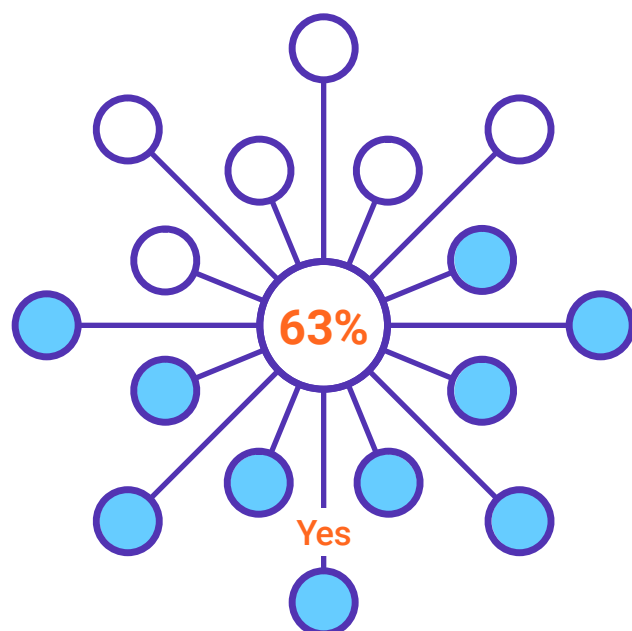
Is your organization using microservices architecture today?

Sample Size : 580 (96% of Respondents)



Does your microservices architecture rely on Kafka to preserve and communicate state?

Sample Size : 444 (74% of Respondents)



KAFKA'S FEATURE SET ENABLES THE CENTRAL NERVOUS SYSTEM:

Streaming platforms represent a paradigm shift. As companies are building and rebuilding to be digitally-native, they are putting data at the core and fundamentally transforming how applications are built. Kafka's features, like Kafka Connect and the Streams API, enable this transformation and interaction to be possible.



KAFKA CONNECT:

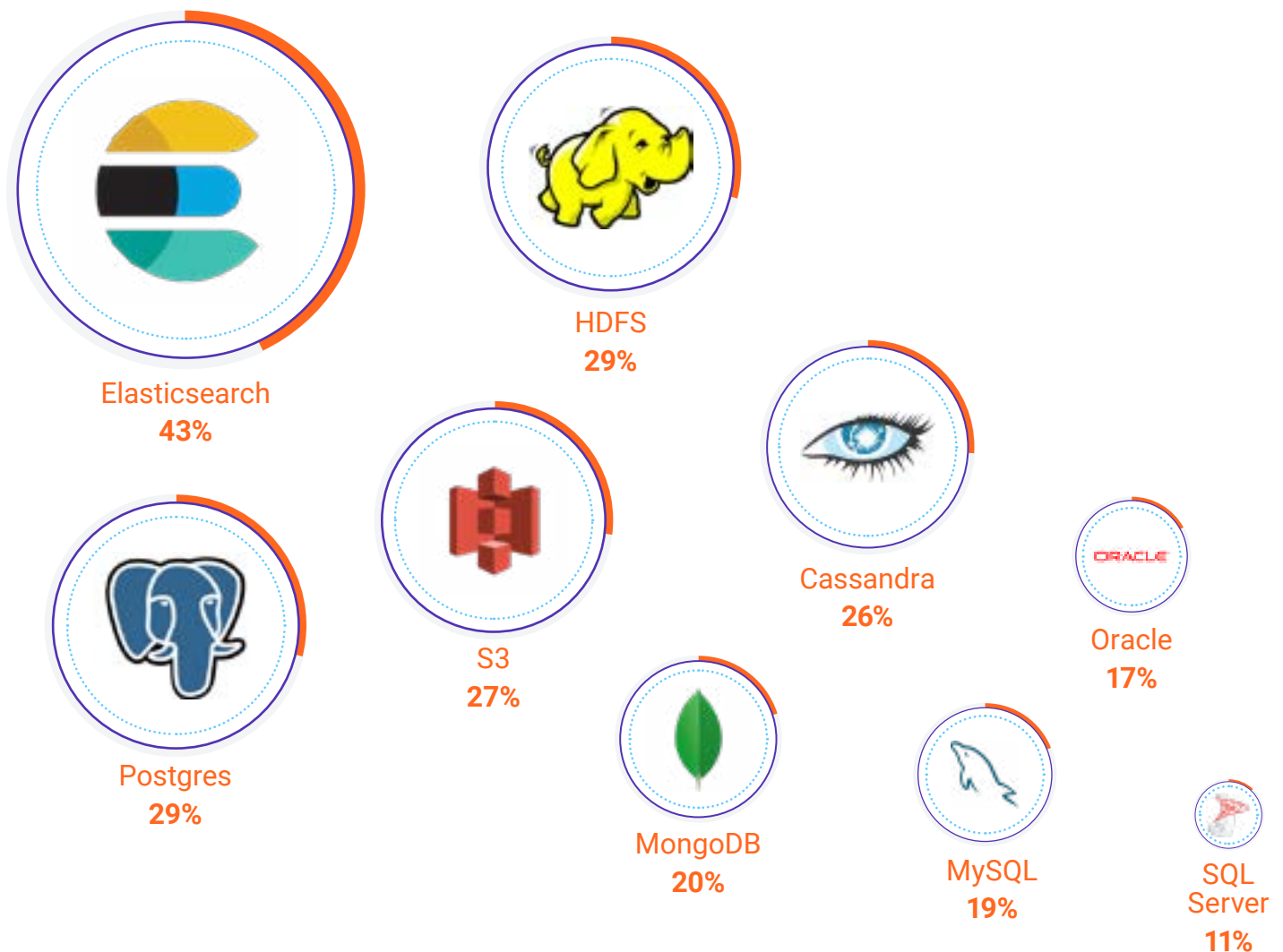
Kafka Connect is a framework included in Kafka that integrates Kafka with other systems. Its purpose is to make it easy to add new systems to create scalable and secure stream data pipelines.

When integrating data stores, organizations from this year's survey use Kafka most often with Elasticsearch (43%). Just under a third use it to integrate Postgres (29%), HDFS (29%), S3 (27%), and Cassandra (26%).

When selecting "other" data stores, the most common were HBase, Redshift, Clickhouse, Druid, and Vertica.

**Which data store(s) do you integrate with Kafka? - Top 10**

Sample Size : 603 (All Respondents)



KAFKA STREAMS API:

The Kafka Streams API integrates into a streaming application to build and execute powerful stream processing functions, without needing to set up and run separate infrastructure.

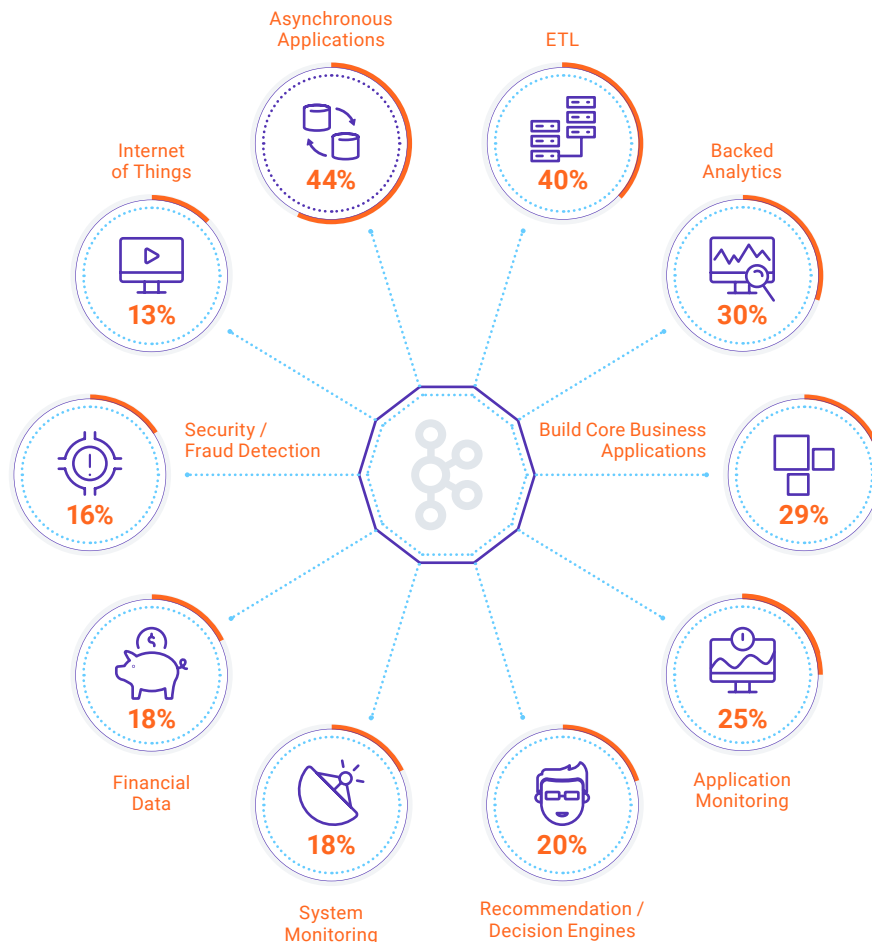
Organizations are using Kafka's stream processing capabilities most often for asynchronous applications (44%) and ETL (40%). Another 3 out of 10 use them for backend analytics (30%) and building core business applications (29%).

Banking and other financial services companies understandably use Kafka's stream processing capabilities for financial data more than other companies (59% and 53%, respectively). Media and entertainment companies, meanwhile, favor using Kafka's stream processing capabilities with backend analytics over other industries (46%).



What are you using Kafka's stream processing capabilities for? - Top 10

Sample Size : 530 (88% of Respondents)



THE INTRODUCTION OF EXACTLY-ONCE SEMANTICS:

Two thirds of organizations (68%) are building applications on Kafka in 2018 that utilize exactly-once semantics.

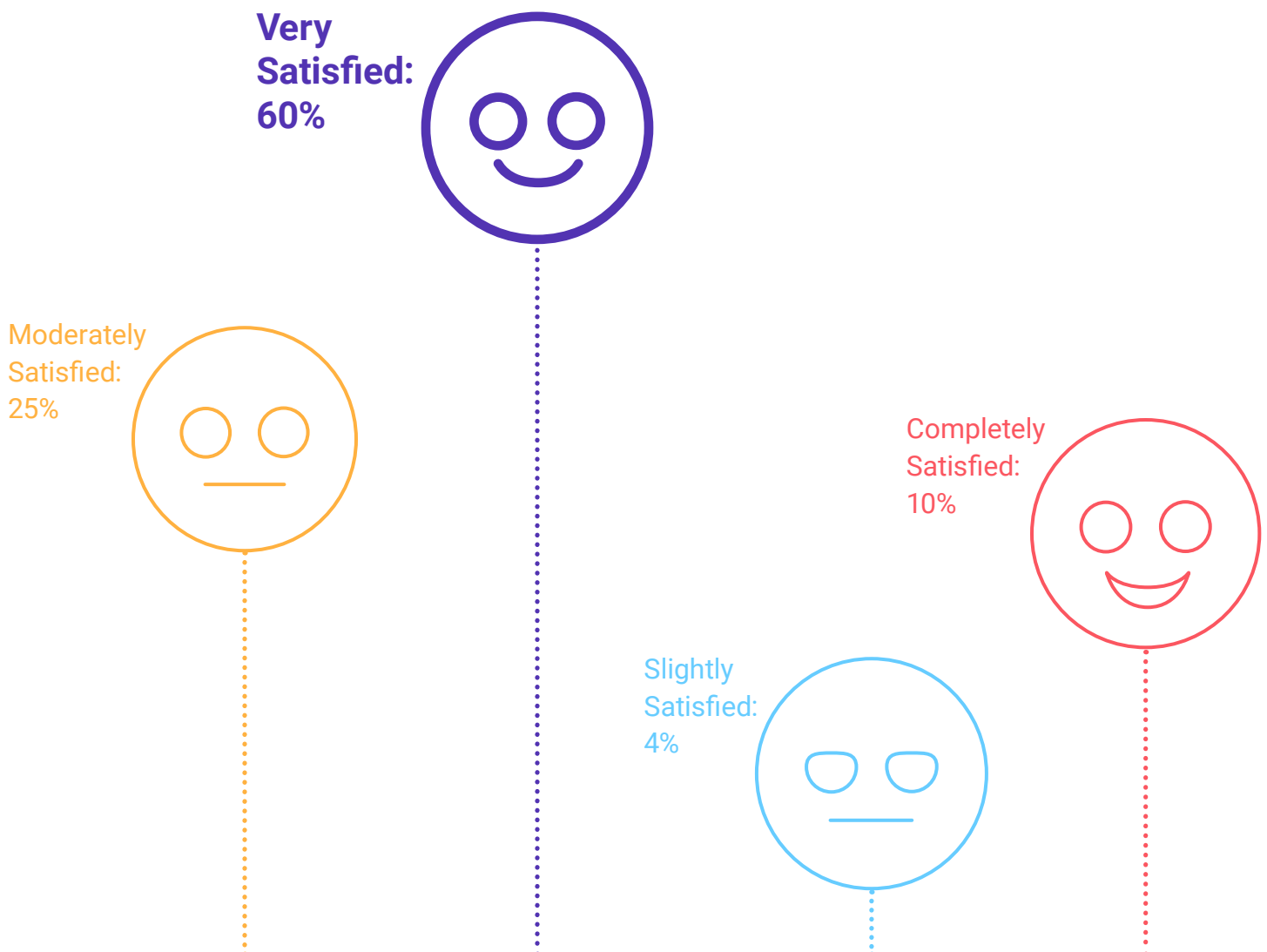
KAFKA USERS REMAIN SATISFIED; TOP BENEFIT IS INCREASED AGILITY:

Seven out of ten users (70%) are very or completely satisfied with Kafka. This satisfaction rating has remained steady since 2016, at 68-70% each year.



How satisfied are you
with Kafka?

Sample Size : 537 (89% of Respondents)

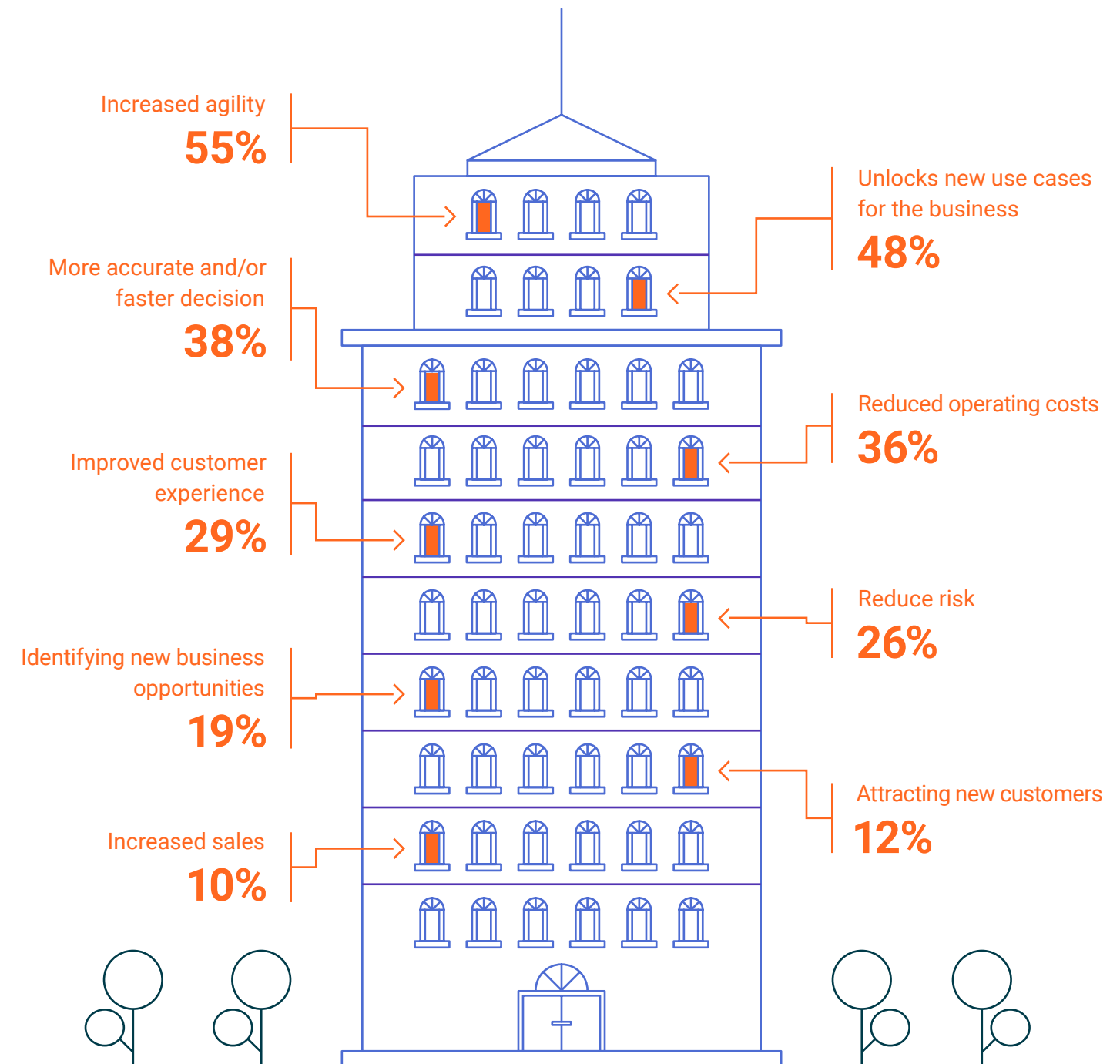


Top benefits cited from Kafka usage are increased agility (55%), unlocking new use cases for the business (42%), and more accurate and/or faster decision-making (38%). About one in three companies mention the benefits of reduced operating costs (36%) and improved customer experience (29%).



What value does your business get from using Kafka and streaming platforms?

Sample Size : 536 (89% of Respondents)



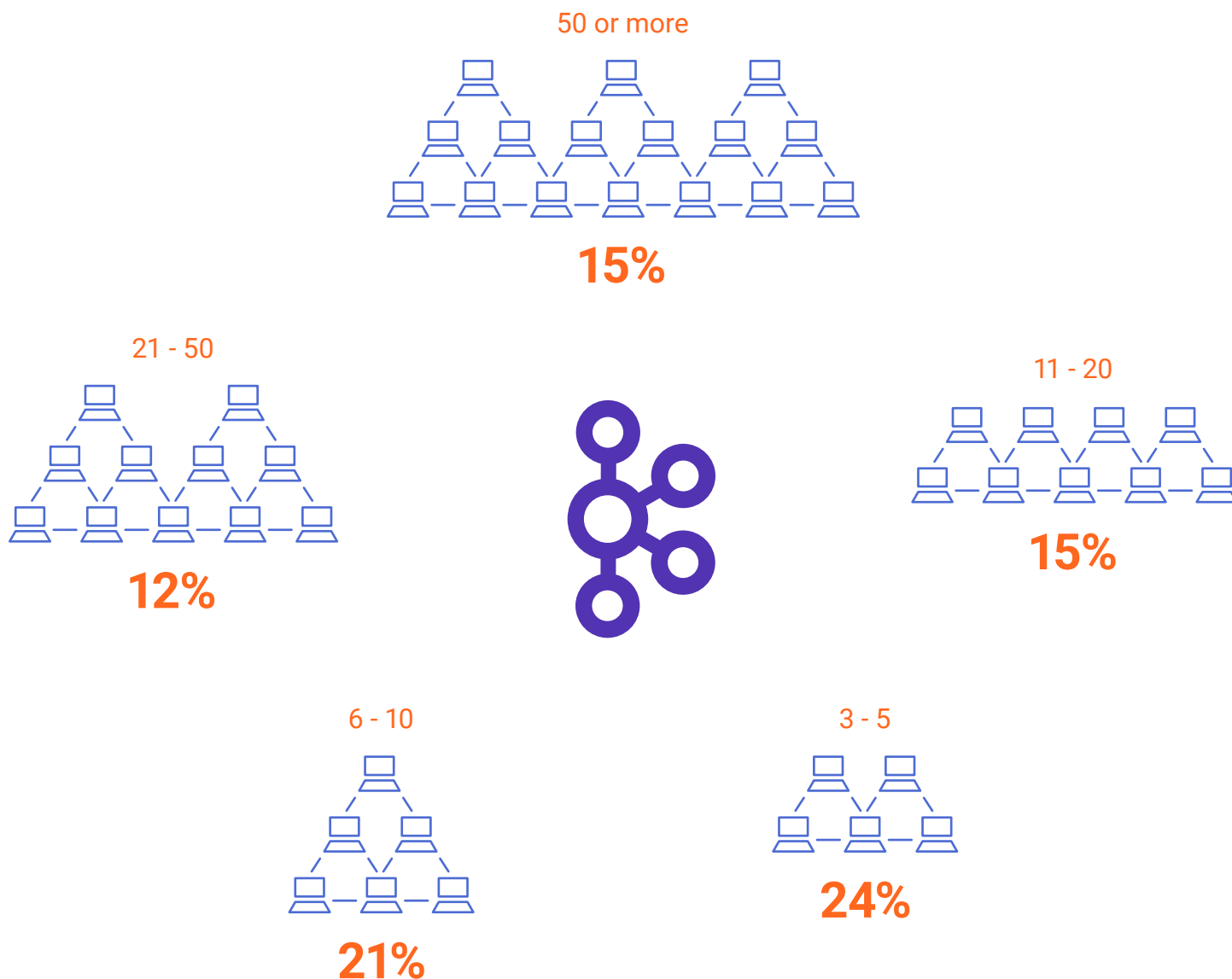
MORE THAN 90% OF COMPANIES WILL IMPLEMENT NEW KAFKA APPLICATIONS:

A wide majority of companies (87%) report at least 3 systems or applications using Kafka today. Nearly half (45%) report between 3 and 10 systems or applications, while 15% report a number of 50 systems or applications using Kafka.



How many systems or applications use Kafka today?

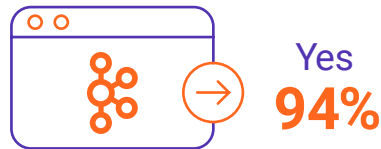
Sample Size : 564 (93% of Respondents)





Are you planning to deploy new applications or services using Kafka in 2018?

Sample Size : 564 (93% of Respondents)



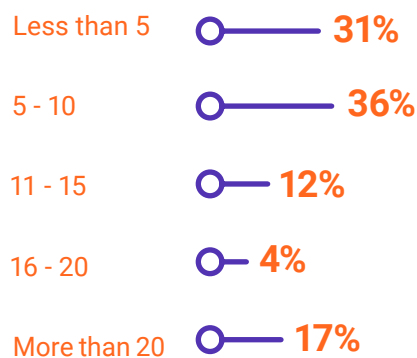
The number of a company's systems or applications using Kafka today has increased dramatically in just the past year: while 14% of organizations reported 21 or more applications or systems using Kafka in 2017, twice as many (28%) did so in 2018.

The vast majority of organizations (94%) plan to deploy new applications or systems using Kafka this year. Two thirds (67%) plan to deploy between 1 and 10 new applications or systems. Asynchronous applications (60%) and data warehouses (40%) are the top two planned new applications for 2018.



How many?

Sample Size : 475 (79% of Respondents)



Today, the more Kafka-powered systems or applications a company has, the more new ones they plan to implement in 2018. 63% of companies with more than 50 existing systems or applications expect more than 20 new systems or applications, in contrast with 9% of all other businesses.

Asynchronous applications are the top type of planned new applications this year – 60% of companies plan to deploy them. Another 40% plan to deploy data warehouses. A third or more of companies will add application monitoring (36%), recommendation and decision engines (34%), and system monitoring (33%).

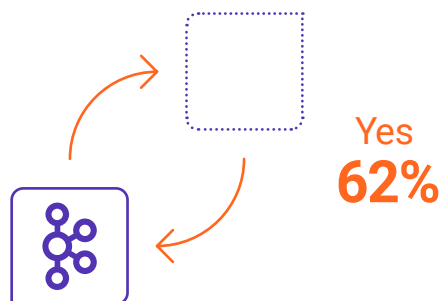
REPLACING TECHNOLOGIES AND RECRUITING STAFF AS KAFKA MATURES:

A majority of respondents (62%) are replacing existing technology within their organization with Kafka deployments. The most common technologies Kafka is replacing are messaging (68%), pub/sub (54%), and ETL (47%).



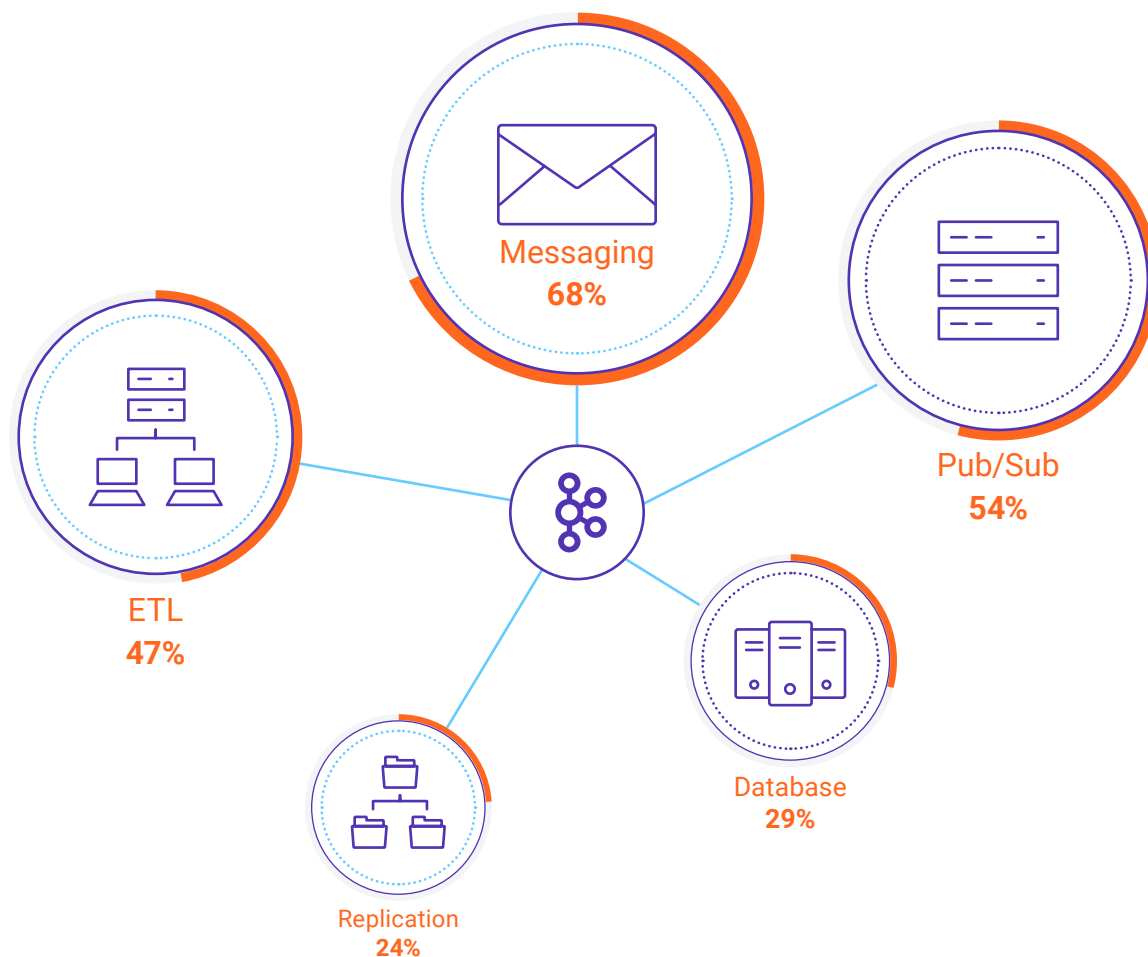
Are your Kafka deployments replacing existing technology within your organization?

Sample Size : 498 (82% of Respondents)



What technologies, if any, are you replacing with Kafka?

Sample Size : 311 (51% of Respondents)

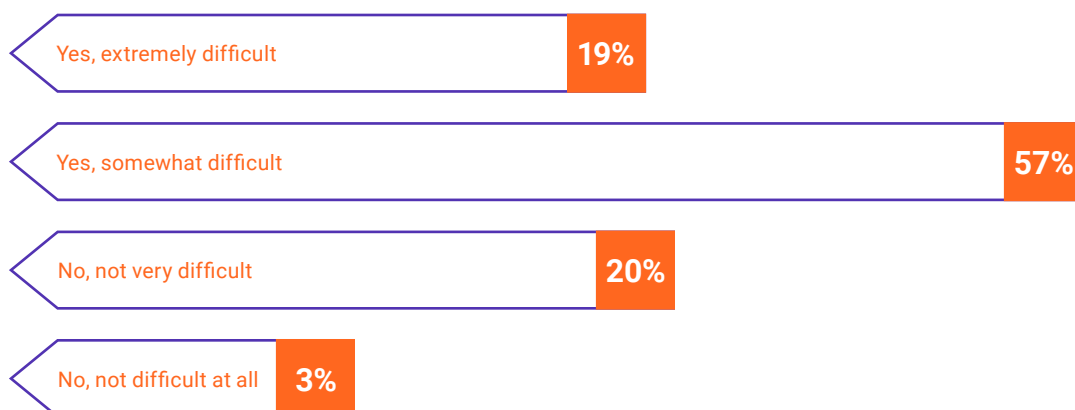


Most Kafka users (85%) say their department is planning to grow in 2018. However, finding the right staff, well versed in Kafka skills, has been somewhat or extremely challenging for most (77%).



Have you found it difficult to find the right talent with Kafka skills?

Sample Size : 537 (89% of Respondents)



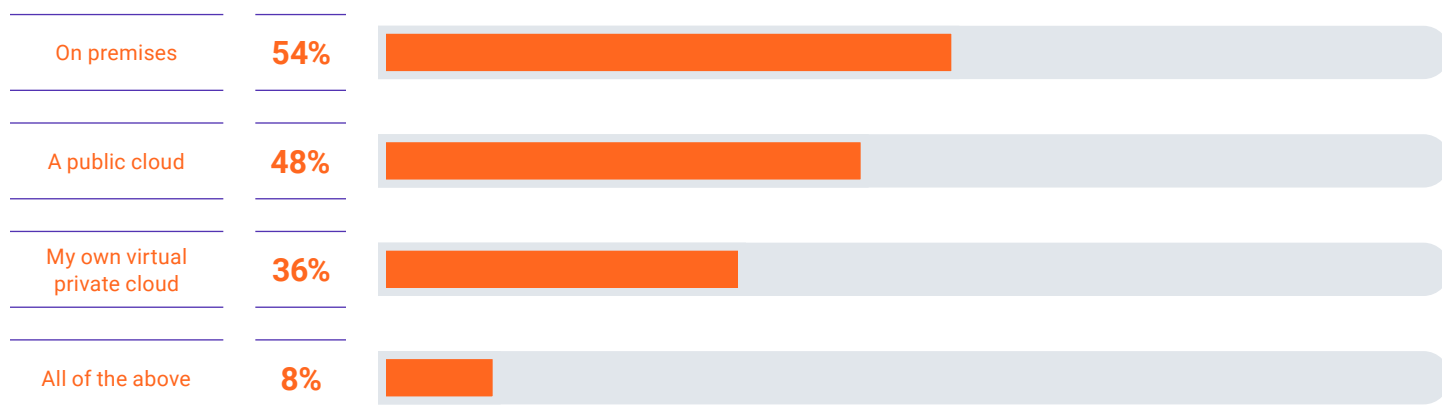
KAFKA USERS DEPLOY THE STREAMING PLATFORM ACROSS A VARIETY OF ENVIRONMENTS:

Just over half of respondents (54%) use Kafka on premises, 48% access it from a public cloud, and a third access it from a virtual private cloud (36%). Another 8% use Kafka in all of these environments.



In which environments are you using Kafka today?

Sample Size : 521 (86% of Respondents)



CLOUD:

Amazon Web Services (56%) is the most common public cloud provider for another year in a row. Google Cloud Platform (17%) and Microsoft Azure (15%) round out the top 3.

In addition, nine out of ten respondents (90%) will store the same amount of data or more in the cloud this year compared to last.



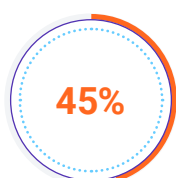
Which public cloud providers do you use, if any?

Sample Size : 521 (86% of Respondents)

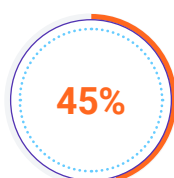


How will your data distribution plans this year compare to last year for the cloud?

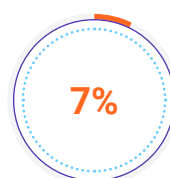
Sample Size : 521 (86% of Respondents)



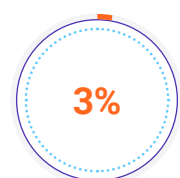
Keep more data
in the cloud



Make no change



Keep less data
in the cloud



Discontinue use of
data in the cloud

ON PREMISES:

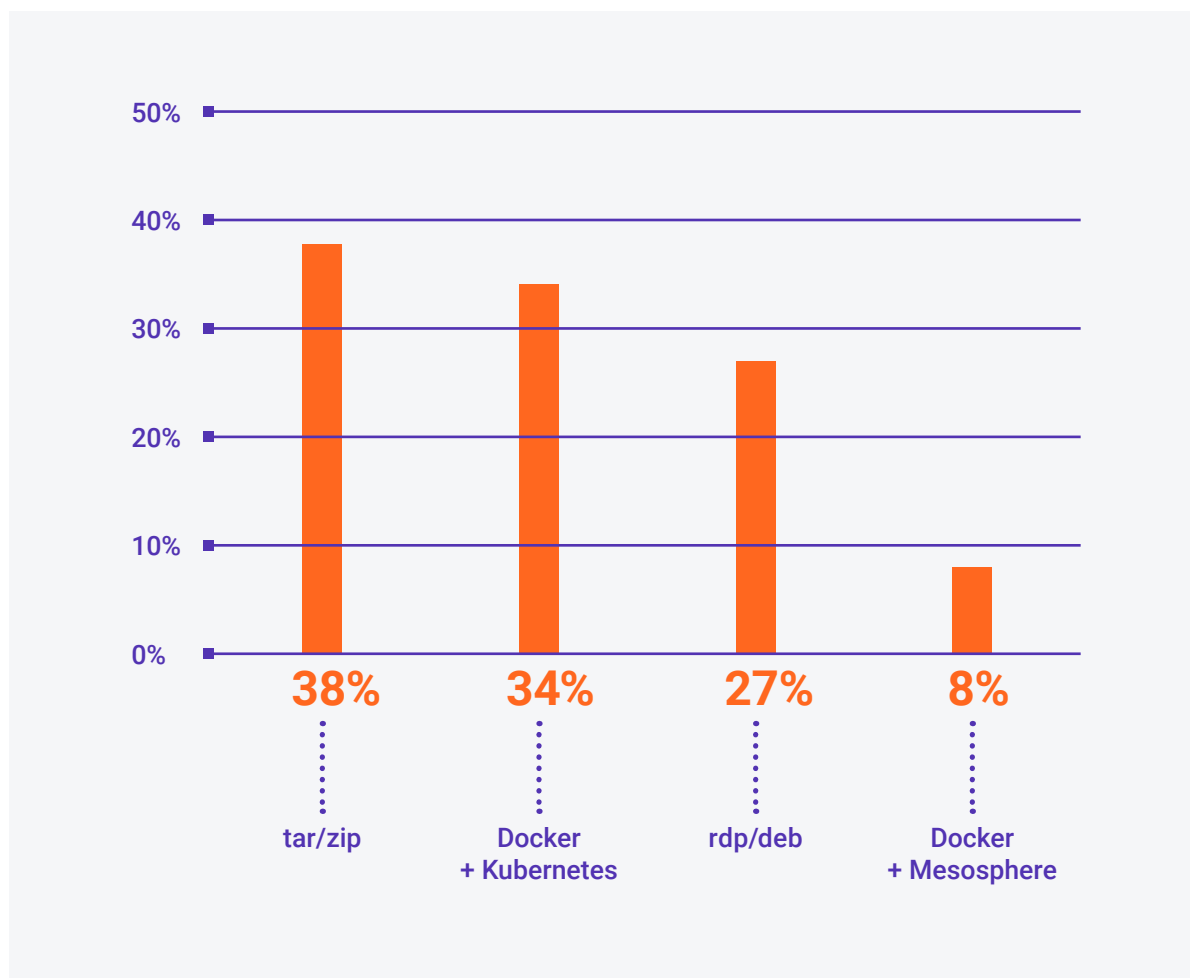
Data distribution plans on premises remain fairly steady from last year: while 50% will keep the same amount of data on premises, about a fifth each will either keep less data there (22%) or keep more (18%). A tenth (9%) will discontinue use of data on their premises.

DEPLOYMENT:

More than a third of companies deploy Kafka with tar/zip (38%) or Docker + Kubernetes (34%).

**How do you deploy Kafka?**

Sample Size : 521 (86% of Respondents)



Conclusion

Today, more than a third of the Fortune 500 and thousands of businesses around the world – ranging from web-scale companies like Yelp, Netflix and Pinterest, to enterprise giants like Goldman Sachs, Cisco and Microsoft – use Kafka to power their business.

As streaming platforms become central to **data strategies**, companies both small and large are re-thinking their architecture with **real-time context at the forefront**.



How large is your organization in annual sales? (U.S. Dollars)

Sample Size : 477 (79% of Respondents)



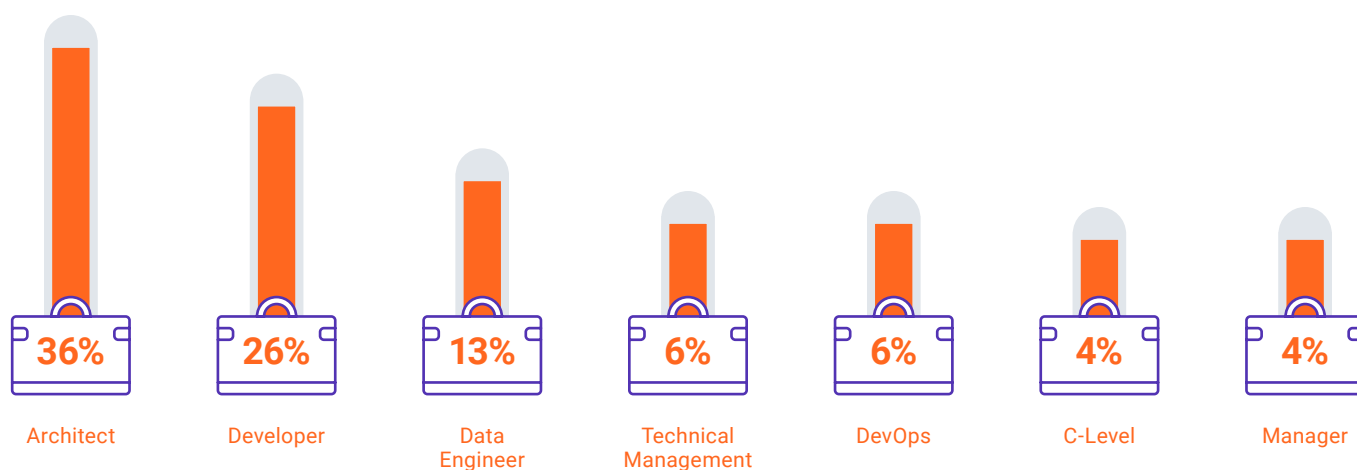
Monoliths are evolving into microservices. Data centers are moving to the cloud. What was once a 'batch' mindset is quickly being replaced with stream processing as the demands of the business impose more and more **real-time requirements** on developers and architects.

Who Took this Survey?



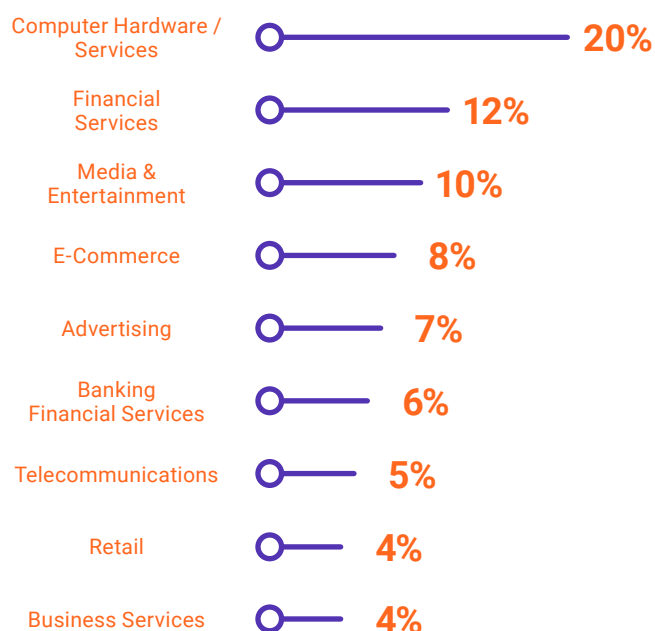
What is your role?

Sample Size : 499 (83% of Respondents)



Which of the following categories best describes your organization's industry?

Sample Size : 477 (79% of Respondents)

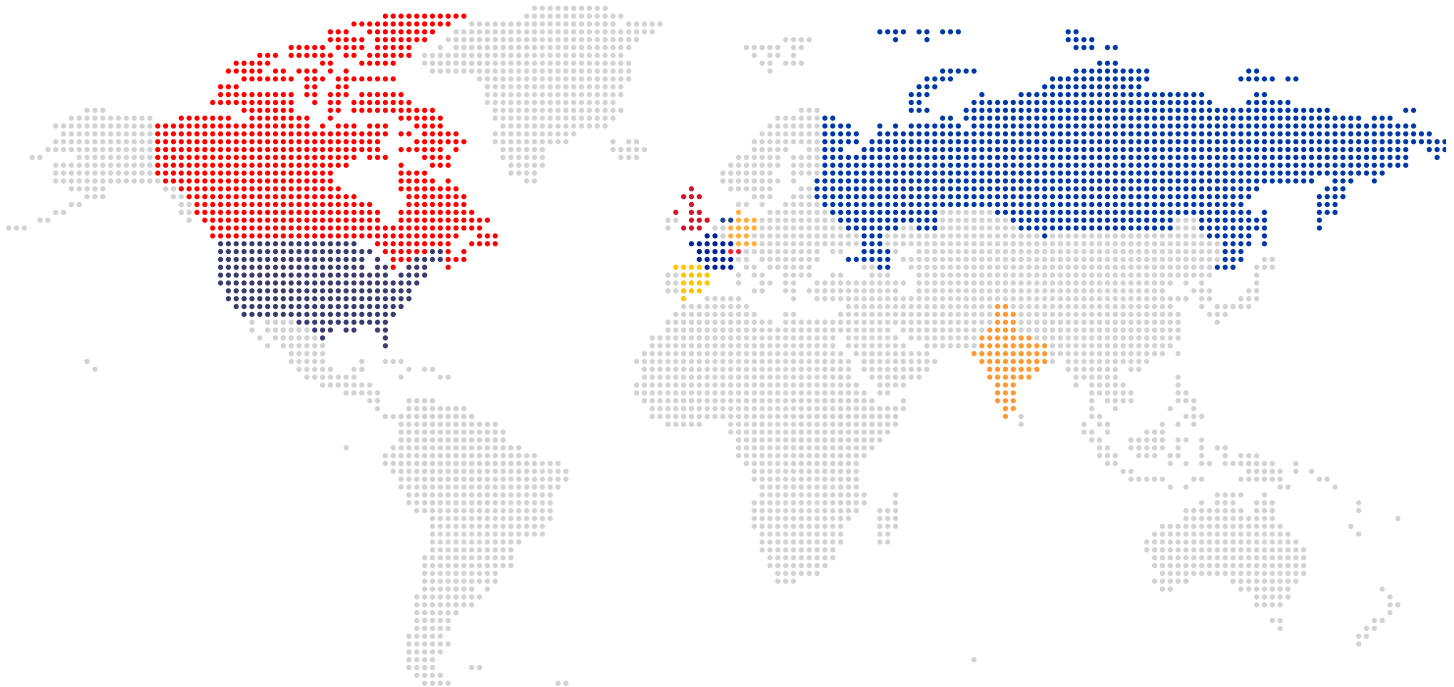




What country is your organization based in?

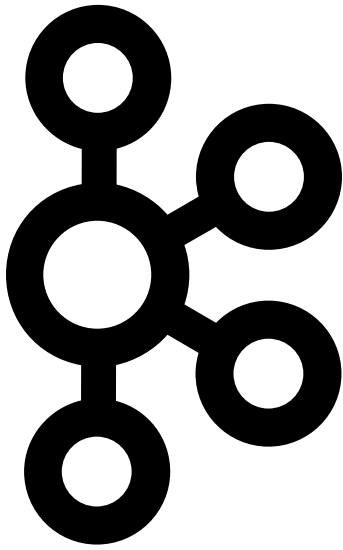
Sample Size : 496 (82% of Respondents)

Top 10 countries using Apache Kafka



UNITED STATES	29%	████████████████████
INDIA	8%	████
GERMANY	7%	████
RUSSIA	6%	████
UNITED KINGDOM	5%	████
NETHERLANDS	4%	████
FRANCE	3%	███
SPAIN	3%	███
CANADA	2%	██
SWITZERLAND	2%	██

Surveyed users and their organizations were most commonly located in the U.S. (29% of respondents), India (8%), Germany (7%), Russia (6%), and the U.K. (5%). **In particular, Russian (6% vs. 1%) and Indian (8% vs. 4%) representation among respondents increased over the past year.**



The Apache Kafka Community

From the project's creation in 2012 until now, the Kafka community has experienced tremendous growth and adoption and is a central nervous system for companies around the world. In fact, Apache Kafka is the 11th busiest Apache repository.

<https://projects.apache.org/statistics.html>

Here is an at-a-glance overview of how the project has grown since inception, its growth over time (at the time of publication), and what's coming up in the future.

KAFKA COMMITTERS

Jun Rao	Jakob Homan	Ewen Cheslack-Postava	Damian Guy
Neha Narkhede	David Arthur	Ismael Juma	Onur Karaman
Joe Stein	Sriram Subramanian	Jason Gustafson	Matthias J. Sax
Jay Krepes	Guozhang Wang	Jiangjie (Becket) Qin	Dong Lin
Joel Koshy	Gwen Shapira	Grant Henke	
Prashanth Menon	Sriharsha Chintalapani	Rajini Sivaram	

PROJECT MANAGEMENT COMMUNITY

Jun Rao	Jakob Homan	Ewen Cheslack-Postava	Damian Guy
Neha Narkhede	David Arthur	Ismael Juma	Onur Karaman
Joe Stein	Sriram Subramanian	Jason Gustafson	Matthias J. Sax
Jay Krepes	Guozhang Wang	Jiangjie (Becket) Qin	Dong Lin
Joel Koshy	Gwen Shapira	Grant Henke	
Prashanth Menon	Sriharsha Chintalapani	Rajini Sivaram	

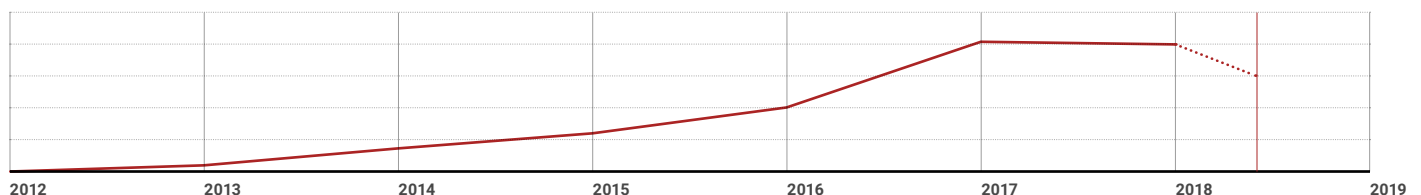


438
Contributors

8491
Stars

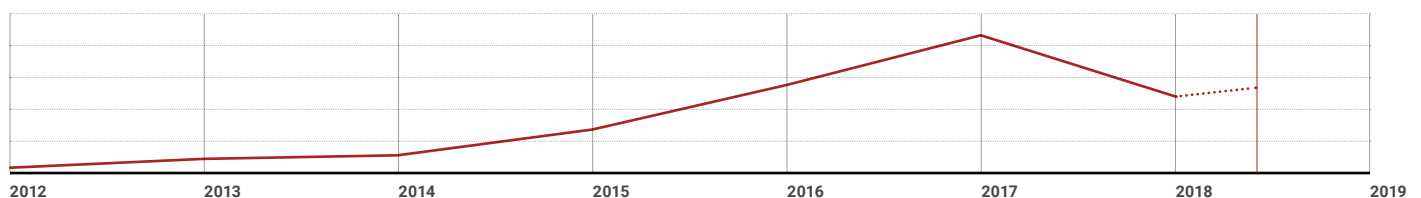
4927
Forks

Number of Contributors Each Year



Number of Commits Each Year

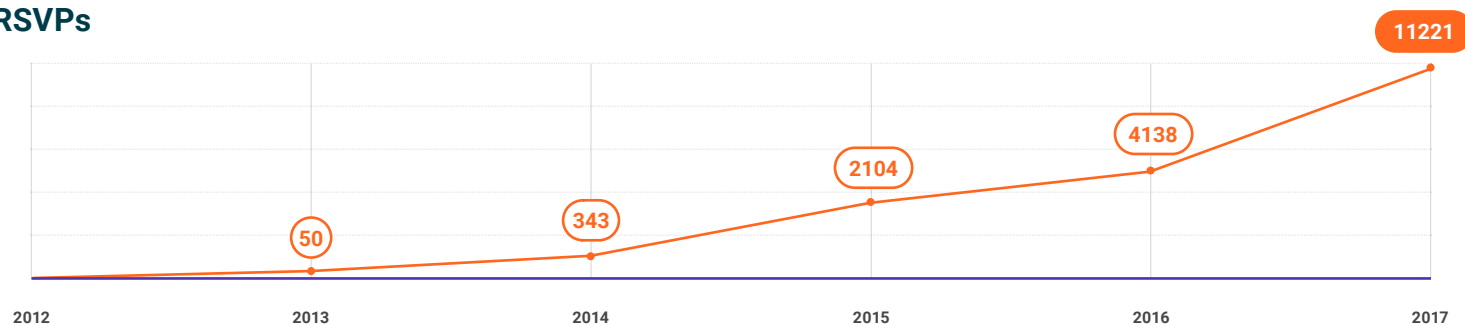
Total Commits: 5,070



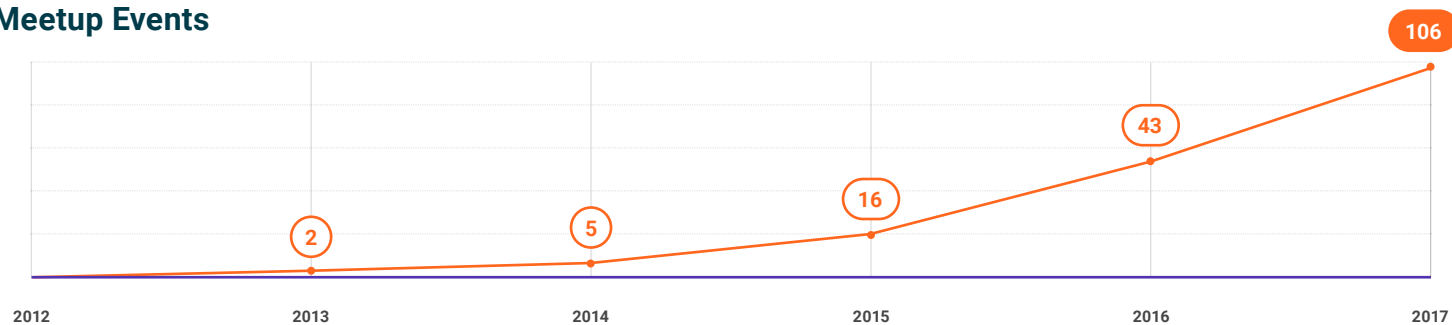
Meetup Stats



RSVPs



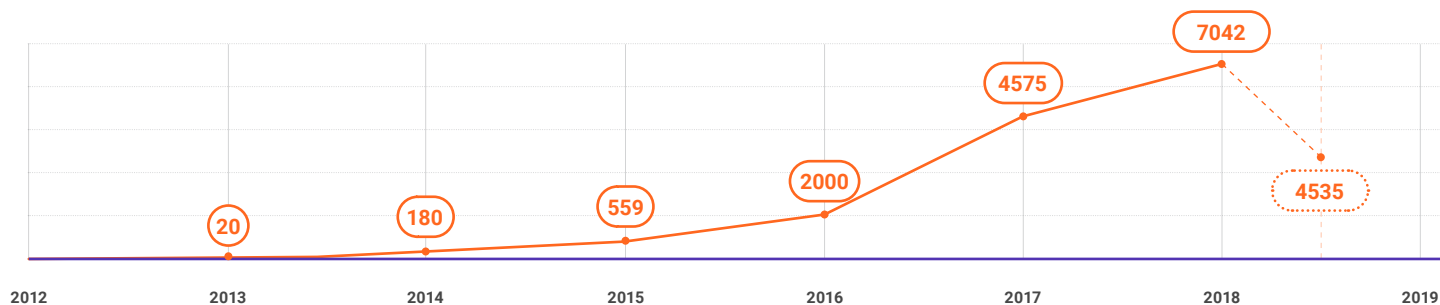
Meetup Events



STACK OVERFLOW

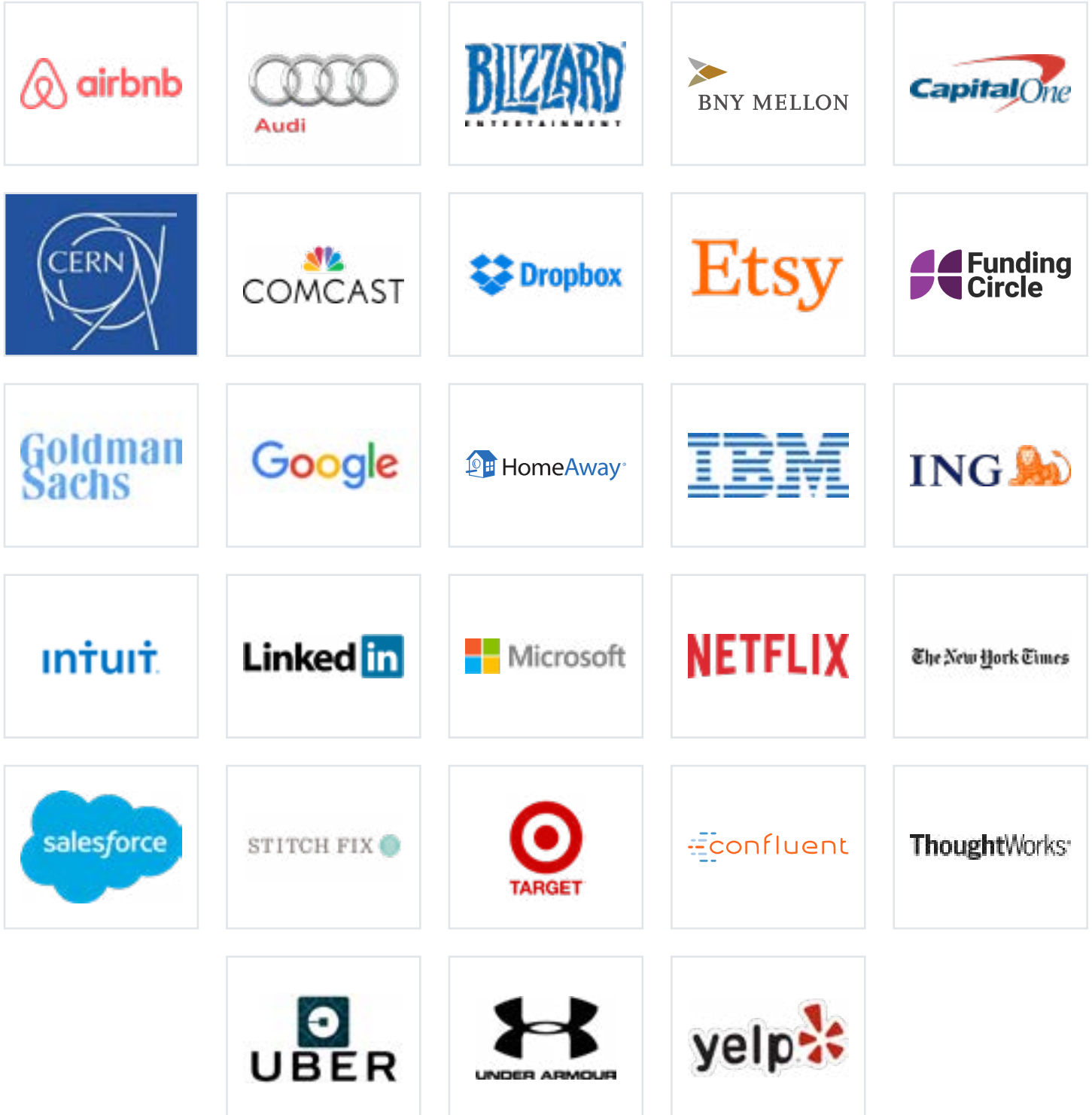
The Kafka community is active on Stack Overflow with 2,445 questions, 16,466 answers and 5,433 people involved since 2012.

Number of questions and answers each year since 2012:



KAFKA SUMMITS

The Kafka community gathers multiple times each year to share best practices, write code and discuss the future of streaming technologies. Since the first event in San Francisco in 2016, the event has hosted over 2,486 Kafka enthusiasts, 68 sponsors and 148 speakers sharing real world use cases of Kafka in action. **Kafka Summit hosted speakers from a variety of companies, including:**



Don't miss Kafka Summit San Francisco taking place October 16-17, 2018. For details, visit:

<https://kafka-summit.org/events/kafka-summit-san-francisco-2018/>

RELEASES AND NEW FEATURES

The Kafka Improvement Proposal (KIP) process ensures everyone's voice in the community is heard. So far, more than 300 KIPs have been proposed and close to 200 KIPs have been adopted.

In the latest 1.1.0 release, 120 people made code contributions.

COMING IN 2018

The Kafka community has been able to contribute to and solve hard problems within Kafka. As the community looks forward, here are some features on the horizon:

Improved replication: the community is working on more efficient replication with semantics allowing for acknowledgement from fewer replicas to enable lower end-to-end latency.

Topic resizing: it will be possible to increase the number of partitions for a topic while preserving ordering guarantees for consumer applications.

Faster controller failover: the controller, which is responsible for electing partition leaders, will be able to recover from failure much quicker than before. This will allow Kafka to support an even higher number of partitions in a single cluster.

Improved ACL configuration: specify ACLs with a prefix and wildcards, so that admins can set ACLs at for groups of topics at a time.

OAuth authentication: Allowing a client to be authenticated by the brokers through an OAuth token.

Apache Kafka, Apache, Kafka, the Kafka logo, and the Apache feather logo are either registered trademarks or trademarks of The Apache Software Foundation. All other products or name brands are trademarks of their respective holders, including The Apache Software Foundation.

About Confluent

Confluent built the central nervous system for the modern real-time enterprise. To get there requires expertise. Knowing the technology and using it to catalyze the next trend. This is not the future state of business. This is now. Confluent is amplifying the power and managing the complexity of Apache Kafka, liberating data, freeing developers, operators and the C-suite to harness the value of real-time streaming events with astonishing results.

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

