

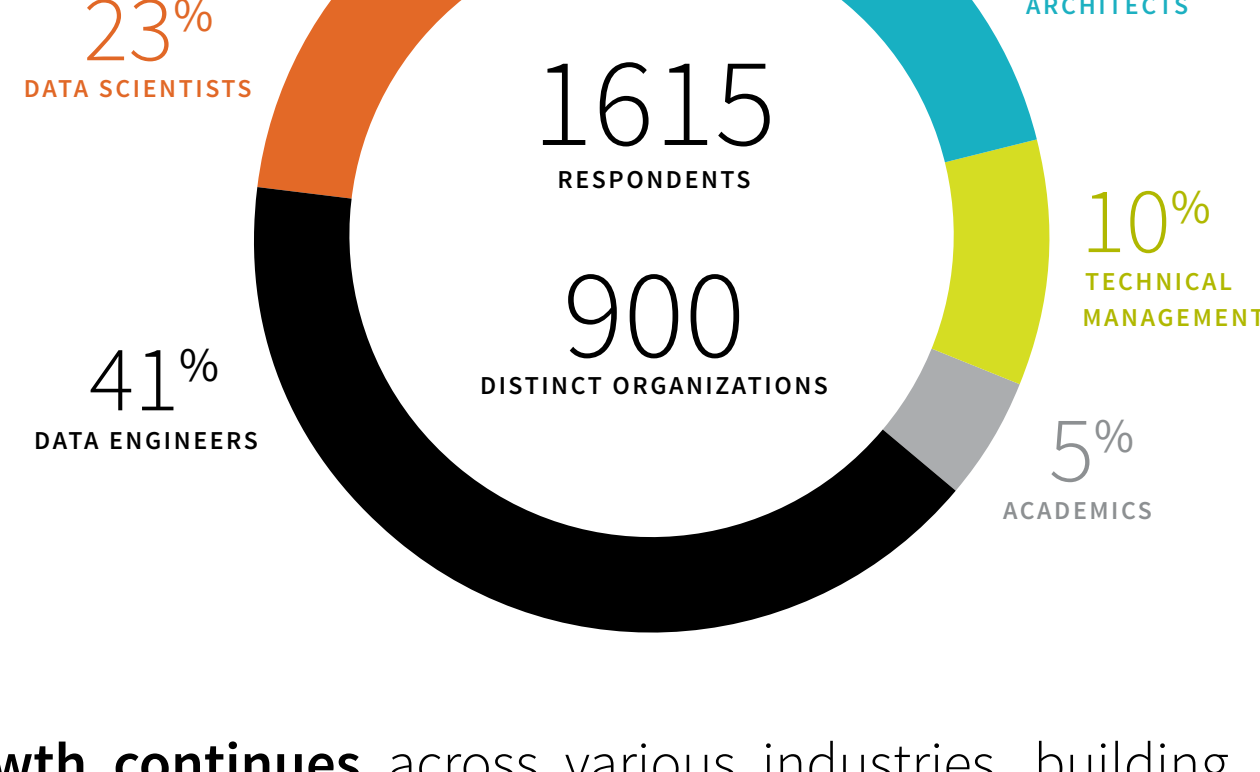
# APACHE SPARK<sup>®</sup> SURVEY 2016

## REPORT HIGHLIGHTS

View the full report at [databricks.com/2016-spark-survey](http://databricks.com/2016-spark-survey)



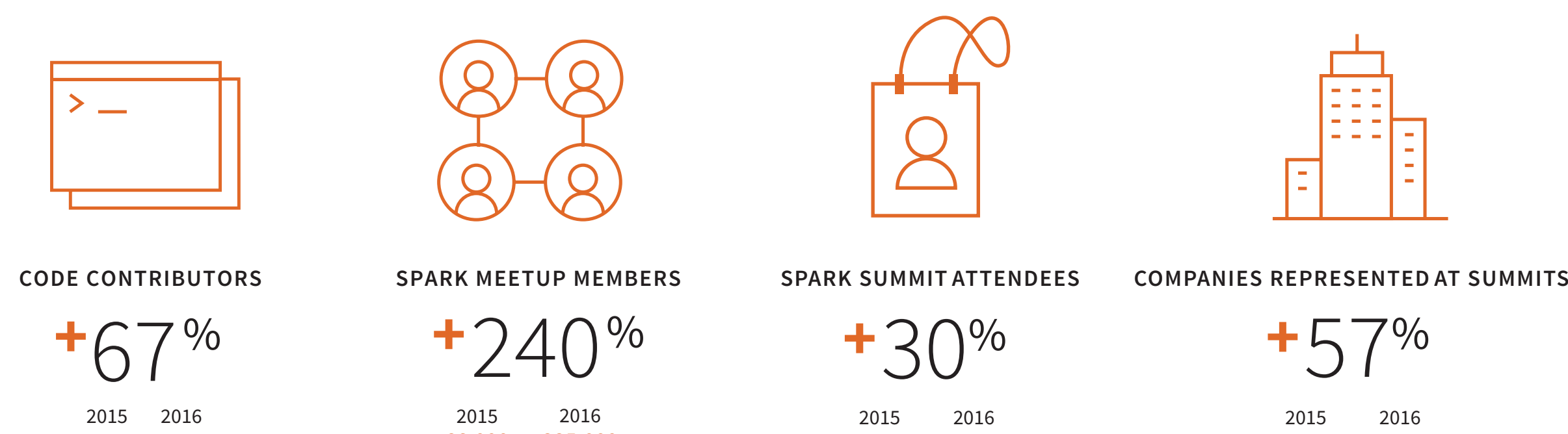
Databricks ran our Apache Spark Survey 2016 this summer to identify how organizations are using Apache Spark. This year's results reflect the answers and opinions of 1615 respondents representing over 900 organizations. Survey respondents were predominantly Apache Spark users.



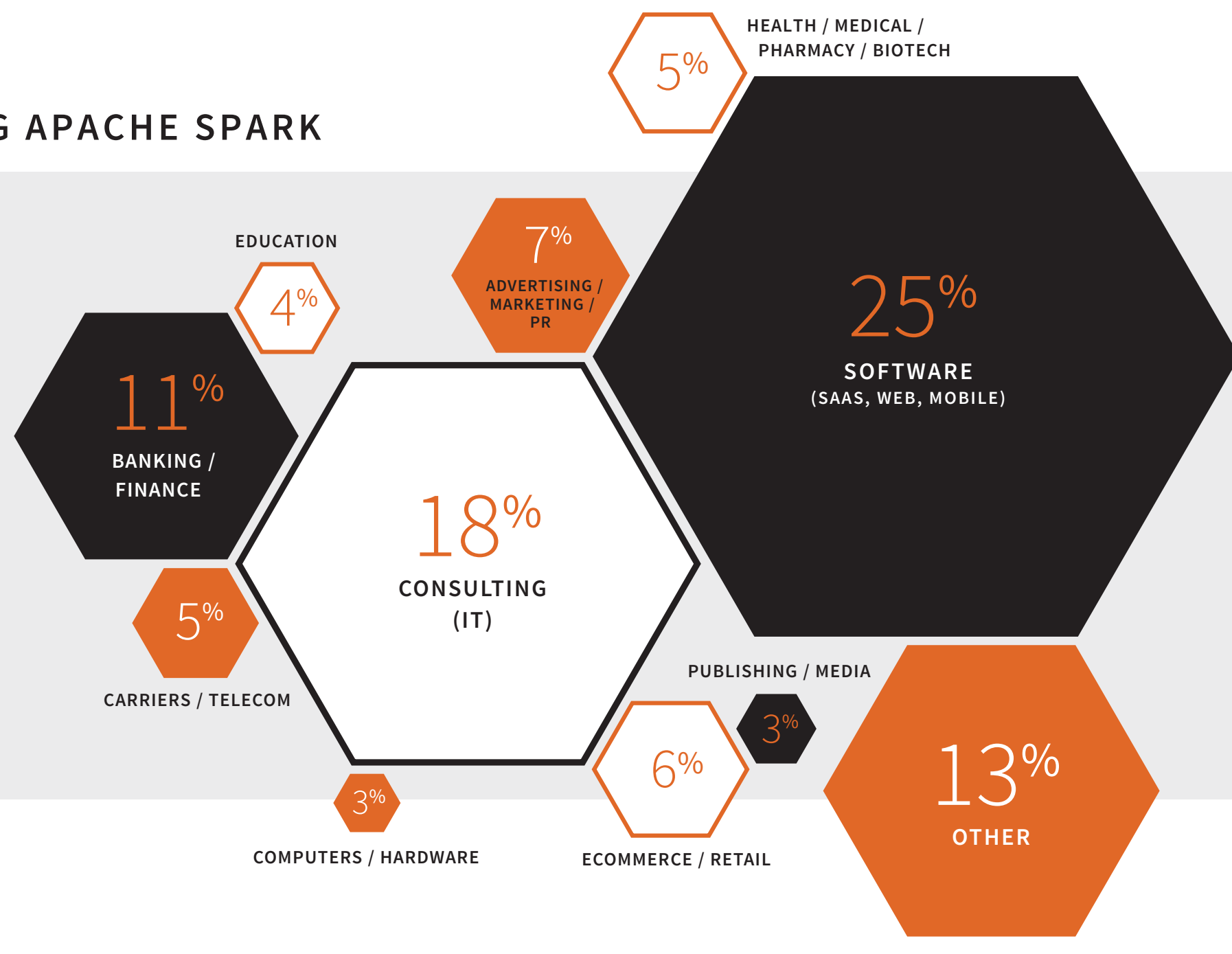
The survey results suggest that Spark's growth continues across various industries, building sophisticated data solutions by people in various functional roles. Using multiple Spark components, many users build different types of products in production. The results indicate that Spark has moved well beyond the early-adopter phase at high-tech companies and is now mainstream in large data-driven enterprises. And with the rise of public cloud computing, the survey findings reflect that users show an increased affinity toward using Spark in the public cloud.

### APACHE SPARK'S GROWTH CONTINUES

Spark today remains the most active open source project in Big Data. Today, there are over 1000 Spark contributors, compared to 600 in 2015 from 250+ organizations. With such large numbers of contributors and organizations investing in Spark's future development, it has engaged a community of developers globally.

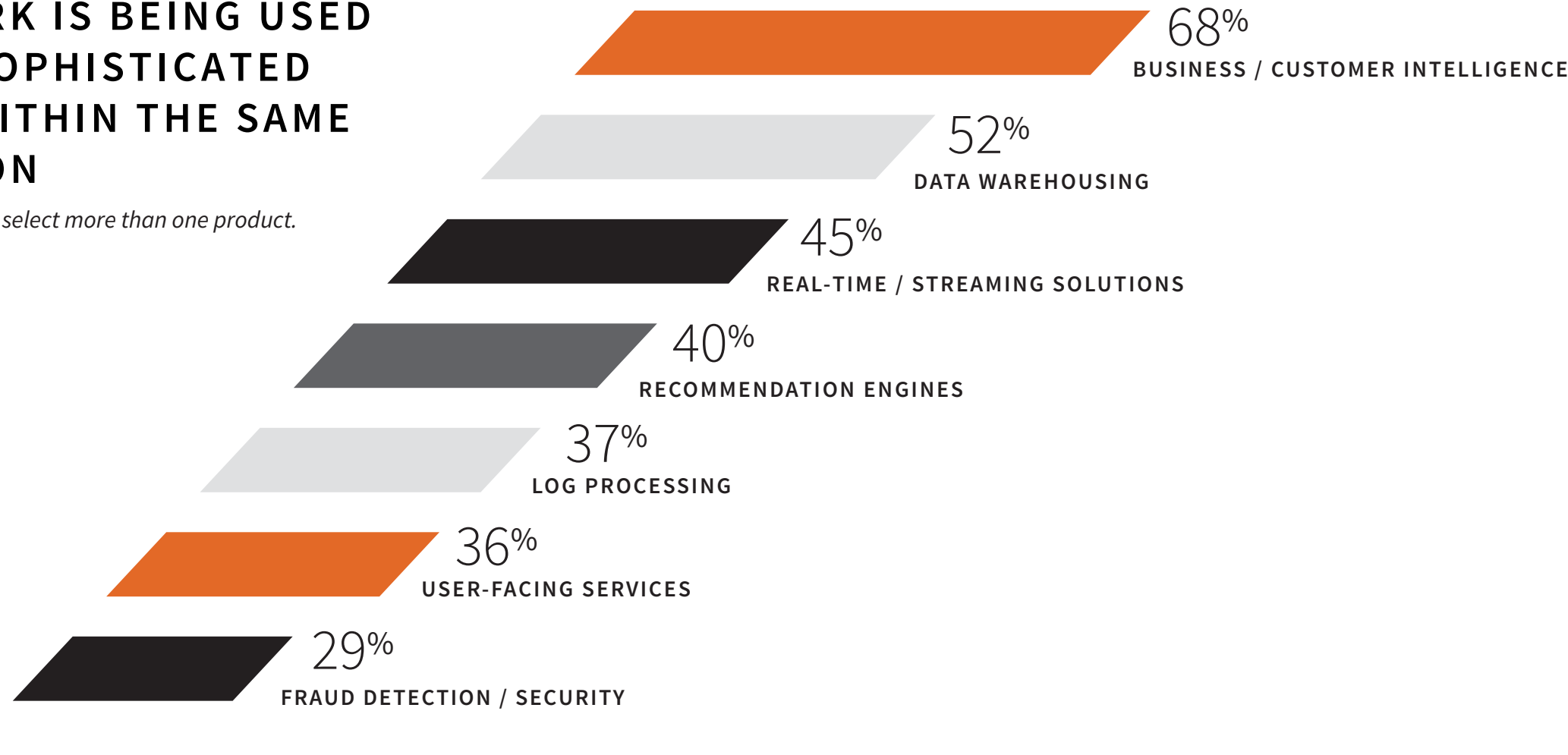


### KEY INDUSTRIES USING APACHE SPARK

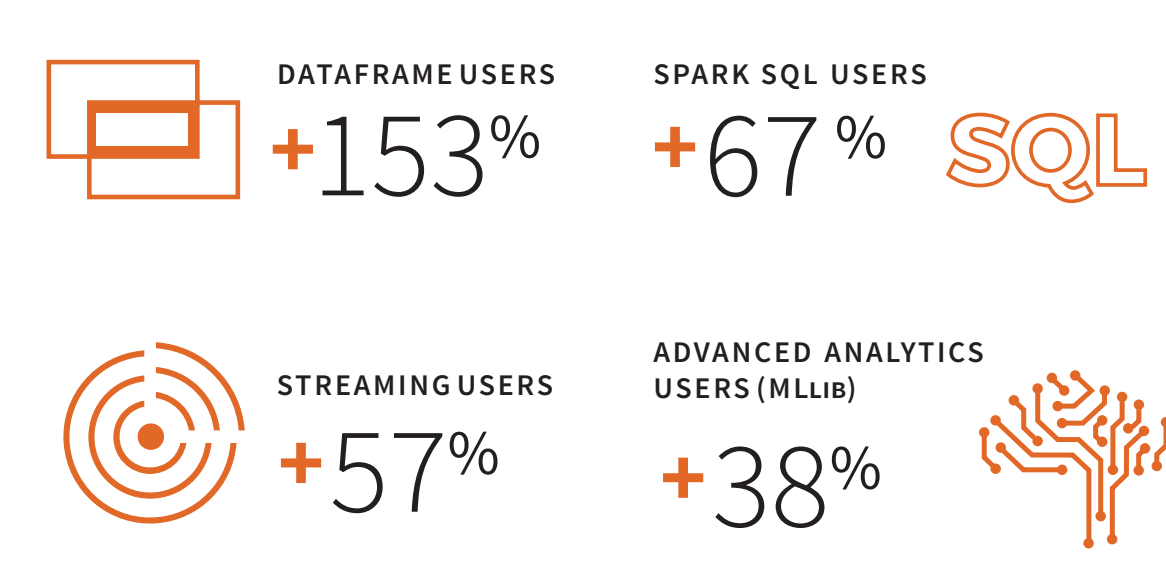


### APACHE SPARK IS BEING USED TO CREATE SOPHISTICATED PRODUCTS WITHIN THE SAME ORGANIZATION

Respondents were allowed to select more than one product.

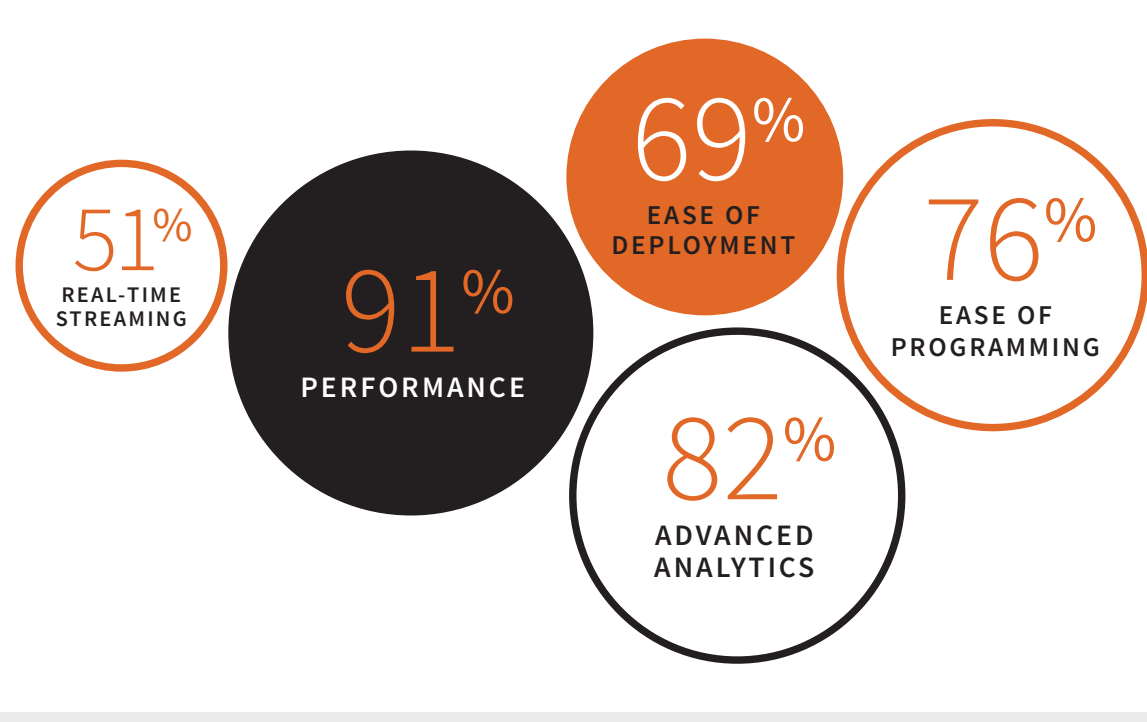


### APACHE SPARK'S FASTEST GROWING AREAS IN 2016



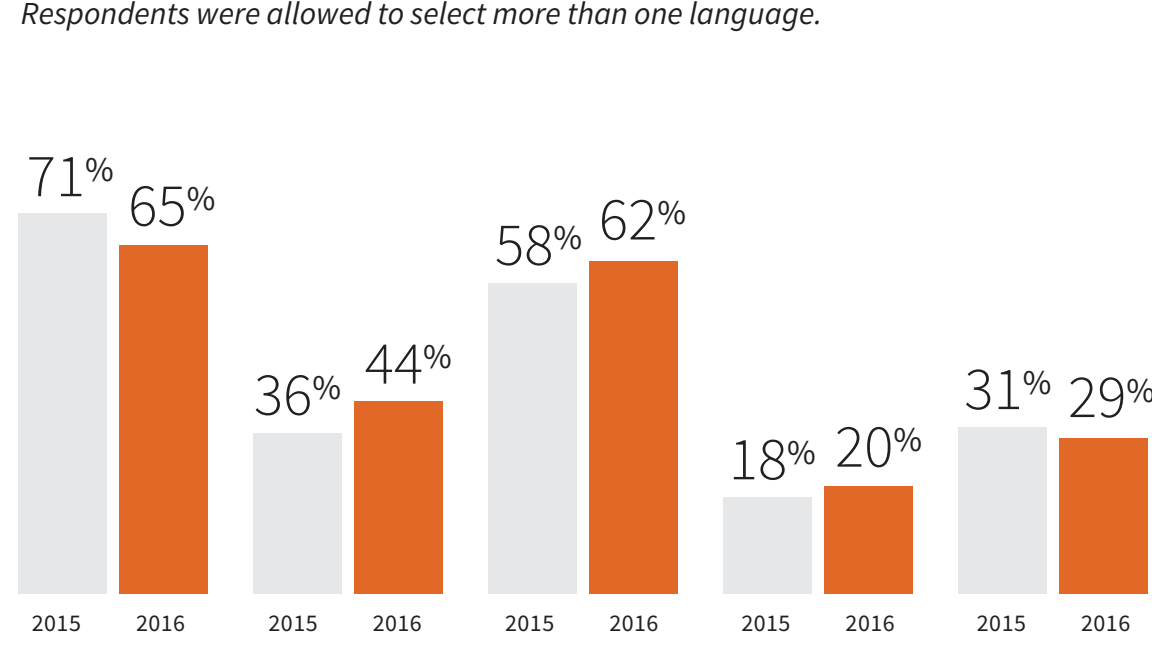
### FEATURES USERS CONSIDER IMPORTANT

Respondents were allowed to select more than one feature.



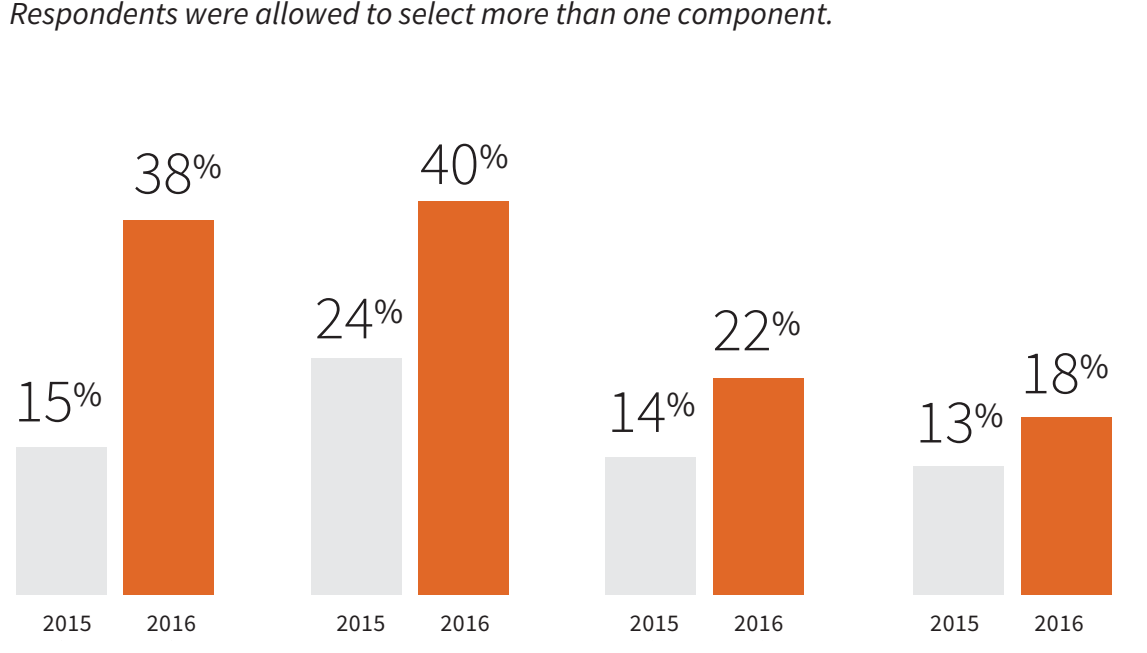
### LANGUAGES USED IN APACHE SPARK

Respondents were allowed to select more than one language.

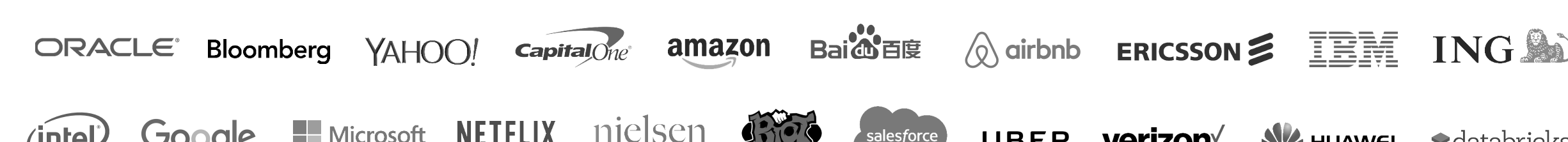


### SPARK COMPONENTS USED IN PRODUCTION

Respondents were allowed to select more than one component.



### NOTABLE APACHE SPARK USERS WHO PRESENTED AT SPARK SUMMIT 2016

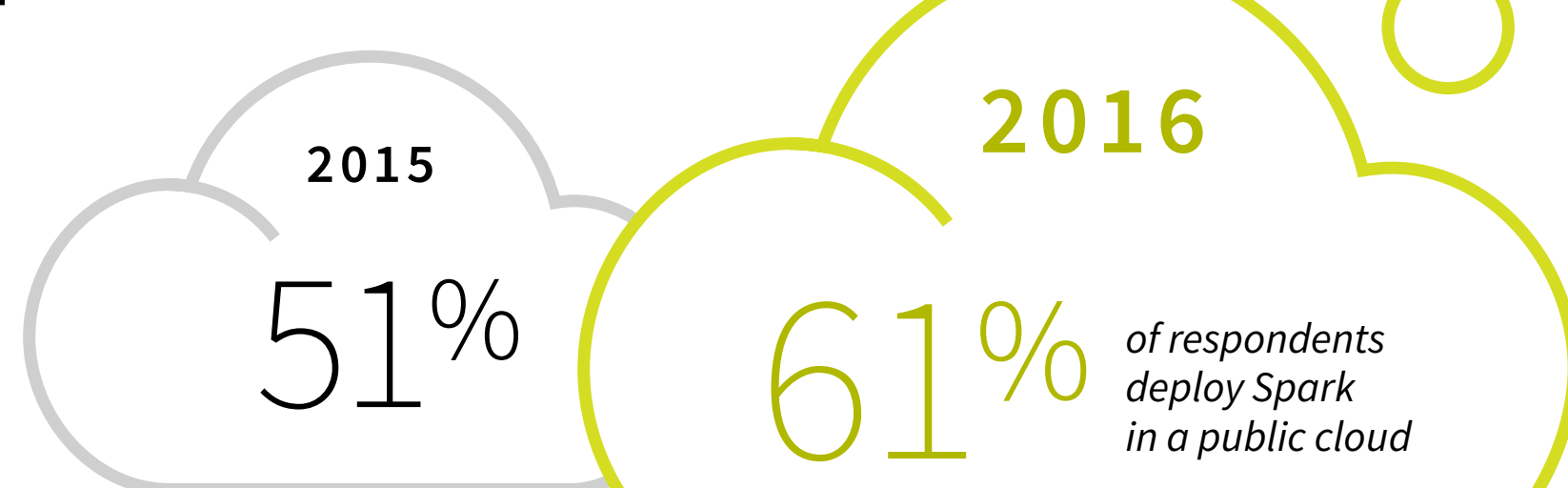


Apache Spark's growth and adoption continues as users, industries, development environments, disciplines, and programming languages embrace its **ease of use and programming**; its **unified compute engine**; and its **performance to solve complex data problems at scale**.

### APACHE SPARK IN THE CLOUD IS GROWING

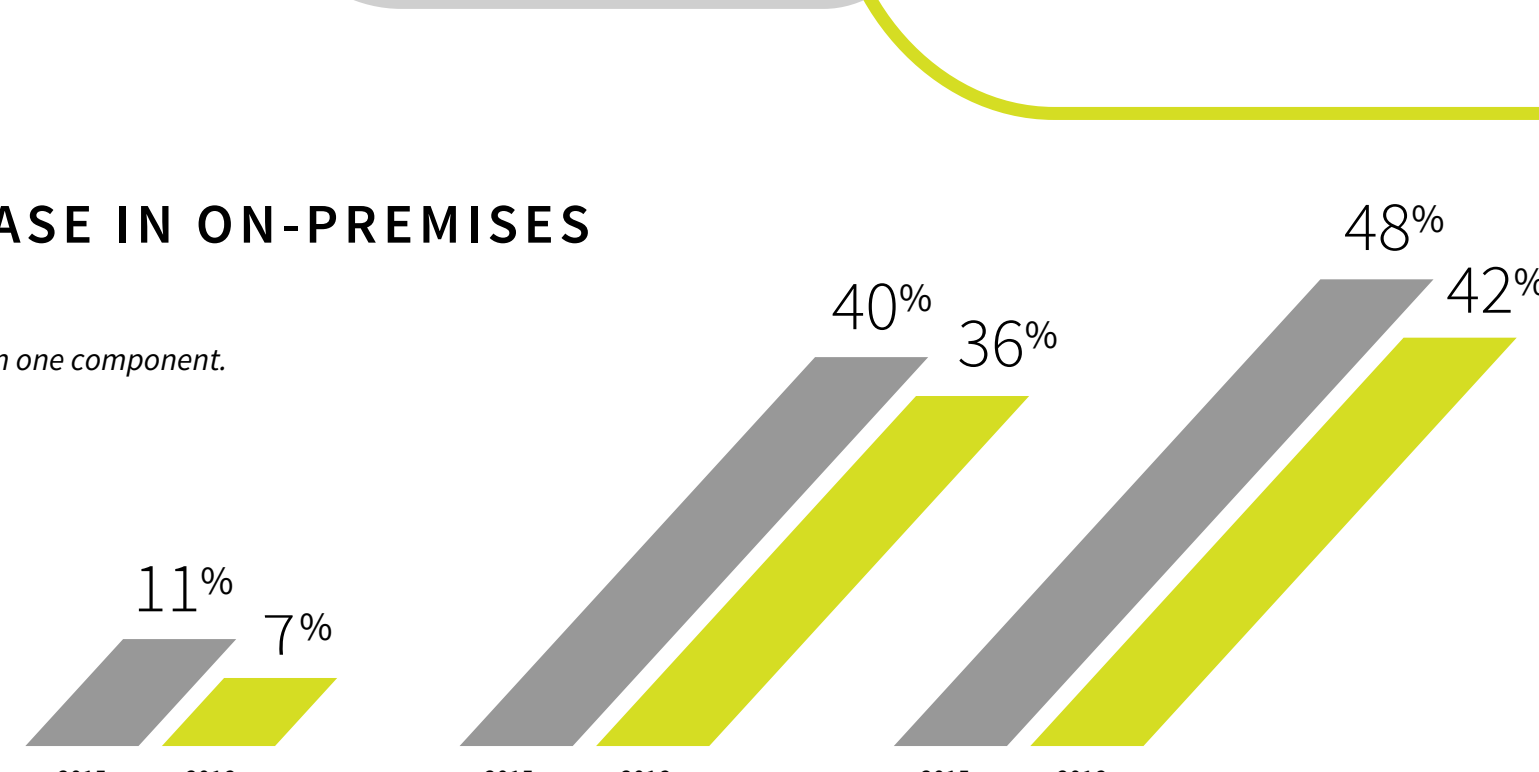
Apache Spark deployments in the public cloud increased in 2016. In contrast, the percentage of Spark deployments on-premises decreased in the past year.

### APACHE SPARK DEPLOYMENT IN PUBLIC CLOUDS HAS INCREASED BY 10% SINCE 2015.



### PERCENTAGE DECREASE IN ON-PREMISES DEPLOYMENTS

Respondents were allowed to select more than one component.

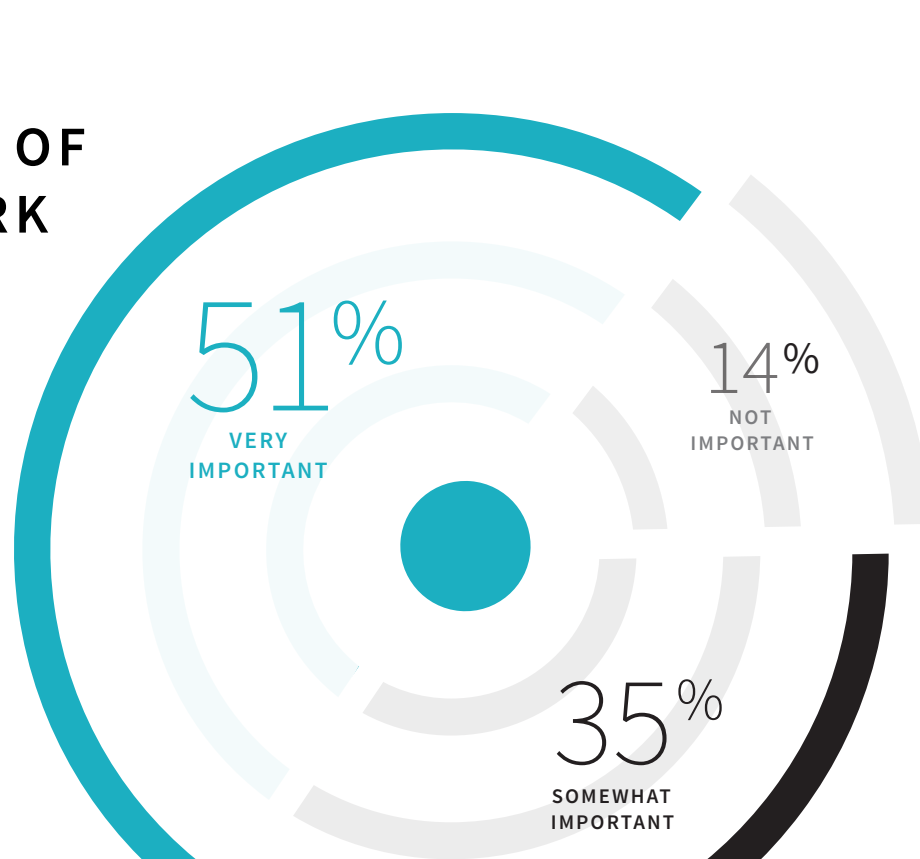


The rise of cloud computing is rapid, inexorable and causing a huge upheaval in the tech industry, writes *The Economist*. We observed this trend reflected in the survey results, as many respondents elect to deploy Spark in the public cloud, reaping its many benefits. Not only do cloud deployments have lower deployment costs and fewer management headaches, they have higher and proven performance benefits.

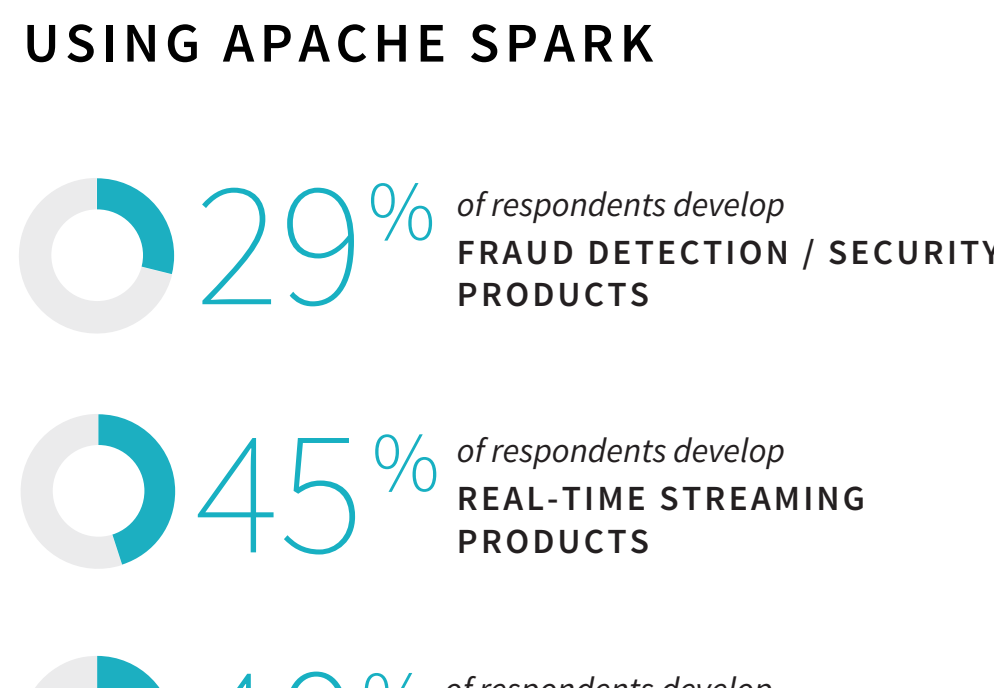
### APACHE SPARK STREAMING AND MACHINE LEARNING SURGE IN USAGE

Apache Spark Streaming is growing. Since its release, Spark Streaming has become one of the most widely used distributed streaming engines. Interest in developing real-time applications and advanced analytics is on the rise.

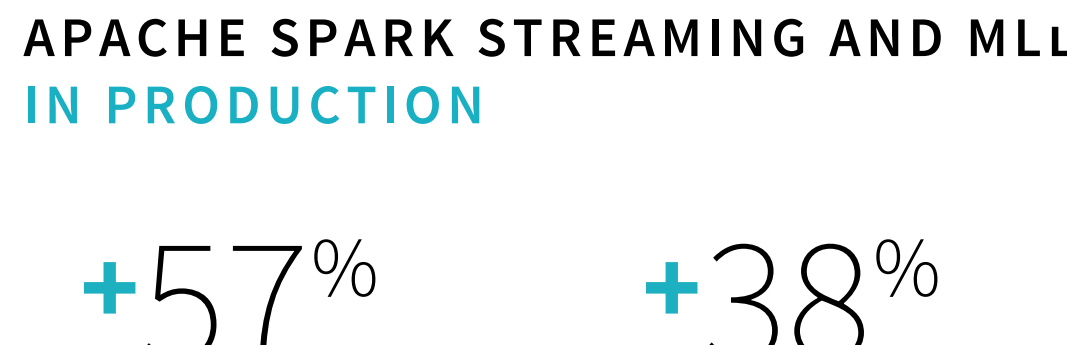
### IMPORTANCE OF APACHE SPARK STREAMING TO USERS



### TYPES OF PRODUCTS DEVELOPED USING APACHE SPARK



### APACHE SPARK STREAMING AND MLlib USE IN PRODUCTION



Spark Streaming is being used for real-time solutions, from evaluation to production, closer in usage to Spark's other commonly used components. As a preferred choice of streaming engine over others, more organizations are building real-time streaming solutions as they consider streaming an important Spark feature.

2015 and 2016 have been exciting years for the adoption and increased growth of Apache Spark and its community. Your voice matters. We got an insightful glimpse into the growth and trends from this year's survey: who's using Spark, how they are using it, what's important, what new features they use, and what they are using it for. Thank you to everyone who participated in the Databricks 2016 Apache Spark Survey!