instaclustr

Apache Kafka

Project Assessment Report

Document Identifier: ICF_KPA

Document Version: 1.0

Issue Date: 11 JUN 2019

Table of contents

1   Introduction

1.1   Overview

The **Instaclustr Certification Framework for Open Source Software (ICF-OSS)** provides an analytical basis for selecting and recommending various open source technologies for production usage. This is achieved by:

a) assessing various open source projects to gain a level of assurance that the project has the foundation and capability to build and sustain production-grade software, and

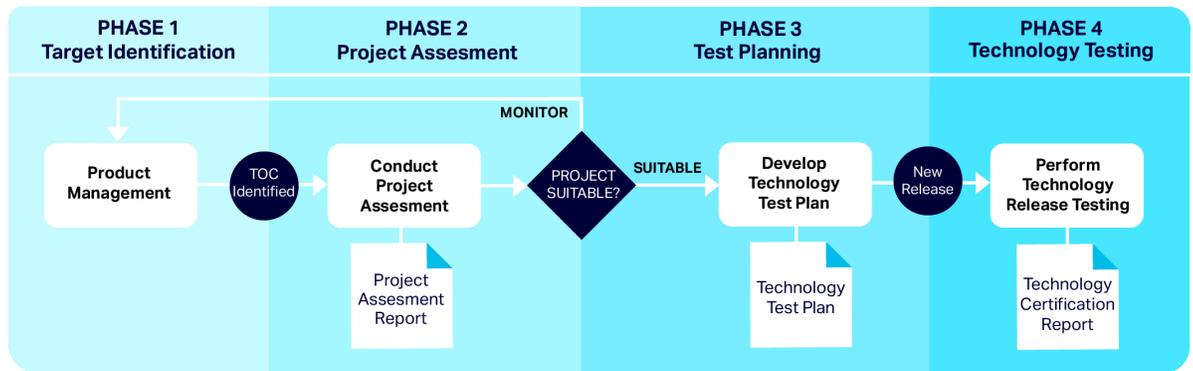b) testing specific versions of open source technologies for function, performance, and interoperability.

Figure 1: Instaclustr Certification Framework Process

1.2   Document Identification Information

Identifier	ICF_KPA – Apache Kafka Project Assessment Report
Description	This document is the Project Assessment Report for Apache Kafka as part of the Instaclustr Certification Framework for Open Source Software (ICF-OSS).
Target of Certification	Apache Kafka http://kafka.apache.org/

1.3   Copyright Statement

© Instaclustr Pty Limited, 2019

Except as permitted by the copyright law applicable to you, you may not reproduce, distribute, publish, display, communicate or transmit any of the content of this document, in any form, but any means, without the prior written permission of Instaclustr Pty Limited.

2 Introduction

2.1 Overview

This document is the Project Assessment Report (PAR) for Apache Kafka that has been developed as part of Phase 2 of the certification process, the Instaclustr Certification Framework.

This document reports the result of Instaclustr's assessment of the health of the Apache Kafka project and community as a source of software for significant production usage.

2.2 Scope

This project assessment uses the following broad criteria to assess the health of the Apache Kafka project:

- Project maturity
- Development & maintenance
- Community
- Licensing model
- Governance

2.3 Purpose

The project assessment phase of the ICF-OSS focuses on the overarching governance and management of the project as a whole. This phase assesses the non-technical characteristics of the project rather than individual technology releases.

The intent is to ensure that all open source technologies that are integrated into the Instaclustr Managed Platform are being effectively managed and controlled, that risks associated with licensing and use are suitable and that there are strong indications the project will enjoy ongoing, open community support in the foreseeable future.

2.4 Review & Updates

This assessment and report will be updated on an annual basis, or if any significant changes occur in the Apache Kafka Project.

3 Assessment Results

3.1 Project maturity

3.1.1 Observations

Kafka was first open sourced by LinkedIn in 2011 and became an Apache Foundation a top-level project in October 2012.

Apache Kafka is very widely adopted with large scale, publicly acknowledge large-scale users including:

- LinkedIn
- New York Times
- Netflix
- Twitter

Some of these companies are know the run thousands of nodes of Kafka.

Other indicators of the adoption of Kafka include:

- Included in 2.2k stacks on stackshare.io (compared to, for example, 13.9k for Java - a very broadly adopted technology or 1.4k stacks for Apache Cassandra).
- AWS (who have very broad visibility in to open source technology usage) introducing a Kafka offering in 2018.

3.1.2 Assessment

The Apache Kafka project is considered to be mature with a significant user base and deployment.

The Apache Kafka project maturity is assessed as **EXCELLENT**.

3.2 Development and maintenance

3.2.1 Observations

As at February 2019, there had been more than 25 releases of Apache Kafka. The project was very active in 2018 with 7 releases include a major version release to 2.x. The level of activity can be seen in the following graph from Github and show a relatively constant strong rate of contribution.

Figure 2: Github activity for the Apache Kafka project.

3.2.2   Assessment

The velocity of development and release has increased significantly in the last 18 months. The project is highly active with a broad community of large-scale users.

The ongoing development and maintenance of the Apache Kafka project is assessed as **EXCELLENT**.

3.3   Community

3.3.1   Observations

Apache Kafka has a broad and active community. Indicators include:

- The Kafka user mailing list sees around 300 emails per month:
  - 2018: 3,510
  - 2017: 5,867
  - 2016: 7,252
  - 2015: 6,045
- 405 Kafka related meetups listed on meetup.com
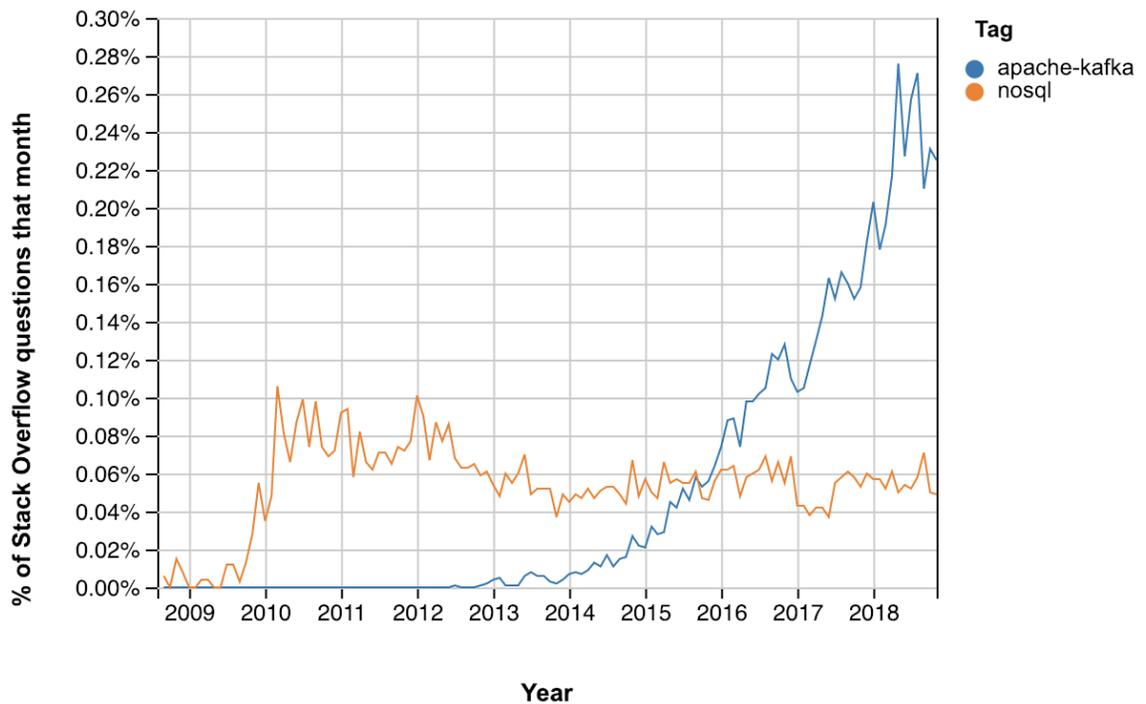- Stack Overflow trends demonstrates an active flow of questions:

Figure 3: Stack Overflow trend for Apache Kafka versus NoSQL

The Stack Overflow statistics in particular show a strong and rapidly increasing level of interest in Kafka.

The broader ecosystem surrounding Kafka is also strong and growing quickly. There are multiple companies with Kafka as a core focus of their business (for example Confluent, Landoop and Instaclustr). There are also many consulting companies offering Kafka consulting services, some with Kafka as a primary focus of their business.

3.3.2 Assessment

The Apache Kafka project has a wide and varied active community which is growing stronger at a significant and rapid rate.

The community is assessed as **EXCELLENT**.

3.4 Licensing model

3.4.1 Observations

Apache Kafka is licensed under the Apache 2.0 license. This license is very open and, in Instaclustr's view, strongly encouraging of use and contribution to the software.

A good summary of licence conditions can be found here:
https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)

3.4.2 Assessment

The Apache Software Foundation's Apache 2.0 license considered the benchmark for open source technologies.

The licensing model is assessed as **EXCELLENT**.

3.5 Governance

3.5.1 Observations

The Apache Kafka project is an Apache Foundation project and governed according to The Apache Way (http://www.apache.org/foundation/governance/). This governance mechanism has been proven to allow multiple parties to contribute to project in a meritocracy of ideas and contributors and has generated many extremely successful projects (for example Hadoop, Open Office and Tomcat).

The Project Management Committee for Kafka has representatives from multiple organisations although a significant majority of PMC members are employed by Confluent.

3.5.2 Assessment

The governance, control and openness associated with the Apache Software Foundation is considered the benchmark for open source technologies.

The governance model is assessed as **EXCELLENT**.

4   Recommendation

The Apache Kafka project is considered **HIGHLY SUITABLE** for inclusion on the Instaclustr Managed Platform.

CRITERIA	ASSESSMENT	RESULT
PROJECT MATURITY	The Apache Kafka project is considered to be mature with a significant user base and deployment.	**EXCELLENT**
DEVELOPMENT & MAINTENANCE	The velocity of development and release has increased significantly in the last 18 months. The project is highly active with a broad community of large-scale users.	**EXCELLENT**
COMMUNITY	The Apache Kafka project has a wide and varied and highly active community.	**EXCELLENT**
LICENSING MODEL	The Apache Software Foundation's Apache 2.0 license considered the benchmark for open source technologies.	**EXCELLENT**
GOVERNANCE	The governance, control and openness associated with the Apache Software Foundation is considered the benchmark for open source technologies.	**EXCELLENT**