

Protective Technology with Apache Kafka



Bogdan Sucaciu

Tech Lead

@bsucaciu

bsucaciu.com







Creator: github.com/apache/kafka

Apache Kafka is an open-source distributed event streaming platform used by thousands of companies for high-performance data pipelines, streaming analytics, data integration, and mission-critical applications.





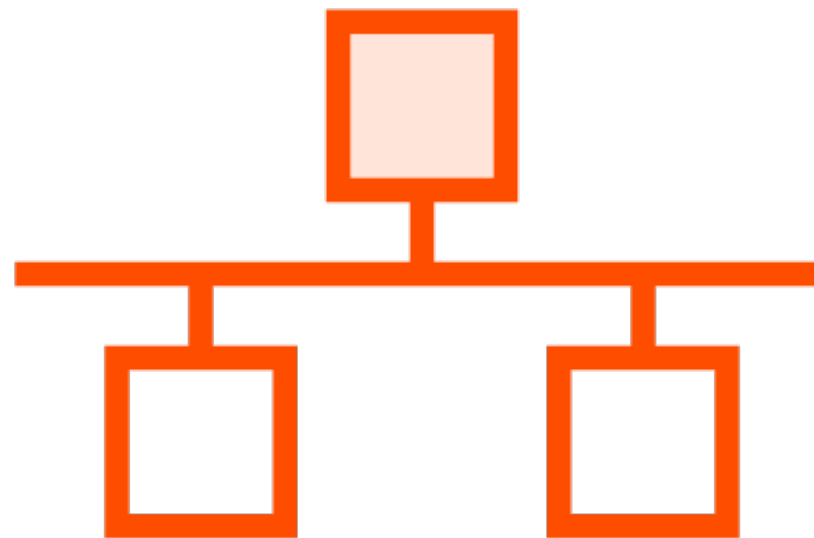
What is it?



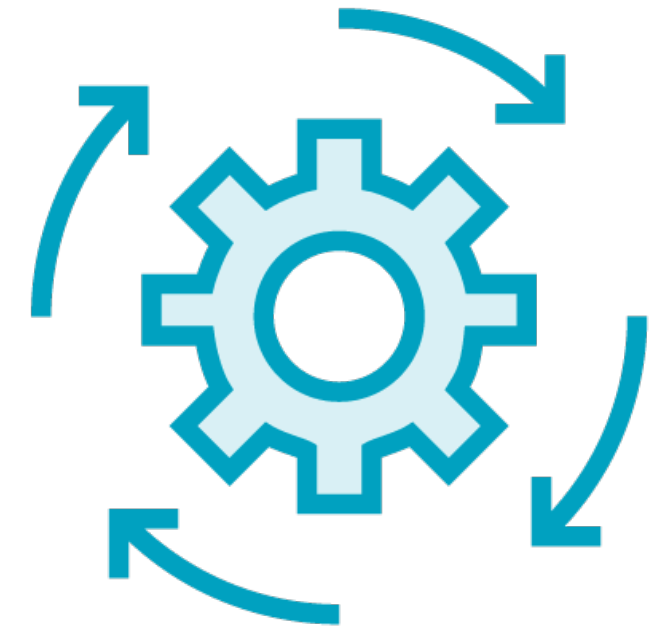
Distributed Streaming Platform



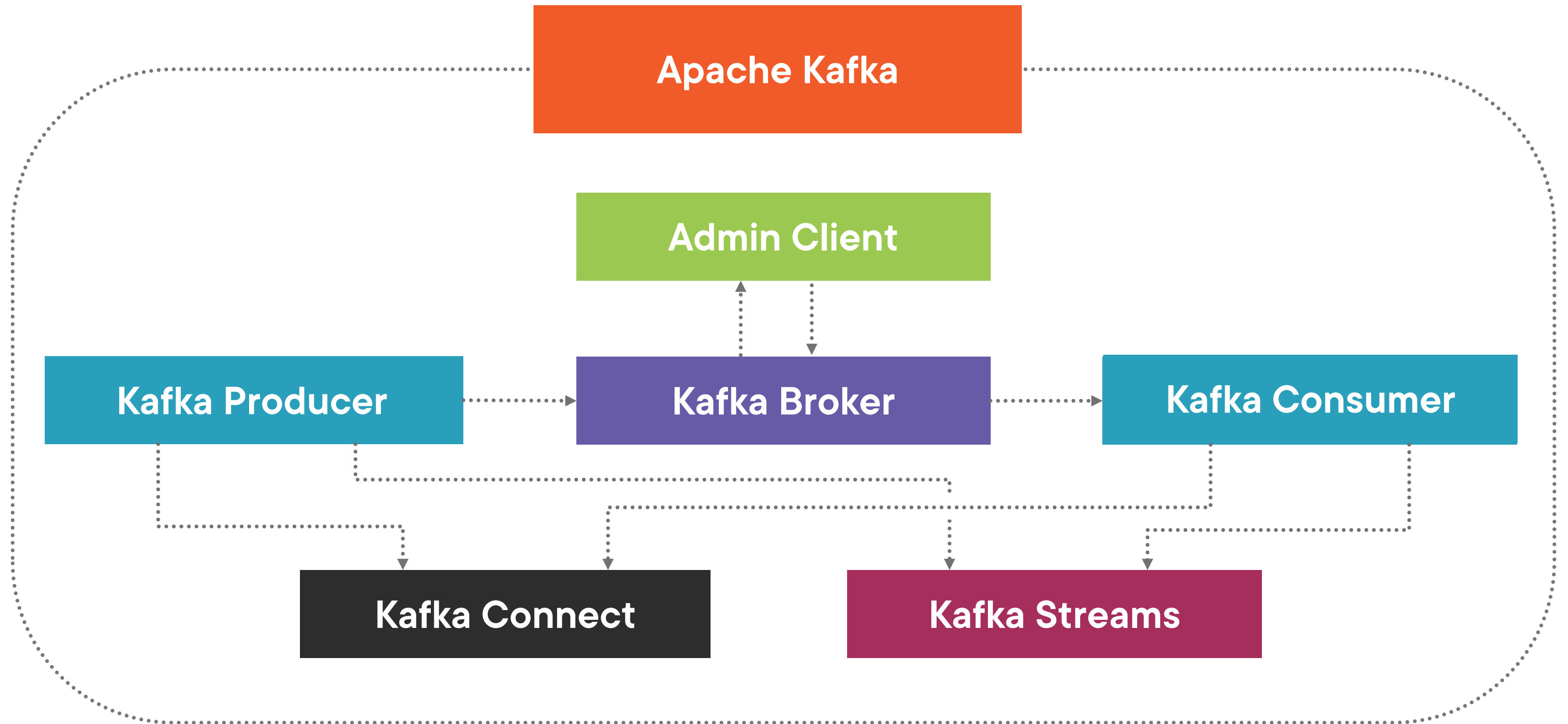
Messaging system
Publish / Subscribe
pattern

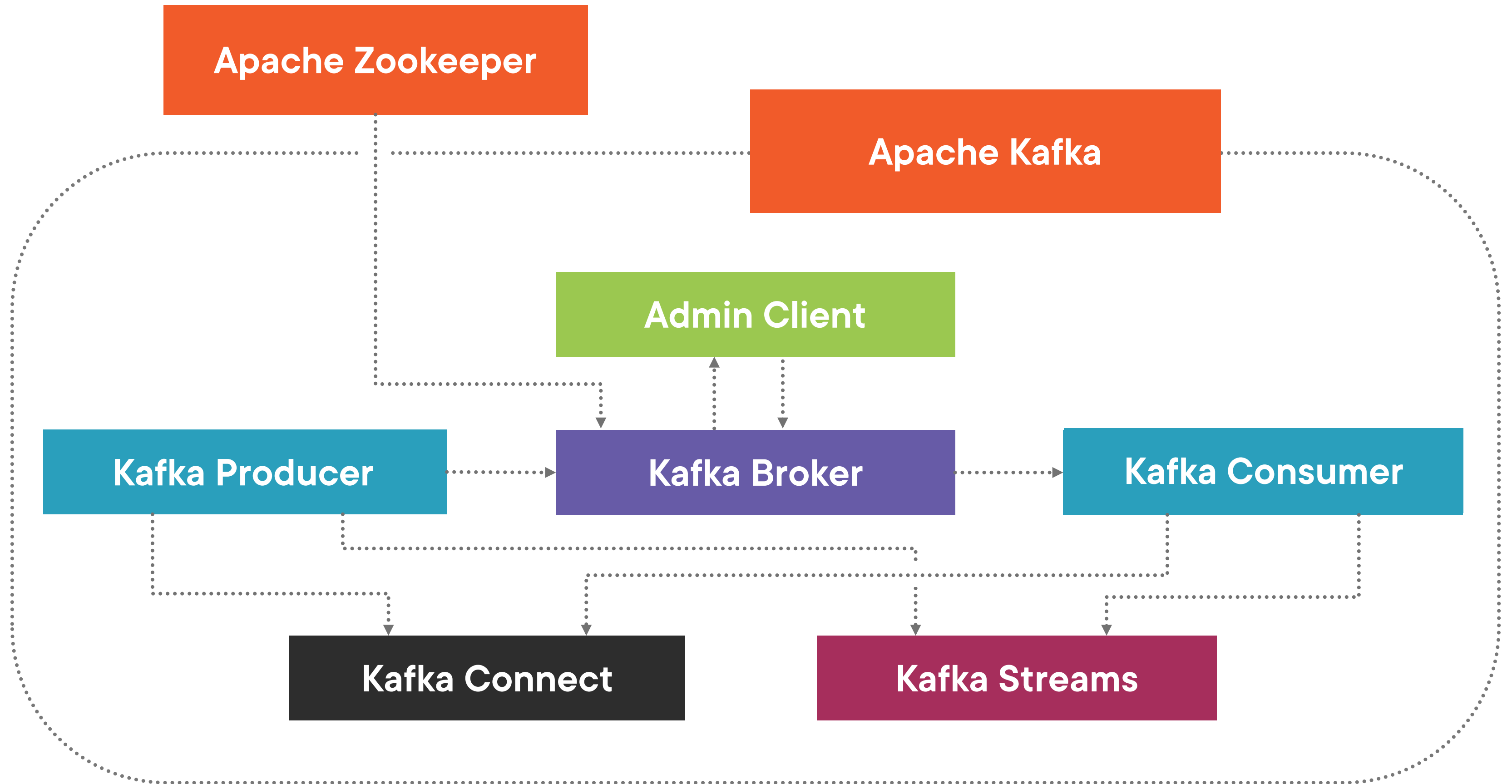


Distributed storage
Store data in a fault-
tolerant, durable way



Data processing
Process events as they
occur







What is it?

Where do I get it?

<https://kafka.apache.org/>



Apache Kafka is **NOT**
enterprise-ready!



Enterprise requirements ...

Data governance

- Creation and deletion of topics
- Enforcing data types
- Dealing with encrypted data

Integration

Testing

Availability

Security



Apache Kafka

**Open-source /
Community**

**Enterprise /
Commercial**

<https://cwiki.apache.org/confluence/display/KAFKA/Ecosystem>



Confluent Platform

Schema Registry

ksqlDB



Cloud Subscriptions



Confluent Cloud

AWS MSK

IBM Event Streams

Aiven

Heroku

CloudKarafka

Instaclustr

Etc.





What is it?

Where do I get it?

What makes it special; why use this one?



Apache Kafka Core Capabilities



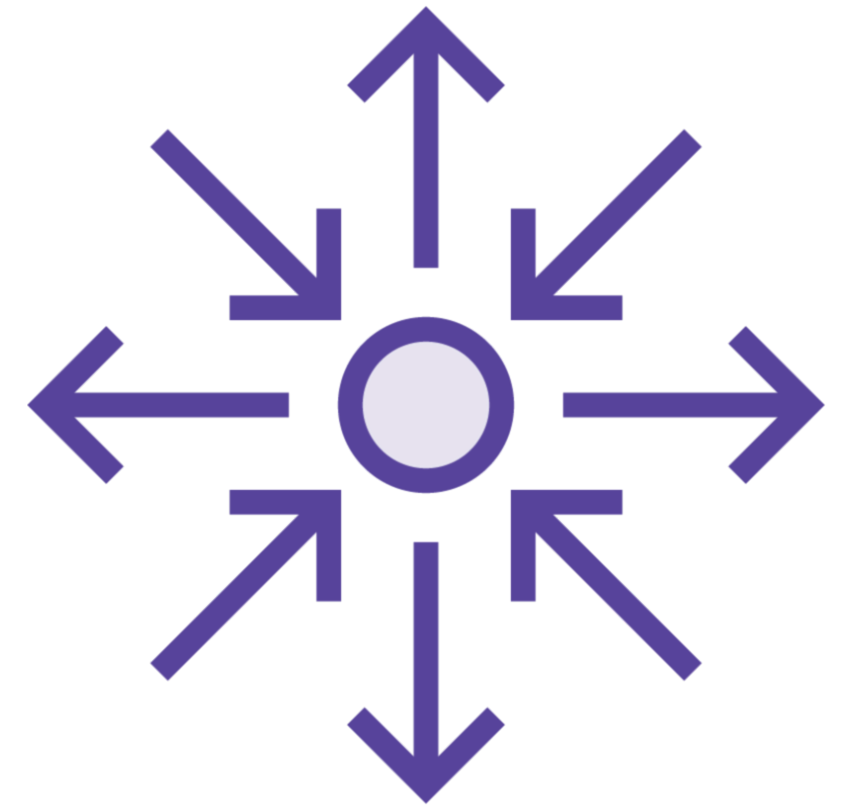
High throughput



Scalable



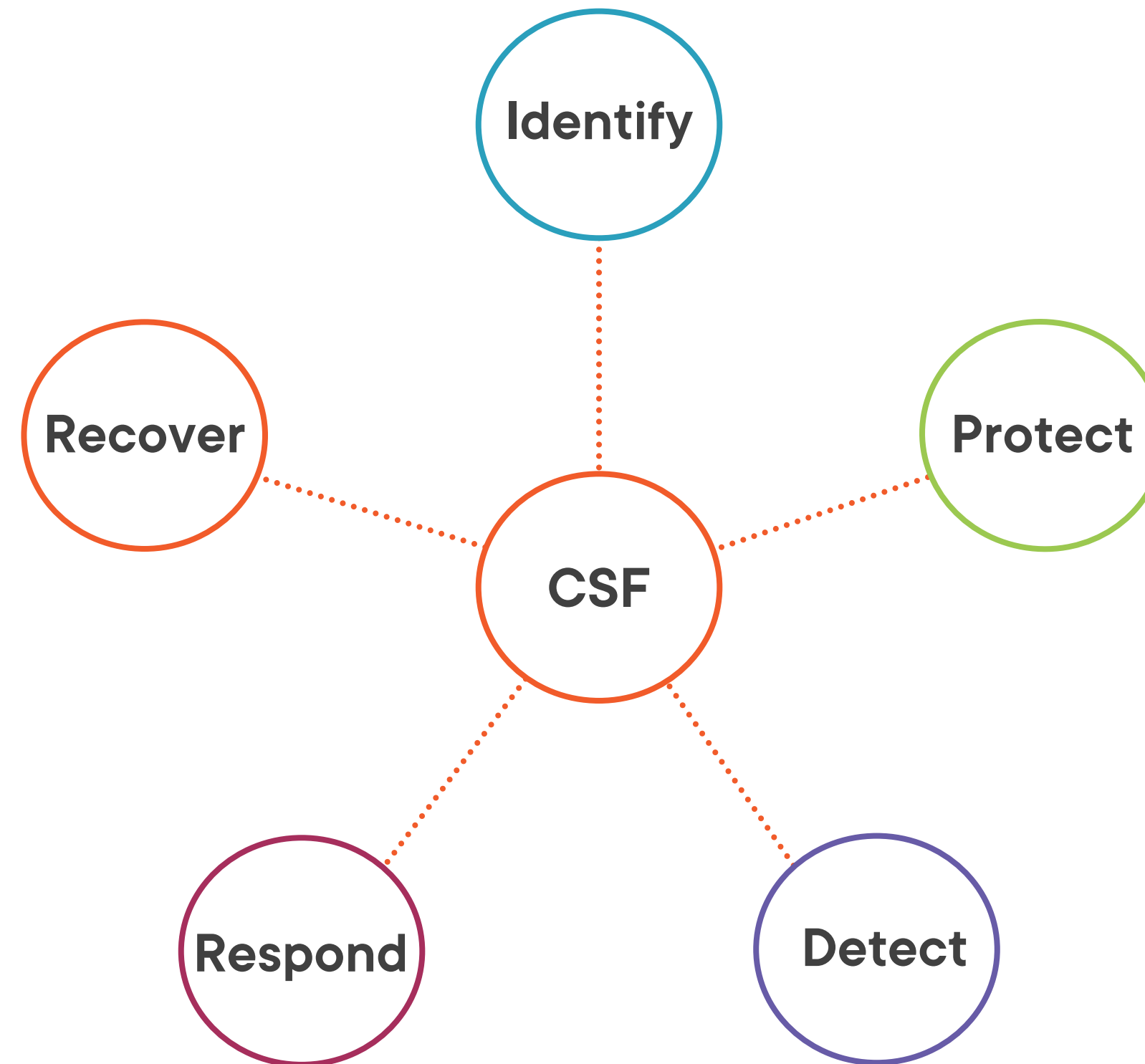
**Permanent
storage**



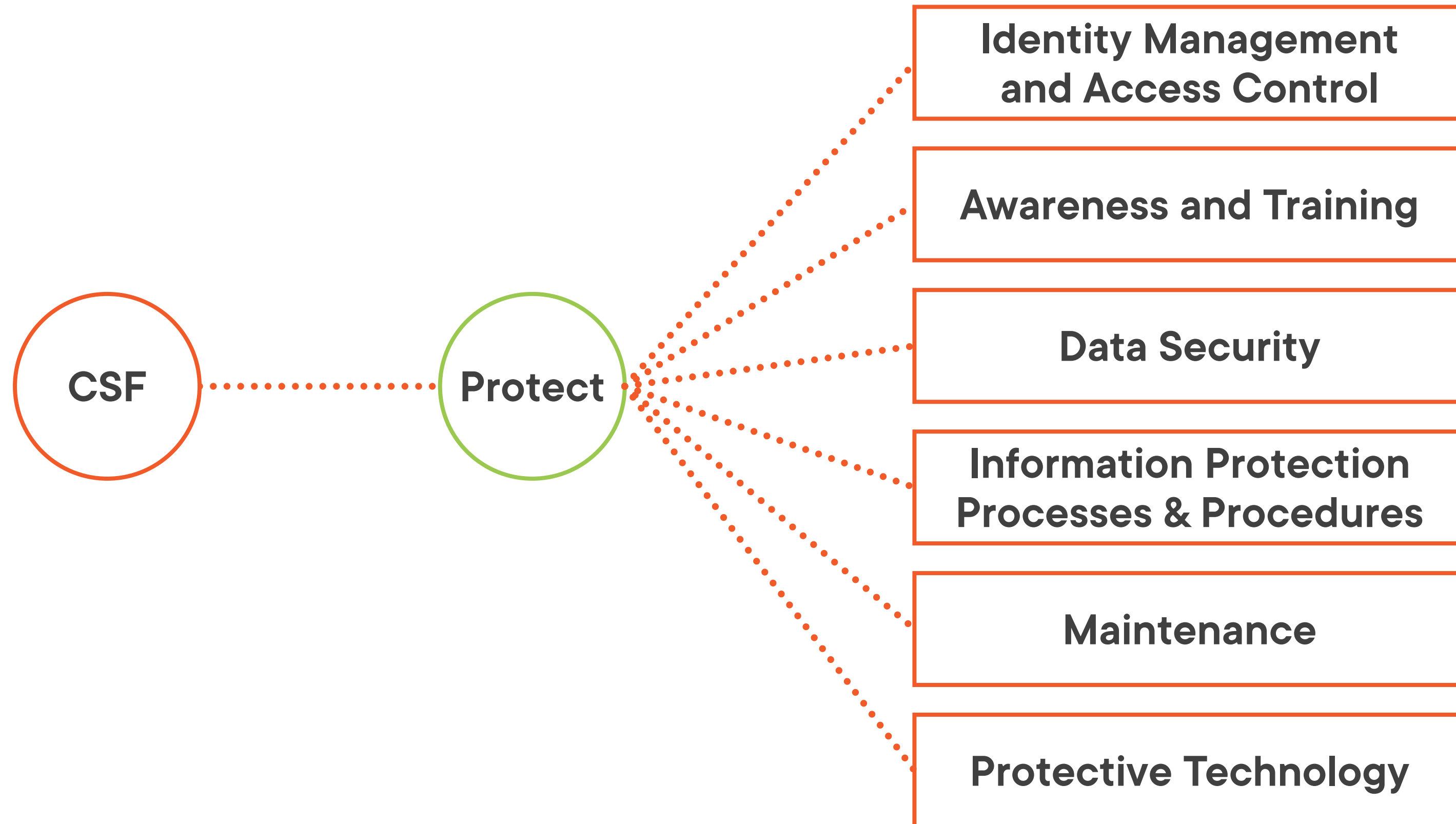
High availability



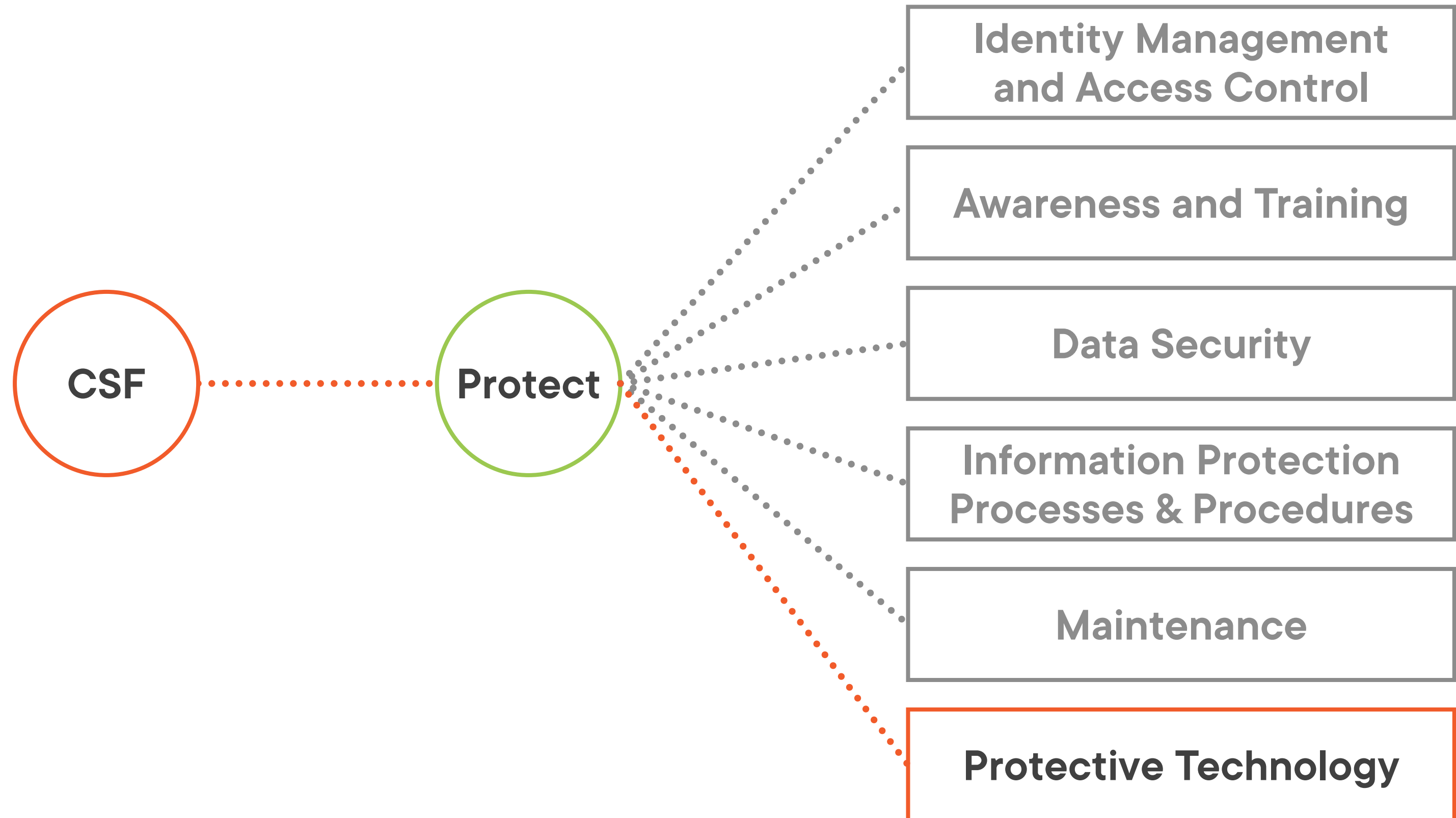
NIST Cybersecurity Framework



NIST Cybersecurity Framework



NIST Cybersecurity Framework



NIST Cybersecurity Framework



Demo: Set up Kafka Cluster

1. Set up Kafka cluster using Docker
2. Ingest logs using Kafka Connect
3. Analyze audit events using KSQL
4. Analyze network logs using Kafka Streams



Summary



What is Apache Kafka?

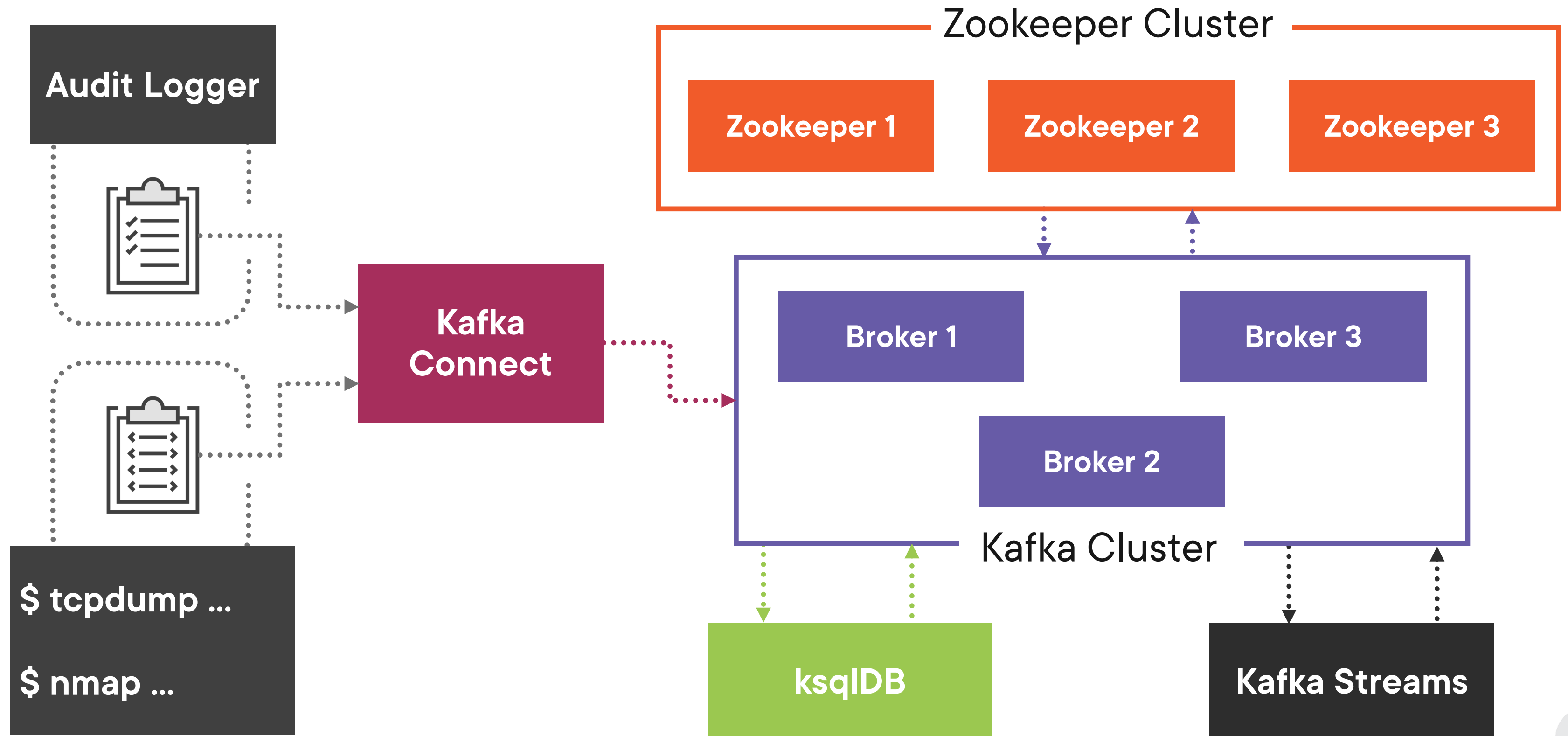
Where do I get it?

Why is it important?

Protective technology



Protective Technology with Apache Kafka



More Information

Capabilities

Data ingestion

<https://kafka.apache.org/documentation.html#connect>

Real-time analytics

Storage

Related information

Kafka documentation

Pluralsight

<https://pluralsight.com/paths/skill/handling-streaming-data-with-messaging-systems>

