## [Course Overview](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd)

### [Course Overview](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd)

[Hello and welcome to my course, Getting Started with Apache Kafka.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=4.54) [I am Ryan Plant, a software engineer, solutions architect and,](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=9.8) [in general, a distributed systems addict.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=14.2) [Chances are, you have a profile on LinkedIn and frequent the site often.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=17.44) [You're also likely to be a subscriber to Netflix and have interacted with](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=22.6) [its service on your smartphone, web browser or TV.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=27.02) [Perhaps you've gotten a ride with Uber or stayed in an Airbnb rental.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=30.94) [If any of these are the case, congratulations! Whether you knew it or not,](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=36.01) [you're a user of a system called Apache Kafka. Apache Kafka provides](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=40.42) [the messaging infrastructure of these and many more massive software as](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=46.17) [a service applications we use every day.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=50.69) [Each of these services produce hundreds of billions of](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=53.94) [messages that amount to several 100 terabytes of data being](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=57.47) [moved around and consumed per day.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=61.3) [That's truly big data.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=64.82) [Whether you're a small or large enterprise in this day and age,](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=67.14) [you're likely drowning in data and struggling to figure out](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=71.06) [how to move it where it needs to be to enable an ever](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=74.47) [increasing number of use cases.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=77.99) [It's not an easy problem,](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=80.24) [but this is why I produced this course to introduce you to Apache](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=82.01) [Kafka and provide you with an overview of how it can help you solve](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=86.54) [these problems today and for the future.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=90.69) [In this course,](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=93.85) [you will get a thorough understanding of Apache KAFTA's architecture](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=95.17) [and how it has adopted proven distributed systems design principles](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=98.74) [that enable it to scale and perform reliably.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=103.82) [We will break down this architecture into individual](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=107.54) [components and cover each in great detail.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=110.4) [We will demonstrate the components in action with](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=113.84) [common scenarios and walkthroughs.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=117.11) [How Apache Kafka Solutions can be developed in Java.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=119.17) [By the end of the course,](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=123.64) [you will have an understanding and appreciation for why Apache Kafka is](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=125.22) [taking the industry by storm, but most importantly,](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=130.63) [you will come away with confidence and knowledge to build your own](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=134.39) [next generation big data solutions with Apache Kafka.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=138.43) [I hope you will invest your valuable time and allow me to show you how to get started with Apache Kafka.](https://app.pluralsight.com/course-player?clipId=3a86bb85-c211-41a9-9e16-76a467723bdd&startTime=143.14)

## [Getting Started with Apache Kafka](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e)

### [Enterprise Challenges with Data](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e)

[Hello.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=0.24) [Welcome to the course, Getting Started with Apache Kafka.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=0.97) [My name is Ryan Plant and I am the course author.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=4.62) [With this first module, my objective will be to answer the question.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=8.14) [Why Apache Kafka? Like most things, to understand the why, we first](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=12.03) [need to learn the what. Apache Kafka is all about data and getting](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=17.23) [large amounts of it from one place to another rapidly, scalably, and](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=22.04) [reliably. Throughout this course, I'll refer to this as data movement](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=27.09) [or data logistics.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=31.89) [In computing, a common term for transferring data is messaging,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=33.97) [and that's how Apache Kafka would describe itself, as a messaging system.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=38.93) [But unlike other messaging systems,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=43.94) [Apache Kafka is tailored for high throughput use cases where vast](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=46.06) [amounts of data need to be moved in a scalable,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=50.75) [fault tolerant way.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=53.52) [So why is something like a Kafka needed?](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=55.34) [This is what a lot of enterprises look like.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=58.3) [Okay, I admit it.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=61.1) [It may be a little dramatic, but if you think about large companies,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=62.7) [your own possibly, there are hundreds of applications](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=66.54) [all needing data to operate. Now, whether it be creating logs,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=70.03) [records and databases, key value pairs,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=74.75) [binary objects or messages all of these applications are](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=77.53) [creating data at an incredible rate.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=81.1) [Oftentimes,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=83.72) [that rate can strain existing data stores and require](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=84.93) [more stores to take on the load.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=88.68) [When that happens,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=91.28) [you have issues related to getting the data where it needs to](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=92.42) [be and enabling applications to find it.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=95.65) [Furthermore, as businesses change, the variety of the data increases,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=98.74) [making the types of applications and data stores change as well. Now,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=103.78) [this obviously doesn't happen overnight,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=108.51) [but it happens,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=111.18) [and the result becomes a complex web of point‑to‑point data](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=112.38) [movements that are very hard to manage and work with.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=116.03) [In this common enterprise scenario,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=119.74) [there are a lot of tools and methods being used to make this](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=122.05) [complex distribution topology possible.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=125.17) [Most of these have been in the technology toolbox for decades,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=128.03) [and each comes with its fair share of trade offs.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=131.57) [Let's take a look at them.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=134.9) [Probably the most common is database replication and log shipping.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=136.74) [Obviously, it could be useful,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=141.17) [but this is limited to a certain kind of data movement between](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=143.05) [relational databases that support replication.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=146.54) [And that's it.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=149.6) [The way a database implements replication is very specific to the](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=150.72) [database and therefore doesn't work across vendors.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=155.05) [So in a heterogeneous database environment,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=158.44) [this becomes a limitation. As a point‑to‑point integration, there is a](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=161.01) [significant amount of coupling between the source and the target.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=165.07) [Changes to the schema have a direct impact on replication,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=168.94) [so as your requirements change, the ripple effect can introduce](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=173.14) [challenges to your replication architecture. For log shipping,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=176.81) [performance could be a challenge,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=180.95) [depending on how big the log is that you're trying to ship and the frequency.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=182.52) [But overall, these methods, while somewhat functional,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=186.69) [are cumbersome to manage and maintain and really only apply to](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=189.84) [a certain type of datastore. For integrating data between](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=193.98) [different sources and targets, ETL or extract, transform and](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=198.48) [load options are implemented, but not without its drawbacks.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=203.16) [Enterprise‑grade ETL is typically very costly and, in some cases, proprietary.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=207.02) [In recent years, viable open source options have sprung up,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=213.24) [but nonetheless,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=217.14) [ETL tools and infrastructure usually requires a bit of](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=218.08) [training and ramp up time to use productively.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=221.7) [Every ETL job that runs is a custom application written](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=224.84) [by a developer who specializes in ETL.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=229.3) [As the data environment increases in complexity,](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=232.64) [so do the jobs and as most ETL systems centralize the execution of these](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=235.66) [jobs, the performance and scalability become strained as concurrent or](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=240.18) [sequential jobs compete for the limited resources, which may require multiple ETL environments to exist, which further increases the complexity.](https://app.pluralsight.com/course-player?clipId=cc0bb949-4ae6-43bd-b8d7-9455bee4584e&startTime=245.02)

### [Messaging Limitations and Challenges](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc)

[The next two areas are where we'll spend a bit more time because,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=0.84) [as I said, Kafka is a messaging system.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=4.51) [So discussing how current messaging systems are applied to enterprise](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=7.46) [scenarios will help you understand how and why Kafka is a viable tool](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=11.16) [to consider in your modern day toolbox.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=16.2) [Messaging makes a lot of sense because it establishes a fairly simple](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=19.24) [paradigm for moving data between applications and datastores.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=23.37) [However, when it comes to a large scale implementation,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=27.94) [traditional message systems can struggle,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=31.64) [namely with scalability.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=34.61) [The means to collect and distribute data as messages](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=36.84) [relies on the role of a messaging broker,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=40.28) [which is oftentimes a bottleneck for reasons we'll cover shortly.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=43.26) [Additionally,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=47.64) [there are a lot of variables that determine the reliability](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=48.77) [and performance of a messaging system,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=51.74) [a big one being message or data packet size.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=53.87) [Larger messages can put severe strain on message brokers,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=57.61) [and this is a challenge because you may not be able to](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=61.39) [control messages coming from some systems.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=63.85) [Furthermore,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=67.04) [a messaging environment is dependent on the ability for message](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=67.98) [consumers to actually consume at a reasonable rate.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=71.87) [There is also the challenge of fault tolerance.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=75.83) [Think about it, if a consumer pops something off the](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=78.82) [queue or reads it from a topic.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=81.96) [It's probably gone.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=84.04) [So if the consumer loses the message or processes it incorrectly,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=85.62) [it is extremely difficult to get it back to reprocess.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=89.84) [Let's go into some of these issues a little further,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=94.54) [since its key to understanding how Kafka provides a better messaging system.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=97.5) [Under ideal circumstances,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=102.12) [you have applications serving as publishers of messages and a broker](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=103.88) [that is like a mailbox whose job it is to deliver or make available](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=108.65) [to messages to consuming applications,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=112.74) [which consume their messages at a reasonable rate.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=115.24) [But under higher volumes and varieties of message sizes,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=118.45) [the publishing applications can run amok and blast the broker with messages.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=121.95) [If the applications have not implemented some sort of throttling,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=127.24) [the broker can be put into a tough situation fast.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=131.17) [Now, most messaging systems are implemented on a single node or host,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=134.94) [which generally relies on a limited amount of local or quota storage.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=139.12) [Generally, this isn't a problem as messaging systems are usually very efficient,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=144.11) [provided they could turn over the messages they're receiving](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=148.55) [fast enough before the storage becomes limited.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=151.29) [This happens when you have lazy, slow or unresponsive application consumers.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=154.64) [For whatever reason, the result can be an outage of disastrous proportions.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=159.96) [The brokers disks get full, the broker croaks, becomes unresponsive,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=164.93) [and now you're publishing applications can't publish their messages.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=170.01) [And,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=174.34) [depending on the error handling, can cause a complete denial](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=174.55) [of service of the application altogether.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=177.81) [Another category of peril is with regard to application](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=181.51) [faults in the consuming applications.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=184.92) [Faults can happen for any reason, but a common reason is a bug of some sort.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=187.58) [Where this becomes a problem is when the bug incorrectly processes the message](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=192.6) [it is getting from the broker either via a queue or a topic.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=197.2) [Why is this a problem?](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=201.34) [As we mentioned,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=202.84) [the broker's job is to turn over the messages. It doesn't](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=204.01) [and can't keep them around for very long.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=208.04) [So if a consumer consumes the message, processes it](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=210.94) [incorrectly and poisons data,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=214.48) [it can't go back to retrieve the message again because it's not there.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=216.76) [Of course,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=221.44) [the consumer wouldn't have to do it if it stashed all of the](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=222.13) [messages somewhere, but that isn't always the case.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=225.61) [Even so,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=228.99) [the work to retrieve the message again and reprocess it once](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=230.23) [the bug has been smashed is a lot of work,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=233.81) [and it may be too late.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=236.81) [Technically, messaging systems are considered a form of middleware.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=238.61) [In this case,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=243.15) [I'm referring to more custom brokering solutions where you](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=244.05) [need to write complex logic to handle data movement between](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=247.87) [applications and data stores.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=251.28) [This is where your code needs to have intimate knowledge of every datastore,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=253.62) [and that knowledge will likely be specific to the datastore type and provider.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=258.33) [Furthermore,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=262.94) [you will likely be in the realm of dealing with distributed](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=263.84) [coordination logic, multiphase commits and error handling](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=266.96) [to consistently manage data.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=270.89) [Anyone who has lived this world will tell you it is extremely complex,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=273.13) [and it never ends. With every application change, new datastore,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=278.29) [new schema, you have to revisit this code,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=283.39) [which is deceiving because on a white board it sounds like a great solution,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=286.14) [and it is tempting to pursue and may even work out under simple conditions.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=290.7) [But when you have to handle multiple sources of data at](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=295.48) [different velocities and use cases,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=298.57) [you run into challenges maintaining data consistency.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=300.67) [The more distributed your system gets,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=304.34) [the harder it is to enforce strict consistency.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=306.63) [Writing your own middleware,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=309.89) [maybe cheap at first, but when considering the overall total](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=311.38) [cost of ownership of a complex code base,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=315.15) [the costs are high.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=317.85) [An alternative is to employ a vendor's middleware solution,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=319.44) [which may or may not work for all of your scenarios](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=323.52) [and in itself can become expensive.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=326.38) [Let's take a look at two leading patterns for using](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=328.93) [middleware for data movement.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=332) [The first pattern is a multi‑write scenario where your application](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=334.04) [relies on code written somewhere to handle data flows](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=337.98) [transactionally on two or more different datastores.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=341.66) [As I said before,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=345.64) [this requires substantial care and maintenance for it to work](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=347.13) [reliably without data consistency issues.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=350.5) [For example, if a target database is not available for whatever reason,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=353.54) [and the transaction cannot commit, what should happen? There are many](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=358.45) [different approaches to this, but if not careful,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=362.33) [it could lead to data inconsistency where the second transaction](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=365.19) [commits without the first. Performance and scalability](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=369.33) [challenges come from this pattern as well,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=373.03) [because transactional consistency requires all participants to commit that can](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=375.49) [cause holds on the part of the application and database.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=381.09) [Furthermore,](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=385.14) [if one were to scale out to more resources to share the load it would](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=386.11) [mean that more and more would have to be party to the coordination](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=390.78) [logic. This can get out of hand fast. An alternative could be to](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=394.81) [leverage a messaging broker in the middle to coordinate application data](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=400.15) [movements to stores and vice versa.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=404.57) [This would be in line with the messaging scenario we discussed](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=406.84) [earlier and therefore subject to the same issues of slow](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=409.75) [consumers or unavailable data stores.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=413.49) [This pattern is also difficult to scale out as more consumers intending](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=416.49) [to share the load, may compete for access to messages and may not coordinate consistently across their peers.](https://app.pluralsight.com/course-player?clipId=c688b5c9-637d-4f17-95d3-655eb97eb4bc&startTime=421.15)

### [LinkedIn's Search for a Better Solution](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16)

[This is the typical enterprise challenge when it comes to handling growing](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=0.53) [data sets, moving faster and faster through systems.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=4.84) [Surely there has to be a better way to move data cleanly without a](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=8.69) [complex web of different integration topologies, reliably as to reduce](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=13.08) [the impact of any one component slowness or availability on the system](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=18.57) [quickly as data movement and access is only getting faster for](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=24.04) [real‑time use cases and finally,](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=28.53) [autonomously reducing the coupling between components so we can improve or](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=30.88) [change parts of the system without a cascading effect.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=36.56) [Is there a better way?](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=40.34) [Well, it just so happens that in 2010 LinkedIn asked that same question.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=41.85) [Now, most of you,](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=48.24) [if not all of you, know what LinkedIn is. It's a](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=49.23) [massive social network for professionals.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=52.85) [Over the years since its founding in 2003 it has grown exponentially.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=55.68) [They are the epitome of big, fast and varied data challenges.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=61.05) [Here are but a few of the most recent stats to give you a sense of the](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=65.64) [scale that they work with on a day‑to‑day basis.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=70.39) [LinkedIn's site offers many different features and capabilities,](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=73.8) [all provided by a portfolio of data creating and consuming applications.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=78.07) [Their ability to direct,](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=84.5) [manage and utilize the data has been a direct enabler of their success.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=86.15) [But this data handling ability wasn't always there.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=91.45) [In fact, up until 2010 things were quite ugly and getting uglier.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=95.15) [Remember this picture?](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=101.49) [Well, it more or less represented, LinkedIn before they invented Kafka.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=102.84) [Under the pressure of hypergrowth,](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=107.84) [LinkedIn needed to find a better way to deal with data.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=110.29) [I suspect many of you actively use LinkedIn and therefore](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=114.04) [will know how much functionality is available on their](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=117.71) [website and through their mobile apps.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=120.79) [All of this functionality is in self‑contained applications that](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=123.54) [autonomously produce and consume data from the data infrastructure.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=127.84) [Over the years, as more applications were written,](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=132.31) [more users used the site. A lot of technology was used](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=135.55) [to get data where it needed to go.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=139.16) [LinkedIn had several relational database types multiple node SQL stores,](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=141.22) [queueing systems, log processors, you name it.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=146.44) [Generally,](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=150.44) [all that was at their disposal to handle their growing data problems](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=151.2) [were the tools and methods that we discussed earlier,](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=154.7) [which LinkedIn found woefully inadequate for solving their data problems.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=157.3) [Given the growth,](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=162.03) [you could imagine how important it was for LinkedIn to come up with a better](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=163.14) [way to make all of this data available without slowing down,](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=167.19) [crashing or further limiting the system.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=171.04) [And, of course, this is where a Kafka comes in,](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=174.08) [where it started as an internal project in 2009.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=177.34) [Incidentally, you may be wondering why LinkedIn named their solution Kafka.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=181.64) [It refers to the German language writer, Franz Kafka,](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=187.04) [whose work was so freakish and surreal it inspired an adjective based on](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=190.54) [his name. In the case of LinkedIn their data infrastructure and the](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=195.17) [ability to work with it had become so nightmarish and scary that they](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=199.65) [named their solution after the author, whose name would best describe the solution they were hoping to escape from.](https://app.pluralsight.com/course-player?clipId=616a09b2-7b4f-4eec-9258-1070000bdf16&startTime=203.86)

### [Apache Kafka as a Viable Solution](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a)

[To guide their design, LinkedIn set forth some goals that a](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=0.04) [solution must meet in order to be a viable tool.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=3.76) [First and foremost,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=7.42) [it needed to be able to handle large volumes of](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=8.73) [data in the terabytes and beyond.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=11.73) [They knew that in order for this to be accomplished,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=14.48) [it would need to be designed to scale out by adding machines to seamlessly](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=17.44) [share the load. They couldn't afford a lossy system.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=21.88) [Data had to be reliably managed,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=25.52) [transmitted and made durable in the case of failure. It was important](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=27.72) [for LinkedIn that all application producers and consumers be loosely](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=32.43) [coupled but engage in common data exchanges.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=37.06) [It would be unacceptable,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=41.14) [for one application's runtime conditions to affect another's.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=42.58) [To enable this loosely coupled paradigm between producers and consumers,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=47.04) [they wanted to embody common and simple messaging semantics](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=51.74) [of publish‑subscribe. Independent data producing](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=55.57) [applications would send data on a topic,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=59.09) [and any interested consuming application could listen](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=62.16) [in and receive data on that topic,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=65.48) [which it could process and in turn, produce on a different topic for](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=68.14) [others to consume. At a high‑level this is what the data architecture](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=72.34) [looked like once Kafka was in place. Now,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=76.94) [the real details of how this works will be covered in the next few modules.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=79.83) [But you can see what role Kafka plays with respect to](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=84.01) [different data producers and consumers.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=87.18) [As a central broker of data,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=89.8) [Kafka enables disparate applications and data stores to](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=91.78) [engage with one another in a loosely coupled fashion by](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=95.52) [conveying data through messaging topics,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=98.9) [which Kafka manages at scale and reliably,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=101.26) [regardless of the system, the vendor,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=104.73) [the language or runtime all can integrate into this data fabric](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=107.27) [provided by none other than Apache Kafka. Because of Kafka,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=111.8) [LinkedIn was able to successfully weather their storm of data and](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=116.2) [establish a foundation upon which to build their next generation of](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=120.42) [applications and services.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=124.36) [It has been openly admitted by LinkedIn engineers that without](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=126.37) [Kafka, LinkedIn would not have been able to sustain their growth](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=130.03) [and achieve the valuation they have today.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=133.84) [As we discussed,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=136.94) [Kafka's development started in 2009, and its first](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=138.19) [deployment was in 2010. Within the next year,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=142.29) [LinkedIn hardened Kafka to a point that they felt it could be released as an](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=146.04) [open‑source project under the Apache Software Foundation.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=150.51) [This they did in 2011. Very soon after its submission to the Apache incubator,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=154.6) [it achieved top‑level status and has become one of the most](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=160.58) [adopted tools in the Apache ecosystem.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=164.36) [Today,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=167.64) [Kafka is responsible for driving 1.1 trillion](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=168.43) [messages per day at LinkedIn alone.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=172.93) [But there are many more big name companies that have adopted Kafka to](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=175.35) [solve the common problems we discussed at the onset.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=179.06) [Here are just a few of the most noteworthy names.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=182.7) [All in all, Since 2015 Apache Kafka's adoption rate has grown 700%.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=186.05) [As the software development community contributes more](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=193.26) [and more capabilities to its code base,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=196.47) [many of the more recent innovations that Kafka has produced will be touched](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=199.04) [on in the last module, and likely in additional courses.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=203.25) [For now,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=207.44) [let's get more into the guts of Apache Kafka and find out just](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=208.17) [how it can pull off such amazing data movement feats, not only](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=212.59) [for these enterprises listed here, but for your own.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=216.17) [Before we move on to the next module,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=220.54) [let's quickly summarize what we've covered in this module,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=222.95) [as it self‑describes on its website, Kafka is a distributed](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=226.3) [messaging system designed for high throughput.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=230.41) [It was intended to address many of the shortcomings that](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=233.52) [traditional data handling tools and methodologies have,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=236.48) [particularly when data is growing and it needs to move](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=240.24) [faster through more and more diverse systems.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=243.37) [Which was the cause at LinkedIn in 2009 when it decided to](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=246.43) [invest R and D effort to solve its data problems.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=251.26) [Luckily,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=254.71) [LinkedIn did a great job and was kind enough to open source it in 2012 and](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=255.38) [make it available under the generous Apache license.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=260.52) [Now,](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=263.64) [many enterprises and internet scale companies can take advantage of Kafka and](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=264.34) [many are, using it as LinkedIn does for addressing their complex data](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=269.34) [infrastructure and preparing themselves to take advantage of the future opportunities for analyzing large amounts of data.](https://app.pluralsight.com/course-player?clipId=3b7b4b83-2760-46d2-826a-77e2172f8f9a&startTime=273.64)

## [Getting to Know Apache Kafka's Architecture](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234)

### [Apache Kafka as a Messaging System](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234)

[In the last module,](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=0.84) [I raved about Apache Kafka and why it is a breakthrough](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=2.18) [tool for managing data movement at scale.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=5.92) [To support my claims,](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=9.33) [I cited many large scale companies that use it and even gave some](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=10.77) [impressive statistics that back up Kafka's ability to move data.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=14.9) [That's all nice.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=19.03) [But now we're going to start getting deeper into Apache Kafka and learn](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=19.96) [just how exactly it is able to deliver on such abilities.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=23.99) [This module will start by examining the overall architecture of Apache Kafka.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=27.91) [As I highlighted in the first module, Apache Kafka is truly a messaging system.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=33.64) [More specifically,](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=39.54) [it is a publish subscribe messaging system in a pub subsystem, there are](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=40.93) [publishers of messages and subscribers of messages.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=45.7) [A publisher creates some data and sends it to a specific](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=48.84) [location where an interested and authorized subscriber can](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=52.21) [retrieve the message and process it.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=56.03) [In Kafka,](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=58.25) [we call these traditional publishers something slightly different ‑](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=59.14) [producers, and the subscribers we call ‑ consumers.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=62.74) [As we'll see in the upcoming modules,](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=66.33) [producers and consumers are simply applications that you write or](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=68.68) [use to implement the producing and consuming APIs.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=72.67) [Now the producer sends its messages to what I said was a specific](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=77.24) [location. In Kafka, this location is referred to as a topic,](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=81.26) [which is really a collection or grouping of messages.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=85.47) [Topics have a specific name that can be defined up front or on demand as](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=88.87) [long as producers know the topic name and have permission to send to it,](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=93.72) [messages can be sent to that specific location.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=97.81) [The same goes for consumers.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=100.74) [Consumers retrieve messages based on the topic it is interested in.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=102.39) [This should be very familiar for those with experience using pub sub messaging.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=106.7) [As these concepts were intentionally carried forward into Kafka due to the](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=111.43) [simplicity of the paradigm. The messages and their topics need to be kept](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=116.01) [somewhere, after all, they are physical containers of data.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=121.39) [The place where Kafka keeps and maintains topics is called the broker,](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=125.67) [as it is in other messaging systems.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=130.4) [We'll cover topics in greater detail later,](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=132.94) [but for now let's look closer at the Kafka broker.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=135.69) [The Kafka broker is a software process also referred to as an](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=141.35) [executable or daemon service that runs on a machine, a](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=145.86) [physical machine or a virtual machine.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=150.36) [A synonym for a broker is also a server,](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=152.97) [but I like to avoid using the term server,](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=156.37) [since it has a tendency to be overloaded.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=158.79) [The broker has access to resources on the machine,](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=161.71) [such as the file system, which it uses to store messages,](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=165.36) [which it categorizes as topics.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=169.04) [Like any executable, you can run more than one on a machine,](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=171.9) [but each must have unique settings so that they don't conflict with one another.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=176.39) [We'll come back to this shortly.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=181.54) [It is in the Kafka broker, where the differences between other messaging systems become apparent.](https://app.pluralsight.com/course-player?clipId=a13351b2-ca91-435b-be38-bad99bc45234&startTime=183.24)

### [The Apache Kafka Cluster](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09)

[How the Kafka broker handles messages in their topics is what](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=0.14) [gives Kafka its high throughput capabilities.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=4.12) [Achieving high throughput is largely a function of how well](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=7.22) [a system can distribute its load and efficiently process it](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=10.82) [on multiple nodes in parallel.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=14.62) [This is a hallmark of scalable design.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=16.84) [With Apache Kafka,](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=19.74) [you can see scale out the number of brokers as much as needed to](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=21.24) [achieve the levels of throughput required.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=24.75) [And all of this without affecting existing producer and consuming applications.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=27.14) [In fact,](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=32.58) [LinkedIn as of April of 2016 has publicly stated that they have upwards of](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=33.52) [1400 brokers processing over two petabytes per week.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=39.76) [You simply can't find another messaging system out there with that capability.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=44.77) [I am not satisfied yet, though, and I hope you're not either.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=49.87) [Let's continue to dive deeper and understand what about Kafka's](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=53.46) [architecture enables it to scale so well and furthermore,](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=57.39) [how it can achieve such high levels of reliability.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=61.34) [It's time to introduce another key concept within the Apache](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=65.54) [Kafka architecture and that is the cluster.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=69.08) [A Kafka cluster is a grouping of multiple Kafka brokers.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=72.18) [As I said,](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=76.85) [you can have a single broker or multiple brokers on a single](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=77.73) [machine or brokers on different machines.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=81.5) [For example, if you had only a single broker on a machine and only one machine,](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=84.74) [it would be said that you have a cluster of one. If you have](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=90.69) [two brokers running on the same machine,](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=94.51) [it would be considered a cluster of two. This would be the same](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=97.06) [if each broker was running on its own machine,](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=101.24) [you still would have a cluster of two.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=104.08) [The important thing to remember here is that a Kafka](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=106.74) [cluster is just a grouping of brokers.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=109.69) [It doesn't matter if they're running on their own machines or not.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=112.49) [What matters is how independent brokers are grouped to form a cluster.](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=115.48) [The grouping mechanism that determines a cluster's membership of brokers is](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=120.35) [an important part of Kafka's architecture, and what really enables its](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=124.61) [ability to scale to thousands upon thousands of brokers and be distributed](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=128.74) [in a fault tolerant way. For the sake of putting down a placeholder, this is where Apache ZooKeeper comes in](https://app.pluralsight.com/course-player?clipId=1ec399bd-2033-4d76-bd54-2a4a4252cf09&startTime=133.76)

### [Principles of Distributed Systems](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718)

[At this point,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=null) [it is important to discuss a little theory about distributed systems.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=1.13) [But along the way I will do my best to associate the principles to Apache Kafka.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=5.35) [So you'll understand how brokers organize into clusters and how Kafka](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=10.4) [clusters achieve amazing feats of scalability and reliability.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=14.42) [First,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=19.02) [I'm going to give you my general definition of a distributed](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=19.64) [system for the purposes of describing how Kafka,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=22.77) [as a distributed system,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=25.82) [works without spending too much time discussing philosophy.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=27.2) [Let's just say a system is a collection of resources that have](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=30.88) [instructions to achieve a specific goal or function.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=35.01) [A distributed system is one that consists of multiple independent resources,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=38.38) [also known as workers or nodes, sometimes even called worker nodes.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=43.74) [Obviously, the reason there are multiple nodes is to spread the work around,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=49.04) [presumably to get more done than what could otherwise be achieved with less.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=53.95) [But in order to do that,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=59.09) [there needs to be coordination amongst all of the available](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=60.68) [working nodes to ensure consistency and optimal progress](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=64.28) [towards the overall task or goal at hand.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=68.64) [Without coordination, it would be difficult,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=71.54) [if not outright chaotic,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=74.14) [to divide up the work appropriately and track progress as to](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=76.06) [ensure the most optimal use of resources,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=80.12) [for example,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=82.86) [proper coordination can avoid duplication of effort or individual worker nodes,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=83.68) [undermining each other's work without knowing it.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=88.78) [Of course,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=91.74) [coordination isn't possible without clear and thorough](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=92.58) [communication between all components within the system.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=95.9) [In Kafka, these worker nodes are the Kafka brokers.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=99.31) [Within a distributed system, there are different roles and responsibilities.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=104.24) [And like any organization of able workers,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=108.55) [there is generally a hierarchy that starts with a controller or supervisor.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=111.42) [A controller is really just a worker node like any other.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=115.84) [It just happened to be elected from amongst its peers to officiate](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=119.28) [in the administrative capacity of a controller.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=123.44) [In fact,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=126.24) [the worker node selected is the controller is commonly](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=127.12) [the one that's been around the longest.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=130.36) [But other factors could be considered as well.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=132.62) [I'll have to leave it at that,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=135.07) [as this specific topic is beyond the scope of this course.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=136.61) [Once selected, the controller has some critical responsibilities.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=140.39) [First, to maintain an inventory of what workers are available to take on work.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=144.52) [Second,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=150.27) [to maintain a list of work items that has been](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=150.91) [committed to and assigned to workers,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=153.65) [and third to maintain active status of the staff and](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=156.4) [their progress on assigned tasks.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=160.13) [Once a controller is established and the workers are assigned and available,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=162.84) [you could say a team is formed and work can now be distributed and](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=167.16) [executed. In Kafka, this team formation is the cluster,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=170.76) [and its members are brokers that have assigned themselves to a](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=175.24) [designated controller within the cluster.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=178.58) [When a task comes in as an example from a Kafka producer,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=181.74) [the controller has to make a decision which workers should take it.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=185.63) [There are a lot of factors at play here,](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=189.14) [many of which are beyond the scope of this module, but let's cover the most important ones.](https://app.pluralsight.com/course-player?clipId=6189f511-92fa-4dbd-887e-402ac995f718&startTime=191.18)

### [Reliable Work Distribution](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5)

[First, the controller needs to know who is available and in good health.](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=0.04) [And very importantly,](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=4.13) [the controller needs to know what risk policy should](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=5.49) [govern its assignment decisions.](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=8.44) [A great example for risk policy is the redundancy level,](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=10.34) [the thing that determines what level of replication to](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=14.43) [employ in case an assigned worker fails.](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=17.37) [In a distributed system that offers redundancy options,](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=21.66) [it has to ensure that if work is assigned to a worker](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=24.72) [and that worker becomes unavailable,](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=28.28) [the work assigned or the work already done is not lost.](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=29.91) [That means each task given to a worker must also be given to at least one of](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=33.92) [the worker's peers in the event of an unexpected catastrophe.](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=39.23) [But amongst a group of worker nodes, which one will get assigned?](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=43.54) [This is how it works. For an assignment, if the](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=47.35) [controller determines redundancy is required,](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=50.03) [it will promote a worker into a leader which will take](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=53.1) [direct ownership of the task assigned.](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=56.37) [It will be the leader's job to recruit two of its](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=58.91) [peers to take part in the replication.](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=62.02) [In Kafka, the risk policy to protect against loss is](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=64.54) [known as its replication factor,](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=68.23) [and we will cover this in more detail within the next module.](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=70.56) [Once peers have committed to the leader,](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=75.84) [a quorum is formed,](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=78.31) [and these committed peers now take on a new role in relation to a leader,](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=79.84) [a follower.](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=84.11) [If, for whatever reason a leader cannot get a quorum,](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=85.21) [the controller will reassign tasks to leaders that can. Up until](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=88.24) [this point in covering the principles of distributed systems and](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=93.28) [how Apache Kafka applies them,](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=96.8) [we have been using the term work somewhat generally. In Apache Kafka, the work](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=98.93) [that the cluster of brokers performs is receiving messages,](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=103.99) [categorizing them into topics and reliably](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=107.54) [persisting them for eventual retrieval.](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=110.58) [As we have already discussed,](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=113.41) [the components that drive this work are producing applications called](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=115.01) [producers that send the messages to the cluster and consuming applications,](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=118.62) [called consumers that retrieve and process messages.](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=123.16) [Comparatively speaking,](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=126.75) [the effort required to handle messages from the producers is](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=128.3) [substantially less than the effort required by consumers.](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=131.97) [We will talk about this more in modules four and five.](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=135.44) [I just wanted to bring this up to highlight that both add load to the cluster, but in very different ways](https://app.pluralsight.com/course-player?clipId=d4c578d1-80cb-43a7-bfc8-14e4fb5d4cb5&startTime=138.57)

### [Distributed Consensus with Apache Zookeeper](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003)

[Virtually every component within a distributed system has to keep](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=0.34) [some form of communication going between themselves.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=4.13) [In distributed computing terms,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=7.45) [this communication is referred to as a consensus or gossip protocol,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=9.61) [and without it distributed system simply cannot operate.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=14.57) [Besides the obvious data payloads being transferred as messages](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=18.34) [There are other types of network communications happening that](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=22.54) [keep the cluster operating normally.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=25.64) [For example,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=27.8) [events related to brokers becoming available and requesting](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=28.77) [cluster membership or broker name lookups.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=32.32) [Retrieving Bootstrap configuration settings and being notified of](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=35.34) [new settings consistently and in a timely fashion.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=39.22) [Events related to controller and leader election and health status updates like](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=42.3) [heartbeat events It is time to talk about Apache ZooKeeper.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=46.85) [I have to give a big disclaimer,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=52.54) [though Apache ZooKeeper could use its own course,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=54.36) [so we can't spend a lot of time on it here.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=58.04) [It is used in a variety of distributed systems.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=60.83) [For all the reasons we have been discussing in the context of Apache Kafka,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=63.9) [specifically,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=68.44) [ZooKeeper serves as a centralized service for metadata about vast](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=69.54) [clusters of distributed nodes needing Bootstrap and runtime configuration](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=74.21) [information, health and synchronization status,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=78.55) [and cluster and quorum group membership,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=81.47) [including the roles of elected nodes.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=83.91) [Some rather large distributed systems depend on](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=86.51) [ZooKeeper, like Apache Hadoop HBase, Mesos,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=89.52) [Solr,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=93.47) [as well as Redis and Neo4j database is they all use it to enable](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=94.38) [their scale out fault tolerant capabilities.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=98.93) [ZooKeeper itself is a distributed system,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=101.84) [and for it to run reliably has to have multiple nodes,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=104.76) [which form what is called a ZooKeeper ensemble.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=107.94) [An ensemble is like saying a cluster for ZooKeeper.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=111.03) [In the case of Kafka, because of the type of work a ZooKeeper ensemble performs,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=114.11) [it is generally not needed to have more than one ensemble](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=119.26) [to power one or many Kafka clusters.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=122.4) [Let's bring it all together in a logical view.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=126.94) [At the heart of Apache Kafka, you have a cluster which,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=129.74) [as we discussed,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=132.94) [consists of possibly hundreds of independent brokers](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=133.97) [closely associated with the Kafka cluster.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=137.22) [You have a ZooKeeper environment which provides the brokers within a cluster,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=139.91) [the metadata it needs to operate at scale and reliably. As this](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=144.24) [metadata is constantly changing, connectivity and chatter between the](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=148.39) [cluster members and ZooKeeper is required.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=152.64) [Of course,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=155.79) [the cluster doesn't do much unless if you put it to work and that's](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=156.6) [where Kafka producers and consumer applications come in.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=160.24) [Each of these components can scale out to take on more demand and](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=164.17) [increase levels of reliability and availability.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=168.51) [I have given some incredible statistics about LinkedIn's Kafka environment,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=172.24) [but Netflix is even more impressive.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=176.28) [Over 4000 brokers across 36 clusters,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=179.06) [processing over 700 billion messages per day.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=183.23) [I hope with this foundation established, you will start to understand how.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=186.86) [But we're not finished.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=191.54) [We still have to dive deeper into the messaging internals,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=192.75) [which happen inside the broker.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=196.54) [And with that, learn more about Kafka topics and partitions.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=198.72) [In this module, we covered a lot of ground.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=204.4) [We started by discussing how Kafka is a publish subscribe messaging](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=207.07) [system with common concepts as other messaging systems.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=211.6) [But we started to differentiate Kafka with slightly different terminology,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=215.3) [especially around producers and consumers that publish](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=219.74) [and receive messages respectively,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=223.34) [and independent Kafka brokers and how they are grouped to form a cluster.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=226.03) [We also introduced some fundamental characteristics of distributed systems,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=231.26) [and we started to describe the worker node roles of controllers,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=236.14) [leaders, peers and followers.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=240.6) [Also,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=242.87) [we discussed how distributed systems enforce and provide](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=243.62) [flexibility and reliability through replication policies.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=247.51) [We rounded it off by highlighting the importance of consensus or](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=251.54) [gossip‑oriented communication within a distributed system and the](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=255.82) [critical role that ZooKeeper plays in that,](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=260.23) [and in enabling Kafka and other distributed systems to function scalably and reliably.](https://app.pluralsight.com/course-player?clipId=2c09bbfc-93c6-4c36-a88b-6c87994df003&startTime=262.77)

## [Understanding Topics, Partitions, and Brokers](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084)

### [Introduction and Apache Kafka Setup Demo](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084)

[At this point,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=1.05) [I am hoping you now have a fundamental understanding of](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=1.89) [the architecture of Apache Kafka,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=4.68) [at least from the standpoint of how Kafka organizes its brokers into clusters](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=7.44) [and distributes work redundantly. We will build upon this foundation and now](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=12.07) [discuss the central concepts of Kafka, message topics,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=17.28) [partitions, and how brokers manage them in line with the](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=21.62) [distributed systems principles we just covered.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=25.26) [Before we go much further,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=29.54) [I want to show you how to get Apache Kafka installed on a](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=30.97) [development machine like a virtual machine running Linux.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=34.62) [We will not go into a lot of configuration details at this point.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=37.98) [Thankfully, we don't need to.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=41.87) [Kafka is ready to go with a basic installation,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=43.73) [and that's what we'll start with because it's the easiest.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=46.5) [We'll download the binary package from the Kafka Apache site,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=49.74) [extract the archive into a directory, and finally,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=53.62) [we'll take a look inside that directory. In this demo](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=56.95) [and throughout the rest of the course,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=60.61) [I'm going to assume a set of prerequisites. First, that you're somewhat](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=62.26) [familiar with the Linux operating system as that is the recommended](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=66.76) [operating system for running Apache Kafka.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=70.55) [We'll be using the terminal and a Bash shell mostly for this. The Java 8](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=73.44) [Development Kit needs to be installed and configured.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=78.53) [It doesn't matter which JDK you use.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=81.11) [It could be the OpenJDK or the Oracle JDK.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=83.9) [It's just important that you have one installed and configured.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=87.5) [Finally,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=91.14) [you have to have Scala installed. Since Apache](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=91.92) [Kafka was mostly written in Scala,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=95.44) [you'll need it's runtime. For this course, we're using the latest version,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=97.5) [which is 2.11.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=101.56) [Let's get started.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=103.34) [Let's grab the binary package from one of the official Apache Kafka mirrors.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=104.6) [We'll use wget for this, but you can also download it using a web browser.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=109.74) [In this case,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=114.07) [we're downloading the version of Kafka that corresponds to the](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=114.87) [version of Scala installed on the system.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=117.81) [We can see that it's downloaded as an archive,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=121.24) [so now we'll extract it using the tar command.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=123.62) [Now I took the liberty to create a Kafka program files directory in](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=127.44) [usr/local/bin and copied the extracted archive contents there.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=131.3) [This step isn't required as you can run Kafka from wherever you would like,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=136.14) [even from the same location that you extracted it from the archive.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=140.6) [But let's explore the Kafka installation directory contents.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=144.44) [The site‑docs folder just contains an archive containing all of](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=149.14) [the documentation that you'll find online.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=152.86) [Let's go into the libs folder.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=155.65) [This folder contains all of the dependencies Kafka has in order](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=157.74) [to run. You'll notice there at the bottom these archive for](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=161.58) [ZooKeeper and its client library.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=165.37) [This enables Kafka to be a self‑contained installation,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=167.94) [not requiring ZooKeeper to be installed prior,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=171.4) [which is convenient to get started quickly.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=174.37) [Next is the config folder.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=177.64) [This is an important folder because here you'll find all of the files you'll](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=179.61) [need to configure all of the components of Apache Kafka.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=183.31) [Some of these files are out of scope for the current course,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=187.54) [but the ones related to the broker,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=191.31) [the producers, and consumers, we'll get into those in due time.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=193.16) [This file, server.properties,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=197.24) [is the configuration file for the Kafka broker itself.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=199.45) [By taking a quick peek into it,](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=203.04) [you'll notice that it's self‑describing and straightforward as](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=204.76) [far as knowing where to add custom settings.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=208.03) [Finally, the bin folder.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=211.34) [This folder contains all of the programs to get Kafka up](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=213.34) [and running in a variety of capacities.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=216.69) [You'll notice the windows folder as well. It contains batch files that more](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=219.35) [or less do the same job as the shell scripts you see here.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=224.01) [Like with the configuration files, many of these scripts](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=227.42) [will be outside of the scope of this course, and we'll stick to the most pertinent ones.](https://app.pluralsight.com/course-player?clipId=e1be1909-1f17-4717-80c6-86bd98aed084&startTime=230.66)

### [Apache Kafka Topics in Detail](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73)

[All right, now that we have Kafka installed,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=0.64) [let's get back to the course content.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=3.58) [At the beginning of the last module,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=6.24) [we briefly mentioned the concept of a messaging topic.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=8.08) [It is the primary abstraction of Kafka because it is](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=12.14) [central to its entire purpose.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=15.12) [Kafka topics are really just a named feed or category of messages.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=17.84) [One way to think of it would be to consider a mailbox.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=22.84) [It's an addressable, referenceable collection point for messages](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=26.13) [that producers send messages to and consumers retrieve messages](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=29.96) [from. In Kafka, a topic is a logical entity,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=34.01) [something that virtually spans across the entire cluster of](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=38.83) [brokers. Producers and consumers alike don't really know or care](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=42.34) [about where and how the messages are kept.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=47.18) [They just care about the topic to work with.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=49.86) [However, behind the scenes, for each topic,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=52.54) [the Kafka cluster is maintaining one or more physical log files.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=56.03) [We'll go into that last statement soon, but before we do,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=60.44) [let's get clear on what a topic is from a logical viewpoint.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=63.92) [Topics can span an entire cluster of brokers for the benefit of](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=67.27) [scalability and fault tolerance.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=72.43) [With the abstraction of a topic, a producer simply needs](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=75.44) [to publish messages to that topic.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=78.73) [How it's maintained and managed over the multiple brokers is not its concern.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=80.73) [Similarly,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=85.7) [consumers simply want to consume from a topic, regardless of where it is.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=86.59) [Let's zoom in a bit and look at what's happening within any given topic.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=91.53) [When a producer sends a message to a Kafka topic,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=96.64) [the messages are appended to a time‑ordered sequential stream.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=99.7) [Each message represents an event, or fact, that from the](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=104.14) [perspective of the producer is worthwhile to preserve and](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=108.19) [make available to potential consumers.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=111.66) [These events are immutable.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=114.25) [Once they are received into a topic, they cannot be changed.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=116.09) [Consequently,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=120.34) [if a producer happens to send a message that is incorrect or](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=121.44) [represent a fact that is no longer valid,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=125.86) [its only recourse is to follow up that previous message with a newer,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=128.7) [more correct one.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=133.42) [It would be the job of the consumer to reconcile between the](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=134.94) [messages when it reads them and processes them.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=138.79) [This style of maintaining data as events is an architectural style](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=143.04) [known as event sourcing and is becoming widely used as a means to](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=147.75) [manage independent caches of data in a reliable,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=152.26) [flexible, and distributable manner.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=155.92) [You'll recall I just said that a Kafka topic stores a time‑ordered](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=159.14) [sequence of messages that share the same category.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=163.49) [Let's look at a Kafka message from a logical standpoint and discuss what it](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=167.32) [contains. At a high level, a Kafka message has a timestamp that it's set](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=171.55) [when the message is received by a Kafka broker.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=177.54) [Furthermore, a message received gets a unique identifier.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=180.94) [The combination of the timestamp and its identifier form its placement](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=185.54) [in the sequence of messages received within a topic.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=189.96) [These identifier itself is referenceable by the consumers in order](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=193.64) [for them to operate autonomously and efficiently,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=197.77) [as you will shortly see.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=200.72) [Of course, the message itself has a binary payload of data,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=202.83) [which is what the producers and consumers really care about.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=207.23) [We'll get into the binary data in detail in the next](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=210.64) [module when we talk about producers.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=213.5) [From the consumer's perspective, they simply read messages from a topic.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=216.34) [One of the truly remarkable things about Kafka and a primary tenant,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=220.94) [if you recall, of its design goals was to make message consumption possible from](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=225.1) [a theoretically unlimited number of independent and autonomous consumers. Again,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=230.18) [there may be several consumers that are all interested in receiving the same](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=236.18) [messages at the same or different times.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=240.33) [Furthermore,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=243.84) [if a consumer happens to erroneously process some messages, that fault should](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=244.8) [not have any impact on any other consumer or the producers,](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=249.75) [for that matter. Each should maintain its own exclusive](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=253.88) [operational boundary from one another.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=258.12) [Even a complete crash should not keep others from operating.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=260.84) [So, how do consumers do this?](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=265.04) [We'll get into the guts of this in module five.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=267.39) [But since our current discussion is still from a logical point of view, let me touch on that.](https://app.pluralsight.com/course-player?clipId=4e929f79-f439-4640-840c-81f9f702df73&startTime=270.22)

### [The Consumer Offset and Message Retention Policy](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2)

[So back to the consumers.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=0.14) [How do they maintain their autonomy as far as message](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=1.85) [consumption from a common topic?](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=5.15) [It's called the message offset.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=6.85) [The message offset is a critical concept to understand](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=9.04) [because it is how consumers can do read messages at their own](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=13.59) [pace and process them independently.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=17.35) [Essentially, the offset is a placeholder.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=20.74) [You can think of it like a bookmark that maintains the last read position.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=23.52) [In the case of a Kafka topic,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=28.02) [it is the last read message. The offset is entirely](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=29.86) [established and maintained by the consumer.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=33.52) [Since the consumer is responsible for reading the](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=36.61) [messages and processing them on its own,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=39.21) [it stands to reason it should also keep track of](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=41.84) [what it has read and has not read.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=44.25) [The offset itself refers to a message identifier.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=47.14) [In fact, it is the same identifier I described a few slides ago.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=51.04) [Let's look at how this works.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=55.41) [When a consumer wishes to read from a topic,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=57.54) [it must establish a connection with a broker.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=60.32) [Upon the connection, the consumer will decide what messages it wants to consume.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=63.44) [This is entirely at the consumer's discretion.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=69.82) [If the consumer has not previously read from the topic,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=73.1) [or if it has, but wants to start over,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=76.57) [it will issue a statement to read from the beginning of the topic.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=79.54) [Behind the scenes,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=83.84) [this is essentially the consumer establishing that its message offset for the](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=85.25) [topic is 0. And as it reads through the sequence of messages,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=89.9) [it will inevitably come to the last message in the](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=94.32) [topic and move its offset accordingly.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=97.12) [Of course,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=100.84) [another consumer is likely interested in the message from the topic as well,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=101.48) [but another consumer could be at a different place in the topic.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=106.11) [It could've already read the messages from the beginning and simply is waiting](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=109.61) [for more messages to arrive so it can read and process them.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=113.35) [The key here is that it knows where it left off and](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=117.66) [can choose to advance its position,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=120.84) [stay put, or go back and reread another previously read message,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=122.97) [all without the producer, brokers, or other consumers needing to know or care.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=127.13) [When other messages arrive, the connected consumer will receive](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=133.04) [an event indicating there is a new message,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=137.05) [and it can advance its position once it retrieves the new message.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=139.96) [When the last message in the topic has been read and processed, the](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=144.44) [consumer can set its offset and at that point is caught up.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=148.07) [Kafka is immune to one of the challenges of most](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=153.14) [messaging systems, slow consumers.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=155.94) [We discussed this in the previous module.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=158.94) [The reason why Kafka is immune is because it has the means to](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=161.74) [retain messages over a long period of time.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=165.04) [The time it can retain messages is configurable and is](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=167.84) [known as the message retention policy.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=171.06) [All published messages are retained by a Kafka cluster, regardless if](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=173.5) [a single consumer has consumed a message. The length of time in which](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=178.25) [messages are retained is configurable in hours. By default, the](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=183) [retention period is 168 hours, or 7 days.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=187.96) [Beyond that, messages will start to fall off,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=192.64) [starting with the oldest messages that have expired or fallen past](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=195.62) [the retention period to make room for the new messages. The](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=199.1) [retention period is set on a per‑topic basis,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=202.33) [which means that within a cluster you could have hundreds of topics,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=206.02) [each with different retention periods.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=209.98) [Besides the retention policy,](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=212.74) [the ability to retain messages is obviously going to correspond to the storage resources available.](https://app.pluralsight.com/course-player?clipId=45f2603c-6f7f-49b1-a494-605ad1f4b3f2&startTime=214.9)

### [Demo: Starting Apache Kafka and Producing and Consuming Messages](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b)

[I think it's about time for a demo.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=0.54) [In this one,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=2.04) [we'll look at the basics of setting up a Kafka cluster with](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=3.3) [a single ZooKeeper server in a single Kafka broker and](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=6.76) [creating an Apache Kafka topic,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=10.64) [sending some messages to it with a producer client and reading](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=13.2) [from that topic with some consumer clients.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=16.63) [The demo will be very basic,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=19.64) [as it's meant to illustrate the quickest way to get](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=21.83) [started and see Kafka actually working.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=24.61) [As we go on,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=27.76) [we'll get more detailed in terms of what's happening behind the scenes.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=28.67) [I'm hoping you'll watch out for how we use the built‑in Kafka producer](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=32.64) [and consumer client applications and the message order in which we'll be](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=36.85) [producing and consuming messages in a topic.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=41.08) [There's a tendency to grasp onto every detail and want to](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=44.74) [understand the what and why for everything.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=47.85) [For example,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=50.84) [you may find yourself getting caught up on the](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=51.6) [command line parameters and options.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=53.79) [Don't worry about that, because we're going to cover those things soon enough.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=55.94) [The purpose for this is to give you a solid baseline](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=59.61) [of a working Kafka environment.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=62.78) [The first thing we need to do in order to use Kafka is to](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=65.53) [start the main components of the cluster.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=68.5) [Hopefully,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=71.11) [you remember what those are, the ZooKeeper instance and at least one broker.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=71.73) [Fortunately, Kafka makes this easy by giving us some shell programs.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=76.5) [Remember the bin folder from the Kafka installation demo?](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=81.21) [That's where we're going.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=84.54) [Notice the ZooKeeper shell programs. We'll use the zookeeper‑server‑start one.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=86.34) [Notice the USAGE hint.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=96.94) [It is expecting a configuration file to know how ZooKeeper](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=98.78) [should behave once started. You can examine and modify this](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=102.49) [file as needed, as it is found in the config folder with the](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=106.13) [other Kafka configuration files.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=109.8) [We'll just use the installation defaults this time.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=111.79) [When you run the shell program,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=121.64) [you'll see a bunch of info messages, but this signals that](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=123.3) [we've successively started ZooKeeper,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=126.72) [and now it is sitting there waiting for processes to connect to it.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=129.18) [To test that a ZooKeeper environment is running as expected,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=132.74) [we can connect to it via telnet and issue a ZooKeeper](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=136.15) ["four‑letter command", such a stat.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=139.16) [This gives us the status of the ZooKeeper server.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=143.44) [You may notice we're running in standalone mode.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=146.54) [That is to say there is only a single instance running for](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=149.33) [testing and development purposes. With ZooKeeper started, we](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=152.68) [can now start a single Kafka broker.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=156.57) [The process is very similar.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=159.08) [We simply use another shell program. Again, notice the USAGE](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=160.81) [hint. Like ZooKeeper it is expecting a configuration file to](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=164.54) [represent a specific broker instance. For this single broker](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=169.09) [demo, we'll use the defaults, but we'll come back to this soon.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=173.12) [Once the Kafka servers start command is executed,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=179.74) [you'll see a bunch of info messages whirling by the terminal again.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=183.2) [Notice the last few. It will say Registered broker 0 at path. That's](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=188.06) [saying that the broker has registered itself with the ZooKeeper server and](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=193.64) [is available to do work, so let's give it some work to do. Now with the](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=197.62) [server started, we'll create a topic.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=202.82) [We will use the handy shell programs to do this for us.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=205.79) [The one we'll use is kafka‑topics.sh. When we execute this,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=209.24) [the USAGE hint here is quite a bit more involved.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=214.29) [There are a lot of commands and actions that can be taken](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=217.64) [for this process of creating a topic.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=220.42) [I encourage you to study these options later.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=222.72) [We'll use a few of them as we go.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=225.66) [Take a look at the command I just typed in to create a topic.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=227.74) [Yes, that's quite a few things needed to successfully create a topic.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=231.31) [But let's talk a little bit about why. aside from the](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=235.94) [obvious create and topic commands,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=239.33) [you'll notice we needed to pass in the ZooKeeper server.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=241.94) [This is because there could be multiple ZooKeeper instances,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=245.54) [each managing their own independent clusters. By specifying the](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=249.18) [ZooKeeper server here, you're basically saying,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=254.04) [I want this topic to be created for this specific ZooKeeper managed cluster.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=257.12) [Remember,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=262.44) [it is the ZooKeeper component that is responsible for assigning](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=263.23) [a broker to be responsible for the topic.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=266.83) [Another important thing to call out is the flags](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=270.24) [regarding replication factor and partitions.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=273.16) [We'll talk about partitions and replication factors in a lot of](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=276.32) [detail later. When the topic was created,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=280.07) [some interesting things happened behind the scenes. First, ZooKeeper](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=285.15) [scanned its registry of brokers and made a decision to assign a broker](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=289.79) [as the leader for the topic, my\_topic.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=293.97) [Second, on the broker there is a logs directory, and in there a](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=298.14) [new directory was created called my\_topic,0.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=302.62) [Within this directory, there are two files, an index file and the log file.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=307.61) [We'll get into this as we talk next about partitions.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=313.14) [Another useful function of the kafka‑topics shell program is](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=319.54) [an option that enables us to inquire about the topics that](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=323.66) [are available on the cluster.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=326.84) [You do this with the kafka‑topics shell program](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=328.74) [with the option of list, and then, of course,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=332.13) [you have to pass the ZooKeeper server.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=334.81) [Now that we have a topic, let's produce and consume some basic messages.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=338.84) [First thing is to instantiate a producer.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=343.44) [By now, you can predict how we'll do this.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=346.32) [That's right, a handy shell program called kafka‑console‑producer.sh,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=349.09) [another case of a lot of usage hints. In the next module, we'll cover](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=353.27) [the majority of these, but for now,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=359.8) [we'll just go with the minimum to produce some basic messages.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=362.02) [Once you've executed this command,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=373.94) [you can keep the terminal window open most of the time,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=376.08) [or as long as you like, and type whatever you want here.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=379.24) [Everything you type and follow with Enter, you will cause the](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=382.49) [producer to send the message to the broker,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=385.96) [which will then commit it to its log,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=388.26) [waiting for a consumer to consume them. So let's get the other](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=390.49) [side working. Like the producer shell program,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=393.83) [there is a consumer shell program, kafka‑console‑consumer.sh.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=398.14) [This is how you execute it.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=404.3) [Immediately, you'll see it pulling the messages from the broker](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=406.65) [and displaying them on the terminal window.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=410.34) [Let's keep the consumer terminal window open and move over to the](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=413.09) [producer terminal window and write some more messages.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=416.79) [Lastly,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=434.14) [let's take a look now at the messages in the log. Warning, they won't make](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=434.79) [a lot of sense because most of the content is binary,](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=438.91) [but you'll recognize the text portion to prove that the messages were actually received and persisted in the log.](https://app.pluralsight.com/course-player?clipId=0283e5c0-ed57-45d2-90c3-fbf889c3ba2b&startTime=442.49)

### [Apache Kafka as a Distributed Commit Log](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185)

[Up to this point, we've discussed,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=0.14) [at a logical level, the critical concept of a messaging topic in Apache Kafka.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=1.96) [Throughout this description,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=7.74) [you may have drawn parallels with other types of data processing systems,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=9.22) [for sure other messaging systems that are also centered](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=13.32) [around the concept of a messaging topic.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=16.53) [But given what we have described as to how messages are](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=19.64) [represented as immutable events that are appended and](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=22.95) [persisted in a time‑ordered sequence, does this sound familiar to you?](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=26.07) [Think of the internals of other data processing systems that you've](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=31.04) [learned about or even worked with in your career.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=34.86) [Can you see some analogies?](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=37.64) [The reason we have spent our initial focus in this module on the](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=40.34) [logical abstract concept of the topic is to impress upon you this](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=43.69) [simple yet powerful basis upon which Apache Kafka's architecture](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=48.56) [was built, the commit log.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=52.77) [If you think about the internals of the database,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=55.64) [you'll see the corollary I am suggesting. A database transaction or](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=58.43) [commit log is the primary record of what happens.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=63.41) [It continually appends events in the order in which](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=67.14) [they are received. Log entries,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=70.44) [new or otherwise, are maintained in the physical log](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=73.07) [file maintained by the database.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=76) [From there,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=78.74) [any number of derivative data structures can be formed to represent the](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=79.61) [contents of the log. In relational databases, for example,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=83.34) [these structures can be things like indexes,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=87.55) [tables, and views.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=90.06) [They are just projections from that authoritative](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=91.64) [source of truth we're calling a log.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=93.84) [This concept is useful for many reasons.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=96.64) [For example, when a database crash occurs,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=99.44) [it is the commit or transaction log that serves as the source from which](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=102.37) [restoration is possible. Because when you read from the log,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=106.54) [you're essentially able to replay all of the things that](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=110.06) [happened in the order in which they happened.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=113.12) [Additionally,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=115.82) [it is also the basis for which replication and](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=116.82) [distribution can occur for redundancy, fault tolerance, and even scalability.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=119.88) [If you go to Apache Kafka's project site at kafka.apache.org,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=126.24) [which I encourage everybody to do by the way,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=130.62) [you'll notice the very first statement made to describe](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=133.6) [what Apache Kafka is and how it works.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=137.14) [This is that statement.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=140.11) [I emphasize the last three words based on the](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=141.74) [discussion we have had up to this point.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=144.56) [Yes, Apache Kafka is a messaging system.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=146.94) [We've already discussed this.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=149.73) [Yes, it uses publish and subscribe semantics as the basis for its data movement.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=151.36) [But what it really is, at its heart, is a distributed commit log.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=157.14) [In a nutshell, if you were to strip off all of the higher abstraction](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=161.74) [layers from a database and be left with its commit log,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=165.75) [and if you were to leverage the principles we discussed in the last module](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=169.14) [as the basis for distributing the contents of that log,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=173.2) [you would have Kafka.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=176.31) [In a way,](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=178.02) [Kafka is really just a highly distributed raw database that](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=178.78) [brokers reads and writes using publish and subscribe semantics. Simple and really cool.](https://app.pluralsight.com/course-player?clipId=1af4853d-9b66-4ba3-864e-109beb8ba185&startTime=183.27)

### [Apache Kafka Partitions in Detail](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6)

[Now we're ready to cover the next most important concept in Apache Kafka,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=0.14) [partitions.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=4.79) [This is where we start to cover how the logical concept of a topic is](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=6.37) [implemented in the physical world and how, within this physical realm,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=11.21) [Kafka is able to operate and deliver upon its amazing promise.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=15.62) [The topic, as a logical concept,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=20.74) [is represented by one or more physical log files called partitions.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=23.37) [The number of partitions in a topic depends on the circumstances](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=28.54) [in which Apache Kafka is intended to be used.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=32.53) [Nonetheless, the number of partitions per topic is entirely configurable.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=35.74) [The partition itself is central to how Apache Kafka achieves](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=41.24) [its amazing nonfunctional capabilities, such as its ability to scale out,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=45.27) [providing for greater levels of fault tolerance, and](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=50.23) [achieving higher levels of overall throughput.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=53.35) [As such, each partition is maintained on at least one or more brokers.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=56.34) [Let's take a closer look.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=61.64) [As you will recall from the demo, we created a topic called my\_topic,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=63.81) [and as part of that process,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=69.49) [we were required to indicate the number of partitions it would have,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=71.07) [as well as the topic's replication factor.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=75.2) [For the simple case of the demo, we chose a single partition.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=78.54) [Obviously, we could've set any reasonable number we wanted,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=82.84) [but I chose a single one to illustrate the basics of how partitions work and](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=86.77) [why they are a critical element of the Kafka architecture.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=91.66) [At a minimum, each topic has to have a single partition because that partition,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=95.84) [as I mentioned,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=101.02) [is the physical representation of the topic as a commit log stored on one or](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=102.09) [more brokers. Referencing the demo, that log was maintained on the broker's](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=106.7) [file system in the directory tmp/kafka‑logs.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=112.4) [For that one topic we created, there was a subfolder called my\_topic‑0,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=117.09) [which contained the log for that single partition.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=123.43) [There are some legitimate reasons for choosing a single partition topic,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=127.11) [even in production.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=130.87) [Those reasons we'll cover and module five.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=132.49) [But like any design choice, there are tradeoffs.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=134.99) [In this case, the tradeoffs can limit scalability and throughput.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=138.4) [Why? As you know,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=142.4) [a physical node, upon which the broker and the partition log resides,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=144.27) [is limited by a finite amount of computational resources, such as CPU,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=148.44) [memory, disk space, and network.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=154.43) [While it is possible to keep adding more faster and bigger](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=157.84) [resources, eventually, under the strain of high use, your](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=161.93) [only real option is to scale out,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=165.84) [especially considering each partition you have must fit on one machine.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=168.39) [You cannot split a single partition across multiple machines.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=174.54) [Therefore, if you only have one partition for a large and growing topic,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=179.02) [you will be limited by the one broker node's ability to capture and](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=184.03) [retain messages being published to that topic,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=188.39) [not to mention the possible I/O constraints it will run into. In Kafka, that](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=191.24) [means you'll need more brokers in the cluster and topics that can leverage](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=196.79) [those brokers by partitioning into multiple partitions.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=200.99) [It is the biggest way to scale Apache Kafka. So let's](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=205.34) [suppose we have created the same topic, but with three partitions instead of one.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=209.28) [How is this different? In that scenario,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=214.94) [with three partitions, we're causing a single topic to be](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=218) [split across three different log files,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=221.79) [ideally managed by three different broker nodes.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=224.5) [Each partition is mutually exclusive from one another in](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=228.02) [that they receive unique messages from a Kafka producer](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=232.01) [producing on the same topic.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=235.28) [This enables each partition to share the burden of the message load](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=237.64) [across multiple different broker nodes and increase the parallelism of](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=241.98) [certain operations like message consumption.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=246.42) [Despite each partition getting different messages,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=249.94) [the manner in which they are managed is the same.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=253.09) [It's a time‑ordered sequence of events that are](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=256.46) [appended to buy the Kafka producer.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=259.52) [It's the same.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=261.56) [I hope you're wondering how the Kafka producer splits the](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=263.04) [messages across the different partitions.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=266.92) [That is a good question and one we will discuss in the next module.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=269.54) [But to tease the answer a little bit,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=273.74) [it is based on a partitioning scheme that can be](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=275.99) [established by the producer. In this illustration, a](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=278.93) [specific partitioning scheme is not used,](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=282.8) [so the producer is just doing it round robin.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=285.74) [It's a very important topic because it has implications on how balanced the partitions are for any given topic.](https://app.pluralsight.com/course-player?clipId=c0b992da-5490-421f-848e-5d4a350945d6&startTime=288.53)

### [Distributed Partition Management in Apache Kafka](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6)

[Let's start to bring this together into a consolidated big picture.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=0.34) [From a work distribution standpoint,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=3.86) [remember in the last module we talked about leaders](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=6.75) [being elected to take ownership of work.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=10.37) [This is how partitions get distributed to the available brokers in the cluster.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=13.54) [For example, when a command to create a topic with three partitions has issued,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=17.64) [it is handled by ZooKeeper, who is maintaining metadata regarding the cluster.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=23.72) [At this stage,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=29.56) [ZooKeeper is specifically going to look at the available brokers and](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=30.6) [decide which brokers will be made the responsible leaders for](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=35.07) [managing a single partition within a topic.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=38.97) [When that assignment is made,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=42.44) [each unique Kafka broker will create a log for the newly assigned partition.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=44.23) [This log will be precisely what we saw in the demo,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=49.52) [and it's hosting directory will be named by the topic and](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=52.62) [the partition in which it represents.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=55.74) [Additionally, as partition assignments are broadcast,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=58.64) [each individual broker maintains a subset of the metadata that ZooKeeper does,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=62.44) [particularly the mapping of what partitions are being](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=67.55) [managed by what brokers in the cluster.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=70.58) [This enables any individual broker to direct a producer client to the](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=73.54) [appropriate broker for producing messages to a specific partition. Along](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=78.49) [the way, status is being sent by each broker to ZooKeeper so that proper](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=84.14) [consensus can be maintained in the cluster.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=88.87) [When a producer is ready to publish messages to a topic,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=92.14) [it must have knowledge of at least one broker in the cluster so it](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=95.67) [can find the leaders of the topic's partitions.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=99.86) [Each broker knows which partitions are owned by which leader.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=103.44) [The metadata related to the topic is sent back to the producer so it](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=108.24) [can begin to send messages to the individual brokers participating in](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=112.46) [managing the topic, or I should say, the partitions in that topic.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=117.08) [On the surface,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=121.34) [consumers operate much the same way a producer does, except it leverages](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=122.61) [ZooKeeper more for some very important ways that I'll have to defer the](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=127.52) [explanation of which until module five.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=131.62) [But one way or another, when consuming messages from the cluster,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=134.84) [the consumer inquires of ZooKeeper which brokers own which](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=138.89) [partitions and gets additional metadata that affects the](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=143.48) [consumer's consumption behavior,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=147.78) [particularly in scenarios where there are large groups of](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=150.14) [consumers sharing the consumption workload.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=153.35) [Once the consumer knows the brokers with the partitions that make up the topic,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=156.64) [it will pull the messages from the brokers based on the message offset per](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=162.07) [partition. Because messages are produced to multiple partitions and at](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=167.19) [potentially different times, consumers working with multiple partitions are](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=172.45) [likely going to consume messages in different orders and will therefore be](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=177.6) [responsible for handling the order if it is required.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=182.28) [We've established that having multiple partitions is a must in order to](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=186.54) [effectively scale out a Kafka cluster's ability to handle large volumes](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=190.63) [of messages and increase the degree of parallelism in which consumers](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=195.47) [can consume those messages.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=199.74) [However, there are drawbacks to having too many partitions.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=202.04) [Like any architectural decision,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=206.51) [the number of Kafka partitions will have its tradeoffs, and](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=208.73) [therefore, a right balance needs to be found based on use case](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=212.31) [needs and resource availability.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=216.55) [As you've seen,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=219.63) [ZooKeeper plays a big role in making sure all of the brokers](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=220.88) [in the cluster are working in concert.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=224.79) [The more partitions there are,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=227.74) [the more entry ZooKeeper has to make to keep track of them.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=229.82) [And since ZooKeeper works on this registry in memory,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=233.64) [the resources on ZooKeeper can become constrained.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=237.13) [Now this can be mitigated by ensuring your ZooKeeper](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=241.04) [ensemble is properly provisioned,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=244.07) [commensurate with the growth of topics and their partitions.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=246.85) [Every message in each partition is totally ordered,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=251.04) [as we discussed earlier, in the sequence in which it is received.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=254.23) [However, as I mentioned, a topic may consist of many partitions,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=258.34) [so there will not be a global order to messages in](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=263.86) [a topic across all partitions.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=266.93) [This can be complex if your consuming application needs to have a global](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=269.94) [messaging order in the topic across all partitions. To get a global order](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=274.33) [without the consumers having to manage the ordering process,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=279.48) [you may need to consider a single partition for a topic.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=283.7) [The tradeoff there is obvious. You'll be limited in terms of scalability](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=287.84) [beyond the single broker managing that single partition.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=292.37) [Alternatively,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=296.14) [you can intelligently use consumers and consumer groups to consume messages from](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=297.44) [the topic partitions and handle the ordering process there.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=302.51) [We'll cover these scenarios more in module five.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=307.04) [Lastly, when you have a higher number of partitions,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=310.39) [the process of a leader falling over to another can start to get time consuming.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=313.81) [Now the failover process is handled very fast, in the low milliseconds,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=319.54) [but in larger clusters with a large number of partitions,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=324.94) [this can start to add up, which is why in big,](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=328.6) [big implementations you'll see multiple clusters on their own.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=331.75) [These are a few major considerations, but not the only ones.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=337) [We'll discuss more of these as we talk about producers in module four and consumers in Module five.](https://app.pluralsight.com/course-player?clipId=24a55c15-ccd3-43f8-9643-10f0f3c3fdc6&startTime=340.49)

### [Achieving Reliability with Apache Kafka Replication](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62)

[So we've covered partitions fairly well.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=0.14) [We know what they are, why they are.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=2.83) [But you remember this slide.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=5.24) [Based on what we've discussed thus far,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=6.52) [what are we missing that we should cover in this module?](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=9.42) [Yes, we need to discuss how partitions enable work not only to be distributed,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=12.94) [but reliably distributed,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=18.49) [as per the last module's distributed computing principles.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=20.54) [There's a lot of different types of faults that can](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=24.74) [happen in a distributed system,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=26.93) [exponentially more than in a non‑distributed system.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=28.85) [For example, what if a broker fails and becomes unresponsive?](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=32.83) [What if there is a network issue that makes a broker unreachable?](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=37.64) [Since the data is stored by the broker,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=41.53) [what happens if the disk fails or the data is otherwise inaccessible?](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=43.75) [Let's suppose we have a multi‑broker cluster,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=48.94) [and each broker is the leader for one or more partitions.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=51.8) [What would happen if one of those brokers were to fail or become unresponsive?](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=55.84) [This could be a potential catastrophe, because it could mean the loss of data.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=61.54) [Due to the responsibility of ZooKeeper, when it determines that a broker is down,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=65.4) [it will find another broker to take its place,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=71.34) [and the metadata used for work distribution for either](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=74.18) [producers or consumers will get updated,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=77.2) [and the system will go on.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=80.05) [But without some redundancy between brokers,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=82.34) [there could be unrecoverable data loss because the data](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=85.12) [being managed by the failed broker, unless replicated to another cluster,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=88.62) [is now inaccessible.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=93.6) [Once ZooKeeper and the brokers have handled the reassignment and the](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=95.23) [producers and consumers in the system are updated,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=99.33) [they can continue to publish and consume messages to](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=102.14) [and from the topic partitions,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=105.1) [but the previous messages that were lost in the partition](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=106.3) [that failed will still be inaccessible.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=110.82) [This is why the designers of Kafka created a facility enabled through a](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=114.34) [configuration property to ensure redundancy of data, so in the inevitable](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=118.52) [event of an individual broker failure or fault,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=123.59) [there isn't any data loss. This configuration](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=126.63) [property is one we've already seen.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=130.02) [It's called the replication factor. You'll recall in the demo we've](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=132.74) [miserably set that to 1, which means that the topic's partitions will](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=137.29) [only have a single replica at any given time.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=142.11) [We can excuse ourselves though because it's just a demo](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=146.24) [running on a virtual machine, but it's an important](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=149.32) [configuration property not to be for gotten,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=152.12) [so let's discuss it further.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=155) [Obviously,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=156.84) [it's a good idea to leverage the replication facility](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=157.64) [designed in Kafka for the reasons we discussed.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=160.78) [It's a critical safeguard that enables reliable work distribution,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=163.54) [which for a distributed system is a must,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=167.85) [as failure of some sort is virtually guaranteed.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=170.54) [This ensures that messages are stored redundantly across more than one broker,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=173.76) [which makes the overall system more resilient and fault](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=178.66) [tolerant when a broker failure happens,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=182.22) [all for the purposes of mitigating the risk of data loss.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=184.73) [By setting the replication factor to N,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=188.32) [you virtually guarantee yourself up to N‑1 server or broker failures.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=192.13) [Therefore,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=198.64) [it is generally recommended that you set the](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=199.34) [replication factor accordingly. In fact,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=201.68) [a minimum of two to three so that failures or machine maintenance](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=204.69) [will not interrupt the cluster's operation.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=209.48) [Finally,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=212.64) [it is important to note that the replication factor can vary from](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=213.38) [topic to topic because it is set at the topic level.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=217.53) [Let's walk through how Kafka operates when the replication factor is set.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=221.74) [This should look familiar. Here we're using the built‑in](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=227.24) [shell program for creating a topic, as I showed a few slides ago.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=230.14) [In this particular scenario, let's simulate the same thing,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=234.51) [but this time set the replication‑factor to 3. This is the same create topic](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=237.82) [scenario as before. The only difference is we'll walk through what happens](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=243.03) [when we set the replication‑factor to 3 instead of 1. With a](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=246.96) [replication‑factor of 3 set, it is the leader's job to get peer brokers to](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=250.96) [participate in a quorum for the purposes of replicating the log to achieve](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=255.69) [the intended redundancy level.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=260.78) [When the leader of a partition has a quorum,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=263.24) [it will engage its peers and start copying the partition log.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=265.92) [When all members of the replication quorum are caught up and a](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=270.74) [full synchronized replica set is in place,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=274.4) [it is reported throughout the cluster that the](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=277.39) [number of in‑sync replicas, or ISRs,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=280.26) [is equal to the replication factor for that topic and each partition within it.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=283.89) [Obviously,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=289.34) [this is an important metric. When the ISR is equal to the](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=290.62) [replication factor, the topic and each partition within that](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=294.72) [topic is considered to be in a healthy state.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=299.08) [If for any reason a quorum cannot be established and/or the](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=302.84) [number of ISRs fall below the configured replication factor for](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=306.58) [the topic, intervention may be required.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=310.59) [There could be legitimate planned and unplanned reasons a](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=314.04) [broker is not able to be replicated to.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=317.23) [And because of that,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=319.94) [Kafka doesn't automatically go out and search for a new](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=321.09) [following peer to replace the quorum member.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=324.09) [Despite how resilient Kafka is, vigilant monitoring and compensating actions](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=327.14) [are needed to eventually replace or tune a lagging or missing‑in‑action member](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=332.55) [of the quorum. In the last demo, you may have noticed I used this useful](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=337.55) [command in the kafka‑topics shell program.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=342.12) [It was a command for describe. This useful command enables](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=345.78) [you to see what's going on within a topic,](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=349.88) [including all of the partitions, the leaders it's assigned to, the replicas, and the ISRs.](https://app.pluralsight.com/course-player?clipId=e1d9e7f3-c54a-4e46-b5e8-8867acc7db62&startTime=352.85)

### [Demo: Fault-tolerance and Resiliency in Apache Kafka](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025)

[Let's go through one last demonstration that will cover](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=0.44) [as much as we have up to this point, which is a lot.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=3.41) [In this demo, we will set up a Kafka cluster with multiple brokers.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=6.7) [We'll go with an odd number of three.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=12.55) [And for simplicity,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=14.31) [we'll configure it with a topic with a single partition](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=16.23) [that has a replication factor of 3.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=20.01) [We'll go with this configuration to illustrate what details are](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=23.24) [provided when using the describe command,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=26.7) [and then we'll simulate a failure scenario and witness how](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=28.61) [Kafka is resilient and enables operation to continue](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=32.99) [transparently to both publishers and consumers.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=36.31) [Setting up a multi‑broker cluster on a single machine is relatively simple.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=39.94) [Remember those server.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=44.34) [property files that we showed at the beginning of the module?](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=45.67) [Well,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=48.9) [you just basically need to create a separate server.properties](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=49.24) [configuration file for each broker you want to instantiate.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=53.05) [So here I already have ZooKeeper running,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=57.34) [so let's just start up the three Kafka brokers in their own terminal windows.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=59.6) [You'll recall that we can do that by executing the shell](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=64.17) [program for kafka‑server‑start and passing the](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=67.94) [configuration properties. And each one, of course,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=71.57) [as I just said, is going to have their own properties file, one per broker.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=74.62) [Now let's create a topic with a replication‑factor of 3 and a single partition.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=82.64) [We'll call this topic replicated\_topic.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=87.78) [Let's run that. Yep, and as you can see,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=91.53) [we successfully created the topic, replicated\_topic.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=94.63) [Now we can check some of the details about the topic we just](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=99.04) [created by using the handy‑dandy describe command.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=102) [Let's run that and take a look.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=105.67) [All right.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=107.01) [Okay, let's expand this a little bit more so we can see everything on one line.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=107.97) [So it shows us that we have only one partition with a](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=114.04) [replication factor of 3, just like we asked. Further in the](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=117.44) [details, it lists the information by partition.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=120.97) [If this were a topic with more than one partition,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=123.98) [there would be multiple lines here,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=126.99) [one for each partition. For this single partition, partition 0, it](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=128.69) [shows the leader host, in this case, node 1.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=133.24) [Additionally, it indicates the replicas in place for this partition.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=136.29) [We see 1, 0, and 2, meaning there is a replica on node 1,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=140.24) [obviously because its leader, node 0, and node 2.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=144.67) [Furthermore, we see the in‑sync replicas in the same order, 1, 0, and 2.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=148.48) [Since the number of ISRs is equal to the replicas,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=154.44) [we can safely say the partition and the quorum](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=158.18) [managing it are in a healthy state.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=161.1) [Now let's create a simple producer and produce some messages to the new topic.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=163.25) [Okay, so here's a producer terminal window.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=167.21) [You'll notice it's the same as before,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=169.74) [but now you understand a little bit more about brokers and](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=171.94) [how they work together, and so you'll notice that we're](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=175.05) [passing a broker‑list parameter.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=177.31) [I'll explain why this is needed in the next module.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=179.36) [So now let's add a few messages, and now let's quickly](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=182.36) [create a consumer to retrieve the messages.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=189.67) [Okay, here's a terminal window for the consumer,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=192.68) [and you'll notice there's not really anything new here.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=195.46) [The messages were retrieved, and now the consumer is waiting for more.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=198.06) [At this point,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=201.47) [we're going to simulate a broker fault, and after we'll use the describe](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=202.21) [command to see how Kafka handled the broker going down as far as the](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=206.49) [topic management goes. Here I have the terminal window that I used to](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=210.44) [launch the first broker, or broker 1.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=214.34) [So why 1? Well, because in the topic details, it says the leader for the topic,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=217.34) [and its single partition is node 1.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=223.3) [So the idea here is to kill it and see what happens.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=226.44) [All right, let's kill this broker.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=230.94) [Okay, so that did it.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=233.54) [Now, you can see here in the INFO message that the node 1 shut down.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=235.21) [All right,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=241.04) [so now let's go to the window where we had the details](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=241.77) [up and pull up the details again.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=245.27) [Okay,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=249.14) [so here you can see the same topic with the same partition](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=249.51) [count, same replication factor, but now the leader has](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=252.43) [changed. The leader is now node 0.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=255.9) [Whereas if we look at the previous details, the leader was node 1.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=258.94) [Now it's again node 0.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=263.42) [So if we look at the replicas at this point,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=265.84) [we still see that there's three replicas.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=268.6) [There's 1, 0, and 2. But if you look at the in‑sync replicas,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=270.69) [there's only two in‑sync replicas.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=276.37) [One is gone because we killed broker 1.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=278.23) [It is no longer available.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=281.6) [So, as an administrator,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=283.21) [if you look at this and you see that there's two in‑sync replicas and](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=284.74) [there's a replication factor of 3, what does that tell you?](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=288.2) [Well, that tells you that your quorum is unhealthy,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=291.13) [that there is a missing replica, and it needs to be replaced.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=294.28) [Now, if there were another broker available,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=298.74) [Kafka would've already added it to the quorum and started](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=301.66) [replicating it to take the place of the lost broker 1.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=305.29) [But if you recall, we only had three brokers to start with,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=308.16) [we killed one, and that was all there is.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=311.39) [But now let's go back to the producer and the consumer](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=313.74) [and see how they handled things.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=316.85) [So here's the producer terminal.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=318.68) [It's just sitting there humming along.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=320.84) [It's as if it doesn't even know anything happened.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=322.61) [Well, behind the scenes it does, but it's still ready to publish messages.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=324.66) [Whereas when we killed broker 1, something happened in the consumer.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=329.5) [If we scroll through the messages printed to the terminal,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=333.97) [you'll see a lot of exceptions and stack trace information with](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=336.84) [warnings indicating that it attempted to pull or fetch records](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=340.68) [from the broker, and it was unsuccessful.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=345.04) [But notice this didn't cause the consumer to fail.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=347.64) [It's still there waiting for new messages to arrive. In module five,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=351.3) [we're going to talk more about the consumer and what causes these kinds](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=355.18) [of errors to happen when there is a change in the brokers that are](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=359.61) [providing the messages. So to prove this,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=363.37) [let's go to the producer terminal window and produce another message.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=365.62) [We'll call it My Message 4. So in the consumer terminal window,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=369.94) [you'll see that it was able to retrieve the message just fine.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=375.44) [Just because a broker falls out of the quorum doesn't mean that any data](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=378.15) [was lost, and that's thanks to the replication factor.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=382.9) [Sure, the cluster could use another broker to be a complete,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=386.37) [healthy quorum, but things are still working.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=389.53) [Okay, I hadn't planned on this, but just for fun,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=392.44) [let's get broker 1 back into the ring.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=395.21) [Let's start him back up.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=397.45) [Now that he started, he's joined himself back into the cluster.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=399.47) [Now, if we go back up here and do a describe on the topics,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=403.44) [we're going to get some information about the new membership. So you'll](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=407.67) [see that it's the same topic, and the leader is still 0,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=412.93) [as it was before, because that didn't change.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=417.51) [Nothing changed about node 0.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=420.04) [The replicas are the same.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=422.24) [We still have 1, 0, and 2. But the difference now is is that our](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=424.21) [in‑sync replicas have gone from two, when we killed the first broker,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=428.58) [now it's back to three, and it 0, 2, and 1.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=433.58) [So I hope that can illustrate how resilient a Kafka cluster can be.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=437.56) [Not only the cluster itself, but the producers and the consumers.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=442.14) [I hope this module's explanation of the brokers,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=446.64) [the partitions,](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=449.54) [topics, and everything has given you an understanding as to why it](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=450.59) [can be that resilient at the broker side. The next two modules on](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=454) [producers and consumers will talk about what happens internally to enable them to be just as resilient.](https://app.pluralsight.com/course-player?clipId=9b1fb36d-da06-4b53-bb65-982862c20025&startTime=457.27)

### [Module 3 Summary](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a)

[Wow!](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=0.24) [We covered a lot in this module.](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=0.67) [We needed to because there are a lot of moving parts in Kafka,](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=3.66) [and it's important to understand more about it before we](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=7.11) [start exploring what it means to build producing and](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=10.15) [consuming applications with it.](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=13.37) [In this module,](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=15.84) [we spent a bit of time discussing topics as logical concepts and partitions as](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=16.8) [the physical commit log that stores the topic's messages.](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=22.32) [We dived deeper into the role of brokers in a Kafka cluster,](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=26.54) [particularly as it pertains to partition management and behavior.](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=30.76) [As we went through these things,](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=35.64) [I tried my best to continually map it to the previous module](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=37.26) [where we discussed distributed systems and how Kafka embodies](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=41.25) [distributed systems principles.](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=45.45) [For example, we covered how brokers become leaders to own and manage partitions.](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=47.51) [We also looked at work distribution and failover.](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=53.24) [We saw Kafka in action.](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=56.74) [We did three demos,](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=58.53) [and I hope they were illustrative enough of the concepts. They were](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=59.87) [intended to motivate you to explore more on your own and to continue](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=63.44) [to go deeper and further into this course.](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=68.07) [There are a lot of things to Kafka,](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=70.94) [but I think we covered a good amount of detail that is relevant](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=72.78) [to understanding how it all works. Of course,](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=76.36) [we'll spend more focused time on this in the few modules remaining.](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=79.04) [All in all, I think we laid down a solid foundation upon which we can build upon,](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=82.74) [and I look forward to discussing more about producers and consumers in the next few modules.](https://app.pluralsight.com/course-player?clipId=7a35a5a0-e8d6-4064-94cb-3a9270f7165a&startTime=87.28)

## [Producing Messages with Kafka Producers](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db)

### [Introduction and Setting up an Apache Kafka Development Environment](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db)

[In this module,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=0.19) [we'll be getting into more details about the Apache Kafka producer.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=1.13) [In reality,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=5.81) [we've covered quite a bit about what the producer does. Here, we'll look at](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=6.62) [how it does it and what resources are available to developers to write](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=11.47) [applications that publish messages to Kafka. I will walk through how to](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=15.72) [build your own Kafka producer and spend some time covering some important](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=19.96) [configuration properties that affect the message sending behavior. As we did](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=24.22) [in the last module, the first thing I'd like to do is get the setup out of](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=28.86) [the way. By setup, I mean specifically getting a development environment](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=33.1) [established to develop Apache Kafka producer and consumer applications.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=37.74) [Getting a development environment set up is really straightforward.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=43.37) [We will essentially just add the Apache Kafka client libraries using a](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=47.02) [dependency manager and import the packages into the environment.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=51.3) [Once we've validated that the dependency manager has properly](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=55.94) [imported the packages, we will then briefly walk through the API.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=59.65) [For a successful setup and subsequent exercises, you will need to](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=62.93) [have a standard integrated development environment such as](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=68.58) [JetBrains IntelliJ or Eclipse.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=71.89) [There is a free version of IntelliJ called the Community Edition, and,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=74.94) [of course, Eclipse is free as well. In this course,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=78.72) [I will show my bias for IntelliJ.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=81.89) [Aside from some of the user interface differences,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=84.64) [the process should be more or less the same for getting set up in Eclipse.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=87.47) [It should go without saying, but you'll need the latest Java JDK.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=92.18) [Currently, the latest is Java 8. You should have the Maven dependency](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=96.49) [manager installed and plugged into your IDE to make things easy. I](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=101.2) [recommend version 3 of Maven. While not required for development per se,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=105.81) [you should have access to a test Kafka cluster.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=111.28) [By that, I mean at least one running Kafka broker.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=114.93) [This will enable you to test the producer and consumer](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=119.04) [applications you build end to end. In this course,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=121.89) [you'll notice I am developing client applications within](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=125.35) [the same virtual machine for simplicity. So the first thing](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=128.47) [is to launch the IDE itself.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=131.98) [Next, select Create New Project. Select Maven.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=133.93) [If you haven't already, select your project SDK or add one.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=137.64) [This is where you should specify the latest version of the](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=141.84) [Java JDK, Java 8 preferably. Select Next.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=144.47) [This is where you will add your own details related to](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=148.58) [your Kafka applications project.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=151.27) [When you're finished,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=153.44) [you continue through and hit Finish. With your project open,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=154.49) [navigate to your project's POM file.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=158.33) [Here we will need to add the Kafka dependencies so Maven](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=160.84) [can import the packages into the project.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=164.11) [Now let's add the dependencies.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=166.93) [This is pretty typical,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=168.96) [but what you're going to do here is you're going to want to add the](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=170.33) [org.apache.kafka as the groupId, kafka‑clients as the artifactId,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=172.86) [and then the version is going to be 0.10.0.1 Now let's take a look at](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=179.84) [the API we'll be using in this and subsequent modules.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=185.18) [Next, locate and expand the kafka‑clients external library node.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=188.7) [Next,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=193.73) [expand the org.apache.kafka node. Here you'll see the clients and common](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=194.16) [namespaces. Feel free to browse the common namespace,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=199.68) [but we'll be going into the clients namespace for most of the time.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=203.86) [In the clients namespace,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=207.74) [you'll find all of the objects you'll be working with directly in](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=209.3) [creating either producer or consumer clients.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=212.9) [For producer development,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=216.64) [you'll be using the Producer namespace. For the consumer development,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=218.18) [you'll use the Consumer namespace. The Producer namespace, you'll see the](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=222.11) [main classes and interfaces will be looking at,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=227.46) [particularly the KafkaProducer class and the ProducerRecord class.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=230.28) [So far, we've really just focused on the Apache Kafka producer externals,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=236.22) [the high‑level component view of a producer and its interaction](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=241.24) [topology with the rest of the Kafka cluster of brokers.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=245.15) [Now let's explore the high level of what goes on within the producer.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=248.84) [Since the producer is a piece of software, what we'll cover](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=253.24) [here is a logical representation of the key components and how](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=256.63) [they work together to send messages.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=260.63) [We'll use this as a map to go further into each](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=263.24) [component throughout this module.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=266.67) [Admittedly, this is a busy diagram,](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=268.94) [so let's go into it piece by piece and cover what each](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=271.74) [component does and at what point in the producer lifecycle they come into the picture.](https://app.pluralsight.com/course-player?clipId=9af48c31-5c9d-4abc-b812-e61094ed22db&startTime=275.09)

### [Basics of Creating an Apache Kafka Producer](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452)

[When creating a Kafka producer client application,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=0.04) [you'll first need an object to represent the required](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=3.27) [configuration properties needed to start up a producer.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=6.69) [As indicated here, there are three required properties needed,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=10.74) [bootstrap.servers and both key and value serializers.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=15.24) [Let's take a look at the code and describe them a bit more.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=19.64) [Configuration items are generally key‑value pairs,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=23.28) [so to construct a dictionary of key‑value pairs that represent the](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=27.3) [configuration settings for the Kafka producer,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=30.91) [the easiest way to do it is to use the Properties class](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=34.1) [from the core java.util's library.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=38.07) [You'll recall from the previous module that when we](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=41.44) [used the Kafka producer shell program,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=44.27) [we simply needed to supply a list of brokers for the producer to connect to.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=46.71) [This corresponds to the bootstrap.servers configuration](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=52.04) [setting needed for the producer to start up.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=56.11) [The producer doesn't connect to every broker referenced in this list,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=59.64) [just the first available one.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=64.14) [It uses the broker it connects to for discovering the](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=66.31) [full membership of the cluster, which, of course,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=69.71) [can change at any time.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=72.61) [It uses this list to determine the partition owners or leaders](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=74.54) [so that when it's ready to send messages,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=79.19) [it can do so immediately.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=81.41) [It is a best practice to provide more than one broker in](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=83.64) [the broker list in the unlikely event that the first](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=87.11) [broker specified is unavailable.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=90.5) [Next is the key and value serializers.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=94.04) [If you recall in the last module,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=96.98) [I mentioned how the message content is encoded as binary.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=99.58) [This is to optimize the size of the messages not only for network transmission,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=104.19) [but for storage and even compression.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=109.46) [Since it is the producer that serves as the beginning of a message's lifecycle,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=112.17) [it is responsible for describing how the message contents are to be](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=117.21) [encoded so the consumer can know how to decode them.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=121.75) [In this example, you'll notice that for both the key and value.serializer,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=126.14) [we're using the StringSerializer class,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=130.83) [which is the most common serializer scheme used in Kafka.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=133.54) [You're probably wondering what is meant by a key and value,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=137.74) [and why are they so important that a producer requires their](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=141.63) [serialization strategy to be established up front?](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=145.19) [That's a good question.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=149.12) [We'll get to that shortly.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=150.53) [These are but three of the many configuration settings that can be set.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=151.63) [For a full list of settings,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=156.38) [always refer to the producerconfigs section of the Kafka documentation site.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=158.56) [We will cover more of the important but optional settings as we continue.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=164.67) [Like any standard application, you need to have an entry point.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=169.34) [In this case,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=172.84) [we'll be hosting a Kafka producer within a standard Java console application,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=174.26) [and the boilerplate code for this should be evident.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=178.66) [Here,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=181.38) [you'll see the creation and setting of the Properties](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=182.01) [dictionary for configuration items, as described in the last slide,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=185.54) [followed by the primary class instantiation statement that makes this generic](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=189.85) [console application an actual Kafka producer application.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=194.92) [There are different approaches to writing this instantiation statement,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=199.53) [but this is by far the most simple.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=203.05) [The other options are based on what values you want to provide to the](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=205.24) [instantiation and parameters you would pass to the constructor.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=208.8) [By exploring the documentation and writing applications,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=213.44) [you'll get to know these options on your own.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=216.6) [When instantiating a Kafka producer with a Properties object,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=219.74) [as illustrated in the last slide,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=223.6) [you are effectively setting things up for the Kafka producer to](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=225.58) [start sending messages with the basic defaults.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=229.74) [In our case,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=232.37) [we instantiated an object of type KafkaProducer and called it](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=234.02) [myProducer and passed it a properties object named props.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=238.56) [If you look inside the implementation of the KafkaProducer,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=243.52) [you will notice a type called ProducerConfig.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=247.06) [When the KafkaProducer object is created,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=249.67) [the properties are used to instantiate an instance of the ProducerConfig class,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=253.22) [and from there, all producer configuration is defined and referenced internally.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=259.04) [It is from this object that the internal fields for key](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=265.04) [and value.serializer are initialized.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=268.48) [So, when providing these values in the Properties object,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=271.44) [you're indirectly, through the ProducerConfig,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=275.15) [setting the internal fields of the producer to expect message](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=278.16) [values for the key and value of type string.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=281.89) [This is essentially establishing a type‑safe contract between](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=285.76) [the instance of the KafkaProducer and the message](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=289.95) [specifications it is configured to produce.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=292.97) [This contract extends to the consumer who,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=296.23) [when reading messages from a topic,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=299.8) [needs to know the message specifications and its type contract,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=302.04) [which is why the configuration properties are required from the onset.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=306.11) [This is good,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=311.24) [but all we've really done here is create a](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=312.04) [KafkaProducer object with its default settings,](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=314.25) [and that's it.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=317.54) [Not very exciting, right?](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=318.64) [A producer's job is to produce.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=320.31) [What does it produce? Messages, of course. So let's get to that next.](https://app.pluralsight.com/course-player?clipId=51d4f408-e0b4-4baf-a5be-5d9c3d168452&startTime=322.16)

### [Creating and Preparing Apache Kafka Producer Records](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df)

[From the point of view of the Kafka Producer, it doesn't really send messages.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=0.14) [In fact, you won't find a single type in the entire Kafka API called message.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=5.64) [What you will find is a critical class called ProducerRecord,](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=12.54) [and it represents what will be published by the Kafka Producer.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=17.27) [A producer record is also fairly basic and straightforward, it only](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=21.94) [requires two values to be set in order for it to be considered a valid](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=27.08) [record that can be sent by the Kafka Producer.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=31.19) [These two values are the topic and the value.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=34.54) [The other optional values of partition,](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=38.64) [timestamp, and key will be covered shortly.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=41.35) [Let's take a closer look.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=44.69) [It doesn't take much to actually get messages flowing to Apache Kafka.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=46.75) [You saw how simple it was using the producer shell](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=51.44) [application, that's because the API was designed to](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=54.08) [require the bare minimum to get started.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=57.67) [This is represented not only by the simplicity of](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=60.44) [getting a Kafka Producer instantiated, but now also the producer record.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=63.06) [The first required value should be self‑explanatory at this point,](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=68.09) [it's the topic to which these record is destined.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=72.32) [The value is really just the contents of the message that are to be serialized](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=76.03) [using the specific serializer in the configuration settings. In the last module](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=80.72) [when we used the Kafka Producer Shell Program,](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=86.03) [you'll recall we just had to specify two parameters,](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=88.83) [the broker List and the topic.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=92.09) [By taking in these values,](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=94.44) [the Kafka Producer was setting its own configuration properties for](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=96.43) [Bootstrap servers and then taking the topic value to set the required topic](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=100.28) [filled in the producer record. In the Shell program,](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=105.31) [it hardcoded the default serializer to be a string serializer class.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=109.05) [The message provided in the input stream from the terminal provided the](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=114.04) [values for the Kafka Producer to send to the broker. Back to that last](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=117.53) [property, the reason it is called the value is because it must](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=121.74) [correspond to the serializer type specified in the configuration](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=125.81) [properties for the Kafka Producer instance. If you were to try and](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=130.25) [create a producer record that didn't match the serializer type](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=134.47) [specification for the producer,](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=138.33) [the producer would generate a runtime serialization exception,](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=140.58) [stating the type provided doesn't match what was expected](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=144.86) [as per the value.serializer property.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=148.38) [When you define and instantiate a Kafka Producer, you are doing](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=151.84) [so with the type of messages it will send.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=156.24) [This is established up front with the requirement of](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=159.21) [setting the key and value serializers. Kafka Producers](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=161.75) [send very specific producer records,](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=165.67) [and the type specification of the key and value must match that](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=168.04) [of the producer that is going to send it.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=172.65) [Trying to attempt otherwise will cause exceptions to be thrown](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=175.44) [and nothing will get sent to the Kafka cluster.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=179.16) [This is something to keep in mind as you're designing your](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=182.24) [Kafka Producer applications. Initially,](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=185.08) [you may think of this as a limitation, but given all of the other configuration](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=187.99) [properties that you can set on a per Kafka Producer basis,](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=192.51) [having the delineation between different Kafka Producer instances for](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=196.18) [very specific categories of messages, in other words,](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=200.63) [topics, it isn't that limiting at all, but rather a powerful ability, allowing you to have per topic flexibility at the producer level.](https://app.pluralsight.com/course-player?clipId=194cd1fd-86f4-4c99-a5a1-ed03d79c11df&startTime=203.94)

### [Apache Kafka Producer Record Properties](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098)

[I told you I would cover the optional properties,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=0.04) [so here are two of them, partition and timestamp.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=2.44) [Yes, the partition refers to a specific partition within a topic.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=6.36) [When creating a producer record,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=11.44) [you can set the value of this to a specific partition](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=13.43) [that you want messages to be sent to.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=17.44) [Doing this is an advanced scenario,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=19.78) [but an important one when it comes to how the producer](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=22.34) [decides which broker to send its messages to.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=25.37) [Hang in there for a minute and we'll get back to this.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=28.69) [The timestamp is a new addition to Kafka,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=31.75) [starting in the current .10 version. It allows for the explicit](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=35.1) [setting of a timestamp to the producer record.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=39.79) [Its presence is somewhat controversial because the timestamp is](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=42.98) [transmitted with the message. And since it is a long data type,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=47.22) [it carries with it the additional overhead of 8 bytes,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=51.38) [which can affect performance and throughput in high‑volume situations.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=55.64) [This property is nuanced, however,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=60.44) [because the actual timestamp that will be logged for a message will be](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=62.55) [based on settings defined in the broker server.properties file,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=66.71) [specifically the log.message.timestamp.type setting.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=72.38) [There are two modes available for determining which](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=77.14) [timestamp the message should have.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=79.8) [If the setting is CreateTime, which is the default,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=82.69) [the timestamp applied to the message is set by the producer](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=86.83) [and will be what is committed to the log.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=90.56) [It doesn't matter if you choose to set this timestamp explicitly.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=93.57) [Even if you ignore this setting,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=97.84) [the producer will automatically apply the timestamp to every outgoing message.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=99.89) [These alternative mode is LogAppendTime,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=105.44) [which will overwrite whatever the timestamp is coming from the](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=109.64) [producer with the timestamp of the broker at the time the message is](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=112.83) [appended to the log. From a design standpoint,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=116.86) [the mode for which to establish the message's time is not a trivial matter.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=120.67) [The ability to establish time, where and what to do with it,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=126.2) [are all very important considerations.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=130.22) [This last optional property is actually pretty important.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=133.08) [Let me define it first, and then I'll discuss why. The key is a value that,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=136.99) [if present,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=141.31) [will determine how and to which partition within a topic the Kafka](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=142.13) [producer will be sending the message to. Do you remember this slide from](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=146.5) [the last module when I taught you about producers writing to multiple](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=150.28) [partitions within a topic? You had a question I said I would answer in](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=154.18) [this module, and now is the time.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=158.71) [Well, almost.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=161.26) [Let me finish out the discussion on producer record first and show how it works](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=162.82) [with the instance of KafkaProducer to make messaging magic.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=167.67) [Even though the key attribute is optional,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=171.92) [I would urge you to avoid leaving it blank or null.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=174.14) [The key serves to very useful purposes.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=177.74) [It can be used as additional information in the message that can](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=180.53) [be used to make processing decisions later. And as we will soon](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=183.97) [see, it can strongly influence the manner in which messages are](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=187.74) [routed to the partitions.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=191.88) [However,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=193.94) [a possible downside to using a key is the payload](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=194.68) [overhead introduced when a key is added,](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=198.15) [which can depend on the type of serializer used. Once again, as with other design decisions, there are trade‑offs to be considered.](https://app.pluralsight.com/course-player?clipId=a3bc2a71-6c3b-4734-a757-47c7eac4f098&startTime=200.28)

### [The Process of Sending Messages, Part One](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d)

[So now we're back in our PowerPoint IDE,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=0.04) [and in it we are adding an object of type,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=2.71) [ProducerRecord, with the basic required attributes,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=5.69) [but with a key, and next,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=9.33) [we're calling the send method on the myProducer instance and passing the](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=12.16) [myRecord object we just instantiated as its required parameter.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=16.94) [Since the send operation can be unsuccessful,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=22.14) [it is always a good practice to wrap the call with a try..catch](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=25.94) [block and use structured exception handling.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=29.68) [I didn't illustrate it here because of space limitations,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=32.74) [but you'll see it in the demo. Now that we have a producer record for the](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=36.26) [producer to send, let's see what actually happens internally.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=40.13) [I like to look at the message sending process in two parts. For the](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=44.44) [next few slides, we'll discuss the first part.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=48.97) [When calling the send method,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=52.14) [the producer will reach out to the cluster using the](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=54.32) [bootstrap.servers list to discover the cluster membership.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=57.34) [The response comes back as metadata, containing detailed](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=61.44) [information related to the topics, their partitions and their](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=65.32) [managing brokers on the cluster.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=69.17) [This metadata is used to instantiate a metadata object in the producer](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=71.45) [and throughout the producer's lifecycle, it will keep this object fresh](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=76.57) [with the latest information about the cluster.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=80.85) [Additionally,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=83.94) [a pseudo processing pipeline within the Kafka producer is engaged. With](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=84.82) [the producer now having an actual producer record to work with, the first](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=89.95) [step in this pipeline will be to pass the message through the serializer](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=94.31) [using the configured serializer.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=98.1) [Remember in our case, we're just using the string serializer.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=101.04) [The next step in the pipeline is the partitioner,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=105.24) [whose job it is to determine what partition to send the record to. Here the](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=108.15) [producer can employ different partitioning strategies,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=113.95) [depending on the values being passed to it in the producer record, and the](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=116.94) [information it has regarding the cluster membership.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=121.66) [This is where I finally get around to answering that all important question](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=125.04) [related to how the producer distributes messages to partitions. Between the](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=129.46) [time the send operation is invoked to the time a message is received by a](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=134.39) [broker, quite a few things happen.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=139.03) [We discussed the serialization step.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=141.74) [Now is the all important partition routing step,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=144.24) [which is determined by four possible strategies.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=147.68) [First, the Kafka producer looks at the producer record contents,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=150.99) [especially the partition field.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=155.5) [It will look if there's a value provided for that partition field.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=157.94) [If it has,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=161.81) [the next question will be if the proposed partition](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=162.86) [is actually a valid partition, for example,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=165.98) [for the topic being requested,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=169.21) [is there a partition that matches the one proposed? For this](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=171.68) [answer, the producer refers to the metadata object that](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=175.91) [maintains the cluster metadata, including a list of topics,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=179.47) [their partitions and the leaders for each.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=183.49) [If the value proposed does not match a known partition for the topic,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=186.64) [or if that partition is unavailable, which is unlikely if replication is enabled,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=191.29) [then an exception will be thrown and the send operation will abort.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=197.14) [If the proposed partition is valid,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=201.38) [then the producer will add the producer record object to the](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=203.69) [specific partition buffer for the topic,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=207.27) [where it will, on a separate thread,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=209.99) [await the actual send to the broker leader of that specified partition.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=212.1) [We'll get into this buffering step as part of the second part of the](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=217.44) [message sending process, but for now let's continue.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=221.43) [If a partition was not specified in the producer record,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=225.34) [the next question to determine the routing strategy is whether a](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=229.64) [key was provided in the producer record, because,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=233.33) [as you will recall, it is an optional value.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=236.6) [If the answer is no,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=239.75) [as was the case in the last module when using the Kafka Producer Shell program,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=241.38) [the message will be routed using a round robin strategy](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=246.12) [that attempts to evenly distribute the message across all](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=249.8) [the partitions in the topic.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=253.29) [Now, technically speaking,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=254.91) [this scheme is defined in the default partitioner class we'll talk](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=256.64) [about in a few more steps. If there is a key provided, the next](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=261.09) [qualifying question is whether a custom non‑default partitioner class](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=265.88) [was provided as part of the configuration properties provided to](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=271.1) [instantiate the Kafka producer.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=275.8) [For this,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=277.84) [the producer references the producer config object and looks for a specific](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=278.75) [value called PARTITIONER\_CLASS\_CONFIG, which represents the optional](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=283.33) [partitioner.class setting provided in the properties object.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=290.22) [If there is nothing provided, which is the common default scenario,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=294.68) [the routing will be done through a key‑based partitioning](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=299.34) [scheme, which Kafka provides as a default implementation](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=303) [of the partitioner interface.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=307.39) [The default partitioner class takes a MurmurHash of the](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=309.57) [key and then applies a modulus function by the total](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=313.03) [number of partitions for the topic,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=316.73) [and that's how it determines what partition to send it to.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=318.81) [I suppose you could call that a fancy way of describing a murmur‑based mod hash.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=322.54) [Some use cases may call for a custom key‑based partitioning scheme,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=328.24) [and that's when you would need to develop your own partition or](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=332.6) [implementation, add that implementation class to the class path and specify](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=336.02) [the class type as the partitioner.class property setting.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=341.5) [If that has been done, it is that custom scheme that will be used.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=345.84) [I hope the way to answer that question was worth it and now](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=349.89) [you have a good idea as to how the producer determines what](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=353.51) [partition to direct messages to.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=357) [This knowledge is very important for designing Kafka applications,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=359.06) [which is why I was keen to spend some time on it.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=362.76) [But this is an advanced topic,](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=365.84) [and like many advanced topics that is slightly beyond the scope of this course, I would encourage further study on it.](https://app.pluralsight.com/course-player?clipId=422c0fce-a7de-47f9-b892-e2ef1737526d&startTime=367.74)

### [The Process of Sending Messages, Part Two](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf)

[Now back to our map.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=0.34) [We left off with the partitioner.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=1.87) [In our walk‑through example, we didn't specify a partition,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=4.37) [but we did provide a key, and therefore,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=8.25) [according to the routing strategy flow we just covered,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=11.03) [the key‑based partitioning scheme will be used,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=14.32) [which again, is defined in the default partitioner class.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=16.84) [This officially brings us to the second and final part of describing the](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=21.4) [message sending process inside of the Kafka producer.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=25.34) [With the partitioning scheme established,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=29.39) [the producer can now dispatch the producer record onto an in‑memory](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=31.71) [queue‑like data structure called a RecordAccumulator.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=35.99) [The RecordAccumulator is a fairly low level object that has a lot of complexity.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=40.82) [We will not go into it in a tremendous amount of](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=46.29) [detail during this introductory course,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=48.77) [but I will describe at a high level what it does and why.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=51.36) [But first, let's talk about efficiency.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=55.44) [Each time you send,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=58.07) [persist, or read a message, resource overhead is](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=59.47) [incurred. In high throughput systems, this overhead can](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=62.94) [dramatically impact the performance,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=66.97) [reliability, and overall throughput of the system. And the more that](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=69.46) [overhead is incurred on handling fewer units of work,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=74.29) [the less efficient that system is.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=78.13) [Think of it this way, suppose you have a garage full of boxes and you](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=80.74) [need to move all of those boxes to a new destination.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=84.91) [If the goal is efficiency, as far as how much you can get moved](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=88.34) [using the least amount of resources, like time and energy, what](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=92.22) [type of vehicle would you choose?](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=96.42) [Would it be a four passenger car, or would it be a moving truck?](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=98.54) [Overall,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=103.14) [the answer would be a moving truck because you can transport more at once.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=104.24) [Assuming you have an equal number of loaders and unloaders,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=109.17) [you'll likely consume less time and energy with the truck because the](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=112.46) [smaller vehicle will need to make more trips. Thus likely incurring](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=116.97) [higher overall costs of time and energy.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=121.09) [Of course, this metaphor can get out of hand,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=124.54) [but I hope the point is illustrated nonetheless.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=127.22) [This is Kafka's approach to addressing common](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=130.08) [inefficiencies in messaging systems,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=133.11) [micro‑batching. Whether it be on the producing side,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=135.62) [the broker side, or the consumer side.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=138.99) [Apache Kafka was designed with the means of being able to](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=142.26) [rapidly queue, or batch up requests, to send,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=146.04) [persist, or read in flexibly bound memory buffers that can take](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=149.86) [advantage of modern day operating system functions, such as](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=154.98) [Pagecache and the Linux sendfile() system call. By batching, the](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=158.58) [cost overhead of transmission,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=164.22) [flushing to disk, or doing a network fetch is amortized over the entire batch.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=166.29) [The RecordAccumulator gives the producer its ability to micro‑batch records](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=172.39) [intended to be sent at high volumes and high frequencies.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=177.35) [When a producer record has been assigned to a partition through the partitioner,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=180.97) [it will get handed over to a RecordAccumulator,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=185.64) [where it will be added to a collection of record batch objects for each](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=188.67) [topic partition combination needed by the producer instance.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=193.58) [Each of these RecordBatch objects, as the name suggests,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=197.6) [is a small batch of records that is going to be sent to the](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=202.24) [broker that owns the assigned partition.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=205.51) [There are a lot of factors that determine how many producer](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=208.42) [records are to be accumulated and buffered into a RecordBatch](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=211.8) [before it is sent off to the brokers.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=215.92) [Most of these factors are based on advanced configuration settings to](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=217.89) [find at the producer level that error set using a properties object,](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=222.32) [similar to the way the other properties were set. Let's take a look at a few of the important settings.](https://app.pluralsight.com/course-player?clipId=8b9d0019-4c26-4959-99b2-9e0e78ffebdf&startTime=226.38)

### [Message Buffering and Micro-batching](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa)

[Each RecordBatch has a limit of how many ProducerRecords can be buffered.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=0.56) [This limit is not based on the number of records,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=5.77) [but rather by a configuration setting named batch.size,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=9.03) [whose value represents the maximum number of bytes that](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=13.29) [can be buffered per each RecordBatch.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=17.2) [Furthermore, across all buffers,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=20.54) [there is a configuration setting that establishes a ceiling or](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=23.12) [threshold for how much memory can be used to buffer records](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=26.76) [waiting to be sent to the brokers.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=30.43) [This setting is called buffer.memory, and like batch.size,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=32.84) [its value represents the number of bytes.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=37.32) [If the high volume of records being buffered reaches the](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=40.54) [threshold established by the buffer.memory setting,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=44.26) [the max.block.ms setting comes into effect.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=47.23) [This setting determines how many milliseconds the](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=51.93) [send method will be blocked for.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=55.02) [This blocking contingency is intended to force back](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=57.47) [pressure on the thread the producer is using to send more](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=61.5) [ProducerRecords onto the buffer.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=64.7) [The hope is that within the provided number of milliseconds,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=67.14) [the buffered contents will be transmitted and free up more buffer](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=70.89) [memory to enable more records to be enqueued.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=74.67) [When records get sent to a RecordBatch,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=77.65) [they will wait around until one of two things happen.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=80.54) [First,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=84.24) [if record accumulation occurs and the total buffer size](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=85.13) [reaches the per buffer batch size limit,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=89) [the records are sent immediately in a batch.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=92.21) [This optimizes the overhead associated with transferring the](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=95.2) [page cache bytes to the network socket.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=98.91) [This is the micro‑batching intention at its best.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=101.39) [Simultaneous to this,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=106.04) [new records are being dispatched to other accumulators and other record buffers.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=107.7) [The second threshold that determines when buffered messages are](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=113.84) [sent is a configuration setting called linger.ms,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=118.03) [which represents the number of milliseconds an unfull buffer should](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=122.84) [wait before transmitting whatever records are waiting.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=126.95) [For example, if in one buffer,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=130.25) [there is a single record waiting to be transmitted rather than](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=132.47) [to incur the overhead for a single message,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=136.33) [the linger.ms setting will wait around for the specified number of](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=139.23) [milliseconds to pass before the actual transmission.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=143.05) [For high‑frequency partitions whose buffers are being filled rapidly,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=147.04) [the linger.ms setting generally does not come into play.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=151.44) [We covered a lot of details and complexity with regard to](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=155.06) [the various configuration settings and how they can be set](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=158.84) [to affect the producer behavior,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=162.3) [which will have a big influence on the overall performance of the system.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=164.24) [Again, I don't expect you to come away from this an expert.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=168.44) [In fact, after your head stops spinning,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=172.07) [I would once again encourage further study and experimentation](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=174.42) [on this subject as it is an advanced topic.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=178.19) [Finally,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=181.72) [the last part of the message‑sending flow is when the batched](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=182.45) [records finally get transmitted to the brokers,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=186.04) [and the result of the transmission is sent back as a RecordMetadata object,](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=188.94) [which essentially contains information about the records that were successfully or unsuccessfully received.](https://app.pluralsight.com/course-player?clipId=6c578f77-3f19-4083-aafd-f6f515fcd5fa&startTime=194.05)

### [Message Delivery and Ordering Guarantees](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746)

[To ensure the best chance of delivery,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=0.04) [there are some additional settings that should be considered,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=2.32) [which are set at the producer level.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=5.56) [We'll cover some of these here.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=7.6) [First, when sending messages,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=9.78) [the producer can specify what level of acknowledgement](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=11.75) [it expects from the receiving broker.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=15.75) [This is a setting appropriately named acks and can be set using the](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=18.73) [property‑setting method discussed at the beginning.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=23.87) [The first and most risky option is setting the acks value to an integer of 0.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=26.67) [This essentially represents a fire‑and‑forget mode of sending messages](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=33.34) [because no acknowledgement whatsoever is sent by the broker.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=37.91) [This approach is definitely the fastest in terms of request latency,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=41.91) [but not very reliable,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=46.33) [especially if there's an issue with a broker that](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=48.31) [prevents it from logging the message.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=50.49) [The producing application has really no way of knowing if the message got there.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=52.69) [Now this may be okay if the type of messages being](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=57.84) [sent with this mode can be lossy, such as possible clickstream data.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=61.32) [The second middle‑of‑the‑road option is setting the property value to 1.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=66.41) [With this,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=71.26) [the producer is only asking for the leader broker to confirm](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=72.32) [receipt and persistence instead of waiting for all replica](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=75.75) [members in the quorum to confirm. This option offers a good](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=79.17) [balance of performance and reliability,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=83.72) [providing the cluster settings employ appropriate replication.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=86.23) [The third and final is when the property is set to 2 and thereby](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=90.23) [requesting from the cluster that all in‑sync replicas confirmed the](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=95.7) [receipt before counting the message as successfully sent.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=100.38) [Obviously,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=104.14) [this option offers the highest level of assurance that the](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=105.12) [message was successfully sent and received,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=108.41) [but at the cost of performance,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=110.77) [which can be unpredictable based on the possible changes in the](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=112.79) [cluster membership and thus replication topology.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=116.7) [When any error is sent back, the producer needs to decide what to do with it.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=120.15) [The first line of defense is to employ the retries configuration setting,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=125.34) [which controls how many times a producer will, you guessed it, retry](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=130.24) [to send the message before aborting. Closely associated with the retry](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=134.81) [setting is the retry.backoff.ms setting,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=139.13) [which allows you to specify the wait period in milliseconds](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=142.74) [between retries. Depending on your application, message](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=146.47) [ordering can be important. If it is,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=150.05) [these points are important to consider. I made a brief](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=153.28) [mention in the last module that message order is only](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=156.45) [preserved within a given partition.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=159.87) [If the producer sends messages to a partition in a specific order,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=162.44) [that order will be the order in which the broker appends](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=167.17) [them to the log and it will be the order that the consumers](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=170.61) [will read them from the log.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=173.7) [Messages sent to multiple partitions, however, will not have a global order.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=175.43) [Now this should be expected and understood at this point given](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=180.24) [our discussion about partitioning strategies. But to derive a](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=183.88) [global order across partitions,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=187.41) [the order logic will have to be handled at the consumer level or even beyond.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=189.67) [Regardless of the ordering assurances at the partition level,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=194.68) [errors can complicate matters for expected reasons.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=198.5) [If the configuration setting retries is enabled, and the retry.backoff.ms](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=201.94) [setting is set too low, you may have a situation where the first message is](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=207.33) [sent and a success acknowledgement is not received,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=212.16) [causing a retry to happen.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=215.81) [But before the retry can be sent,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=218.04) [the second message is sent and successfully received while the retry](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=220.22) [first is sent and ultimately acknowledged. Now the result would be a](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=224.68) [reverse order within a single partition.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=229.94) [The only way to avoid this,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=232.75) [but at a high cost of throughput, would be to set the](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=234.28) [max.in.flight.request.per.connection setting to 1,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=237.8) [which would effectively tell the producer that at any given](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=241.84) [time, only one request could be made.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=244.94) [Ouch. But that may be what is needed.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=248.01) [A combination of these settings will determine the message](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=251.64) [delivery semantics required by your system.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=255.31) [It is possible to achieve either an at‑least‑once,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=259.24) [an at‑most‑once, and an only‑once message delivery assurance,](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=263.15) [but only with the design that carefully considers the settings available at](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=268.74) [all three component members of the system, the producer, the broker, and the consumer.](https://app.pluralsight.com/course-player?clipId=dd491cbc-3105-431d-aae4-82c0539e4746&startTime=273.46)

### [Demo: Creating and Running an Apache Kafka Producer Application in Java](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6)

[It's the long‑awaited demo time where I will endeavor to show you how you can](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=0.94) [start to build a custom Kafka producer application in Java.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=5.58) [In this,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=10.14) [I will cover some of the highlights of what we've](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=10.87) [been discussing in this module.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=13.4) [But overall,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=15.37) [the scenario will closely resemble that of using](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=16.2) [the Kafka producer shell program.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=18.85) [We will use a basic producer configuration against a cluster setup,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=21.54) [consisting of a topic with three partitions,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=26.2) [three member nodes, and ensured with a replication factor of 3.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=29.22) [In this, look for evidence of the default partitioner being used.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=33.95) [This will be seen when the single consumer that we use reads from the topic.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=38.64) [You'll notice that there will not be a global order.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=43.5) [Okay, so in this demo,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=47.24) [I've got the details of our topics up here. As you can see, we have our topic,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=48.94) [my‑topic, with a partition count of 3, replication factor of 3,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=58.14) [as we said. And you'll notice that each partition has its](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=62.89) [own line. And because there's three nodes,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=67.02) [each one is a leader for a partition. And over here, we've got](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=70.41) [a terminal window listening for messages.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=74.55) [Additionally, within the IDE, we have the terminal,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=77.91) [which is also running a consumer waiting for messages to arrive.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=80.84) [Okay, so let's minimize that and look at the code.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=86.54) [Alright, so in the code window, you'll see that we have Kafka producer.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=89) [But before that, we have our properties,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=96.34) [and we're instantiating some of those important required properties,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=98.67) [just like we saw in the slides.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=102.6) [But here, we have a Kafka producer instantiation,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=104.81) [and you're going to notice that the signature is a little bit different.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=109.74) [The signature,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=114.24) [basically, is indicating that it wants strings for keys and values in the](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=114.94) [signature itself, and that's one of those optional instantiations that I](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=119.76) [suggested you could do earlier in the slides moving beyond the Kafka producer](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=124.36) [instantiation. Here within a try, catch, finally block, we have basically a](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=129.22) [bounded loop where we take my producer instance,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=134.88) [and we actually call the send method, passing in a new producer record.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=137.63) [You'll also notice that again its signature is specifying the key and](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=143.36) [the value serializer upfront. And in this case,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=147.73) [it's passing the my‑topic and passing in a key and value.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=150.76) [Both of them are strings as the signature suggests. The key](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=155.45) [is being derived from within the loop as just an integer](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=159.39) [that we're casting to a string.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=162.76) [And then my message with the integer represents the](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=164.65) [message and the number of that message.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=168.33) [All right, so let's look at the catch.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=171.84) [We're just doing a standard catch, catching for an exception](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=174.51) [in case of an error occurs in the send.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=177.24) [And then we have our finally block here. And the finally block we didn't](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=180.35) [talk much about in the slides, but we should have.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=184.17) [Really, it's the opportunity to gracefully close down the producer.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=186.35) [If you do not do that, like other types of resource‑intensive,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=190.34) [network‑aware code,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=194.62) [you can cause memory leaks, and it can cause all sorts of problems.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=196.51) [So it's just a general good idea to make sure you're closing the producer](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=199.76) [gracefully as to avoid any sort of leaks whatsoever.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=204.56) [All right, now let's actually run our application.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=208.34) [Let's run the producer itself and see what happens.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=211.69) [Okay, so it's compiling, it's running,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=215.04) [and as you can see in this terminal window over here, and](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=217.38) [let me make that a little bit bigger,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=220.18) [you can see all of the messages. Now they're not in order, and the reason](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=221.7) [why is because the partitioner is taking the rights from the producer and](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=226.33) [spreading them across three different partitions,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=231.22) [and it's doing it in a different order.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=233.94) [It's not exactly even. And you can verify that in this terminal window as well,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=235.74) [which was also a consumer, and you can see that it is definitely not in order.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=240.61) [But I hope that this illustrated what we intended to do, and that is](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=246.04) [to show a very simple Java‑based producer that is producing messages](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=249.62) [and using the default partitioner,](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=255.44) [which is that key‑based partition because we did provide a key, and that it's doing so across multiple partitions.](https://app.pluralsight.com/course-player?clipId=c97993f4-0a06-4f74-9a38-c52137a24ff6&startTime=257.83)

### [Advanced Topics and Module 4 Summary](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657)

[We covered a lot,](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=0.04) [but there are some things we just weren't able to](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=1.43) [cover in this introductory course.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=3.71) [But I wanted to give you a highlight of what they are so you](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=5.53) [can use them as topics to explore more.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=7.92) [Custom serializers, why and how to create customer serializers.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=10.19) [Customs partitioners, why and how to create custom partitioning schemes.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=14.55) [There's options to send messages using an asynchronous callback and a future.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=19.34) [It would have been nice to show you these,](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=24.74) [but I think it's something that you can explore on your own.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=26.56) [Applying compression options,](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=29.57) [which also falls into the category of advanced settings and](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=31.26) [combinations for optimal throughput and performance.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=34.65) [Throughout this module, we focused on the internals of a Kafka producer.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=39.24) [We started with the high‑level map and started to](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=43.67) [drill down component by component and, in the process,](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=46.46) [covered properties and how they are represented as ProducerConfig objects.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=49.6) [How we think of a message is really an instance of a producer record class.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=55.04) [We discussed the processing pipeline when sending a message using the](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=60.74) [KafkaProducer class and how the producer record goes through a](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=65.15) [serialization step and a partition assignment process.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=69.18) [We spent some time talking about how Kafka optimizes](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=73.34) [message throughput through microbatch ing,](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=77.16) [and we walked through the internals of the Kafka producer with the](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=79.89) [record accumulator and record buffer as the means it can accomplish](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=83.61) [micro batching with related configuration settings.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=87.92) [We touched on message delivery and ordering guarantees offered](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=91.21) [by Kafka and some relevant configuration properties to consider](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=95.24) [when designing your applications.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=98.98) [Finally, we ended with a brief demo on a basic Java‑based producer.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=100.77) [Next, we'll cover the other type of client application, the Kafka consumer.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=105.57) [I'll most likely take the same approach as I did with](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=110) [the Kafka producer in this module.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=113.41) [So I hope you enjoyed it and learned enough to be anxious to start exploring on your own and continuing in this course to learn more.](https://app.pluralsight.com/course-player?clipId=15c393bc-a02a-4248-add5-08fa75fed657&startTime=115.18)

## [Consuming Messages with Kafka Consumers and Consumer Groups](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432)

### [Introduction and Apache Kafka Consumer Overview](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432)

[We've arrived at the third and final major component of Apache Kafka,](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=0.04) [the consumer.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=4.89) [As consumers share a lot in common with their producer sibling,](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=6.24) [we will cover consumers with a very similar approach.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=10.2) [We will take a look at the internals and use that as a basis for](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=13.61) [understanding how consumers work and behave, while throughout,](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=16.83) [we'll look at samples and demos to help you understand how to](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=20.24) [create your own Kafka consumers.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=23.49) [Finally, towards the end,](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=25.54) [we will cover consumer groups and how they enable Apache Kafka](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=27.13) [to scale on the message consumption side.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=30.94) [Up to this point, when we've discussed the consumer,](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=34.14) [it has mostly been from an external superficial perspective related to](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=37.29) [its role in relationship to other Kafka components.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=41.27) [In this module,](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=44.84) [we're going to look deep into the consumer internals. Similar to how we](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=45.83) [started with the producer, we'll use this diagram as a map to help](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=49.93) [illustrate and guide for the remainder of the course.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=54.31) [Don't try to understand everything you see here now, because we'll](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=57.44) [cover each component in detail in its due time.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=60.51) [This is almost exactly the same as the producer,](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=64.35) [because the function these initial required property serve is nearly identical.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=67) [Remember, in the consumer's world,](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=72.24) [you are reading messages and therefore need to specify a](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=74.38) [deserializer class for both the key and value.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=77.69) [This of course,](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=81.64) [must correspond to the type specification of the producer message that](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=82.54) [was serialized with the same serializer classes.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=86.48) [As with the producer, there are a lot of configuration settings available.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=89.66) [I encourage you to use the link above and familiarize yourself with the](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=93.94) [various consumer configuration settings and options.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=97.68) [Of course, we'll be covering the most relevant ones in this module.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=100.84) [When it comes to creating a consumer,](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=104.74) [you'll once again notice the similarities with creating a producer.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=106.99) [This is by no coincidence,](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=110.74) [as the designers of the Kafka client APIs wanted to make](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=112.14) [working with Kafka familiar and productive.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=115.71) [Here, we are using a standard main entry point because we'll be using a](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=118.44) [console‑based Java application in our consumer samples.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=122.4) [Of course,](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=126.24) [the difference here is that you're creating an instance of a Kafka consumer,](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=126.97) [but passing the required configuration is needed.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=131.42) [With the Kafka consumer object in place, we now need to give it something to subscribe to so it can start working.](https://app.pluralsight.com/course-player?clipId=5309a4c7-5999-4570-a366-562c0de17432&startTime=135.04)

### [Subscribing and Unsubscribing to Topics](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660)

[Subscribing to topics programmatically is extremely](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=0.14) [easy and somewhat straightforward.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=3.23) [To subscribe to a topic, you call a specific method.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=5.64) [Want to take a guess what it's named? That's right.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=9.21) [The subscribe method.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=12.26) [The method signature for subscribe takes in a collection of strings,](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=14.24) [which represent a list of topics.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=17.98) [A single consumer can subscribe to any number of topics](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=20.84) [from one to theoretically infinity.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=24.55) [Seriously,](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=27.34) [I don't know if there is a limit to the number of](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=28.07) [topics that can be subscribed to.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=30.11) [In this case, we're just using a list of one.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=31.99) [Alternatively,](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=34.65) [you could subscribe by passing a regular expression as the parameter,](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=35.94) [which is a useful overload for the subscribe method.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=39.84) [There is a noteworthy nuance to the subscribe method,](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=43.04) [and it is evident when you want to add another topic to the subscription list.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=46.07) [Let's look at this code example.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=50.23) [Initially, we subscribe to my‑topic by passing the subscribe method,](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=52.25) [a single item in a list of strings.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=56.82) [Now we want to add another new topic using the same technique.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=59.54) [By doing it this way,](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=63.84) [you would think that it would work and that you would now](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=64.99) [have two topics to subscribe to, right? Wrong! Calls to](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=67.5) [subscribe are not incremental,](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=72.27) [meaning that any subsequent call to subscribe will](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=74.96) [overwrite whatever it had in there before.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=78.31) [Therefore,](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=80.89) [the best approach would be to maintain the topics of](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=81.78) [interest into a separate structure,](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=84.75) [manage them there, and pass in the reference to the topic list like this.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=87.07) [While we're here, we may as well cover the opposite of subscribe,](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=94.54) [and that is unsubscribe.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=97.77) [It's as basic as it gets. Notice, there isn't a parameter list here.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=99.4) [This suggests that you don't really unsubscribe from individual topics.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=103.89) [You're basically unsubscribing from all topics.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=108.05) [Here's another option for unsubscribing. You just pass an empty list to the subscribe method.](https://app.pluralsight.com/course-player?clipId=4c7a6e94-9621-414e-a59e-a51de6230660&startTime=111.44)

### [Comparing Subscribe and Assign APIs](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74)

[There are some important points about Kafka consumers that I want to teach](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=0.13) [you at this point. We just covered the basics of creating a list of topics](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=3.91) [we want a single consumer to subscribe to.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=8.46) [It's as simple is calling the subscribe() method and passing it](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=11.8) [a list. By calling this method, you are asking for automatic or](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=15.08) [dynamic partition assignment.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=19.84) [That is to say that you're enlisting the single consumer instance to](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=22.34) [eventually pull from every partition within that topic,](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=26.54) [which can be at least one, but likely many. When](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=30.31) [adding multiple topics to the list,](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=34.02) [you're enlisting the consumer instance to pull from](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=36.36) [every partition within every topic, which is guaranteed to be many.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=39.56) [This has very important implications,](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=44.92) [which shouldn't be taken lightly for reasons we'll cover shortly.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=47.24) [Besides subscribing to topics,](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=51.34) [there's another option, subscribing to individual partitions.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=53.53) [This is done through the assign() method.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=57.98) [The assign() method is only valid for subscribing to a list](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=60.42) [containing the class topic partition, as we'll see next.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=64.02) [But first,](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=68.24) [let me explain a key difference between subscribe() and assign() methods.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=69.07) [There's a reason why the API designers decided to call this operation](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=73.45) [out as a separate method as opposed to overloading the subscribe](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=77.79) [method. By asking for specific partitions,](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=81.84) [you're basically taking on the responsibility of assigning](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=84.71) [yourself specific partitions. More specifically,](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=88.41) [assigning specific partitions to a single consumer instance.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=91.99) [Once you've assigned yourself a list of partitions,](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=96.74) [the consumer will then start pulling these individual partitions,](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=99.53) [regardless of the topic those partitions are part of. Both have](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=102.94) [one thing in common, they take lists,](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=107.15) [and they cannot be added to incrementally, as we covered earlier. As we'll](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=109.88) [cover later with consumer groups, this assignment responsibility is](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=114.91) [generally managed for you and for a good reason.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=119.39) [I suppose you can say using the assign() method is an advanced case and](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=122.34) [therefore needs to be treated with respect for reasons we'll discuss later.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=127.35) [To manually assign partitions to a Kafka consumer,](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=131.44) [you will first need to create a list containing](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=134.77) [your manual partition assignments.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=137.18) [This is done by instantiating topic partition classes with the](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=139.71) [appropriate information. You'll see the topic partition class](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=143.68) [referenced in a lot of places in the Kafka API.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=147.8) [It simply provides a type safe data structure to represent](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=151.44) [individual partitions within a topic.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=154.78) [Second,](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=158.14) [you're going to invoke the assign() method, passing at the](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=158.77) [list of topic partitions just created.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=161.8) [That's about it. I'm sure you'll notice the general similarities with the subscribe() method.](https://app.pluralsight.com/course-player?clipId=d2d976ce-b438-4822-b9c2-163901c70e74&startTime=164.9)

### [Single Consumer Subscriptions and Assignments](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a)

[This is a good time to teach you about how individual Kafka consumer](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=0.24) [instances interact with their subscribe topics.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=4.18) [When a single consumer subscribes to a topic using the subscribe method,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=8.04) [it will constantly pull any and all partitions within](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=13.12) [the topic for new messages to consume.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=16.46) [This is the case for all of the topics for which the consumer is](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=19.74) [subscribed. Depending on the number of topics and the number of](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=23.19) [partitions within each of those topics,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=27.53) [that could be a lot of message polling by a single consumer.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=30.12) [We will discuss the challenges of this approach soon,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=34.64) [but I wanted to teach you how this works first.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=37.72) [The benefit to using the subscribe method to retrieve data is that](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=40.54) [partition management is entirely managed for you.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=44.61) [For example, suppose there is a new partition added to an existing topic,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=48.54) [presumably because the administrators wanted to](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=53.94) [increase the scalability of the topic.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=56.31) [When that happens, the metadata about the cluster will have changed,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=59.04) [and it will be sent to the consumer.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=62.83) [Since the consumer maintains an internal object that manages](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=65.24) [its subscriptions, called SubscriptionState,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=68.91) [it will know if the change has affected its subscriptions.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=72.26) [In this case,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=76.34) [it has, so it will know to automatically add the](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=77.16) [new partition to the topic list,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=80.42) [which the consumer will start polling for messages.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=82.43) [We'll cover how this happens towards the end of this module.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=85.74) [Pretty convenient, isn't it?](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=89.04) [Just remember, this capability is only available through the subscribe method.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=90.84) [In slight contrast,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=95.24) [a single consumer instance may want complete control over what](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=96.86) [partitions it wants to poll messages from.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=101.21) [There are legitimate reasons this may be called for,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=104.08) [but they are advanced use cases that we won't have a](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=107.23) [lot of time to cover in this course.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=109.66) [By specifying a list of topic partition objects,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=112.24) [the consumer is assigning itself to specific partitions.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=116.17) [This is a lot like hard coding a list of specific partition IDs in a watch list.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=119.94) [At this point,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=126.04) [the fact that a partition participates as part of a topic is](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=127.15) [less relevant, because as far as the consumer is concerned,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=131.45) [it doesn't know or really care.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=135.3) [It knows what topic each partition is in,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=137.64) [but it doesn't really do anything with that information](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=140.76) [once it has assigned itself a partition.](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=143.21) [So if a partition is added to a topic,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=146.24) [the consumer instance may be notified of it, as per the](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=149.24) [protocol of retrieving metadata from the cluster,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=152.45) [but it doesn't really care, and why would it? If it had good](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=155.64) [reason to assign itself specific partitions, why would it](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=159.04) [care what happens in the topic?](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=162.47) [If it was interested in what happens at the topic level,](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=164.84) [it would have used the subscribe method instead, right?](https://app.pluralsight.com/course-player?clipId=b0b45042-c177-438e-a191-b210e68c796a&startTime=168.25)

### [The Poll Loop](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a)

[We've talked about subscribing to topics and assigning partitions,](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=0.24) [which is really important to understand how the Kafka consumer works.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=4.1) [You may have noticed that several times in that](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=8.84) [discourse, I used the term poll or polling.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=10.92) [This was intentional because now it is the time to understand](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=15.01) [what polling means within the consumer context.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=18.56) [It would be natural to think that by invoking the subscribe or](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=21.95) [assign methods that we just talked about,](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=25.39) [you would be actually kicking off the consumer to start receiving messages.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=27.94) [That's not how it works.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=32.23) [Nothing happens until you start the most critical](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=34.04) [process in the entire consumer component,](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=37.02) [which is the poll loop.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=39.84) [The poll loop is the heart and soul of the Kafka consumer,](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=41.74) [and it is what enables the consumer to realize its purpose,](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=45.29) [and that is to continuously and reliably poll the brokers](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=48.66) [in the cluster for messages.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=52.46) [It's a single and simple method,](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=54.74) [but don't let that fool you. From that simple method, all of](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=56.9) [the complex interactions between the consumer and the broker](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=60.76) [are kicked off and coordinated.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=64) [We'll get to the details of this soon,](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=66.54) [but you will see how it goes way beyond just receiving messages.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=68.73) [Let's take a look how to start the polling process.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=72.82) [First of all, the pole loop can't be a loop without a loop to run it in.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=75.47) [As funny as that sounds, it's true.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=79.96) [A Kafka consumer is a long‑running application, or at least it](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=81.85) [should be, whose job is to always be looking for new messages and](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=85.62) [process them from the Kafka cluster.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=90.29) [There should be very few reasons you would stop polling once you've started.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=92.51) [So think of the loop as an infinite loop that we will](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=97.08) [only be interrupting for valid reasons, whether intentional or unintentional.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=100.34) [You start the loop by calling the poll method on the Kafka consumer object,](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=105.22) [passing it a long typed number representing a very](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=110.04) [important value that we'll cover shortly.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=113.4) [You'll notice the direct output of the poll method is to](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=115.96) [return an object of type ConsumerRecords,](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=119.27) [which contains any records the consumer was able to retrieve from the broker.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=122.42) [From there, what you do with the ConsumerRecords is entirely up to you,](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=127.64) [and this is where the diversity of Kafka consumer applications come in.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=132.34) [Otherwise, they're fairly generic as far as what it takes to](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=136.24) [get it polling for messages. You'll notice that I enclosed the](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=139.16) [call to poll within a try block.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=142.65) [Since the Kafka consumer is idle until the poll method is invoked,](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=144.79) [there really isn't anything that can throw a runtime exception to be handled.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=149) [But because the invocation of the poll methods starts any](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=153.09) [and all network activity with the cluster,](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=156.84) [it is a good idea to enclose it with the means to do](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=159.24) [structured exception handling.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=161.93) [As the poll operation opens network resources, it is always a good idea to make sure it closes in the end.](https://app.pluralsight.com/course-player?clipId=62ed39ea-f80d-4c9f-a934-ce762519851a&startTime=163.84)

### [Demo: Simple Kafka Consumer](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d)

[Let's go through a demo of a custom consumer application written in Java.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=0.94) [The development environment will be the same as before.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=6.33) [The only difference is that we'll have a consumer app class](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=9.64) [to run the consumer console application.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=12.77) [The demo cluster configuration will consist of a single broker with two topics,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=15.9) [each with three partitions.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=20.96) [We'll only do a single replication factor for this demo.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=23) [Look for the use of the handy kafka‑producer‑perf‑test](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=26.84) [shell program to generate messages.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=31) [Even though we're not planning on blasting our Kafka](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=33.56) [environment for performance tests,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=36.69) [I want to use this as an opportunity to show you how to use this tool.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=38.33) [We'll demonstrate two different consumers,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=42.72) [one using the subscribe method for retrieving](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=45.64) [messages and the other using assign.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=48.16) [We'll observe the output from each consumer and note the differences,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=50.74) [and then we'll add a new partition to a topic.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=54.46) [With this configuration, we'll look at the output of both consumers.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=57.94) [You'll notice the differences between the assign](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=62.54) [consumer and subscribe consumer.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=64.9) [I've already started a single Kafka broker and created](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=67.36) [two topics with three partitions each,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=70.62) [and you can see this with the results of issuing a](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=73.25) [describe command against the cluster.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=75.75) [Here's the topic, my‑other‑topic, again,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=77.84) [PartitionCount of 3, single ReplicationFactor,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=80.59) [and there they are with their leaders.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=83.43) [And then we have the second topic, my‑topic, with the same configuration.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=85.89) [Before we do anything,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=90.74) [let's take a look at our sample KafkaConsumer Java](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=91.98) [applications that we've got so far.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=95.08) [So let's open up our IDE here.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=97.16) [We have two applications actually. We have one for subscribing to topics and](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=98.9) [the other for getting specific partitions assigned to it.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=105) [As we go through this,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=109.11) [you'll notice that our code is aligned with what we've](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=110.08) [covered in the slides up to this point.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=112.46) [For example,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=114.84) [here we've established the required properties for the consumer and](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=115.64) [passed them into an instance of the KafkaConsumer class,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=120.05) [thus creating myConsumer object.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=124.93) [And here, for the subscribe consumer,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=127.33) [we need to create a list of topics that we're interested in watching.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=130.49) [These topics are simply my‑topic and my‑other‑topic, as you saw in the terminal.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=134.74) [And to add these topics to the consumer,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=140.33) [we simply invoke the subscribe method here,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=142.8) [passing in the list of topics.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=145.82) [Next is where the action happens,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=149.26) [starting responsibly with the try block and setting up a loop for](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=150.93) [which we can enter into the consumer poll loop.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=155.73) [Here you'll see that the poll method has been set with a timeout value of 10 ms.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=158.69) [Now, we'll get back to what this parameter means shortly.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=164.57) [And within this loop,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=167.47) [we're going to be taking each record that we get from the poll method,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=169.44) [and we're going to be iterating over it and processing it minimally.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=173.91) [In this case,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=178.55) [we're just taking the values that are present and formatting](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=179.18) [them and outputting them to the console.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=183.39) [And finally, you'll see that we literally have a finally block,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=185.61) [so that when it exits we can responsibly close the consumer](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=190.57) [and free up the resources we need to.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=196.02) [The assign consumer is virtually identical.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=198.27) [The only difference is how we construct a list of specific topics](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=200.83) [that we want to assign to our particular consumer,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=205.31) [and here you'll see those.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=208.35) [We have to create specific TopicPartition objects,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=211.36) [one representing each of the specific partitions within a](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=216.62) [specific topic that we want to assign.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=220.81) [Here we are going to have to, we're going to have the first partition,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=223.2) [Partition 0 from the my‑topic partition, excuse me,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=227.16) [my‑topic topic, and we will have another representing the second partition,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=231.44) [or I should say the third partition from my‑other‑topic.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=238.33) [And we will be adding them to our list of partitions,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=242.29) [and instead of calling the subscribe method,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=247.24) [obviously we're invoking the assign method and](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=250.22) [passing in that list of partitions.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=253.56) [Everything else, as far as how we're going to be processing the](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=255.33) [records that are retrieved from those partitions,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=259.05) [is identical.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=261.84) [Let's go ahead and run these two consumer programs so that](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=263.04) [they can be listening for messages to arrive in their](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=266.54) [respective subscriptions or assignments,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=269.23) [and then after that, we'll start producing some messages.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=271.2) [So let's start with the SubscribeApp.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=274.27) [All we have to do is go in here and hit Run,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=276.6) [and we'll do the same thing for the ConsumerApp.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=278.89) [Jumping back out to a terminal,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=283.24) [I wanted to show you a handy tool for creating lots of messages.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=285.39) [It's called the kafka‑producer‑perf‑test shell program,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=289.02) [and this is roughly how you get it to work.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=292.56) [So, as it says, this tool is used to verify the producer performance,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=294.29) [so you can really just pump a bunch of messages in here.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=298.3) [In order to get it to work,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=301.21) [you have to pass in a topic. You give it the number of](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=302.31) [records that you want it to produce, the size of those records,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=304.88) [and you can also determine what the throughput you want it to be.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=309.06) [So in this case,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=313.07) [you would be setting a value to represent how many messages per second.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=313.9) [And finally, you would be passing to it a list of properties.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=317.62) [Now, at the bare minimum,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=321.31) [it would be the minimum required properties, such as](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=323.15) [bootstrap.servers, and then the, in this case,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=326.16) [since we're producing,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=329.03) [it would be the key.serializer and the value.serializer classes.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=329.92) [So if I have this other window up here, it's a little busy,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=333.44) [so you can see that I'm ready to start sending some](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=336.48) [messages to here to my‑other‑topic.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=340.42) [I've said that I want to send 50 records with just 1 byte each,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=342.79) [so very small, with a throughput of 10 per second.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=348.75) [You can't see that, it's a little cut off,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=351.83) [but it's going to be 10 per second,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=353.94) [so it'll take a total of 5 seconds to send all of these messages.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=355.52) [And, of course, I passed in the required properties,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=358.94) [the bootstrap.servers and the string key‑value serializers.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=362.14) [I'm also doing this for the other topic,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=365.93) [so both topics will have the same configuration.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=368.65) [Alright, let's run our producers.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=371.06) [Let's start off with the producer, the test producer for my‑topic,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=373.01) [and we'll run the test producer for my‑other‑topic.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=377.74) [And here you're going to see a bunch of stuff](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=381.95) [getting output to the Subscriber app.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=384.25) [And if we go over here, we'll see a lot of output to the Consumer app.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=387.1) [Okay, so let's look at the results here,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=391.72) [and then I'm going to, I want you to look for certain things.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=393.98) [Okay, so when we ran the first performance tool,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=396.32) [you'll notice that on the SubscribeApp,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=399.7) [you'll remember now that it's subscribing to all partitions by topic.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=401.81) [So, within this,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=406.53) [you'll see for my‑topic it has a mix of messages](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=407.53) [it's getting from all partitions.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=411.51) [And then we ran the other perf tool that started publishing](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=413.04) [messages to the other topic, and just even with this,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=416.84) [you'll start to see that it was looking at all of](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=420.39) [the partitions within that topic, as you would expect.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=424.01) [Now the output is just based upon different producers formatting the results](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=426.64) [differently based upon what values were in the messages.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=430.46) [Now, if we go over here to KafkaConsumer,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=433.2) [where the assign method was used, we'll see some different results.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=435.85) [For my‑topic, all we see is Partition 0,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=440.41) [because, if you remember,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=444.08) [that's all we told this consumer to look for for that topic.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=445.35) [In addition, if we see my‑other‑topic,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=449.84) [which was the other topic whose partition number 2 we assigned to the consumer,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=453.02) [that's all that it would notice, so it's doing exactly what we would expect.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=458.24) [Now, we're going to keep this open, just as is,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=462.08) [we're not going to stop it at all,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=465.33) [and what I'm going to do is go over to another terminal window and I'm going](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=467.23) [to create another partition within one of those topics,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=471.26) [and we're going to see what happens when we start producing data to those,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=474.86) [to that new partition.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=479.32) [Okay, here in this terminal window,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=480.51) [you'll notice that I'm using the kafka‑topics shell program,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=482.54) [and we've seen this in action before.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=486.44) [There's a little bit of a difference here.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=488.5) [Yes, I'm passing in the zookeeper reference,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=490.23) [but here I'm using the command to alter,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=492.45) [because I'm basically saying I want to alter this topic,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=495.34) [which is just the my‑topic topic,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=499.19) [and I'm saying I want it to have four partitions instead of three.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=501.73) [So let's run this, and as you can see it says Adding partitions succeeded.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=506.47) [Now let's go into here real quick and look at our](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=513.07) [describe just to make sure that it took.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=515.36) [Basically, what you're going to see now,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=518.31) [despite all of that whizzing by, we'll get to that later,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=519.9) [is basically you'll see that for my‑topic we now have a PartitionCount of 4,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=522.78) [so 0 through 3.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=527.5) [Okay,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=529.44) [so with that, go back to our producer test tools, and let's produce a new](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=529.99) [round of messages and see what happens in our consumers.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=535.71) [Okay, so if we go back here to our applications,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=539.27) [we more less left off where we had them,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=542.12) [and that is the SubscribeApp and the ConsumerApp,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=544.69) [they're just waiting for new messages.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=547.51) [Now, remember,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=549.14) [we added a new partition to the topic my‑topic, so in this window we are](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=549.84) [going to create more messages that go to my‑topic.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=555.06) [We don't need to add more messages to the other topic,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=558.52) [since that was not changed, so let's do that here.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=561.86) [So we're producing more messages to the](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=564.93) [KafkaConsumerSubscribeApp, and as you can see,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=567.57) [it went through and here within the my‑topic we have now an extra](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=570.93) [partitions‑worth of messages that it's looking at.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=576.04) [It has 0, 1, 2, and 3,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=578.54) [and it successfully subscribed to all of those new partitions.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=581.9) [It became aware that there was a new partition, it added it to its](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=586.11) [subscription, and then actively started listening for it without really any](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=590.02) [intervention on our part other than just creating the new topic.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=593.82) [Now, let's compare that to the ConsumerAssignApp.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=597.62) [Nothing happened. When it got more messages, it definitely received them,](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=600.48) [but it only got the messages for Partition 0. You'll notice there's nothing](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=606.43) [there from Partition 1, nothing there for Partition 2 or 3. All it knows or](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=611.92) [cares about is Partition 0; it has no knowledge that there is other](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=616.72) [partitions other than this. So that hopefully illustrates the differences a](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=620.84) [bit between the assign and the subscribe methods when polling for messages in a consumer.](https://app.pluralsight.com/course-player?clipId=1a68ae6f-6611-4b21-8ce1-415de6a9ef4d&startTime=625.08)

### [Walkthrough: Consumer Polling](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963)

[When the subscriber assign method is invoked,](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=0.14) [the content of the collections they were passed to are used to](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=3.1) [set fields within the SubscriptionState object.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=6.73) [This object serves as the source of truth for any and all details](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=11.01) [related to the topics and partitions this consumer instance is](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=15.7) [subscribed or assigned to. A lot of what happens within the consumer](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=19.27) [invariably crosses paths with this object.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=24.64) [This object also plays a very important role with the](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=28.34) [consumer coordinator in managing the offsets,](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=32.26) [a topic we covered briefly in module three and we'll spend](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=35.52) [a bit more time on it later in the module. When poll is](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=39.24) [invoked, consumer settings,](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=42.32) [particularly those referring to the bootstrap servers,](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=44.18) [is used to request the metadata about the cluster.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=47.54) [This kicks off everything within the consumer.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=51.54) [The fetcher serves as the responsible object for most of the](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=54.78) [communication between the consumer and the cluster.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=59.2) [Within it, there are several fetch‑related operations that are](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=62.94) [executed to initiate communication with the cluster, but the fetcher](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=67.13) [itself doesn't actually communicate with the cluster, that is the job](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=71.45) [of the consumer network client.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=76.33) [With the client open and sending TCP packets,](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=79.14) [the consumers start sending heartbeats,](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=82.77) [which enable the cluster to know what consumers are still connected.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=85.08) [Additionally, the initial request for metadata is sent and received.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=89.64) [The response is used to instantiate its internal metadata object,](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=95.04) [which will keep up to date,](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=99.77) [while the poll method runs, getting periodic updates from the cluster](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=101.31) [when cluster details change. With metadata available,](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=105.24) [other major elements become more involved. With information about](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=109.12) [the cluster, the consumer coordinator can now take responsibility](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=113.26) [to coordinate between the consumer.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=117.74) [This object has two main duties.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=119.76) [First,](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=122.47) [being aware of automatic or dynamic partition reassignment and](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=123.25) [notification of assignment changes to the subscription state object,](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=127.64) [and second, for committing offsets to the cluster,](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=131.81) [the confirmation of which will cause the update of the subscription state so](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=136.04) [it can always be aware of the status of topics and partitions. To actually](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=140.51) [start retrieving messages, the fetcher needs to know what topics or](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=145.19) [individual partitions it should be asking for.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=149.54) [It gets this information from the subscription state object](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=152.49) [and with it, starts requesting messages.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=156.08) [Here is where I'll explain what that value that is being passed to](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=158.65) [the poll method means. It is a timeout setting, representing the](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=162.48) [number of milliseconds the network client is to spend pulling the](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=166.77) [cluster for messages before returning.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=170.61) [This is an important setting because it establishes the minimum](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=173.05) [amount of time each message retrieval cycle will take.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=176.66) [I'll cover this shortly.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=180.55) [When the timeout expires,](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=182) [a batch of records are returned and added to an in‑memory](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=183.84) [buffer where they are parsed, deserialized, and grouped into](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=187.55) [consumer records by topic and partition. Once the fetcher finishes this process, it returns the objects for processing.](https://app.pluralsight.com/course-player?clipId=356dc73c-7534-43a5-8187-45e98dfea963&startTime=191.37)

### [Walkthrough: Message Processing](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440)

[An important thing to understand about Kafka consumers is](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=0.04) [that they are essentially single‑threaded.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=3.29) [There is one poll loop per Kafka consumer,](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=5.94) [and you can only have a single thread per Kafka consumer.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=8.81) [With all of the responsibilities that stem from the poll method,](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=12.94) [this may be surprising to you, if not downright troubling.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=16.31) [The Kafka consumer was designed this way mainly to keep its operations simple](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=20.64) [and to force parallelism of message consumption in another,](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=26.01) [more scalable way that we'll see shortly.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=29.73) [Again,](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=32.74) [knowing this is important for you and your approach to designing Kafka](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=33.55) [consumer applications because the reality of only a single thread available](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=37.21) [for record processing will have implications on how much you can reasonably](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=41.76) [expect a single Kafka consumer to do.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=46.05) [Let's discuss this further.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=49.54) [Let's continue our walkthrough of the consumer internals](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=51.11) [by discussing what happens after the poll method has](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=53.99) [returned messages for processing.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=56.74) [Since the return type of the poll method is a collection of consumer records,](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=58.85) [we will need to iterate through them to process them individually.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=63.13) [Now,](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=67.52) [what logic to apply to each individual record is entirely up](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=67.96) [to the developers working on the consumers.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=71.86) [But careful consideration should be made to how each record should be processed.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=74.54) [Remember, when calling the poll method,](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=80.34) [you can only do so much within a single thread.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=83.06) [If you were to spend too much time in processing records,](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=87.04) [it could have big implications on the environment in which the](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=90.33) [consumer application process is running.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=93.55) [Thankfully, because of Kafka's architecture,](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=96.74) [a slow consumer doesn't have an impact on the cluster,](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=99.58) [producers, or other consumers.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=103.8) [Nonetheless,](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=106.34) [it's important to remember that any one consumer can](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=107.49) [subscribe to any number of topics and partitions.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=110.9) [The more the consumer signs up for,](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=113.84) [the more it has to process and all within a single polling loop.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=116.77) [Given the possible load that a Kafka cluster can be required to handle,](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=120.64) [having a single consumer may not be a feasible or rational idea.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=124.71) [Let's get into some more details about the consumer so we can explore options for developing and configuring consumer applications at scale.](https://app.pluralsight.com/course-player?clipId=ea8cc6e5-fbcf-49c8-b484-25a32c220440&startTime=129.48)

### [The Consumer OFfset in Detail](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62)

[It's been a few modules since we discussed the all‑important offset.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=0.14) [If you recall,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=4.54) [the offset is the critical value that enables consumers to operate](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=5.78) [independently by representing the last read position the consumer](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=10.08) [has read from a partition within a topic.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=14.14) [When you think about the business of consuming messages,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=17.64) [you realize just how important the offset is and,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=20.66) [more importantly, whether it is accurate.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=24.25) [How Kafka manages the consumer offset is one of the](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=27.44) [more important things to understand,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=30.8) [and that's why we're going to spend a bit of time on it right now.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=33.03) [First, there is some important terminology to learn about the offset.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=36.94) [There are different categories of offsets,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=41.64) [with each representing the various stage they are in.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=43.86) [When an individual is reading from a partition,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=47.74) [it obviously needs to establish what it has and hasn't read.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=50.81) [This definitive answer is called the last committed offset,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=55.54) [and it represents the last record that the consumer](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=60.01) [has confirmed to have processed.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=63.2) [We'll get into this confirmation process shortly,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=65.84) [but this is the starting point for a consumer within any given partition,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=69.07) [depending on the configured offset/reset behavior,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=73.38) [which we'll also cover later. You will notice we're really looking at it](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=76.42) [from a partition viewpoint, and that is because each partition is mutually](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=81.4) [exclusive with regard to consumer offsets.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=85.74) [So for any given topic,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=88.89) [a consumer may have multiple offsets it's tracking, one](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=90.88) [for each partition within a topic.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=94.57) [As the consumer reads records from the last committed offset,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=97.94) [it tracks its current position. As we illustrated in module 3,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=102.25) [this position advances as the consumer advances in the log](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=107.18) [towards the last record in the partition,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=111.11) [which is known as the log end offset.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=113.64) [There is a notable difference, however,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=117.54) [between the current position and the last committed offset,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=119.51) [and it represents potentially uncommitted offsets.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=123.43) [The success of robust and scalable message consumption in Apache](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=127.72) [Kafka largely depends on your understanding of what creates this](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=131.87) [gap and what can be done to narrow it.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=136.45) [Every application has different processing requirements,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=139.44) [functional and nonfunctional.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=142.75) [It is the job of the application designer and developer to](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=145.04) [find the appropriate trade‑offs that work.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=148.94) [Next, I will walk through a scenario that illustrates this gap.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=152.04) [There are two very important configuration properties that govern the default](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=156.34) [behavior of the consumer offset. These properties are optional because their](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=160.55) [defaults are sufficient for getting up and running.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=165.66) [The first is enable.auto.commit,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=169.14) [which is basically giving Kafka the responsibility to manage when current](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=172.04) [position offsets are upgraded to full committed offsets.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=176.58) [This is a fairly blind setting because Kafka isn't going to](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=181.34) [know under what logical circumstances a record should be](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=184.73) [considered a committed record.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=188.36) [The only thing it can do is establish an interval of time between commit](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=190.74) [actions that faithfully commit based on a frequency.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=195.23) [That frequency is established by the auto commit interval](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=199.34) [property, and, by default, it is set to 5,000 milliseconds or 5](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=203.29) [seconds. Now for high‑throughput scenarios, 5 seconds is an](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=207.88) [eternity and likely sufficient.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=212.69) [But let's consider the biggest variable here for a moment,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=215.05) [and that is your processing logic.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=218.2) [When a record is in processing scope,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=220.94) [let's say it has a current offset position of 4 because the](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=223.42) [last successfully committed record was 3.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=226.76) [Let's also suppose that for whatever reason,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=230.24) [the processing of the current record takes longer than 5,000](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=232.71) [milliseconds or whatever that interval is set to. Faithfully.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=236.63) [Kafka is going to commit that record's offset regardless if](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=240.72) [it is finished processing or not because unless If you tell](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=244.21) [it explicitly when it's done, how is it supposed to know? Now this may be fine,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=247.46) [but it's not entirely consistent.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=253.38) [Generally, large‑scale systems operate within eventually consistent boundaries.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=256.24) [Now that's okay most of the time provided there is](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=260.98) [something else that's very important present,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=264.42) [and that is reliability.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=266.88) [Sorry, but I have to stand on my little soapbox for a minute.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=269.09) [The gap between what is considered committed and what is](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=272.2) [actually committed isn't entirely bad. As I said,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=275.41) [many large‑scale distributed systems aren't 100% consistent.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=278.84) [They are eventually consistent and Kafka. And consumers don't](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=283.06) [have to be an exception to that.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=286.67) [But the extent in which you can tolerate eventual consistency is](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=288.3) [based on your application's functional requirements,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=292.57) [of course, but also on the degree in which you can ensure reliability.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=295.16) [If you can't provide reliability and robustness assurances, then an](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=300.53) [eventually consistent ideal becomes a never‑consistent reality,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=304.67) [which can be a disaster.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=309.04) [So to continue with the offset gap illustration, suppose an are occurs that](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=310.75) [causes the message processing to fail for whatever reason.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=315.5) [Now what? Depending on how far behind the consumer was when it failed,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=319.08) [it may be very hard to know where you may need to go back](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=323.64) [to to start processing again because, according to Kafka,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=327.28) [the records were committed.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=330.92) [Knowing the current position at the time will be a start,](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=332.22) [but it could be messy to recover from.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=335.11) [The impact varies largely based on your consumer topology.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=337.05) [So far, we've only discussed a single consumer.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=341.33) [The issues for a single consumer are different for a topology where multiple](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=345.34) [consumers exist within what is called a consumer group. I keep pushing this down, but we'll talk about this soon enough.](https://app.pluralsight.com/course-player?clipId=3649d319-2b27-449f-8182-85b45298df62&startTime=349.68)

### [Offset Behavior and Management](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe)

[So to recap and summarize for now on offset behavior. Remember,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=0.24) [just because something is red, doesn't mean it's committed.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=4.73) [A lot of things determine this and it is very subjective depending on the](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=8.05) [offset management mode you're operating in. The offset management mode is](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=12.45) [determined by the offset configuration properties.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=17.73) [First and foremost is whether you want Kafka to manage your commits for you.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=21.64) [The default is true because it is very convenient from a development standpoint,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=26.74) [but as we saw, depending on the situation,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=31.59) [it can be operationally inconvenient if there is an issue. It's a lot](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=34.31) [like garbage collection in modern programming languages,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=39.12) [it's very convenient until it is inconvenient.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=42.55) [The challenge is generally to have some sort of control to](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=45.84) [govern when it is tolerable to be inconvenient.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=48.97) [Fortunately, in Kafka,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=52.84) [you can adjust the commit frequency to be in line with](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=54.39) [your particular consumer application.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=57.71) [This is the commit interval we discussed earlier.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=60.54) [Lengthening this interval will provide an upper bound in which you can ensure](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=63.84) [your record processing will be finished, but it could also create an offset gap](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=68.74) [in the opposite direction where the commits are lagging behind your processing](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=73.39) [positions. As long as there is a gap,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=77.78) [there is some risk exposure to failure and the possible](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=81.24) [inconsistent state you may be left with to clean up,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=84.48) [not to mention the possible duplication of records when reprocessing. Another](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=87.5) [property we haven't covered yet, but will, is the strategy to use when a](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=93.1) [consumer starts reading from a new partition. The default is to start reading](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=98.19) [from the latest known committed offset.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=103.65) [In contrast, this could also be set to the earliest.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=106.74) [There is also a setting for none,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=111.04) [which basically you're asking Kafka to throw an exception to the](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=112.88) [consumer and let you decide what to do with it.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=116.99) [The offset behavior and the issues related to it vary depending on whether](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=120.44) [you're in a single consumer or a consumer group topology.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=125.21) [All this time we've been talking about offsets,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=129.44) [and I haven't taught you how and where they are stored in Kafka.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=132.1) [The only thing I've said at this point is that consumers](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=136.64) [track the offset in terms of what it has or has not read, but](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=139.79) [where does it actually store them?](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=144.27) [Any guesses?](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=146.52) [Think about how Kafka stores data, period.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=148.04) [If you guessed a topic, you would win a prize.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=151.57) [Kafka stores the committed offsets in a special topic called \_\_consumer\_offsets.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=154.48) [If you were to issue a describe command to the cluster asking it to](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=162.64) [show you all of the topics and their partitions,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=166.55) [you would notice this consumer offsets topic and it would have 50 partitions.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=169.47) [Yeah, that's a lot of partitions for a single topic.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=174.91) [Now, why they chose the default of 50 is beyond me, but that's what it is.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=178.44) [Okay,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=183.44) [using the demo from the last time where we already have a](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=183.95) [couple of topics in place and some data in them,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=187.71) [I wanted to show you how to take a look at the \_\_consumer\_offsets topic,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=191.03) [which again, is these designated topic to store all of the consumer](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=198.73) [offsets throughout the entire Kafka cluster. Now,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=203.26) [in this scenario, I only have one single broker running,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=206.98) [but let's take a look at what this offset describe would do.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=210.47) [So here, you're going to see it listing all of the](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=218.64) [partitions that are in this particular topic,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=222.01) [which again, is \_\_consumer\_offsets, it has a partition count of](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=225.13) [50, and it only has a single replication factor,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=230.3) [which is a little bit dangerous,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=233.43) [but if we tried to set that higher, at this point, with only one broker running,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=235.25) [we would get an error.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=240.14) [So it's probably only doing a replication factor based on the number of](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=241.79) [nodes available to it, at this point, by default.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=247.45) [So we now know the committed offsets are stored in a topic](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=250.74) [on the cluster, but how does the committed offset values](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=253.96) [get produced into the topic?](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=257.33) [Remember the class consumer coordinator we touched upon earlier?](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=259.7) [This is the responsible object for communicating to the cluster and](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=263.61) [ensuring the committed offsets are produced into the topic.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=267.66) [This means that a consumer is also a producer of sorts.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=271.65) [We've covered quite a bit more about offsets,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=275.54) [but there are a few more points I want you to add to your](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=278.63) [growing Kafka encyclopedia and the offset mode I](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=281.23) [mentioned a couple of slides back.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=284.85) [There are effectively two modes, automatic and manual. Automatic](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=287.64) [being the default. To switch to manual mode,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=291.92) [you simply set enable.auto.commit property to false.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=294.72) [Of course, by doing this,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=298.84) [the property for auto commit interval is irrelevant, and, therefore ignored.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=300.7) [When you do this,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=305.24) [you are taking full control of when you want Kafka to](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=306.25) [consider a record to be fully processed.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=309.55) [This is a fairly advanced, but not uncommon scenario.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=312.34) [We won't get into it in depth in this course,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=316.14) [but I will give you a high‑level overview of why,](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=318.6) [how, and what it means to use it.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=321.28) [The API for manual offset management consists of two methods, commitSync and commitAsync.](https://app.pluralsight.com/course-player?clipId=627953a2-499c-49f8-b11f-78f7c9e073fe&startTime=323.81)

### [CommitSync and CommitAsync for Manual Offset Management](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28)

[You would use the commitSync method when you want precise control](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=0.04) [over when to consider a record truly processed.](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=4.1) [This is common under circumstances where higher consistency](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=8.04) [and message processing fidelity is required,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=11.95) [where you wouldn't want to retrieve and process new records until you're](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=14.52) [sure the ones you've currently processed are committed.](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=18.6) [It is suggested that you invoke this method after you have iterated and](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=22.72) [processed a batch of consumer records in the for loop,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=27.11) [not during.](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=30.56) [I mean,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=31.94) [you can invoke it after every single message, but that level of paranoia](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=32.43) [may not buy you anything extra other than added latency,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=36.76) [because the call is, as the name suggests,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=40.54) [synchronous, and will block the thread until it](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=43.51) [receives a response from the cluster.](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=46.3) [Hopefully,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=48.64) [the response is a successful confirmation, because if it is an exception,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=49.47) [there's not much you can do and you'll just have to](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=54.15) [start the process of recovery.](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=56.78) [The good news about commitSync is that it will](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=59.44) [automatically retry the commit until it succeeds,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=62.19) [or again,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=65.79) [if it were to receive an unrecoverable error. To control the](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=66.68) [retry attempt interval, you would work with the](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=70.83) [retry.backoff.ms setting, and it's similar to the setting found](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=73.59) [in the producer configuration as well.](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=78.36) [The default is 100ms, so it will retry a lot.](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=80.54) [With this manual offset management mode,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=84.95) [you may be trading throughput and performance for control over the consistency.](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=87.59) [The synchronous blocking nature of the call can add a measure of](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=93.04) [latency to the overall polling process. Like the commitSync method,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=96.33) [you would use it's asynchronous sibling to control when to consider](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=100.48) [your message as truly processed.](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=103.95) [The difference here is due to the asynchronous nature of the call, you](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=106.44) [may not know exactly when the commit succeeded or not.](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=110.3) [Because of this,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=113.94) [the commitAsync method does not automatically retry](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=115.08) [when a commit doesn't happen.](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=118.56) [Retrying without knowing whether the first attempt](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=120.84) [succeeded or failed can lead to ordering issues and](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=123.39) [possible duplication of records; however,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=126.62) [there is a useful option to pass in,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=129.82) [and that is a callback. That callback will be triggered](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=132.14) [upon the commit response from the cluster.](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=135.07) [With this callback,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=137.82) [you can determine the status of the commit and act accordingly.](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=139.04) [Since this is a non‑blocking option,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=142.94) [the throughput and overall performance is going to be better because you will](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=145.41) [not have to wait for a response to continue processing.](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=148.91) [However,](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=152.54) [I wouldn't recommend this option unless you register a callback](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=153.28) [and can handle the responses accordingly. Otherwise, you could end up in a worse situation altogether.](https://app.pluralsight.com/course-player?clipId=3db27bec-7be2-44c2-8e7b-499d31ed7f28&startTime=157.45)

### [When to Manager Your Own Offsets Altogether](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3)

[So we've nearly completed our journey through our Kafka consumer map.](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=0.14) [We've discussed at length the important process of managing](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=4.74) [offsets as part of the overall consumer's responsibility for](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=8.29) [reliably processing messages.](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=12.46) [The place where offset management occurs is after the poll method has](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=14.94) [timed out and presented records for processing.](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=19.25) [Whether this is an auto commit operation happening behind the scenes](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=22.64) [or an explicit call to one of the commit APIs,](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=26.53) [the commit process will take a batch of records,](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=29.84) [determine their offsets,](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=32.75) [and ask the consumer coordinator to commit them to the Kafka](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=34.25) [cluster via the consumer network client,](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=37.7) [which it does immediately. When the offsets have been](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=40.73) [confirmed to be committed, the consumer coordinator updates](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=43.5) [the subscription state object accordingly,](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=47.15) [so the fetcher can always know what offsets have been committed](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=49.8) [and what next records it should be retrieving.](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=53.5) [There are a lot of things Kafka can do for you out of](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=56.54) [the box. But for advanced scenarios,](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=59.36) [you may need to go outside the box entirely and leverage](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=61.62) [Kafka's APIs for complete offset self‑management.](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=64.96) [We discussed many of the facilities for doing this,](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=68.54) [and I would encourage you to explore the APIs further.](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=71.33) [The question is what are some common reasons for taking](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=75.04) [control of the offsets? As we touched upon already, one of](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=78.57) [those big reasons is consistency control.](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=83.03) [Depending on your consumer application's purpose in the larger system,](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=86.34) [you may need finer‑grain control over when a message is](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=90.33) [processed and considered ready to commit.](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=93.84) [If you leave it to the auto commit behavior,](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=97.15) [the only determination of done will be when the auto commit interval](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=100.15) [expires, and that may not be enough to ensure higher levels of consistency.](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=104.3) [Being able to treat the steps of message consumption and processing as a](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=108.7) [single atomic operation, that's a good reason.](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=113.47) [This is commonly understood and known in transaction](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=117.64) [processing systems as atomicity.](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=120.71) [It is an important attribute of highly consistent systems and](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=123.84) [may be required by your particular system.](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=127.78) [The main reason independent offset management becomes a common](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=131.44) [scenario with Kafka is the desire to achieve exactly](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=135.23) [once‑semantics of message processing. Because of what can go on](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=139.46) [within a distributed system like Kafka,](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=144.07) [there is quite a bit of surface area to get messages](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=146.53) [out of order or have duplicates.](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=149.24) [This surface area is largely attributed to the scalable nature of how](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=152.04) [Kafka handles partitions and automatic partition reassignment and](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=155.96) [rebalancing, topics which we're going to cover next.](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=160.13) [But in order to get an exactly once system,](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=163.42) [you will likely need to manage offsets and the content of the message](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=166.77) [and/or the result of its processing in the same store where you can have full transactional control of the scope.](https://app.pluralsight.com/course-player?clipId=834a7fe9-ec95-4a56-9014-3be160f7f5f3&startTime=171.04)

### [Scaling out with Consumer Groups](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a)

[Up to this point, we've discussed a lot about consumers,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=0.14) [mostly single consumers.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=3.6) [In these discussions, we've been faced with a scary reality.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=6.03) [A single consumer may be required to consume from](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=10.04) [dozens or possibly hundreds of topics, each with countless partitions.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=13.27) [That's a lot for a single anything to manage,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=18.63) [let alone having to do it with a single execution thread for](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=21.18) [both retrieving and processing messages.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=24.96) [As I said,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=27.69) [it simply isn't realistic to expect a single consumer application](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=28.74) [to take on the entire burden of message processing from a](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=33.24) [potentially large Kafka cluster environment.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=36.75) [The solution is to be able to scale out the number](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=40.04) [of consumers consuming messages.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=43.07) [But having a bunch of consumers independently consuming messages from topics and](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=45.64) [partitions won't alone solve this challenge of scalability,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=50.05) [they have to work in concert with one another.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=54.22) [Throughout this course,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=56.92) [we've seen how each component of Apache Kafka has a solution for scaling out.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=58.1) [If more message production is needed,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=63.74) [the solution is to add more and more producers.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=65.69) [If we need more message retention and redundancy, we add more and more brokers.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=69.12) [If we need more metadata management facilities, we add more zookeeper members.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=74.61) [But What about scaling the ability to read and process](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=80.34) [messages beyond a single consumer?](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=83.34) [Consumer groups is the answer.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=85.96) [A consumer group really is a collection of individual independent](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=88.14) [consumer processes working together as a team.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=92.51) [The only thing required to join a consumer to a consumer](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=96.14) [group is to use the group.id setting as a configuration](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=99.63) [property before starting the consumer.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=104.51) [When a consumer is part of a consumer group,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=106.96) [the task of processing the messages for an entire topic is distributed as](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=109.64) [evenly as possible amongst the number of consumers.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=114.53) [Like any work distribution system,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=117.6) [a consumer group can enable higher levels of overall throughput](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=119.9) [through multiple consumers working in parallel.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=124.21) [It can increase the levels of redundancy as the failure or limitation of a](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=127.04) [single consumer is automatically handled and balanced by Kafka,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=131.62) [and with an increased number of working consumers working in parallel,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=135.79) [the overall performance can improve as far as the ability](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=140.6) [to process a large backlog of messages.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=144.06) [Let's look at how this works.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=147.24) [A consumer group is formed when individual consumers with a common group ID](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=148.96) [invoke the subscribe method and pass in a common topic list.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=153.81) [Behind the scenes,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=158.84) [a designated broker is elected to serve as a group coordinator, whose job it](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=160.05) [is to monitor and maintain a consumer groups membership.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=165.02) [In addition,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=169.41) [the group coordinator works with the cluster coordinator and](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=170.49) [zookeeper to assign and monitor specific partitions within a topic to](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=173.83) [individual consumers within a consumer group.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=179.03) [From the second a consumer group is formed,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=181.61) [each consumer is sending regular heartbeats at an interval](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=184.34) [defined in heartbeat.interval.ms property setting.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=188.03) [The group coordinator relies on this heartbeat to determine whether an](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=191.87) [individual consumer is alive and able to participate in the group.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=195.87) [The session.timeout setting is the amount of total time a group](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=200.54) [coordinator will wait after not receiving any heartbeats before it will](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=204.95) [consider the consumer failed and take corrective action.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=209.22) [The group coordinator's main priority is to ensure that](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=212.97) [the purpose of the group is being met,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=216.61) [and that purpose is sharing the load of a topic's messages](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=219.02) [amongst all of its consumer group members.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=222.59) [If there is a consumer that isn't available to share in that load,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=225.74) [the group coordinator will remove that consumer and reassign its](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=229.94) [partitions to another consumer in the group.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=233.48) [This is called a consumer rebalance, and it is,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=236.54) [as you can imagine, quite a process.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=239.97) [If there aren't any additional consumers in the consumer group,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=242.74) [the first consumer in the group will get the new assignment and in this case,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=246.48) [end up taking on twice the load to compensate for the failed consumer.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=250.72) [When this happens,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=254.68) [the first consumer now has to figure out where the](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=255.94) [failed consumer left off and catch up,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=258.63) [hopefully without processing duplicate records.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=261.16) [This is why offset management can make or break the Kafka](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=264.5) [consumers because if it is not handled correctly,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=268.21) [the ability for the consumer group to failover and](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=271.52) [rebalance itself can be compromised.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=274.44) [Consider the case if the failed consumer processed messages](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=278.04) [but failed to commit them before it failed.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=282.17) [The first consumer will likely reprocess the messages because it had no](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=284.87) [idea what records were actually committed or not.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=289.35) [If and when a new consumer joins the group,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=292.24) [another rebalance will occur and the same rebalance protocol will be followed.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=294.84) [It's not just a consumer coming in and out of a consumer group that will](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=299.94) [cause a rebalance, it's also the addition of a new partition.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=304.3) [Generally,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=308.5) [a consumer group is planned for each application that](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=309.27) [requires message flow from one or more topics.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=312.56) [For example, in this case, we have a consumer group called orders,](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=315.95) [and it could subscribe to any number of topics related to order](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=320.14) [management, because the application backend it is intended to serve is an order management system.](https://app.pluralsight.com/course-player?clipId=007cd947-5396-44de-b02c-29909985876a&startTime=323.71)

### [Consumer Group Coordinator](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463)

[Let's spend some more time on what happens during a consumer group](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=0.44) [rebalance, specifically, when a new consumer in the group is assigned a](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=3.95) [partition that was previously assigned to another consumer.](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=8.49) [When the new consumer is assigned a partition, in this case partition 0,](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=12.94) [it needs to know what offset it should start from because it does not](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=17.58) [have a current position for this particular partition.](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=21.9) [Fortunately,](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=25.84) [the consumers subscription state object has cached the last committed](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=27.02) [offset from the previous consumer and can now instruct the new consumer](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=31.66) [that on its first poll on the new partition that it will start with offset](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=36.36) [5 since the last committed offset was 4. This behavior of determining](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=41.02) [where the new consumer should reset its offset to is configured in the](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=47.17) [auto offset reset setting.](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=51.9) [Since the default is latest,](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=54.64) [the new consumer will start reading from the latest known committed position.](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=56.75) [Of course,](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=61.24) [this assumes that the committed offset was accurately and completely](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=62.06) [committed when the previous consumer was rebalanced.](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=66.23) [If the previous consumer was in the middle of processing](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=69.74) [records and didn't have the chance to commit its offsets](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=72.89) [when the rebalance happened,](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=76.49) [then there could be a chance that when the new consumer picks up,](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=78.32) [it could be reading from already processed records,](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=81.74) [thus creating duplicates. To finish up the discussion on](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=84.82) [consumer groups, let's highlight some of the important](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=88.6) [duties of the group coordinator,](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=91.66) [without which consumer groups wouldn't be possible. The primary](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=93.29) [purpose is to make sure each consumer in the consumer group is](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=97.44) [sharing the partition load across the group. Whenever it can, it](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=101.65) [will assign one consumer to one partition if there is an equal](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=105.8) [number of consumers and partitions.](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=109.98) [However, if there are more consumers in the group than there are partitions,](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=112.11) [the extra consumers will be idle,](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=116.84) [creating a consumer over‑provisioning scenario that the group](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=119.34) [coordinator can't change unless partitions become available.](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=122.86) [When a partition does become available,](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=127.34) [the group coordinator will initiate the rebalancing protocol by](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=129.83) [engaging each consumer coordinator in the impacted consumers to start](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=133.77) [the process of rebalancing so the newly added partition can be](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=139.3) [assigned to an appropriate consumer.](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=143.34) [The rebalancing protocol is also initiated during a consumer failure, as we just illustrated.](https://app.pluralsight.com/course-player?clipId=7d6869df-235b-441d-b8bb-ea6949dbb463&startTime=146.11)

### [Demo: Consumer Groups](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37)

[In this final demo,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=0.74) [I will extend the Java‑based consumers we've already seen to take on](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=2.24) [teaming responsibilities within a consumer group.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=6.32) [We will have three independent consumers,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=10.04) [each sharing the same group ID and each participating in the task of](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=12.55) [processing messages from a single topic with three partitions.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=16.79) [In this simple example,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=21.24) [look for how each of the consumers are assigned a partition and](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=22.76) [are sharing the work of processing messages.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=26.26) [Also,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=29.04) [look for what happens when we add an additional consumer](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=29.62) [and when we add an additional partition.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=33.13) [Finally, we'll observe what happens when a rebalance is forced.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=35.84) [In this demo, we're going to launch three basic consumer applications,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=40.44) [which contain identical polling and processing logic.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=43.98) [Here it is.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=48.34) [It's basically pretty simple.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=49) [We're just taking the string that's coming through in the producer,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=50.81) [and we're taking the value and upper‑casing it,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=55.26) [and that's it.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=58.81) [And each one of these three is identical,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=60.84) [and as far as producing, we have a loop running,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=64.14) [and we'll produce 99 records of the alphabet in lowercase,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=67.58) [and all the processing that it will do on the consumer side](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=73.46) [is just to turn those into uppercase.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=76.92) [All right, so let's get these consumers running.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=79.44) [We'll start one by one in each window,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=82.08) [and they'll sit there and wait in a polling loop](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=85.25) [until our producer starts running.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=89.61) [And we'll do that right now.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=92.02) [Now, remember, each of these consumers are in a consumer group,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=93.08) [all subscribing to the same topic,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=96.64) [and the producer is going to be publishing to that topic.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=98.75) [So here we go. First one, the second one, and the third one.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=101.59) [So as we can see here is this particular consumer as part](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=106.84) [of the consumer group was getting all of the messages being](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=113.1) [sent to the first partition, and it took the lowercase alphabet and uppered it.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=116.71) [The same thing could be said of the second consumer in the consumer group.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=122.25) [It was taking partition 2, taking the value, and turning it to uppercase.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=125.94) [And then consumer 03 from the consumer group was taking partition 0. Okay,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=131.39) [let's take a look at what happens when we add a](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=137.52) [fourth consumer to the consumer group.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=142.1) [Now, remember, we have three consumers to the consumer group,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=144.95) [and we have one topic with three partitions.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=148.85) [We have an over‑provisioned consumer group. But let's see what happens.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=151.79) [So we'll start each one of these consumers,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=156.64) [let them sit and wait for messages, and then we will produce to them.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=160.14) [So now we have four in the consumer group, and we can see it in](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=163.56) [here that basically there are four consumers in the consumer group](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=167.01) [test group. And as I was adding them,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=172.97) [they added to them and registered to the group coordinator.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=175.16) [So now let's produce some records.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=178.46) [So we see that the first one receives some records.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=180.74) [The second one receives some records.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=184.19) [The third one did not receive any records,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=185.74) [and I believe the fourth one will have received some records.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=188.58) [So why didn't the third one receive records?](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=192.37) [Well,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=196.04) [because there are only three consumers in the consumer](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=196.57) [group and there are three topics.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=200.17) [So one of them is just sitting idle, and that one happened to be the third one.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=202.59) [The first one here got partition 1.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=208.25) [The second one got partition 2, and the fourth one got partition 0,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=211.94) [and they did their job by taking the value and](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=216.85) [turning it to uppercase all the same.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=219.49) [Okay,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=221.13) [so now that we've added an additional partition, and now we have](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=221.44) [even numbers of consumers to partitions to consume,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=224.94) [let's rerun the producer and see how it now distributes the messages](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=230.34) [across an even number of partitions because, basically,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=235.61) [when we added the new partition, it forced a rebalance.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=238.91) [So let's see what happens.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=242.68) [So for App01 got messages, App02 got messages, App03 got](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=244.41) [messages, and App04 got messages. Four got partition 0, 03](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=250.88) [got 1, 02 got 3, and 01 got 2.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=255.95) [So it did rebalance.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=262.22) [It did recognize the new partition,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=263.88) [and it did assign the new consumer in the consumer group to that partition.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=266.39) [All right, now that we have an even consumer group,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=273.34) [even with the number of partitions, let's force a rebalance.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=276.91) [Now the way you do that is by a basically killing a couple of](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=281.14) [consumers because if they are no longer present,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=284.92) [the group coordinator will not find them.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=288.64) [They won't get their heartbeat, and they will](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=290.9) [remove them from the consumer group.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=293.13) [So let's kill, let's kill 04 and 03. So now they're sitting](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=295.34) [there, and let's go back over to here.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=305.66) [Now, remember,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=309.64) [it takes a little while for the rebalance to occur because, basically,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=310.52) [at this point in time, the group coordinator is waiting around,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=315.65) [waiting for some heartbeats from consumer 04 and 03,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=319.4) [and so far it's not getting them.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=324.14) [And then after a while, it's going to now say,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=326.24) [Oh,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=329.95) [my session's going to start timing out, and I'm going to remove](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=330.36) [the dead consumers from the consumer group.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=334.64) [So these consumers are still running.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=337.68) [So let's now produce some more records.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=339.44) [So now you'll see that consumer 001 took some records.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=341.74) [Consumer 02 took some records. And, of course,](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=346.75) [03 and 04 did not because they were unresponsive.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=349.64) [But what happened was 01 and 02 basically were](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=353.01) [reassigned the partitions that 03 and 04 had.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=356.39) [In this case, it was partitions 1 and 0 that was reassigned to consumer 01, and 2 and 3 were reassigned to consumer 02.](https://app.pluralsight.com/course-player?clipId=1d78ad6d-f152-4f47-8508-e19b5465ad37&startTime=360.9)

### [Configuration and Advanced Topics](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72)

[We have covered a lot of consumer configuration properties in this module,](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=0.04) [but we haven't covered all of them due to time and focus.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=3.97) [Since the performance and throughput of consumer processing can be](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=7.44) [affected by various settings and combinations of settings,](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=11.3) [I thought I would list a few of the more prominent settings here.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=14.74) [My goal is to call some of these out so that you can spend some](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=19.01) [extra time studying them and experimenting with them to](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=22.61) [understand how the consumer behavior varies.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=26) [These settings fall into a category I would call](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=28.66) [performance and overall efficiency.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=31.68) [The fetch.min.bytes setting sets the minimum number of](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=34.14) [bytes that must be returned from the poll.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=37.89) [This ensures that you don't have wasted cycles of processing](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=41.01) [if there aren't enough messages to process.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=44.62) [This setting is analogous to the batch size setting on the producer.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=47.74) [The max.fetch.wait.ms setting establishes the amount of time to wait.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=52.5) [If there isn't enough data to meet the threshold set](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=56.8) [by the fetch.min.bytes setting.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=59.7) [This is somewhat analogous to the linger.ms setting in the producer.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=61.87) [To ensure that each poll isn't retrieving more data than](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=66.64) [your processing loop can handle safely,](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=70.22) [you can set the maximum number of bytes per partition](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=72.4) [that the poll can retrieve per cycle.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=76.64) [Related to this is the setting to establish the maximum](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=79.74) [number of records allowed per poll cycle.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=83.71) [These last two settings are useful to throttle the number and size](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=86.42) [of each incoming batch of records should your processing loop be](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=90.19) [such that a lot of time is spent in processing and you don't want](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=94.2) [to risk a session timeout.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=97.51) [We covered a lot,](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=99.11) [but there are some things we just weren't able to](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=100.64) [cover in this introductory course.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=102.99) [Each of these fall into the category of taking complete](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=105.34) [control of the consumer's behavior.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=108.4) [You can specify how you want a consumer to read a](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=110.83) [partition's messages by using the consumer position control](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=114.18) [API. It comprises of three methods.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=117.98) [First is the seek' method,](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=121.15) [allowing you to specify the specific offset you want to](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=123.04) [read in a given topic and partition.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=126.63) [There's also seekToBeginning,](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=129.54) [which indicates that you want to start from the beginning of a](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=131.29) [group of a specific topics and partitions.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=134.2) [Obviously, seekToEnd is the opposite of seekToBeginning.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=137.09) [And then, there's the ability to literally control the flow](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=141.04) [of messages through pause and resume APIs.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=144.38) [These allow you to determine which topics and partitions you](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=148.09) [may want to pause while focusing on other topics and](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=151.42) [partitions considered a higher priority.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=154.78) [This is useful for situations where a single consumer has to read](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=157.68) [from multiple different topics and partitions.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=161.81) [Finally,](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=164.24) [there are the rebalance listeners that you can leverage when](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=165.1) [subscribing to topics in a consumer group.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=168.36) [These listeners will notify you when a rebalance event occurs so you can manage how you want to handle the offsets yourself.](https://app.pluralsight.com/course-player?clipId=934ee197-dc9e-4b8e-85ca-6bc704abcb72&startTime=171.54)

### [Summary](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89)

[Throughout this module, we focused on the internals of a Kafka consumer.](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=0.94) [We started with a high‑level map and started to drill down component by](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=6.14) [component and in the process covered a lot of things,](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=10) [such as the required consumer properties and their internal consumer](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=13.63) [representation as the ConsumerConfig object, which was analogous to the](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=17.46) [ProducerConfig object in the producer, once again how the term message is](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=22.07) [really a reference to a ConsumerRecord,](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=26.95) [much the same way it was called the ProducerRecord from the point of view](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=29.44) [of the producer in the last module. We talked about how to subscribe to](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=33.44) [topics and how to assign yourself partitions.](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=37.54) [We discussed the important differences between the two and when it](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=40.57) [may be appropriate to use one over the other.](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=44.22) [We also talked about the end‑to‑end consumer polling process,](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=47.24) [complete with the poll method,](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=51.24) [the for loop for processing records, and all the internal consumer](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=52.9) [objects that enable to consumer to function.](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=56.96) [We also discussed the various different modes and options](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=59.76) [for managing offsets in the consumer.](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=63.28) [Additionally,](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=66.04) [we covered the way Kafka consumers can scale out through consumer](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=66.94) [groups and how using consumer groups can increase the overall](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=70.7) [throughput possible through parallel consumers,](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=74.87) [but also the degree in which consumers can be fault tolerant](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=77.96) [and robust amidst failure or cluster changes.](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=81.71) [We covered throughout the various configuration settings and how they control](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=85.24) [the behavior and non‑functional outcomes of the consumer.](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=89.91) [And we had some demonstrations showing how to create and operate a](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=93.47) [Java‑based consumer and consumer group. With the core components of](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=97.41) [Apache Kafka now covered, in this last module,](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=101.6) [we'll cover the broader ecosystem that Apache Kafka finds itself in,](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=104.94) [including its current challenges, opportunities, and the most recent areas of continued development and evolution.](https://app.pluralsight.com/course-player?clipId=95275751-36e1-4494-8b86-88282500ae89&startTime=109.18)

## [Exploring the Kafka Ecosystem and Its Future](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d)

### [Apache Kafka's Success and Challenges](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d)

[At this point, we've covered the major components of Apache Kafka.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=0.11) [I hope you're feeling equipped with enough knowledge to really start](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=3.79) [exploring and building big data solutions using Kafka.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=7.02) [This module is about taking a step back and surveying the](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=10.44) [landscape in which Apache Kafka exists.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=14.35) [We will discuss the success it has enabled to continue challenges it faces](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=17.64) [and how it is evolving to meet those challenges head on.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=22.26) [The main use cases for Apache Kafka today have more or less remained](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=25.84) [the same since it was first created by LinkedIn.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=29.81) [It's hard to go anywhere and have discussions about the challenges of](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=33.09) [data management in modern day enterprises without the mention of](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=36.87) [Apache Kafka. It is generally regarded as a primary solution for](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=40.58) [connecting disparate sources of data.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=45.4) [With its flexible client APIs,](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=47.84) [it is possible to write data connectors and syncs](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=50.07) [for practically any data source.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=52.5) [Many of these have been shared and commercialized at this point,](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=55.2) [and we'll discuss more about them in the coming slides.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=58.76) [Apache Kafka is becoming the de facto option for building data supply](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=62.84) [chains and pipelines that can displace long‑standing,](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=68.15) [expensive, and fragile ETL environments.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=71.52) [Within this context.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=75.44) [Apache Kafka fits really well with other "Big Data"](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=76.97) [solutions like Hadoop and Spark, amongst others,](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=81.71) [because of its ability to integrate, move, and store data at massive scale.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=85.22) [Essentially, reference architectures for data management has](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=90.44) [started to become established within the industry,](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=94.2) [and Kafka is a central piece to many of them. However,](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=97.14) [sometimes new solutions introduce new problems and](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=100.94) [reinforce old unsolved problems.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=104.1) [Despite the vast utility that Kafka offers today's organizations,](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=107.14) [there are still a lot of gaps that the industry is being pressured to solve.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=111.76) [For example,](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=115.97) [having the ability to unmask and manage more data actually makes it](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=117.01) [harder to govern data and manage its rapid evolution.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=121.61) [The commoditization of technology and business specialization](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=126.24) [demands lower overhead and less investment.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=130.16) [So regardless of how useful something in technology is,](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=133.64) [it will always be a challenge the more inconsistent or costly it is to wield.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=137.28) [Data is becoming more and more of a strategic differentiator.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=143.34) [In the last five years,](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=147.57) [there has been an arms race for anyone and anything](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=149.2) [that can manage more and more data.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=152.77) [The next 5 years is going to be all about fast data,](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=155.64) [how to rapidly gain utility from it,](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=159.53) [particularly in predictive, deep learning contexts.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=162.29) [Over the next few slides,](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=167.04) [I will use these three challenge areas to describe how Kafka is evolving to address these pressures.](https://app.pluralsight.com/course-player?clipId=06c5b1c6-8fc7-4e27-9e72-aa9835a1457d&startTime=168.71)

### [Challenges and Solutions for Data Governance](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d)

[I'll start off with Kafka's challenges with data governance and evolution.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=0.14) [Let's consider the common case of a large and growing](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=4.84) [network of Kafka producers and consumers.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=8.57) [As you know by now, each producer is defining its message contract to publish.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=12.35) [You'll recall from module 4 that that contract is based on a](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=17.94) [fairly rigid type dependant serialization system.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=22.34) [We didn't talk a lot about this or nearly in as](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=26.64) [much detail as I would have liked.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=29.69) [But in advanced cases, it becomes infeasible to restrict message](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=31.98) [contracts solely based on the built‑in serializer types.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=36.26) [Eventually, as more data diversity is introduced from different systems,](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=40.92) [custom serializers come into play. Throughout the message lifecycle, there](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=45.62) [can be hundreds of different contract versions in motion,](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=50.81) [with each producer publishing massive amounts of data into Kafka.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=54.06) [Of course,](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=58.94) [it takes consumers to derive any sort of value from the data being produced,](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=59.75) [but they have to be able to do it by reading the data first,](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=64.82) [which they're able to do through deserializing the message content.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=68.4) [This means that with a growing diversity of producers](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=72.92) [and the data they're publishing,](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=76.12) [there is an increased complexity all around because consumers have](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=77.74) [to work with the data being produced and the specifications for](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=82.07) [each type of method it's consuming.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=85.81) [The challenge with Kafka in this common scenario is the](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=88.44) [lack of some common means of cataloging,](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=91.81) [registering, and reconciling the disparate message](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=94.56) [specifications and compatibility mappings between the](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=97.95) [serializing producers and the deserializing consumers.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=101.18) [Confluent is one of the biggest Apache Kafka ecosystem contributors,](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=105.23) [and they have recognized the challenge we just covered.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=110.54) [Fortunately,](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=113.34) [they have started to take steps to address it by](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=114.52) [introducing the Kafka Schema Registry.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=117.33) [This welcome addition to the Kafka family deserves its own course](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=120.84) [because of the richness of its functionality.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=125.13) [But for now, let me introduce how it addresses the challenges we just covered.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=128.04) [One of the more universal data serialization formats out](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=132.94) [there today is called Apache Avro.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=136.72) [It was created to address the challenges with disparate data](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=140.04) [formats and serialization schemes that make integration and](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=143.66) [interoperability difficult.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=147.23) [It is a self‑describing version format that has broad industry adoption.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=150.24) [With Avro, producers can serialize their messages in an Avro‑versioned and](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=155.83) [self‑describing format and expect them to be deserialized seamlessly by the](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=161.65) [consumers. As the name suggests, the schema used by both producers and](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=166.91) [consumers can be registered and version managed centrally within the Kafka](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=172.22) [cluster environment, allowing for easy,](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=177.22) [RESTful service‑based discovery and version compatibility reconciliation.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=179.53) [Now the great thing is the source is fully available on GitHub and available through the generous Apache version 2 license.](https://app.pluralsight.com/course-player?clipId=225d0808-e4cf-41ab-bdd3-d9473b82d85d&startTime=185.14)

### [Challenges and Solutions for Consistency and Productivity](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293)

[Let's consider a typical enterprise data environment.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=null) [There are many sources and targets for data.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=3.7) [Kafka has made quite a reputation for itself in being the](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=6.54) [conduit between these sources and targets.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=10.22) [But the challenge has been a lot of duplication of effort in](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=13.22) [terms of writing producer and consumer applications that](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=16.82) [connect the sources and targets together.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=20.28) [The crazy thing is, when you think about the work to integrate data stores,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=23.23) [they're all more or less the same.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=27.69) [I mean, look at relational database management systems, for example.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=29.53) [They've been around forever,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=33.85) [and there's only so many mainstream database vendors](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=35.35) [out there, yet, across the industry,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=38.33) [it seems that within every company there's the same duplicated effort](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=41.02) [to write integrating producers and consumers for those very same data](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=45.05) [stores. Talk about reinventing the wheel.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=49.35) [The same could be said about file systems,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=52.93) [NoSQL databases, search engines and even Hadoop,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=55.44) [amongst others not mentioned.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=59.15) [The challenge with Kafka in this scenario has been the lack](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=61.54) [of consistency in providing a common framework for](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=64.99) [integrating data sources and targets.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=68.78) [It was always left to the individual engineers to create their own solutions,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=71.53) [using the generic producer and consumer client](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=75.87) [APIs. With each integration effort,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=78.49) [there is cost not only to develop, but to maintain,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=81.16) [and that isn't a very efficient or even productive use of](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=84.78) [time or effort to do something so common.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=88.02) [Furthermore,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=91.74) [not every company has the resources to develop and maintain these things,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=92.73) [which are really becoming commodities at this point.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=96.96) [With the 0.10 release of Apache Kafka,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=99.94) [a new framework and marketplace was introduced to](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=102.9) [address this challenge head‑on.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=105.79) [It's called Kafka Connect and the Connector Hub.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=107.84) [As with the case of the schema registry,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=111.13) [this new innovation deserves its own course to give it justice.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=113.42) [The Connect framework is an API for developers.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=116.99) [It is intended to make the job of connecting data sources](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=120.29) [and targets easier and more consistent.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=123.41) [The goal is to standardize on a common approach for integrating diverse data](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=126.64) [sources with standard producer and consumer applications.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=132.03) [This is awesome because writing highly performant and reliable consumers,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=135.65) [for example, can be really hard and complex, as we covered in the last module.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=139.85) [So having a framework to simplify and standardize this is](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=144.31) [a huge step forward, Now currently,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=148.3) [many of the developers using this framework are those that work for the](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=151.1) [leading technology data providers who have started to include a Kafka](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=155.31) [connector as part of their product roadmaps.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=159.45) [Oracle and HP are some noteworthy examples of this.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=162.29) [Currently,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=165.87) [there are over 50 platform connectors available that are designed to](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=166.59) [connect to many different products and services,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=170.2) [and that list is growing.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=172.94) [Confluent itself has created many of these connectors,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=174.44) [and they also provide an online portal they call the Connector Hub.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=177.93) [They invite anyone and everyone to develop and contribute a Kafka connector](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=182.14) [using the API and that online portal for distribution.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=187.14) [As adoption grows,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=191.13) [this is bound to drive more consistency and greater productivity](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=192.67) [in Kafka‑based data integration initiatives.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=196.73) [Overall,](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=199.84) [it's going to get cheaper and faster than ever to get Kafka integrated in enterprises.](https://app.pluralsight.com/course-player?clipId=40618ec2-6d55-4db1-87b6-c0981e248293&startTime=200.75)

### [Challenges and Solutions for Fast Data](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e)

[Within the last few years,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=0.04) [there has been a lot of hype around predictive analytics,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=1.8) [machine learning, real‑time, stream‑based,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=5.02) [whatever buzzword of your choosing.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=7.73) [There are multiple technology platforms that all propose to](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=10.64) [offer a unique ability to deliver upon this hype for](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=14.01) [real‑time or stream‑based analytics.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=17.74) [Some of these platforms are legitimate, viable solutions such as Apache Storm,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=21.04) [Hadoop, Cassandra, and Apache Spark.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=26.23) [Again, Kafka is generally found in the middle.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=29.54) [But the problem is each one of these technologies introduces a unique and](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=32.47) [mostly complex model for development and operation.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=38.34) [Each have their own API and cluster‑based management](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=41.91) [approach to distributed systems.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=45.15) [Kafka, itself, as we've covered in this course,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=47.84) [has its own API and vast cluster‑based model.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=50.81) [So if you have all of these technologies under the same roof,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=54.44) [so to speak,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=58.15) [that's a tremendous amount of technology to manage and](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=59.26) [maintain all for the same goal of achieving the ability to](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=61.97) [process and analyze data in real time.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=65.75) [Touching on the last challenge,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=68.69) [this introduces consistency and productivity](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=70.44) [challenges and integrating it all together.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=73.24) [With Kafka generally being positioned between these](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=75.86) [technologies for integration,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=79.25) [it would need an army of producers and consumers to](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=81.14) [keep the streaming pipes flowing.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=84.37) [The challenge here is pretty obvious.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=87.24) [Now I'm not saying all of these different platforms](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=89.58) [are present in each environment,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=92.91) [but many are because they each have their own strengths and](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=95) [advantages that complement the weaknesses of the other.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=98.43) [But regardless of whatever of these systems come and go,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=101.83) [one thing is becoming more consistent.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=104.84) [And that is the place Apache Kafka finds itself within these organizations.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=107.14) [The 0.10 release of Apache Kafka was a huge one.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=112.34) [In addition to Kafka Connect,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=116.51) [a new client library for real‑time, stream‑based processing was introduced.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=118.51) [This library is called Kafka Streams.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=123.74) [The real value proposition of this is that for organizations that](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=126.95) [have already made an investment in Apache Kafka,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=131.27) [they can now have streaming data capabilities without having to install,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=134.73) [run, and maintain all of those different platforms.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=139.13) [All they need is their existing Kafka environment.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=142.82) [Given everything we've learned about Kafka in this course,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=146.74) [I am sure you can see how adding this capability to](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=150.02) [Kafka wasn't that much of a stretch.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=152.92) [I mean,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=155.44) [consider what Kafka already does with data in motion and how it does it. This](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=156.01) [is significant because it doesn't require anything more.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=161.4) [I mean, theoretically,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=164.96) [Apache Kafka could be the only infrastructure solution required. But in reality,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=166.45) [many enterprises have good reason to additionally invest in Apache,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=171.9) [Hadoop, and Spark.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=175.83) [So it may be that Kafka itself isn't the only big data](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=177.17) [system in place. But at the very least,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=181.18) [it can be the only system needed for stream‑based processing.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=183.95) [Regardless,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=187.69) [the potential to reduce and consolidate into fewer](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=188.53) [systems is now a very real possibility.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=191.9) [Think of what that can do to lower the initial](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=194.99) [investment and overall total cost.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=197.82) [As I said, Kafka Streams is a client library that works with the Kafka cluster.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=200.42) [As you've learned in this course,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=206.44) [that's exactly like Kafka producers and consumers.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=208.09) [They are client libraries too. And just like we did with the](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=211.26) [producer and consumer client libraries,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=214.69) [Kafka streams can be embedded within Java‑based applications,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=216.86) [making the barrier to adopt lower than any other](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=221.11) [platform offering stream‑based processing.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=224.16) [Think of it this way.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=226.84) [If you already have producers and consumer applications,](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=228.17) [why not just extend them with the Kafka Streams library to provide](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=231.36) [stream‑based processing capabilities all within the same place? This is an exciting area that I hope you'll continue to explore.](https://app.pluralsight.com/course-player?clipId=25b17350-e676-4947-bb1d-751c17a55a4e&startTime=235.75)

### [Apache Kafka's Ecosystem and Summary](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673)

[Everything that Kafka is today,](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=0.14) [and what it will be tomorrow, is made possible through the growing and](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=2.29) [healthy ecosystem of adopters and source code contributors.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=6.15) [These are but a few of the big names that not only have based](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=10.41) [significant parts of their business on Apache Kafka but also](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=13.95) [make generous contributions back,](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=18.12) [allowing all companies and organizations, big and small, to benefit.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=20.44) [That's the beauty of the open source ecosystem in which Apache Kafka is](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=25.52) [firmly placed. In this module we covered the undeniable success that](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=30.62) [Apache Kafka has had since the beginning.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=35.73) [It has enabled organizations to solve some of data](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=38.84) [management's biggest problems.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=41.72) [But as I said, in the process, it has introduced new challenges.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=44.14) [Many of the challenges facing Kafka and the data management industry in](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=48.42) [general stem from the rapidly growing and changing landscape.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=52.62) [Data volumes, velocity and variety are increasing exponentially,](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=57.14) [and as a significant player in this landscape,](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=62.06) [Apache Kafka can't rest on its laurels.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=64.87) [Luckily, with the vast and supporting ecosystem it has,](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=68.04) [Kafka has evolved to meet these challenges and establish reinforced](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=71.77) [foundations upon which to build further for many years to come.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=75.98) [We discussed some of these recent innovations, like the Schema Registry,](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=80.64) [Kafka Connect and Kafka Streams.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=84.82) [I hope you'll agree with me that there is a promising future](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=87.75) [ahead for Kafka and the many technology professionals and](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=91.27) [organizations that invest in it.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=94.41) [We have come to the end of this course.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=96.57) [I hope you learned a lot, at least enough to](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=98.61) [continue your journey in learning more.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=101.12) [It's always hard to decide where to invest your limited amount of time.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=103.82) [I personally faced this challenge, as the course author, in](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=108.4) [determining what details to focus on and what details to](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=111.95) [sacrifice, because I wanted you to get the most out of this](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=115.58) [course within a limited amount of time.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=118.87) [I hope I succeeded, but it is hopefully just the beginning for you.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=121.54) [I encourage you to continue learning about Apache Kafka and trying it out. It's a solid bet to make as a technology professional.](https://app.pluralsight.com/course-player?clipId=6c6a6f59-9b5a-49de-822a-a694db284673&startTime=125.9)