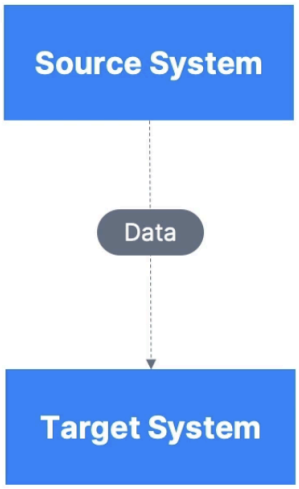
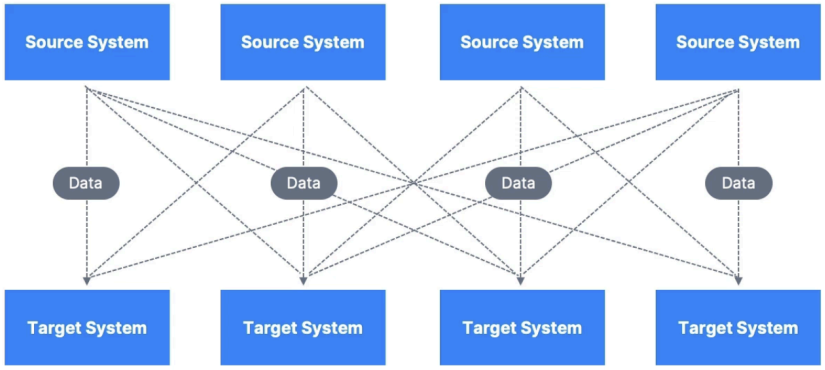
How companies start



Simple at first!

==========
==========

After a while...



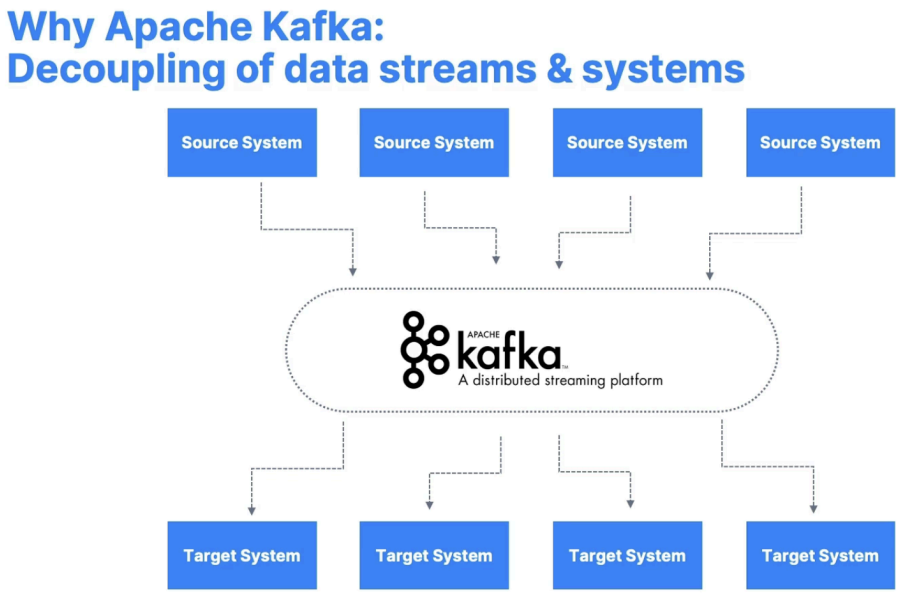
Very complicated!

==========
==========

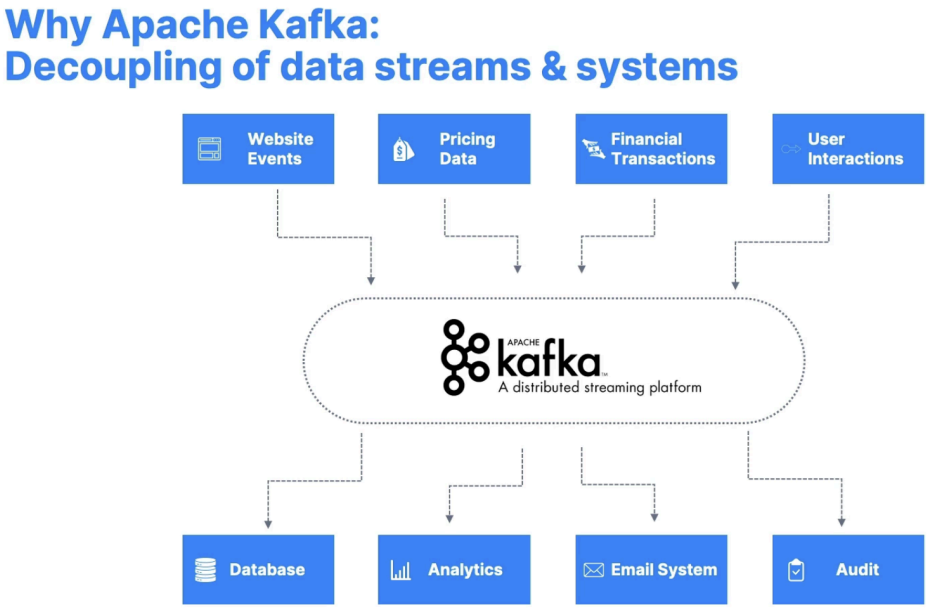
- If you have 4 source systems, and 6 target systems, you would need to write 24 integrations!
- Each integration comes with difficulties.
    - Protocol: how the data is transported(TCP/HTTP/REST/FTP/JDBC
    - Data Format: how the data is parsed(Binary, csv, json, avro, protobuf…)

    - Data schema and validation: how the data is shaped and may change
- Each source system will have an increased load from connections.

==========
==========



==========
==========



Why Apache Kafka?

- Created by LinkedIn, now open source project and maintained mainly maintained by Confluent, IBM, Cloudera
- Distributed, Resilient architecture, fault tolerant.
- Horizontal scalability:
    - Can scale to 100s of brokers
    - Can scale to millions of messages per second
- High performance (latency of less than 10ms) ~ real time
- Used by 2000+ firms, 80% of fortune 100
- Consumers: airBnb, Netflix, LinkedIn, Walmart, Uber

Use Cases:

- Messaging System
- Activity Tracking
- Gather metrics from many different locations
- Application Logs Gathering
- Stream processing(with Kafka streams Api for example)
- Decoupling of system dependencies
- Integration with Big Data tools like spark, Flink, Storm, Hadoop.
- Microservices pub/sub

For Example:

- **Netflix** uses Kafka to apply recommendations in real-time while you're watching TV shows
- **Uber** uses Kafka to gather user, taxi, and trip data in real time to compute and forecast demand, and compute surge pricing in real time.
- **LinkedIn** uses Kafka to prevent spam, collect user interactions to make better connection recommendations in real time.
- **Remember that Kafka is only used as a transportation mechanism, which allows HUGE data flow in your company!**