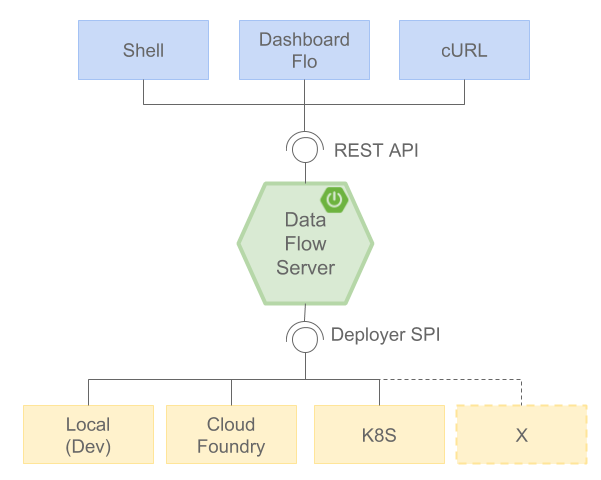
Spring Cloud DataFlow

Spring Cloud DataFlow用于开发和执行大范围数据处理包括ETL、批量计算和持续计算的统一编程模型和托管服务，其简化应用程序的开发和部署而将精力集中在数据处理的用例上。Spring Cloud DataFlow是Spring XD升级版，继承了Spring XD优点的同时提供了基于云技术的原生方式来提供结合流计算和批量计算的混合计算模型。开发者可以通过Spring Cloud DataFlow，在数据获取、实时分析、批处理等常见用例中执行数据流的创建和编排。Spring Cloud DataFlow的主要架构如下所示：



其使用原生云平台引进了新的服务提供上接口（SPI），如Cloud Foundry、K8S、Apache Yarn等现有运行环境，在这些运行平台上执行基于微服务的分布式流处理及批处理Pipeline

1. 使用DSL(领域特定语言)来描述多个应用数据流的流转逻辑，通过Shell/Dashboard/Curl提交给DataFlow Server。
2. DataFlow Server负责部署应用程序到运行环境，其是一个嵌入式的Servlet容器并提供rest API来创建、部署、卸载和销毁Streaming/Task程序，并查询运行状态、数据分析等。在Spring中，Streaming和Task程序都是基于微服务并将Spring boot作为基础类库，这些微服务都支持健康检查、安全、可可配置日志、监控及管理等，程序打包成可执行JAR包
3. DataFlow Server通过Deployer SPI(服务提供上接口)将程序提交到不同的运行环境，比如Lattice、Cloud Foundry、YARN及Local中，Deployer SPI在这些环境中部署模块的抽象层，用于启动DataFlow作业。
4. Spring Cloud DataFlow编程模型

Spring Cloud Stream

Spring Batch

1. Spring Cloud Deployer

2.1 spring-cloud-dataflow-server-yarn

app register --name timestamp --type task --uri maven://org.springframework.cloud.task.app:timestamp-task:1.3.0.RELEASE

task create --name myStamp --definition "timestamp"

task launch myStamp

https://docs.spring.io/spring-cloud-dataflow/docs/current/reference/htmlsingle/#spring-cloud-dataflow-launch-tasks-from-stream

https://docs.spring.io/spring-cloud-stream-app-starters/docs/Celsius.SR2/reference/htmlsingle/#spring-cloud-stream-modules-load-generator-source

https://www.aliyun.com/jiaocheng/802567.html

https://docs.spring.io/spring-cloud-dataflow/docs/current/reference/htmlsingle/