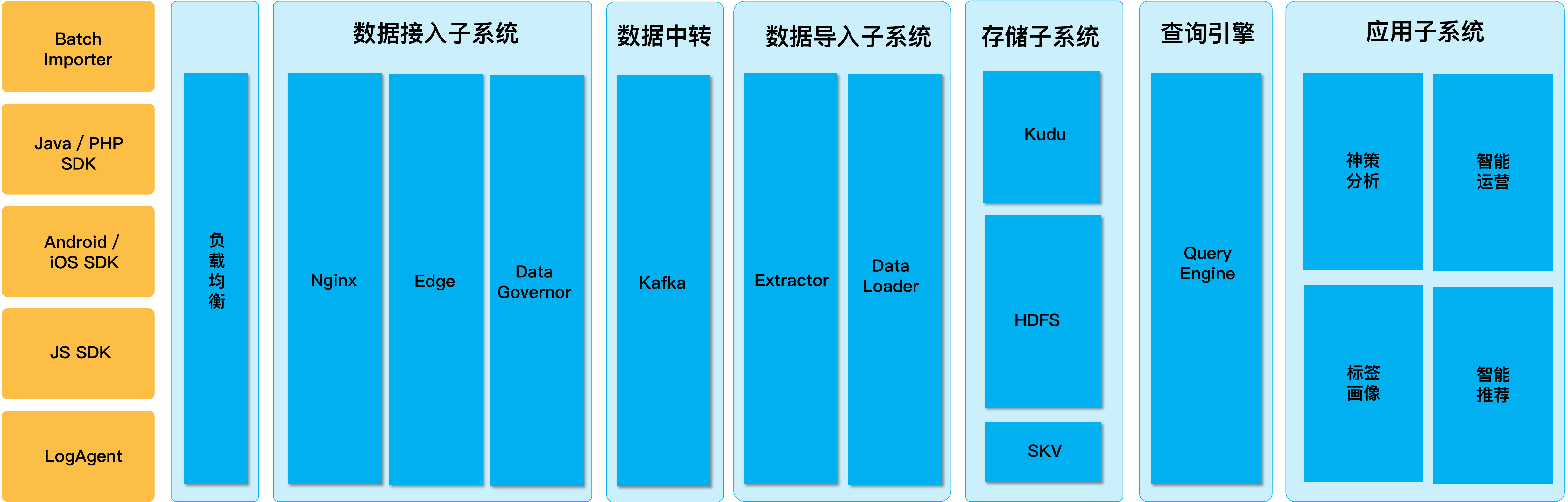
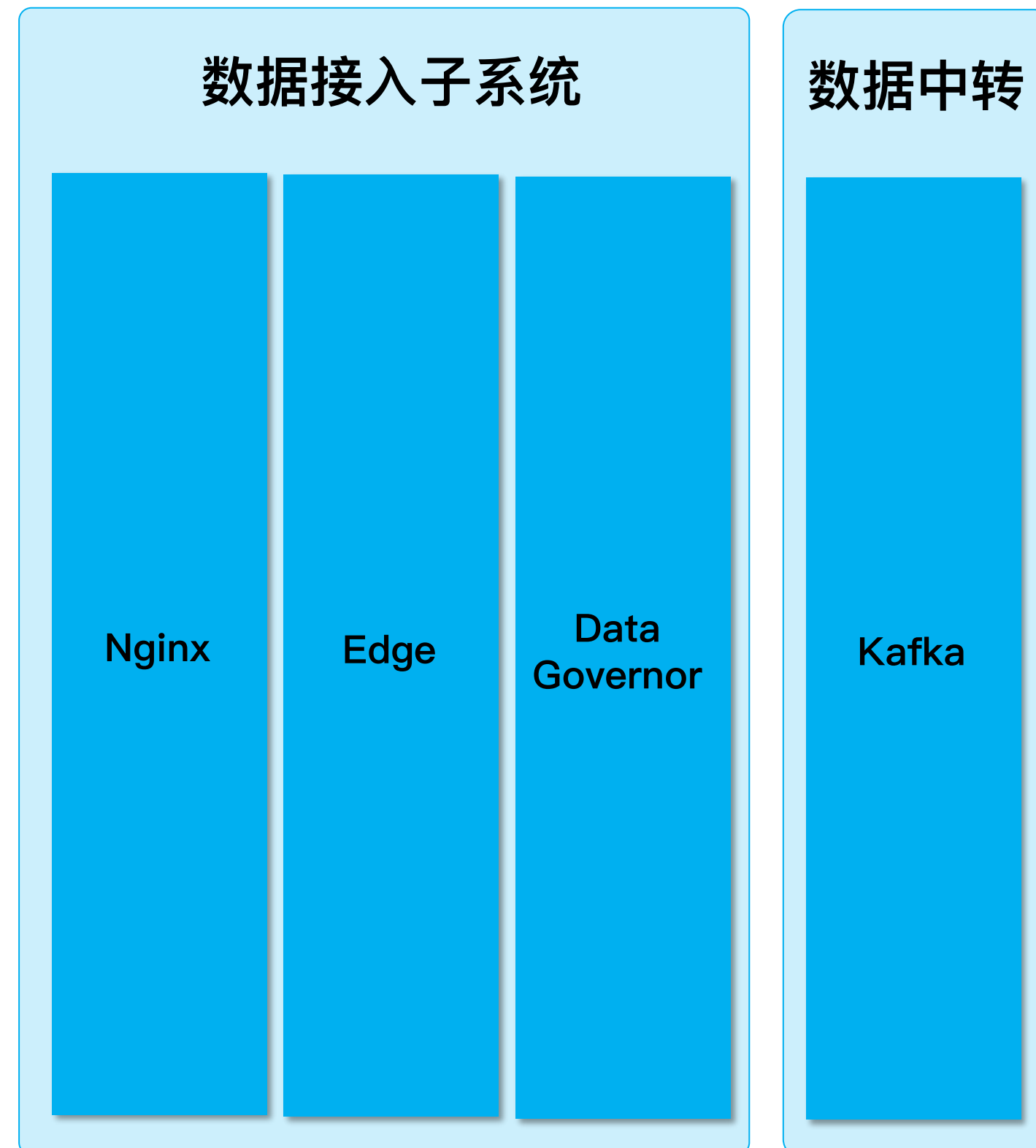


# 神策技术架构

## 数据流向

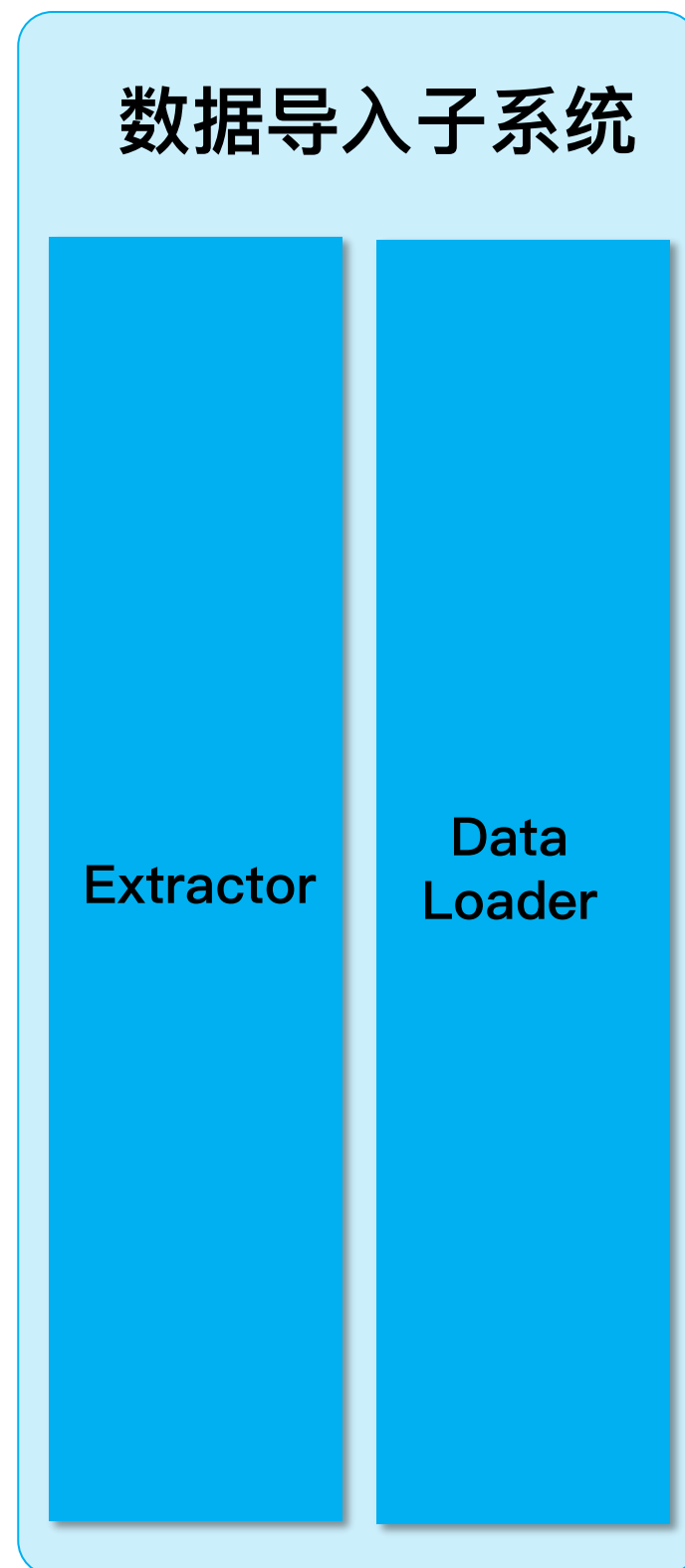


# 神策技术架构



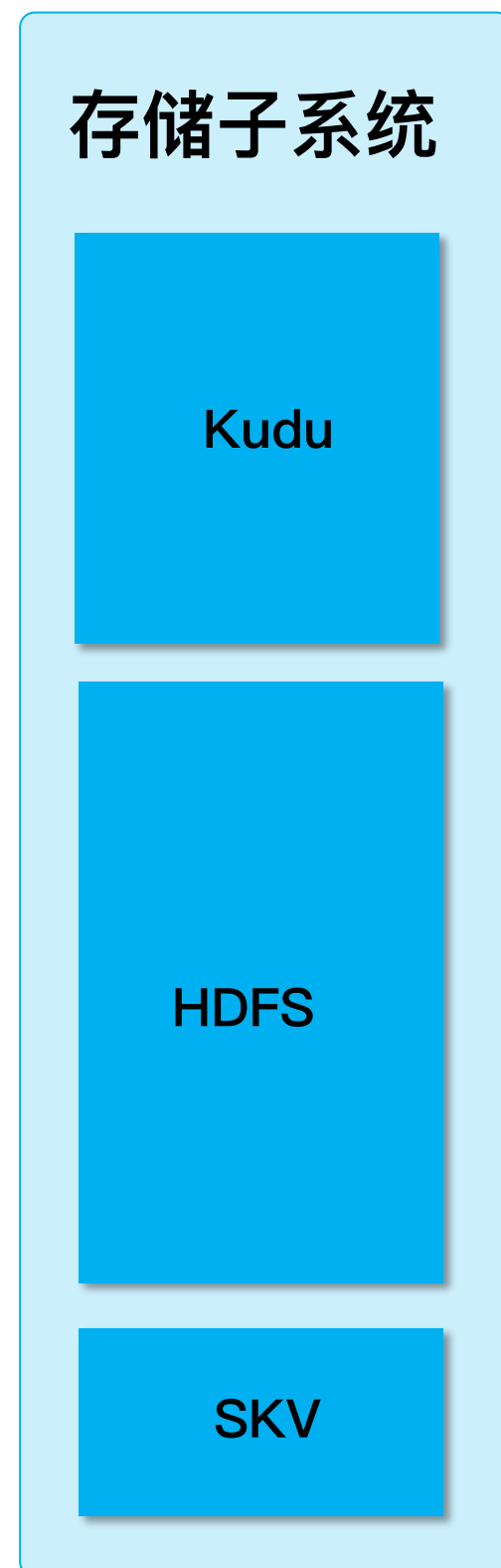
- ***Nginx***
- 数据接收，落盘
- ***Edge***
- 数据初步处理分发
- ***Data Governor***
- 数据治理组件，对应神策的SDG产品
- ***Kafka***
- 消息队列组件，实现数据流上下游的解耦，开放数据订阅

# 神策技术架构

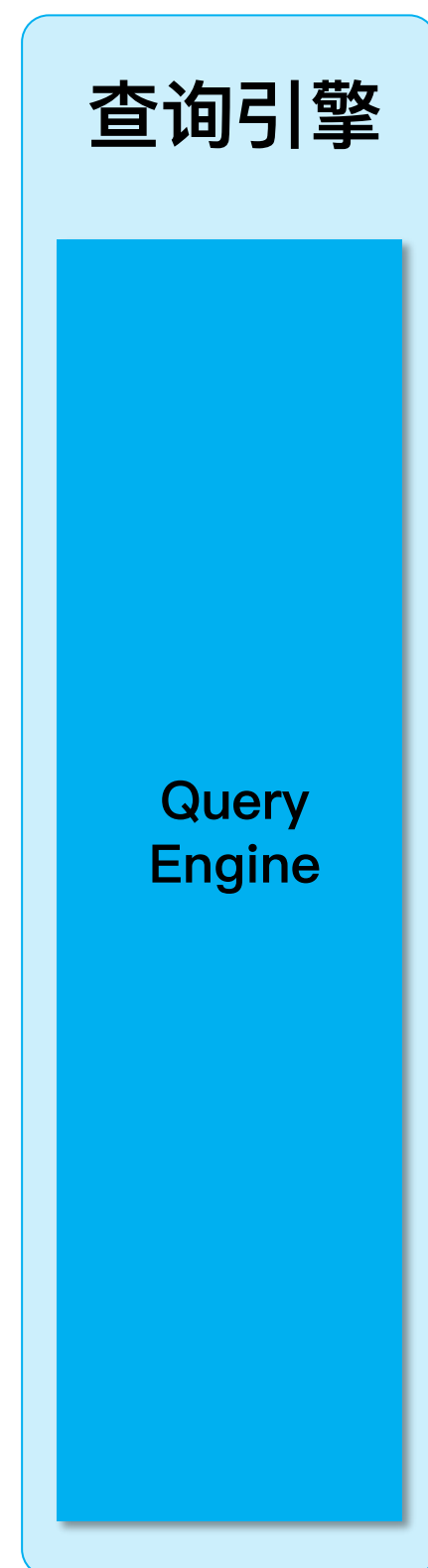


- ***Extractor***
  - 神策预处理模块
  - ID-Mapping
  - 广告渠道信息匹配、首日首次修正
- ***DataLoader***
  - 实时消费Kafka的数据，并写入Kudu表
  - 定期将Kudu数据写入Hdfs，以Parquet列式存储格式存储

# 神策技术架构



- ***Kudu***
- 分布式数据库，写优化存储
- 用户数据持久化存储，事件数据热存储 / 临时存储
- ***Hdfs***（单机版无Hdfs）
- 分布式文件系统，读优化存储
- 事件数据持久化存储，采用parquet列式存储
- ***SKV***
- 全名叫Sensors Key Value，神策 KV 数据库
- 在线标签服务数据、Id-mapping数据、首日首数据



- ***Query Engine***
- 将Web查询请求翻译成 SQL
- 发起SQL查询，访问 Kudu 和 Hdfs 组成的 view
- ***性能优化***
- 事件数据按日期、事件名称分区存储
- 采用列式存储
- 文件内按用户id有序存储
- 特定分析模型查询逻辑优化