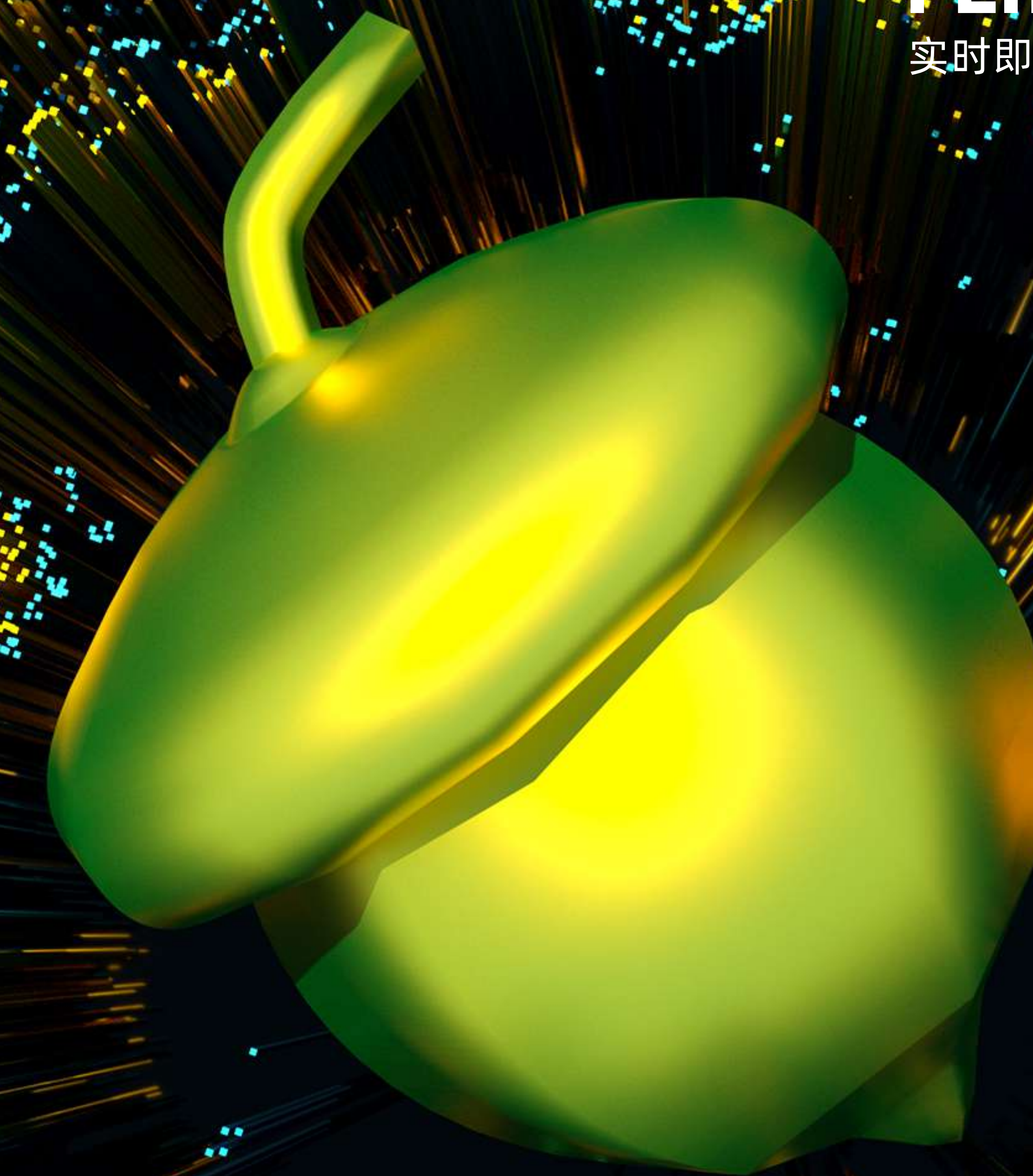


Flink Heading Towards a Unified Engine

北京·国家会议中心
2019.11.28-11.30

王峰（莫问）

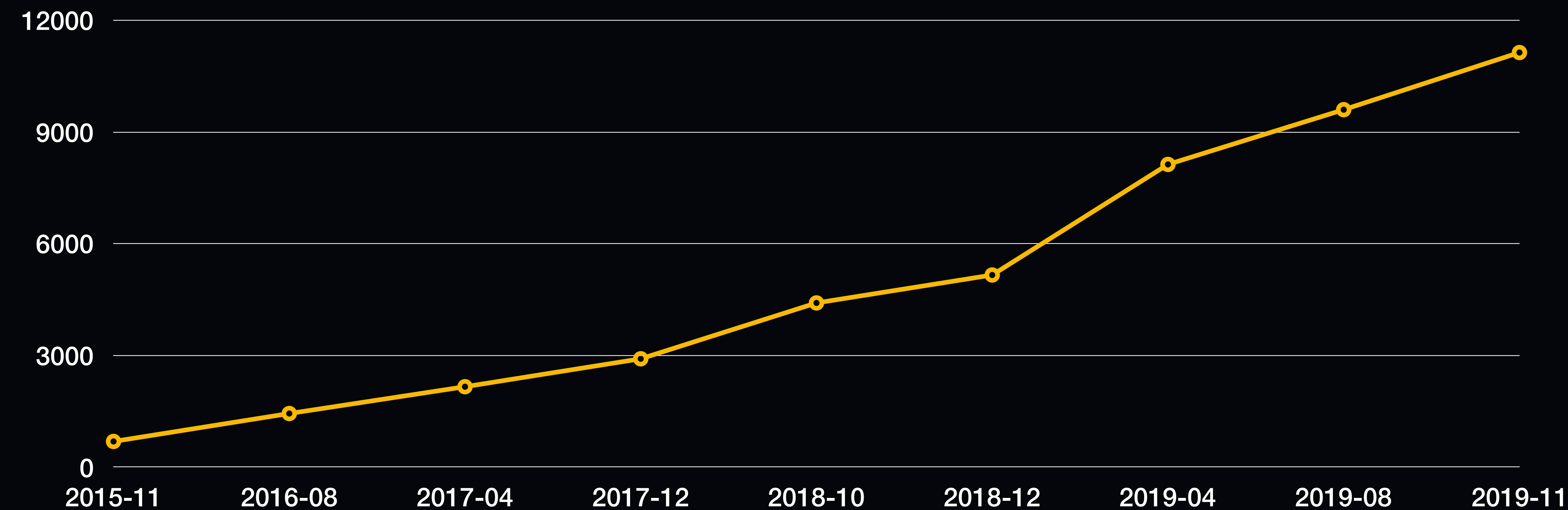
阿里巴巴实时计算负责人



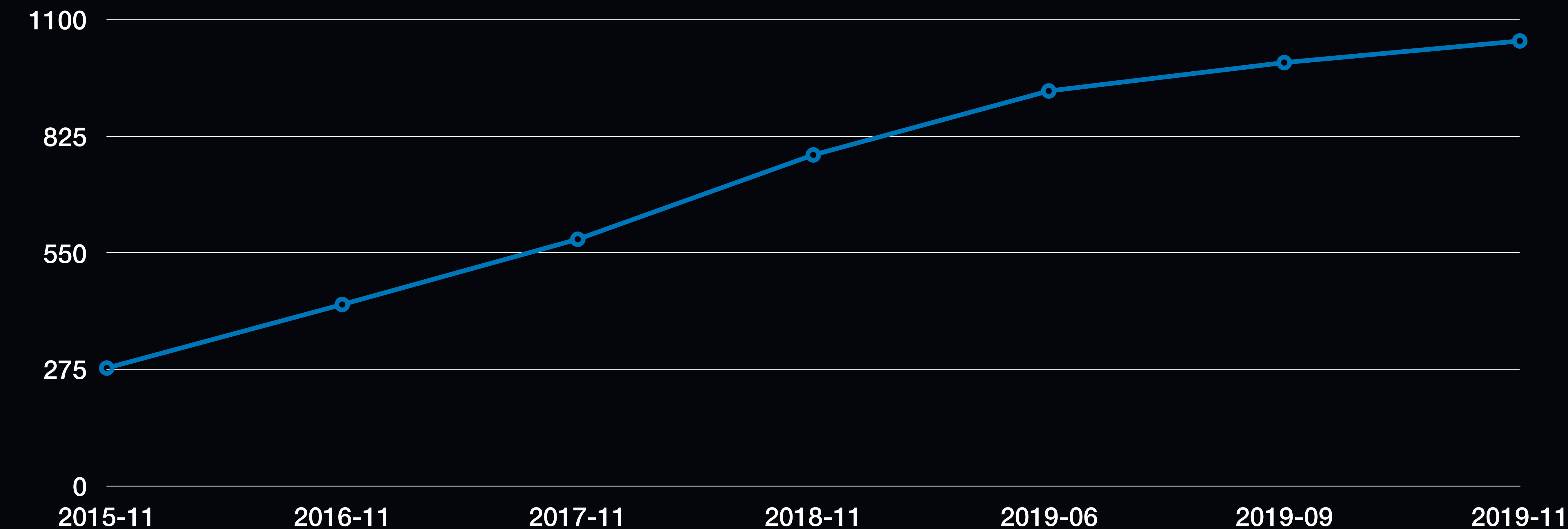
Flink - The Most Active User + Dev Mailing List Apache Project

<https://s.apache.org/FY2019AnnualReport>

Flink Github Stars



Flink Contributors



Apache Flink 在中国的应用

汽车之家
看车·买车·用车

bilibili

DiDi

Tencent 腾讯

Alibaba Group

蚂蚁金服
ANT FINANCIAL

Baidu 百度

iQIYI 爱奇艺

唯品会
品牌特卖

奇安信
QI-ANXIN

58同城

快手

美团
美团点评

ByteDance
字节跳动

mi

斗鱼
DOUYU.COM

虎牙直播
HUYA.COM

趣

千寻位置
QXWZ.COM

贝壳
找房大平台

oppo

360

新浪微博

网易
NETEASE
www.163.com

Ctrip
携程

VIP KID

TAL 好未来

众安保险

The Future of Apache Flink



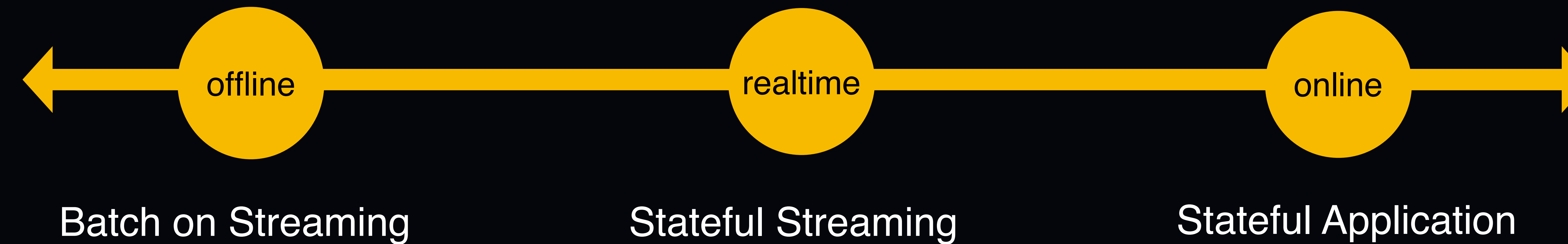
Unified Data Analytics



AI Integration



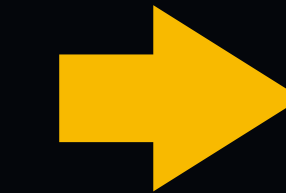
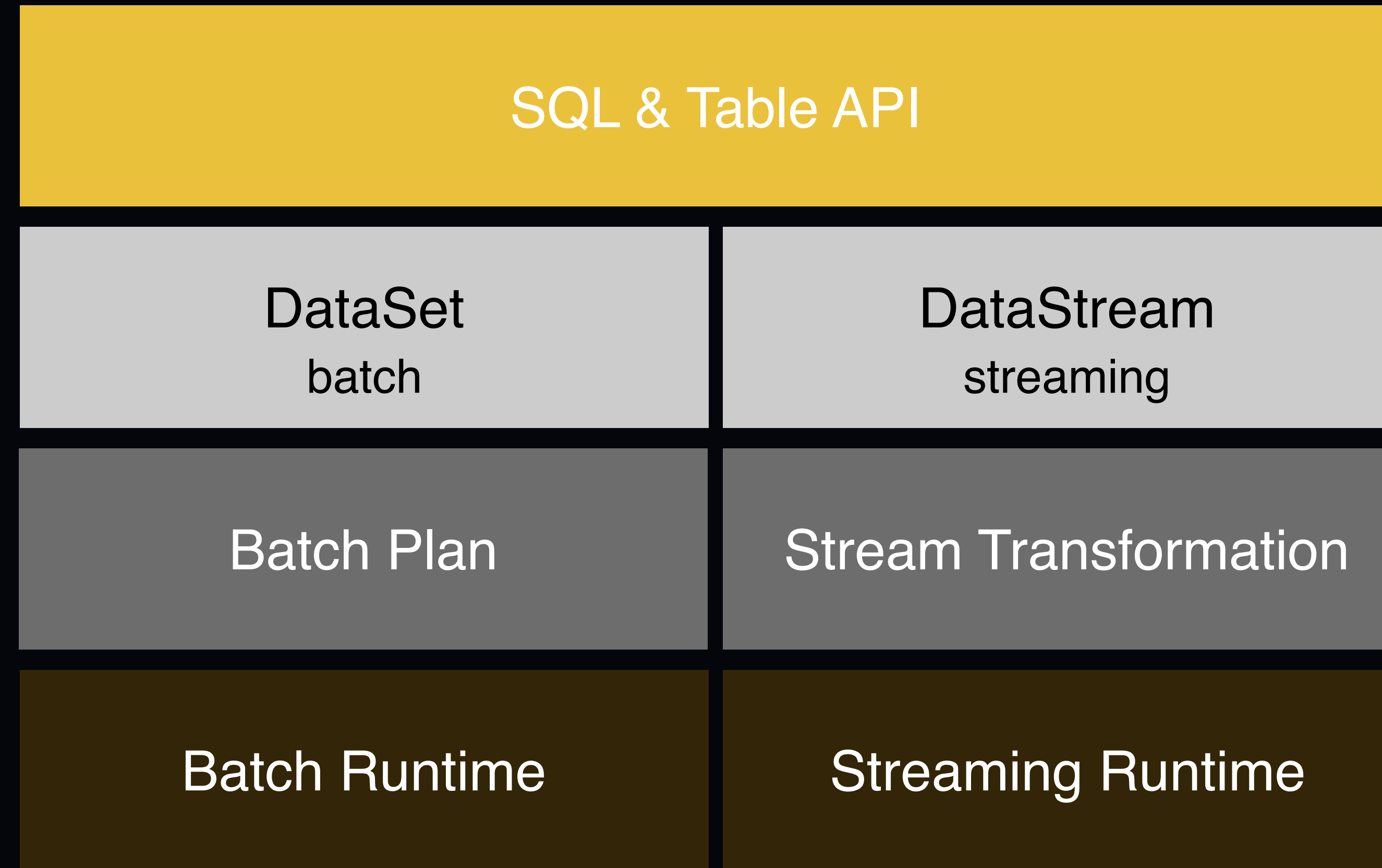
Event-Driven Function



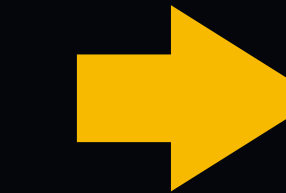
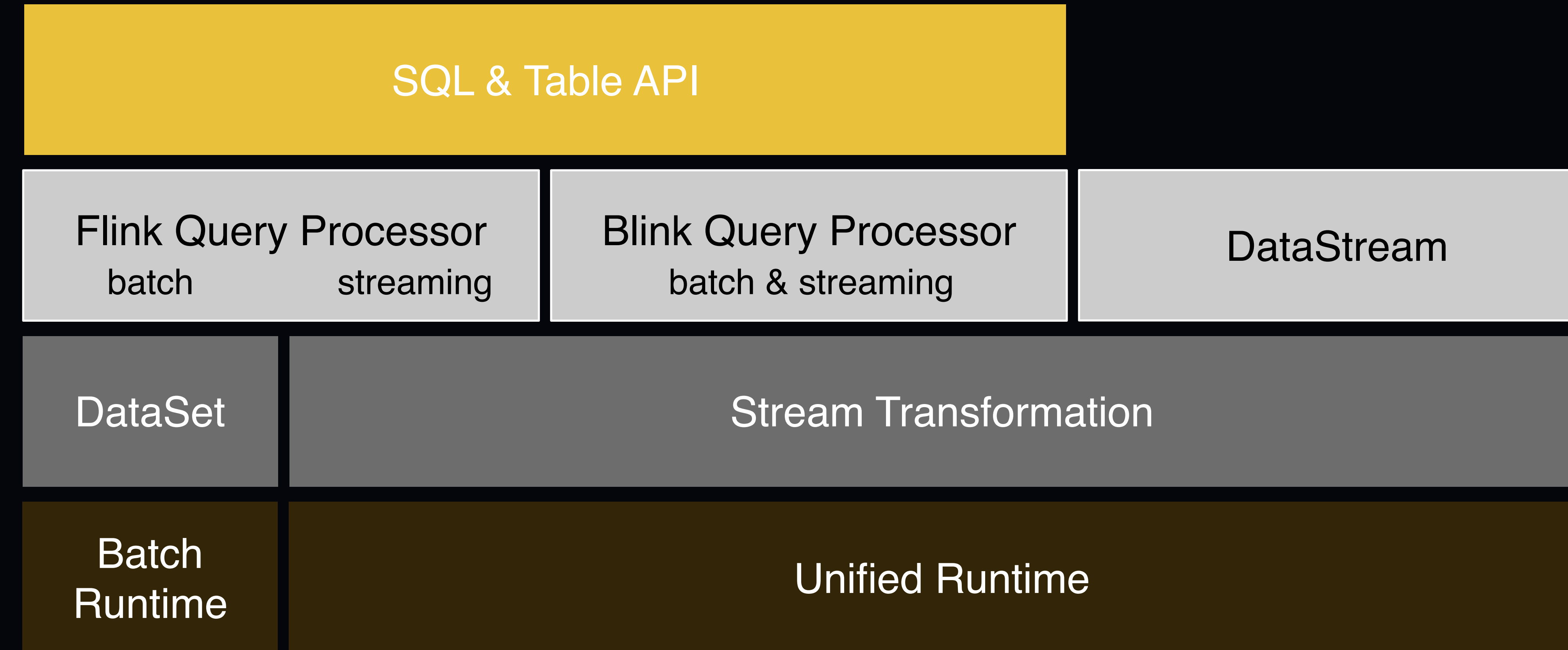
Unified Data Analytics

Flink 批流一体发展路线 (The Roadmap of Flink Unified Architecture)

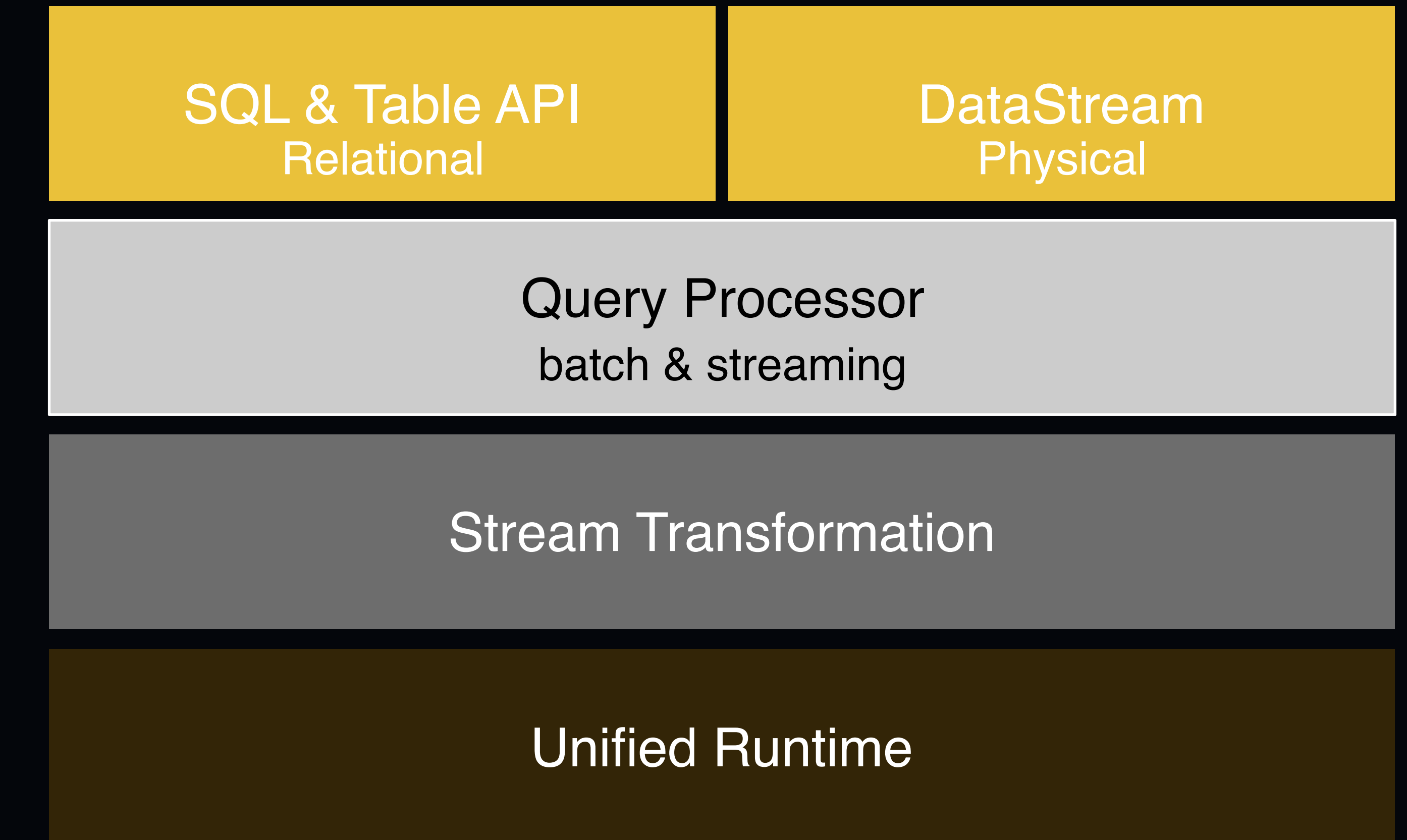
< 1.9



1.9 & 1.10



Future



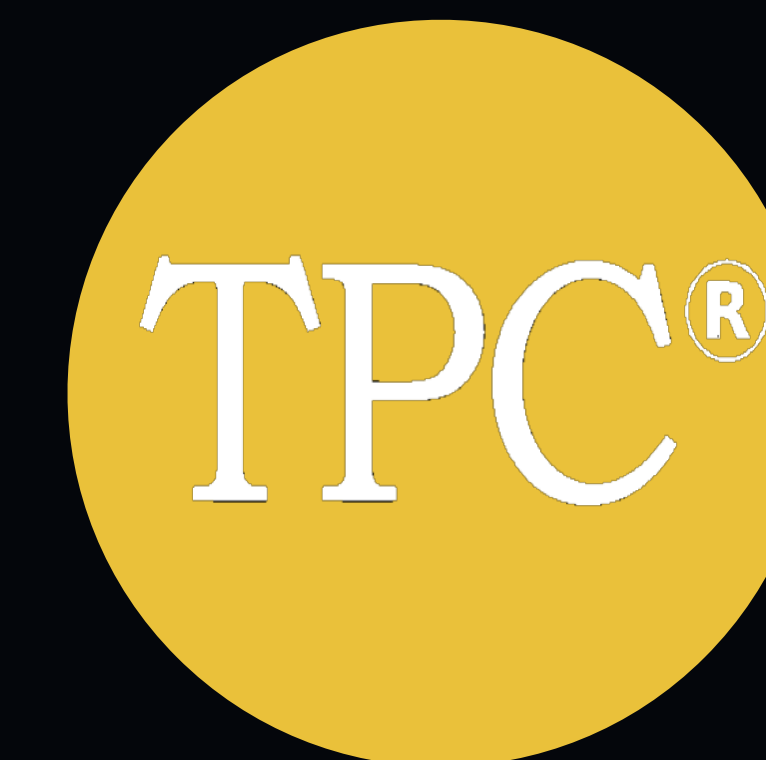
Flink Unified SQL



完整的DDL功能
DDL Functionality



批流统一的Query处理器
Unified Query Processor



通过TPC-H / TPC-DS验证
TPC-H / TPC-DS

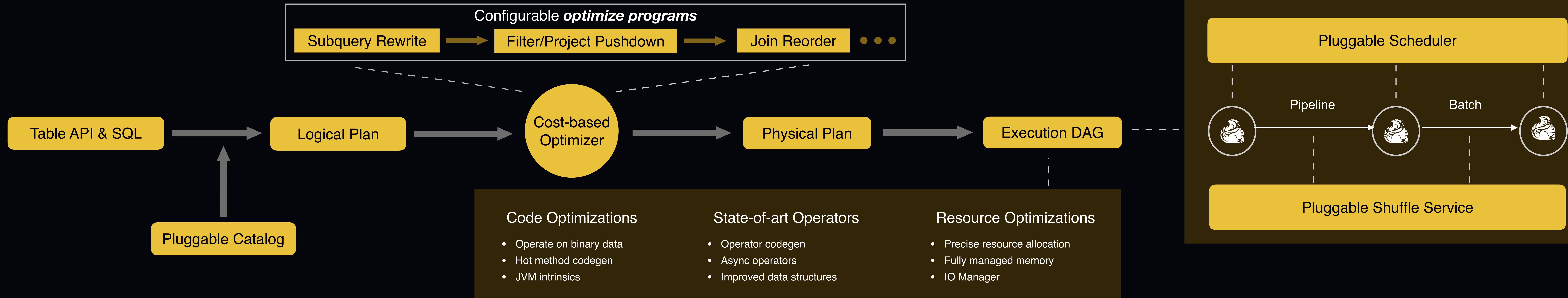


兼容Hive生态系统
Hive Compability



支持Python UDF
Python UDF

Unified SQL Architecture



Flink Integration with Hive Ecosystem



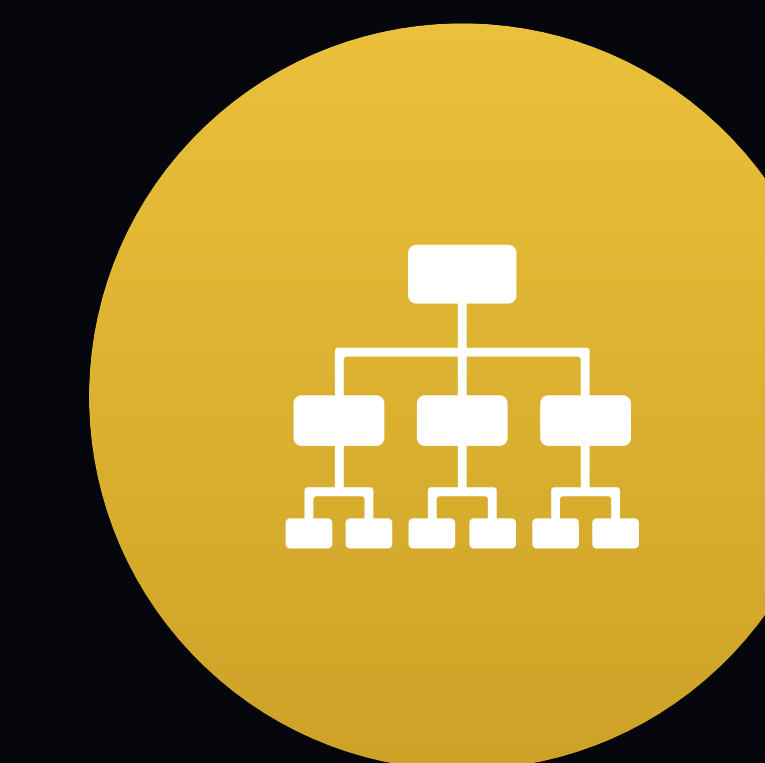
直接读写Hive数据
Directly Read/Write Hive File



无缝对接Hive MetaStore
Hive MetaStore as Catalog



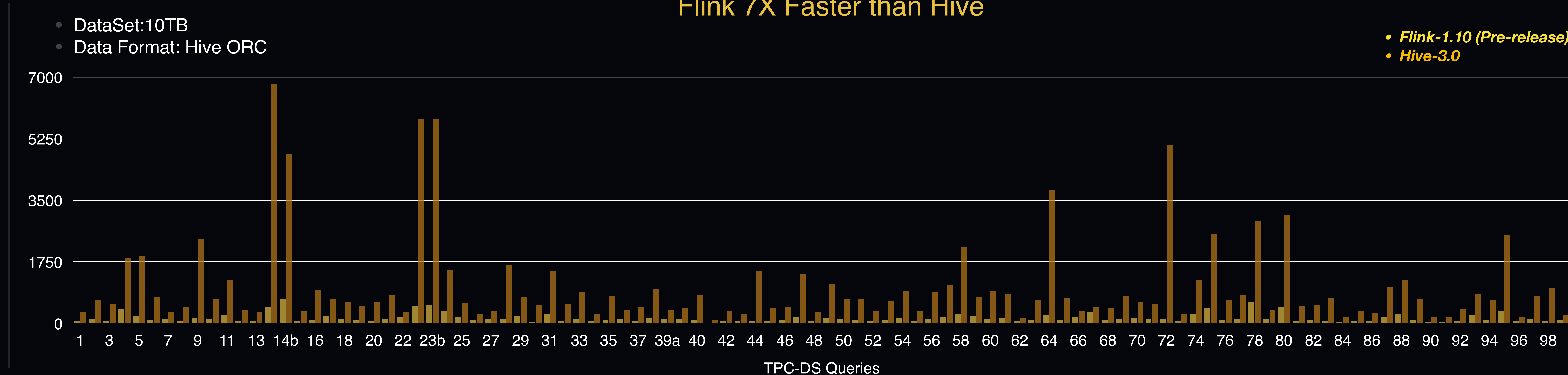
兼容Hive UDF
Compatible with Hive UDF



可与Hive共享资源
Co-Deployment with Hive

TPC-DS Benchmark

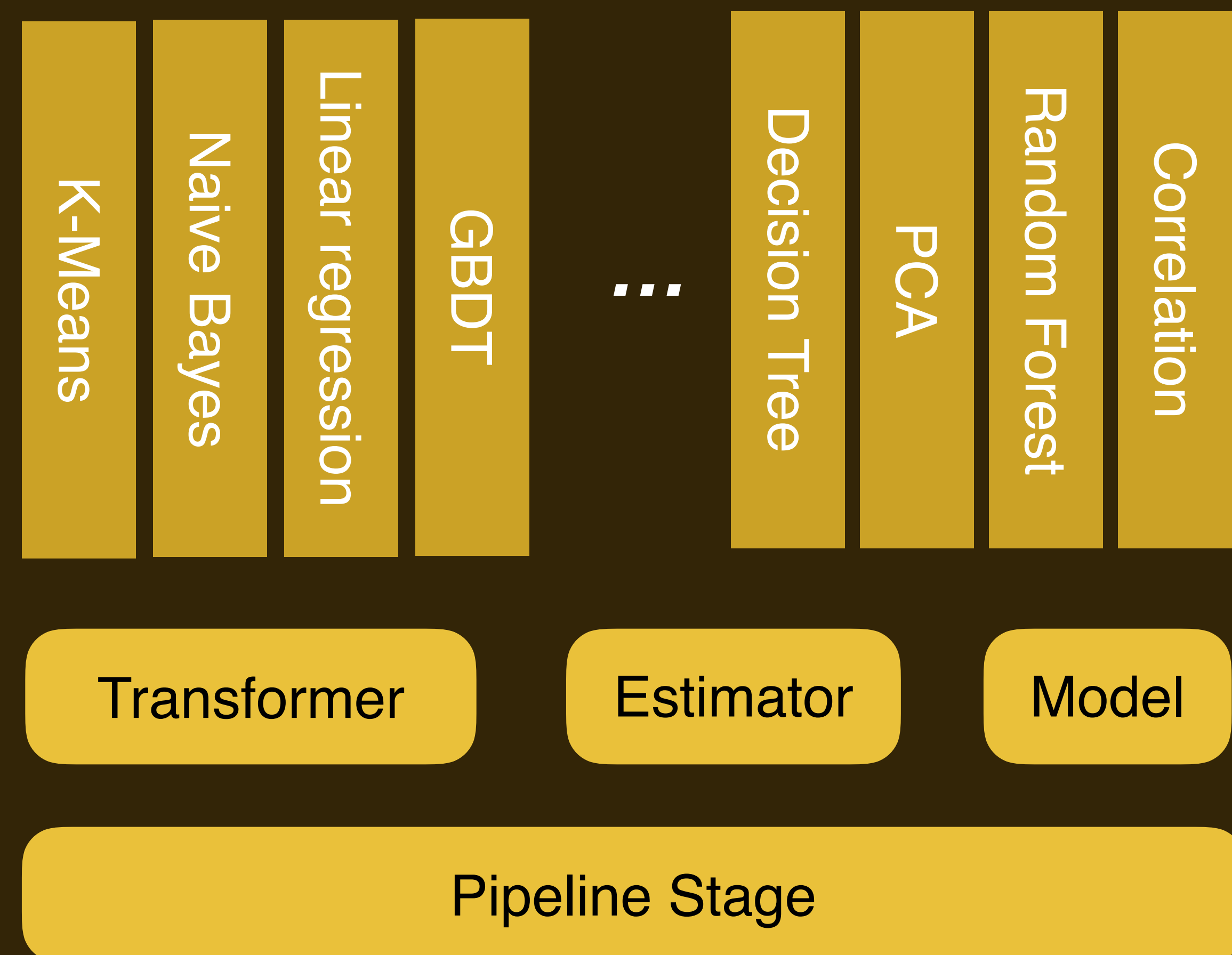
Flink 7X Faster than Hive



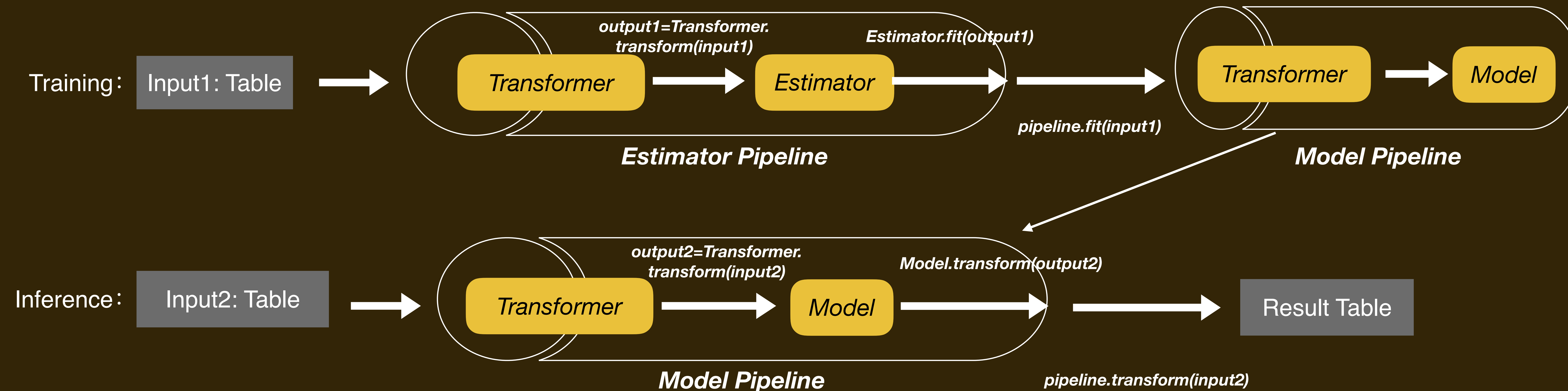
Flink SQL Demo

Flink 拥抱 AI (Flink Embrace AI)

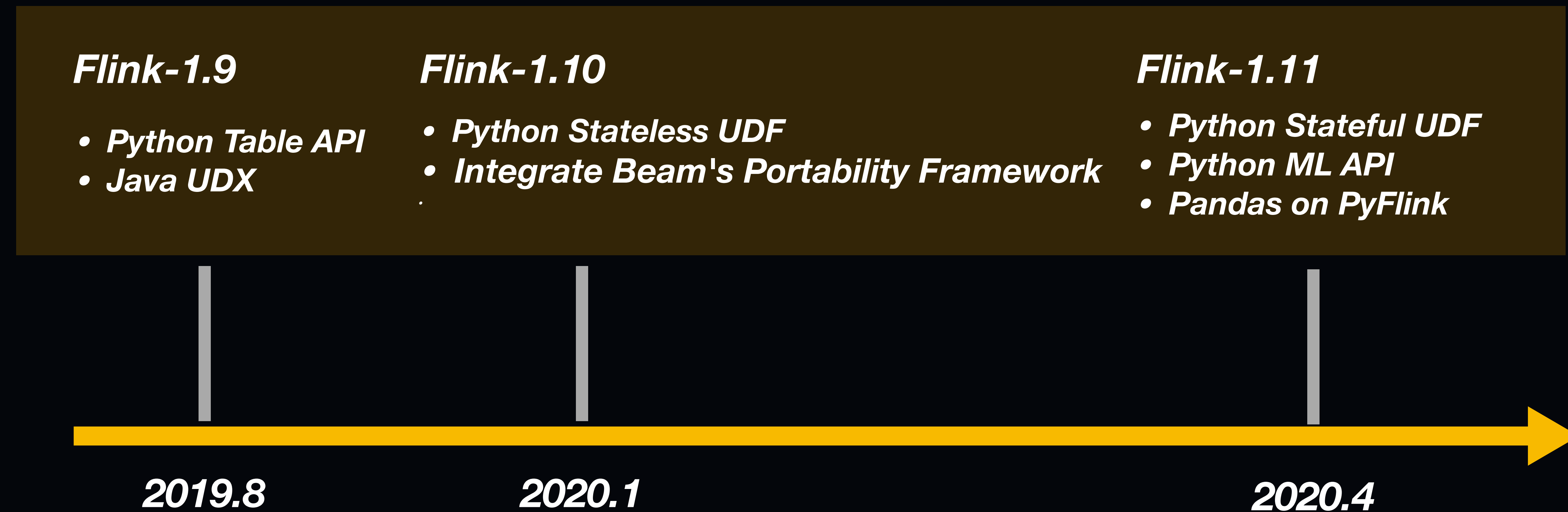
Flink ML Lib



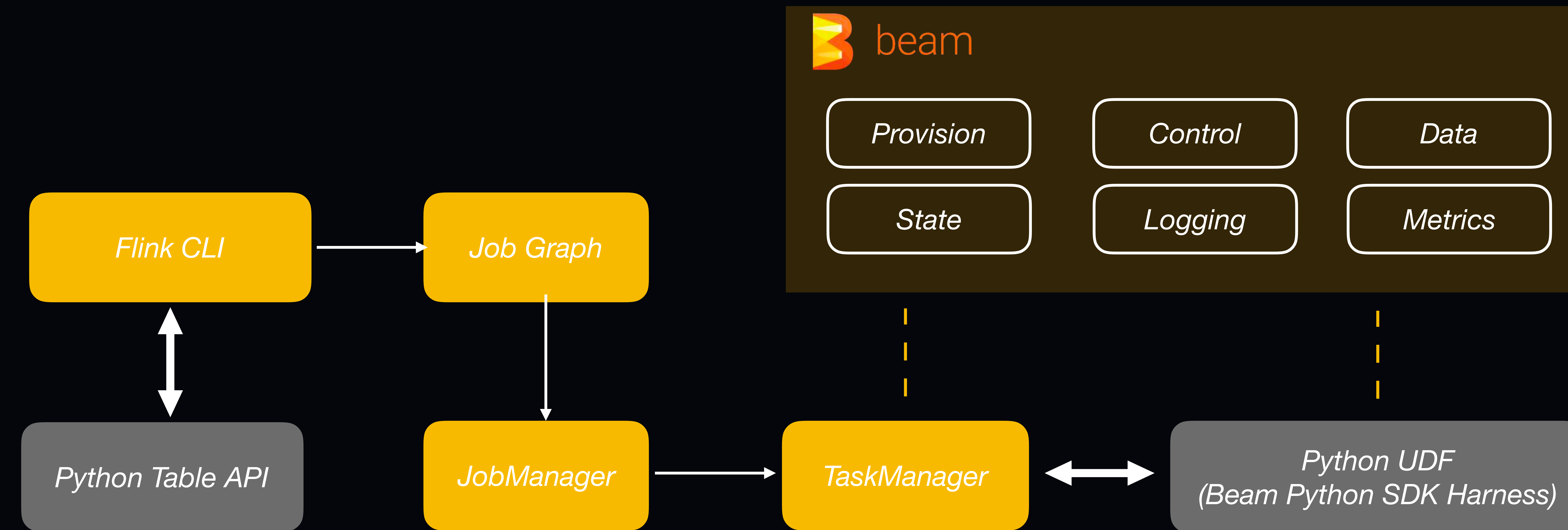
Flink ML Pipeline



PyFlink Roadmap



PyFlink Architecture





Alibaba Alink

- ML Algorithms on top of Flink
- Flink ML Pipeline Interface
- Python API

Open Source:

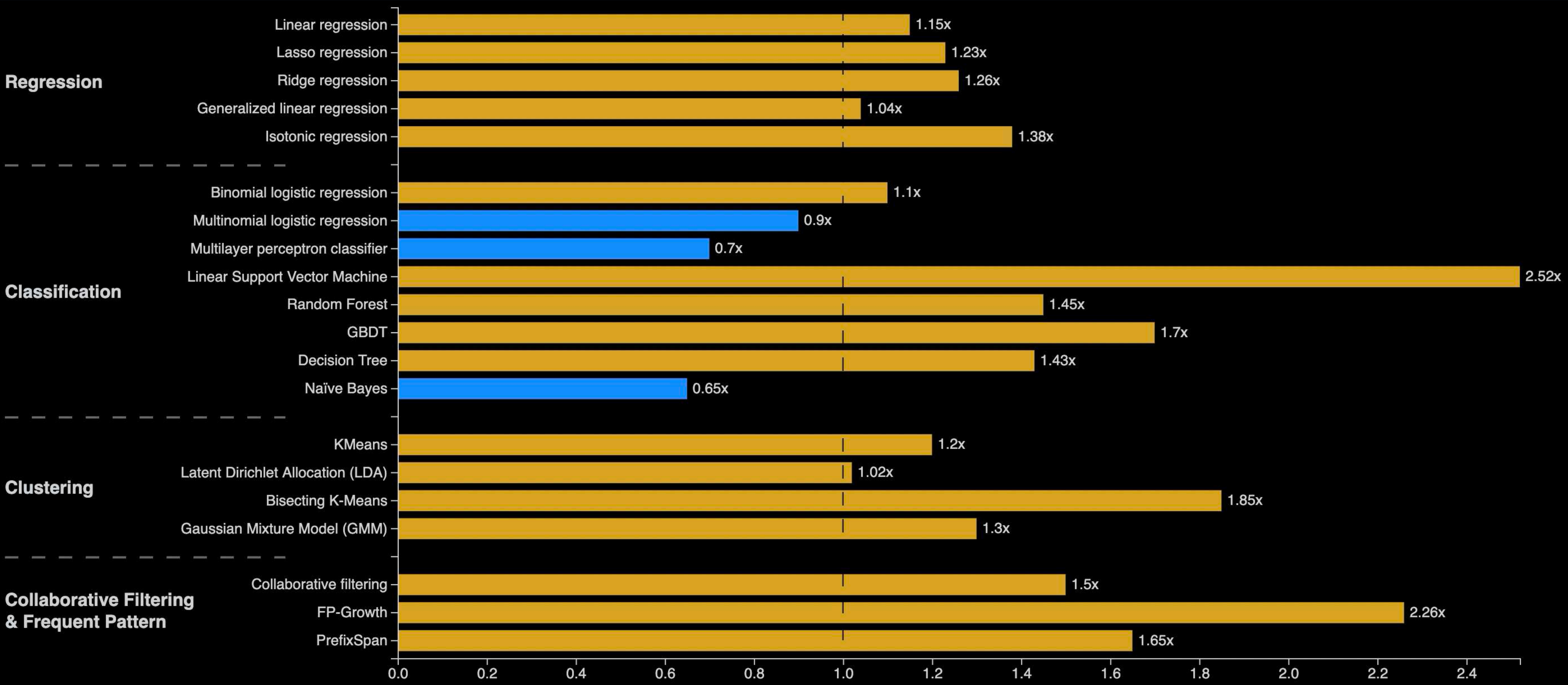
- Contributing to Apache Flink
- New Version of Flink ML

<https://github.com/alibaba/alink>

 Online Learning	 Outlier Selection	 Feature engineering	 Text	 Clustering	 Statistical Analysis	 Evaluation	 Data Processing	 Regression	 Similarity	 Association rule & collaborative filtering	Model Tuning	 Classification
FTRL	SOS	Bucketizer	Word2Vec	KMeans	ChiSquareTest	EvalBinaryClass	VectorInteraction	LinearRegression	ApproxVectorSimilarityTopNLSH	FPGrowth	GridSearchCV	LogisticRegression
		ChiSqSelector	Tokenizer				VectorSlice					Softmax
		Binarizer	NGram				VectorNormalizer	RandomForestRegressor				MultilayerPerceptronClassifier
		OneHotEncoder	Segment	StandardScaler	GbdRegressor	PrefixSpan	LinearSvm					
		FeatureHasher	DocHashCountVectorizer	MinMaxScaler			GeneralizedLinearRegression	GbdtClassifier				
		PCA	DocCountVectorizer	VectorElementwiseProduct	AftSurvivalRegression	ApproxVectorSimillirityJoinLSH	GridSearchTVSplit	DecisionTreeClassifier				
		QuantileDiscretizer	StopWordsRemover	CmdSteamOp	IsotonicRegression			OneVsRest				
				VectorSizeHint	DecisionTreeRegressor			NaiveBayes				
				Imputer								
				IndexToString								
				vectorAssembler								
				DCT								
				StringIndexer								
				VectorPolynomialExpand								

Alink vs Spark ML

	Streaming Mode	Batch Mode	Online Learning	Python API
Spark ML	NO	YES	NO	YES
Alink	YES	YES	YES	YES



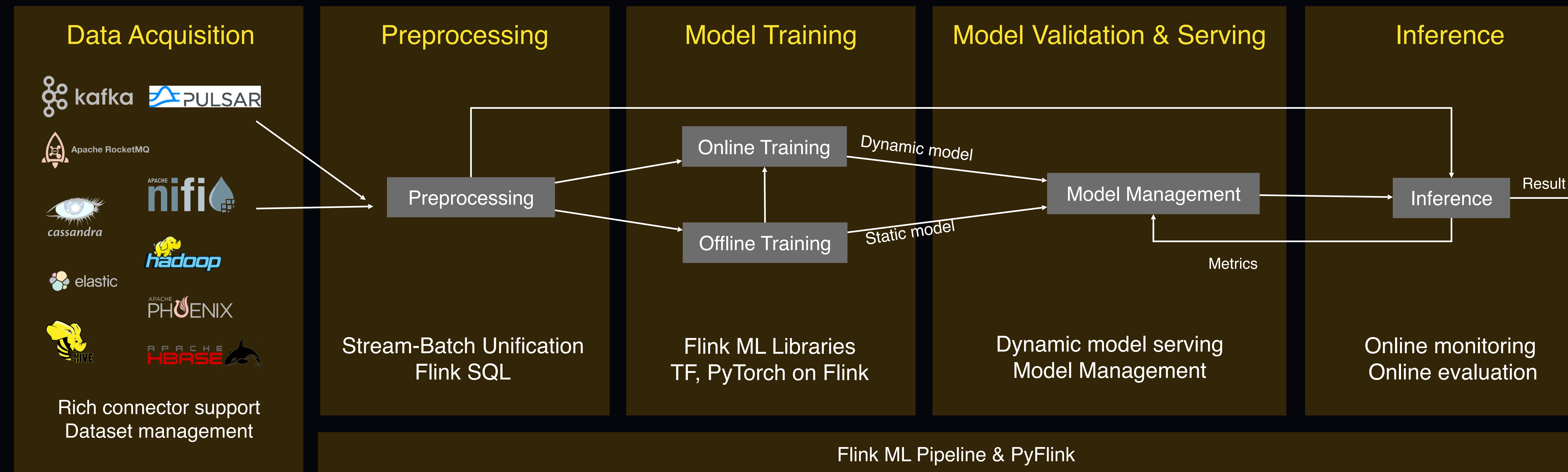
Alink ML Demo

AI Flow - Workflow for Unified Analytics + AI

- Online & Offline Unification
- Engine & Platform-Agnostic
- Components Relation Definitions
- DataSet and IO Specs

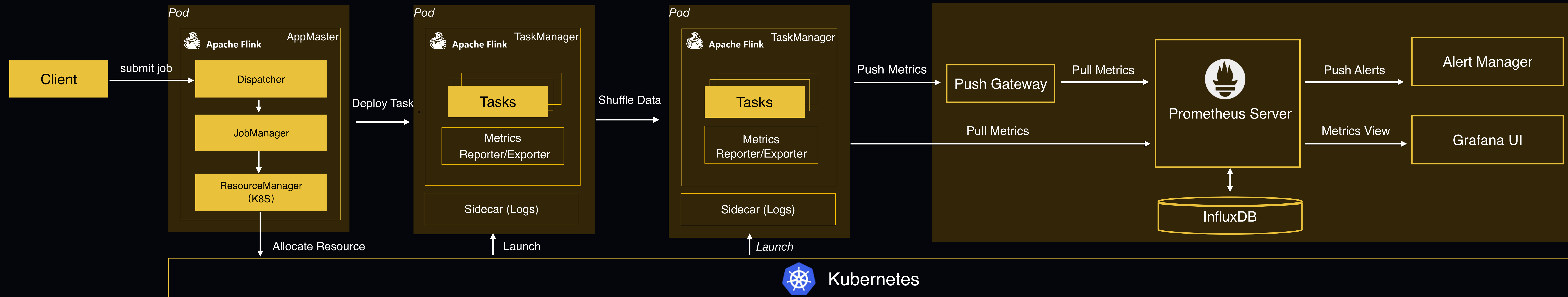
Flink AI Flow

- Flink as Analytics Engine
- TF/PyTorch on Flink
- OpenSource at 2020.Q1



云原生 (Cloud Native)

Flink Integration with Kubernetes Ecosystem



阿里巴巴Blink贡献给Apache Flink社区

Alibaba Blink Contributed to Apache Flink



Blink

2019.3



Flink-1.9 & Flink-1.10

> 1000,000 Lines

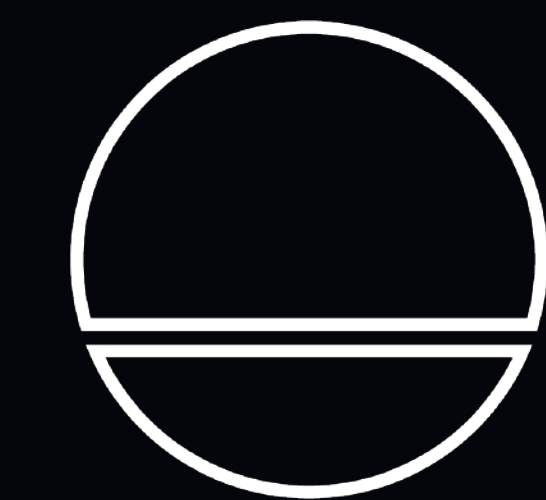
2020.1



Apache Flink



阿里云实时计算 = Ververica Platform on Alibaba Cloud



VERVERICA
PLATFORM

=



Apache Flink

+

Enterprise Features

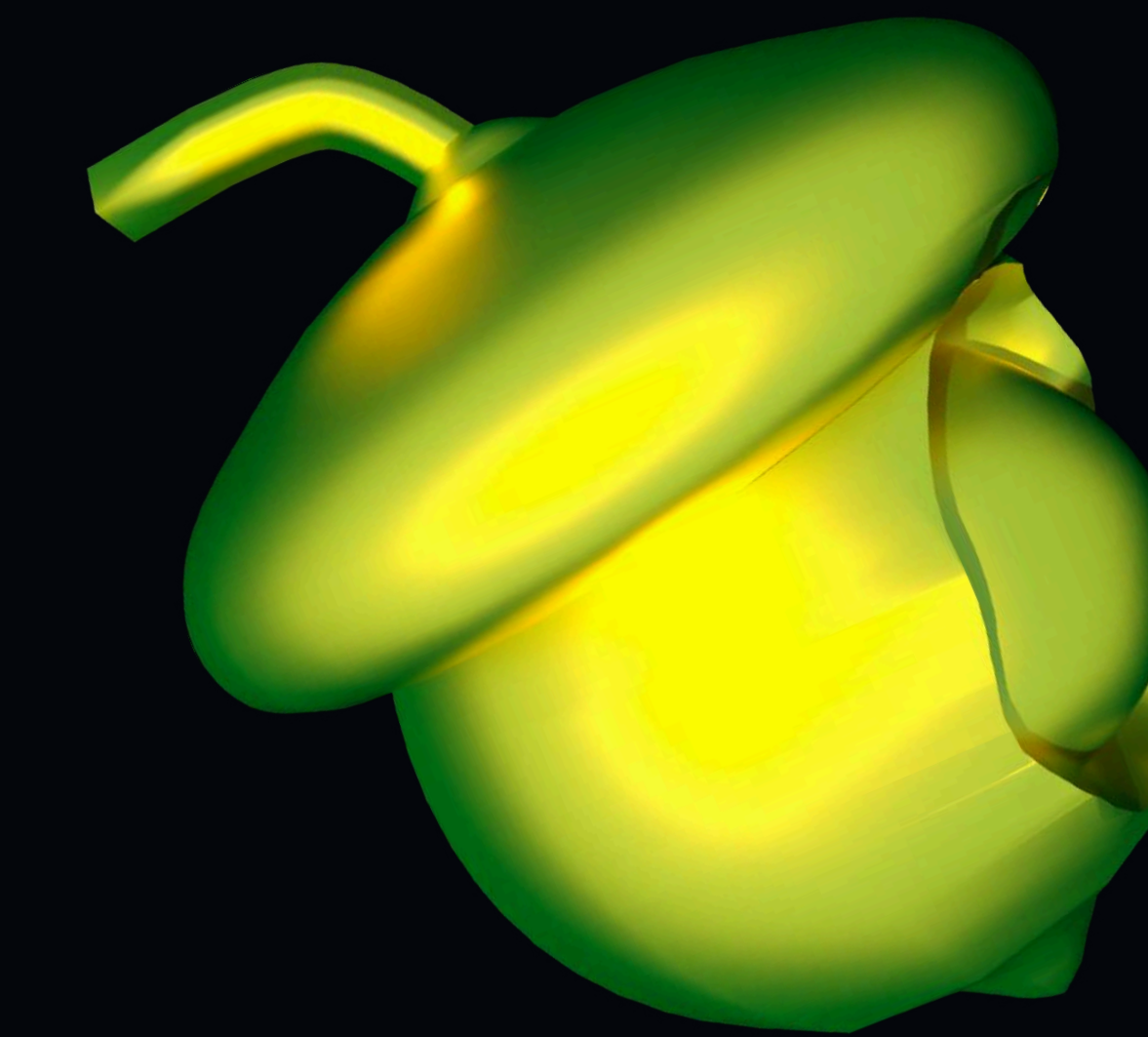
- Management & Operations
- Multi-Tenancy & Security
- Auto Scaling & Tuning
- Better Performance

 **Alibaba Cloud**

EMR (Hadoop)

ACK (Kubernetes)

Flink Forward Asia



2019

Thanks!