



## 专题出品人：孙斌

爱奇艺 VP

负责爱奇艺的商业智能、大数据、个性化推荐和用户增长等部门和方向。在互联网产品设计和研发方面有近二十年经验，曾就职于微软、hulu、雅虎等公司，也曾在国内知名互联网公司负责过创业项目。

## 专题：现代数据架构演进

地点：中优大宴会厅 1（三层）



## 专题出品人：孙斌

爱奇艺 VP

负责爱奇艺的商业智能、大数据、个性化推荐和用户增长等部门和方向。在互联网产品设计和研发方面有近二十年经验，曾就职于微软、hulu、雅虎等公司，也曾在国内知名互联网公司负责过创业项目。

## 专题：现代数据架构演进

地点：中优大宴会厅 1（三层）

在开源、云、AIGC 等技术的影响下，大数据架构正向现代数据架构演进。这些转变升级包括，从传统数仓到云原生数仓，从 Hadoop 到数据湖，从 ETL 到 ELT，从以 IT 为中心的交付到业务自助分析等等。当前敏捷 BI 已经开始成熟，面向业务人员的指标中台正快速发展，DataOps 备受关注，以 GPT 为代表的 AIGC 技术有望提升数据加工、分析效率，降低数据门槛。数据领域开始了新一轮的升级，人人用数据、时时用数据这样的数据平民化时代即将到来。

火山引擎 DataLeap 资深架构师

by 丁桂涛

字节跳动  
火山引擎 DataLeap 技术专家

- 数据资产行业痛点
- LLM应用：DataLeap-找数助手
- LLM 在 DataLeap 数据开发方向的实践与应用
  - 当前数据开发的现状
  - 数据开发当前痛点
  - LLM应用：DataLeap-开发助手
- DataLeap LLM 未来规划
  - LLM 在 DataLeap 迭代规划
  - 全链路 AI 能力

### 听众收益点：

- DataLeap 大数据研发治理套件的核心能力和优势
- LLM 技术在大数据研发与大数据资产消费领域的实践
- LLM 技术目前在大数据平台产品应用的局限

## 作业帮大数据湖仓架构和实践

随着数据湖相关技术的逐渐成熟、湖仓理论的发展，生产环境落地已经非常广泛。相比较传统数仓在数据更新时效性、查询索引增强等方面有明显优势，在基于云对象存储的存算分离架构下表现更为突出。本次演讲将从作业帮现有架构面临的挑战出发，结合真实的业务场景，阐述相关数据湖仓技术方案和实践经验。

### 演讲提纲：

- 现有架构的挑战
  - 作业帮现有数据架构
  - 架构缺陷及历史问题
- Iceberg 介绍
  - 元数据结构&特性
- 湖仓整体架构
  - 典型场景
  - 架构方案
- 湖仓实践
  - 采集入湖实践
  - 离线计算实践
- 展望和不足

### 听众收益点：

- 作业帮的业务场景及可复制的湖仓架构设计；
- HIVE 到 Iceberg 迁移及 Iceberg 应用实践；

by 孙建业

作业帮  
大数据平台架构负责人

## 爱奇艺数据治理中的数据湖应用实践

爱奇艺数据中台团队在管理和服务公司内数据资产，实施数据治理的过程中，不断吸收新的理念，引入新的工具，精细化爱奇艺的数据体系管理。

“数据湖”作为近年来数据领域广泛热议的概念，其技术层面也受到了业界的广泛关注。我们的团队对相关数据湖的理论和实践进行了深入研究，认为数据湖不仅是一种集成和处理数据的极具前景的技术，更是一种治理数据的全新视角，可以对数据管理、数据存储、数据链路等治理工作带来显著提升。

### 演讲提纲：

- 爱奇艺数据体系与数据治理工作介绍
- 数据治理工作中遇到了哪些问题
- 数据湖技术是如何解决数据治理工作中问题的
- 数据治理工作中应用数据湖技术获得了哪些收益
- 未来规划

### 听众收益点：

- 了解数据治理中遇到的一些问题的一种解决思路
- 了解数据湖技术如何在数据治理起到作用

by 杜益凡

爱奇艺  
商业智能部/研究员

## 基于软件开发新理念的新一代的数据基础设施

随着云原生技术的发展，数据库与大数据行业趋势也有了一些转变。Serverless、全托管，以及自建数据库均有其适用的场景。然而，如何在可控性和成本之间找到平衡、在复杂性和易用性之间找到演进的方法，是亟需关注的问题。DevOps 和云原生带来了新的管理范式，容器化、容器编排、基础架构即代码、声明式 API 等给业界带来了新的启发。将 Kubernetes 作为业务与云的中间适配层，将多种数据库引擎统一管理，做好抽象能够更好的管理好多场景下使用云数据库的复杂度。

本次演讲，会从背景和历史教训出发，基于软件开发的新理念，为大家介绍如何做好多数数据库引擎的抽象，如何用领域来封装复杂度、如何用低代码扩展新引擎？我们应该如何理解 HTAP？向量数据库可以被统一管理吗？混合云专有的 DBaaS 如何实现等等。

### 演讲提纲：

- 历史与教训：云原生数据库与大数据领域的研究
- 企业对数据基础设施的需求和本质
- 抽象：如何更好的管理复杂度
- 易用性：如何获得丰富的Day-2 Operation

### 听众收益点：

- 了解新一代的数据基础设施理念与实践
- 了解基于 Kubernetes 的有状态服务的抽象

by 曹伟

杭州云猿生数据  
创始人/CEO

### 关注主办方（InfoQ）



### 联系我们

购票热线：+86 18514549229  
票务微信：18514549229  
票务咨询：ticket@geekbang.com  
商务赞助：hezuo@geekbang.com  
媒体支持：media@geekbang.com  
议题申请：lucien@geekbang.com

### 交通指南

