

敏捷实时大数据平台全程落地应用实践

讲师 卢山巍 王东 徐翔 赵平 曹海霞

【课程特点】

1. 基于敏捷大数据开源平台工具，从无到有构建一套完整的实时大数据平台体系架构

这是一门针对如何构建、使用完整的大数据平台系统，为期 2 天的深度培训课程，由宜信的大数据团队，根据团队的实践经验，将大数据平台的整个开发、部署、应用的过程，进行了深度的分享讲解。本课程，包括了从数据的抽取，数据加工，数据存储运算（实时数据仓库/数据集市/ETL），再到实时数据应用，以及与数据仓库平台相关的元数据管理、血缘分析等全部环节。实现数据平台和数据应用的飞跃进阶。不管您的现在处于哪个级别，敏捷大数据平台直接带您升级到最高层级。

2. 核心技术解析、开源自主可控、二次开发扩展，让用户真正掌握驾驭整个大数据平台建设和发展

本次课程的大数据平台实践，全程采用开源工具，你可以深入进行二次开发，从而实现对整个大数据平台的真正自主可控。本次课程，培训专家团队，也将会深入讲解二次开发的方法与技巧。更重要的是，期望受训者，通过开源代码，指导文档和视频，可以为开源社区贡献代码，有能力把该平台运用到实际项目中。

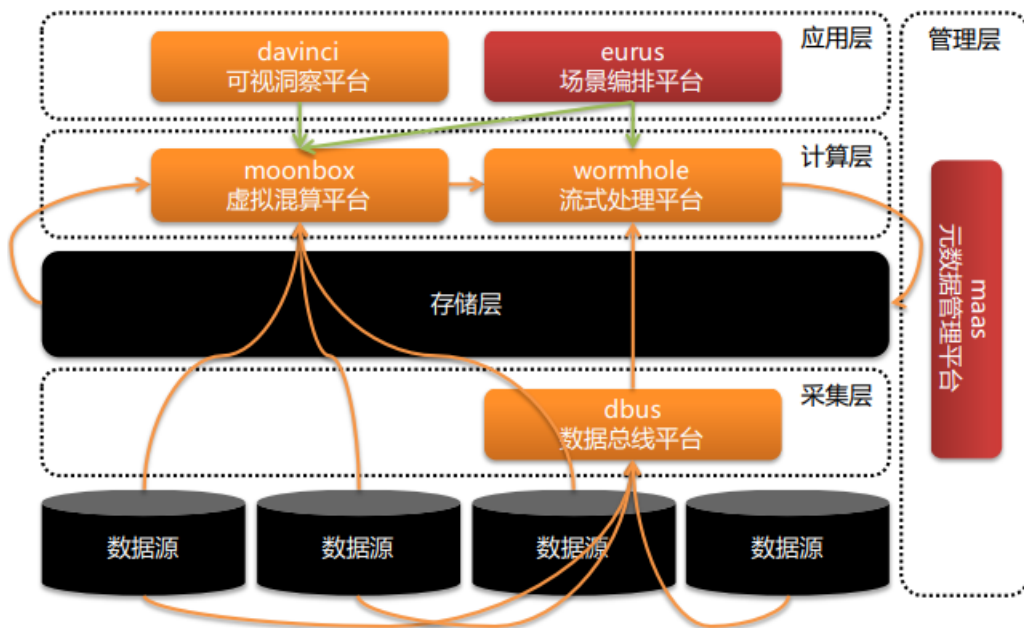
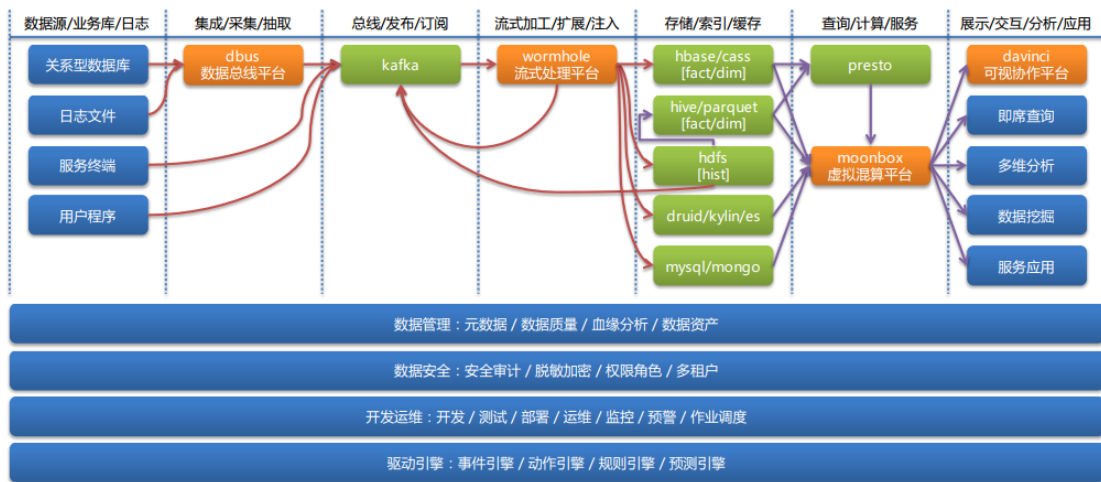
3. 常见大数据实时场景案例分析，典型架构模式探讨

本课程还会就一些典型的实时数据处理场景进行举例和案例分析，讲解场景需求背景、挑战、痛点、解决方案和项目效果，还会讨论一些通用的切面问题，如数据质量、运维监控等。过程中也会探讨不同的典型架构模式及其适用性。

4. 一些常见的主流的大数据开源工具的选型和剖析，为您提供深度参考

本次培训，还会提供各种大数据平台的相关开源工具选型剖析，这是宜信大数据团队在开发部署大数据平台的过程中，对多种工具进行试用，并总结的经验，这将是一份极其珍贵的参考建议。例如，数据存储方面，会涉及到 HBase、Cassandra、Kudu 的实际选型建议，数据查询方面，会涉及到 Spark SQL、Presto、Impala Hive 等工具的实际选型建议等。

本次大数据平台全程落地应用的整体结构图：



【培训安排】

	培训内容	讲师	时长 (小时)
10月20日上午	敏捷大数据理念及架构	卢山巍	3
10月20日下午	数据总线平台 Dbus	王东	3
	可视化应用平台 Davinci	徐翔	2
10月21日上午	流式处理平台 Wormhole	赵平	3
10月21日下午	计算服务平台 Moonbox	卢山巍	2
	实践应用案例解析	曹海霞	3

【培训对象与目标】

培训对象：

针对有一定基础的技术人员，懂得数据仓库平台的基础概念，了解 ETL、Hadoop 等基础技术概念，所在企业有计划部署结构化与非结构化的大数据仓库平台的技术工程师。

培训目标：

- ◆ 了解敏捷大数据的设计理念和架构设计
- ◆ 了解各个平台能提供的能力
- ◆ 对有技术背景的人员，通过开源代码，指导文档和视频，可以为开源社区贡献代码，有能力把该平台运用到实际项目中

【培训大纲】

1) 敏捷大数据理念

1.1 实时数据处理之痛

1.2 实时数据处理之理想世界

1.3 敏捷大数据定义

1.4 敏捷大数据理念

1.4.1 组件平台产品化

1.4.2 开发集成标准化

1.4.3 开发配置可视化

1.4.4 多层统一及开放化

1.4.5 自助、自动及智能化

1.4.6 引擎驱动化

2) 敏捷大数据处理架构

2.1 概念模块总览

2.2 数据管理（元数据，数据分析，血缘分析，数据资产）

2.3 数据安全

2.3.1 安全审计

2.3.2 脱敏加密

2.3.3 权限角色

2.3.4 多租户

2.4 开发运维监控及预警

2.5 驱动引擎

2.6 敏捷大数据之 CAP

2.7 开源平台地址

3) 数据集成、采集与抽取：深入洞察数据总线平台-Dbus

3.1 背景介绍

3.2 各类型数据源数据采集原理

3.2.1 关系型数据库

3.2.2 日志类数据库

3.2.3 奇异的 DB2

3.2.4 Mongo

3.3 系统健康检测

3.3.1 心跳机制（超时，丢数据、某环节掉链子等）

3.3.2 健康预警

3.4 开源平台地址

4) 数据加工处理：wormhole 流式处理平台

4.1 wormhole 架构设计理念

4.1.1 wormhole 设计目标

4.1.2 wormhole 设计规范

4.2 技术详解

4.2.1 多 flow 支持

4.2.2 多消息协议支持

4.2.3 lookup 实现与优化

4.2.4 Event Time 策略

4.2.5 异构 sink 幂等

4.3 开源平台地址

5) 数据分析展示：Davinci 可视化应用平台

5.1 背景介绍

5.2 Davinci 系统介绍

5.2.1 用户与权限系统

5.2.2 数据源、数据视图与数据模型

5.2.3 配置可视化图表

5.2.4 数据应用

6) 数据查询计算：Moonbox 计算服务平台

6.1 平台设计思想

6.2 总体架构和模块介绍

6.3 功能特点

6.3.1 MQL 介绍

6.3.2 集群管理

6.3.3 用户及权限体系

6.3.4 执行优化策略

6.3.5 多数据源支持

6.3.6 函数支持

6.3.7 定时事件

6.3.8 交互方式

6.3.9 与外界系统集成

6.4 开源地址

7) 实践应用案例解析

7.1 敏捷大数据轻模型案例之实时仪表盘

7.2 敏捷大数据全模型案例之实时即席查询

7.3 开源平台地址

【讲师简介】

卢山巍



卢山巍，宜信技术研发中心高级架构师，提出了敏捷大数据方法学，并围绕该套方法学，设计出了一系列的大数据相关平台，包括流式处理平台，可视化应用平台，计算服务平台。在加入宜信之前，就职于亿贝，是亿贝相关数据项目的高级软件架构师，在大数据处理相关领域有丰富的经验和深刻的见解。毕业于澳大利亚伍伦贡大学大学计算机专业，获得硕士学位。

总述 & 特长

- ◆ 10 年以上的互联网大数据工作经验和银行软件开发经验，对大数据相关技术有深刻的理解，擅于多维度高阶抽象来设计更通用的产品；
- ◆ 全栈工程师。包括前端，服务端，业务框架，数据仓库，大数据平台等；
- ◆ 优秀的业务分析和问题解决能力，出色的客户沟通能力；
- ◆ 热爱技术并乐于跟随技术的发展，具有超强的把各种新技术融合落地实现的能力。
- ◆ IT 技能
 - 语言：Java；Scala；C#；
 - 大数据：Spark，Hadoop；Kafka；Storm；Akka；分布式系统设计等；
 - 数据库：Teradata；MySQL；Ms SQL；SQL/NoSQL；数据实体关系设计；
 - 系统：SOA；DDD；Design Pattern；UML；Web Services/REST；XML/JSON；
 - 前端：HTML/CSS/JavaScript；GWT；Swift

讲师相关信息：

2018 思科金融行业峰会 金融领域敏捷大数据实践 演讲讲师

参考：http://science.china.com.cn/2018-06/08/content_40375694.htm

开源项目 wormhole 负责人：<https://github.com/edp963/wormhole>

开源项目 davinci 负责人：<https://github.com/edp963/davinci>

开源项目 moonbox 负责人：<https://github.com/edp963/moonbox>

王东



王东，宜信技术研发中心高级架构师，主要负责日志归集、流式处理和大数据业务产品解决方案，包括实时敏捷大数据技术栈基础组件——DBus 实时数据总线平台的建设和 NLP 自然语言处理技术解决方案的建设和探索等。在加入宜信之前，就职于 Naver(韩国最大搜索引擎公司)，担任中国研发中心资深研发工程师，拥有多年从事分布式数据库引擎研发经验，负责开源项目 CUBRID-cluster 分布式数据库开发和 CUBRID 数据库引擎开发工作。毕业于北京大学软件工程专业，获得硕士学位。

总述 & 特长

- ◆ 10 年以上行业经验和管理经验，对开发、测试和项目管理有丰富经验；
- ◆ 擅长 c/c++，Java, Scala, python 开发，扎实的数据结构和算法基础，很好的设计和开发能力，丰富跨平台开发经验 (Linux, Windows, AIX)；
- ◆ 熟悉开源流式处理引擎引擎，包括 Storm, Spark Streaming, Kafka 等；
- ◆ 熟悉大规模分布式系统开发、维护经验和常用开源分布式系统，包括 Hadoop、Hive、Spark、HBase、Kudu、ElasticSearch, mango 等
- ◆ 熟悉系统级编程，包括系统调用/API, 多线程，多进程编程，网络 socket 编程，IO 操作；
- ◆ 熟悉 Mysql, SQL server, CUBRID 等数据库, 对数据库理论和开发有深刻的理解；
- ◆ 熟悉分布式数据库集群，熟悉 sharding，高可用性 (HA) 和分布式一致性(2PC)；
- ◆ 熟悉和阅读源码包括 Mysql, Redis, Leveldb, Nginx, libevent 等，对 Nosql 相关技术有一定研究；
- ◆ 熟悉 python, Linux/Windows shell, js 等脚本；
- ◆ 擅长在百万行代码中定位问题，缩小问题规模，解决内存泄露，性能调优等问题；

讲师相关信息：

@ITPUB 学院

网址：<http://edu.itpub.net>

2018 全球软件与运维技术峰会(WOT) 大数据处理技术会场 演讲讲师

参考：http://wot.51cto.com/act/wot2018/dev/page/theme?hall_id=34

dbaplus 线上分享嘉宾讲师

如何基于日志，同步实现数据的一致性和实时抽取？

参考：<http://dbaplus.cn/news-21-872-1.html>

基于可视化配置的日志结构化转换实现

参考：<http://dbaplus.cn/news-134-1860-1.html>

开源项目 DBus 负责人: <https://github.com/BriData/DBus>

开源 RDBMS 数据库项目 CUBRID Commiter: <http://www.cubrid.org/>

开源分布式数据库项目 CUBRID-cluster 发起人

<https://www.slideshare.net/cubrid/cubrid-cluster-introduction>

<https://sourceforge.net/projects/cubridcluster/>

其他：

- ◆ 2013，雪鸟企业管理（上海）有限公司 培训 通过《Scrum 敏捷软件开发》课程
- ◆ 2011, 作为演讲嘉宾，参加 Naver 韩国首尔 workshop 大会，介绍和演示 CUBRID cluster 项目
- ◆ 2005/2006/2007/2008，多次参加微软 Tech-ed 大会培训，学习微软最新技术和工具.
- ◆ 2008，获得《网页搜索的显示方法及其客户端设备》专利证书，专利号：ZL 2005100002308.0
- ◆ 2005，获得 软件设计师 证书

徐翔



徐翔，宜信技术研发中心高级研发工程师，davinci 可视协作平台技术负责人。互联网金融行业从业 5 年，在数据可视化、前端、用户体验领域有着丰富的经验。毕业于北京工业大学计算机专业。

总述 & 特长

- 7 年以上互联网软件从业经验，对软件产品设计研发周期有着深刻的理解；
- 前、后端全栈从业经历，参与过 c 端与 b 端产品的设计和开发工作；
- 专业的敏捷开发管理经验，善于使用 scrum 管理产品研发周期；
- 扎实的可视化理论基础，对行业内可视化软件发展现状有着独到的理解；
- IT 技能

语言 : JavaScript; TypeScript; Java

前端 : React; Redux; Flux; FP; Vue; CSS-Modules; NW; PhoneGap

后端 : Node.js; SpringBoot;

讲师相关信息 :

开源项目 Davinci 技术负责人: <https://github.com/edp963/davinci>

赵平



赵平, 宜信技术研发中心高级架构师, 负责流式处理平台——Wormhole 的设计和实现, 并负责流式处理项目的解决方案设计和落地工作。加入宜信之前, 在东信北邮任项目经理一职, 拥有丰富的分布式设计和研发经验, 负责过多个重要项目, 包括中国移动“魔百盒”项目、*省移动家信通平台、*省企信通平台等

总述 & 特长

- 十年以上的分布式及大数据项目建设和管理经验, 对流式计算和分布式方面积累了大量的设计和开发经验, 并对流式计算有深刻的理解;
- 善于系统抽象及架构设计, 同时拥有大量系统问题的解决经验;
- 擅长数据的应用, 利用大数据技术、结合业务场景来挖掘数据的价值;
- 热爱创新, 特别在技术和设计方面, 并能够将新技术在项目中快速融合并落地;
- 技术栈:

语言 : Java、Scala、Golang

分布式 : Zookeeper、RocketMQ、Redis、Dubbo 等

大数据 : Spark、Flink、Hadoop、Kafka、Hbase、Kudu、ElasticSearch 等

数据库 : Mysql、Oracle

开源项目 wormhole 负责人: <https://github.com/edp963/wormhole>

曹海霞



曹海霞，宜信技术研发中心解决方案架构师，主要负责敏捷大数据平台的解决方案设计和落地实施。理解并分析客户需求，根据需求给出合理的解决方案并落地实施。负责 wormhole 1.0 版本的设计开发及实施运维。加入宜信前，就职于 Websense（韦伯森斯）北京研发中心，资深研发工程师，负责邮件安全网关数据和日志系统的架构设计、开发及技术支持。在数据处理系统设计、分布式数据处理、数据分析、反垃圾邮件技术、网络安全等方面有丰富的经验。毕业于江西理工大学计算机应用专业，获得硕士学位。

总述 & 特长

- ◆ 10 年以上数据处理及日志分析经验，对数据问题敏感，能快速理解需求并提供专业稳定的解决方案。
- ◆ 具有扎实的数据结构和算法基础，全面的专业基础，多年的系统架构设计经验，对方案的可行性和落地，能整体把握、有效推进。
- ◆ 对大数据相关技术有深刻理解，对各类大数据技术的应用场景有深刻的理解。
- ◆ 对网络安全及反垃圾邮件技术有深入研究，对大数据安全有深刻的理解及实践。
- ◆ IT 技能
- ◆ 语言：C/C++；Scala；Java；
- ◆ 大数据：Spark，Hadoop；Kafka；Akka 等；
- ◆ 数据存储：MySQL，SQL Server，HBase，Kudu，Elasticsearch，Redis 等；

开源项目 wormhole: <https://github.com/edp963/wormhole>

附录：

ITPUB 学院是为企业和个人提供企业应用方案培训咨询（包括企业内训）、个人实战技能培训（包括认证培训）在内的全方位 IT 技术培训咨询服务新平台，隶属于盛拓传媒 IT168 企业级平台。

IT168 企业级平台包括 ITPUB、ChinaUnix 两大社区；中国系统架构师大会（SACC）、中国数据库技术大会（DTCC）两大会议品牌。其中，ITPUB 目前是中国专业的数据库技术社区，ChinaUnix 是国内专业的 Linux/Unix 系统、开源软件技术社区。中国系统架构师大会、中国数据库技术大会是每届规模超过 1500 人的技术培训大会。

ITPUB 学院旨在提供更多精彩课程，详情请查看官方网站 <http://edu.itpub.net>

联系人：黄老师

电话：010-59127187

邮箱：edu@itpub.net

微信：13699173494

