

高能物理通用数据传输系统 设计与实现

高能物理研究所计算中心 曾珊

2017年6月5日

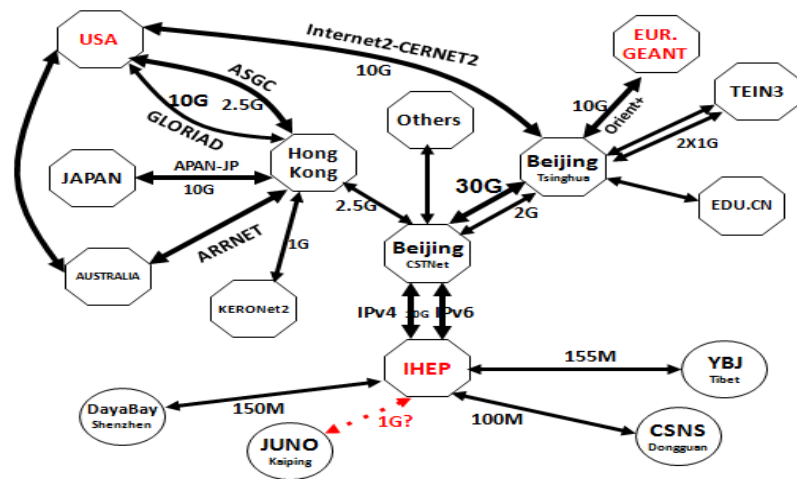
成都

大纲

- 背景介绍
- 设计目标
- 架构设计
- 模块功能介绍
- 部署模式
- 目前进展与下一步计划

背景介绍

- 高能物理实验产生海量数据
 - CSNS: 每年500TB
 - LHAASO: 每年4PB
 - JUNO: 每年2PB
- 实验现场异地建设
 - 数据具有共享需求
 - 高能所为保证远程数据传输提供了良好的广域网环境
- 对数据传输系统提出新的需求
 - 通用型（部署统一、维护方便）
 - 实时性（数据实时传输）
 - 稳定性（重传、监控...）



大纲

- 背景介绍
- 设计目标
- 架构设计
- 模块功能介绍
- 部署模式
- 目前进展与下一步计划

设计目标

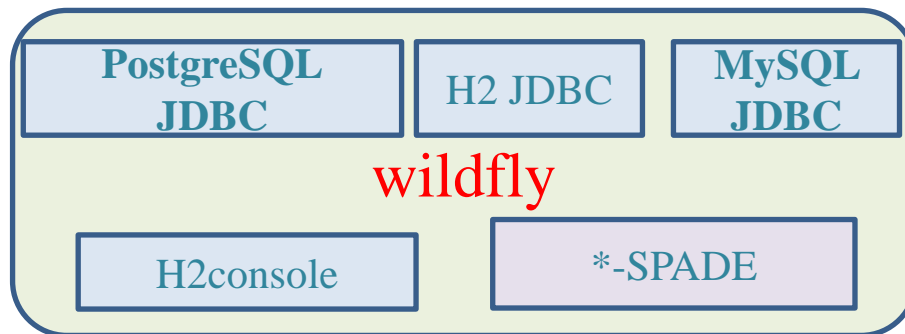
- 设计一种通用的数据传输系统，满足多个高能物理实验的数据传输需求
 - 实现海量数据的共享与分发（1对1或者1对多）
 - 从实验现场传输到高能所计算中心/其他计算中心
 - 从高能所计算中心分发到其他合作单位计算中心
 - 该系统能够应用于所有的高能物理实验，通过配置文件的修改来满足不同实验的需求
 - 功能灵活扩展
 - 系统稳定性
 - 缓冲区机制
 - 重传机制
 - 良好的接口设计

大纲

- 背景介绍
- 设计目标
- 架构设计
- 模块功能介绍
- 部署模式
- 目前进展与下一步计划

架构设计(I)

- 基于DYB数据传输系统SPADE进行开发实现
 - 该系统已稳定运行8年
 - 完成从大亚湾现场到高能所再relay到LBNL的源数据传输
- 部署于Wildfly（JEE server）上的Java application
 - COMMON-SPADE
 - JUNO SPADE/LHAASO SPADE/CSNS SPADE....
 - 根据配置文件中参数的不同，实现不同的功能
 - 发送：Sender Instance
 - 接收：Receiver Instance



架构设计(II)

- 从JBoss7到Wildfly10.0.0.1
 - DYB SPADE部署于JBoss7之上（基于JavaEE的开源的应用服务器）
 - 存在的问题：启动很慢，代码结构复杂
 - Wildfly是Jboss被Redhat公司收购后的更名
 - 代码重整，采用模块化设计
 - 非常轻量，内存占用非常少（解决启动慢的问题）
 - 方便的配置和管理接口
 - 严格遵守Java EE7和OSGi规范
 - 安装和卸载的方法
 - 安装：在Wildfly上运行`deploy --force --name=spade.war ~/dybapps/dyb-spade-2.2.6/target/dyb-spade-2.2.6.war`
 - 卸载：在Wildfly上运行`undeploy --force --name=spade.war`

大纲

- 背景介绍
- 设计目标
- 架构设计
- **模块功能介绍**
- 部署模式
- 目前进展与下一步计划

模块化设计 (I)

• 数据传输通用功能模块

– 传输协议

- 支持SCP/GridFTP协议的数据传输（多流传输）
- 为简化使用，其中Gridftp认证采用的是ssh-key，并未使用网格的证书（与目前dyb的用法一致）

– 缓冲区机制

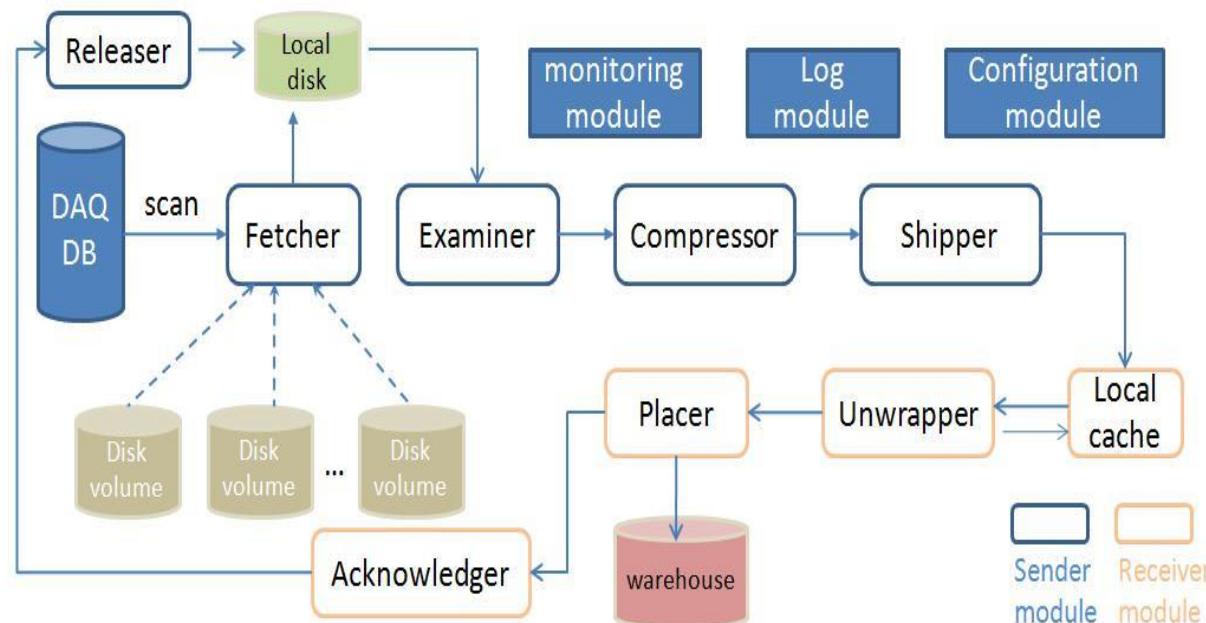
- 数据传输方和接收方都可以设置本地缓存，保证数据传输的有效性

– 反馈机制

- 数据传输成功
 - » 缓冲区清空
- 数据传输失败
 - » 启动重传机制

– 监视模块

- 按文件个数
- 按文件大小



模块化设计（II）

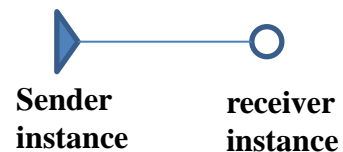
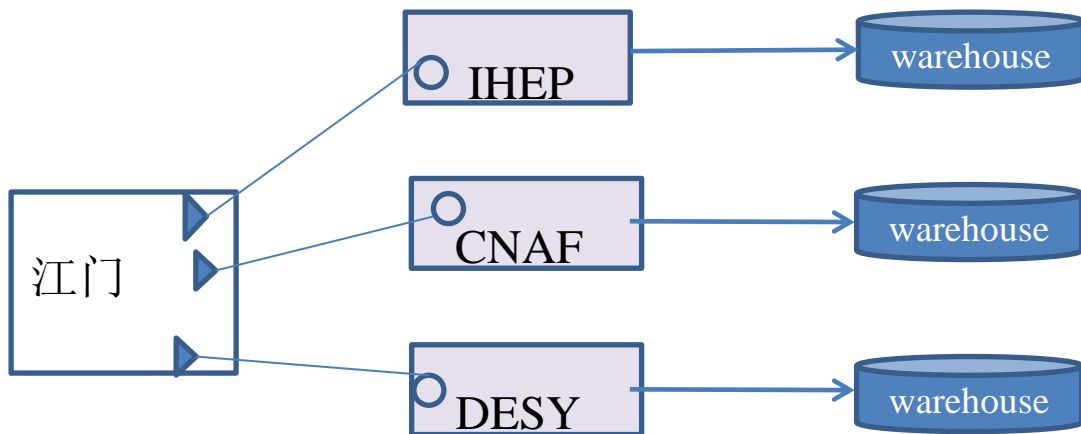
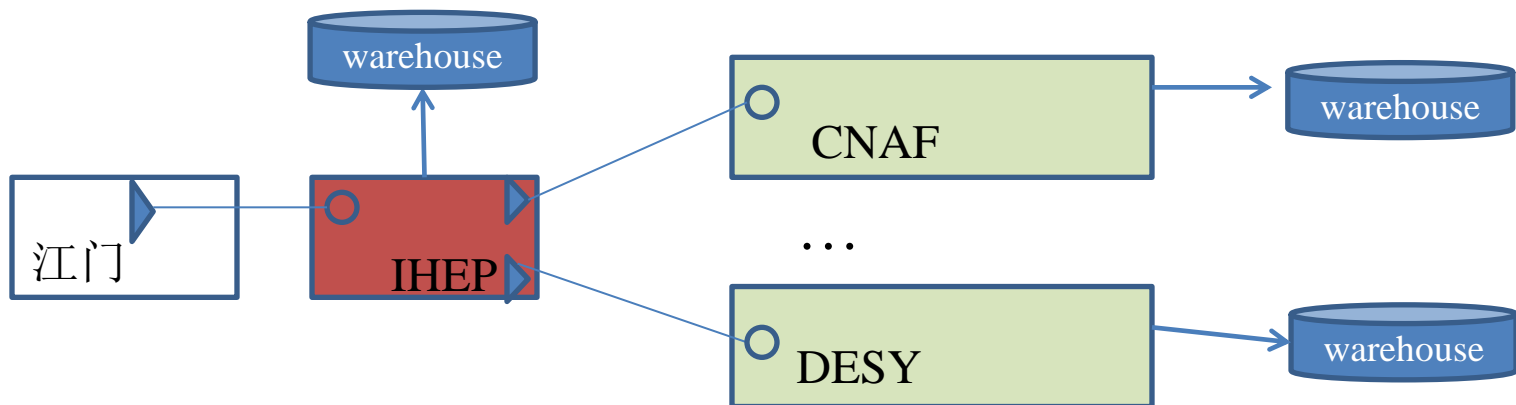
- 为了和其他系统进行交互（例如：DAQ, Bookkeeping等）
 - 底层提供多种数据库支持
 - **Postgresql/MySQL/H2**
 - 数据传输系统会为传输文件创建Metadata（xml文件）
 - 生成时间，参数配置，数据源等信息
- 配置模块中提供与实验相关的定制化接口
 - 通过配置文件的方式进行定义
 - Metadata的文件内容
 - 需要传输文件类型（前缀、后缀、扫描文件的存放路径等）
 - 文件在远端warehouse的存放路径

大纲

- 背景介绍
- 设计目标
- 架构设计
- 模块功能介绍
- **部署模式**
- 目前进展与下一步计划

部署模式

- 支持两种模式：Relay模式和Direct模式
- 以JUNO为例



大纲

- 背景介绍
- 设计目标
- 架构设计
- 模块功能介绍
- 部署模式
- 目前进展与下一步计划

目前进展与下一步计划

- 完成系统功能模块设计与测试床部署
- 分别在中山、IHEP和CNAF（一台虚拟机）部署完成
 - 完成中山到IHEP的数据传输功能测试（模拟数据）
 - 由于目前中山到IHEP并没有专线，数据传输性能受限于广域网的性能
- 下一步计划
 - 采用Relay方式与CNAF进行数据传输测试
 - 专线ready后，需要使用真实数据进行测试（性能）
 - 与DAQ、Bookeeping等联调测试
 - 考虑采用Docker技术进行SPADE的部署,不同的物理实验启用不同的Docker APP

总结

- 通用数据传输系统能够支持多个高能物理实验
 - 数据传输通用功能模块继承大亚湾数据传输系统的实现
 - 基于不同实验进行定制化配置实现
 - 提供了与其他交互的接口与兼容性考虑
- 下一步
 - 需要进一步与DAQ等系统的联调测试
 - 系统性能测试
 - 考虑采用Docker技术进行SPADE的部署

谢谢

Q&A