

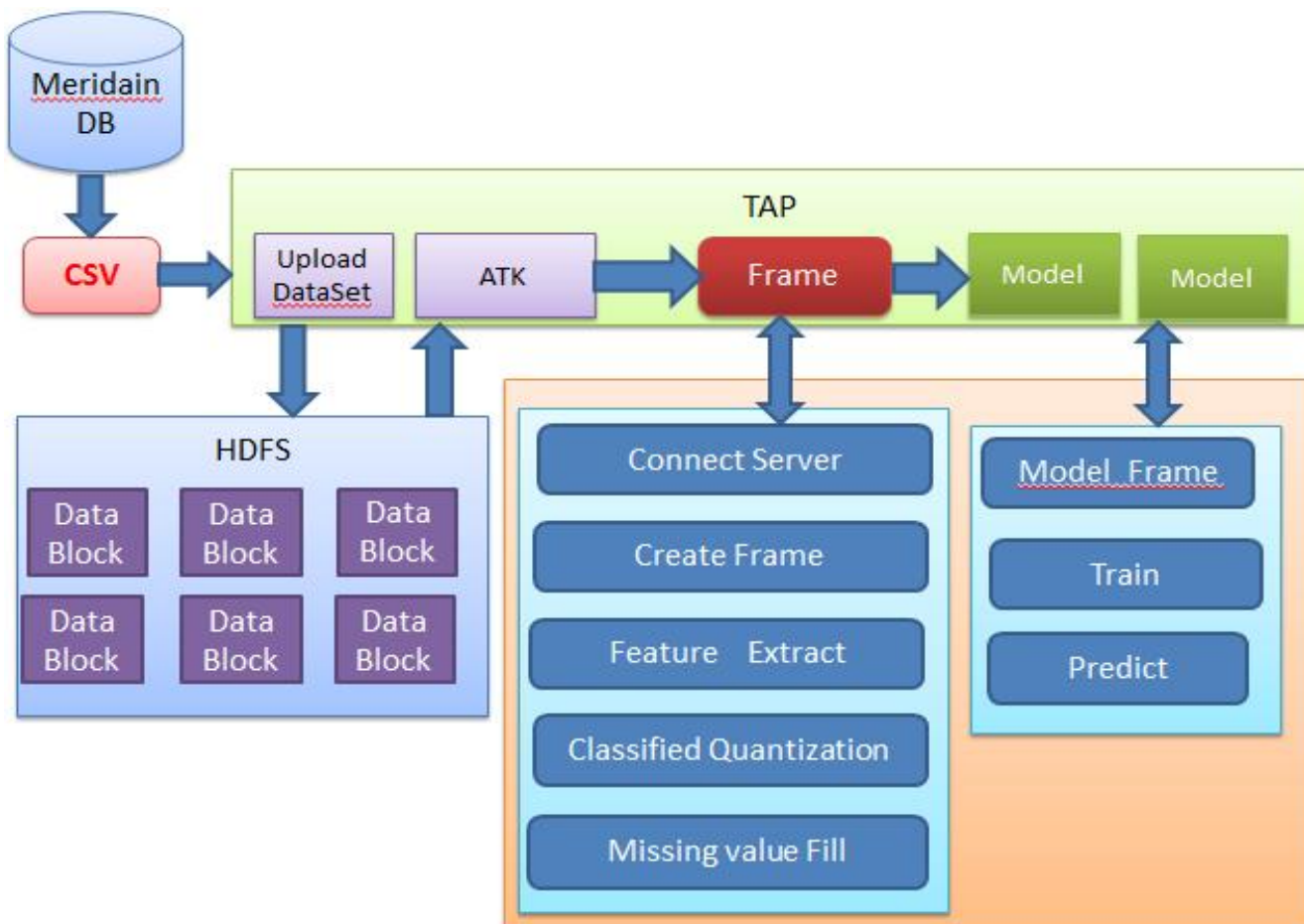
Li, Jun A

From: Lin, XiangX
Sent: Friday, January 29, 2016 2:50 PM
To: Li, Jun A
Subject: FW: Meridian TAP project status ww05.5

Hi, guys:

下图是通过TAP平台处理和量化流程示意图，另外在数据上传到HDFS之前需要注意以下几点：

- 1 feature名称不能有重复；
- 2 feature名称不能包含 %, #, / 等特殊符号；
- 3 feature值不要包含中文字符；
- 4 feature值中不要包含与CSV默认分割符相同的字符；



From: Zhang, Lijuan
Sent: Friday, January 29, 2016 1:36 PM
To: Lin, XiangX
Subject: FW: Meridian TAP project status ww05.5

林翔

我看你之前画过相关的图，能不能针对atk这个版本修改下？

From: Wang, AlbertZQ
Sent: Friday, January 29, 2016 1:24 PM
To: Zhang, Lijuan <lijuan.zhang@intel.com>; Cao, Buddy <buddy.cao@intel.com>; Lou, Ming <ming.lou@intel.com>; Li, Carl <carl.li@intel.com>; Liu, Huanglingzi <huanglingzi.liu@intel.com>; Gao, Fengqian <fengqian.gao@intel.com>; Ding, Danny <danny.ding@intel.com>; Lu, Yuan Y <yuan.y.lu@intel.com>; Yi, Lan <lan.yi@intel.com>
Subject: RE: Meridian TAP project status ww05.5

Hi Lijuan

Thanks a lot! Can we have a diagram for the data processing module's functionality and data flow specification?

Cheers,
Albert

From: Zhang, Lijuan
Sent: Friday, January 29, 2016 1:17 PM
To: Cao, Buddy; Wang, AlbertZQ; Lou, Ming; Li, Carl; Liu, Huanglingzi; Gao, Fengqian; Ding, Danny; Lu, Yuan Y; Yi, Lan
Subject: RE: Meridian TAP project status ww05.5

Hi, guys

本周数据量化从c#转到atk python code基本功能点已经实现，所有步骤目前均利用atk现有函数实现，内容如下表所示：

处理步骤	完成功能
分类量化	1通过数组来指定列名处理； 2 建立一个工厂方法，处理需要分类的feature； 3 每处理一个feature新生产一个列，列名命名方式为：HCT -> Classify_HCT 4 处理完成以后删除原有feature； 5 将Classify_命名feature恢复原名称；
数据填补	1将整列数据提取到一个数组中 2 该数组去掉默认值“32768” 3建立一个工厂方法，执行中位数、均值、频次等操作，取到填补值； 4将填补值，替换到数组中； 5创建新frame，包括原有feature和填补过的feature；填补的feature名称为Fill_HCT； 6 与原有frame进行join操作，此时会生成一个新frame（model_frame），包括了原来所以feature，填补过feature，和填补feature的副本；

7根据填补维护数组进行删除操作; 8 将填补frame进行重命名;

下一步需要完成的工作:

- 1 针对feature处理列表动态维护数组序列下标, 与frame结构对应;
- 2 完善脚本整个流程调试工作;

Thanks
Lijuan

From: Cao, Buddy

Sent: Friday, January 29, 2016 11:05 AM

To: Wang, AlbertZQ <albertzq.wang@intel.com>; Lou, Ming <ming.lou@intel.com>; Li, Carl <carl.li@intel.com>; Liu, Huanglingzi <huanglingzi.liu@intel.com>; Gao, Fengqian <fengqian.gao@intel.com>; Ding, Danny <danny.ding@intel.com>; Zhang, Lijuan <lijuan.zhang@intel.com>; Lu, Yuan Y <yuan.y.lu@intel.com>; Yi, Lan <lan.yi@intel.com>

Subject: Meridian TAP project status ww05.5

In this week, team puts most of efforts to develop 数据量化工具, research ATK Plug-in and work with ATK/TAP engineers for deployment and app defects/issues.

- **Done** - (Deployment - FlexLab) Latest TAP deployment in Flex lab - **Basically done except for non-critical scoring-engine**
 - [TRACS-80 - 'Yarn-cluster' error at customer side](#)
 - 1 bug resolved – atk instance creation fail
 - 1 bug WIP – scoring-engine start fail but will not block us.
 - Andrej/Yuan are debugging. Poland team (Andrej) started to help on the deployment from ww03.2. Over 10 failed apps are started successfully. Yuan is working with Andrej daily basis. However, the latest TAP has significant upgrades on most of the apps/services, it generates risks for the deployment success. Yuan continues work closely with Andrej on the deployment and share status on a daily basis.
- **Done** - (Defect) Meridian ATK周四出现了 “404 not found” 问题 – **Fixed in ww05.4**
- (Development) 数据量化工具 Migration from C# to ATK Python code – **WIP**
 - Plan to deliver in 4 weeks
- (Research & Solution Proposal) ATK plugin Verification – **WIP**
 - Study done and sample code is ready.
 - The plug-in sample code cannot be pushed to atk successfully, ATK engineer is helping Yuan to debug.
- (Document) China version of TAP Deployment Guide – Yuan sent to Andrej for review – **Waiting Andrej's reply**
- (Trial) Albert - 记录一下每次modelling的时间, 做个粗略的bench mark。。比如把10K数据复制成几倍的, 看看执行效率有什么变化 – **TBD**
- (Document) Albert - TAP类似的sizing的估算 (这个其实反应了对应不同的计算规模, 大概可以带动多少compute能力的投入 (IA产品: Xeon, SSD 等)) – **TBD**
- (Defect & Feature Request) Status keep tracking
 - (Upgrade) ticket [TRACS-75 upgrade the TAP to a newer TAP release](#). Poland team (Adam) replied that there is no automatic way to upgrade, they will provide the instruction to upgrade manually
 - (Meridian App) ticket [TRACS-93 - "spark-assembly" file missing in Jan-1 release](#) Meridian Hypertension model with Latest TAP deployment in Flex lab. - Missing spark-assembly.jar package – will fix in JAN-2 release
 - (Deployment) ticket [TRACS-80 - 'Yarn-cluster' error at customer side](#) - 1 bug WIP – scoring-engine start fail but will not block us. - will fixed in JAN-2 release

- (Deployment) ticket [TRACS-81 – failed to push all of apps JAN-1 2016 package](#) - will fixed in JAN-2 release
- Done - (Development)数据量化工具dpp app improvement for Meridian – Delivered to Meridian in ww03.5
 - 分类符号涵盖了“<,>,>=,<=,=”, 主要组合使用可以满足不同的边界问题;
 - 增加了每个步骤的出差提示, 主要包括: xml文件格式错误, 分类配置错误, feature选取配置错误, 列名重复提示错误等;
 - 对涵盖"/", "#", "%"的列名进行了处理, 列名更改规则: “#”替换为“_V”, “%”替换为“_R”, “/”替换为“_”;
 - help中增加了classify.xml配置的示例说明;
- Done - (Research & Solution Proposal - Meridian)数据量化工具Migration from C# to ATK Python code – [Proposal and Evaluation](#) delivered to HLS in ww04.5
- Done - (Document)China version of TAP Deployment Guide – Delivered in ww04.5
- Done - (Development - Meridian)Meridian填补数据的问题, 目前的填充方式应该是自动填充, 是否能够不自动填充 – Delivered to Meridian in ww04.4
 - 可以为每个feature设置不同的填补方式; 填补方式有: 均值、中位数、计算频次; 可以支持不填充设置
- Done - (Development - Meridian)DPP工具FeatureList.txt文件中第一行并未在工具中显示出来 – Delivered to Meridian in ww04.4
- Done - (Deployment – Meridian)Meridian TAP model.train()错误- Fixed by Soila/Fengqian in ww04.5
 - [TRACS-81 – failed to push all of apps JAN-1 2016 package](#)

Thanks
Wei Cao (Buddy)

From: Cao, Buddy
Sent: Friday, January 22, 2016 3:22 PM
To: Wang, AlbertZQ <albertzq.wang@intel.com>; Lou, Ming <ming.lou@intel.com>; Li, Carl <carl.li@intel.com>; Liu, Huanglingzi <huanglingzi.liu@intel.com>; Gao, Fengqian <fengqian.gao@intel.com>; Ding, Danny <danny.ding@intel.com>; Zhang, Lijuan <lijuan.zhang@intel.com>; Lu, Yuan Y <yuan.y.lu@intel.com>; Yi, Lan <lan.yi@intel.com>
Subject: Meridian TAP project status ww04.5

Hi, Folks,

We have a very productive week and completed 5 critical tasks within plan. Thanks for your dedication and contribution, let's keeping focus on the upcoming tasks and proactively communicate with Meridian and ATK/TAP teams.

- Done - (Research & Solution Proposal - Meridian)数据量化工具Migration from C# to ATK Python code – [Proposal and Evaluation](#) delivered to HLS in ww04.5
- Done - (Document)China version of TAP Deployment Guide – Delivered in ww04.5
- Done - (Development - Meridian)Meridian填补数据的问题, 目前的填充方式应该是自动填充, 是否能够不自动填充 – Delivered to Meridian in ww04.4
 - 可以为每个feature设置不同的填补方式; 填补方式有: 均值、中位数、计算频次; 可以支持不填充设置
- Done - (Development - Meridian)DPP工具FeatureList.txt文件中第一行并未在工具中显示出来 – Delivered to Meridian in ww04.4
- Done - (Deployment – Meridian)Meridian TAP model.train()错误- Fixed by Soila/Fengqian in ww04.5
 - [TRACS-81 – failed to push all of apps JAN-1 2016 package](#)
- (Deployment - FlexLab)Latest TAP deployment in Flex lab - no ETA yet

- [TRACS-80 - 'Yarn-cluster' error at customer side](#)
- Stuck on 2 bugs— scoring-engine start fail and create atk instance fail.
- Andrej/Yuan are debugging. Poland team (Andrej) started to help on the deployment from ww03.2. Over 10 failed apps are started successfully. Yuan is working with Andrej daily basis. However, the latest TAP has significant upgrades on most of the apps/services, it generates risks for the deployment success. Yuan continues work closely with Andrej on the deployment and share status on a daily basis.
- (Research & Solution Proposal) ATK plugin Verification – **TBD, would probably start to research from ww05.**
 - 另外最好记录一下每次modelling的时间，做个粗略的bench mark。。比如把10K数据复制成几倍的，看看执行效率有什么变化
- (Upgrade)Upgrade Meridian TAP – **WIP** by Poland team
 - TAP request Ticket [TRACS-75](#) was submitted
 - Poland team (Adam) replied that there is no automatic way to upgrade, they will provide the instruction to upgrade manually.
- Done - (Development)数据量化工具dpp app improvement for Meridian – **Delivered to Meridian in ww03.5**
 - 分类符号涵盖了“<,>,>=,<=,=”，主要组合使用可以满足不同的边界问题；
 - 增加了每个步骤的出差提示，主要包括：xml文件格式错误，分类配置错误，feature选取配置错误，列名重复提示错误等；
 - 对涵盖“/”，“#”，“%”的列名进行了处理，列名更改规则：“#”替换为“_V”，“%”替换为“_R”，“/”替换为“_”；
 - help中增加了classify.xml配置的示例说明；

Thanks
Wei Cao (Buddy)

From: Cao, Buddy
Sent: Monday, January 18, 2016 3:28 PM
To: Wang, AlbertZQ <albertzq.wang@intel.com>; Lou, Ming <ming.lou@intel.com>; Li, Carl <carl.li@intel.com>; Liu, Huanglingzi <huanglingzi.liu@intel.com>; Gao, Fengqian <fengqian.gao@intel.com>; Ding, Danny <danny.ding@intel.com>; Zhang, Lijuan <lijuan.zhang@intel.com>; Lu, Yuan Y <yuan.y.lu@intel.com>; Yi, Lan <lan.yi@intel.com>
Cc: Cao, Buddy <buddy.cao@intel.com>
Subject: Meridian TAP project status ww04.1

Guys,

Here you go, the latest status from Flex Services.

- (Development)数据量化工具dpp app improvement for Meridian – **Delivered to Meridian in ww03.5**
 - 分类符号涵盖了“<,>,>=,<=,=”，主要组合使用可以满足不同的边界问题；
 - 增加了每个步骤的出差提示，主要包括：xml文件格式错误，分类配置错误，feature选取配置错误，列名重复提示错误等；
 - 对涵盖“/”，“#”，“%”的列名进行了处理，列名更改规则：“#”替换为“_V”，“%”替换为“_R”，“/”替换为“_”；
 - help中增加了classify.xml配置的示例说明；
- (Research & Solution Proposal)数据量化工具Migration from C# to ATK Python code
 - Research is WIP
 - Solution Proposal and Evaluation for the Implementation will be provided to HLS and Meridian by **ww04.5**
- (Deployment)Latest TAP deployment in Flex lab - **ETA is uncertain**
 - Stuck on pushing app for ATK and Console
 - Poland team (Andrej) started to help on the deployment from ww03.2. Over 10 failed apps are started successfully. There are still 2+ key apps are not started including atk and console.

- Yuan is working with Andrej daily basis. However, the latest TAP has significant upgrades on most of the apps/services, it generates risks for the deployment success.
- Yuan will continue work closely with Andrej on the deployment and share status on a daily basis.
- (Document)China version of TAP Deployment Guide – Plan to deliver **in ww04.5**
 - Yuan has completed 50% of the deployment guide document.
- (Upgrade)Upgrade Meridian TAP – **WIP** by Poland team
 - TAP request Ticket [TRACS-75](#) was submitted
 - Poland team (Adam) replied that there is no automatic way to upgrade, they will provide the instruction to upgrade manually.

Thanks
Wei Cao (Buddy)