LCC-Optimizer产品操作说明书

版本号:v1.1.0 版权所有:贝联珠贯科技(浙江)有限公司

1. 产品介绍

1.1 概述

LCC-Optimizer是贝联珠贯科技(浙江)有限公司推出的一款大数据作业优化产品。旨在帮助企业大数据团队充分优化现有的大数据作业,在保证作业稳定运行的前提下,以低侵入的方式提升作业执行效率,提高机器利用率。LCC-Optimizer凭借强大的优化引擎,用户只需要简单的安装设置,即可自由的选择现有集群中的大数据任务进行自动优化,让现有大数据集群释放出更多潜力。

1.2 名称解释

1.2.1 任务

• 优化的对象,当前支持的类型有: MapReduce Job, Spark Application。

1.2.2 执行

• 任务的具体执行,对应Yarn上一个或多个Application。

1.2.3 规则

 每一类优化逻辑被定义为规则,它本质上是针对Application历史数据的一类分析方法和对提交任务 的动作。

1.2.4 优化指标

目前产品中常用的优化指标包括:

• 运行时长(s):

Spark计算逻辑: Spark History上的Application结束时间 - 开始时间

Hadoop计算逻辑: Jobhistory Overview上的结束时间 - 开始时间

CPU使用量(vcore-seconds):

Spark计算逻辑: spark driver 数量 * 每个driver申请的核数 * 运行时长 + spark executor 数量 * 每个executor申请的核数 * 运行时长

Hadoop计算逻辑: Jobhistory Counters上的 (Total vcore-milliseconds taken by all map tasks) + (Total vcore-milliseconds taken by all reduce tasks)

内存使用量(GB-seconds):

Spark计算逻辑: spark driver 数量 * 每个driver申请的内存 * 运行时长 + spark executor 数量 * 每个executor申请的内存 * 运行时长

Hadoop计算逻辑: Jobhistory Counters上的(Total megabyte-milliseconds taken by all map tasks) + (Total megabyte-milliseconds taken by all reduce tasks)

2. 用户指南

在产品安装完成后,客户端只需要一个浏览器即可开始使用产品。

🥐 🔹 登录地址:由lcc-hbo-pc和客户提供的Domain Suffix组成。比如:

https://lcc-hbo-pc.rc.io

2.1 用户登录与管理

首先是登录界面,如下图所示:

默认管理员账号密码: admin/123456,登录后可以进行修改



2.1.1 个人中心

登录产品后,可以点击用户名下拉选项的【个人中心】进入个人中心,进行用户名称修改,密码 修改,邮箱绑定等操作。

• 如果忘记密码可以通过绑定的邮箱进行密码重置

2.1.2 用户管理

登录产品后,点击系统管理图标 〇 下的【运维管理】后点击左侧导航栏中的【用户管理】进入。可以在界面中进行如下操作:

- 查找现有人员
- 单个或批量添加新人员
- 变更人员角色
- 删除现有人员

当前系统角色包括:系统管理员,管理员,普通用户

2.2 主页

登录主界面后可以看到界面上左边「主页」「执行列表」「任务列表」「分析报告」「优化设置」五个大类,上方是账号设置。默认展示的是主页,如下图所示:



版权声明: Copyright 2022 LCComputing Inc.

主页包括内容:

- 系统优化效果:包括任务总时长,CPU使用量,内存使用量
- 任务统计:包括总任务数,已优化任务数,已生效规则,最近7天优化任务执行趋势图
- 任务加速排行:查看哪些任务运行时长优化的比例最高
- 任务类型占比:不同任务类型的占比

2.3 执行列表

执行列表可以看到配置了LCC-Optimizer HOOK的Gateway的最近作业提交列表。包括不同任务类型的执行列表,比如: MapReduce Job, Spark Application。可以通过筛选执行时间,优化情况进行查询,如下图所示:

4	LCC-Optimizer				admin
۳	执行列表 / MapReduce Job				
≣					
≔	开始时间: 2022-08-17	20:5〔 → 2022-08-18 20:5〔 🗎 是否优1	七: 已优化	∨重置 查询	
\odot					
ŝ	执行列表				
	Yarn Application ID	开始时间	执行时长 💠 CPU使用	量(vcore-secs)	状态 任务ID
	application_1660050842344_10943	22-08-18 20:50:32 22-08-18 20:50:44	12sec 6	14.6	SUCCEEDED 7972ac904ec
	application_1660050842344_10942	22-08-18 20:50:14 22-08-18 20:50:20	6sec 3	3.5	SUCCEEDED 703955f45c5
	application_1660050842344_10941	22-08-18 20:49:10 22-08-18 20:49:23	13sec 8	8.7	SUCCEEDED f0e0698d0fc7
	application_1660050842344_10939	22-08-18 20:48:39 22-08-18 20:48:55	16sec 18 ²⁶	dmin 18.7	SUCCEEDED 39d60c99a46
	application_1660050842344_10937	22-08-18 20:48:29 22-08-18 20:48:43	14sec 9	9.3	SUCCEEDED 69b006c8917
	application_1660050842344_10936	22-08-18 20:48:01 22-08-18 20:48:15	13sec 17	17.4	SUCCEEDED 88222514d6
	application_1660050842344_10934	22-08-18 20:47:17 22-08-18 20:47:28	10sec 6	14	SUCCEEDED 396b5a8c613
	application_1660050842344_10933	22-08-18 20:46:58 22-08-18 20:47:03	5sec 2	2.9	SUCCEEDED 637b3b9207(
	application_1660050842344_10932	22-08-18 20:46:17 22-08-18 20:46:39	22sec 17	17.1	SUCCEEDED f8961a04b80
) I	application_1660050842344_10931	22-08-18 20:45:45 22-08-18 20:52:22	6min 37sec 1566	3128.2	SUCCEEDED 740a3dc28c2

- 通过点击【任务ID】,可以直接跳转到任务详情。
- 通过点击【优化详情】,可以查看该次执行具体应用了哪些优化规则,调整了哪些优化参数。如下 图所示:

4	LCC-Optimizer								admin
Ľ	执行列表 / MapReduce Job								
≣		优化详情				×			
≣	开始时间: 开始								
Ø		规则名称	规则类型	动作	原参数	优化后参数			
ŝ	执行列表				{"mapreduce.reduce.java. {"mapreduce		napreduce.reduce.java.		
		Reduce内存设置	任务	提交参数	opts":"-Xmx4915m","ma preduce.reduce.memory.	opts":"-Xmx410m","mapr educe.reduce.memory.m		1 10 11	12.45
	◆ 结束时间 ◆				mb":"6144"}	b":"512"}		优化	操作
	37:38 22-08-18 20:37:51	Map内存设置			{"mapreduce.map.java.op ts":"-Xmx2457m","mapre duce.map.memory.m b":"3072","mapreduce.ta sk.io.sort.mb":"1719"}	{"mapreduce.map.java.op	41fc91	已优化	优化详情
	37:15 22-08-18 20:37:27		任务	提交参数		uce.map.memory.mb":"6	5304	已优化	优化详情
	36:49 22-08-18 20:37:02					ort.mb":"100"}	b6a05	已优化	优化详情
	36:22 22-08-18 20:36:37	10000 10		1010			. 2fc59	admin 已优化	优化详情
	36:20 22-08-18 20:36:34	14sec 9		9.1	SUCCEEDED	0772fa06074ac003acba5c1d49	59b01f	已优化	优化详情
	35:56 22-08-18 20:36:06	10sec 7		7.2	SUCCEEDED	32d7336ca14e117abbaddb8aaa	a569d8	已优化	优化详情
	35:51 22-08-18 20:36:04	13sec 16		17	SUCCEEDED	979d3e8171af4a2b58c163e292	054d09	已优化	优化详情
	35:51 22-08-18 20:36:06	15sec 12		12.8	SUCCEEDED	f4f44776a8b7aaaf675fa3522fa7	bce4	已优化	优化详情
	35:18 22-08-18 20:35:30	12sec 7		7.3	SUCCEEDED	c8ccb17877c1cfbf8d9d8221593	3b7a38	已优化	优化详情
Ē	35:16 22-08-18 20:35:28	12sec 7	admin	8	SUCCEEDED	38d9e1f3bc5809bf8fee839c2f9	4dd4a	已优化	优化详情

2.4 任务列表

任务列表可以通过任务类型,任务ID, Yarn Application Id,任务内容,任务执行时间进行筛选查 询任务,如下如所示:

4	LCC-Optimizer			admin
m	任务列表			
≔				
∷⊟	类型: 全部 🗸 任	务: 请输入任务	Yarn APP ID: 请输入任务的某次执行Yarn A	Q 查询
\odot	内容: 「「清输入任务中的内容关键字 时间区	间:1 2022-07-19 20:55:	2022-08-18 20:55:	admin C 重置
\$				
	任务列表			
	任务	类型	内容	执行次数
	9485c138f4c2d1572636dc64783d5bf9	MapReduce Job	select cs_sold_date_sk,max(cs_sold_date_sk),min(cs_sold	1
	95ee5a98e051686e3dd0d030581b4355	Spark Application	use tpcds;select sum(ss_ext_sales_price) ext_price from stor	3
	20c142d2430894afa1e567bb70c40f14	Spark Application	use tpcds_200g;q75.sql WITH all_sales AS (SELECT d_y	1
	8d11a95bb03f7f615067397e9b302c61	Spark Application	use tpcds_200g;q50.sql select s_store_name, s_compan	1
	a47aea761feb6cb7a243ce381ffb5396	Spark Application	use tpcds_200g;q5.sql WITH ssr AS (SELECT s_store_id,	1
	726ab92d9d327059092c2cdb20116f8b	Spark Application	use tpcds_200g;q74.sql with year_total as (select c_cus	1
	301242ff4fd1bc97078b35131b68798a	Spark Application	use tpcds_200g;q49.sql select 'web' as channel, web.ite	1
	6da4ef49a16363c30afff77db33b0172	Spark Application	use tpcds_200g;q73.sql select c_last_name, c_first_nam	1
	1b178210231e7d550a5c275c6750fe7e	Spark Application	use tpcds_200g;q71.sql select i_brand_id brand_id, i_br	1
	5debd7ea8ded57827d0cfc0d71b8d795	admin Spark Application	use tpcds_200g;eq48.sql select sum (ss_quantity) from st	1 admin
			< 1 2 3 4 5 *** 28 >	10 条/页 >> 跳至 页
		版和美丽。 0		

2.4.1 任务详情

通过点击任务链接,即可进入【任务详情】,如下图所示:



版权声明: Copyright 2022 LCComputing Inc.

- 任务基本参数
 - 任务ID:任务的唯一标识
 - 任务类型:任务类型
 - 任务内容:任务的命令或者SQL语句
- 执行统计: 只展示最近30天的数据
 - 统计值: 总执行次数, 平均执行时长

- 指标变化趋势,包括:运行时长,CPU使用量,内存使用量
- 执行列表
 - 任务视角的执行列表: 该任务的所有执行。

2.1 分析报告

2.1.1 性能对比报告

经过一段时间的优化,我们可以通过【性能对比报告】分析查看任务执行的具体优化情况,如下 图所示: 么 LCC-0

回 主页

≔ 执行列转

三 任务列表

④ 分析报告

愈 优化设置

□ 运维中心



- 对比对象:可以选择【基准批次】和【对比批次】两个执行批次进行对比,查看优化效果。 •
 - 基准批次:用作对比基准的执行批次,一般可以选择未优化的执行批次作为基准(通过"未优 化"标签筛选)。
 - 对比批次:用作对比的执行批次,一般选择同一批任务在不同条件下的执行作为对比(如通过 "已优化"标签筛选)。
- 执行批次选择
 - 任务类型:只有相同的任务类型才可以进行对比。 0
 - 时间范围:通过执行所在的时间进行选择。 0

- 标签:通过自动打上的标签进行选择,比如不同的引擎版本、是否优化等标签,可以同时选择 多个标签。
- 分析内容包括:
 - 对比效果
 - 任务总时长:汇总所有执行的运行时长,进行两个批次的对比
 - CPU使用量:汇总所有执行的CPU使用量,进行两个批次的对比
 - 内存使用量:汇总所有执行的内存使用量,进行两个批次的对比
 - 优化分布
 - 运行时长优化分布
 - CPU使用量优化分布
 - 内存使用量优化分布
 - 优化排名
 - 时长优化排名:查看哪些任务运行时长优化的比例最高
 - 对比报告
 - 同一个任务在两个对比执行之间的具体指标的对比。
 - 对比报告可以下载进行二次分析。

2.2 健康告警

根据任务执行历史可以对任务的执行情况进行健康检查,从而产生异常情况的告警事件,及时发 现问题。点击「健康告警」进入查看。

- 「APP类告警统计」展示统计期间内的不同告警事件的分布与趋势。
- 「告警列表」展示告警明细,包括:
 - 告警ID,告警事件,告警类别,告警对象,告警内容,产生时间

2.3 设置

2.3.1 优化设置

2.3.1.1 黑白名单模式

全局名单模式包括以下两种:

• 全局白名单模式

- 全局黑名单模式
- 2.3.1.1.1 全局白名单模式

全局白名单模式: 仅加入全局白名单的任务才会进行优化;

场景: 在测试阶段逐步增加优化的任务范围。

第一步:登录系统后,前往「优化中心」->「设置」->「优化设置」页面,选择「黑白名单模式」
 Tab页,点击开关,打开「全局白名单」。如下图所示:

👌 LCC-Optimizer	优化中心	尊 chengxi
凹 主页	设置 / 优化设置	
≔ 执行列表 ~	黑白名单模式 任务指纹	
≔ 任务列表		
 ② 分析报告 > 	全局黑白名单模式	
② 设置 ^	全局白名単模式,仅加入全局白名单的任务才会进行优化; 全局黑名单模式,被加入全局黑名单之外的任务都会进行优化。	
优化设置	全局白名單 具体任务名单,您可前往任务列表设置>	
优化规则		
系统配置		
Ē		

第二步:点击右侧提示中的「任务列表设置>」,或者点击左侧菜单栏「任务列表」,通过任务列表
 表复选框选中需要优化的任务,点击「加入全局白名单」按钮并二次确认。如下图所示:

LCC-Optimizer	优化中心							鐐 chengxi
四 主页	任务列表							
≔ 执行列表 >		○ 确定要	ᄡᆕᇳ᠈ᄼ					
≔ 任务列表	任务类型 ⑦: 全部	······································	加重加八公		12-02 16:45:	26 -> 2022	2-12-16 16:45:26	
② 分析报告 >	任务ID _{chen} gi ^{Xi})	青输入	4X)	開定	查询 _o	6月94里置	∨展开	
② 设置								
	任务列表						退出全局白名单	加入全局白名单
								清空
	■ 任务	类型	内容	全局白名单	隔离次数 💲	执行次数 🍦	执行频率 ② 💲	内存使用量波动 ② 💲
	chengxi testtestettest	Spark Application		• 否	0	13	3次/天	79
	v7_286b8554	196f6 MapReduce Job	L29wdC9oYWRvb3	• 否	0	11	11次/天	67
	v7_a4428609	ae96 MapReduce Job	/opt/hadoop-exam	• 否	0	11	0次/天	5
E	v7_04bab14f	85ff8 MapReduce Job	/root/changfan/ben	• 否	0	10	0次/天	9

📌 全局黑名单模式:除全局黑名单之外的任务都会进行自动优化。

场景:在**生产阶段**逐步剔除优化效果暂时不佳的任务。

第一步:登录系统后,前往「优化中心」->「设置」->「优化设置」页面,选择「黑白名单模式」
 Tab页,点击开关,打开「全局黑名单」。如下图所示:

\land LCC-Optimizer	优化中心	龄 chengxi
□ 主页	设置 / 优化设置	
≔执行列表 ~	黑白名单模式 任务指纹	
≔ 任务列表		
② 分析报告 ~	全局黑白名单模式	
② 设置 ^	全局白名单模式,仅加入全局白名单的任务才会进行优化; 全局黑名单模式,被加入全局黑名单之外的任务都会进行优化。Chengy	
优化设置	▲局黑名单 具体任务名单,您可前往任务列表设置>	
优化规则		
系统配置		
Ē		

第二步:点击右侧提示中的「任务列表设置>」,或者点击左侧菜单栏「任务列表」,通过任务列表
 表复选框选中需要优化的任务,点击「加入全局黑名单」按钮并二次确认。如下图所示:

싥 LCC-Optimizer	优化中心				段 chengxi
凹 主页	任务列表				
≔执行列表 ~		() 确定要批量物) 化			
≔ 任务列表	任务类型 ⑦: 全部	明定安加重加八公		12-02 17:01:18 - 202	22-12-16 17:01:18 📋
② 分析报告 >	任务ID _{ch} engx ¹) 请输入		取 消 備 定	查询 _{ch} er ^g 重置	✓展开 cheng ^{xi}
	任务列表				退出全局黑名单加入全局黑名单
	● 已选择1项				清空
	王 任务	类型内容	全局黑名单	隔离次数 💠 执行次数 🗧	執行频率 ⑦ ↓ 内存使用量波动 ⑦ ↓
	chengxi testtestettest	Spark Application	○ 否	o 13	3次/天 79
	v7_286b855496f6	MapReduce Job L29wdC	9oYWRvb3 ● 否	0 11	11次/天 67
	v7_a4428609ae96	MapReduce Job /opt/hado	oop-exam ● 否	0 11	0次/天 5
ē	v7_04bab14f85ff8	MapReduce Job /root/cha	ngfan/ben ● 否	0 10	0次/天 9

2.3.1.2 任务指纹

进入任务指纹管理页面:前往「优化中心」->「设置」->「优化设置」页面,选择「任务指纹」 Tab页进入,再该页面可以对【任务指纹】做如下管理:

- 添加新的任务指纹版本: 在添加页面可以对当前指纹算法进行测试运行, 查看结果是否符合预期。
- 编辑任务指纹:对现有的任务指纹进行编辑
- 启用任务指纹:由于任务指纹只能启用一个作为当前版本,启用后原指纹将自动停用并迁移。

2.3.2 优化规则

本产品针对不同优化任务预设了一些列优化规则,如下图所示:

么 LCC-Op	otimizer									admin
四 主页		优化设置 / 优	化规则							
Ⅲ 执行列表	~									
Ⅲ 任务列表		规则列	表							
② 分析报告	~	编号。	规则名称	描述	类型 admin	对象	启用状态	更新时间	操作	
◎ 优化设置	^	1	Executor个数优化		Spark Application	任务		2022-08-18 17:12	配置 黑白名单	
优化规则		4	Executor内存优化		Spark Application	任务		2022-08-18 17:12	配置 黑白名单	
优化名单		7	Executor CPU优化		Spark Application	任务		2022-08-18 17:12	配置 黑白名单	
口、法维由心	Ū.	16	Reduce内存设置		MapReduce Job	任务		2022-08-18 17:12	配置 黑白名单	
區 连维中心		19	Map内存设置		MapReduce Job	任务		2022-08-18 17:12	配置 黑白名单	
		160	Spark并行度优化		Spark Application	任务		2022-08-18 17:12	配置 黑白名单	
		8								条/页 ^{solmin}
Ξ					版权声明:C	Copyright 2022 LC	CComputing Inc.			

- 启停优化规则
 - 可以通过启用状态变更优化规则是否启用。
- 配置优化规则
 - 可以通过配置变量的方式修改优化规则配置。
- 黑白名单:可以对具体优化规则设置黑白名单,黑白名单只能二选一。
 - 黑名单:表示名单中的任务不会参与该优化规则的优化。

• 白名单:表示只有名单中的任务会参与该优化规则的优化。

2.3.3 系统配置

「系统配置」统一管理各类系统的配置参数,点击「设置」->「系统配置」进入管理。

- 添加配置:点击「添加配置」弹出添加配置对话框,选择「配置类别」「集群」「版本」,输入 「配置项」和「配置值」进行添加。
 - 配置值支持两种形式: Value值 和 Yaml 文件
- 「配置列表」中可以对配置项进行修改和删除操作。

2.3.4 诊断查询

当遇到系统问题或者需要查看历史时,可以通过「诊断查询」进行数据明细的查询和下载。可以 查询的类型包括:

- YARN Application,查询范围包括:Loki日志,MySQL数据记录,ClickHouse数据记录。
- 任务,查询范围包括:Loki日志,MySQL数据记录。
- 服务日志,查询范围包括: Loki日志。

3. 快速入门

3.1 安装部署

参见《LCC-Optimizer部署运维手册》

3.2 查看历史作业

部署安装完成后系统会进行历史作业的采集,10分钟左右(默认会从history server抓取过去一段 时间的job数据,具体等待时间取决于历史作业的多少)可以在【执行列表】和【任务列表】中查看到 系统现有任务的执行情况。

3.3 开启优化

- 设置优化名单:在【优化名单】中开启生效优化白名单,这样保证只有在优化名单中的任务才会进行自动优化。
 - 特别注意:初始使用时,强烈建议打开全局白名单以灰度的方式逐步开启优化。否则默认在 关闭全局名单的情况下,所有作业都会自动进行优化!

• 打开优化规则:在【优化规则】中开启想要的优化规则。

3.4 提交测试作业

- 先在客户的自己的大数据系统中提交一个测试作业,默认首次提交不进行优化
- 在【执行列表】确认可以看到该测试作业
- 在【优化名单】中添加该测试任务
- 再次提交相同的测试任务,此时任务由于是第二次运行,会进行自动优化。

3.5 查看任务详情

- 测试作业执行完成后,在【执行列表】中可以看到该测试任务的执行。
- 在【执行列表】中通过点击【任务ID】可以查看到该测试任务的详情页面。

3.6 查看优化报告

 在【性能对比报告】中,通过条件选择这两次测试任务不同的执行批次,可以对比查看具体的优化 效果。

4. 常见问题

4.1 如何关闭优化?

有两种程度的关闭,在大多数情况下,用户只需要关闭优化功能(参照1)即可。

1. 关闭优化功能

可以通过产品界面关闭优化功能。在关闭优化功能后,对应的任务不会受到优化的作用; LCC-Optimizer仅会为其同步History的执行记录等信息。

- a. 关闭所有任务的优化
 - i. 开启全局白名单,清除优化名单中的所有任务。该设置表示没有任务会进行自动优化。
- b. 仅关闭指定任务的优化
 - 在开启全局优化名单的情况下,从优化名单中删除对应的任务,该任务就不会再进行自动优化
 - 在没有开启全局优化名单的情况下,可以在所有优化规则中将指定任务添加到黑名单中,则 该任务不会进行这些优化规则的优化。

2. 完全关闭LCC-Optimizer

由于对作业的优化应用发生在提交作业处安装的HOOK,所以通过关闭HOOK,可以完全关闭LCC-Optimizer产品。关闭HOOK后,作业提交过程不再有LCC-Optimizer任何作用、对应的执行记录不 再同步到LCC-Optimizer。

具体针对不同引擎关闭HOOK的方法,参见《LCC-Optimizer部署运维手册》

4.2 什么样的任务合适优先加入优化?

- 可以优先选择具备如下特征的任务进行优化设置
 - 执行具备周期性的任务
 - 处理数据相对稳定的任务
- 在任务列表中去搜索Yarn Application ID找到对应任务ID
 - 。 然后在优化名单中通过任务ID加入优化白名单

4.3 作业优化不符合预期怎么处理?

- 可以通过优化名单暂时关闭对该任务的自动优化
- 联系我们针对该任务进行具体诊断