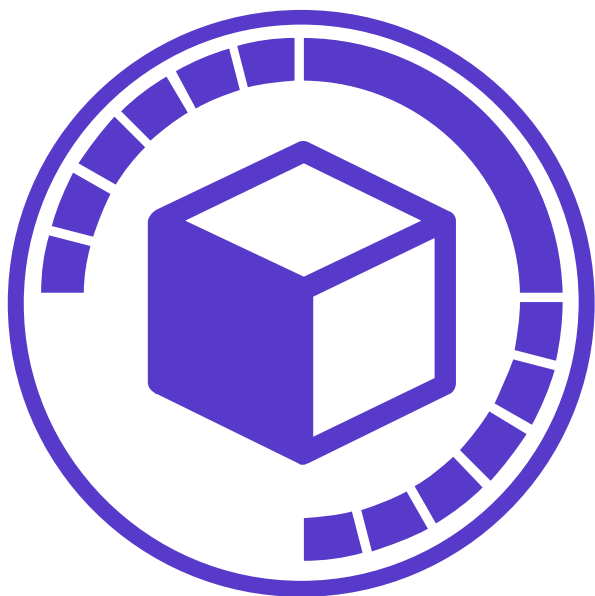


## 欢迎加入阿里云 Dala Lake Analytics 团队



如果想及时了

解Spark、Hadoop或者Hbase相关的文章，欢迎关注微信公共帐号：iteblog\_hadoop

回望过去10年，数据技术发展迅速，数据也在呈现爆炸式的增长，这也伴随着如下两个现象。一、数据更加分散：企业的数据是散落在不同的数据存储之中，如对象存储OSS，OLTP的MySQL，NoSQL的Mongo及HBase，以及数据仓库ADB之中，甚至是以服务的形式提供。二、2015年之前，只有大公司在谈数据赋能业务，如今，随着云的普及，外部及内部等原因，企业主动或者被动数字化，中小企业也在大量使用数据赋能业务，这就需要云提供更加实惠的所见即所得的分析能力。为了解决数据分散带来的分析困境、普及分析能力、挖掘隐藏的数据价值，阿里云数据库分析Data Lake Analytics应运而出。

工作地可以open to 北京，杭州，深圳

产品地址：<https://www.aliyun.com/product/datalakeanalytics>

目前已经服务数千家企业，为广大的大中小企业 探索数据湖 挖掘隐藏的数据价值。

**平台岗位：**

**岗位职责：**

1. 分析型数据库-商业化管控平台研发，面向用户提供简洁、高效、自助化、自动化的数据库商业管控平台。
2. 分析型数据库-监控告警体系研发，为多种分析型数据库提供一致的、全方位的、可自定义的、自运维的监控告警与分析平台。
3. 分析型数据库-数据接入平台研发，为多种分析型数据库提供多数据源、多数据格式灵活适配、低成本、便捷的数据接入和转换能力。
4. 分析型数据库-商业化输出研发，包括公有云、专有云、混合云等多种形态的商业化输出研发。

### 职位描述：

1. 精通java服务端编程，熟悉前端react或vue等框架技术，有3年以上大型互联网开发经验
2. 熟悉linux、docker以及k8s等系统和容器技术，对进程间通信，内存管理，网络编程等比较了解。
3. 精通tcp/ip、http网络协议和java中间件技术，对MQ、RPC、微服务、数据层技术有丰富的实践经验。
4. 熟悉常用的开源监控体系架构，比如：prometheus、grafana等，对如何定义数据库监控指标有实战经验。
5. 在分布式数据库、OLAP数据分析领域有开发和优化经验者优先，熟悉Impala/Presto/Greenplum等MPP架构数据库产品。
6. 良好的沟通和团队协作能力，能够熟练编写各类技术文档。

### 内核岗位：

#### 岗位职责：

数据湖分析 Data Lake Analytics 产品专注提供Serverless化的分析能力，提供联邦查询能力，构建数据湖存储，以提供极低分析存储成本服务广大客户。

- 1、负责产品云化，Serverless化，支持统一meta，元数据发现等关键技术
- 2、负责计算引擎的开发工作，支持SQL优化，执行及相关连接器实现及优化，新硬件加速等关键突破等
- 3、负责数据湖存储研发，实现对对象存储的适配，增量更新的数据格式，缓存加速等
- 4、负责为技术&产品代言，在各种场景打造影响力

## 职位描述

- 1、熟悉JAVA，较好的系统知识、性能意识、清晰的代码能力
- 2、在大数据计算和存储相关领域有3年以上的研发经验，并有大规模落地应用者优先
- 3、精通分布式存储、分布式一致性、SQL计算、迭代计算的一个或者多个领域者优先
- 4、熟悉开源的Spark/Kylin/Hadoop/Flink/Presto或类似分布式计算引擎者优先
- 5、熟悉开源的HBase/cassandra/Kudu/Infobright/Rocksdb或类似列式存储引擎者优先
- 6、熟悉K8s，有一定的调度算法经验优先
- 7、具备一定的抗压能力、自我驱动力、皮实、自省、良好的团队沟通及学习能力

联系人：封神

微信：fengshenwukong

邮件：dragon.caol@alibaba-inc.com

Dala Lake Analytics介绍，更多细节详聊：

Data Lake Analytic主要提供以下几个方面的能力：

### 一、元数据的统一管理

异构数据源中关系、文档、键值等数据散落在各个数据库及存储服务中，形成了数据孤岛，异构数据计算的一个核心目标是帮助用户挖掘异构数据关联后的数据价值。而在进行异构数据计算之前，首先需要构建异构数据源的统一元数据视图，支持数据统一管理、数据清洗与转换，并保障数据安全。

### 二、Serverless化弹性

随着企业的数字化转型，各种异构数据不论是数据类型，还是数据量都在快速增长。伴随着数据的快速增长，需要有更多的计算资源来做数据的探索及分析；同时数据湖的分析计算请求具有随机性，没有特定时段及资源需求量的预先规划。需要有一套云原生Serverless形态的弹性计算资源管理平台，来解决瞬时大量计算资源的需求。

### 三、联邦分析引擎

分析引擎需要提供外接数据源的抽象接口，该接口能方便连接各类异构数据源（宽表、文档、对象存储、键值、RDBMS等），并**具备数据库的体验**

。该接口需要能够表达和充分利用存储引擎本身具备的能力，以及将计算引擎算子下推到数据源，减少不必要的网络传输。比如，在读取数据库中的数据时，直接将计算引擎的过滤或者聚合算子传递给数据库，数据库只需简单本地计算即可大大降低数据的传输量。分析引擎需要利用高效的数据格式来读取数据湖中的数据。该格式需要具备高效的序列化反序列化能力与压缩能力，尽可能减少序列化与反序列化开销，减少网络开销，同时该格式需要对CPU计算友好，以方便执行引擎执行层能更好地利用CPU的硬件能力。要满足这些要求，内存结构需要是列式的，相比于行式，列式天然更适合分析，容易实现裁剪和向量化等技术。

**本博客文章除特别声明，全部都是原创！**

**原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。**

**本文链接: [【】](#)（）**