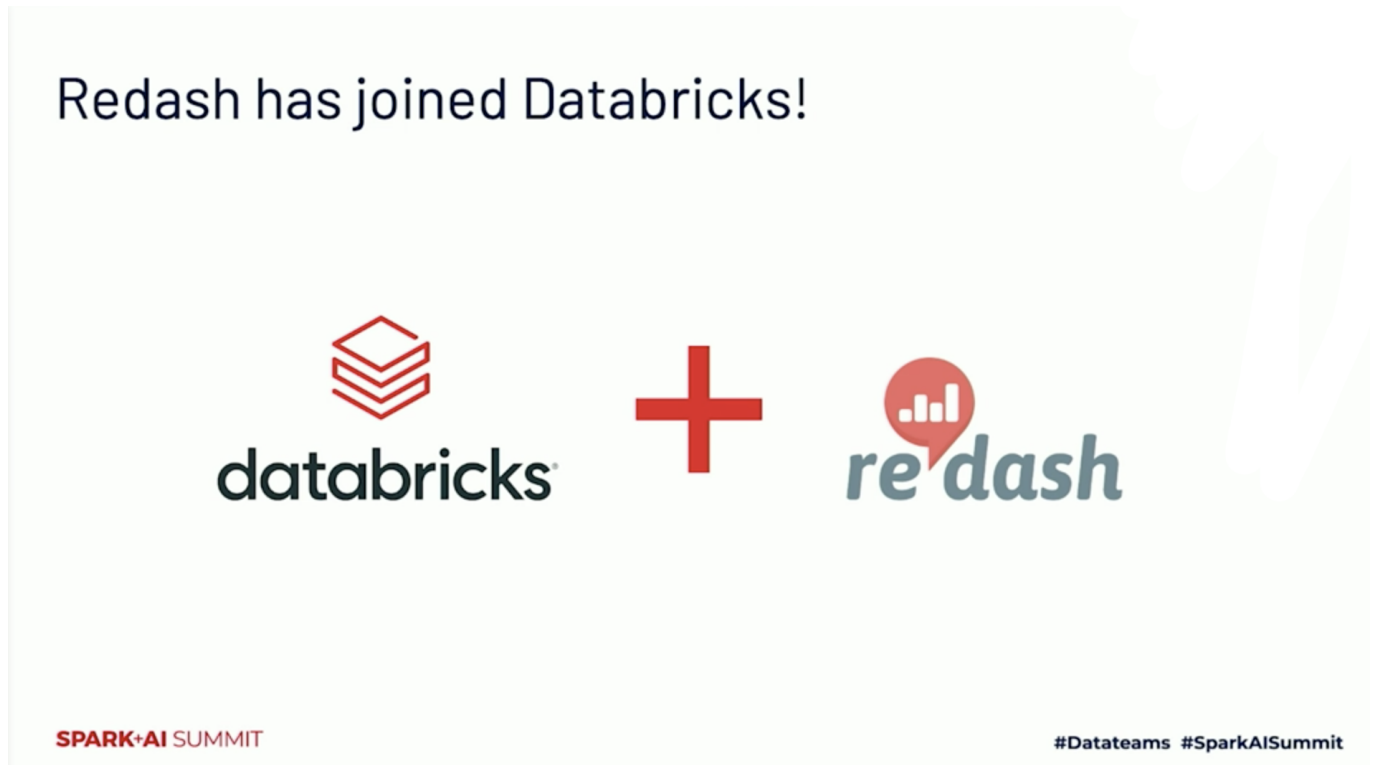


Spark 背后的商业公司收购的 Redash 是个啥？

在2020年6月24日的 Spark AI summit Keynote 上，数砖的首席执行官 Ali Ghodsi 宣布其收购了 [Redash 开源产品](#) 的背后公司 Redash！



如果想及时了解Spark、Hadoop或者HBase相关的文章，欢迎关注微信公众号：iteblog_hadoop

通过这次收购，Redash 加入了 Apache Spark、Delta Lake 和 MLflow，创建了一个更大、更繁荣的开源系统，为数据团队提供了同类中最好的工具。为什么数砖会收购 Redash 公司？

Redash 是什么？

Redash 是一个可以协作的可视化和仪表盘平台，旨在让任何人，无论其技术水平如何，都可以在团队内部或跨团队共享见解。一直使用 SQL 的用户可以利用 Redash 来探索、查询、可视化和共享来自任何数据源的数据。世界各地每天有数千个组织的数百万用户使用 Redash 来做出数据驱动的决定。

Redash 主要包含以下特点：

- Query editor: 使用模式浏览器和代码提示来快速编写 SQL 和 NoSQL 查询。
- 可视化和仪表盘：使用拖放创建漂亮的可视化界面，并将它们合并到单个仪表板中。
- 分享：
通过共享可视化仪表盘及其关联的查询，可以轻松实现协作，并支持对报告和查询的同行

审查。

- 进度刷新：在用户定义的间隔内自动更新图表和仪表板。
- 报警：定义条件，并在数据更改发出相应的警报。
- REST API: 所有的事情都可以通过 UI 和 REST API 完成。
- 支持广泛的数据源：提供可扩展的数据源 API，并且内置提供大量常见 SQL、NoSQL 数据库和平台的数据源接口。

下面我们来看看 Redash 的可视化界面

The screenshot shows the Redash interface for a dashboard titled "Patient Outcome by Age Group". The interface includes a sidebar with a schema browser, a central SQL editor, and a results table. The SQL query is as follows:

```
1 select age_group, outcome, count(*) from (  
2 select  
3   round(year(current_date())-birth_year, -1) as age_group,  
4   case  
5     when (released_date is not null) then "Recovered"  
6     when (deceased_date is not null) then "Death"  
7     else "In Progress"  
8   end as outcome  
9 from  
10  covid.patient_delta  
11 where birth_year is not null  
12 ) x  
13 group by age_group, outcome  
14 order by age_group, outcome
```

The results table displays the following data:

age_group	outcome	count(1)
0	In Progress	1
10	In Progress	6
10	Recovered	1
20	In Progress	67
20	Recovered	5
30	In Progress	109
30	Recovered	11
40	Death	2

The interface also shows a "Refresh Schedule" set to "Every minute" and a "Refreshed just now" status.

轻松地对 Delta Lake 或任何其他数据源中的数据进行 SQL 查询

☆ Patient Outcome by Age Group Show Data Only

Demo

Search schema...

- covid.covid_19_containment_measu...
- covid.covid_19_ecdc
- covid.covid_19_geographic_distrib...
- covid.covid_confirmed_usafacts_csv
- covid.covid_usafacts_csv
- covid.covid_with_census
- covid.patient_csv
- covid.patient_delta
- covid.policy
- covid.route_csv
- covid.route_delta
- covid.search_trend
- covid.sk_case
- covid.sk_patient_info
- covid.sk_patient_route
- covid.sk_region
- covid.time
- covid.time_age
- covid.time_csv
- covid.time_series_covid19_confirm...
- covid.time_series_covid19_deaths_...
- covid.time_series_covid19_recover...

Add description

James Nguyen created a minute ago

James Nguyen updated a minute ago

Refresh Schedule Every minute

```

1 select age_group, outcome, count(*) from (
2 select
3 round(year(current_date())-birth_year, -1) as age_group,
4 case
5 when (released_date is not null) then "Recovered"
6 when (deceased_date is not null) then "Death"
7 else "In Progress"
8 end as outcome
9 from
10 covid.patient_delta
11 where birth_year is not null
12 ) x
13 group by age_group, outcome
14 order by age_group, outcome

```

Save Execute

Table Outcome by Age Group x Add Visualization

23 rows 2 seconds runtime Refreshed a minute ago

将结果进行可视化操作

★ Global Patient Dashboard Saved Done Editing

Global Cases 2 minutes ago

Outcome by Age Group - Outcome by Age Group 2 minutes ago

Avg/Stddev Age vs infection reasons 2 minutes ago

Cohort Analysis of patient recovery (with parameter) 2 minutes ago

Begin date range 2020-01-28

Users	1	2	3	4	5	6	7	8
11	0.00%	0.00%	27.27%	63.64%	90.91%	100.00%	100.00%	100.00%

分享计算出来的结果

Redash and Databricks

Databricks 是通过一些早期客户第一次听说了 Redash。随着时间的推移，越来越多的用户要求数砖改进 Databricks 和 Redash 之间的集成。所以在今年早些时候，数砖邀请了 Redash 的创始人兼首席执行官 Arik Fraimovich 访问 Databricks，讨论两个项目如何合作，如何让数据更容易分析。从这里可以看出，很多东西是客户需求推动的，客户需求才是最重要的。

从前面对 Redash 的介绍，可以看出 Redash 和数砖的产品在功能上有很多共同点，他们的核心价值观都是试图使数据从业者能够更容易地围绕数据进行协作，并使所有团队都能民主化地访问数据，所以这也是数砖收购 Redash 的一个很大的原因。

Databricks 集成 Redash 之后使得数据科学家和数据工程师可以轻松查询和可视化 Delta Lakes 和其他数据源中数据。

Redash 与现有的 Databricks 平台可以达到无缝集成：Databricks 运营的所有数据中心都可以使用 Redash；另外身份管理和数据治理是统一的，无需额外配置；最后，这两个产品的 catalogs 和元数据也都是共享的。

Databricks 的愿景是提供一个统一的数据分析平台，帮助公司的每个数据团队解决世界上最棘手的问题——包括数据分析师、数据工程师、数据科学家和机器学习工程师。通过为每个团队提供他们自己工作所需的工具，同时拥有一个可以协作的共享平台，每个数据团队都可以在一起获得成功。

收购 Redash 之后，Databricks 将更好的构建 lakehouse，它将数据湖和数据仓库的最佳功能结合在一起，在统一的体系结构中，每个团队都可以在相同的数据源上一起工作。

本博客文章除特别声明，全部都是原创！
转载本文请加上：转载自过往记忆 (<https://www.iteblog.com/>)
本文链接: 【】 ()