



# Running Big Data Applications in production with Airflow + Firebolt



Boaz Farkash  
Firebolt CPO

# Agenda

---

- The shift from Analytics to Data Applications
- Modern Cloud DWs to the rescue
- The shift from ETL to ELT
- Airflow to rule it all
- Firebolt
- Example

# The shift from Analytics to Data Applications

---

Once upon a time...	Today
GB scale	TB scale
Batch, Historical	Real-time, Operational
Internal	Internal, Customer-facing
Low concurrency	High concurrency
Multi-second response time	Sub-second response time
Delivered by: <ul style="list-style-type: none"><li>• DW specialists/Architects</li><li>• Analysts</li></ul>	Delivered by: <ul style="list-style-type: none"><li>• <b>Data Engineers</b></li><li>• <b>Developers</b></li><li>• Analysts</li></ul>

# The shift from Analytics to Data Applications

---

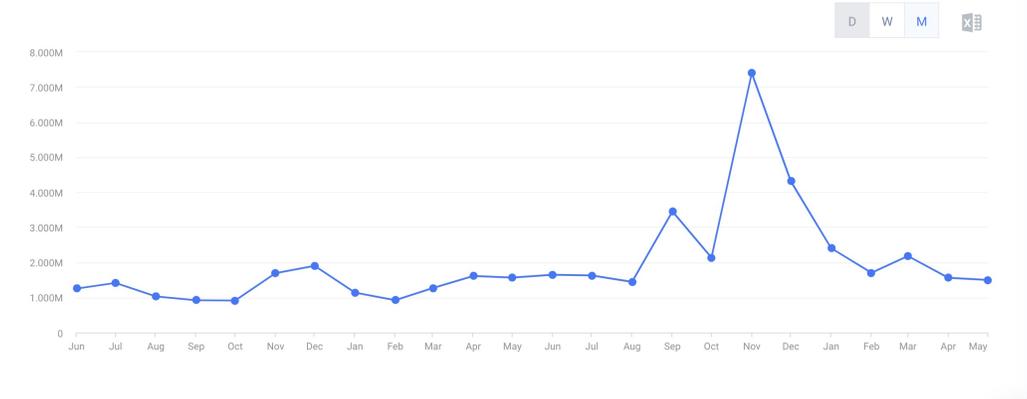
## **Demand for data-applications is exploding**

- Data is literally the product for more and more companies
- Companies compete over deeper/faster analytics to end users
- Home-grown data-applications for operational excellence are on the rise
- Traditional Analytics has been commoditized, Data-Applications are still a challenge

### Traffic & Engagement

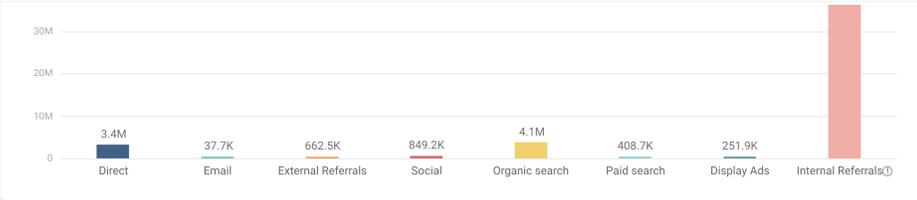
Jun 2019 - May 2021 (24 Months) Worldwide

Avg. Monthly Visits	Avg. Page Views	Avg. Pages Per Visit	Avg. Visit Duration	Bounce Rate	Segment Share
1.958M <span>↑ 19.84%</span>	6.270M <span>↑ 5.83%</span>	3.20 <span>↓ 12.29%</span>	00:02:48 <span>↓ 18.54%</span>	45% <span>↑ 22.39%</span>	0.18% <span>↓ 0.80%</span>



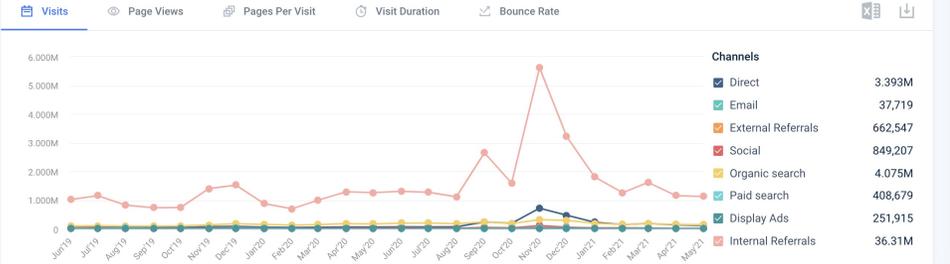
### Marketing Channels

Jun 2019 - May 2021 (24 Months) Worldwide



#### Marketing channels traffic and engagement

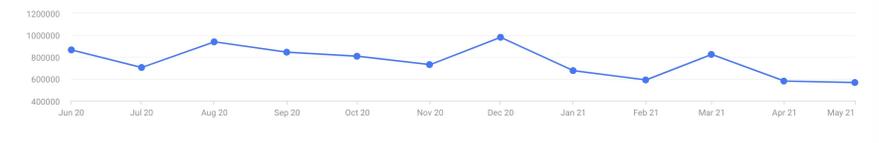
Jun 2019 - May 2021 Worldwide Desktop



### Keyword Generator

Jun 2020 - May 2021 (12 Months) United States

#### YouTube Volume Trend for "fortnite"



Phrase Match (201,698) Related Keywords (156,013)

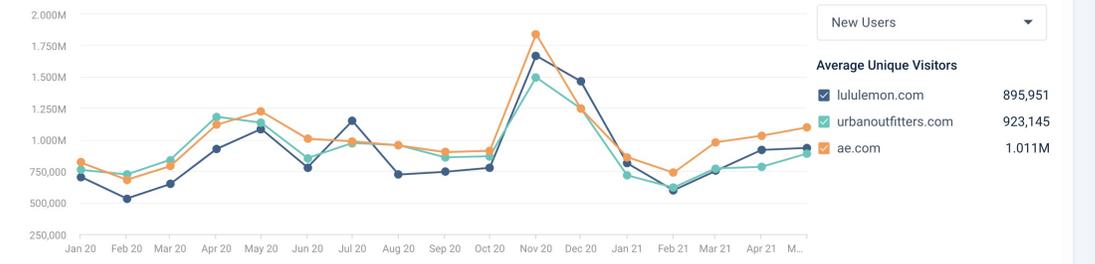
	Keyword (156,013)	Relevance Score ↓	Volume	Volume Trend	Clicks	Top Country
1	fortnite	●●●●●	230,369		240,793	United States
2	typical gamer	●●●●●	1,897,565		1,287,655	United States
3	fresh	●●●●●	2,919,849		2,246,950	United States
4	syphepk	●●●●●	1,492,346		1,078,822	United States
5	tg plays	●●●●●	1,077,238		908,246	United States
6	fortnite live	●●●●●	311,460		464,975	United States
7	fortnite item shop	●●●●●	934,451		1,079,768	United States
8	haxx	●●●●●	4,270,911		2,771,406	United States

### New vs. Returning

Jan, 2020 - May, 2021 (Custom) United States Desktop

#### New vs. Returning Users

Jan 2020 - May 2021 United States Desktop



Filters DATE FILTER is from 2018/03/01 until 2018/03/07 App Name is any value Acc Name is any value Platform is any value Media Source is "3bc36bb9f1bde2703684dae170581dc"

## 20,965,924

Clicks

## 1,454,042

NOI

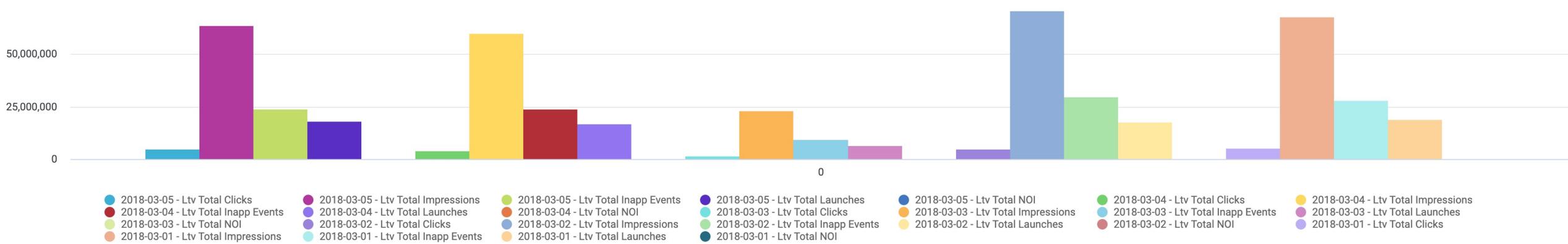
## 286,382,694

Impressions

## 115,941,404

Inapp Events

### Daily Totals



App Name	Media Source	Platform	Total Clicks	Total Impressions	Total Inapp Events	Total Launches	Total NOI
64d26b5fb55376b039c97db79e...	3bc36bb9f1bde2703684dae17058...	android	8,270,409	89,349,895	3,269,035	7,877,190	360,999
70928775a1f5be26817f0b916355f...	3bc36bb9f1bde2703684dae17058...	android	1,097,707	9,097,533	285,487	851,117	34,599
c08a2d14f5cffe1e380e9522f1018...	3bc36bb9f1bde2703684dae17058...	ios	1,092,151	14,439	648,430	358,398	21,439
9cae489d880fed31690c1972bd37...	3bc36bb9f1bde2703684dae17058...	android	503,349	2,312	167,868	287,487	12,976
8c3ee640dd7614d981746774201a...	3bc36bb9f1bde2703684dae17058...	ios	439,922	80,328	45,711	264,036	9,723
22536b1e21c6f133760a7664717c...	3bc36bb9f1bde2703684dae17058...	android	320,930	5,461,301	18,916	65,247	9,775
2aa526d02c0c6da0c2dcf150ddb0...	3bc36bb9f1bde2703684dae17058...	android	247,008	6,211,940	26,161	84,208	8,807
54fbbd537aefdfb105b07cacb6698...	3bc36bb9f1bde2703684dae17058...	android	226,053	2,785,743	406,938	348,030	5,616
079ae651d8344066c941143c4a08...	3bc36bb9f1bde2703684dae17058...	ios	216,010	5,356,054	273,402	1,334,482	17,003

# Modern cloud DWs to the rescue

---

## What is a modern cloud DW anyway?

- Cloud-Native, SaaS
- Built for data-lake architectures
- Elasticity through **decoupled storage & compute**



# Coupled / Decoupled storage & compute

---

Coupled	Decoupled
Everyone is competing for the same resources	Simply spin up isolated resources per workload
Queue and priority management hell	Simply spin up isolated resources per workload
One giant cluster always up	Start/stop various sizes of clusters as needed
Scaling requires complex migrations	Scaling is 1-click
Doesn't fit development practices	Unlocks CI/CD practices



# The shift from ETL to ELT

---

ETL	ELT
Scale-out workloads only possible in the data lake	Scale-out workloads easy to run in the CDW
Setting up and developing ETL pipelines is labor intensive and complex	For ELT all you need is SQL – simple and quick
Running transformation logic in the DW is risky and can slow production queries / take cluster down	Running transformation logic is done on isolated compute without risk
Slow iterations to implement changes, cross team effort	Same team implement changes with SQL only, fast iterations

# Benefits of separate compute & ELT

---

## Unlocking a new development paradigm for data



Multiple environments and easier environment setup



Easier data testing and data testing becomes standard



Bring Devops best practice to DataOps with CI/CD, version control, and more

*Allows teams to move faster with more confidence*

# Airflow to rule it all

---

- The modern CDW is the enabler of modern Data-Ops
- The ecosystem is exploding with tools for Data-ops



-  Apache Airflow orchestrates everything

# The shift from Analytics to Data Applications

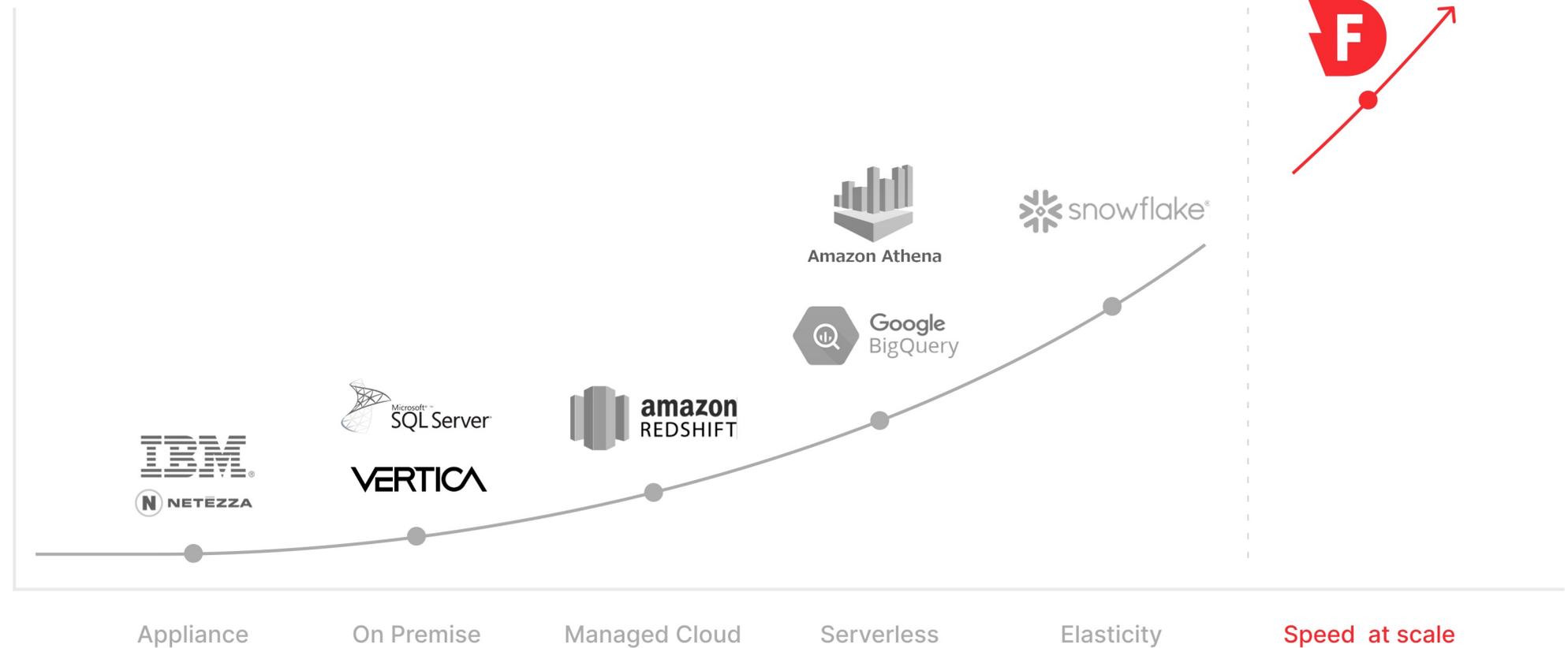
---

Once upon a time...	Today
GB scale	TB scale
Batch, Historical	Real-time, Operational
Internal	Internal, Customer-facing
Low concurrency	High concurrency
Multi-second response time	Sub-second response time
Delivered by: <ul style="list-style-type: none"><li>• DW specialists/Architects</li><li>• Analysts</li></ul>	Delivered by: <ul style="list-style-type: none"><li>• <b>Data Engineers</b></li><li>• <b>Developers</b></li><li>• Analysts</li></ul>



# A new leap in the evolution of data warehouses

Simplicity





# The Firebolt difference

1

## Speed

Up to 182x faster speed at scale with optimized storage, indexing and engines

2

## Scale

Elastic scale at speed across ELT, semi-structured data, and thousands of users

3

## Efficiency

Do more with less. 10x price-performance advantage through greater HW efficiency & choice

*The fastest way to build the fastest data experiences*



# Speed at Scale

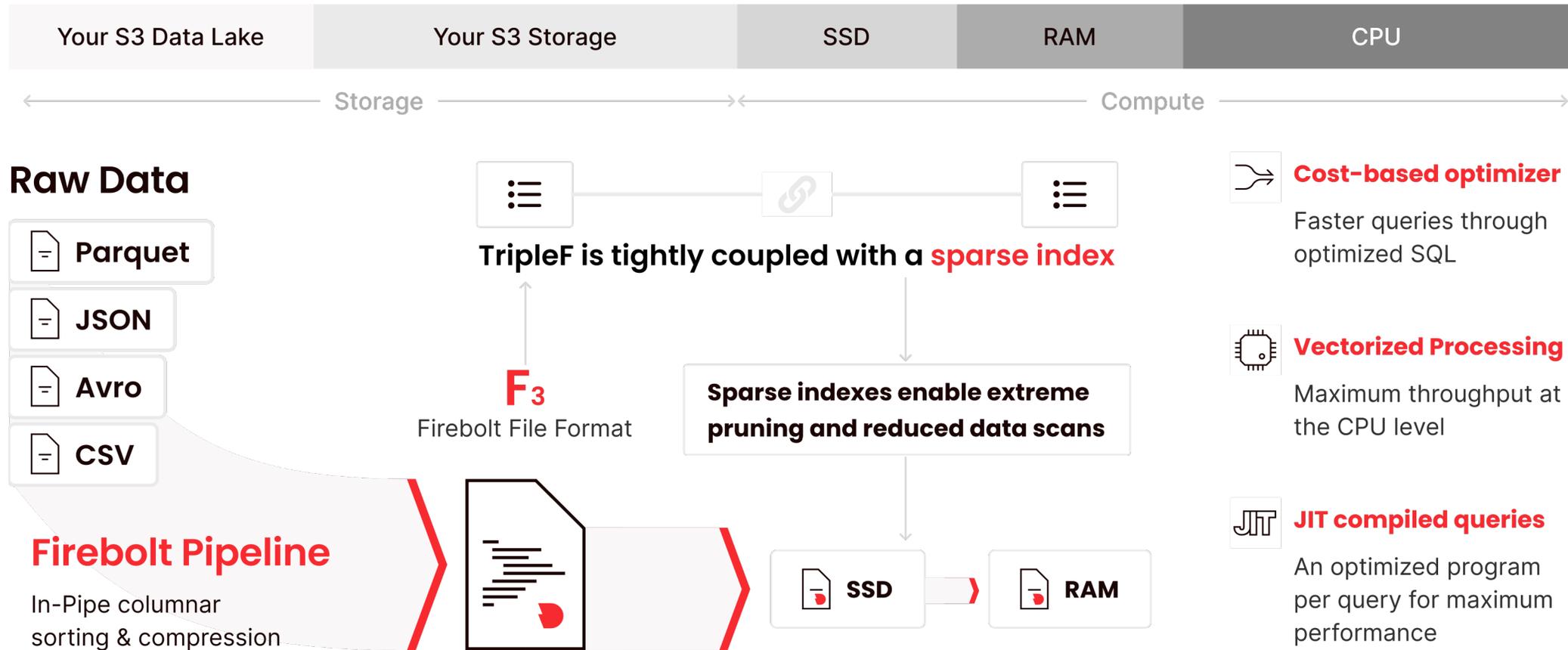
## FiveTran benchmark

- 1 TB of data
- 4 billion rows
- 1 user, 1 query at a time
- 8-11 seconds

**Firebolt** – Consistently delivers **sub-second** performance over dozens of TBs and beyond



# Getting the most out of your compute

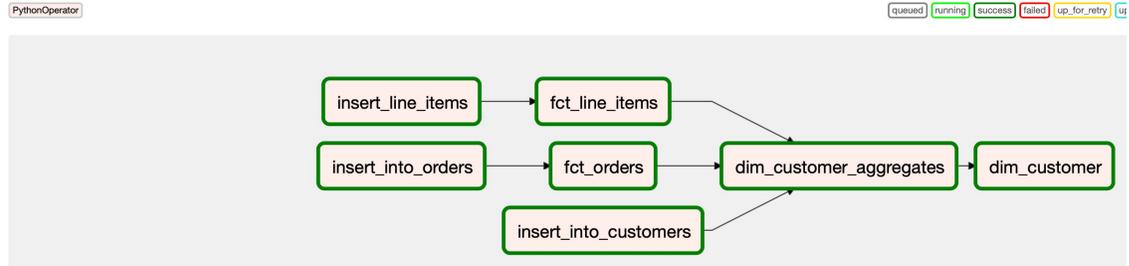


## DAG: demo\_ELT\_process

Tree View Graph View Task Duration Task Tries Landing Times Gantt Details Code

DAG Docs

2021-07-13T16:17:40Z Runs 25 Run manual\_2021-07-13T16:17:39.923597+00:00 Layout Left > Right Update

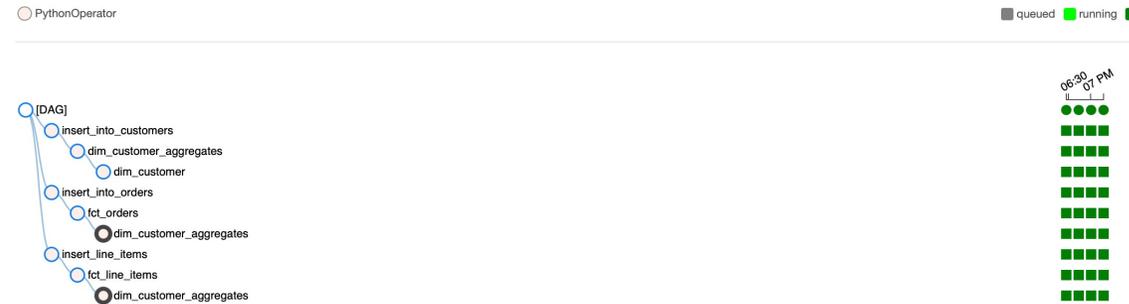


## DAG: demo\_ELT\_process

Tree View Graph View Task Duration Task Tries Landing Times Gantt Details Code

DAG Docs

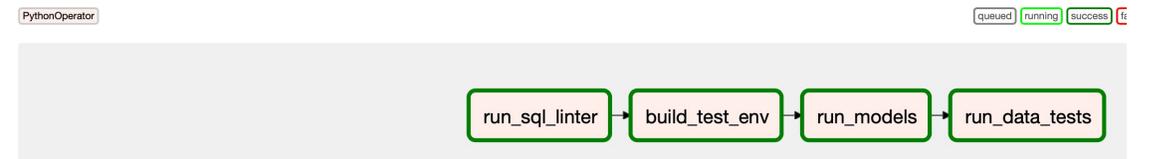
2021-07-13T16:17:39Z Runs 25 Update



## DAG: demo\_CICD\_process

Tree View Graph View Task Duration Task Tries Landing Times Gantt Details Code

2021-07-13T16:21:43Z Runs 25 Run manual\_2021-07-13T16:21:42.052444+00:00 Layout Left > Right Update



## DAG: demo\_CICD\_process

Tree View Graph View Task Duration Task Tries Landing Times Gantt Details Code

2021-07-13T16:21:42Z Runs 25 Update





**Thank you for your time**