

VIENNA DATA
SCIENCE TOOLS

meetup

green
....
tube
NOVOMATIC INTERACTIVE

NOVOMATIC



&



Google
Big Query



The Modern Data Stack

Bogdan Pirvu, Head of Data Science

NOVOMATIC



Machine Learning



0:10 / 7:14











THE MODERN DATA STACK



Big Data

The 3 V's

- Volume
- Variety
- Velocity
- Veracity
- Value

Big Data Technology Stack





The Modern Data Stack

The Cloud DWH

- Arbitrary Scale
- Decoupling of costs for Storage and Compute
- SQL interface
- Providers



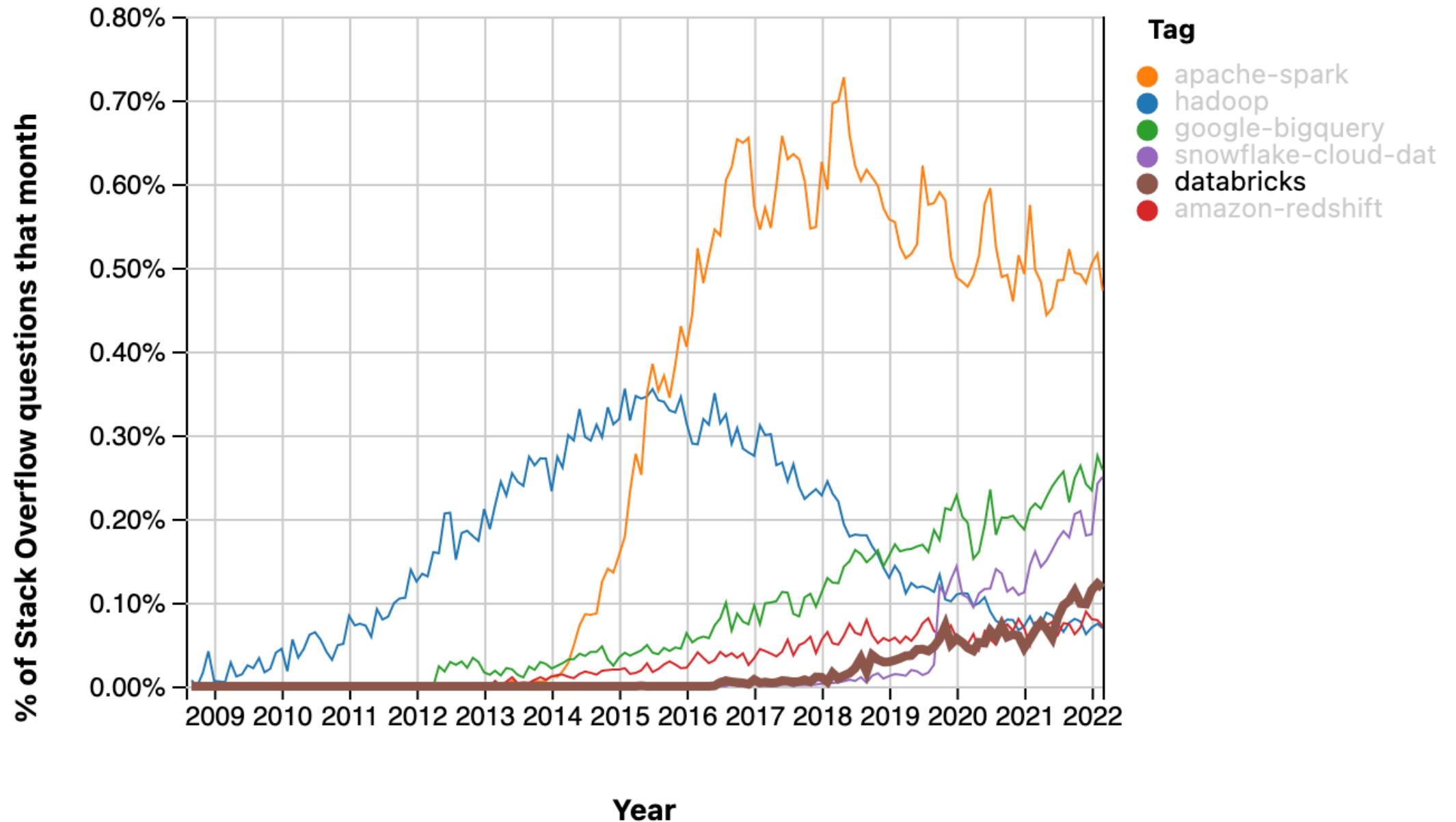
Data Storage – History in a Nutshell

- 1979 Oracle
- 1989-1993 MS SQL Server, IBM DB2
- 1995-1996 MySQL, PostgreSQL
- 2005 M. Stonebraker - One Size Fits All: An Idea Whose Time Has Come and Gone¹
- 2005-2011 Hadoop, MongoDB, CouchDB, Cassandra, Cloudera, MapR, Hortonworks
- 2011-2014 BigQuery, Amazon Redshift, Snowflake
- 2014 Apache Spark
- 2015-2017 Google Spanner, CockroachDB, FaunaDB
- 2019 Databricks Delta Lake, Cloud DWHs gain significant traction
- 2020 Modern Data Stack²

¹ http://cs.brown.edu/~ugur/fits_all.pdf

² <https://www.moderndatastackconference.com/>

Market Share



Source Systems



Ingestion



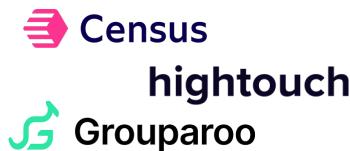
Storage & Processing



Visualization



Reverse ETL



Transformation



Machine Learning



Orchestration



Quality & Observability



Discovery & Governance



Privacy & Security



dbt – Data Build Tool

- Transformation Layer as Code
 - Models defined in SQL
 - Jinja templating for additional logic on top of SQL
- Version control with git
- Automated deployment
- Automated testing
- Automated generation of documentation and data lineage
- Analyst -> Analytics Engineer
- Easy to learn

Advantages of the Modern Data Stack

- Easy to spin up
- Cost efficient
- Analytics Engineering



Thank you!

bpirvu@novomatic.com

NOVOMATIC

References

- Modern Data Stack
 - <https://blog.getdbt.com/future-of-the-modern-data-stack/>
 - https://en.wikipedia.org/wiki/Draft:Modern_Data_Stack
 - <https://towardsdatascience.com/the-beginners-guide-to-the-modern-data-stack-d1c54bd1793e>
- dbt
 - <https://docs.getdbt.com/docs/introduction>
 - <https://docs.getdbt.com/docs/quickstarts/dbt-cloud/bigquery>
 - <https://docs.getdbt.com/reference/dbt-commands>
 - <https://docs.getdbt.com/guides/best-practices/how-we-structure/1-guide-overview>

Keep in Touch



Modern Data Stack
Group



linkedin.com/groups/9234718/



Vienna Data Science Tools
Meetup



meetup.com/vienna-data-science-tools



Personal



linkedin.com/in/bogdan-pirvu/