

# Big Data Tools: Fantastic Beasts and What Do They Do

Pasha Finkelshteyn,  for Big Data 

# **What is Big Data?**

## **Nobody knows**

# What is Big Data?

- Doesn't fit a single node
- Scales without rearchitecturing when grows (3V)
  - Volume
  - Velocity
  - Variety
- Enough data to make reliable business decisions

# Who handles this data? Data Engineers!

# Responsibilities: Pipelines

Transfer data source → sink

Fix leaks

Data engineer is pipeliner and  
plumbr™ in one





# Responsibilities: architecture

Architecture of storage

Architecture of pipelines

Tool selection

# Responsibilities: automation

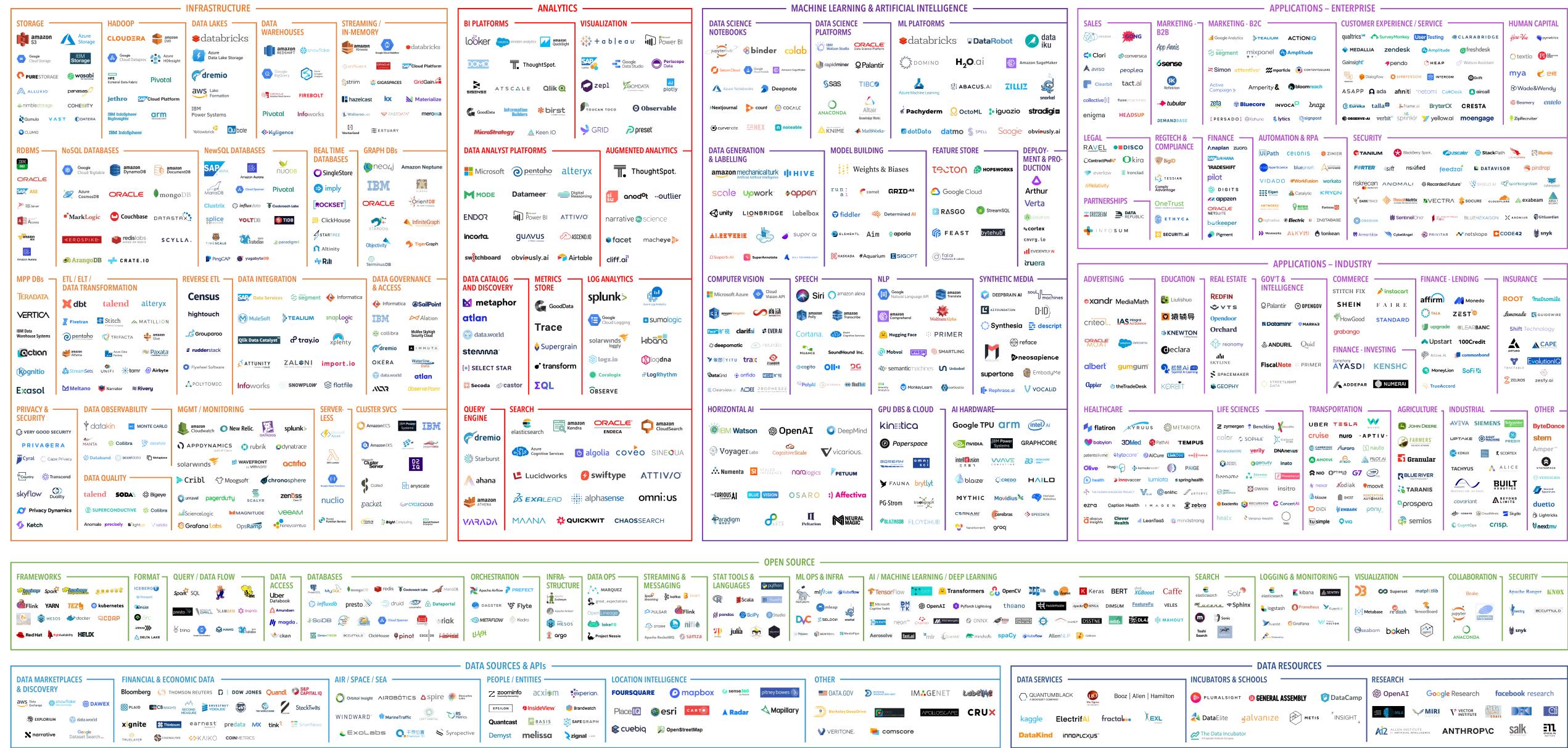
Work together with SRE, Ops,  
analysts, and DS

Help in automation of chore  
tasks

Orchestration



MACHINE LEARNING, ARTIFICIAL INTELLIGENCE, AND DATA (MAD) LANDSCAPE 2021



# What languages DEs use?

-  SQL
-  Python
-  Scala
-  Java
-  R

# What languages DEs use?

-  SQL
-  Python
-  Scala
-  Java
-  R

 folks, we cover all major languages DEs use!

# What languages DEs use?

But also... NoCode/LowCode tools!



And many more

# NoCode is so popular that...

At  SmartData\* conference we had a separate talk on hardcore vs "casual" DEs

<small>\* I'm humble member of PC</small>



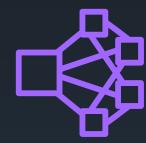
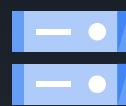
# What do fantastic beasts from BDT team do?



# What do fantastic beasts from BDT team do?

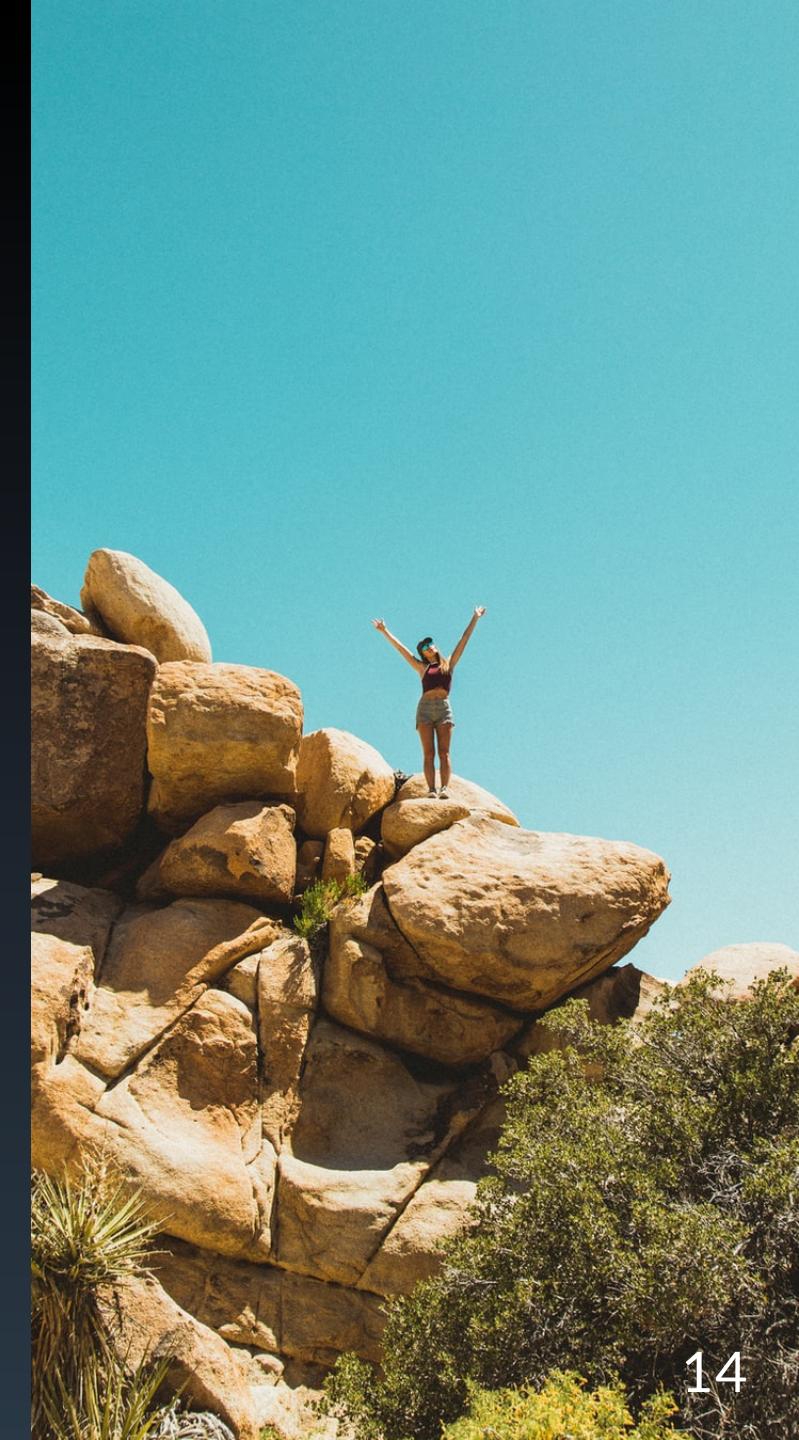
We integrate all the tools we can inside our plugin.

Currently they are:



and more

Why these ? We believe they're widely used!



# Supported functionalities

- Remote (and Local) file systems
- Binary formats view
- Notebook support
- Monitoring support

# Remote (and Local) file systems

- Copy-paste
- Rename
- Drag&Drop
- Metainfo show

# Binary formats view

- avro
- orc
- parquet

# Notebook support

- Zeppelin
- DataBricks (never released)

Why do we need notebook support?

- Autocompletion
- Refactoring

# Monitoring support

- Spark
- Hadoop
- Kafka

# Kafka

Kafka is WIP, will be released soon.

We've added **producers** and **consumers**