

From Zero to Hero

How to Start and Get your Data Ready for AI

Dr. Thomas Petrik – Sphinx IT Consulting

Dirk Beerbohm – Exasol



Exasol in a minute

The **Worlds Fastest** Analytical Database

for secure, time critical analytics supporting data services/products

- More than 20 years of experience in the field
- 200+ employees
- Public company since 2020
- In-memory, columnar analytics
- Low TCO
- Deploy anywhere
- ML at scale



200+
PROVEN SUCCESS Customers

100+
PARTNERSHIPS
Technology or consultancy partners

30+ COUNTRIES
Worldwide presence: North America, EMEA and Asia



Mercedes-Benz



T-Mobile



MONSOON

About Sphinx

Joy, Performance & Passion for Perfection

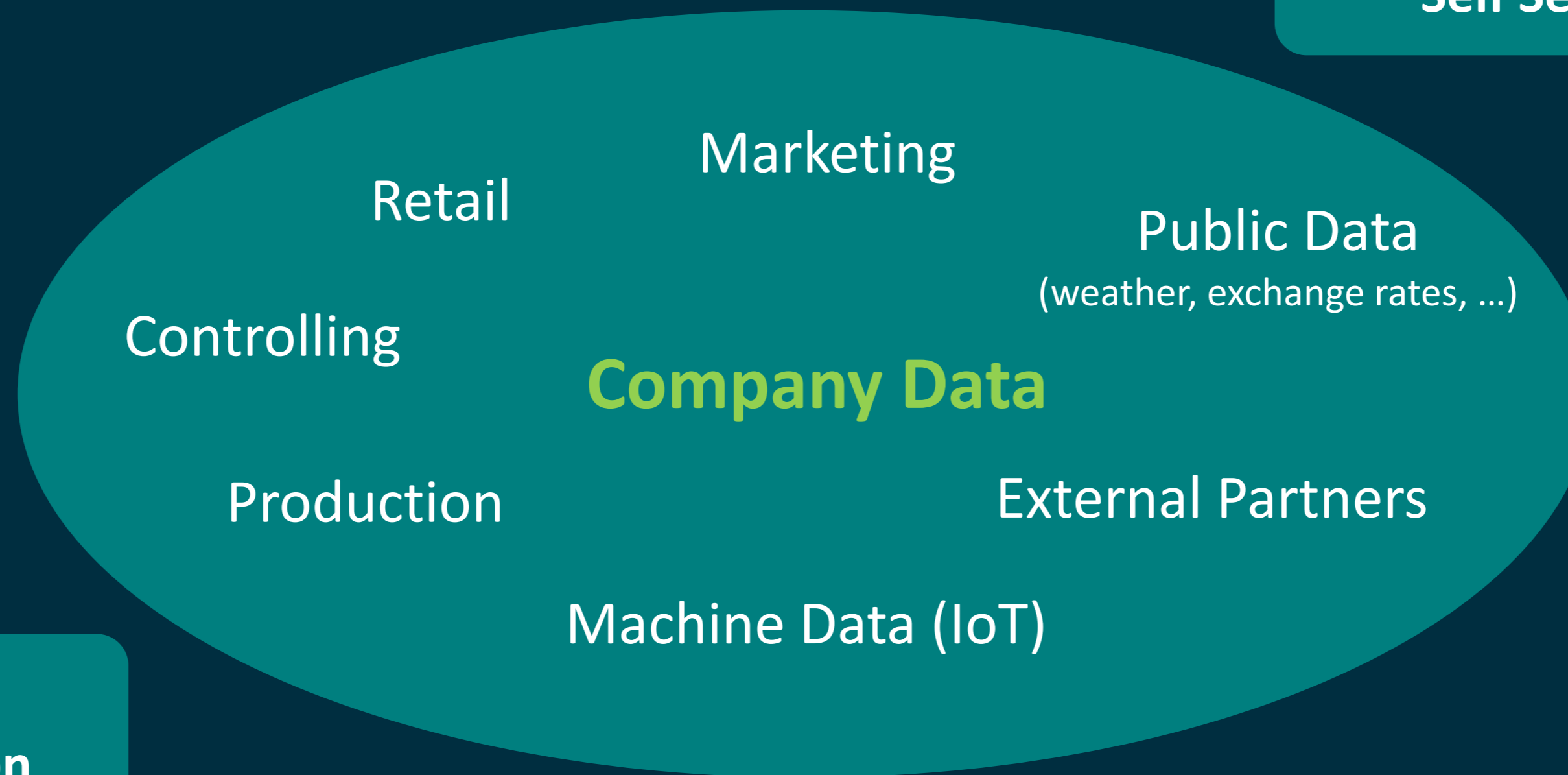
- **Founded:**
1994 Sphinx IT-Consulting GmbH
2003 Sphinx Managed Services GmbH
owner-managed
- **Head Office:** Vienna
- **Our Philosophy:** Holistic consulting, customized solutions, appreciation of existing systems & for the people behind them.
- **Business Areas:**
 - Analytics & Business Intelligence
 - Infrastructure & Database Technology
 - IT- & Data Security
 - IT- & Process Automation
 - AI in Business Solutions
 - Strategic Consulting & External IT



AI Hot Spots in a BI-Infrastructure



**Data Science & AI
Modelling & Prediction**



**Company Dashboards
Self Service BI & AI**



Self Service BI

Goals

- Business Users are able to build a meaningful report via "drag & drop" on a user friendly interface.
- Although the user is free to combine all the attributes and metrics presented **there is no risk of getting wrong results** (although the result may be empty in case that attributes do not match).

Prerequisites

- All attributes must be consistent at any point in time. The status of all data must be tracked continuously.
- Users must be presented only data they allowed to see (Need to Know Principle).
- Results must be reproducible for a specific point in time.
- Performance is a key factor: Response times must not be higher than 1-2 seconds.
- Data must be actual: Omit any caching or additional ETL-processes.

Data Science

Goals

- Define a use case – find the data - train a model - predict
- Data scientists are able to operate on the full width and full history of data.
- Data scientists have the freedom to build and train models in an iterative process with production data in a safe playground.
- "Model to data" instead of "data to model".
Enable high speed predictions.

Prerequisites

- All attributes must be consistent at any point in time. The status of all data must be tracked continuously.
- Users must be presented only data they allowed to see (Need to Know Principle).
- A complete history of data is a key factor for modelling.
- Real time predictions need extreme performance.
- Data scientists must be provided with a proper tool set in-database.
- Sandboxes must be provided as safe playgrounds.

Attributes of a Solid Data Foundation

Correctness

Data Quality

Completeness

Consistency

*Reporting
Consistency*

Load Status

Integrity

*consistent, fine
grained Security*

*Need-to-Know
&
Least Privilege*

*strict bi-temporal
Historisation*

*reproducible
Reporting*

Availability

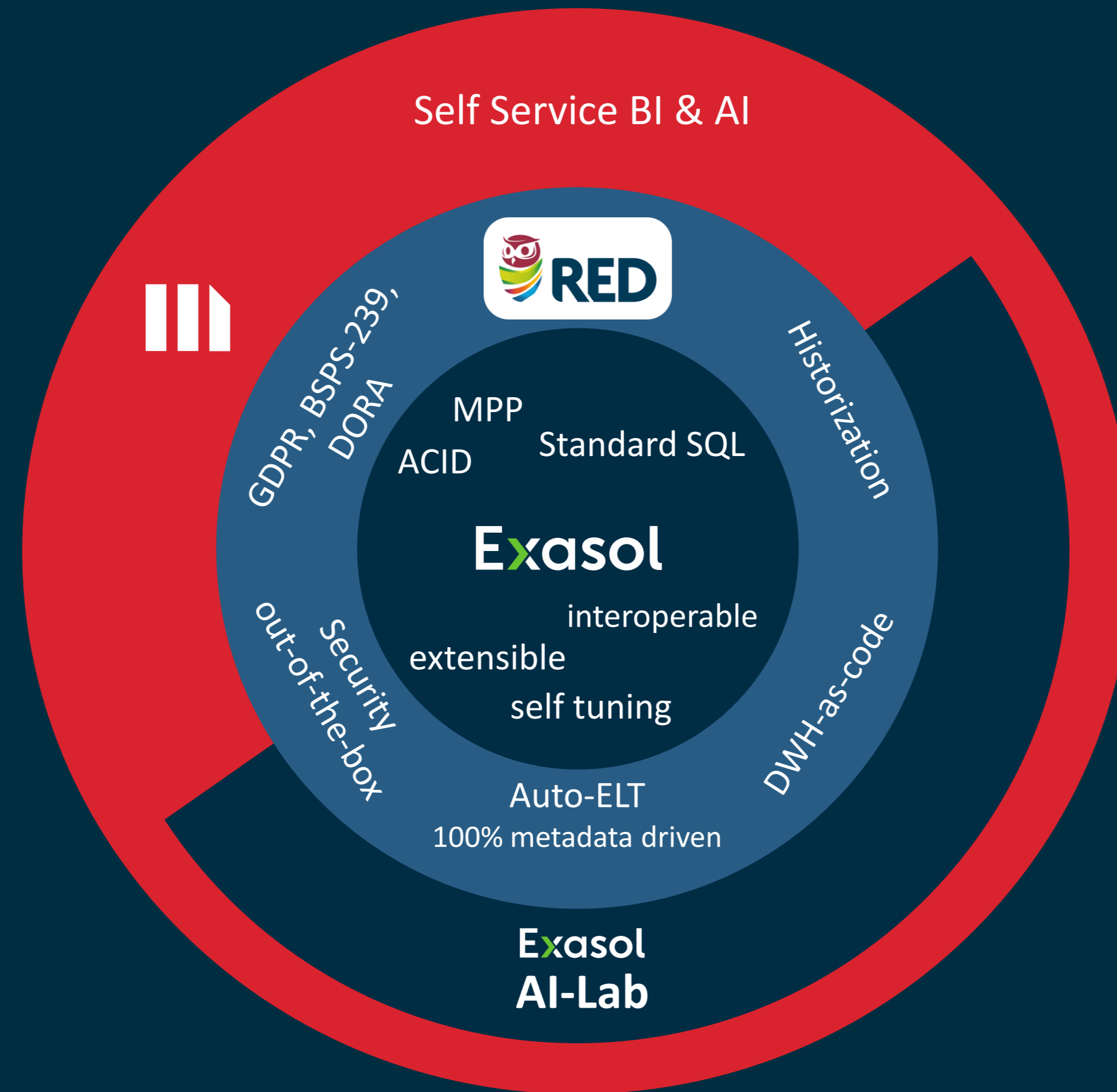
*Architecture (HA & DR)
&
Operations (SLA)*

Performance

*Delay
Source - DWH*

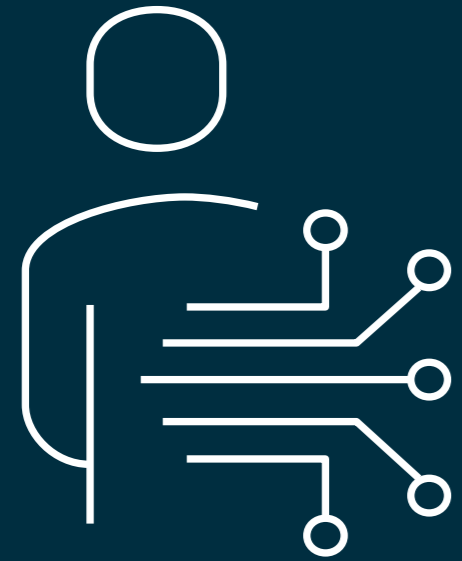
Fully Integrated Solution Stack

Exasol – Sphinx - MicroStrategy



Data Science: Advanced In-database Analytics

POWERFUL UDF FRAMEWORK



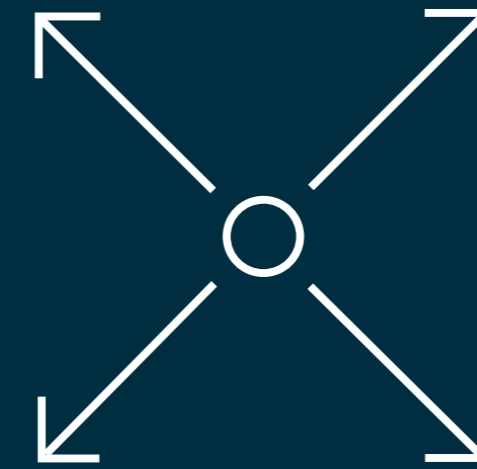
- SQL + Python ...
- Bring algorithms to the data

YOUR ENVIRONMENT



- Any language
- Any library
- Any version

SCALABILITY



- Built-In parallelization
- Support for big datasets

INTEGRATION



- Use your common tools
- Easy model deployment with BucketFS

Live Demo AI-Lab

Thank you!

