PDF SOLUTIONS™

2023 PDF Users Conference:

AI for tomorrow's manufacturing and R&D

Santa Clara Marriott - 2700 Mission College Boulevard Santa Clara, California 95054 USA

Operationalizing and Scaling AI/ML with ModelOps

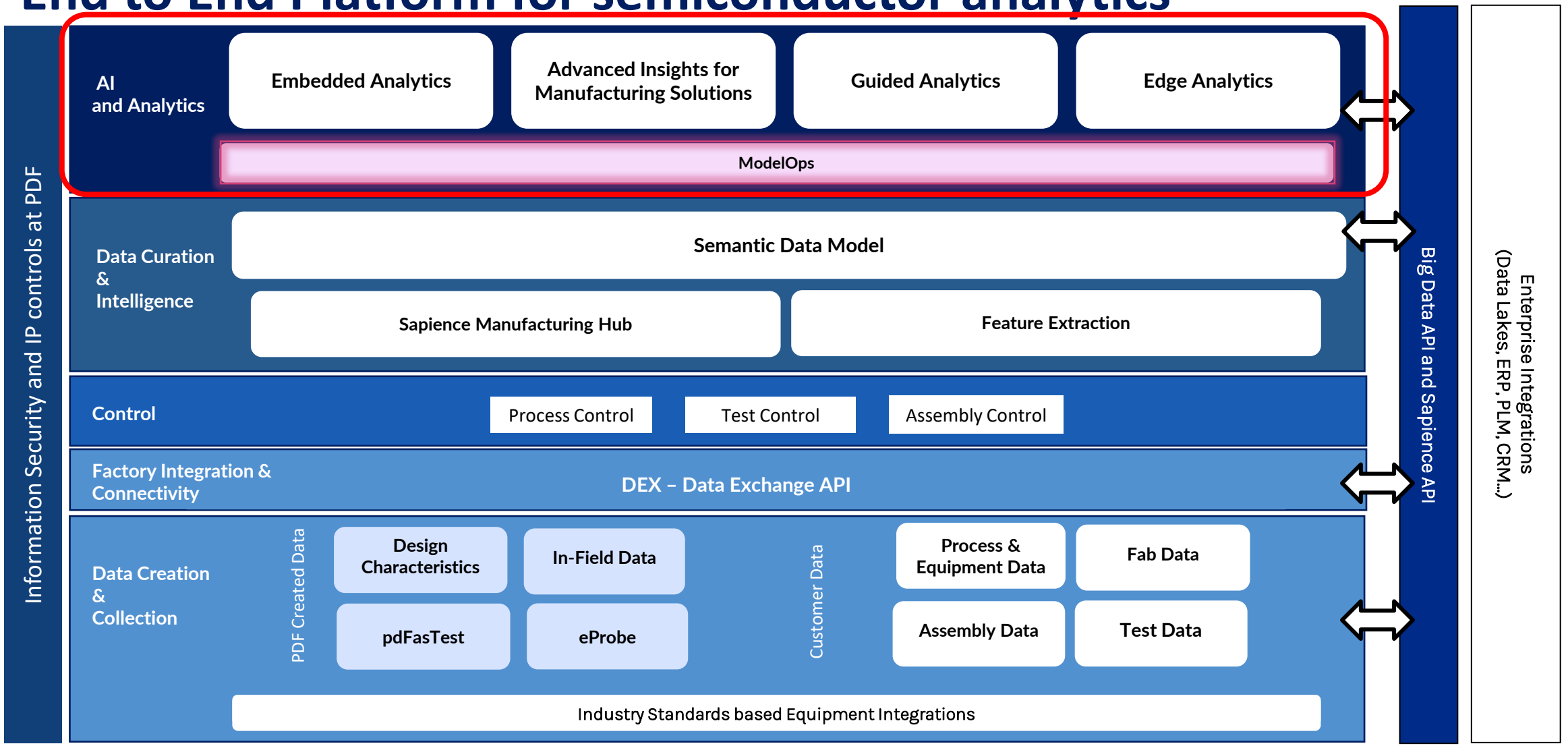
25th October 2023

This presentation and discussions resulting from it may include future product features or fixes, or the expected timing of future releases. This information is intended only to highlight areas of possible future development and current prioritizations. Nothing in this presentation or the discussions stemming from it are a commitment to any future release, new product features or fixes, or the timing of any releases. Actual future releases may or may not include these product features or fixes, and changes to any roadmap or timeline are at the sole discretion of PDF Solutions, Inc. and may be made without any requirement for updating. For information on current prioritizations and intended future features or fixes, contact sales@pdf.com.

PDF Solutions, Exensio, CV, Cimetrix, the PDF Solutions logo, and the Cimetrix logo are registered trademarks of PDF Solutions, Inc. or its subsidiaries. All other trademarks cited in this document are the property of their respective owners.

© 2023 PDF Solutions, Inc. or its subsidiaries. All rights reserved.

End to End Platform for semiconductor analytics



Fully integrated solution to accelerate production ramp, improve overall yield and quality for Semiconductors

© 2023 PDF Solutions, Inc.

Proof of Concepts (PoC's) in AI/ML are becoming easier, and are not by themselves valuable

- Lots of people in the semi industry can do PoC's now
- *More rare* -> productionalizing the AI/ML PoC for just one chip product
- *Very rare* -> a generalized platform that can bring AI/ML to production across multiple chip products and use cases in a heavily automated way

PoC's in AI are difficult to scale for semiconductor use cases

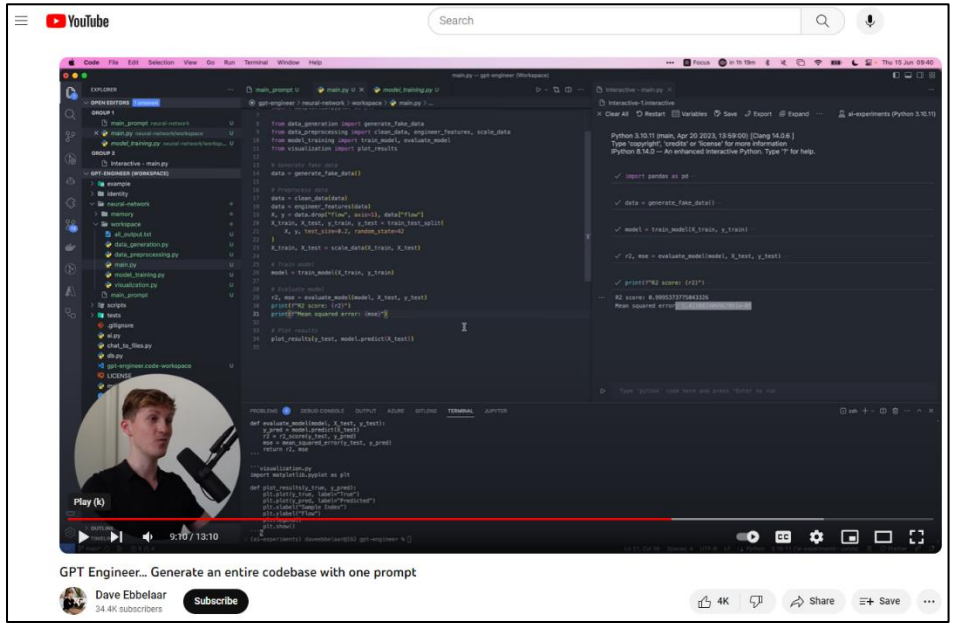
PDF/SOLUTIONS™

© 2023 PDF Solutions, Inc.

Examples of PoC's getting easier (and changing)..

- **"GTP Engineer" is a new offering, here is a nice demo/overview with a relevant use case. Interesting sections 4:00-7:00.**

Prompt: Create code to train and test a neural network using a typical machine learning pipeline. Use a fake dataset (regression), process it with the pipeline, split it into train/test and evaluate the performance by creating line plots. Use sklearn and matplotlib.



https://www.youtube.com/watch?v=gWy-pJ2ofEM

"AI won't take your job, it's somebody using AI that will take your job."

- Richard Baldwin, economist and professor at the Geneva Graduate Institute, Switzerland

… but PoC's by themselves don't generate value

© 2023 PDF Solutions, Inc.

Enabling you to realize value with AI

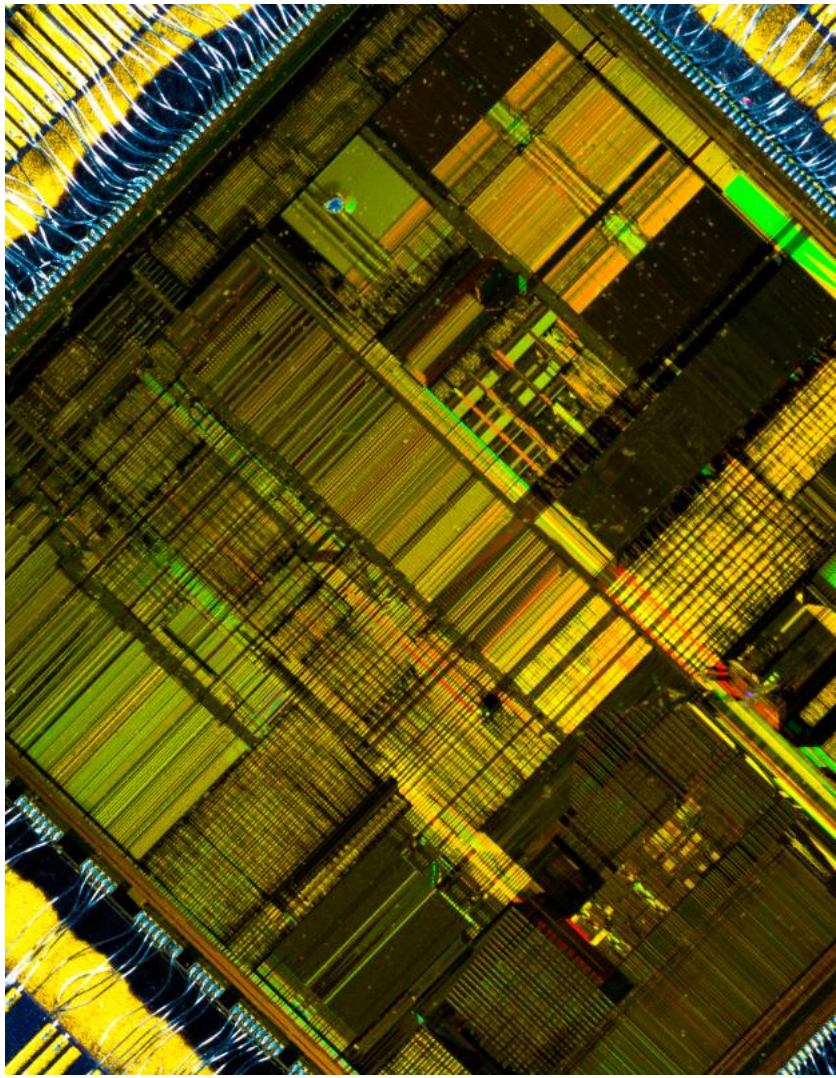
- **Single platform for all data in Exensio**
- **Infrastructure for semiconductor-specific data**
- **The same data science platform for all semiconductor data types – this saves data scientists' time and effort**
  - Bring your own algorithm/model, predefined solutions

PDF provides a platform and solutions that brings ***better*** models to chip production ***faster***

© 2023 PDF Solutions, Inc.

Pain Points for model training in the Semiconductor Industry

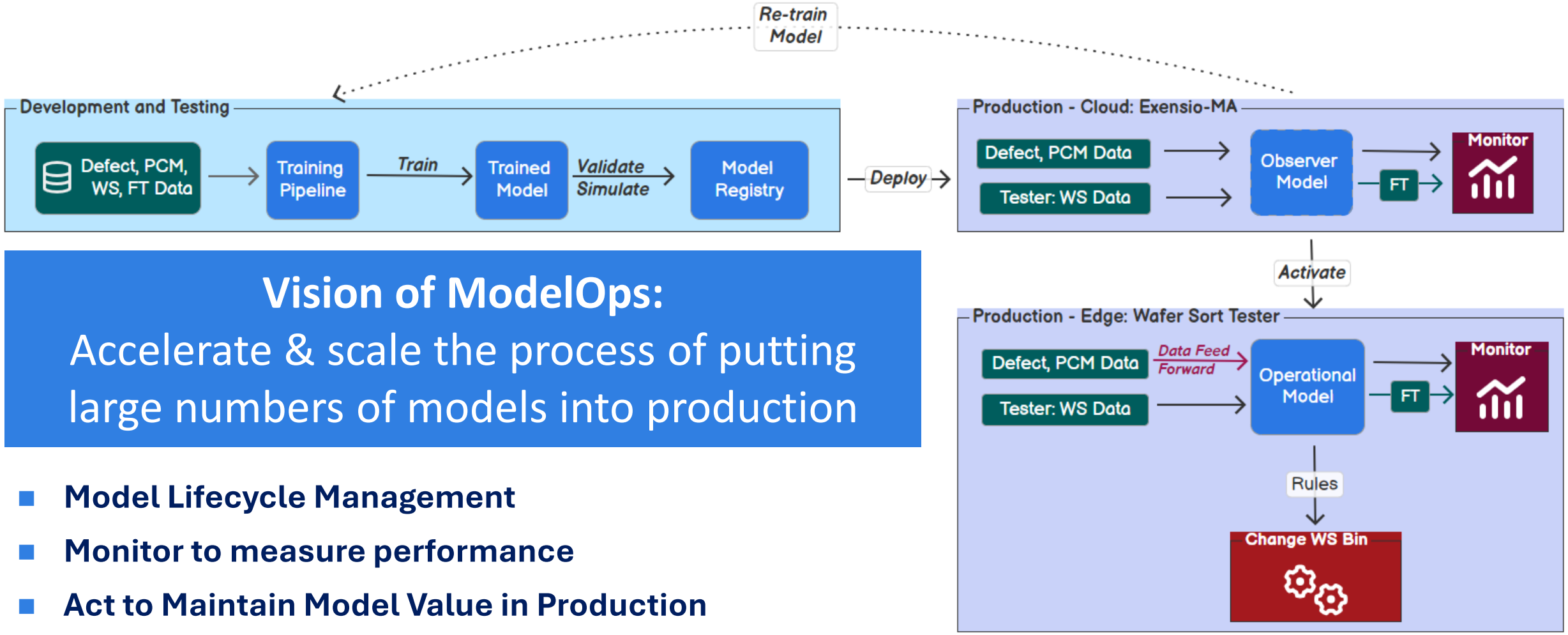
- Semiconductor data makes it difficult to apply ML pipelines designed for other industries
- Models need to be deployable to the edge
- Upstream data preprocessing and staging
- Agnostic to model development and deployment environments (e.g. AWS, Azure, GCP, etc..) and enable BYOA/BYOM



© 2023 PDF Solutions, Inc.

System Capabilities: Create, manage & control models across their lifecycles

Example: Predict Post-Burn-in Test results at Wafer Sort from Defect, PCM & Wafer Sort data

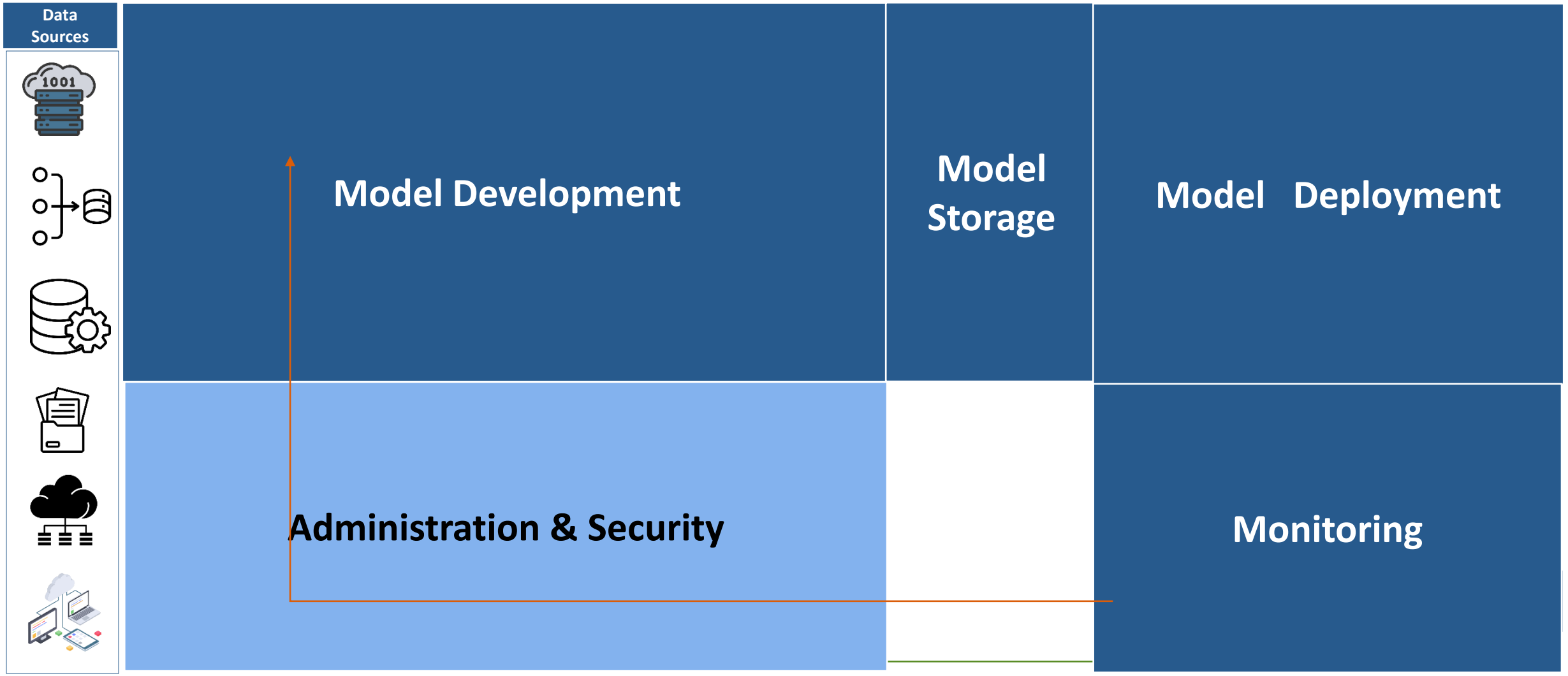


Vision of ModelOps:
Accelerate & scale the process of putting large numbers of models into production

- **Model Lifecycle Management**
- **Monitor to measure performance**
- **Act to Maintain Model Value in Production**

© 2023 PDF Solutions, Inc.

ModelOps Lifecycle with PDF capabilities & Extension Points



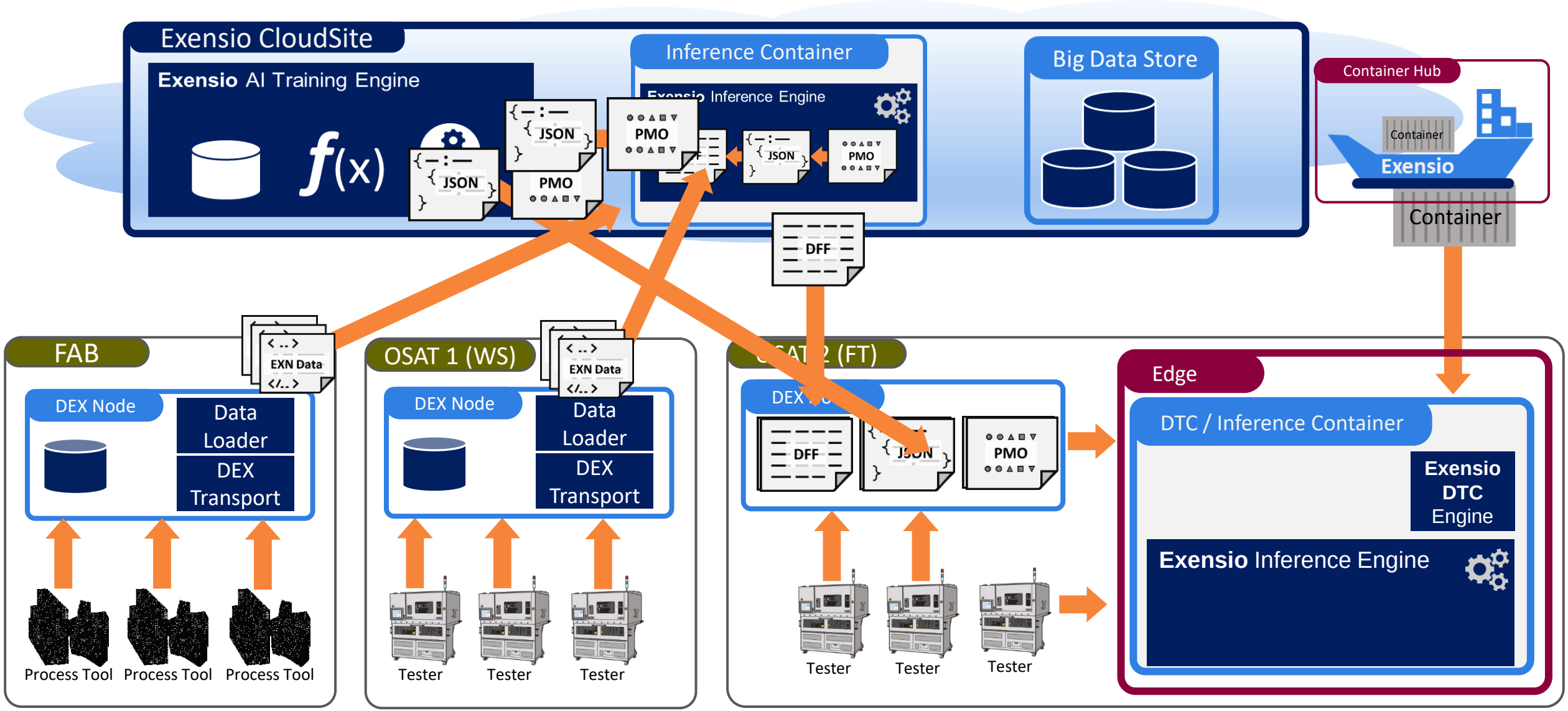
PDF ModelOps capabilities

Extension Points to Bring-your-own Algo / Model / Inference Engine

PDF/SOLUTIONS™

© 2023 PDF Solutions, Inc.

Exensio ML: Test Example



© 2023 PDF Solutions, Inc.

ModelOps Platform Capabilities

Build

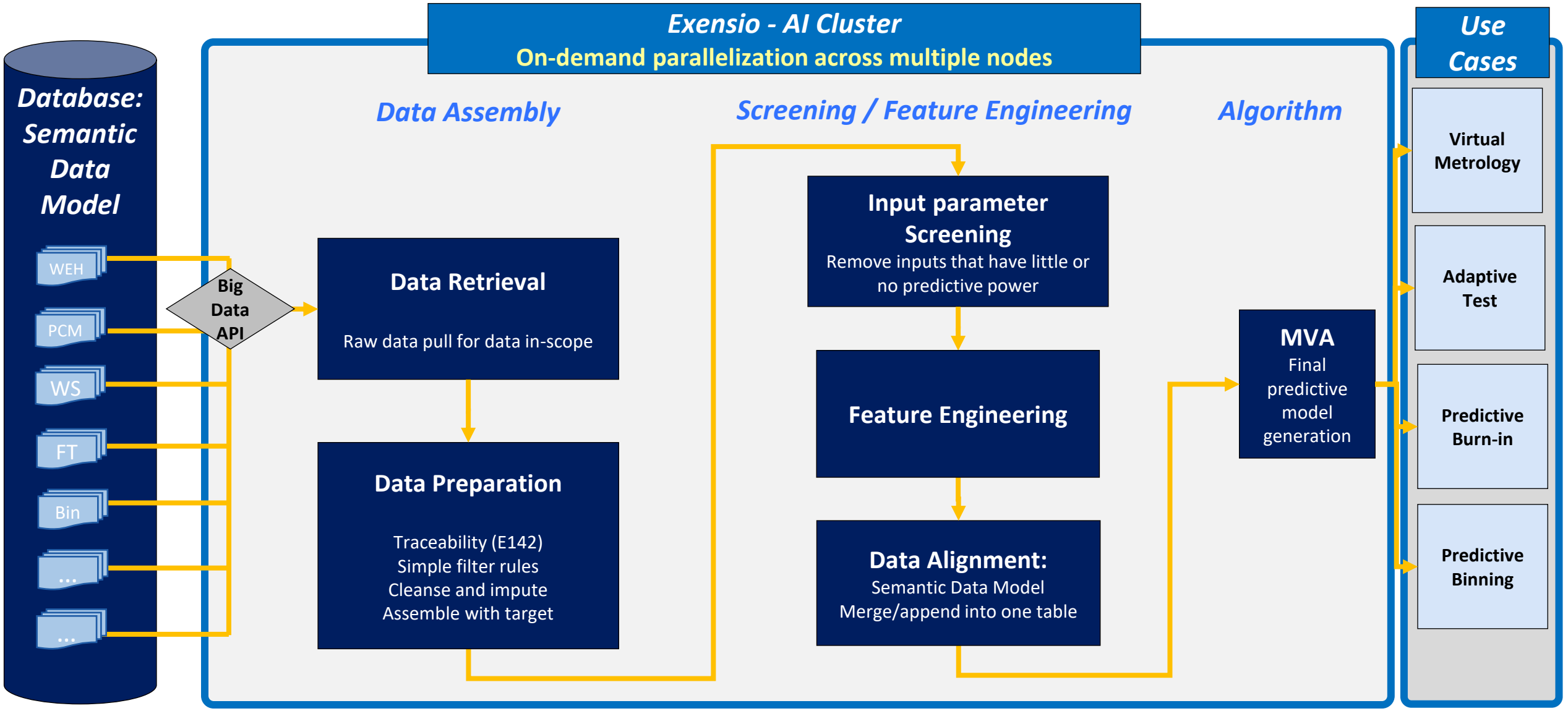
- Dynamically scaled compute
- Leverages any Exensio data
- Modularized pipeline
- Configurable data types / products
- Models saved for deployment
- BYOA - python
  - Most technologies supported
- Integrates with 3rd party platforms

Deploy

- Model routing
- Automated upstream data processing
- Production Monitoring
- Performance skew handling:
  - Re-train/replace models
- Model validation
- Automated safety nets
- BYOM

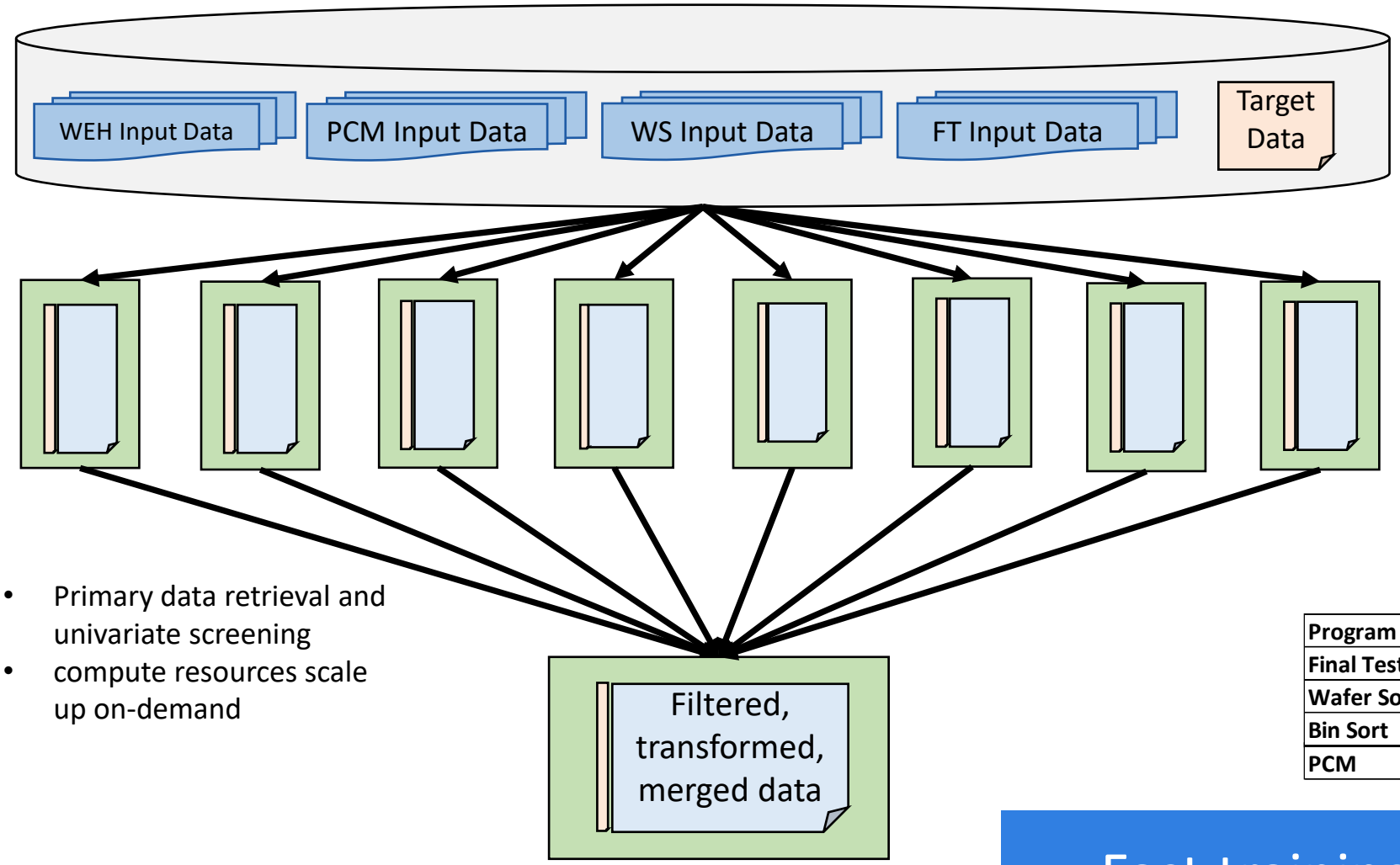
© 2023 PDF Solutions, Inc.

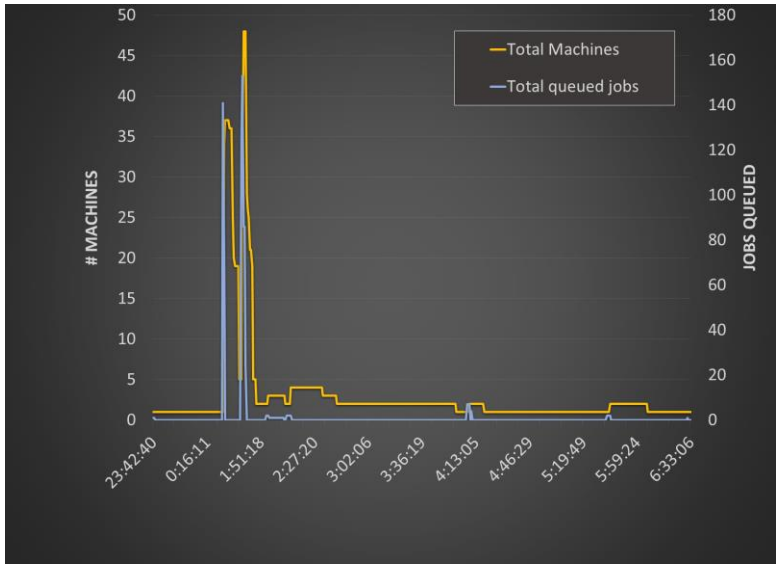
Dynamically-Scalable Training Pipeline



© 2023 PDF Solutions, Inc.

Dynamically-Scalable Training Pipeline





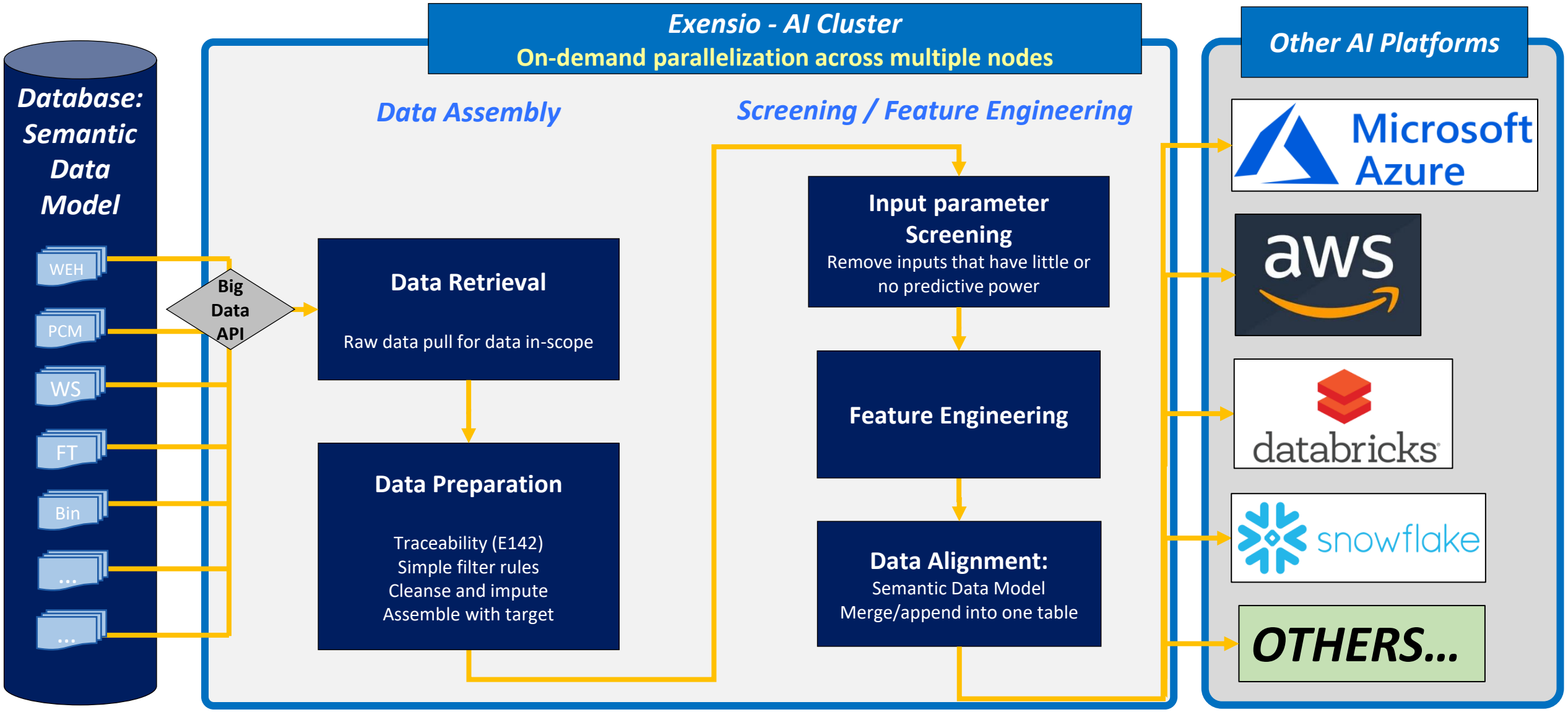
- Primary data retrieval and univariate screening
- compute resources scale up on-demand

Program Class	Program	Columns	Rows
Final Test	1	2,039	11,168,651
Wafer Sort	1	2,039	12,134,526
Bin Sort	1	(Thin)10	12,136,960
PCM	1	312	671,832

Fast training and resource efficient

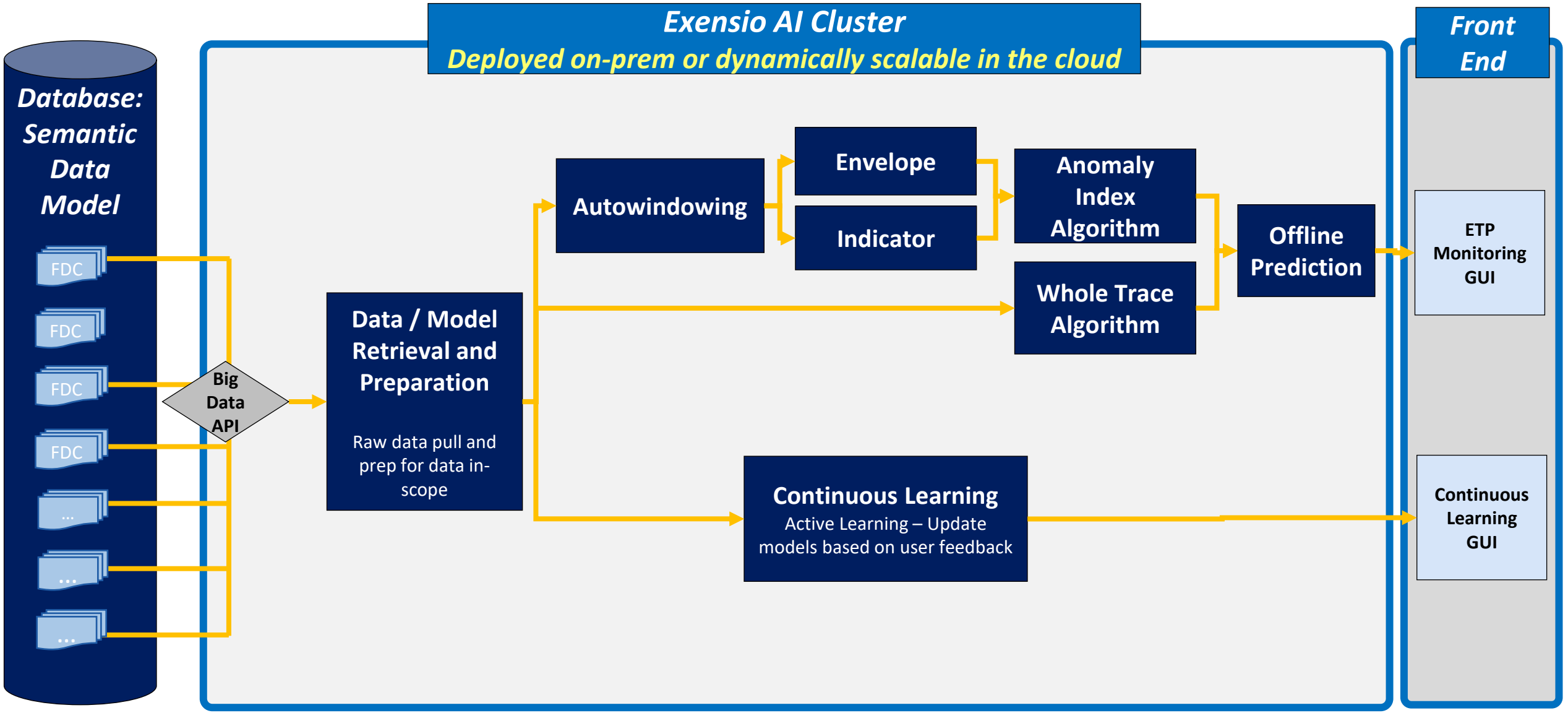
© 2023 PDF Solutions, Inc.

Designed to easily integrate with other platforms



© 2023 PDF Solutions, Inc.

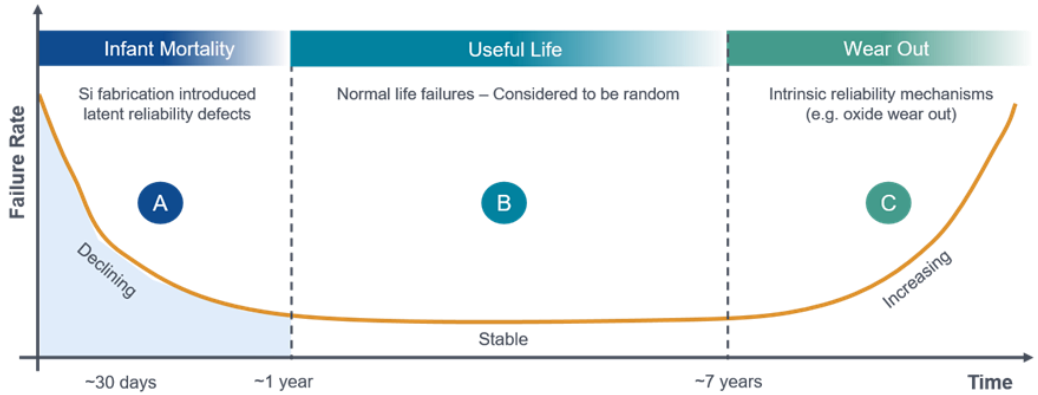
Training Pipeline for Equipment Trouble Protection (ETP)



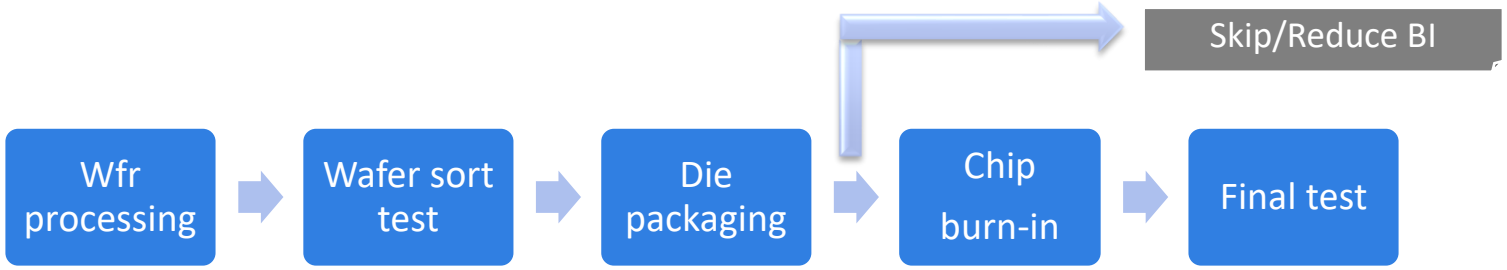
© 2023 PDF Solutions, Inc.

Exensio AIM Predictive Burn-In

- **Product burn-in**
  - High temperature/voltage stress test to screen out weak die (section A of the figure)
  - Done on packaged dies (wafer level burn-in is rare)



https://semiengineering.com/the-importance-of-product-burn-in-test/



- **AIM Predictive Burn-In**
  - ML solution to build models to predict burn-in fails
  - **Save package and burn-in test costs**

Chip Cost	$45	
Die Size	107.7	mm^2
Tech Node	**5nm**	
Wafer Cost ASP (USD)* 2021 ASP	$17,000.00	
Gross Die Per Wafer	553	
Net Die Per Wafer (Assume 95% Yield)	530	
Die Cost	$32.08	71%
Assembly / Package Test Cost	$12.92	29%

Illustrative die assembly and package test cost

https://semiengineering.com/using-machine-learning-to-increase-yield-and-lower-packaging-costs/

© 2023 PDF Solutions, Inc.

Solution Example for Exensio Test : Predictive Burn-in

- **Use Case Results (escapes < 2 dppm):**

BI Failure Bin A Prediction	BI Failure Bin B Prediction	Combined Prediction	Die Population	ACTION
PASS	PASS	BOTH_PASS	< 35%	Completely Skip BI
FAIL	PASS	FAIL_BIN_A_ONLY	< 25%	Partial BI Test
PASS	FAIL	FAIL_BIN_B_ONLY	< 20%	Partial BI Test
FAIL	FAIL	BOTH_FAIL	< 30%	Complete BI

Burn-in cost savings realized from a combination of completely skipping burn-in and reduced testing at burn-in

© 2023 PDF Solutions, Inc.

In summary…

- **Single platform for all data in Exensio**
- **Infrastructure for semiconductor-specific data**
- **The same data science platform for all semiconductor data types – this saves data scientists' time and effort**

With PDF ModelOps, you can bring
the value of AI to chip production *faster*

© 2023 PDF Solutions, Inc.

Thank You

PDF/SOLUTIONS™

pdf-solutions

pdfsolutionsinc

pdfs.inc

pdf_solutions

pdfs_cn

PDF/SOLUTIONS™