

MLOps, meet OPI

Debo Dutta - VP Engineering

About me

Current: VP Engineering, Nutanix (multi cloud platform co)

AI4Infra (P&L engg) & Infra4AI (opensource)

Founding member, MLCommons.org

MLOps/Kubeflow.org (c2017)

Openstack (c2012)

Past:

Visiting Scholar, Stanford

Distinguished Engg.@Cisco

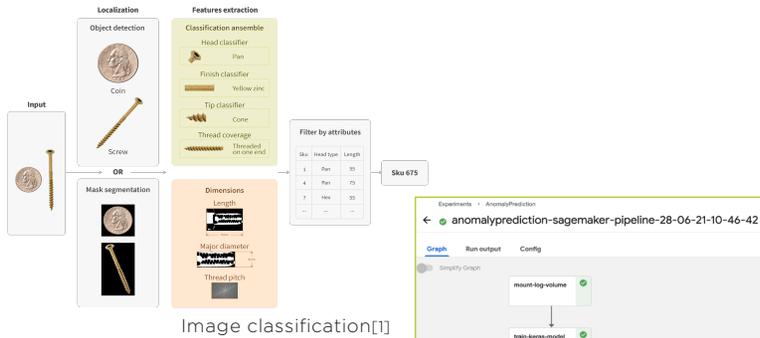
Researcher in Computational Biology

PhD CS, USC/ISI (NS2 group)

Btech, IIT Kharagpur

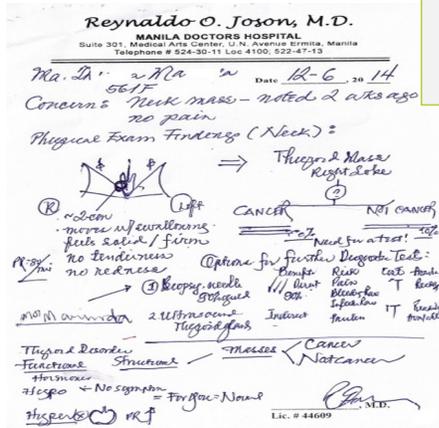


AI fueling digital transformation



- Data -> AI -> Insights -> Actions

- Enterprise data explosion - databases, conversations, logs, metrics, knowledge graphs
- Digital transformation: use enterprise data to increase operational efficiency
- AI accelerates digital transformation by providing learned insights which leads to better actions.



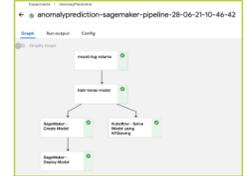
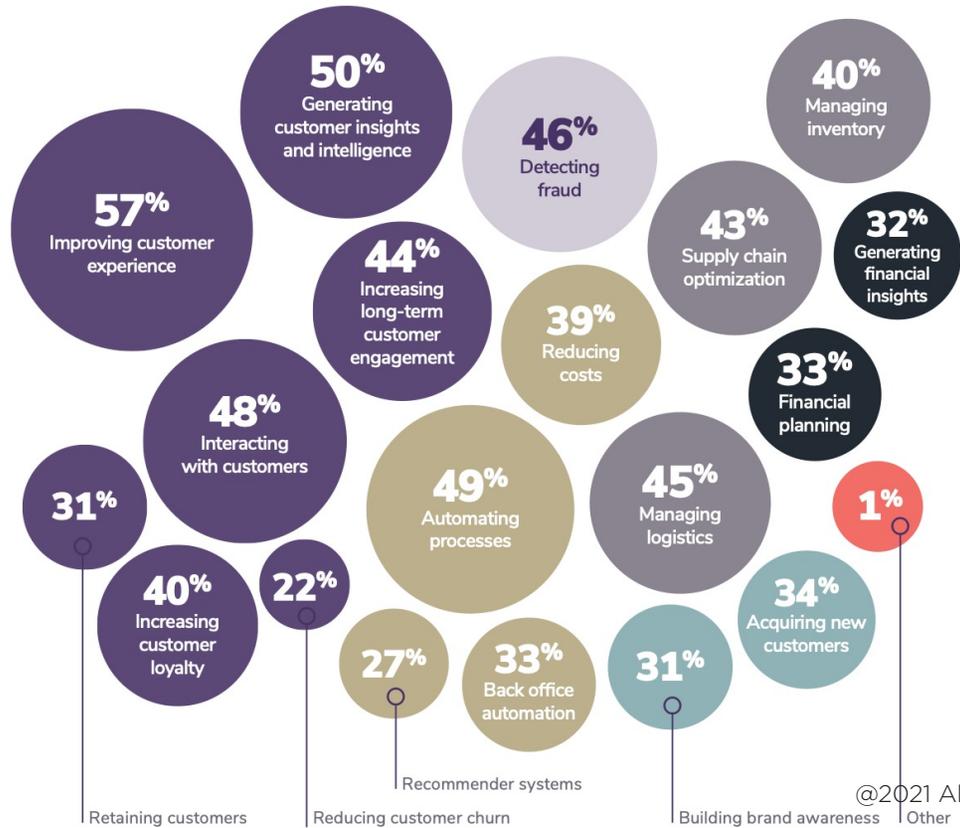
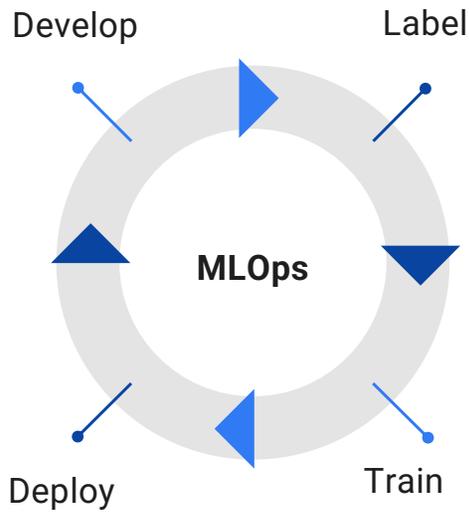
- Key challenge: Accelerating traditional businesses with MLOps

- Finance, retail, pharma, healthcare, public sector
- Manage data and ML

[1] <https://blog.griddynamics.com/identifying-screws-a-practical-use-case-study-for-visual-search/>
 [2] <https://link.springer.com/article/10.1007/s11042-020-10151-w>



Enterprise AI/ML use cases (similar infra needs)

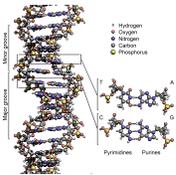


@2021 Algorithmia AI survey

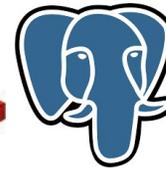
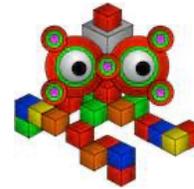
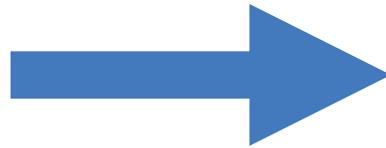
- Customer experience, engagement, loyalty
- Detecting fraud
- Finance
- Other
- Automating processes and reducing costs
- Supply chain and logistics
- Customer acquisition



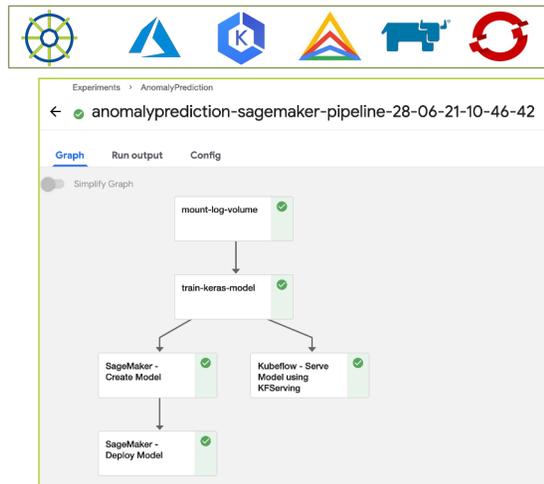
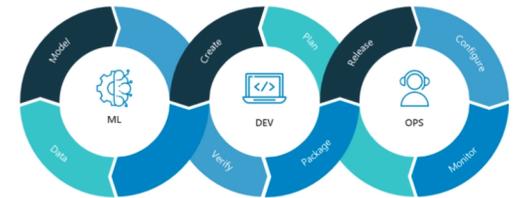
Where is the “valuable” enterprise data?



Source: images.google.com



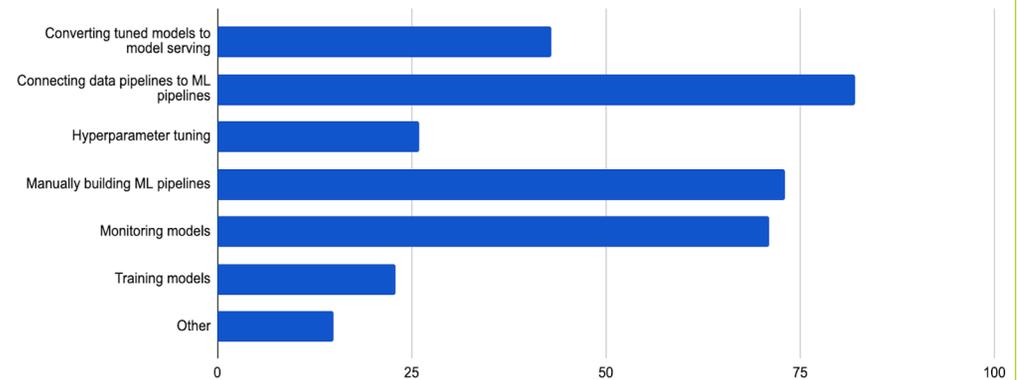
OPI Opportunity: MLOps



End to end Workflow Orchestration

- Repeatable
- Composable
- Re-usable

Where do your teams encounter gaps in your ML activities & workflows?



Survey: Kubeflow PM Team



argo

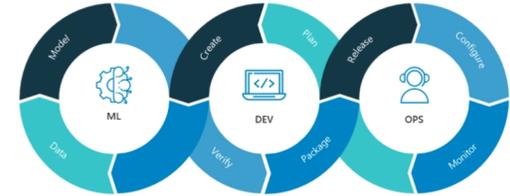


Kubeflow

mlflow™



MLOps Market Opportunity

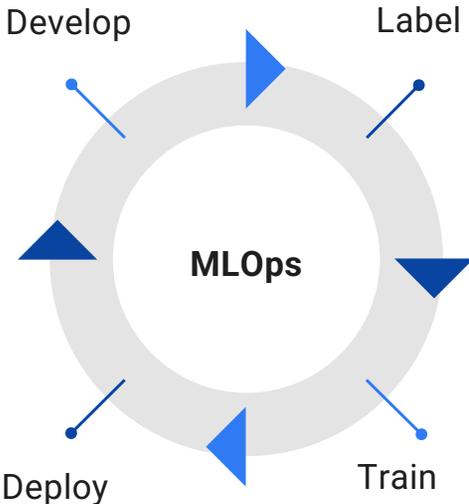


*According to Cognilytica, the global MLOps market will be worth **\$4 billion** by 2025.*

*The industry was worth **\$350 million** in 2019.*

Source: <https://askwonder.com/research/historical-global-ai-ml-ops-spend-ulshsljuf>

MLOps ecosystem: confusion



ML / AI Infrastructure

DATA PREPARATION	MODEL BUILDING	PRODUCTION
Data Exploration & Processing Paxata, TRIFACTA, alteryx, databricks, Superb AI, iguozio, Quobale, HIVE, SAS	Hosted Notebooks Management Amazon SageMaker, Azure Machine Learning, databricks, CLOUDERA, CO, DDMING, Deepnote	Model Observability arize, DataRobot, Other startups
Data Version Control Pachyderm, data iku, DVC	Model Management, Version Tracking and Storage databricks, Amazon SageMaker, Google AI, Azure Machine Learning, iguozio, ALGORITHMIA, DDMING, data iku, SAS	Model Compliance & Audit fiddler, SAS, Other startups
Feature Engineering and Storage FEATURE LABS, Featuretools, TACTON, Google Feast	Experiment Tracking Weights & Biases, comet, TensorBoard, mlflow	Model Deployment and Serving Amazon SageMaker, Kubeflow, ALGORITHMIA, DataRobot, Google AI, Azure Machine Learning, PerceptiLabs, iguozio, SAS
Data Labeling scale, lb, Labelbox, figure eight, Amazon SageMaker, Google AI, Azure Machine Learning	Model Optimization Hyper Parameter SIGOPT, Weights & Biases, Amazon SageMaker, anyscale, comet	Model Validation arize, fiddler, SAS, Other startups
Data Quality Checks great_expectations	Auto ML H2O, DataRobot, Amazon SageMaker, Google AI, Azure Machine Learning, data iku	Platform Specific Model Builds OctoML
	Model Training Amazon SageMaker, Azure Machine Learning, Google AI, Kubeflow, anyscale, SAS, iguozio, PerceptiLabs	
	Model Evaluation TensorBoard, Streamlit	
	Model Explainability fiddler, TensorBoard	

Source: images.google.com

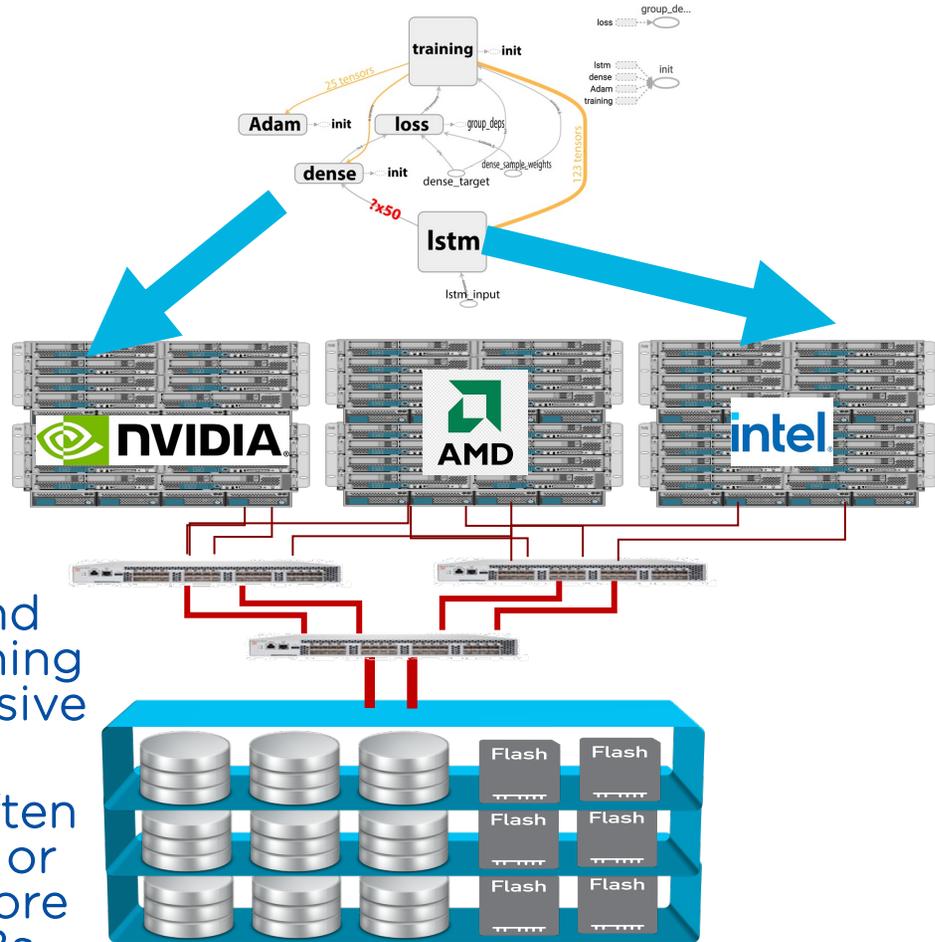




OPI Opportunity: Optimizing MLOps

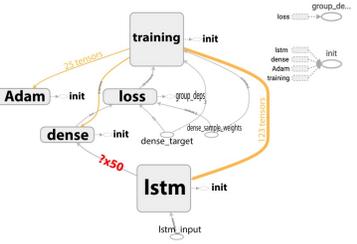
Current ML Infra is disaggregated

Training is inefficient due to communication complexity



ETL and Pre-training Is expensive

Data is often in a filer, or object store or in DBs



Ref: https://www.researchgate.net/figure/Example-of-TensorFlow-graph-some-parts-of-the-graph-are-enlarged-for-better-visibility_fig3_324362772



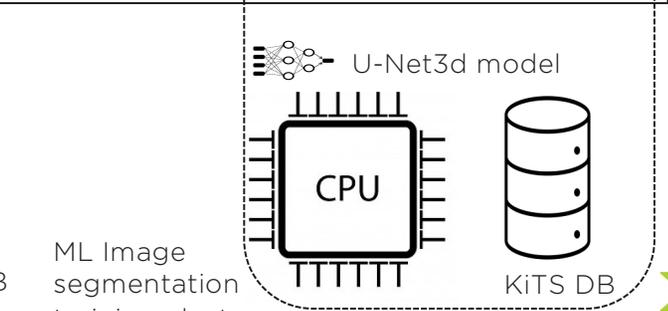
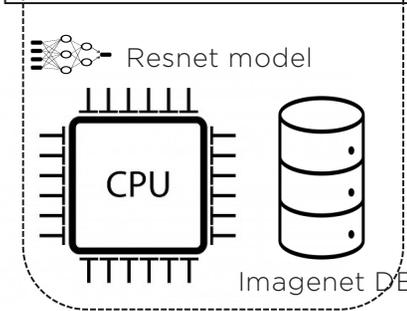
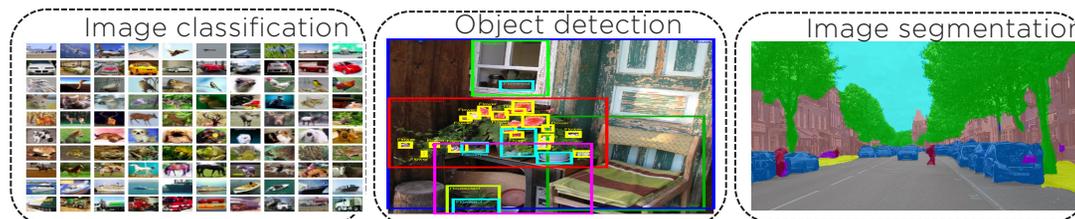
Potential OPI directions:

Reference Arch for ML With OPI

MLPerf workloads

MLCommons “cubes”

MLPerf Storage-ML



@johnugeorge, co-chair storage/ML

ML Image classification training cluster

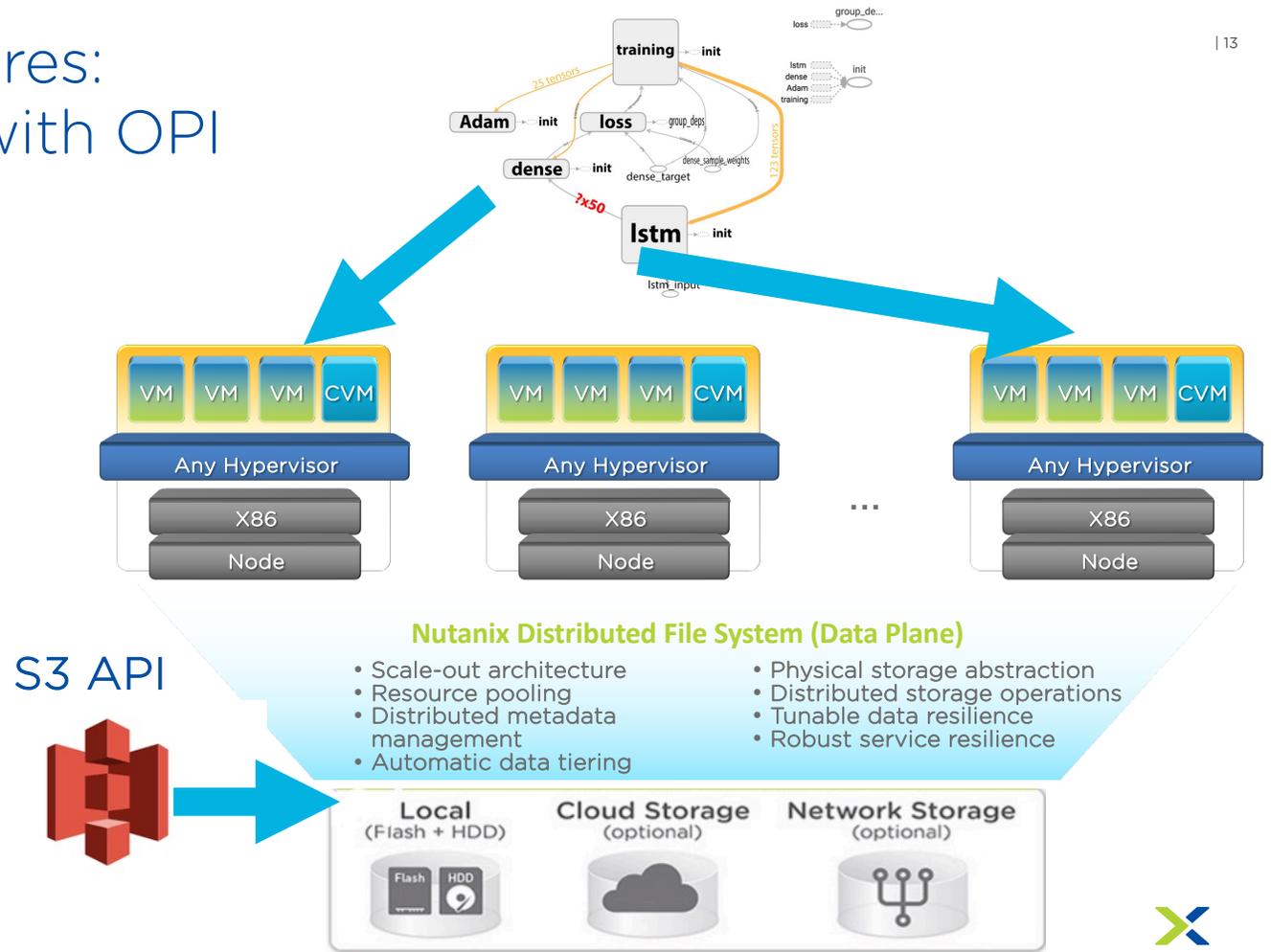
ML Image segmentation training cluster



Research-y Futures: Push down ML with OPI

Collaboration opportunities

- Distributed training
- Split learning
- Federated learning
- SPDK primitives for storage+ML
- Infra optimization - scheduling



The Nutanix logo is centered on the slide. It consists of the word "NUTANIX" in a bold, white, sans-serif font, followed by a small "TM" trademark symbol. The logo is set against a background of a blue-to-green gradient with large, semi-transparent, overlapping arrow shapes pointing in various directions.

NUTANIX™