

MATLAB EXPO

2021

Cloud Data Workflows for Scientists and Engineers: What You Should Know

Igor Alekseev, Prasad Kona, Arvind Hosagrahara



Agenda

	Time	Presenter
Introduction (why we are here)	3 minutes	Arvind H (SA)
AWS – getting engineering data to the cloud	6 minutes	Igor A (SA)
Databricks – helping everyone use the data in the cloud	6 minutes	Prasad K (SA)
MathWorks – connecting and operationalizing the data in the cloud	5 minutes	Arvind H (SA)
Q&A	10 minutes	

Improve Predictive Maintenance Performance using Transient Data

Challenges

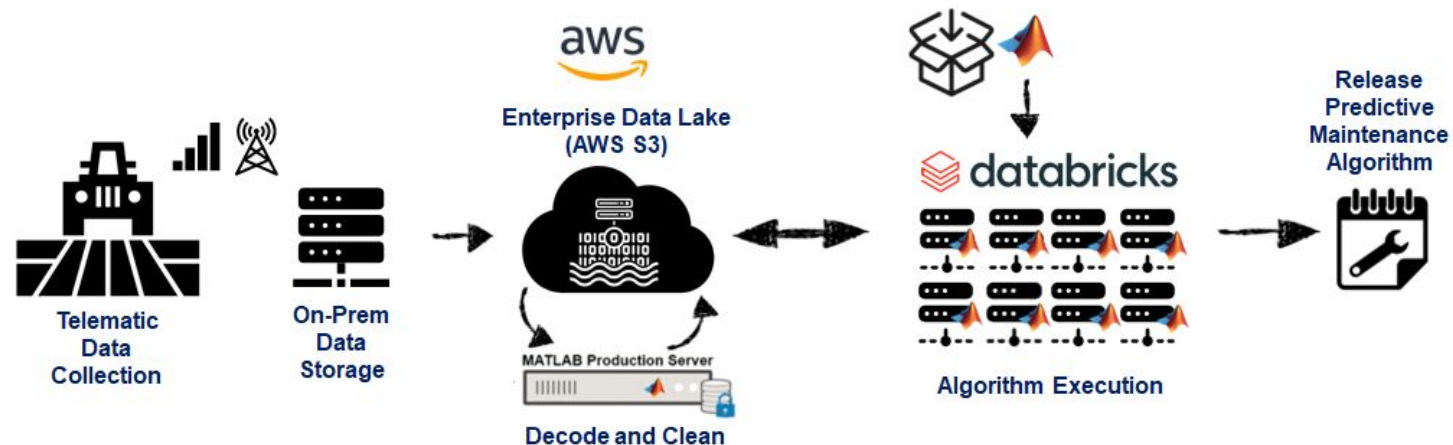
- Embedded controllers have limited computing power
- Engineers don't have easy access to data in Enterprise Data Lake

Solution

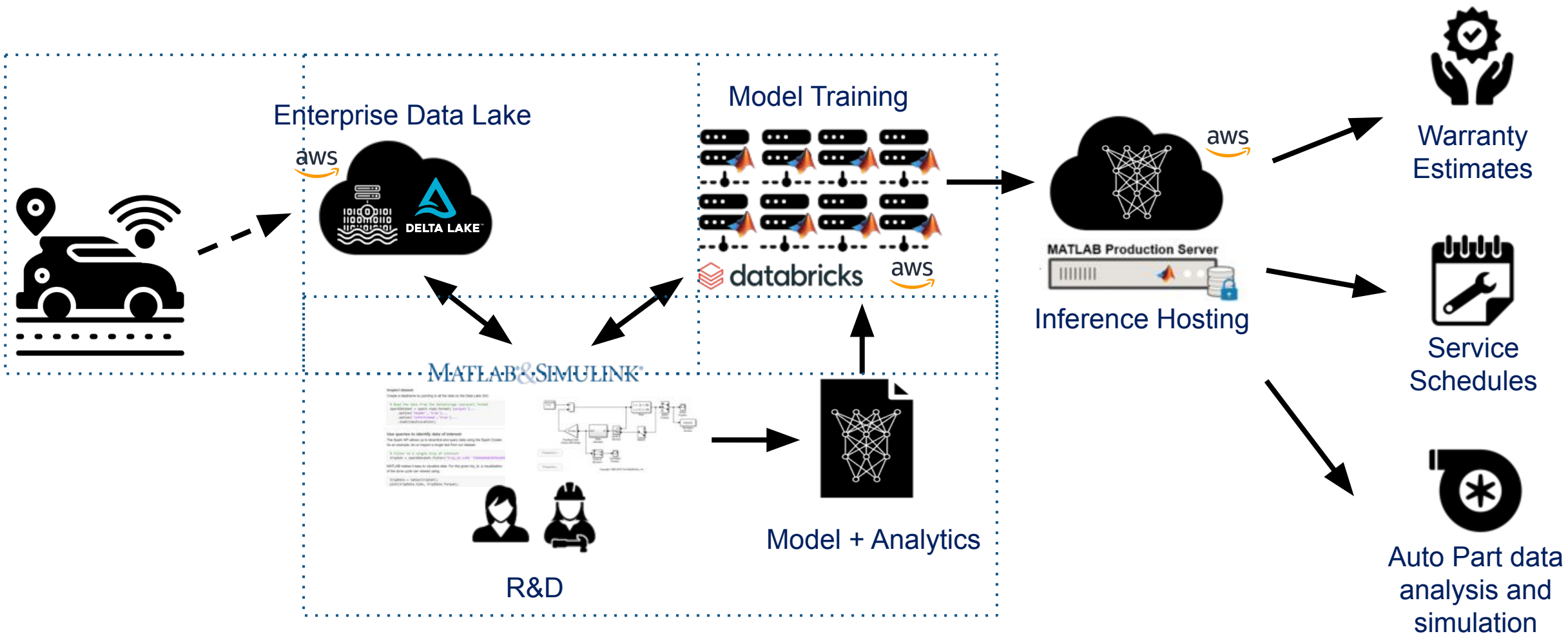
- Integrate MATLAB with cloud-based data storage and compute environment to access data and run analytics

Results

- Speed up algorithm development and deployment
- Domain experts can focus on developing algorithms



Our Data Journey Framework





Getting Engineering Data to the Cloud

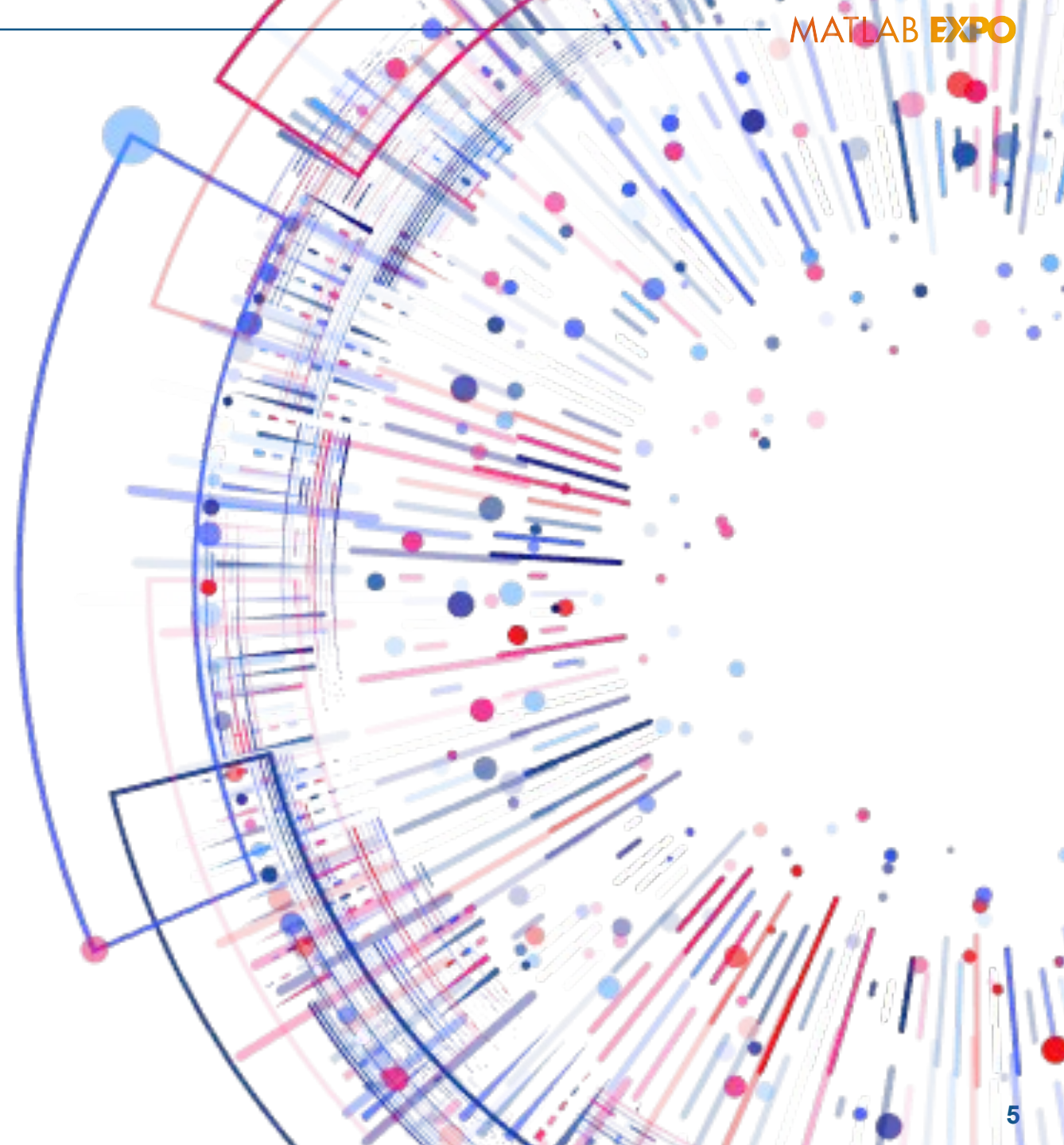
Igor Alekseev



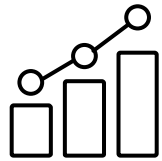
MATLAB EXPO 2021

Critical drivers for modern architecture

- Data volume
- Data variety
- Use case complexity
- Data access
- Breaking down silos



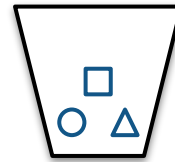
Get more value from R&D data



Growing
exponentially



From new
sources



Increasingly
diverse



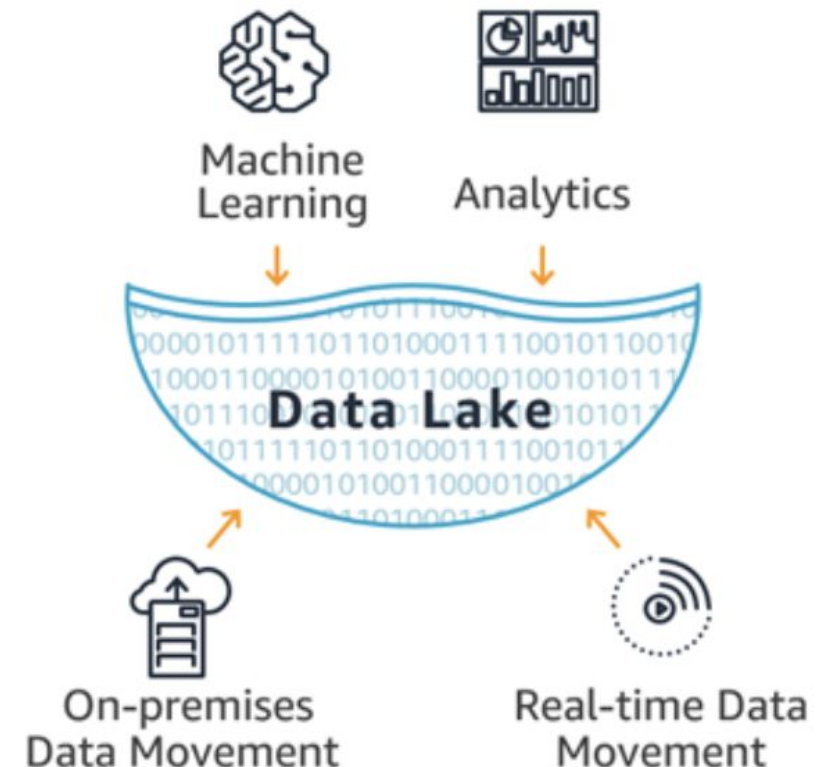
Used by
many people



Analyzed by
many applications

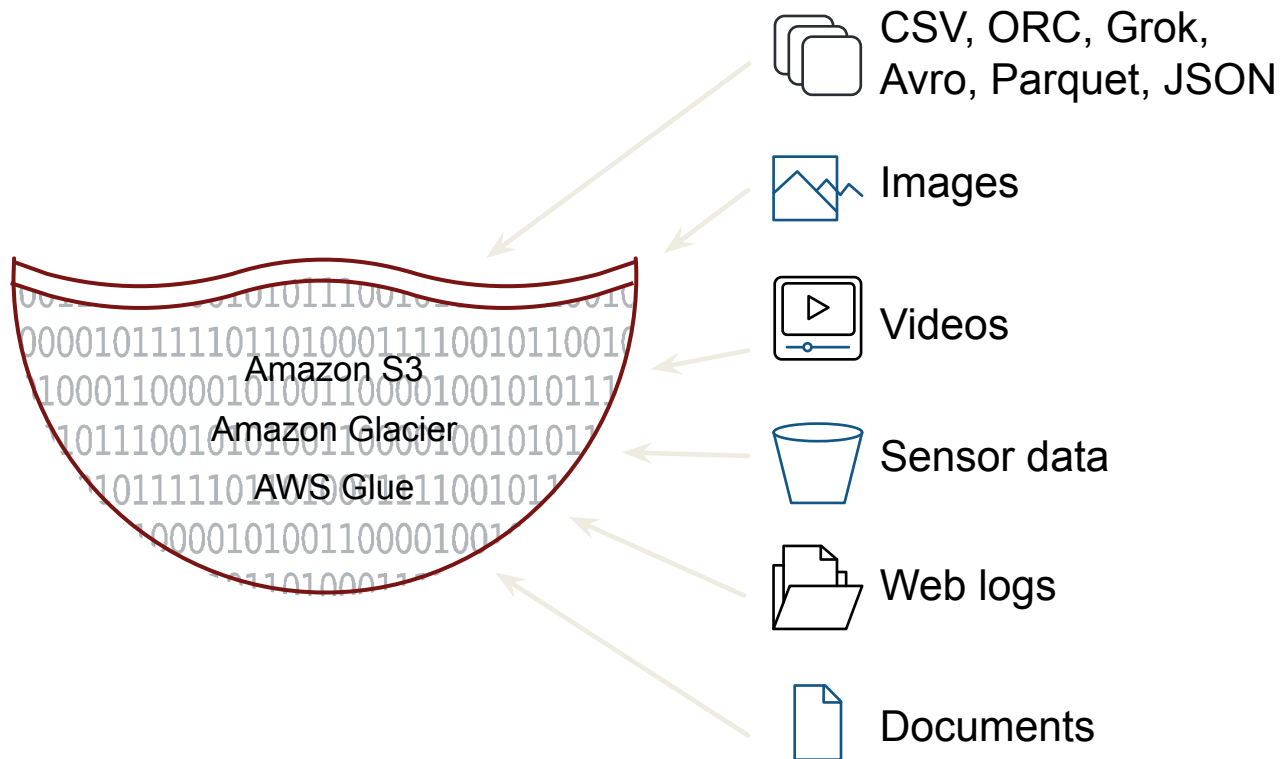
The Data Lake: Storage without limits

A data lake is an architectural approach that allows you to store massive amounts of data into a central repository, so it's readily available to be categorized, processed, analyzed, and consumed by diverse groups within an organization.



Data Lakes with Amazon S3 object storage empowers research

Secure, highly scalable, durable object storage with millisecond latency for data access



Store data in the format you want

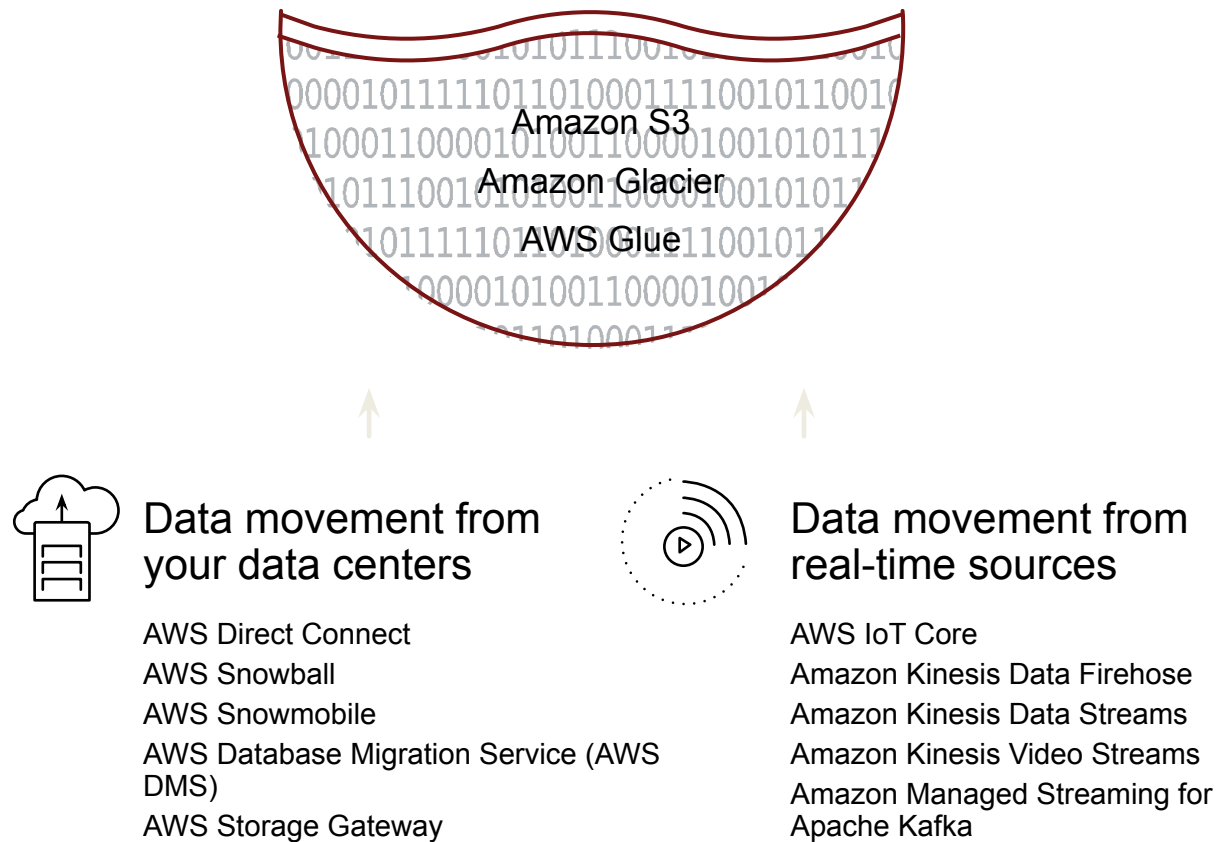
Built for 11 nines of durability

Three different forms of encryption

Run analytics and ML on data lake without data movement

Classify, report, and visualize data usage trends

How to get your data into AWS



Data movement from on-premises datacenters

- Dedicated network connection
- Secure appliances
- Ruggedized shipping container
- Database migration
- Gateway that lets applications write to the cloud

Data movement from real-time sources

- Connect devices to AWS
- Real-time data streams
- Real-time video streams



databricks

Helping everyone use the data in the cloud

Prasad Kona

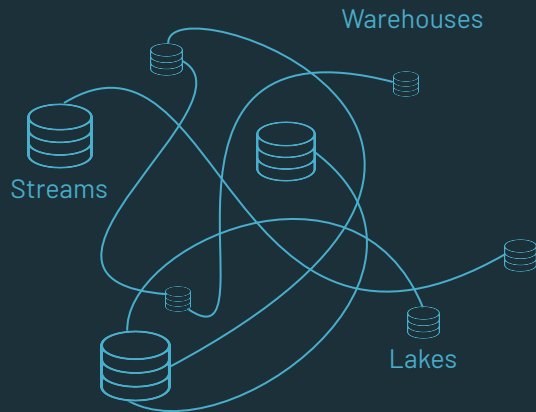


MATLAB EXPO 2021

Unlocking business value: Four challenges

1

Data is messy,
siload and slow



2

ML is hard,
Production is
harder



Data Engineers
in IT

Data Scientists
in Business



3

BI is limited to a
fraction of data

```
11000100011000100010001
000010111000100101010000
1110010101001111110011001
11010100011100110001100011
00010001000100001011100
01001010100001111001010
```

4

Lack of enterprise
readiness



Fragmented
security

Poor
reliability

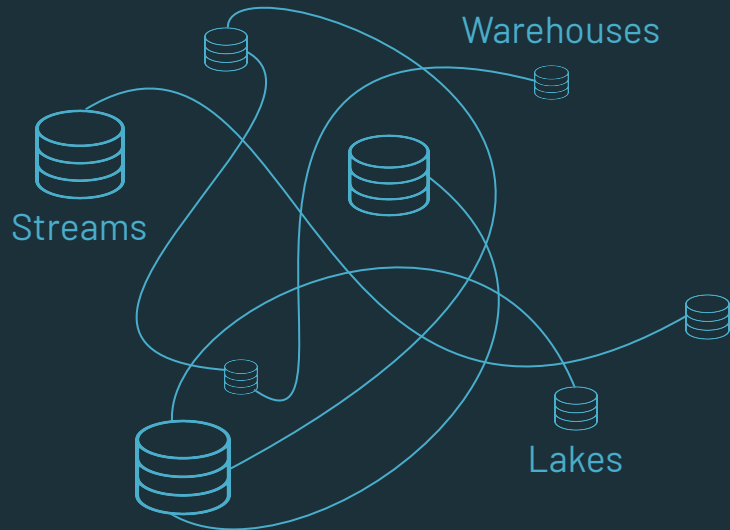


Disjointed
governance

Make all your data ready for analytics and ML

1

Data is messy,
siload and slow



Unified Data Service

Build open, reliable, fast data lakes with all your data

- Big Data
- Business Data
- Applications



Open High Quality Fast





Delta Engine
with
APACHE
Spark
API's

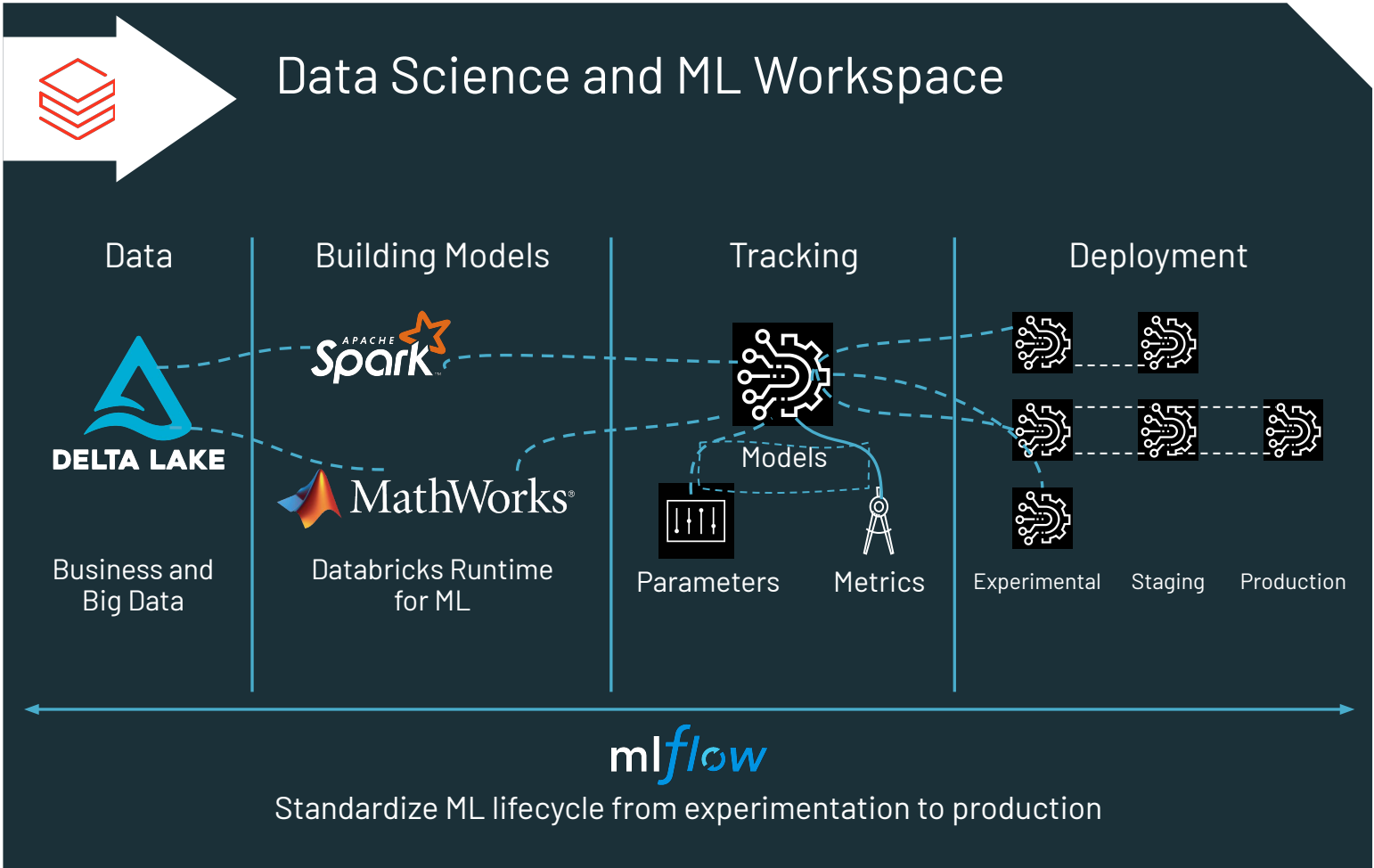
- MATLAB & Simulink
- BI Reporting
- Machine Learning

Unify data and ML across the full lifecycle

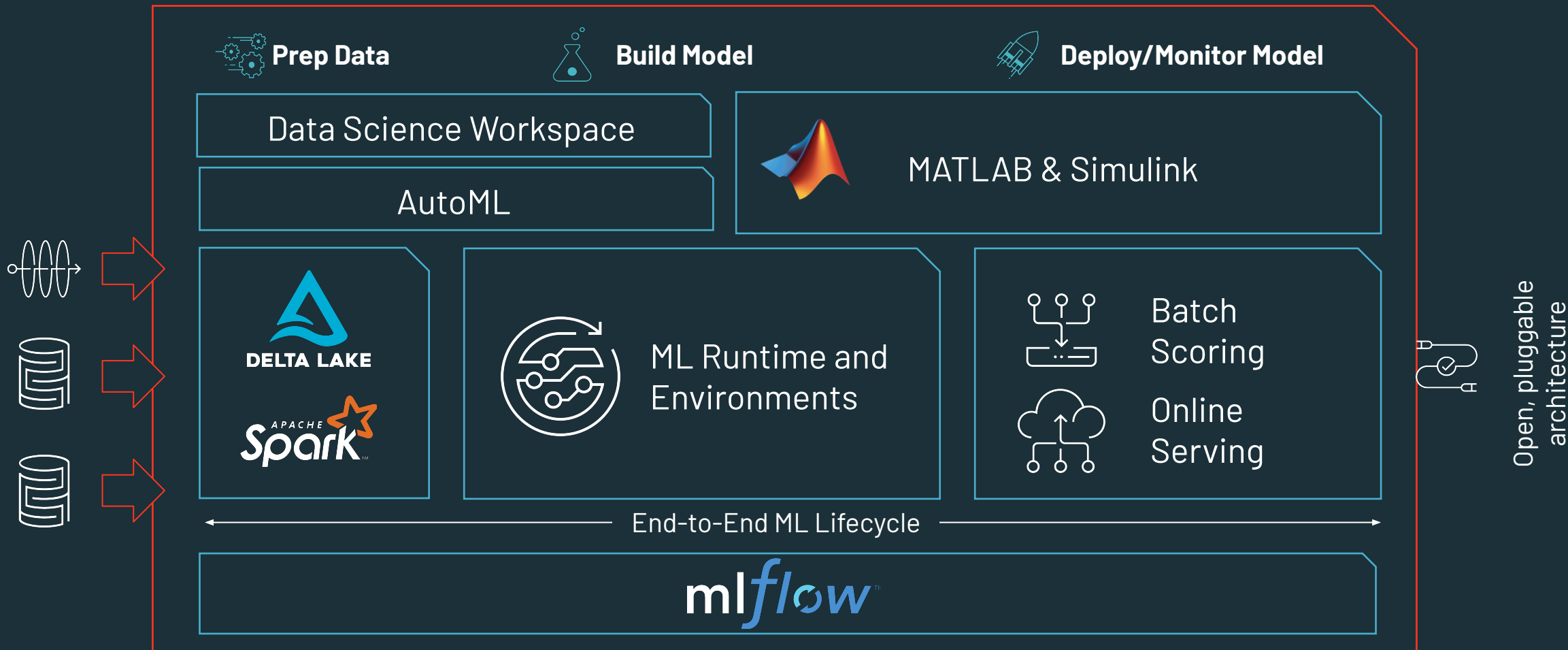
2 ML is hard, production is harder

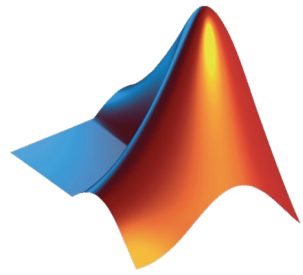
 Data Engineers in IT

 Data Scientists in Business



End-to-End Data Science and ML on databricks





MathWorks®

Leveraging cloud data and compute in production

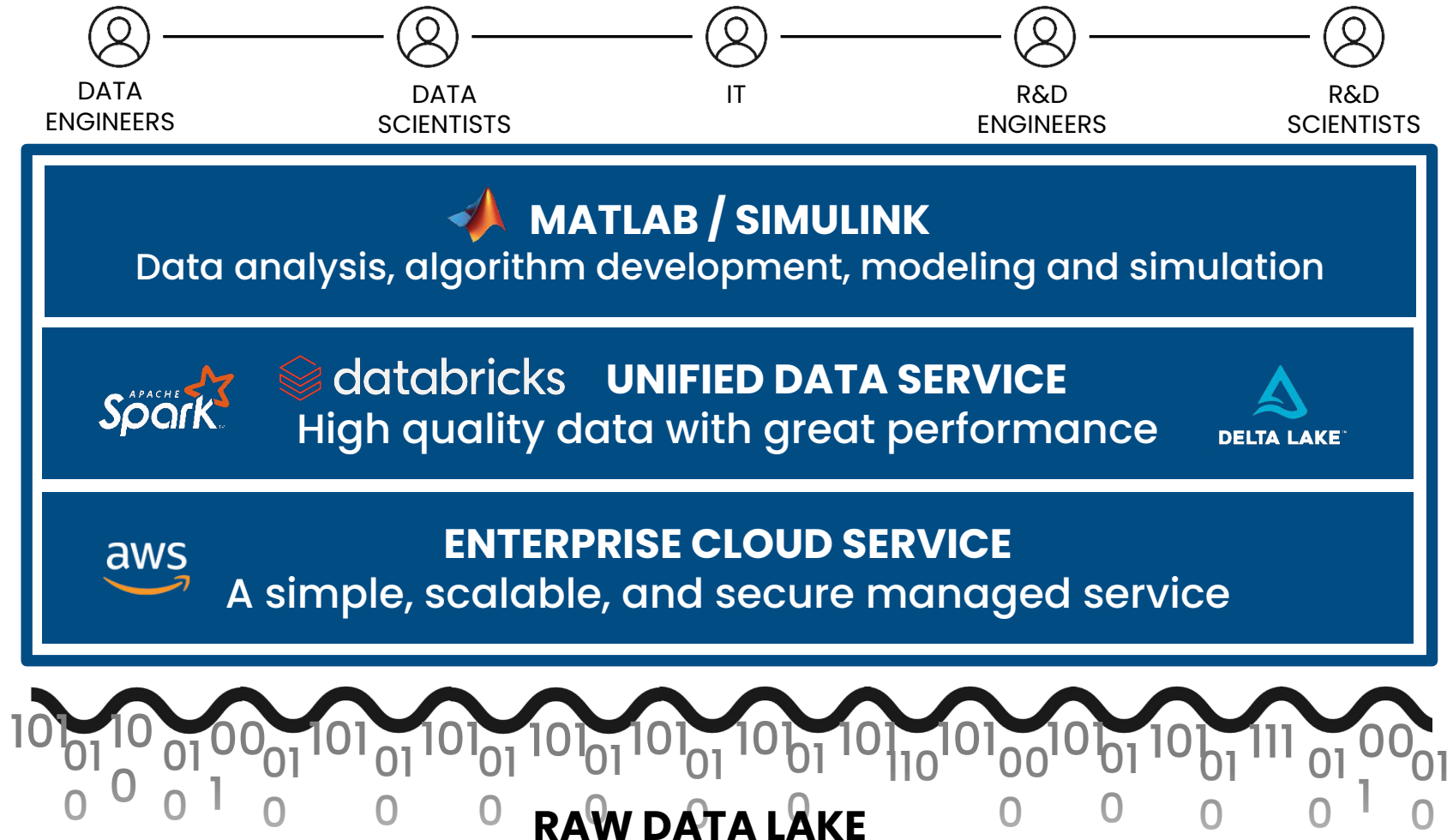
Arvind Hosagrahara



MATLAB EXPO 2021

Integration Approach

“The entire company on one data and compute platform in the cloud”



Improve Predictive Maintenance Performance using Transient Data

Challenges

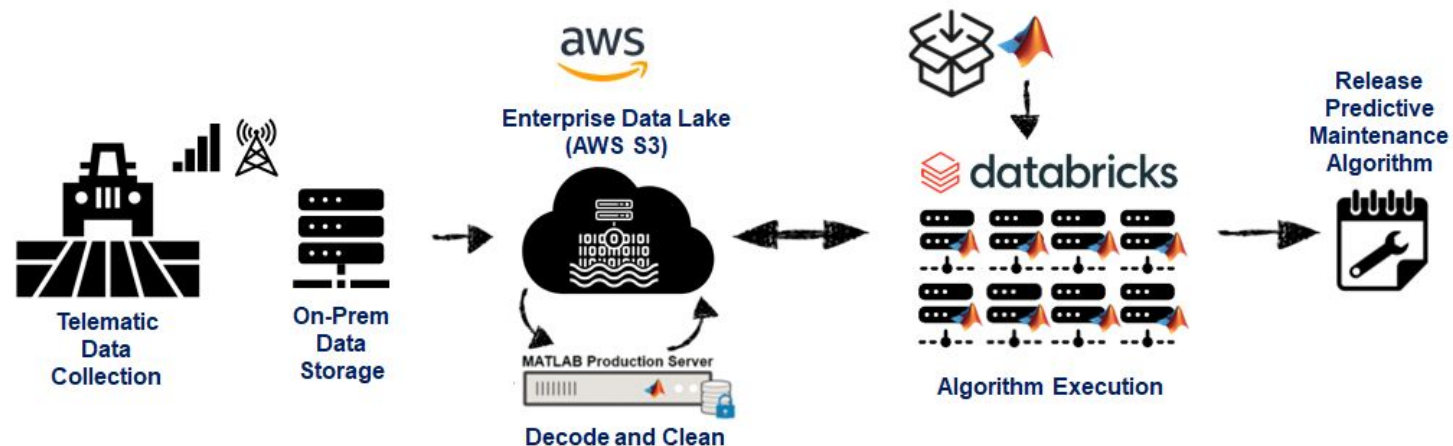
- Embedded controllers have limited computing power
- Engineers don't have easy access to data in Enterprise Data Lake

Solution

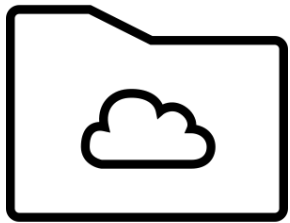
- Integrate MATLAB with cloud-based data storage and compute environment to access data and run analytics

Results

- Speed up algorithm development and deployment
- Domain experts can focus on developing algorithms



Ways to work with Databricks



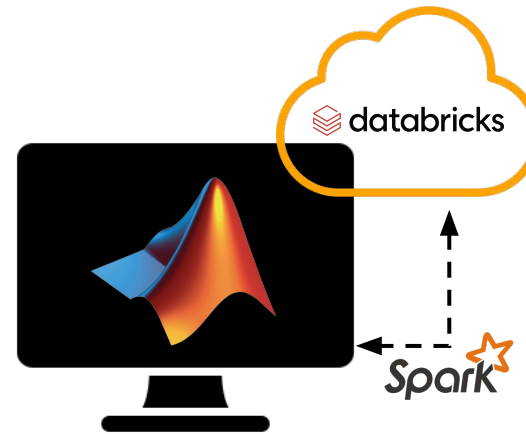
DBFS



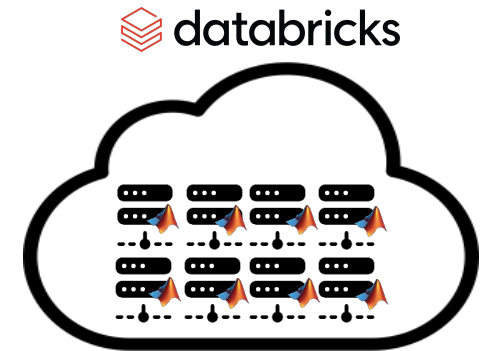
SQL / JDBC



Spark SQL



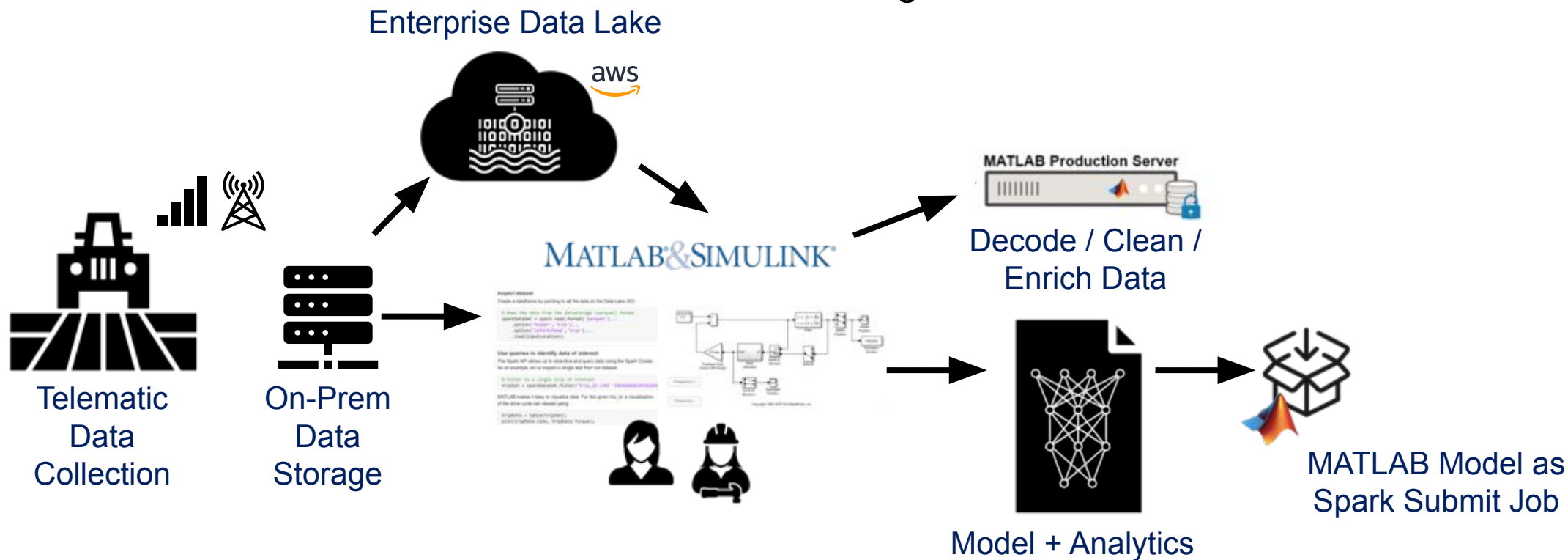
Databricks Connect



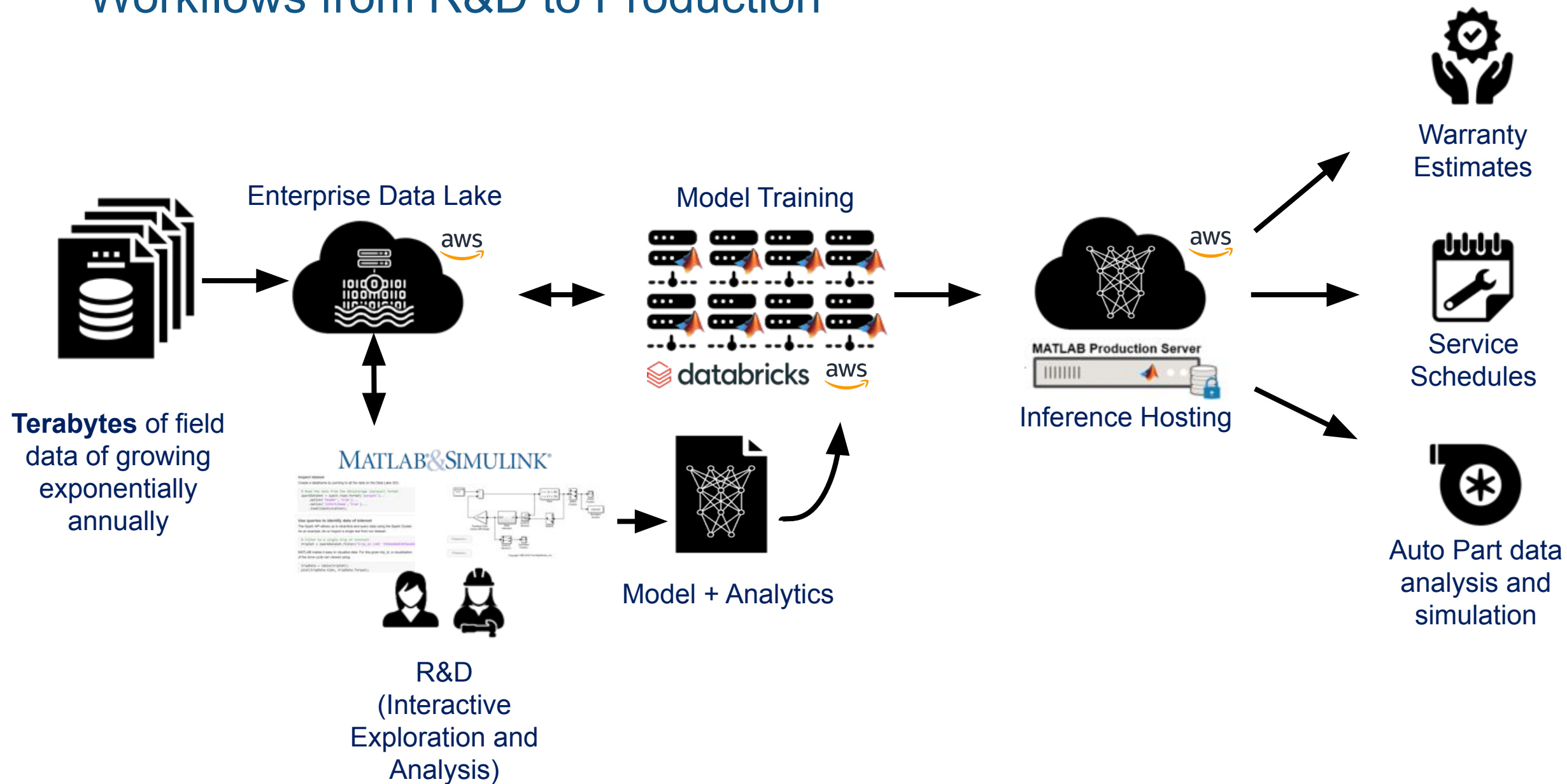
MATLAB Cloud
Execution

Data Engineering workflows

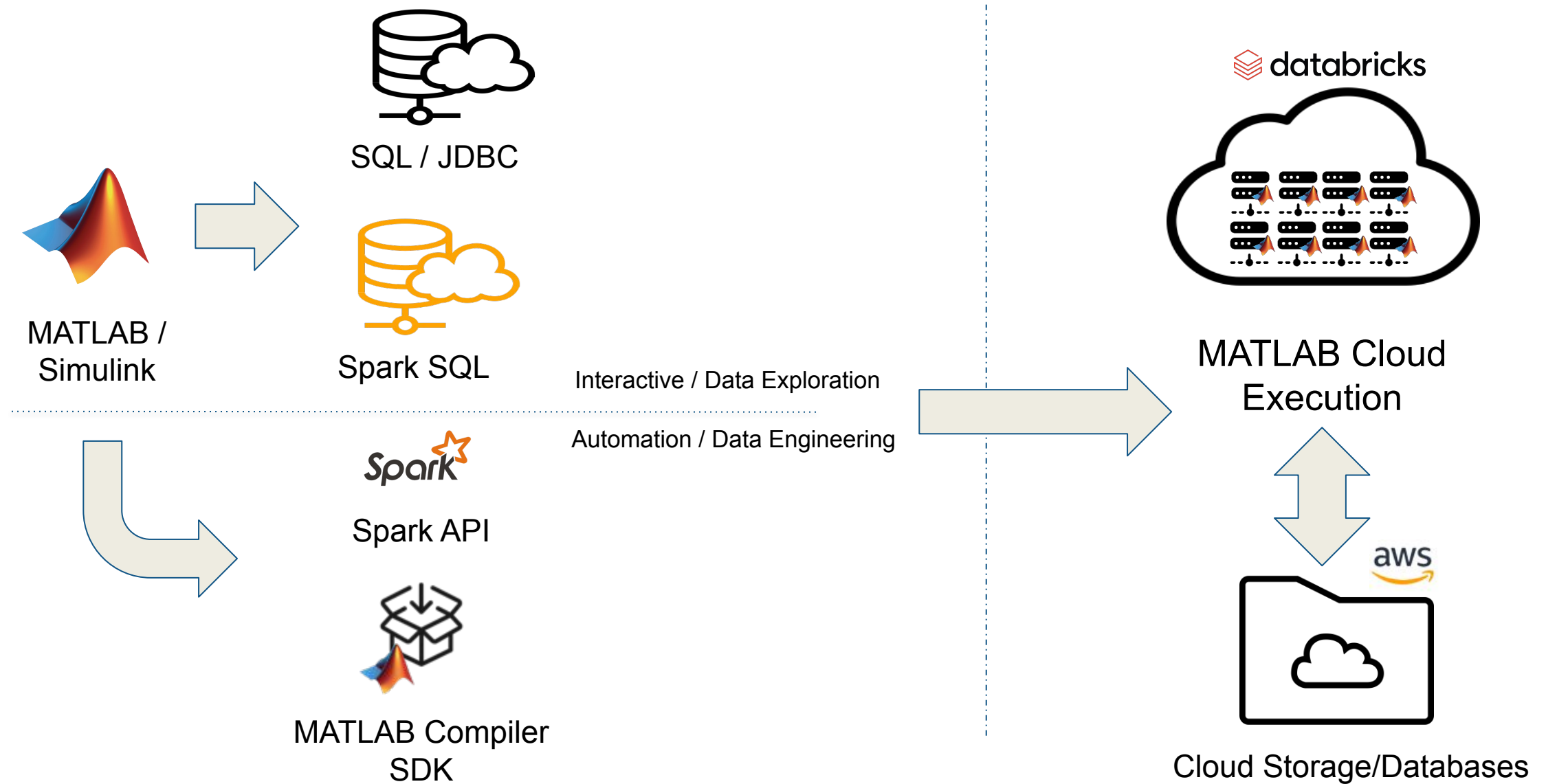
Helping engineers with data preparation algorithms



Workflows from R&D to Production



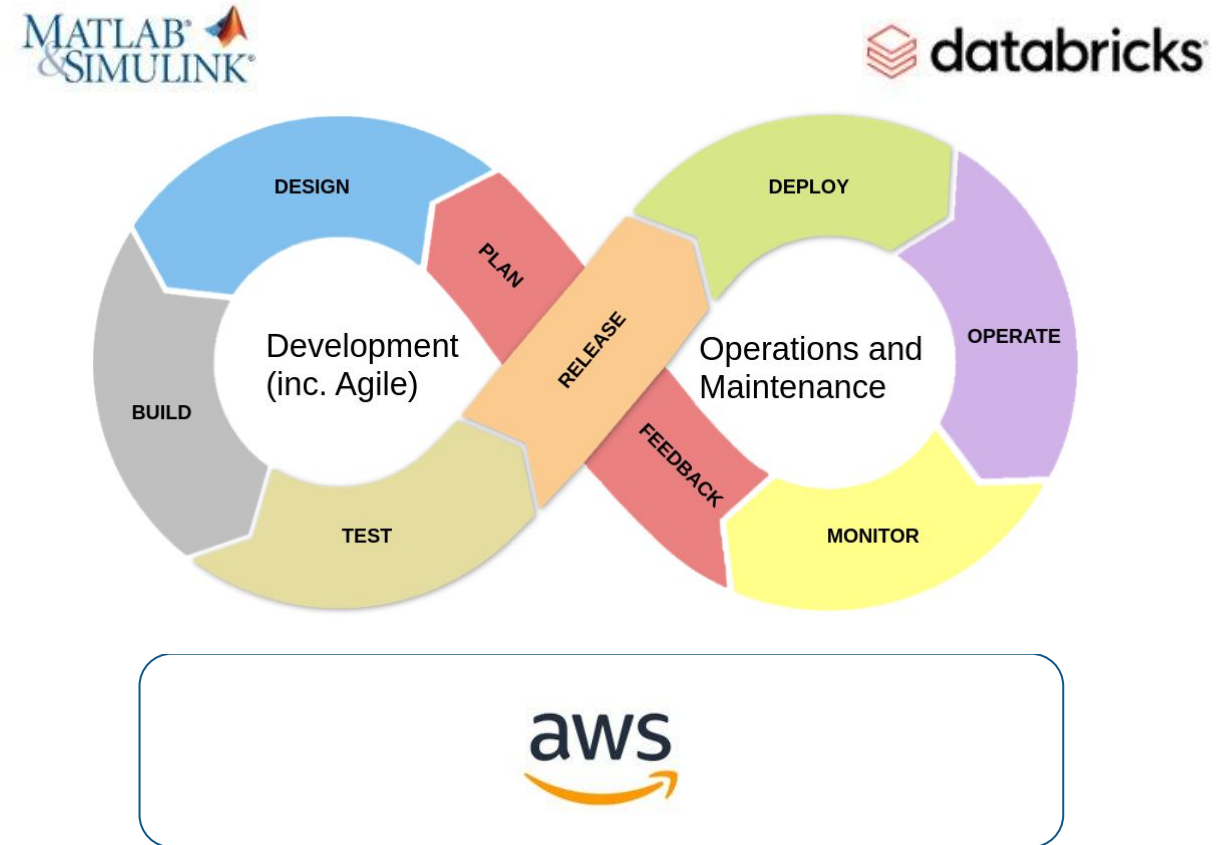
Databricks Workflows



Model DevOps

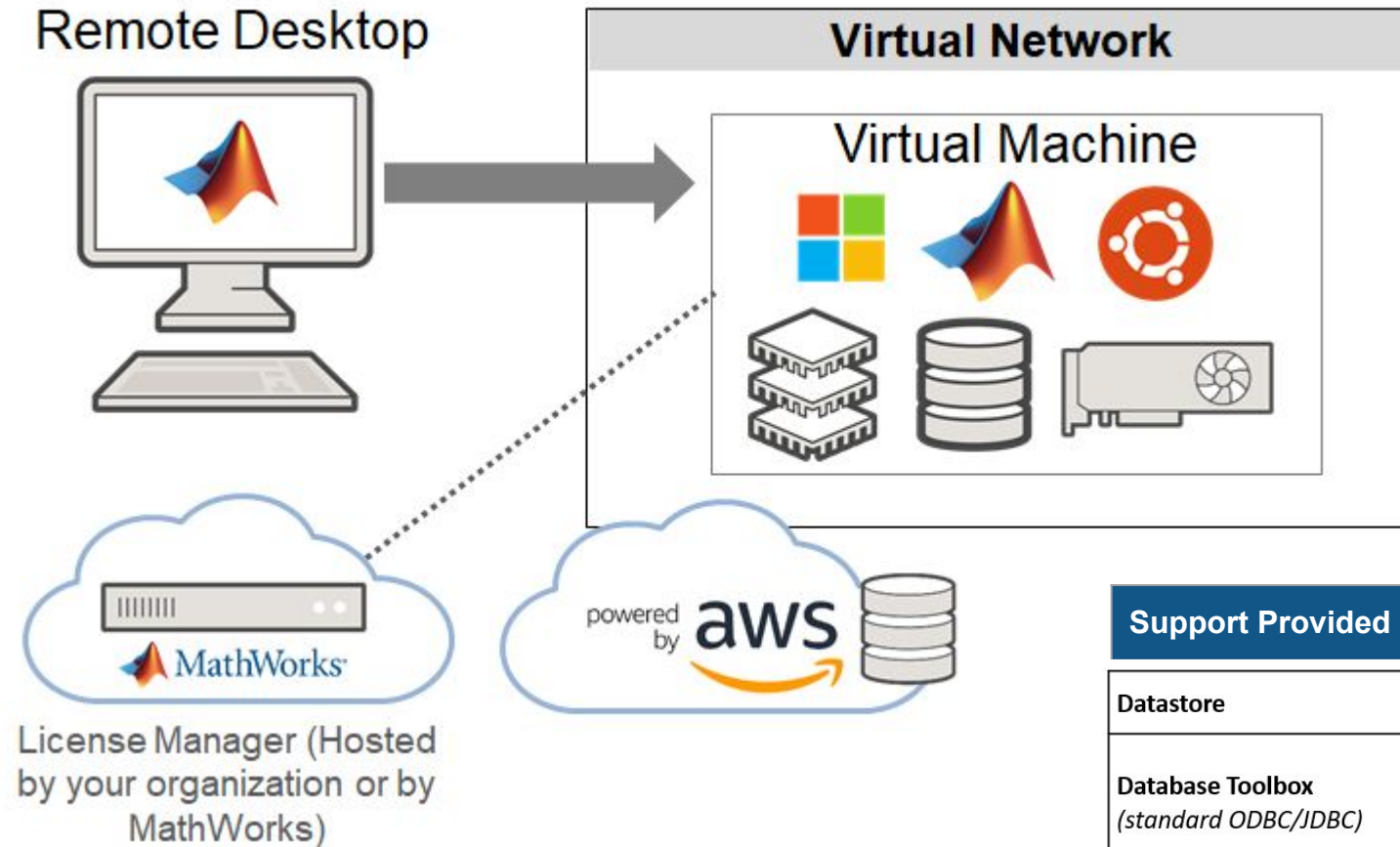
Workflows optimized for engineers to self-serve their operational needs

- Take machine learning and AI models from interactive development to production
- Build, Test and Deploy via automated CI/CD pipelines following best practices for security and scaling
- Monitor performance and refine designs quickly and painlessly
- Align with DevOps maturity best practices throughout the process



MATLAB Integrations for AWS

MATLAB Reference Architecture



Support Provided By:	Data Services		
Datastore	Amazon S3 <i>(read, out of memory data)</i>		
Database Toolbox <i>(standard ODBC/JDBC)</i>	Amazon Aurora	Amazon RDS for PostgreSQL/MySQL/MariaDB/Oracle/SQL Server	
Support Packages	Amazon S3 <i>(read/write/delete, encryption, access control)</i>	Amazon EFS <i>(NFS/Linux)</i>	Amazon Athena <i>(Query of S3 data)</i>

- <https://github.com/mathworks-ref-arch/matlab-on-aws>
- <https://github.com/mathworks-ref-arch/matlab-parallel-server-on-aws>
- <https://github.com/mathworks-ref-arch/matlab-production-server-on-aws>

Summary

- MATLAB offers an integrated experience to use the best-in-class storage and compute platforms
- Enable engineers to analyze datasets at scale in the cloud
- Support a variety of workflows to support Model DevOps best practices
- Leverage partnerships with AWS and Databricks to support your use of MATLAB in the cloud
- Please talk to us if you want to learn more on how to make this happen for your company

MATLAB EXPO 2021

Thank you

