

# DataWorks Challenge 2024

# Welcome!





# DataWorks Challenge 2024

# Ice Breaking



Poll/Survey



# DataWorks Challenge 2024

# Welcome Speech

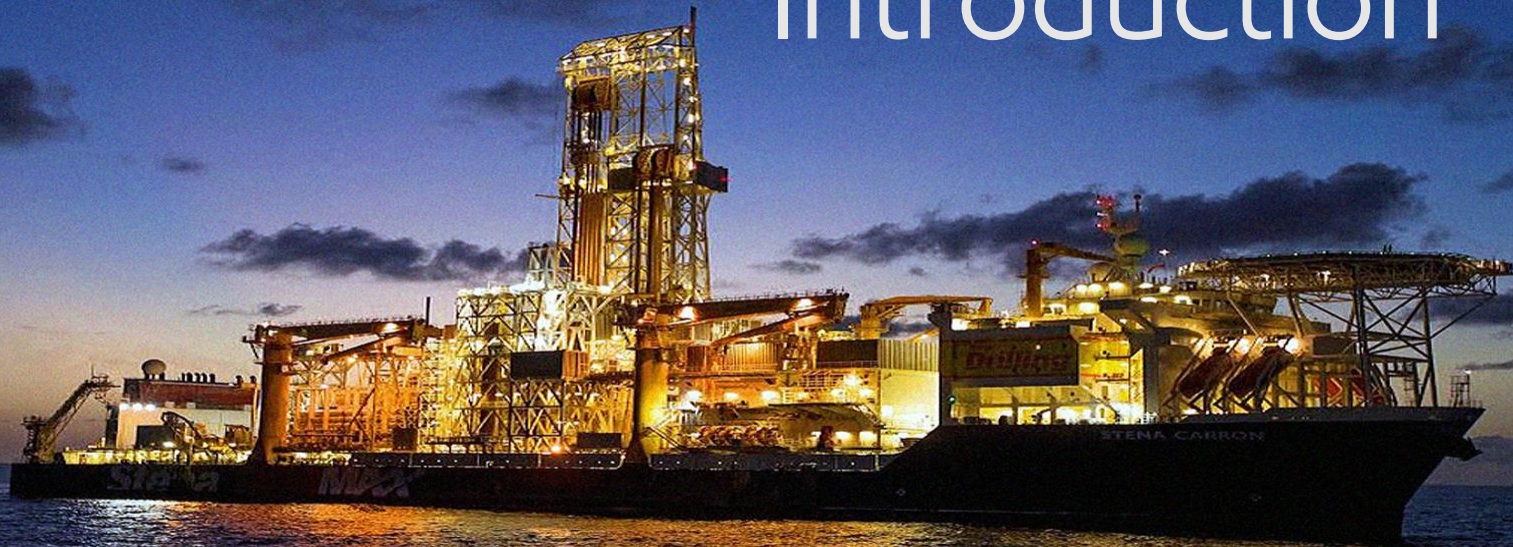


Azmil Ab Aziz  
Upstream IT  
General Manager



# DataWorks Challenge 2024

# DataWorks Introduction



# DataWorks Challenge | The DataWorks



## About DataWorks

- ExxonMobil KL Global Business Center (KLGBC)'s 2<sup>nd</sup> data showcase involving undergraduate students from local universities all over Malaysia



## Benefit

- Opportunity to explore, analyze, and visualize a solution to a real use-case in ExxonMobil.
- Allow students to offer a fresh perspective and be creative, collaborative, and innovative for a particular effort.



## Objective

- Discover local talents with various data & analytics skillsets.



# DataWorks Challenge 2024

# Key Events & Timeline



Muhammad Haikal Iman Osman  
Data Visualization Engineer

# DataWorks Challenge | Timeline

## Registration

- Targets **Malaysian undergraduate students**
- Across **9 public & 9 private** universities respectively

**Aug 7 - 26**

## Solution Submission

- Students to **submit machine learning model & video pitch**
- Top 5 shortlisting conducted by a team of data experts

**Sept 21 @  
11am**

## Virtual Pitching Mentoring

- Top 5 finalist teams to share challenge period experiences
- **Virtual pitching mentorship** by team of industrial experts

**Oct 8 & 10**

**Sept 3**

**Oct 1**

**Oct 15**

## Virtual Launching Program

- Reservoir Well Use Case briefing by Reservoir Engineer
- **DataWorks Challenge 2024 solutioning kickstarts** with ~2.5 weeks

## Finalists Announcement

- **Top 5 finalist** teams will be announced for Finale
- Finalists to be invited for virtual pitching mentoring

## DataWorks Finale

- **Finalists (5 lecturers & 20 students)** to pitch for championship physically
- Location: **Menara ExxonMobil TR 1 & 2**



# DataWorks Challenge 2024

# Quiz



Poll/Survey



# DataWorks Challenge 2024

# Introduction to Use Case



Chun Yan Lim (CY)  
Reservoir Engineer



# DataWorks Challenge | Introduction



Property Agent, Chris

Noticed increase in demand for residential space near City Centre in Kuala Lumpur

Malaysia poised for economic growth with a surge in expats

By Sharen Kaur - March 6, 2024 @ 11:03am

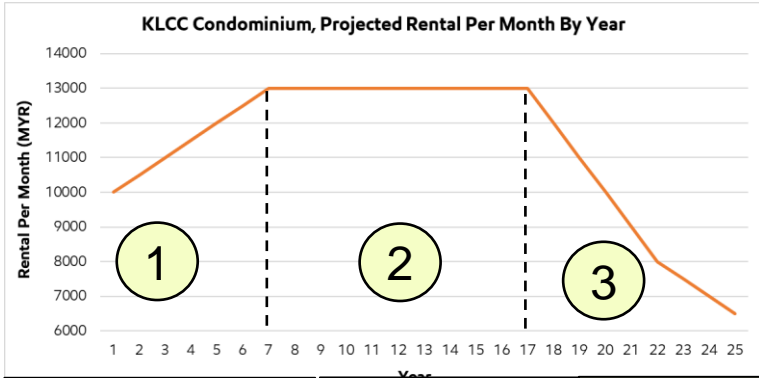


KUALA LUMPUR: The surge of expatriates in Malaysia indicates that the country is becoming more and more of a top choice for investors and international companies looking for growth and expansion prospects.



Walking distance to Petronas Twin Towers, ideal for working Expats in KL

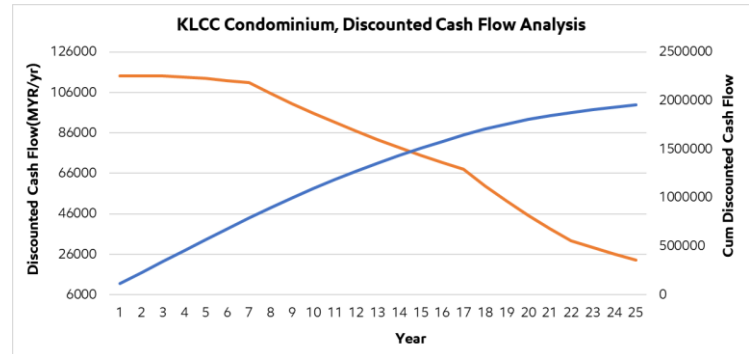
Asking Price **2 Million**



Growth in rental income, underpin by growth in population of expats / living standard

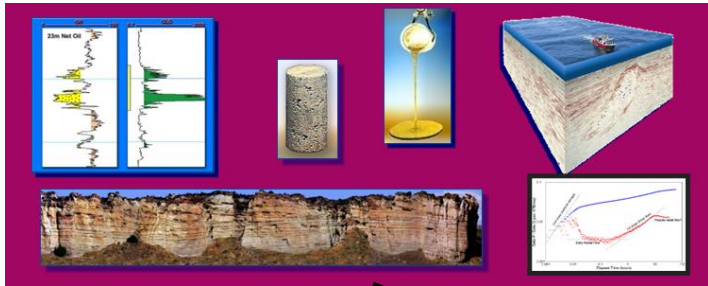
Income plateaued, rental inflation offset by newer condominiums

Decline in rental income, as place is getting older with poor maintenance

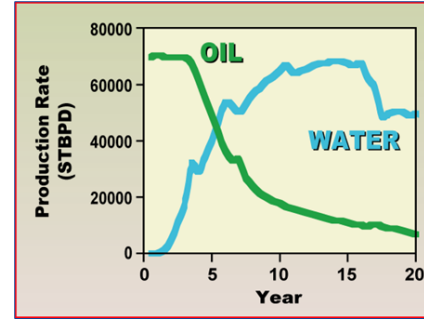




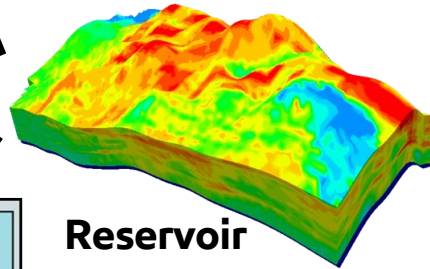
# DataWorks Challenge | Forecasting in Oil & Gas



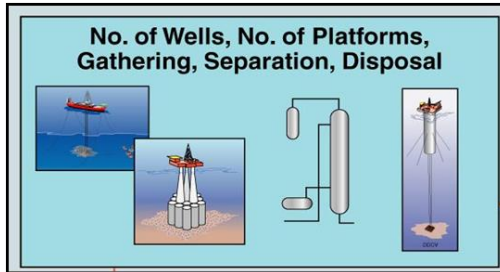
**Characterize**



**Production  
Forecast**



**Reservoir  
Simulation  
Model**



**Design**



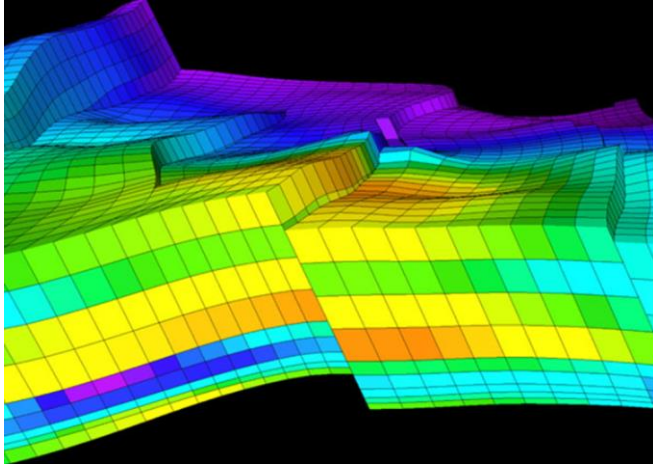
**History Match &  
Optimize**



**Evaluate!**

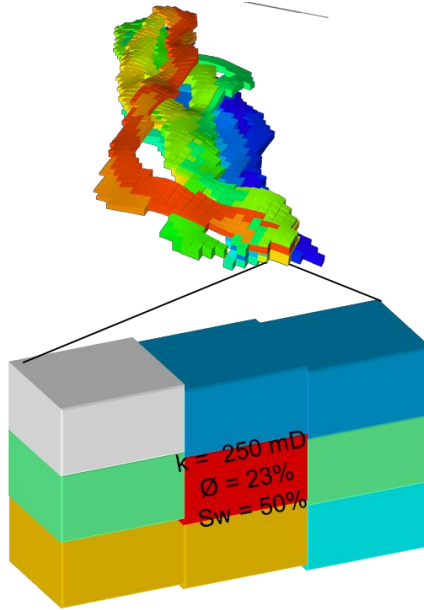


# DataWorks Challenge | The Reservoir Simulation Model



## Reservoir Simulation Model

- 3D grid representing the field, separated into thousands/millions of data cells
- The data in each cell represents the properties of the rock and fluids at that point in space
- Each cell gets one value for each property



## Rock Properties:

- Permeability
- Porosity
- Transmissibility

$$\nabla \cdot \left[ \frac{k_r k}{\mu_o B_o} \left( \nabla P_o - \rho_o \frac{g}{g_c} \nabla z \right) \right] = \frac{\partial}{\partial t} \left[ \frac{\phi S_o}{B_o} \right] + \text{source}$$

"Geologic" properties:  $\phi$   $k = f(P, T)$

$$\nabla \cdot \left[ \frac{k_r k}{\mu_w B_w} \left( \nabla P_w - \rho_w \frac{g}{g_c} \nabla z \right) \right] = \frac{\partial}{\partial t} \left[ \frac{\phi S_w}{B_w} \right] + \text{source}$$

Fluid properties:  $\mu$ ,  $\rho$ ,  $B$ ,  $R_s = f(P, T)$

$$\nabla \cdot \left[ R_s \frac{k_r k}{\mu_o B_o} \left( \nabla P_o - \rho_o \frac{g}{g_c} \nabla z \right) + \frac{k_r k}{\mu_g B_g} \left( \nabla P_g - \rho_g \frac{g}{g_c} \nabla z \right) \right] = \frac{\partial}{\partial t} \left[ \phi \left( \frac{R_s}{B_o} S_o + \frac{S_g}{B_g} \right) \right] + \text{source}$$

Displacement properties, depend on rock-fluid interaction:  $k_r$  and  $P_c = f(S)$

$$S_o + S_w + S_g = 1$$

$$P_{con} = P_o - P_w$$

$$P_{cog} = P_g - P_o$$

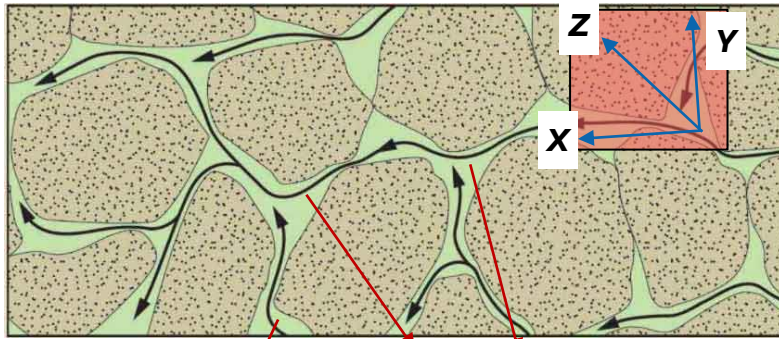
Partial Differential Equation (PDE) for solving mass balance

## Key Takeaways:

- Model output / forecasts are dependent on a variety of input data
- Rock properties such as permeabilities / porosities / fault transmissibility's are often uncertain



# DataWorks Challenge | Basic Concepts on Reservoir Engineering

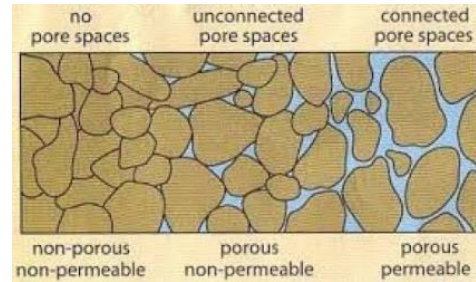


Pore space  
containing fluid

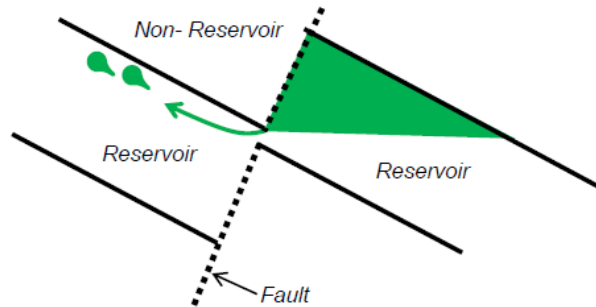
Pathway for fluid  
flow

**Permeability** is the property of a porous material that determines **how easily fluid flows through that material**—a basic measure of the producibility and injectivity of subsurface formations.

Anisotropy is the property of being directionally dependent. A rock is considered to show anisotropy of permeability when the **measured permeability varies depending on the direction of fluid flow**



**Porosity** determines **reservoir storage capacity**. It is defined as the ratio of void space, commonly called pore volume, to bulk volume and is reported either as a fraction or a percentage



In geology, a **fault** is a planar fracture or discontinuity in a volume of rock across which there has been significant displacement as a result of rock-mass movements (wiki)

**Fault transmissibility** determine the level of connectivity of a separate reservoir due to fault segments

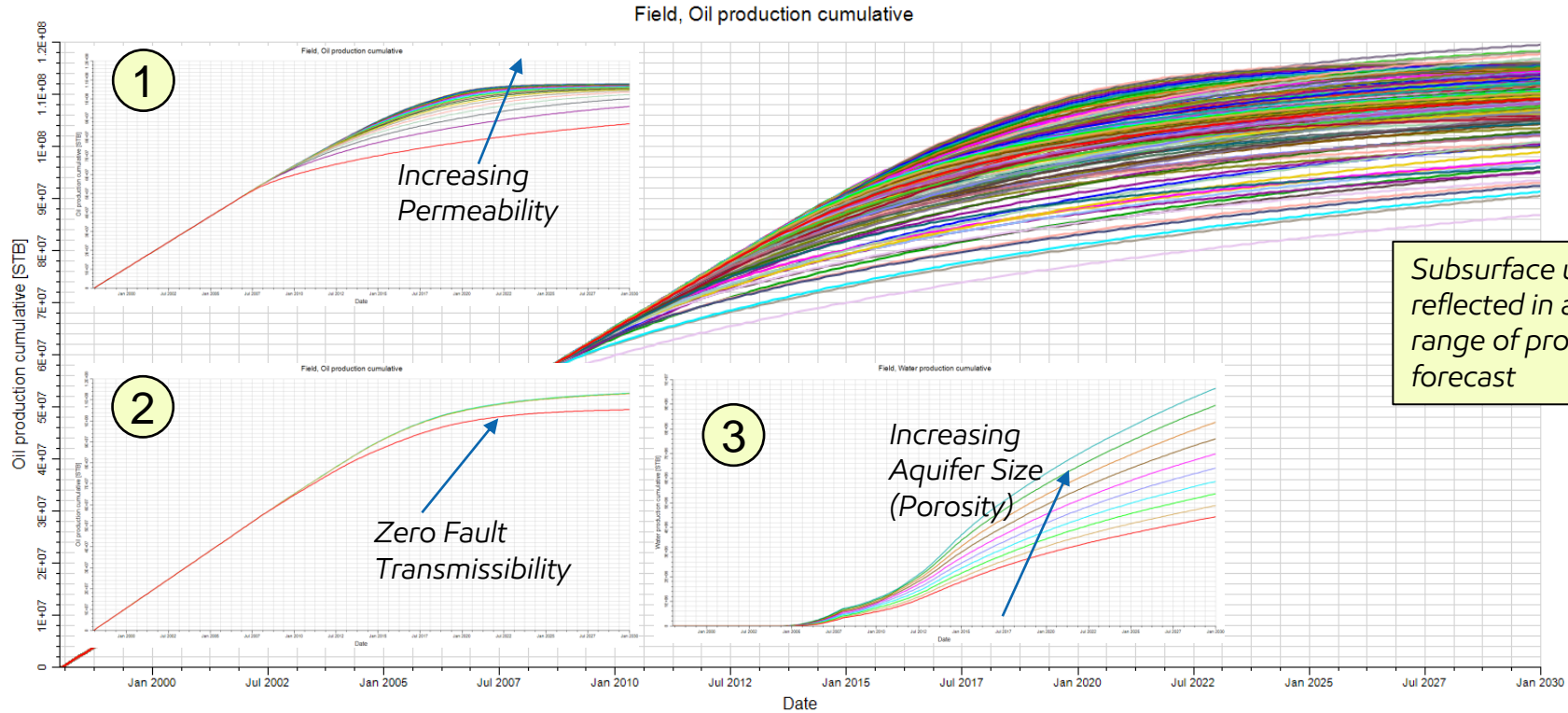


# DataWorks Challenge | The Dataset Description

Columns	DataType	Unit	Descriptions
index_x	integer	dimensionless	representing a point location in a 3D space (x-direction)
index_y	integer	dimensionless	representing a point location in a 3D space (y-direction)
index_z	integer	dimensionless	representing a point location in a 3D space (z-direction)
PERMX	float	milliDarcy(Md)	Permeability in the x-direction, representing the ease of fluid flow within the pore space
PERMY	float	milliDarcy(Md)	Permeability in the y-direction, representing the ease of fluid flow within the pore space
PERMZ	float	milliDarcy(Md)	Permeability in the z-direction, representing the ease of fluid flow within the pore space
PORO	float	dimensionless	Denotes the amount of pore space - representing the amount of fluid within the subsurface
Fault_Trans	float	dimensionless	Fault represents one of the geological subsurface barrier to flow. Transmissibility indicate how permissibile the barrier is to flow; 0 means the fault is completely sealing while 1 means the fault is completely open
Date	Date	mm/dd/yyyy	
Cumulative Gas Production	float	Kilo Standard Cubic Feet Per Day (KSCF/D)	Gas Production Volume on a cumulative basis
Cumulative Oil Production	float	Stock Tank Barrels Per Day (STB/D)	Oil Production Volume on a cumulative basis
Cumulative Water Production	float	Stock Tank Barrels Per Day (STB/D)	Water Production Volume on a cumulative basis



# DataWorks Challenge | The Dataset Analysis





# DataWorks Challenge | The Business Case

## Inputs

Inputs

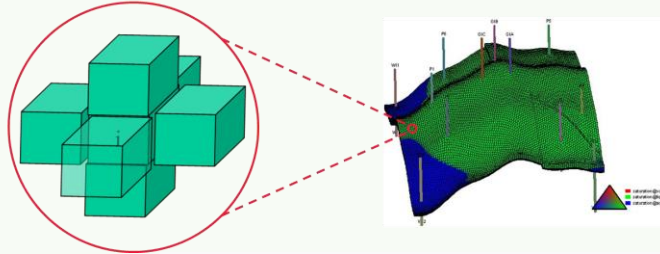
Subsurface  
Uncertainty

## Model

$$\left(\rho_o S_o V_p\right)_i^{t+\Delta t} - \left(\rho_o S_o V_p\right)_i^t = \Delta t \sum_m \left( \rho_o \left( \frac{k_{ro}}{\mu_o} \right) T \Delta P \right)_{i,m} + \Delta t \rho_o q_{o,well_i}$$

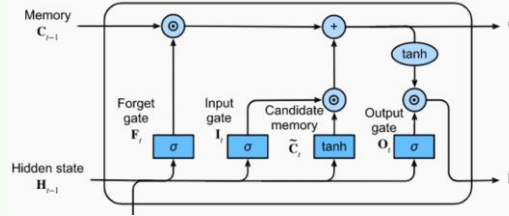
Mass of oil at end of timestep      Mass of oil at start of timestep      Net flow through faces during timestep      Flow through well during timestep

Physics  
Based  
Model

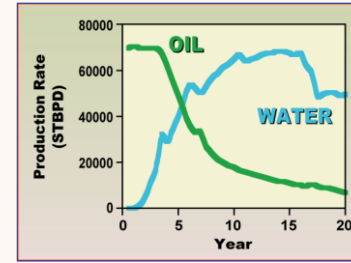


ML Model

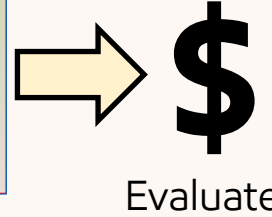
~ Reduced Order Model  
~ Faster and retains  
sufficient predictability



## Output



Forecast



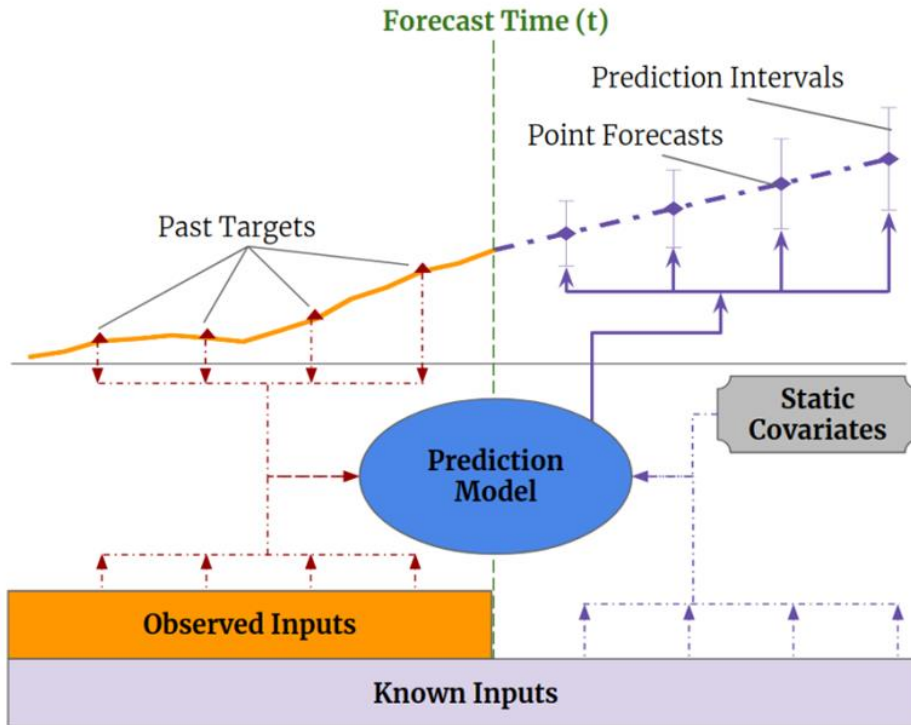
Evaluate

### Problem Statement:

- Physics Based Model is computationally expensive
- High degree of uncertainty associated with subsurface representation



# DataWorks Challenge | The Use Case Expectation



## Remember:

- Goal of the model is to make prediction on production forecast so that we can evaluate if the field will make money
- Rock properties are assumed to be static (i.e. will not change with time)
- There are other dynamic properties (time varying) such as equipment downtime that will also affect forecast but that will not be covered in this exercise for simplicity

# DataWorks Challenge | Scene Setting

## The Five Classical Stages of Machine Learning

### ExxonMobil

1. Decode on Problem  
(What are we modeling?)

Problem: High Fidelity Physics Model Reduction

2. Curate Data  
(What DATA will inform the model?)

Data: Simulation Input and Output  
Dataset generated from running physics-based model

### Students

3. Design an Architecture  
(RNN, Autoencoder, DMD?)
4. Craft a Loss Function  
(What models are "GOOD"?)
5. Employ Optimization  
(What Algorithms to Train Model?)





# DataWorks Challenge | Business Case

## Problem Statement

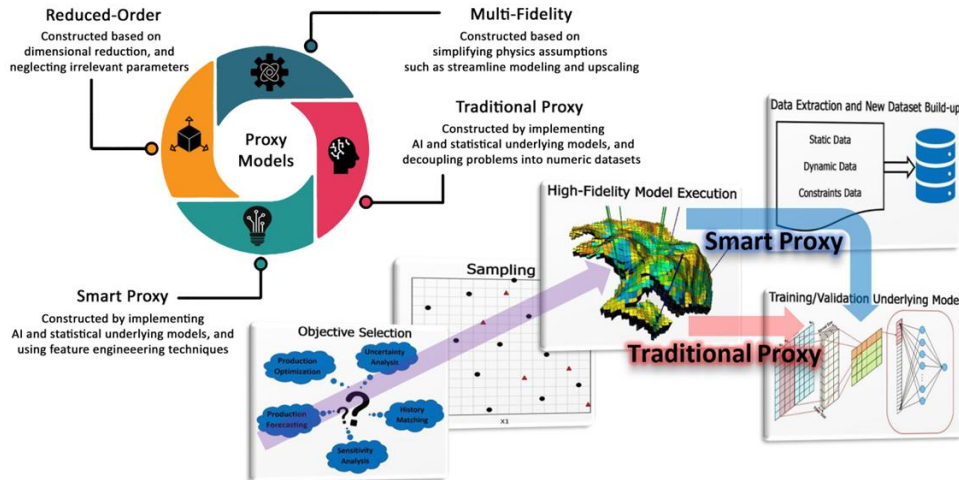
- Develop Proxy Surrogate Model that is significantly faster than the full field physics model and retains sufficient predictability

## Expected Results

- Proxy, reduced order model that able to accurately produce forecasting results for a number of test cases

## Business Value

- Leverage Proxy model to :
  - Explore uncertainty space
  - Generate History Matching insights
  - Probabilistic Assessment of problems by running the model a million times in a monte-carlo manner



## DataWorks Challenge | References

- Hu Huang et al. "A Deep-Learning-Based Graph Neural Network-Long-Short-Term Memory Model for Reservoir Simulation and Optimization with Varying Well Controls, SPE 2023
- Zhen Zhang et al, "Robust Method for Reservoir Simulation History Matching Using Bayesian Inversion and Long-Short-Term Memory Network-Based Proxy", SPE 2023



# DataWorks Challenge 2024

## Starter Pack



Teo Sheng Pu  
Data Engineer



# DataWorks Challenge 2024

## Judgement Criteria FAQ ShareFile Access



Muhammad Haikal Iman Osman  
Data Visualization Engineer



# DataWorks Challenge | Judging Criteria

## **Statistics & Analytics**

The solution utilizes fundamental principles of statistics and analytics. The solution is as accurate and unbiased as the data allows.

EG. We will test the accuracy of your model on Testing Dataset

## **Data Visualization**

The solution is easy to understand for its intended user.

## **Effective Communication**

Understanding of the business needs and is able to tell a compelling story with the data. The solution has the potential to generate real business value.

## **Insights**

The solution is applicable, insightful, and thorough for given use cases.

## **Bonus**

The solution should be robust, Innovative/Creative and perform stably.

## DataWorks Challenge | FAQ

1. Can I use other tools for the analytics and machine learning modeling?
  - Yes, please state the tools being used and provide any files related for your task.
2. Can I work on either one of Data Analytics or Machine learning modeling?
  - Yes, teams can choose to pick either one or both of the domains that suits their skill sets.
3. Regarding the shared folder, will other teammates able to access the shared folder as well?
  - For your own team, only the supervisor and team lead has been granted the access to the shared folder in ShareFile.
  - For other competitors' team, they will not be able to view the documents in your respective shared folder in ShareFile.
4. May I ask, is it permissible if the team get assistance or supervision from lecturer?
  - Yes, it is highly encouraged for the team to work closely with your supervisor.



## DataWorks Challenge | FAQ

5. If we use python fully, do we still need to visualize the data in tableau

- You can use tableau if you are comfortable for visualization, if you use python for both ML and Visualization, then you can just submit the python file and any other visualization in docs/appendix (in jpeg)

6. What's the difference between test and train data?

- Train data - is the dataset used for the training of machine learning model.
- Validation - is the dataset used to evaluate the machine learning model performance.
- Test data - to test robustness of the model.

7. What is the data type of the dataset used for the challenge?

- Numerical.

8. Is the dataset structured or unstructured?

- Structured.



## DataWorks Challenge | FAQ

9. Will there be any dataset provide for this competition, and if so, where can I get the dataset?

- Can be found within the ShareFile folder shared to supervisor & team lead.

10. What would be the timeline of the challenge, and when will be the submission deadline?

- For preliminary stage, the submission deadline would be September 21, 2024 (11:00am).

11. What tool are we going to use?

- You are allowed to use any tool to develop the solution as long as it follows the requirements on the deliverables.

12. May I know if our team supervisor will need to be physically present for the final showcase?

- Yes, it is necessary for top 5 supervisors and teams to be present at Menara ExxonMobil for the Finale .

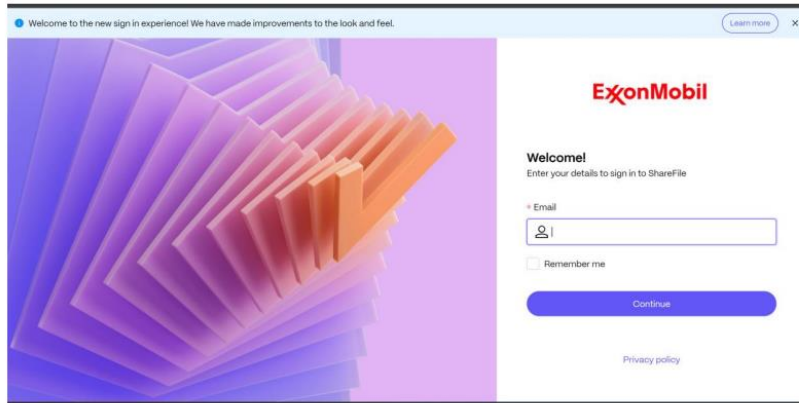




# DataWorks Challenge | ShareFile Access

## Creating Account & Accessing ShareFile

1. Access the URL <https://exxonmobil.sharefile.com>



Welcome to the new sign in experience! We have made improvements to the look and feel. [Learn more](#) X

**ExxonMobil**

**Welcome!**  
Enter your details to sign in to ShareFile

+ Email

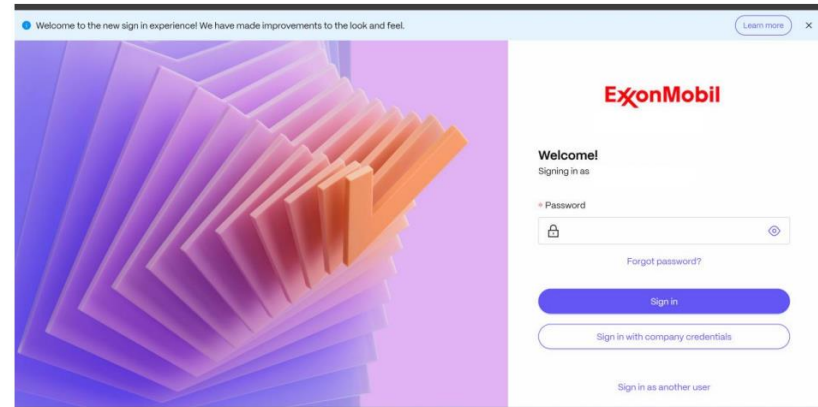
☐ Remember me

[Continue](#)

[Privacy policy](#)

2. Proceed to input your email that you have registered with DataWorks, as that email has been granted the access. Then, select "Forgot password?" to set your password following the instruction in ShareFile.

Note: Only university's supervisor & team lead has been granted access.



Welcome to the new sign in experience! We have made improvements to the look and feel. [Learn more](#) X

**ExxonMobil**

**Welcome!**  
Signing in as

+ Password

[Forgot password?](#)

[Sign in](#)

[Sign in with company credentials](#)

[Sign in as another user](#)

# DataWorks Challenge | ShareFile Access

3. Follow the process of "Sign In". You may be required to read & accept the terms and conditions of ShareFile usage.
4. Upon successful "Sign In", the following Folders view will be available.  
There would be 2 folders, namely  
DataWorks 2024 UseCase Kit – Where you could find the challenge starter kit  
<Your Team Name> – Where you need to upload solution submission



## Accessing DataWorks Challenge 2024 Use Case Kit

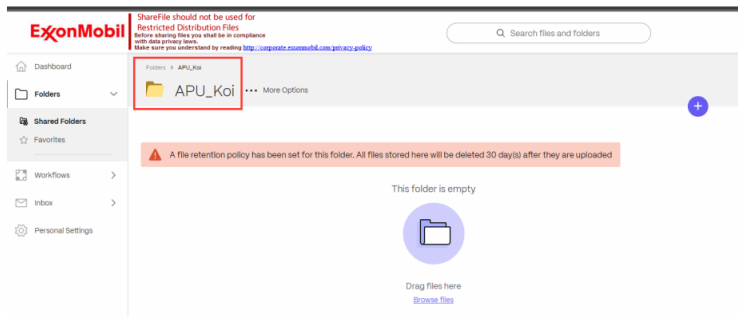
1. Select "DataWorks 2024 UseCase Kit" folder to access the challenge starter kit.  
It contains all necessary essential dataset & information needed for you to ace the challenge.



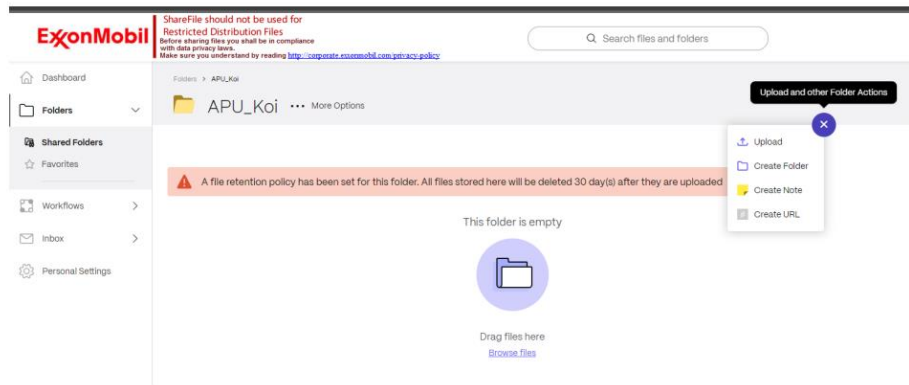
# DataWorks Challenge | ShareFile Access

## Uploading Solution for DataWorks Challenge Submission

1. Select folder with your <Team Name>. In the sample below, <Team Name> is "APU\_Koi".



2. To upload, select the purple "+" button to reveal options



3. Select "Upload" option to upload your solution submission by the deadline.

Note: Please follow the naming conventions as instructed within the "DataWorks 2024 UseCase Kit" folder's starter kit.

4. To "Delete" or perform other editing, you may right-click on the file upload to reveal the available options.
5. Once finalized submission, please email [GSC-EMIT-DATAWORKS@exxonmobil.com](mailto:GSC-EMIT-DATAWORKS@exxonmobil.com) to notify your submission.

## Note

1. In event of technical errors with the ShareFile folder access, do reach out at [GSC-EMIT-DATAWORKS@exxonmobil.com](mailto:GSC-EMIT-DATAWORKS@exxonmobil.com)

# DataWorks Challenge 2024

# Q & A





# DataWorks Challenge 2024

# Closeout



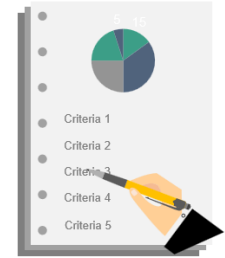
# DataWorks Challenge | The Challenge Structure

21<sup>st</sup> Sept

- **Deadline to submit solution** falls on 21<sup>st</sup> Sept 2024 (Saturday) at 11.00am (GMT +8)
- Upload your team submission to ShareFile & email [GSC-EMIT-DATAWORKS@exxonmobil.com](mailto:GSC-EMIT-DATAWORKS@exxonmobil.com) to notify your submission

1<sup>st</sup> Oct

- **Top 5 teams shortlisted will be announced** through email & invited for Finale
- Shortlisting will be conducted based on listed judging criteria by industry experts



8<sup>th</sup> & 10<sup>th</sup>  
Oct

- Top 5 teams shortlisted will be **invited for Pitching Mentorship** prior to Finale virtually
- Pitching Mentorship will be conducted by industrial experts for 2 days (2 hours / day)

15<sup>th</sup> Oct

- Top 5 teams shortlisted will be invited to **Menara ExxonMobil, KLCC for Championship pitch**
- Closing ceremony and winning teams' announcement will be conducted with the presence of ExxonMobil leaders



DataWorks  
Challenge 2024

Thank you  
&  
All the Best!

