

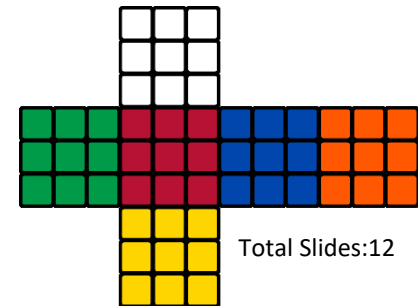
# Master Of Technology

## Spark on Commercial Cloud Vendors

Suria R Asai

( [suria@nus.edu.sg](mailto:suria@nus.edu.sg) )

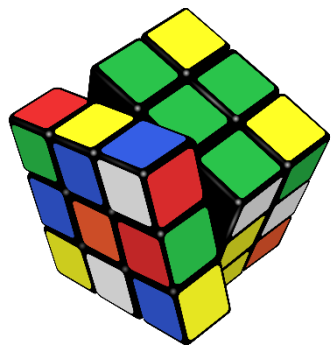
NUS-ISS



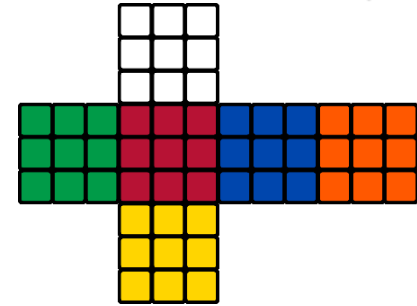
© 2016-2023 NUS. The contents contained in this document may not be reproduced in any form or by any means, without the written permission of ISS, NUS, other than for the purpose for which it has been supplied.

# Learning Objectives

- Understand and use commercial cloud offering to submit a Apache Spark (PySpark) job.
  - Google Cloud Platform
  - DataProc Offerings
  - Google Qwiklabs Skills Boost

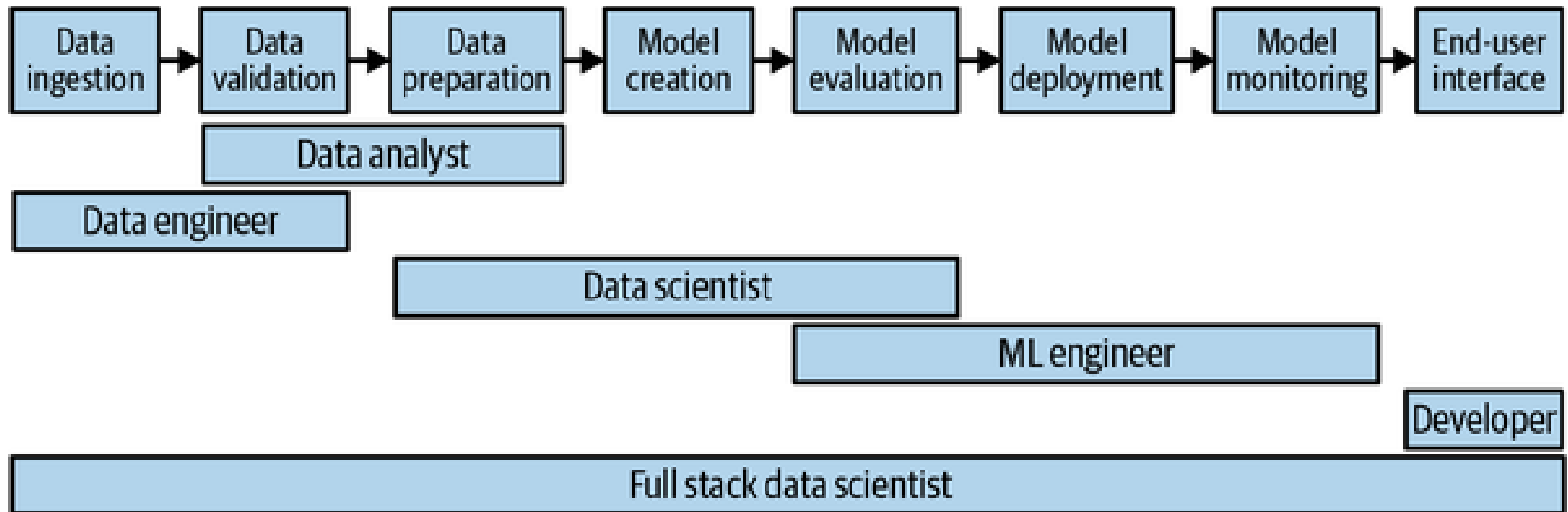


Big Data  
Engineering  
For Analytics



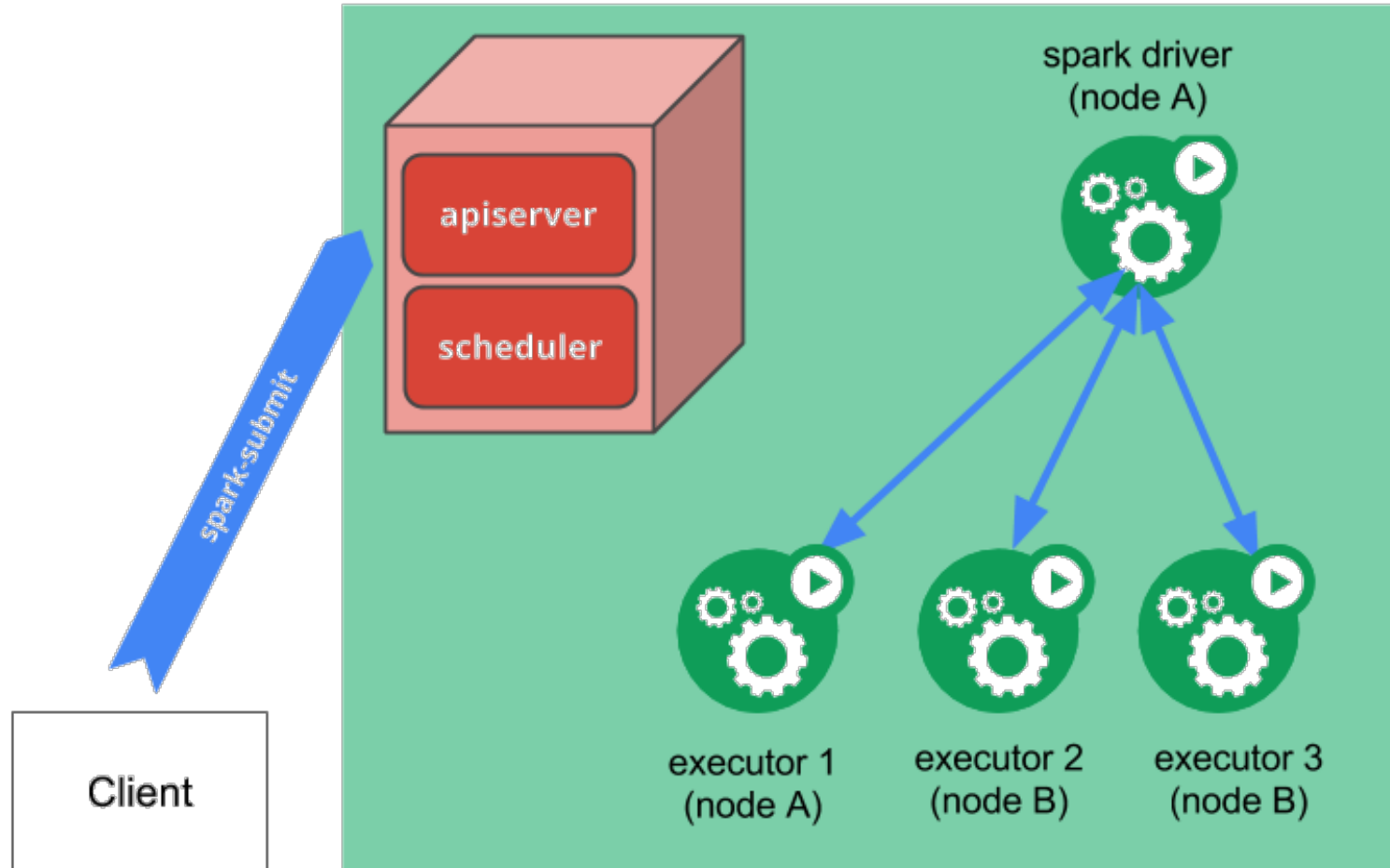
# Spark on Google Cloud Platform

# The many roles ...



# The kube Path . . .

kubernetes cluster



<https://spark.apache.org/docs/latest/running-on-kubernetes.html>

# Google Cloud Platform

- Google Cloud Platform is a suite of cloud services hosted on Google's infrastructure.
  - From computing and storage, to data analytics, machine learning, and networking, Google Cloud offers a wide variety of services and APIs that can be integrated with any cloud-computing application or project, from personal to enterprise-grade.
- To work with GCP, developer/engineer/scientist would need to familiarize with the two key tools
  - Cloud Shell
    - <https://www.cloudskillsboost.google/focuses/2794?parent=catalog>
  - Cluster Creation
    - <https://www.cloudskillsboost.google/focuses/878?parent=catalog>

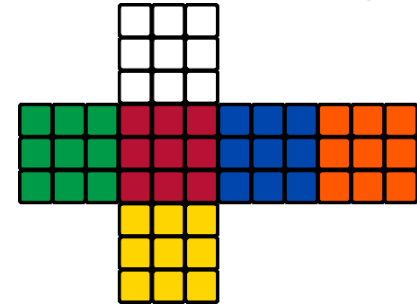
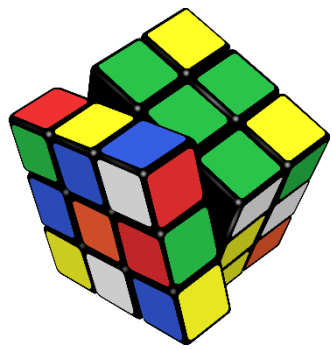
# GCP Dataproc

- Dataproc is a fully managed and highly scalable service for running Apache Hadoop, Apache Spark, Apache Flink, Presto, and 30+ open source tools and frameworks.
- Use Dataproc for data lake modernization, ETL, and secure data science, at scale, integrated with Google Cloud.
  - Introductory Video [Link](#).
  - Start by creating a cluster through the Web UI or Cloud SDK or REST APIs or with SSH access.
  - Once a cluster is provisioned you can submit the job using open source framework of your choice.
  - Scaling becomes a managed service and hence DataProc easens engineering and administrative needs to run a cluster with a pay based on use billing model

# Google Cloud Skills Boost

- Google Cloud Skills Boost is where you can access Google Cloud's entire catalog of labs and courses.
  - You can discover learning paths, build in-demand cloud skills, track your activity progress, and validate your knowledge with badges. Qwiklabs is the technology platform the labs and courses sit on.
  - This learning path contains labs that provide hands-on activities related to breakout session content that introduces you to the fundamentals of Google Cloud.
  - At times the site also contains launches announced at Google Cloud Next conferences.
  - <https://www.cloudskillsboost.google/>





# Appendix

*“Always code as if the guy who ends up maintaining your code will be a violent psychopath who knows where you live”*

— John Woods

# Google Certifications

## Which Google Cloud certification is right for you?

<p><b>Foundational certification</b></p> <p>Validate your broad knowledge of cloud concepts and the products, services, tools, features, benefits, and use cases of Google Cloud.</p> <p>RECOMMENDED CANDIDATE:</p> <ul style="list-style-type: none"> <li>✓ Collaborative role with technical professionals</li> <li>✓ No technical prerequisites</li> </ul> <p>ROLE</p> <p>Cloud Digital Leader</p>	<p><b>Associate certification</b></p> <p>Assesses the fundamental skills to deploy and maintain cloud projects.</p> <p>RECOMMENDED CANDIDATE:</p> <ul style="list-style-type: none"> <li>✓ Has understanding fundamentals of Google Cloud</li> <li>✓ Has experience deploying cloud applications and monitoring operations</li> <li>✓ Has experience managing cloud enterprise solutions</li> </ul> <p>ROLE</p> <p>Cloud Engineer</p>	<p><b>Professional certification</b></p> <p>Assess key technical job functions and advanced skills in design, implementation and management of Google Cloud products.</p> <p>RECOMMENDED CANDIDATE:</p> <ul style="list-style-type: none"> <li>✓ Has in-depth experience setting up cloud environments for an organization</li> <li>✓ Has experience deploying services and solutions based on business requirements</li> </ul> <p>ROLE</p> <p>Cloud Architect</p> <p>Cloud Database Engineer</p> <p>Cloud Developer</p> <p>Data Engineer</p> <p>Cloud DevOps Engineer</p> <p>Cloud Security Engineer</p> <p>Cloud Network Engineer</p> <p>Google Workspace Administrator</p> <p>Machine Learning Engineer</p>
---	---	--

<https://cloud.google.com/certification>

# Azure certifications

## Browse certifications by role



### Administrator

Administrators implement, monitor, and maintain Microsoft solutions.



### AI Engineer

AI engineers use Cognitive Services, Machine Learning, and Knowledge Mining to architect and implement Microsoft AI solutions.



### App Maker

App makers build apps with low-code techniques to simplify, automate, and transform business tasks and processes.



### Business User

Use Microsoft applications to move your business forward. Learn how tools like Microsoft 365, Dynamics 365, and AI can support your team an...



### Data Analyst

Data analysts make data easy to understand through designing and building data models to drive meaningful business value.



### Data Engineer

Data engineers design and implement the management, monitoring, security, and privacy of data using the full stack of data services.



### Data Scientist

Data scientists apply machine learning techniques to train, evaluate, and deploy models that solve business problems.



### Developer

Developers design, build, test, and maintain cloud solutions.



### DevOps Engineer

DevOps engineers combine people, process, and technologies to continuously deliver valuable products and services that meet end user needs...



### Functional Consultant

Functional consultants leverage Microsoft Dynamics 365 and Microsoft Power Platform to anticipate and plan for customer needs.



### Security Engineer

Security engineers implement security controls and threat protection, manage identity and access, and protect data, applications, and...



### Solution Architect

Solution architects have expertise in compute, network, storage, security.

<https://learn.microsoft.com/en-us/certifications/>

# AWS Certifications

## FOUNDATIONAL

Knowledge-based certification for foundational understanding of AWS Cloud.

**No prior experience needed.**



## PROFESSIONAL

Role-based certifications that validate advanced skills and knowledge required to design secure, optimized, and modernized applications and to automate processes on AWS.

**2 years of prior AWS Cloud experience recommended.**



## ASSOCIATE

Role-based certifications that showcase your knowledge and skills on AWS and build your credibility as an AWS Cloud professional. **Prior cloud and/or strong on-premises IT experience recommended.**



## SPECIALTY

Dive deeper and position yourself as a trusted advisor to your stakeholders and/or customers in these strategic areas. **Refer to the exam guides on the exam pages for recommended experience.**



<https://aws.amazon.com/certification/>