The Why and How of Running a Self-Managed Airflow on Kubernetes

Parnab Basak (he/him)
Senior Solution Architect and Airflow SME
Amazon Web Services

Airflow Summit
Let's flow together

September 19-21, 2023,
Toronto, Canada

Agenda

for the next 20 mins

- Available deployment options
- Why should you need to run self-managed Airflow
- How to deploy Airflow on Kubernetes using automation
- Developer & Operator experience
- Owned Responsibilities and tradeoffs
- Questions?

Available Deployment Options

On-premises

Sources on Server | Containers | Containers with Kubernetes | Virtual Machines

Personal Devices

Cloud

Amazon EC2

Amazon EKS

Amazon MWAA

Azure VMs

Azure Kubernetes Service

Azure Data Factory

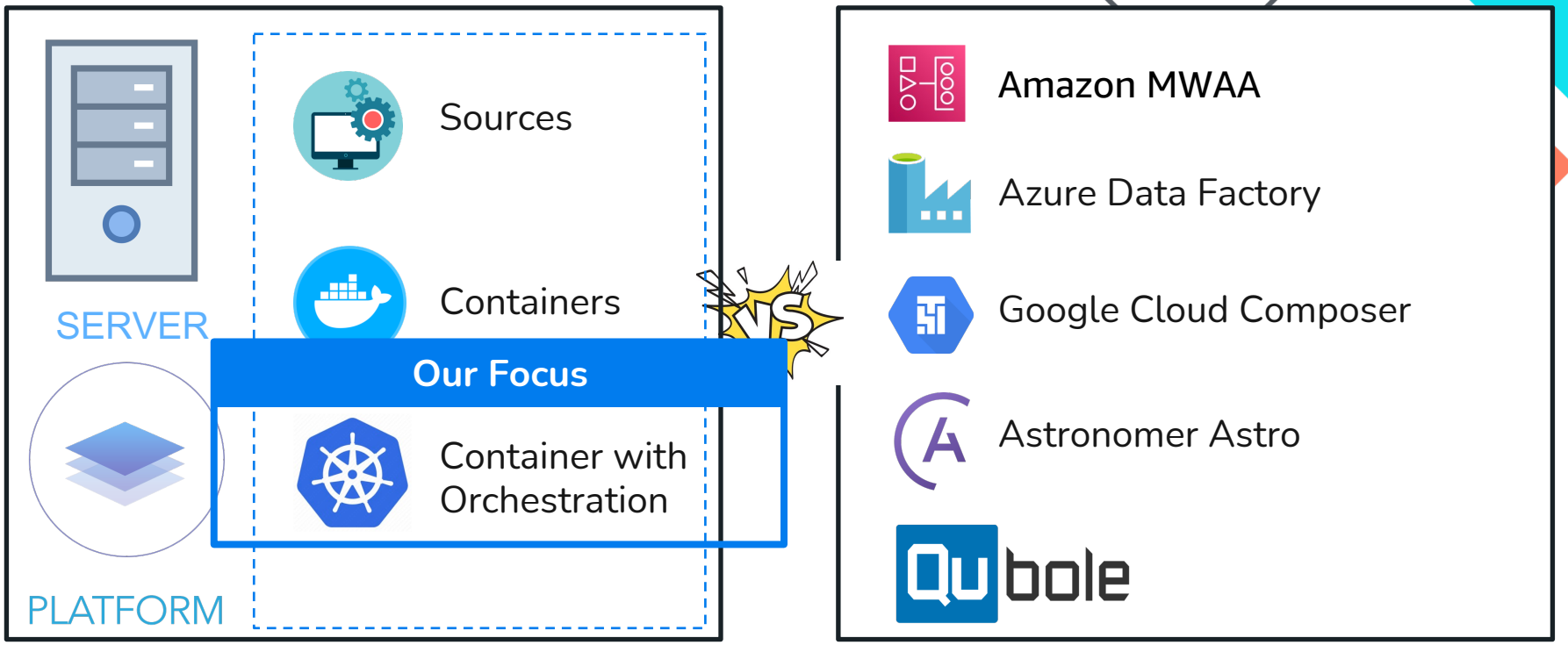
Google Compute Engine

Google Kubernetes Engine

Google Cloud Composer

....Others

Self-Managed vs Platform as a Service



Advantages with Airflow as a Service

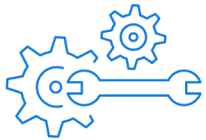
Setup

Security

Scaling

Maintenance

High Availability



Upgrades

Why Self-Managed

Reasons for Self-Managed Airflow

Flexibility and Choices

- Available Versions
- Diff Executor/Database
- Custom Code

Feature Gaps

- Parity with Open-Source
- Abstractions

Avoid Lock-ins

- On-Prem / Non-Cloud / Non-Docker
- Contracts/Agreement

Infrastructure Costs

- Re-use Existing Infra
- Savings / Discounts
- Pause / Freeze

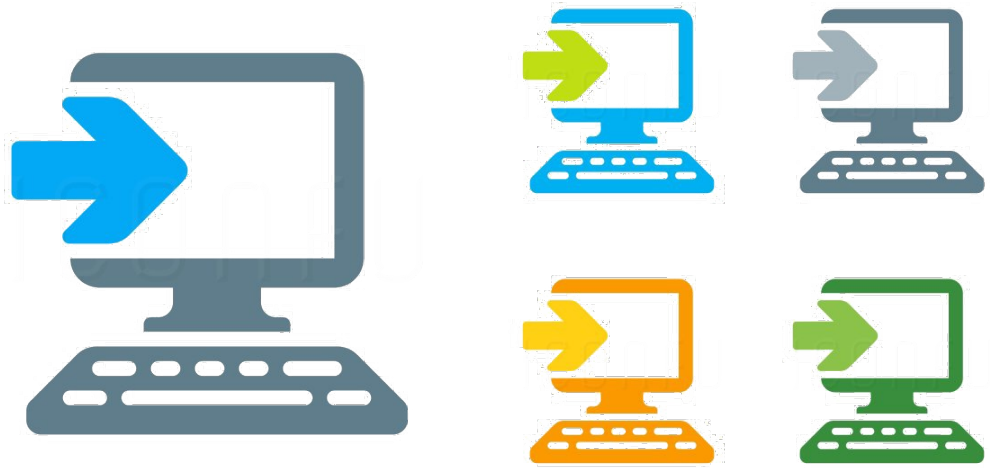
Security

- Data Sovereignty
- Internal Governance

Compliance

- FedRAMP/CJIS
- HIPAA / PCI
- SOC

How to install Self-Managed

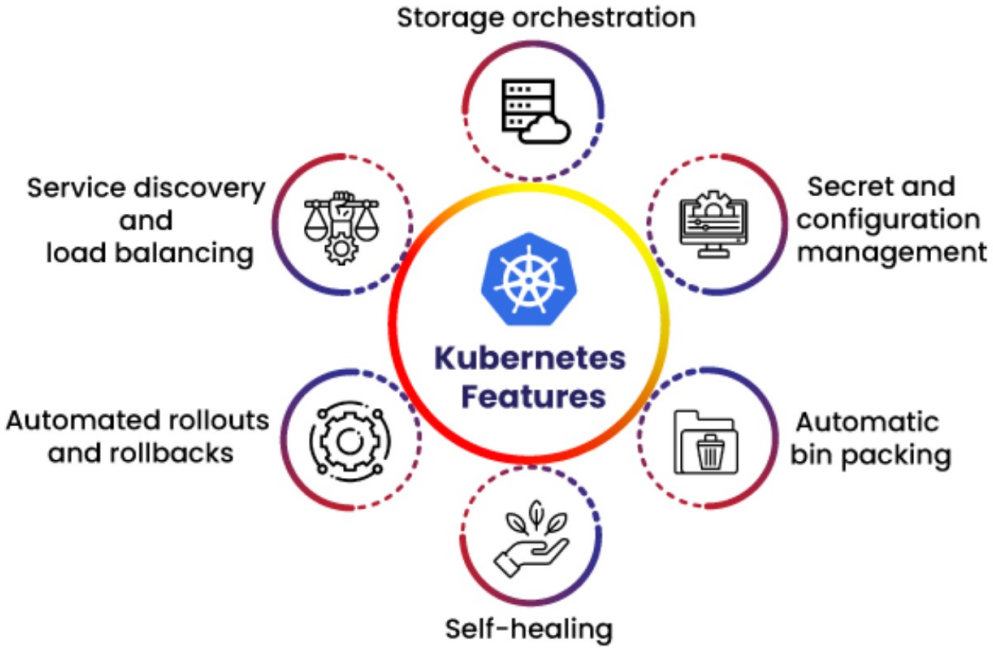


How to Self Install Apache Airflow

	Released Sources	PyPi	Docker Images	Kubernetes
How	Download Code > Build Code > Install	pip install apache-airflow package published in PyPI	Download Images > Customize > Deploy	Download & install via Helm charts
When	Build & Deploy everything	Do not want to build, or not familiar with Containers. Need custom deployment	Familiar with Docker/ Containers. Run Airflow in isolation	Familiar with Docker/ Containers & Kubernetes. Run Airflow in isolation
What to Handle	Code build and all component/dependency configuration	All component/ dependency configuration	Image build & all component, dependency config	Deployment with defaults via Helm charts.
Effort	★★★★	★★★	★★	★

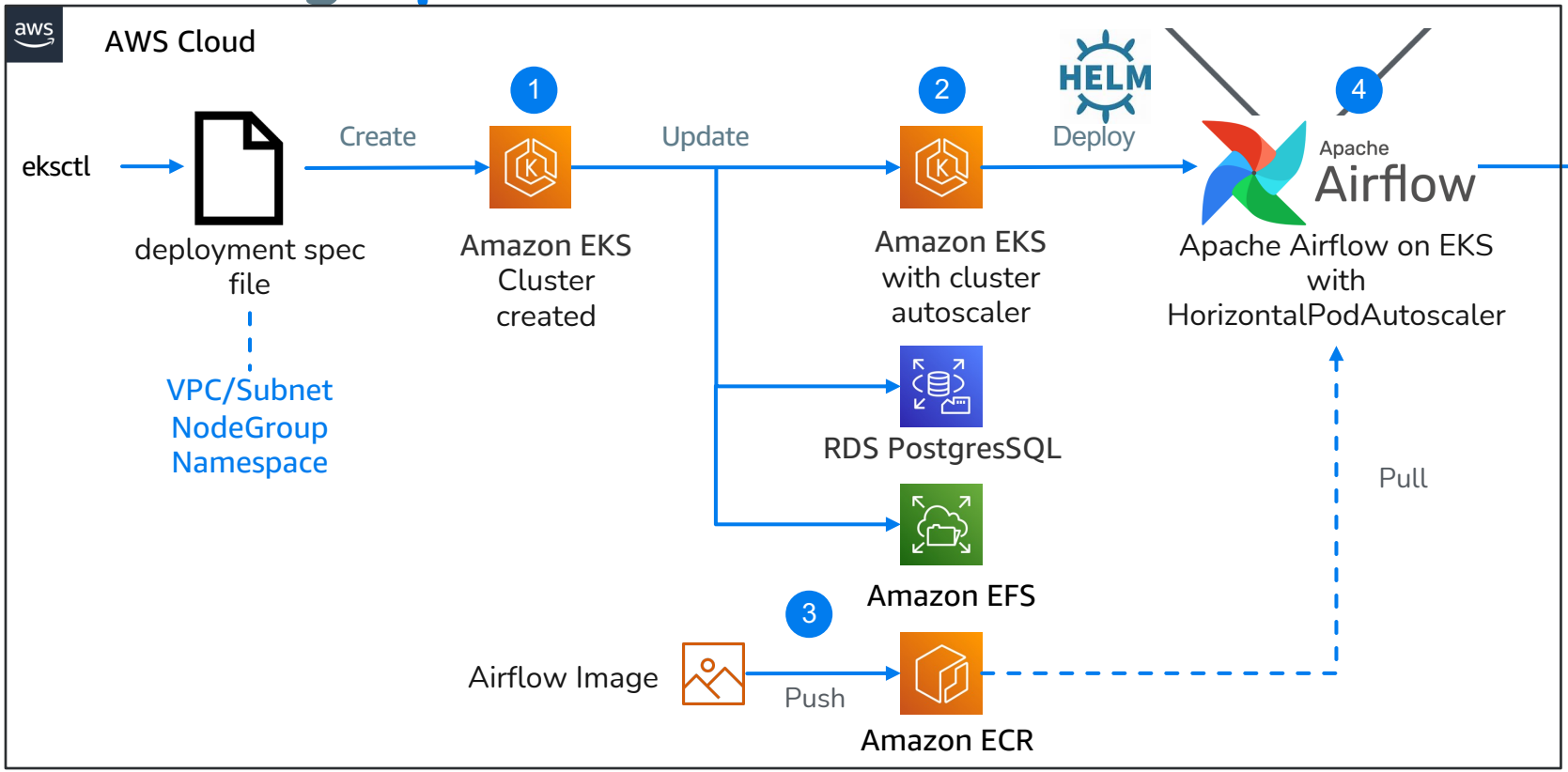
Why Kubernetes

- Open Source
- Rich Eco System
- Faster Time to Market

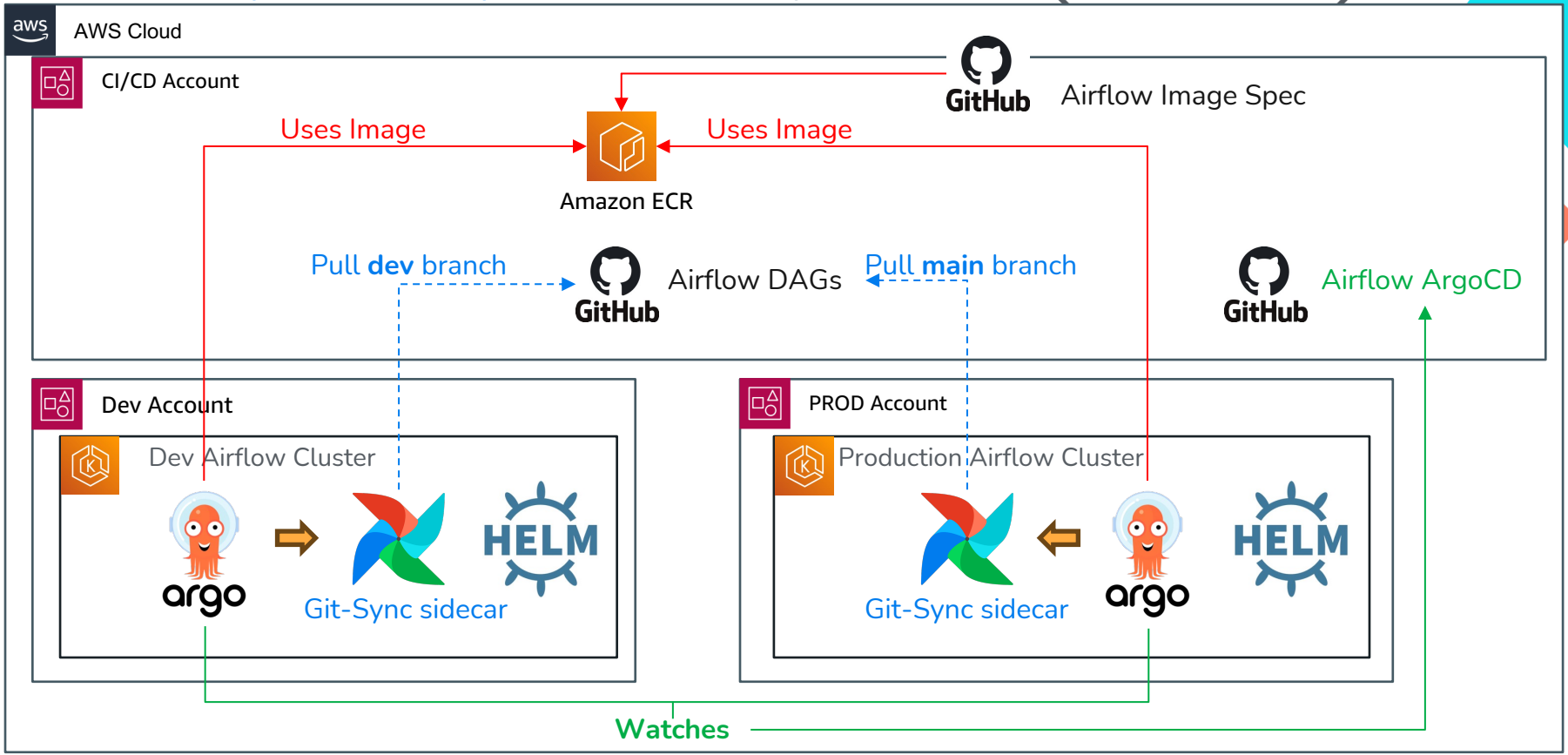


https://www.bestdevops.com/easy-way-to-know-kubernetes/

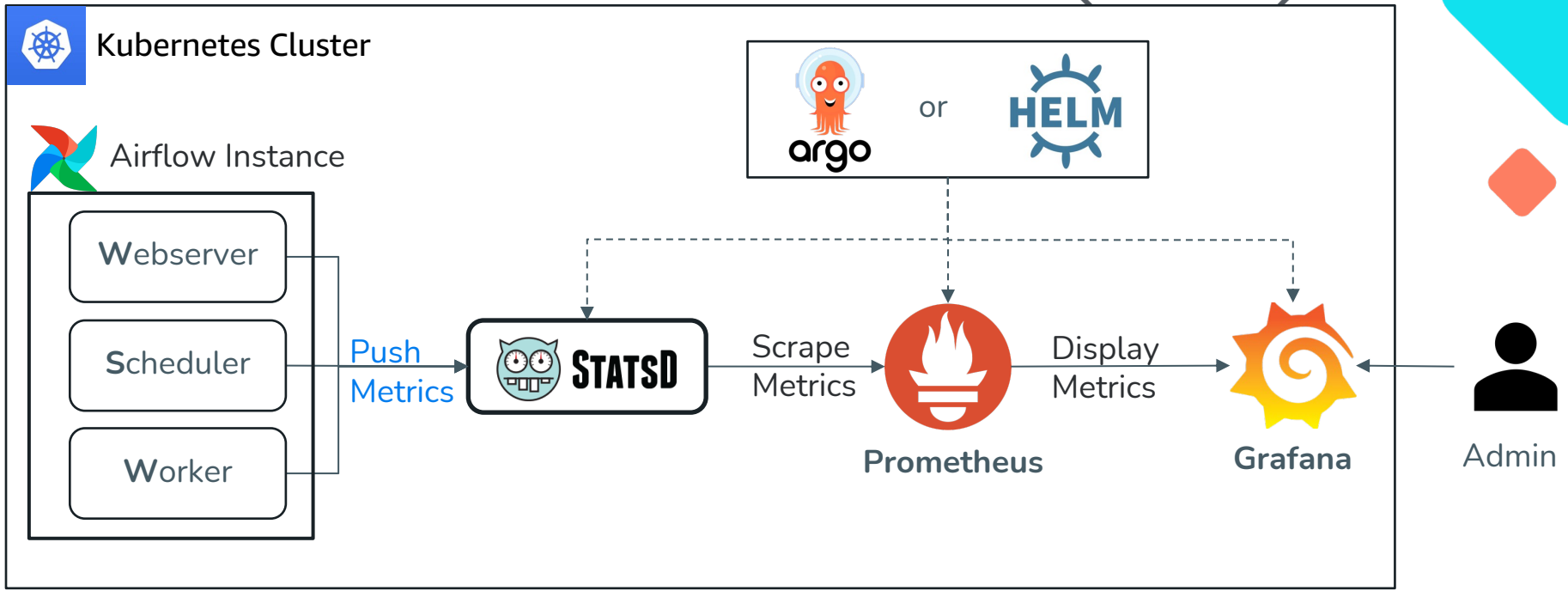
Installing Apache Airflow on Amazon EKS - Helm



Developer & Operator Experience



Operator Observability



Self-Managed Responsibility

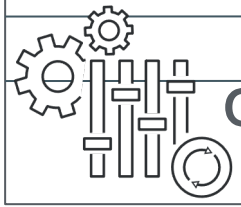
Knowledge

- Apache Airflow
- Kubernetes
- Networking

+

Responsibility

- Setup, Upgrade, Maintenance
- AuthZ and AuthN,
- High Availability,
- Autoscaling, CI/CD
- Logging,
- Observability



...Just to name a few

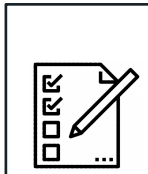
=

Gain

Airflow with **Custom**

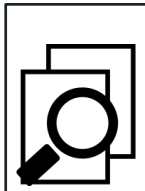
Capabilities

Key Takeaways



Key Requirements

- List and Prioritize
- Must Haves
- Nice to Haves

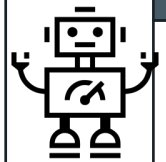


Evaluate MSP

- Value Propositions
- Features Supported
- Roadmap

Decide Hosting

- Where
- How
- Why
- Owned Responsibilities



Automation

- What / How (Re-use)
- Scale
- Maintainability

Session Survey

https://www.pulse.aws/survey/STI90ZJG

- SIX Questions = Less than a Minute
- Completely Anonymous

- 1 star – least favorable
- 5 star – most favorable

Questions?

https://www.linkedin.com/in/parnab-basak/