



StatCom 7 – Private-Sector Experience with Big Data in Africa: Use Cases, Potential and Challenges

Addis Ababa, Ethiopia (Virtual) – 13-15 Oct 2020

Hisham Arafat Shehata, *B.Sc. Engineering, M.Sc. Informatics*

Digital Transformation Consulting Director, EMEA, innovay.io

Industry Practices Director, Data Science, AI and Big Data, Nile University

Lecturer & Research Projects Supervisor, Professional Masters, Cairo University

Solutions Architected, Technology Strategist and Research Engineer



Our Agenda

- Big Data in Statistics Context
- Data in the Big Data – DaaS
- Private Sector Use Cases in Africa
- Potential Applications
- Challenges and Solutions
- Q & A – Discussion



My Background & Experience

24 Years

Mentored
40+ Digital
Innovation
Startups

Professor of
Practice in
Data Science,
Big Data and AI
Applications

Cross 51+
Countries

Mentored
100s of Data
Professionals

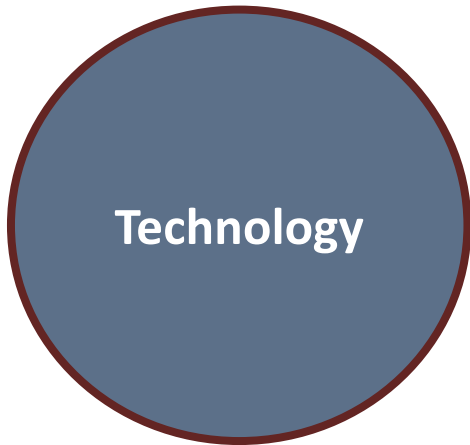
Implementation
Experiences - tens
of real-life
Enterprise
Customer Projects

Private Sector: Big Data in Statistics Context...



Image source: The Enterprise Project

BIG DATA & AI LANDSCAPE 2018



INFRASTRUCTURE

HADOOP ON-PREMISE
 cloudera Hortonworks
 MAPR Pivotal
 IBM InfoSphere
 bluedata jethro

HADOOP IN THE CLOUD
 AWS Microsoft Azure
 Google Cloud IBM InfoSphere
 Treasure Data Databricks Altilscale
 CAZENA CenturyLink

STREAMING / IN-MEMORY
 AWS Databricks Strim
 Confluent GridGain
 dataArtisans Hazelcast
 TERRACOTTA Ixq
 Wallarooops

ANALYTICS

DATA ANALYST PLATFORMS
 Microsoft Pentaho Alteryx
 DataRobot GUAVUS AYASDI
 ATTIVO Datameer Quid Incoia
 Interona ClearStory Origami
 ENDOR MOORE Bortolozzi

DATA SCIENCE PLATFORMS
 IBM KNIME dataiku
 Domino Rapidminer
 Continuum Algorhythmia
 DataRobot ANGILOB SAS

APPLICATIONS - ENTERPRISE

SALES
 InsideSales.com Chorus
 Conversica IBM
 Clari Avaso tact.ai
 FuseMachines Troops

MARKETING - B2B
 Radius App Annie
 Eversting Lattice
 Hentigo Sense
 Tubular DataRobot
 Engagio Mype

MARKETING - B2C
 Zeta Bloomreach SendGrid
 BlueYonder (Persado) Eloqua
 ActionIQ Sailthru Bluecore
 Quantifox Imparticle Ampero
 Emperity Streamline

CUSTOMER SERVICE
 Medallia Zendesk
 Clarabridge
 Gainsight NO DATA
 DigitalGenius Ahni
 Automatt Frame AI
 Omega Intercom CallDesk

NoSQL DATABASES
 Google Cloud AWS
 Oracle Microsoft Azure
 MongoDB MarkLogic
 Aerospike DataStax
 AmazonDB Couchbase
 Redis Scylla

NewSQL DATABASES
 SAP Clustring Pivotal
 Cockroach Labs
 MemSQL InfluxDB
 Yotpo
 Citidata Splice
 Paragim

GRAPH DBs
 Amazon Neptune
 IBM Oracle
 Neo4j
 Amazon Neptune
 IBM Oracle

MPP DBs
 Vertica
 IBM Data Warehouse Systems
 Google Cloud
 Microsoft Azure
 Pivotal
 Snowflake

CLOUD EDW
 AWS

BI PLATFORMS
 Microsoft AWS
 SAP
 Tableau
 Qlik
 ZEPH
 MicroStrategy

VISUALIZATION
 Tableau SAP
 Google Cloud
 Qlik
 ZEPH
 CHARTIO

MACHINE LEARNING
 AWS
 Google Cloud
 DataRobot
 H2O
 Element
 Vizeer
 Gamolon

HUMAN CAPITAL
 Entelo
 HiQ
 iCIMS
 Mya

LEGAL
 Ravel
 Bluebeam
 Judicata
 RASS

FINANCE
 Anaplan
 Zuora
 TradeShift

ENTERPRISE PRODUCTIVITY
 Slack
 Oracle
 Luma
 Clara
 Butter

BACK OFFICE AUTOMATION
 UiPath
 Blueprints
 Bluebeam
 WorkFusion

SECURITY
 Cylance
 Zscaler
 StackPath
 Anomali
 Vectra
 Datavisor
 Avast
 Symantec
 McAfee
 Trend Micro
 Sophos
 Avira
 Avast
 Avast
 Avast

DATA TRANSFORMATION
 Talend Pentaho
 Alteryx
 Informatica
 SAP Data Services
 IBM
 Informatica
 SAP Data Services

DATA INTEGRATION
 SAP Data Services
 Informatica
 IBM
 Informatica
 SAP Data Services

DATA GOVERNANCE
 IBM
 Collibra
 Alation
 IBM
 Collibra
 Alation

MGMT / MONITORING
 AWS New Relic
 Rubrik
 Splunk
 Dynatrace
 SignalFx
 PagerDuty
 New Relic
 Rubrik
 Splunk
 Dynatrace
 SignalFx
 PagerDuty

COMPUTER VISION
 Microsoft Azure
 Amazon Rekognition
 Clarifai
 EverAI
 Microsoft Azure
 Amazon Rekognition
 Clarifai
 EverAI

HORIZONTAL AI
 IBM Watson Cortana
 Microsoft
 IBM Watson Cortana
 Microsoft

SPEECH & NLP
 Google Cloud
 Amazon
 Microsoft
 Google Cloud
 Amazon
 Microsoft

APPLICATIONS - INDUSTRY

ADVERTISING
 AppNexus
 Oracle
 Rubicon
 DoubleClick
 Rubicon
 DoubleClick

EDUCATION
 Blackboard
 FutureLearn
 FutureLearn
 Blackboard
 FutureLearn

GOVERNMENT
 EnergyMark43
 GridSmart
 GridSmart
 EnergyMark43
 GridSmart

REAL ESTATE
 Redfin
 Opencor
 VTS
 Redfin
 Opencor
 VTS

FINANCE - LENDING
 OnDeck
 Affirm
 Lendix
 OnDeck
 Affirm
 Lendix

INSURANCE
 Lemonade
 Cvent
 Lemonade
 Cvent

STORAGE
 AWS
 Microsoft Azure
 Pure Storage
 EMC
 AWS
 Microsoft Azure
 Pure Storage
 EMC

CLUSTER SVCS
 AWS
 Docker
 Mesos
 AWS
 Docker
 Mesos

APP DEV
 AWS
 Heroku
 AWS
 Heroku

CROWD-SOURCING
 Amazon Mechanical Turk
 Upwork
 Amazon Mechanical Turk
 Upwork

HARDWARE
 Google TPU
 ARM
 Intel
 Google TPU
 ARM
 Intel

GPU DBs
 KineticDB
 KineticDB
 KineticDB

SEARCH
 Oracle
 Elasticsearch
 Oracle
 Elasticsearch

LOG ANALYTICS
 Splunk
 Sumologic
 Loggly
 Splunk
 Sumologic
 Loggly

SOCIAL ANALYTICS
 Hootsuite
 Sprinklr
 Synthesio
 Hootsuite
 Sprinklr
 Synthesio

WEB / MOBILE / COMMERCE ANALYTICS
 Google Analytics
 Mixpanel
 Sumo
 Google Analytics
 Mixpanel
 Sumo

HEALTHCARE
 Flatiron
 Clover
 Flatiron
 Clover

LIFE SCIENCES
 Benchling
 Verily
 Benchling
 Verily

TRANSPORTATION
 Uber
 Tesla
 Uber
 Tesla

AGRICULTURE
 FarmLogs
 FarmLogs
 FarmLogs

COMMERCE
 Stitch Fix
 Stitch Fix
 Stitch Fix

INDUSTRIAL
 AVEVA
 Siemens
 AVEVA
 Siemens

CROSS-INFRASTRUCTURE/ANALYTICS

AWS Google Cloud Microsoft IBM SAP SAS 1010DATA VMware TIBCO TERADATA ORACLE NetApp Synopsys MAPR cloudera

OPEN SOURCE

FRAMEWORK
 TensorFlow PyTorch Keras
 TensorFlow PyTorch Keras

QUERY / DATA FLOW
 Apache Spark Presto
 Apache Spark Presto

DATA ACCESS
 Apache Cassandra MongoDB
 Apache Cassandra MongoDB

COORDINATION
 Apache ZooKeeper
 Apache ZooKeeper

STREAMING
 Apache Flink Apache Storm
 Apache Flink Apache Storm

STAT TOOLS
 Apache Mahout
 Apache Mahout

AI / MACHINE LEARNING / DEEP LEARNING
 TensorFlow Theano
 TensorFlow Theano

SEARCH
 Elasticsearch Solr
 Elasticsearch Solr

LOGGING & MONITORING
 Prometheus Kibana
 Prometheus Kibana

VISUALIZATION
 Tableau
 Tableau

COLLABORATION
 Jupyter
 Jupyter

SECURITY
 Apache Ranger
 Apache Ranger

DATA SOURCES & APIs

HEALTH
 Apple
 Validic
 Fitbit
 Garmin
 Kinso

IOT
 GE Digital
 Uptake
 Helium
 Somarva

FINANCIAL & ECONOMIC DATA
 Bloomberg Thomson Reuters Dow Jones
 S&P Capital IQ
 Premise
 Xignite
 Quandl
 Plaid
 Earnest

AIR / SPACE / SEA
 OpenSky
 Airware
 Spire
 Windward
 DroneDeploy

PEOPLE / ENTITIES
 Acxiom Experian
 Epsilon
 Basis
 Quantcast
 SafeGraph

LOCATION INTELLIGENCE
 Foursquare
 Sense360
 Esri
 Factual
 Cuebio
 Radar

OTHER
 Qualtrics
 Data.gov
 Enigma
 CRUX

DATA SERVICES
 Palantir
 Cerner
 Fractal
 Kagglo
 DataKind
 Wipacius

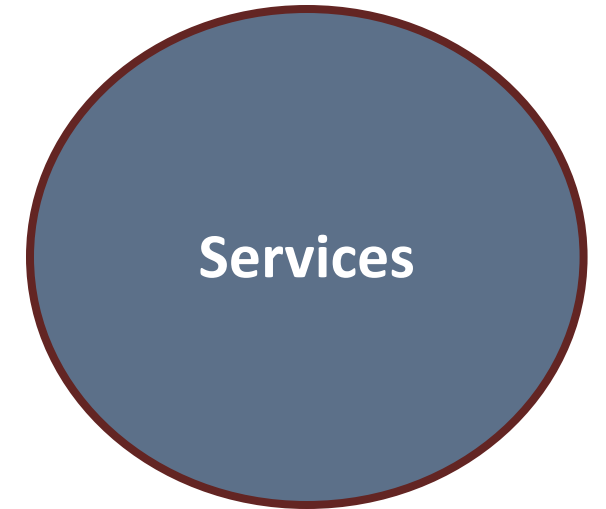
INCUBATORS & SCHOOLS
 Pluralsight
 DataCamp
 DataEpic
 Galvanize
 The Data Incubator

RESEARCH
 Facebook Research
 OpenAI
 MIRI
 Vector Institute
 A2

Private Sector: Big Data Services

Consultancy, Design, Development, Implementation,
Hosting, Cloud, SW/HW, Support and Operations
Providers:

- Dell EMC, Cloudera, Oracle, SAP...
- Accenture, McKinsey and Deloitte...
- AWS, GCP, Azure...
- Apache, Linux Foundation...
- Local and Regional Players...



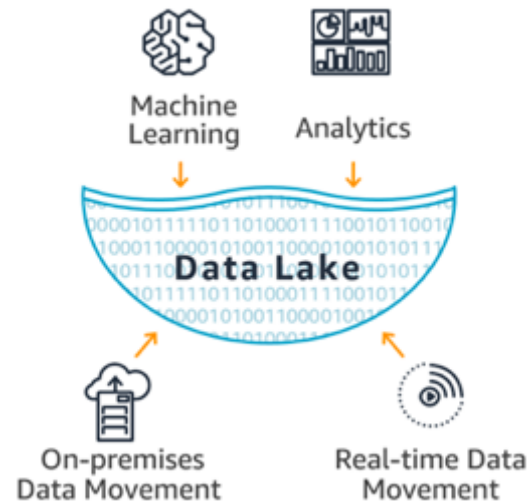


Fading Off Hype...Focus on Data

Statistics Exploit Data

The Data in the Big Data – DaaS

- Sourcing Quality Big Data
- Storing and Managing Big Data
- Processing of Big Data
- Consuming Big Data: Data-as-a-Service



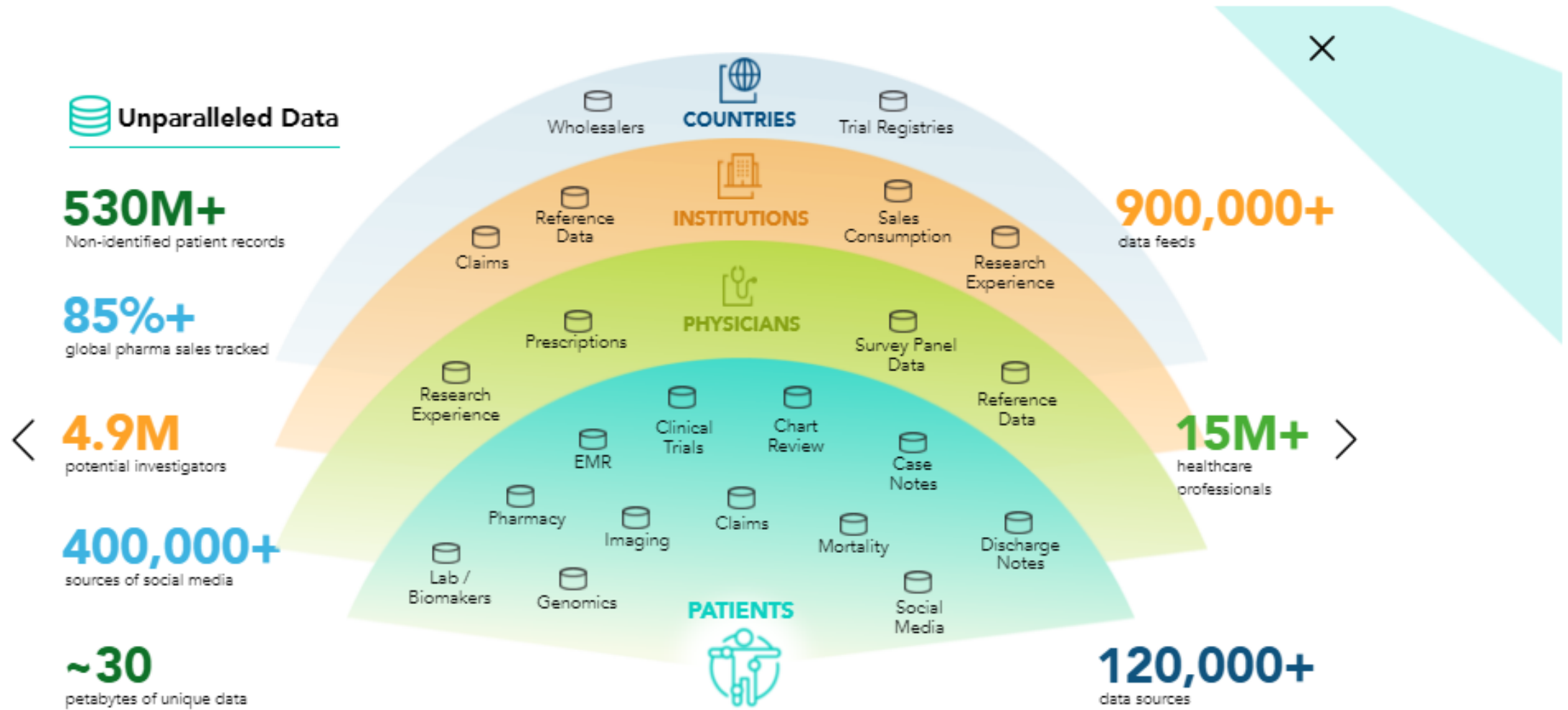
The Data in
the Big Data

Use Case #1: IQVIA

- Approach: Collection and Aggregation
- Vertical Industry: Healthcare
- Global entity with operation in Africa
- Business Model: Reports and Insights
- Examples:
 - Pharmaceutical Market - Quarterly Insight Report
 - Understanding Patient Behavior for Improved Experience
 - The Medical Technology Industry Post-COVID-19
- <https://www.iqvia.com/locations/middle-east-and-africa>



Closer Look into IQVIA...The CORE



Use Case #2: Africa Geoportal by esri

- Approach: Maps Satellite, and Open Data
- Vertical Industry: Cross Industry
- Focused on Africa with global support
- Business Model: Maps and Spatial Analytics
- Examples:
 - COVID-19 Resources for Africa are now available
 - The Horn of Africa is facing the worst desert locust crisis in over 25 years
 - Urban Africa: a rapid migration from rural areas to expanding cities.
- <https://www.africageoportal.com/>

Africa Geoportal Sources



Senegal



Cote d'Ivoire



Ethiopia



Rwanda



Digital Earth
AFRICA

Digital Earth Africa



NASA Earth Science



Use Case #3: Positium



- Focus on Mobile Networks Data for Statistics

Africa is revolutionising mobile big data, led by the UN Statistics Division

The United Nations Statistics Division (UNSD) hosted the 5th International Conference held in Kigali, Rwanda. At the event, the potential for using innovative mobile positioning data sources in Africa was showcased, reflecting on projects that are currently going on around the world and on how businesses and governments can benefit from it.

After five days of interactive workshops, presentations from worldwide big data professionals along with expert panel discussions, Siim Esko, our Head of International Sales, talked exclusively to Ronald Jansen, Chief of the UNSD's Data Innovation and Capacity Branch.



Use Case #4: AfricArXiv

- Focus on Federating and Opening Scientific Content pan Africa.





Potential and Challenges

The Data in the Big Data – Sourcing

- Official Statistics (NSOs)
- Satellite Images & Geological
- Maps and GIS Applications
- Gov & NGO Organizations
- Telemetry and IoT Deployments
- Telecom Operators, Banks, Retailers
- Data Collector and Aggregators
- Social Networks and Platforms
- Content Networks and Open Data



The Data in the Big Data – Magmananet

- Hardware and Software
- Cloud Services
- Platform Technologies
- Consultancy
- Implementation
- Operationalizations
- Support and Sustainability



The Data in the Big Data – Commercialization

- Legalism and Agreements
- Data Privacy
- National Security
- Political Climate
- Sharing Cost and Risk
- Pricing and Profit Sharing
- Ownership and Management
- Simply for Private Sector: What's In It For Me?





Additional Slides



The Practice... Big Data, Data Science and AI Apps

The Practice of Data Science: Approach, Lifecycle, Use Cases, Types of Data, Data-Driven Models

Tools, Packages & Libs, Features/Dimensionality Reduction, Sampling, Model Evaluation, Model Scoring, ROI Estimation

Preparing Data	Statistical Modeling	Mining Methods	Classification Techniques	Deep Learning	Network Analysis	Online Learning
<ul style="list-style-type: none"> • Exploration • Data Quality • Visualization • EDA • Dims / Features Engineering 	<ul style="list-style-type: none"> • Inference • Linear Reg. • Logistics Reg. • Multinomial • Time Series 	<ul style="list-style-type: none"> • Association Rules • Clustering • Text Mining • NLP • Recommenders 	<ul style="list-style-type: none"> • LS / KNN • Naïve Bayes • Decisions Trees • SVM • Ensemble Learning 	<ul style="list-style-type: none"> • Perceptron • Art. Neural Networks • CNN • DBN 	<ul style="list-style-type: none"> • Graphs • Big Graph Theory • Social Network Analysis (SNA) 	<ul style="list-style-type: none"> • Streams • Stochastic Gradient Descent (SGD) • Reservoir Sampling

Programming Languages, Tools, Packages/Libs, Big Data Platforms, Ingestion, MPP, Agile/DevOps

Implementing and Deploying Analytics Models on Big Data Platforms - Applications

Digital Transformation...Key Enablers

- Big Data Analytics...Provides Unprecedented Powerful Data Processing Capabilities
- Data Science...The Innov., Models, Tools, Techniques, Methods to Unleash Value
- Blockchain & DLT...Backbone for Enabling Trusted Digital Assets and Immutable Data
- Bots, RPA & AI...Innovative Approach to Develop Automations and Cognitive Services
- Internet of Things (IoT)...Getting People, Machines and Systems Connected Online
- Gamification, VR & AR...Interactive, Motivational and Progressive Engagement
- Cloud-Native...Methodologies, Architecture, Platforms and Tools for Enterprise Apps





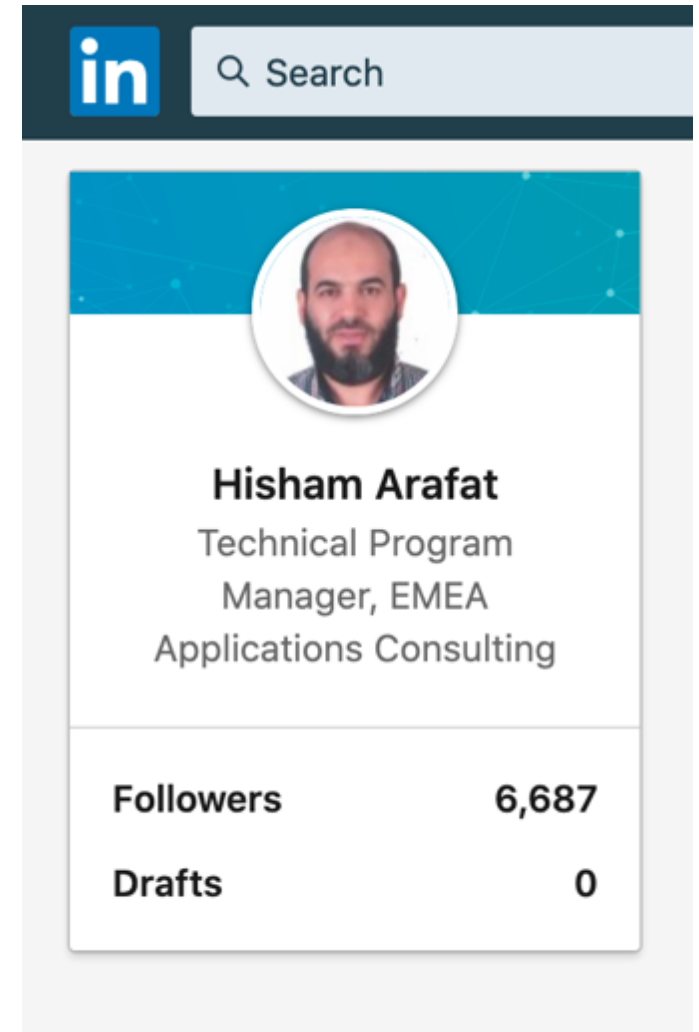
Q & A – Discussion?

Welcome to Connect and Reach Out...

Got Inquiries, Comments, Ideas?

Email: hiarafat@Hotmail.com

LinkedIn: <https://eg.linkedin.com/in/hisham-arafat-a7a69230>





Thank You

innovay.io