



coc coc

/ WEBINAR

Learn more about

DATA SCIENTIST ANALYST ENGINEER



cốc cốc

/ WEBINAR

Meet our speakers!



TRỊNH CÔNG TÂN

DATA ANALYST TEAM LEADER at CÓC CÓC

- 8 years working with *data analytics*
- Build up *data and reporting systems* in IT companies with large data scale



NGUYỄN ĐỊNH HÙNG

SENIOR SOFTWARE ENGINEER at CÓC CÓC

- 10+ years working with *large-scale data and backend systems*
- Building *reliable, high-performance systems*



ĐỖ VIỆT PHƯƠNG

DATA SCIENTIST TEAM LEADER at CÓC CÓC

- 10+ years working with *large-scale analysis and machine learning*
- Contribute directly *predictive modeling* to the company productions

01

Roles in Business Intelligence

02

Learn more about Data Analyst + QA

03

Learn more about Data Engineer + QA

04

Learn more about Data Scientist + QA

05

Minigame





TRỊNH CÔNG TÂN

DATA ANALYST TEAM LEADER at CỐC CỐC

- **8 years** working with *data analytics*
- Build up *data and reporting systems* in IT companies with large data scale



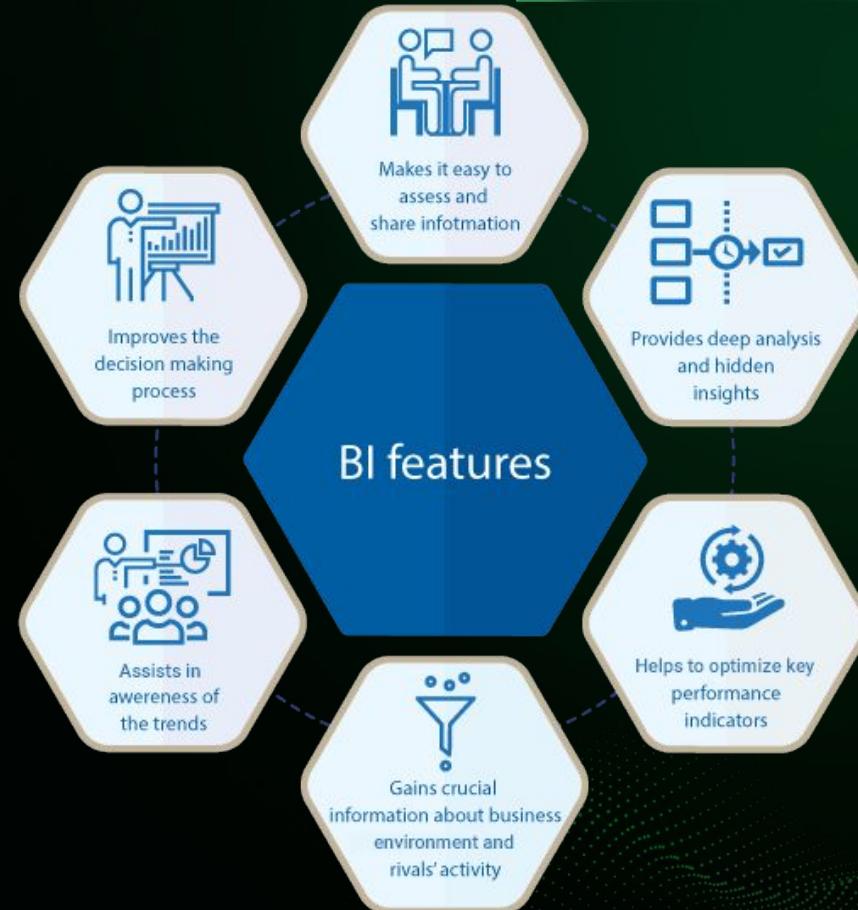
coc coc

WEBINAR

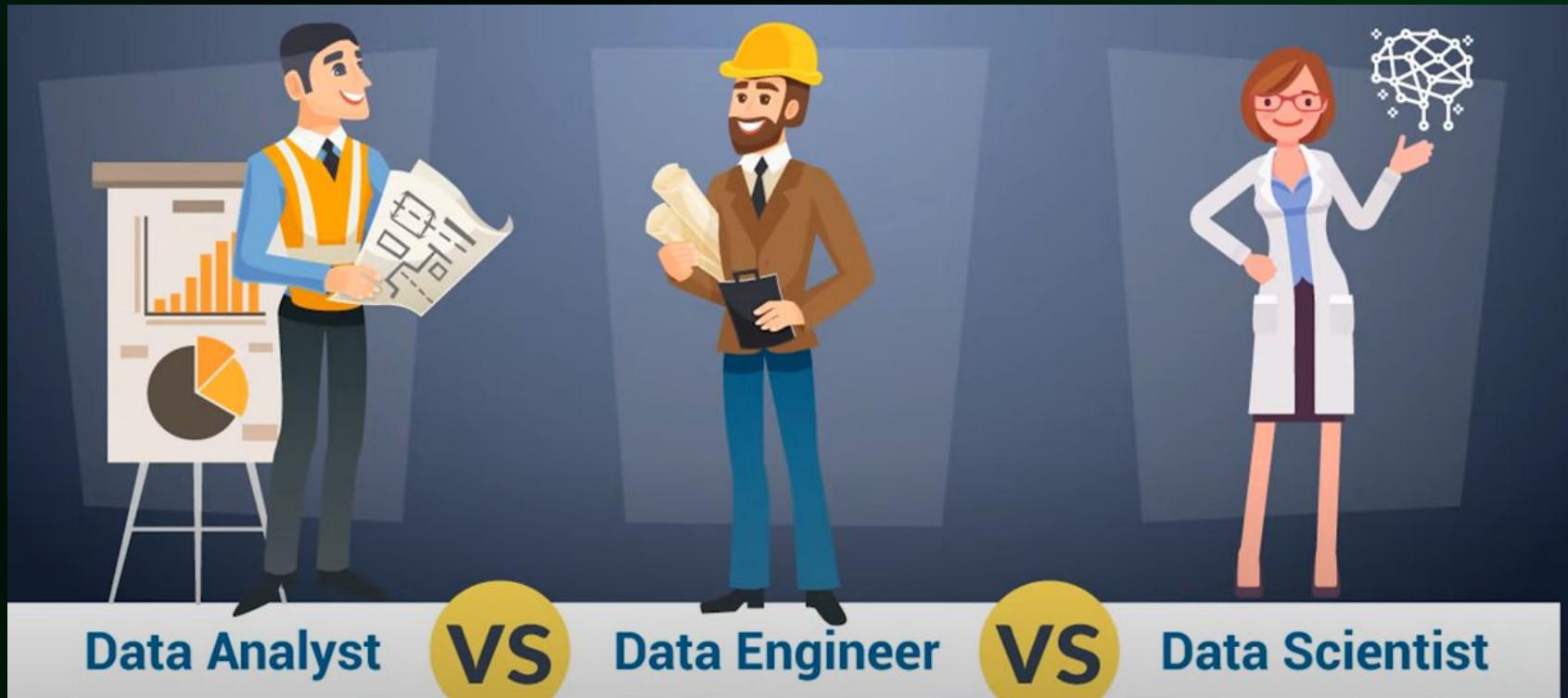
01

ROLES IN BUSINESS INTELLIGENCE

BENEFITS OF BUSINESS INTELLIGENCE



ROLES IN BI





coc coc / WEBINAR

02

Learn more about
**DATA
ANALYST**

01

What is Data Analytics?

02

How to become a Data Analyst?

03

Skill set of Data Analyst

04

Career Path

05

Challenges & Lessons Learned

06

Q&A



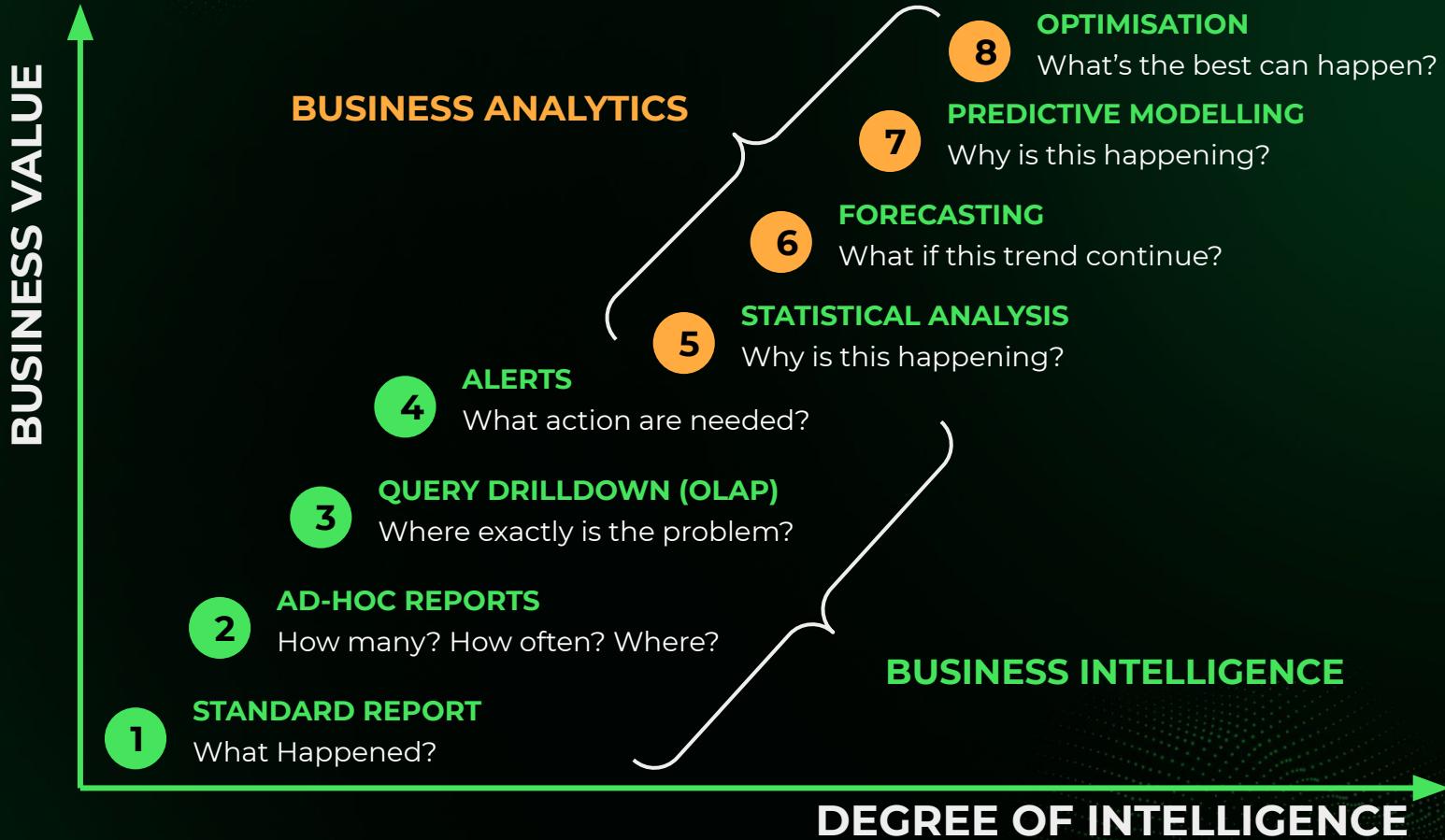
1

WHAT IS DATA ANALYTICS?

Data analytics is the **science of analyzing raw data** to **make conclusions** about that information, then, help a business **optimize** its performance.



```
graph LR; A[RAW DATA] --> B[SCIENCE OF  
ANALYZING]; B --> C[OPTIMIZE  
PERFORMANCE]
```



SOME PROBLEMS TO SOLVE IN CỐC CỐC

1. How to organize reports/dashboards for easy usage of other team?
2. How to define meaningful metrics?
3. Why do users use browser more active in a specific period?
4. What features do user use most frequently? How to improve them?
5. How does a new feature impact to user behavior?
6. Why do user churn?
7. Which groups of user retains most?
8. How to trigger aha moment to increase retention?

*“If you want to know widely, go data analyst
If you want to go deeply, go data scientist...”*

2

HOW TO BECOME DATA ANALYST?

- Learn what market needs: Check Job Description ([Demo](#))
- Learn from other data analysts: Research CV/Linkedin

Data Analyst Lead (BI, SQL)

[Apply Now](#)

Job Description

- Lead for building standardized data model fit for operational, analytical, forecast and support decision making
- Builds analytical capabilities and manages modeling, forecasting, reporting and/or data distribution tools and processes
- Establish and promote Data Ops Best Practices in collaboration with Application and Tech teams
- Lead new data/analytics capabilities and tool evaluations
- Partner with business teams to further the overall mission of modernizing data capabilities and meet line of business needs
- Mentor existing team members and provide training and career growth opportunities
- Recruit and retain industry leading talent for a growing organization

Your Skills and Experience

Knowledge & experience:

- Retail industry experience with Retail operation, Marketing, Supply chain, Merchandise, Financial
- Have experience on analysts transform data into insights that drive business value
- Have experience of data analytics, data visualization and data modeling techniques and technologies
- Have experience to identify trends that can help other departments, managers and executives make business decisions to modernize and improve processes in the organization.

CENTRAL RETAIL

Central Retail

To be the best, sustainable, and most admired customer-centric omni-channel retailer

Product

Monday - Friday

No OT

1000+

Thailand

[View our company page](#)



coc coc

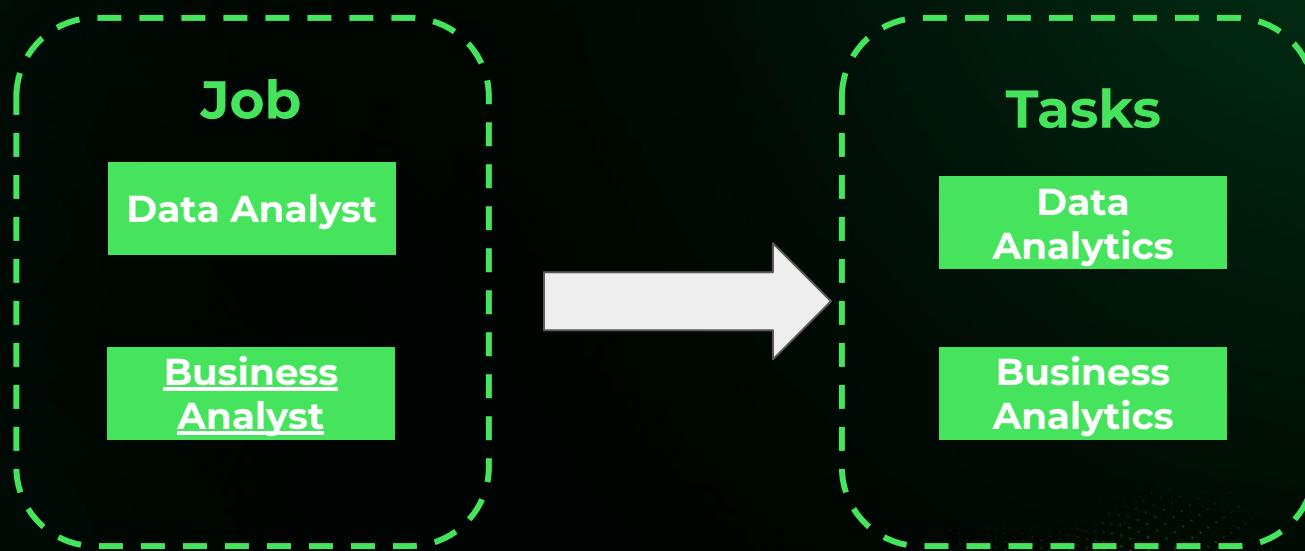
/ WEBINAR

2

HOW TO BECOME DATA ANALYST?

- Learn what market needs: Check Job Description ([Demo](#))
- Learn from other data analysts: Research CV/Linkedin

Some confusing terms:



3

SKILL SET OF DATA ANALYST

RAW DATA



DATA
ANALYST



BUSINESS

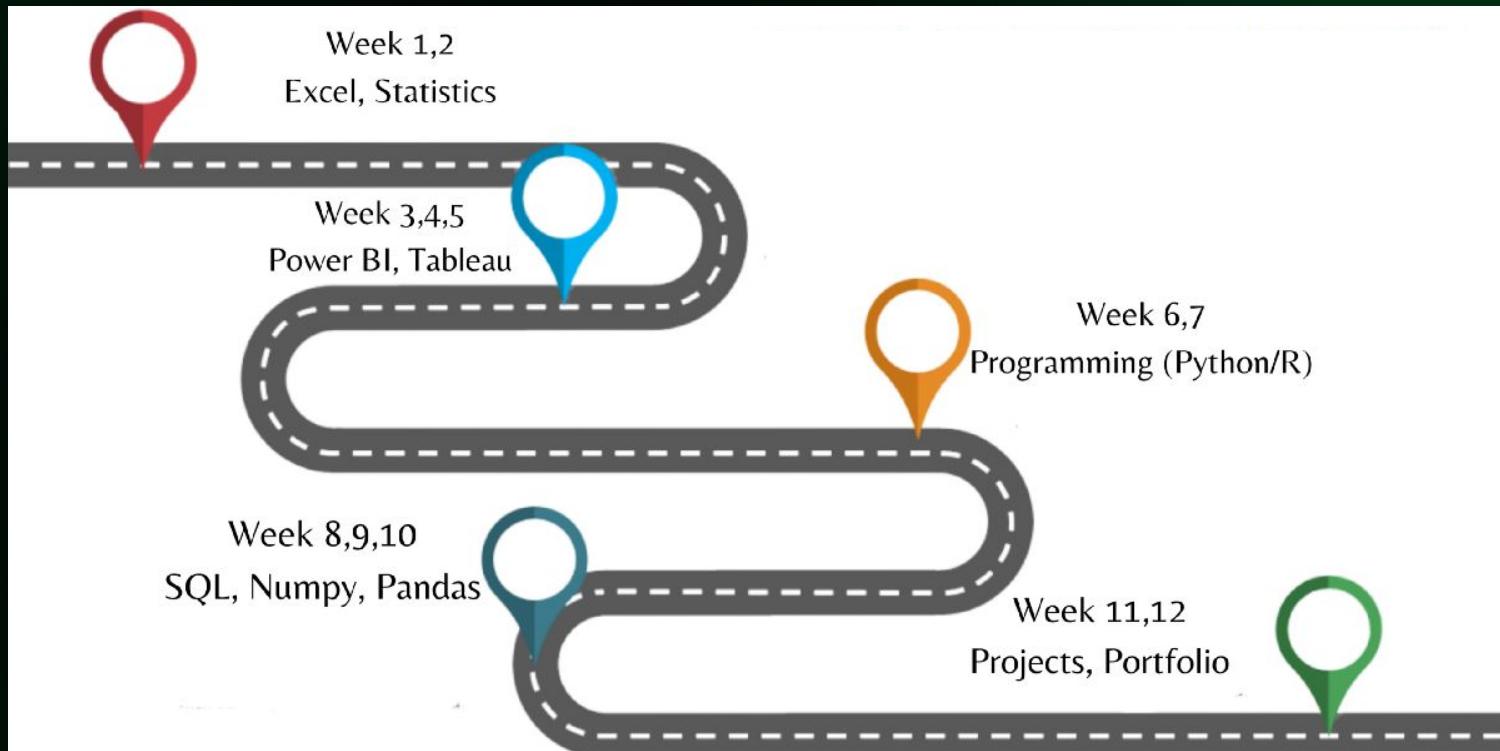
Technical skills:

- Statistics
- BI Tools: Power BI, Excel, Tableau
- Programming: R, Python

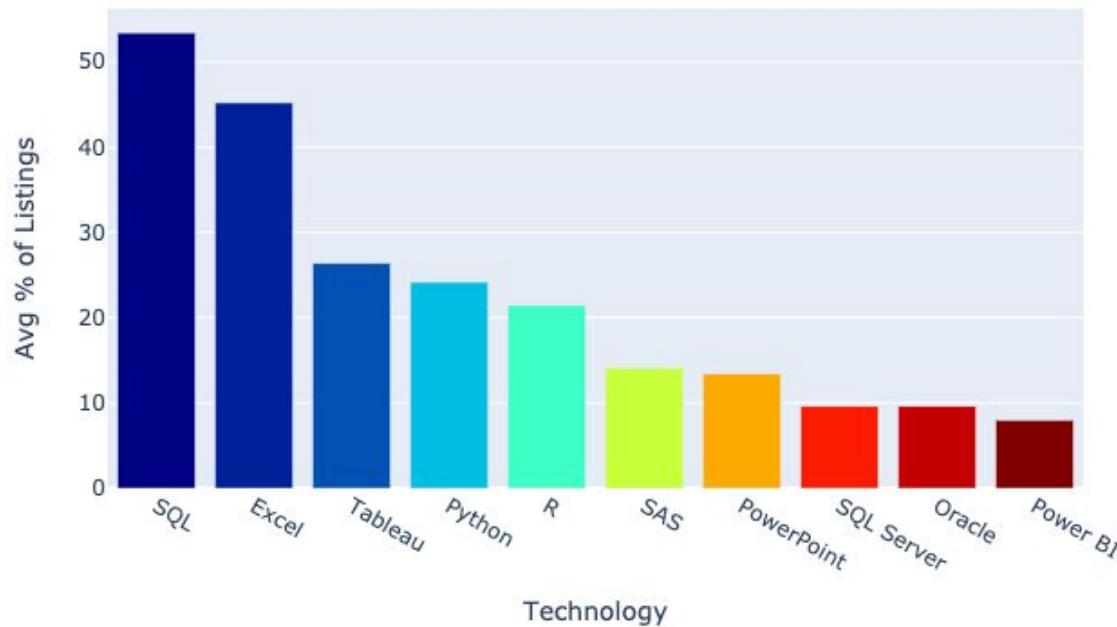
Soft skills:

- Critical thinking and problem solving
- Writing and communication
- Domain knowledge

DATA ANALYST ROADMAP

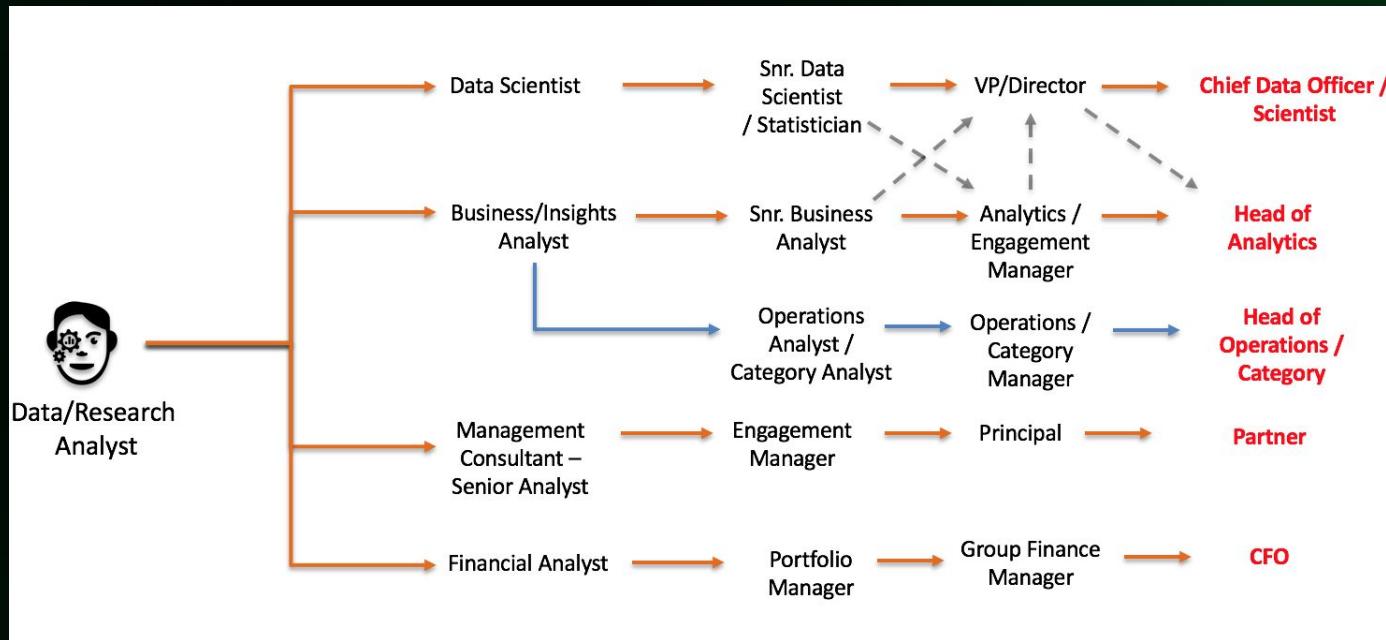


Technologies in Data Analyst Job Listings 2020



4 CAREER PATH

.... Depend on how do you know yourself.



5

Data Analytics in Cốc Cốc CHALLENGES & LESSONS LEARNED

DATA ANALYST IN REAL WORLD

- What is the key metrics we should measure? How to calculate?
- Where is the data?
- How to communicate insights to other team?



5

Data Analytics in Cốc Cốc **CHALLENGES & LESSONS LEARNED**

Challenges:

- Huge amount of data
- Various data types:
transactional, log, NLP,...
- Various domains: business cases, departments
- High needs of data insights to make decisions

Lessons learned:

- Expand our technical skill sets: not Excel anymore
- Improve domain knowledge
- Increase working efficiency
- Improve communication/presentation skills
- High touch with data engineer and data scientist



coc coc

/ WEBINAR

Q&A





cốc cốc

/ WEBINAR

**Làm DA có cần phải học chuyên ngành
công nghệ thông tin không?**

**Sinh viên kinh tế muốn làm DA thì cần
học thêm những gì?**

**Cách tìm kiếm thực tập cho người có background tài
chính kế toán - nên học những phần mềm nào, tay
ngang thì nên tìm hiểu và học ở đâu**



coc coc

/ WEBINAR

QUESTION FOR WEBINAR ATTENDEES

3 players who has correct and fastest answer
will receive a voucher 50K

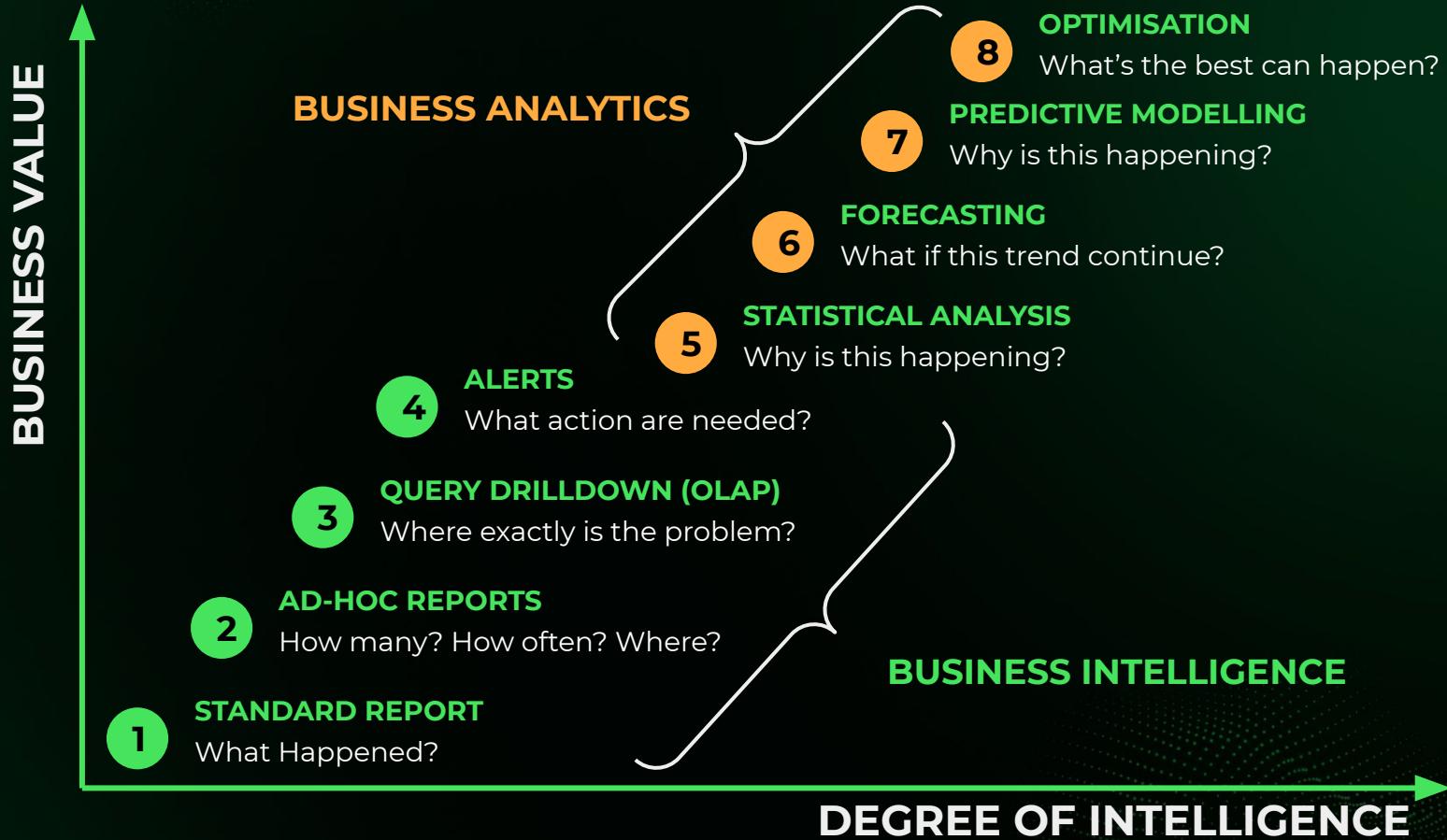


coc coc

/ WEBINAR

How many questions can be solved by Data Analyst?

8 questions





coc coc

/ WEBINAR

03

Learn more about
**DATA
ENGINEER**



NGUYỄN ĐÌNH HÙNG

SENIOR SOFTWARE ENGINEER at CỐC CỐC

- 10+ years working with *large-scale data and backend systems*
- Building reliable, high-performance systems

01

Who are Data Engineers?

02

A day of a Data Engineer

03

Job Market & Career Path

04

Skill set & How to start with DE?

05

Challenges & Lessons Learned

06

Q&A

1

WHO ARE THEY? WHAT THEY DO?

They build and maintain their organization's data ecosystem.

- Build data pipelines for bringing together information from different data sources
- Distribute data throughout the organization
- Prepare data for analytical, machine learning and operational uses
- Understand business objectives



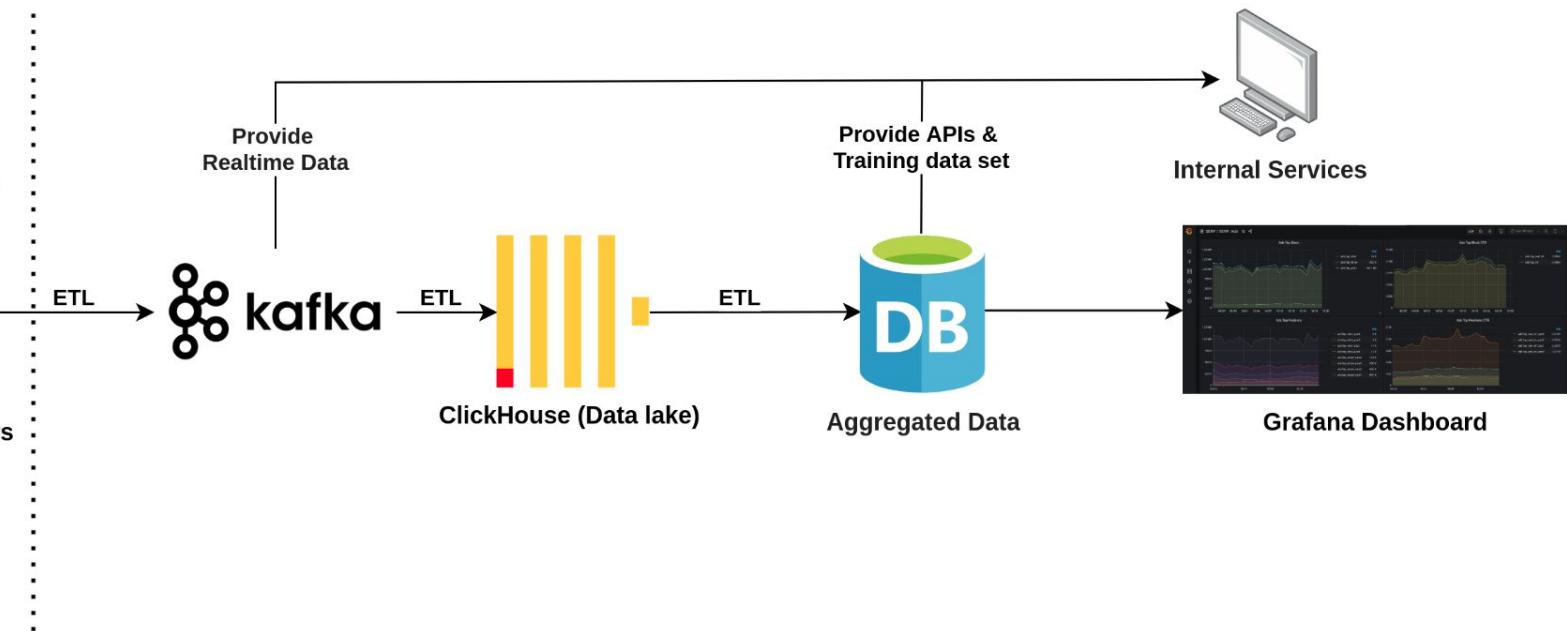
2

A DAY OF DATA ENGINEER

- **Daily meetings**
- **Data loading:** SQL, ETL tools, Cron jobs taking data from data sources
- **Data manipulation:** Coding Java, Python, C++ tasks to extract, aggregate raw data into usable form docs, tables,...
- **Data pipeline:** create tables, index storing the results into data warehouse
- **Optimizing:** monitor the jobs and results, optimize performance,...
- **Data servicing:** create web APIs for serving our data to teams, clients

DEMO - DATA PIPELINE

Data Sources



DEMO - DATA PIPELINE

Data sources:

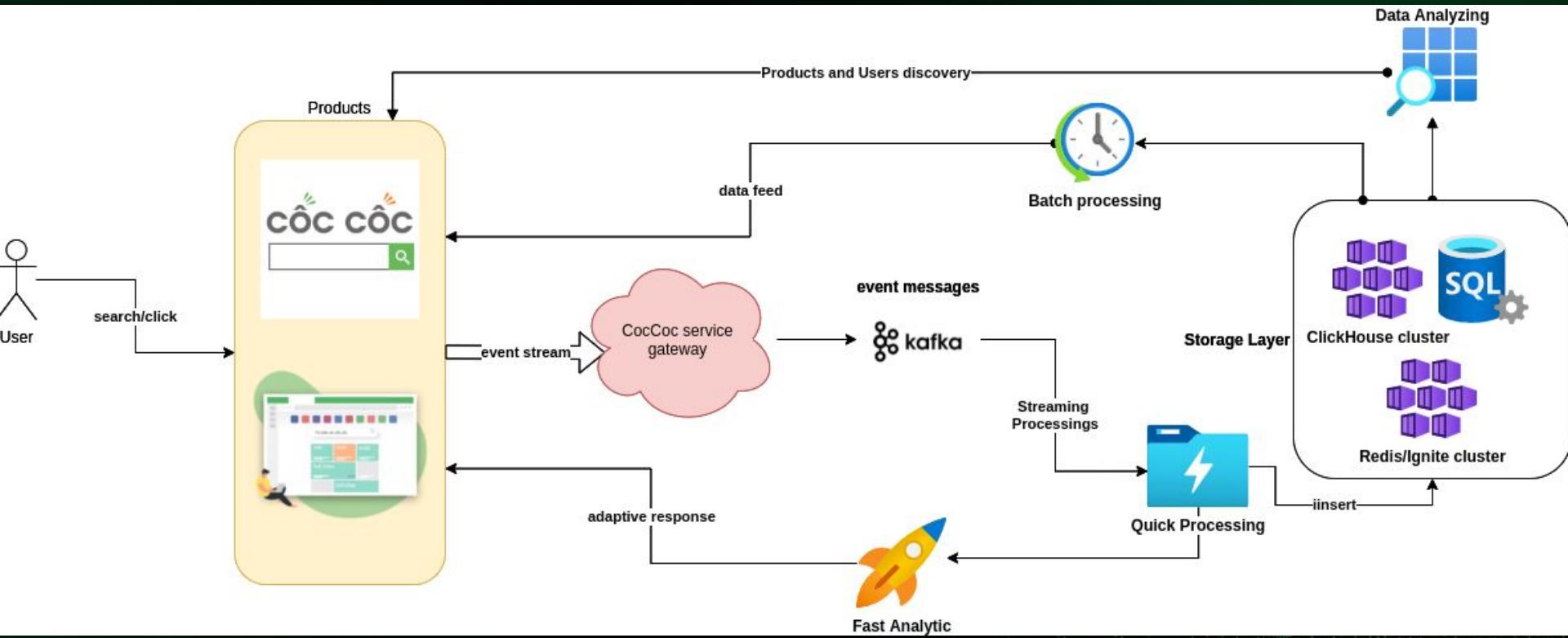
- 3rd APIs: SafeForce, AccessTrade, CityAds...
- Cloud Storage: Coc Coc Apps on Google Play, App Store,...
- Browser metrics: Coc Coc browser usage, crash reports,...
- Service's Database: Advertisement, Coc Coc Account,...

ETLs: inhouse ETL tools Java, C/C++, Go, Python

Database: ClickHouse, NoSQL (MongoDB, ElasticSearch,...), SQL (MySQL)

Visualization: Grafana, inhouse

DEMO - NEW FEED



3

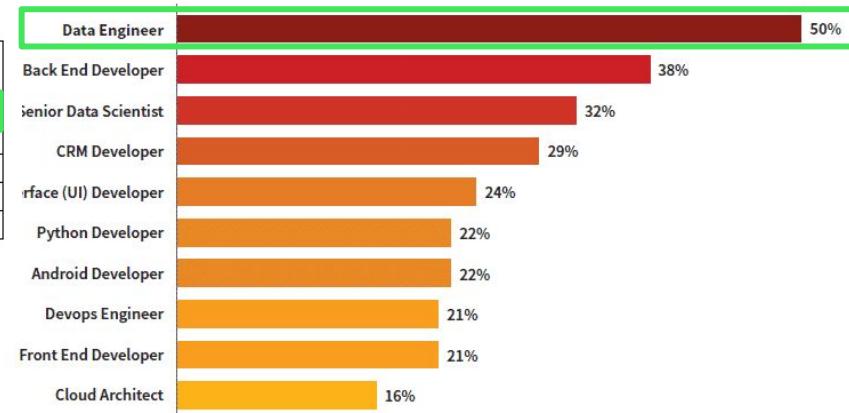
JOB MARKET & CAREER PATH

Data Engineer Shortage

	Open Positions*	Candidates**	Number of candidates / open positions
Data Engineer	16022	40645	2.53
Data Scientist	20756	98935	4.76
Big Data Engineer	1736	4300	2.47
Web Developer	27098	292830	10.8
Marketing Manager	15253	820534	53.79

Data Engineer Job Growth

FASTEST GROWING TECH OCCUPATIONS
YEAR-OVER-YEAR GROWTH



VIETNAM JOB MARKET

Top 3 emerging positions



AI/ML Engineer



Big Data Engineer



DevOps Engineer

Top trending technical skills required



DevOps



Java



Python

Top trending soft skills required



PMP



English



Japanese

“Data Scientist and Big Data Engineer are expected to be at the top of the list of most in-demand positions in the next 5 years with their rapidly growing rate.”

Adecco Vietnam Salary Guide 2020

CAREER PATH

Start off as Software engineers, BI analysts

Engineering Grades									(*) Cốc Cốc Grading
Grades	P8	P9	P10	P11	P12	P13	P14	P15	
Willis Towers Watson grouping	Entry	Intermediate		Career		Specialist		Master	
Coc Coc grouping	Junior	Middle		Senior		Expert			●

Autonomy

Train team members

Own disciplines

Tech leader

Process

Other disciplines

Approach

Business sense

Scale

4

SKILL SET & HOW TO START WITH DE

Programming languages: Java, Python, SQL,...

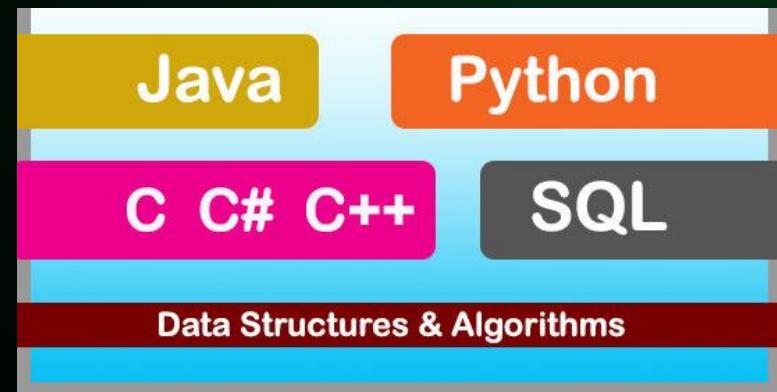
Good understanding of DS&A

*.nix environment

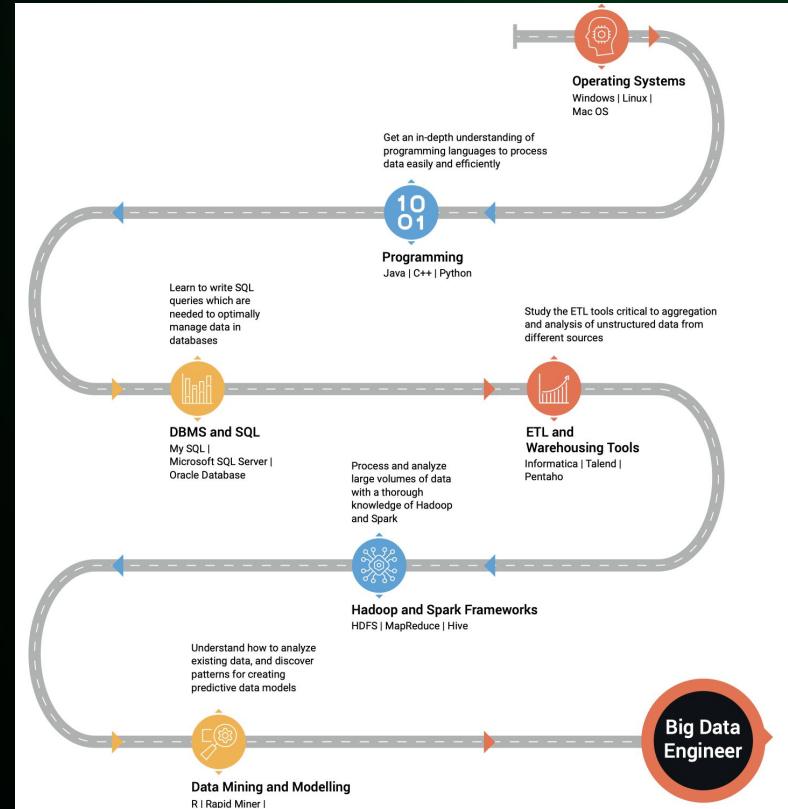
Understanding of ETL tools

Database systems (SQL, No-SQL)

Web services



ROADMAP



5

DATA ENGINEERING IN CỐC CỐC CHALLENGES & LESSONS LEARNED

- Data in various domains Browser, Ads, Search,...
- Serve our 25 mills VNese users !
- Terabytes databases
- Billions events/day
- Thousands of rps systems
- Talented teams



coc coc

/ WEBINAR

Q&A





cốc cốc

/ WEBINAR

Việc chuyển từ 1 software engineering sang data engineering cần hoàn thiện thêm những kỹ năng nào?

Em muốn hỏi là sinh viên năm 2 thì có thể làm công việc gì bổ trợ cho công việc data engineer ạ

Mình năm nay 26 tuổi. Vừa bắt đầu tìm hiểu và học các kiến thức cơ bản để trở thành 1 Data Engineer. Tương lai Cốc Cốc có mở các chương trình tuyển các thực tập sinh hoặc JUNIOR không ạ?



coc coc

/ WEBINAR

QUESTION FOR WEBINAR ATTENDEES

3 players who has correct and fastest answer
will receive a voucher 50K



coc coc

/ WEBINAR

Which phase that “ETL” stands for?



coc coc

/ WEBINAR

Extract - Transform - Load



coc coc

/ WEBINAR

04

Learn more about
**DATA
SCIENTIST**



ĐỖ VIỆT PHƯƠNG

DATA SCIENTIST TEAM LEADER at CỐC CỐC

- 10+ years working with *large-scale analysis and machine learning*
- Contribute directly *predictive modeling* to the company productions



Tab mới +

coccoc.com

Ứng dụng Tool Cốc Cốc Khiên xanh 2021 Report - Check list Ambassador eGov

Giao diện và quyền lực như 'Nam Tào' Xuân Bắc: 1 vợ 3 con, nhà biệt phủ xây bằng gỗ tự nhiên, là Phó giám đốc Nhà hát kịch Việt Nam

VN-index lập đỉnh lịch sử trọng phiên 27/10, vốn hóa thị trường tăng thêm 5,4 tỷ USD.

Các thế giới khôn đùn, một quốc gia không cần nhập khẩu dầu, ngày càng giàu có hơn nhờ khung hoàng năng lượng

Giai mộng là cuộc kỳ quan trọng & chiếc gối ngủ quyết định giấc ngủ | Ví thẻ người Nhật chỉ tin tưởng gối này

Được phát triển từ công nghệ mới - Bản hảng nghìn chiếc vẫn cháy ầm ầm | Hiệu quả đã được công nhận | Hàng chính hãng 100% tại đây giá sale 40%+ FreeShip 0Đ

GÓI NGỦ Bán chạy

Cuộc sống của Diệp Lâm Anh hiện tại ra sao?

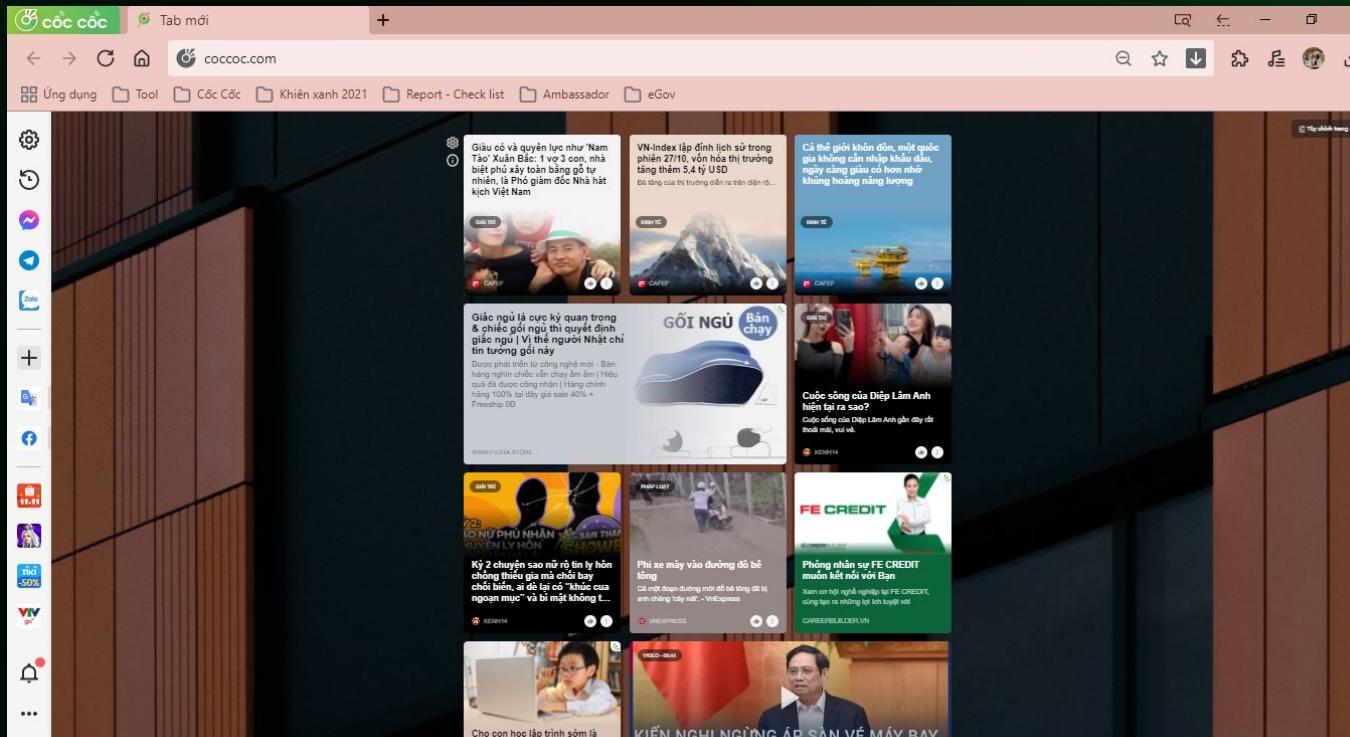
Ký 2 chuyện sao nữ rõ tin ly hôn với chồng cũ: Vì sao cô "không cưa ngoan mục" và bí mật không t

Phi xe máy vào đường đỗ bê tông

Phòng nhân sự FE CREDIT muốn kết nối với Bạn

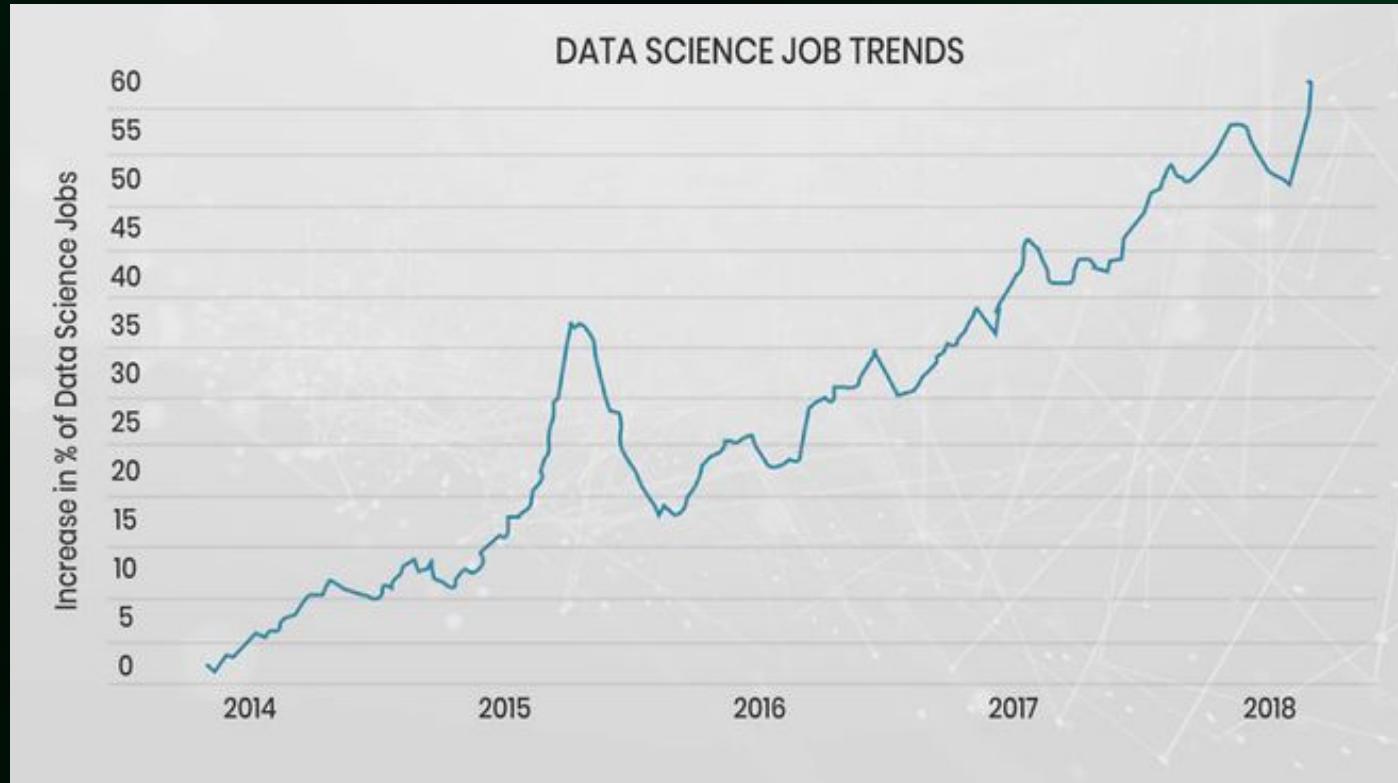
Cho con học lập trình sớm là

KIẾN NGHỊ NGỪNG ÁP SÁN VỀ MÁY Bay





Increase in Demand for Data Scientist



Source: kickstarter.com

What Industries hiring us?

Technology/IT



43%

Industrial



39%

Financial



16%

Healthcare



2%



coc coc

/ WEBINAR

ROLES OF DATA SCIENTIST

Contribute to Recommendation News Feed

The click through rate increased up to **20%** compared to the human setting.



côc côc

/ WEBINAR

ROLES OF DATA SCIENTIST

Contribute to Monetization Advertising Department

The ad relevances increased up to **8%** compared to the non-optimization solution



coc coc

/ WEBINAR

ROLES OF DATA SCIENTIST

Contribute to Anomaly Detection and Early Warning System

Reduce upto **100%** human workload for rule-based setting



cốc cốc

/ WEBINAR

01

Who are Data Scientists?

02

A day of a Data Scientist

03

How to become a Data Scientist?

04

Career Path

05

Challenges & Opportunities

06

Lesson Learned from Cốc Cốc Project

07

Q&A



côc côc

/ WEBINAR

1

WHO ARE THEY?

Dictionary

Search for a word



da·ta sci·en·tist

noun

noun: **data scientist**; plural noun: **data scientists**

a person employed to analyze and interpret complex digital data, such as the usage statistics of a website, especially in order to assist a business in its decision-making.

"Silicon Valley technology companies are hiring data scientists to help them glean insights from the terabytes of data that they collect everyday"

Translate data scientist to

Choose language ▾

Definitions from Oxford Languages

Feedback



coc coc

/ WEBINAR

2

A DAY IN THE LIFE OF A DATA SCIENTIST

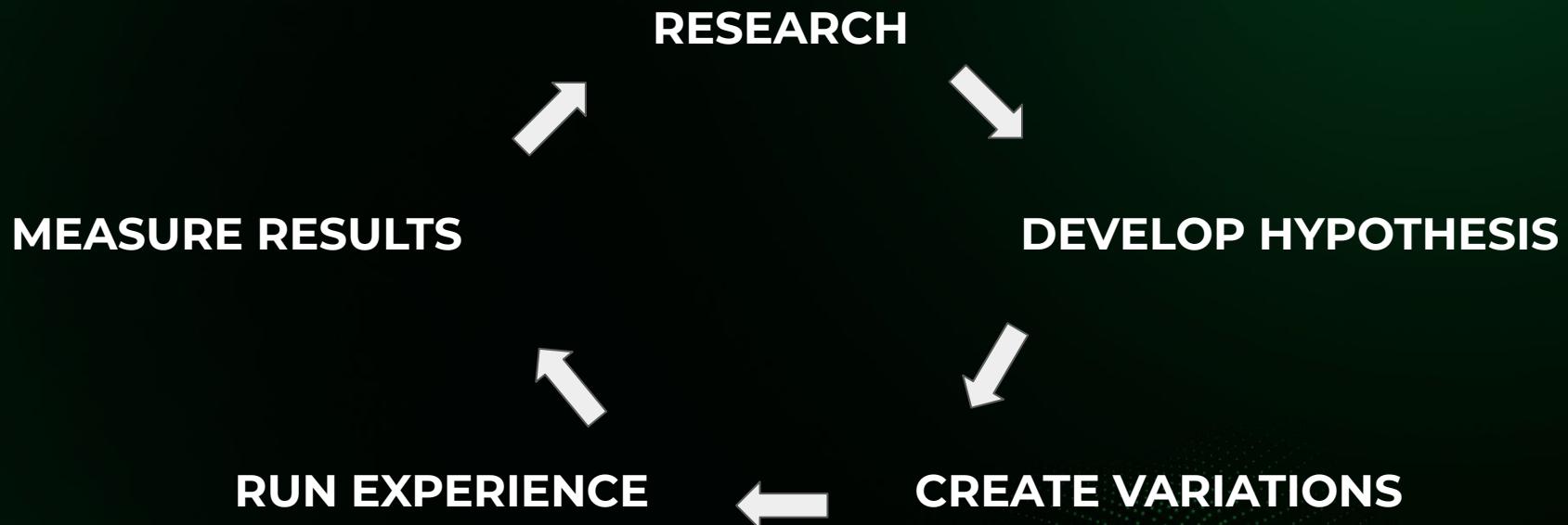
1

What Does
a Data Scientist do?

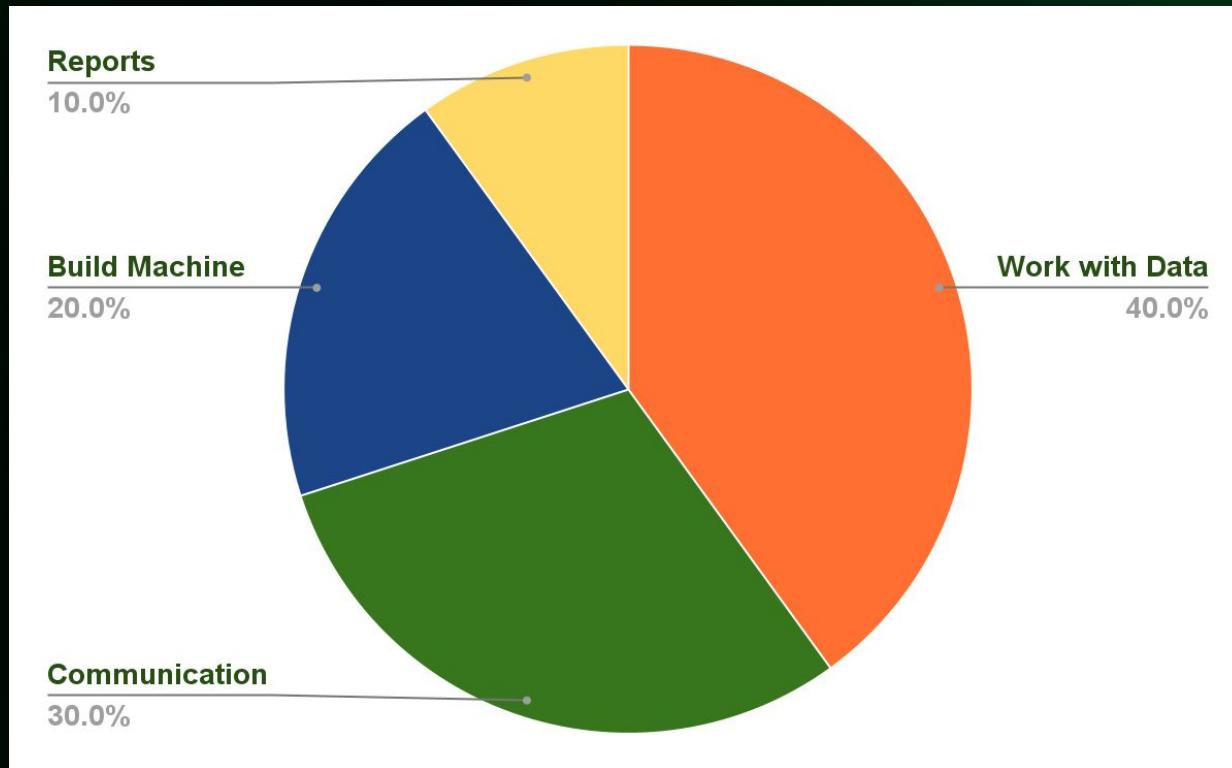
2

Important Skills

WHAT DOES A DATA SCIENTIST DO?



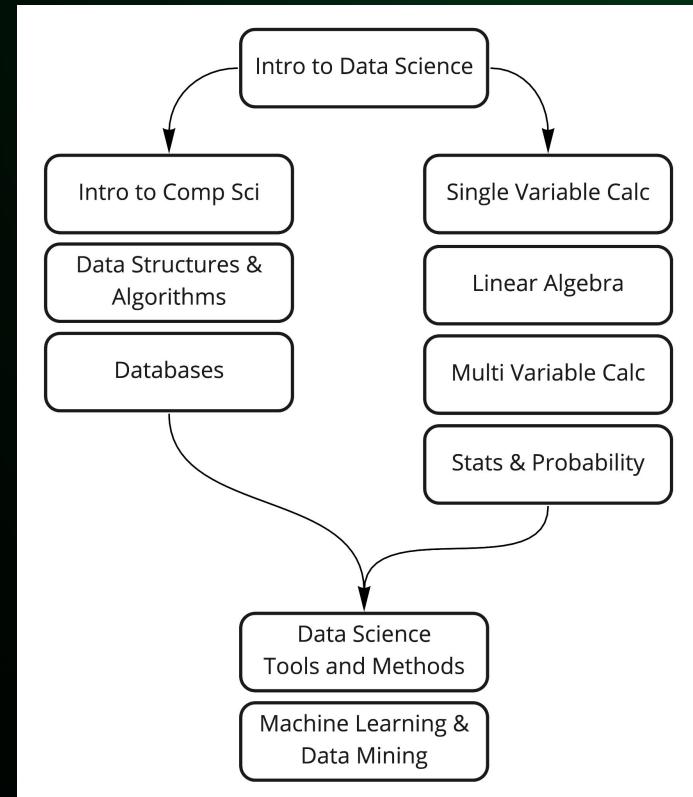
WORKLOAD



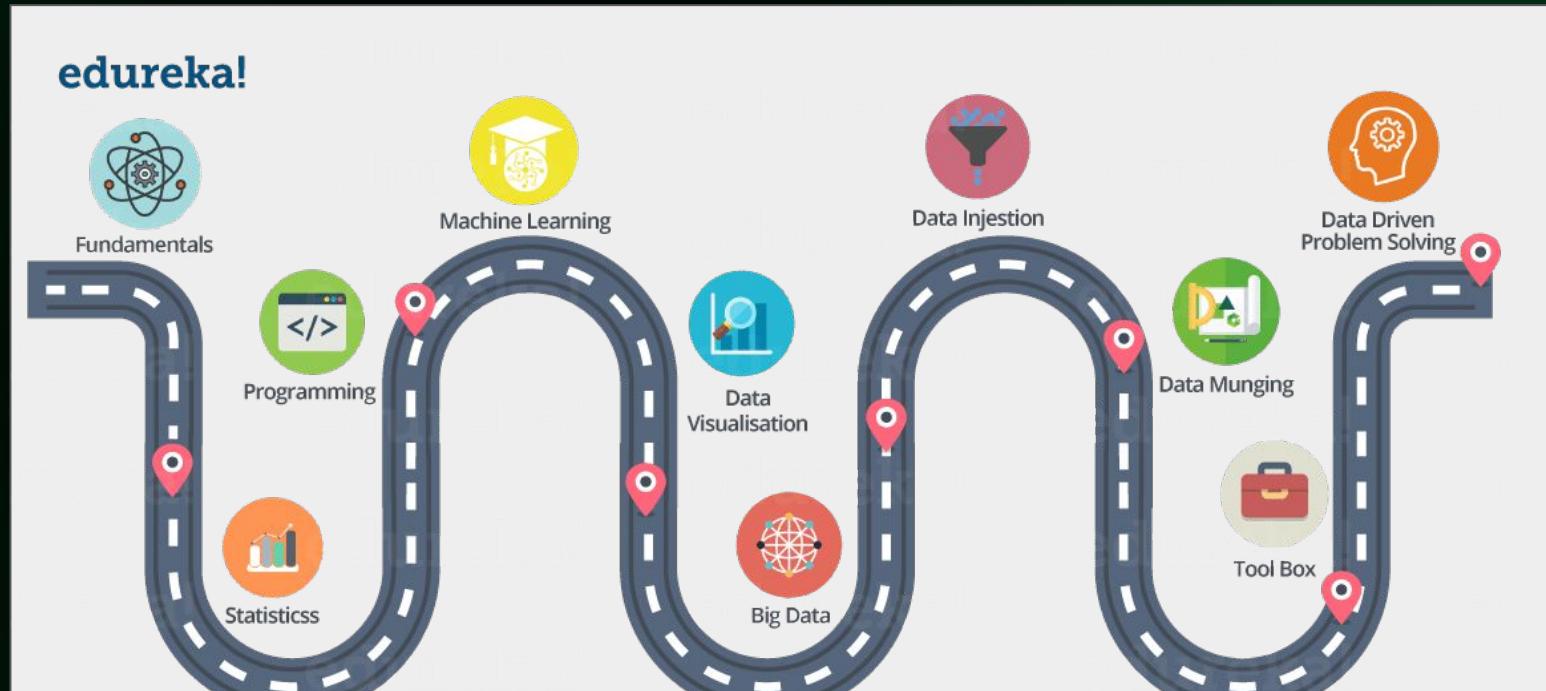
3

HOW TO BECOME A DATA SCIENTIST

- Study Data Scientist courses
- Read Data Science Books: [Link](#)
- Participate in Data Science community



IMPORTANT SKILL ROADMAP 1 - 1.5y



Road map: <http://nirvacana.com/thoughts/2013/07/08/becoming-a-data-scientist/>

A TYPICAL DATA SCIENTIST

Predominantly Male
(69%)



8 years
in the workforce overall

Bilingual



Python/R
(73%)

2.3 years
as a Data Scientist



Master/PhD
(74%)



CODING TOOLBOX



54% 45% 36% 19% 18% 8%



côc côc

/ WEBINAR

4

CAREER PATH & PROGRESSION

Data Engineering

Computer Science

Data Analysis

?

► Entry-Level

Junior, Associate
Data Scientist

AVG.SALARY: 91K
1-3 YEARS

► Mid-Level

Data Scientist

AVG.SALARY: 108K
3-6 YEARS

Machine Learning,
Ai Engineer

AVG.SALARY: 127K
3-6 YEARS

Data Science
Manager, Architect

AVG. SALARY: 140K
5-8 YEARS

Chief, Principal,
Sr. Data Scientist

AVG. SALARY: 146K
5-8 YEARS

► Advanced-Level

Director of
Data Science

AVG. SALARY: 169K
5-8 YEARS

5

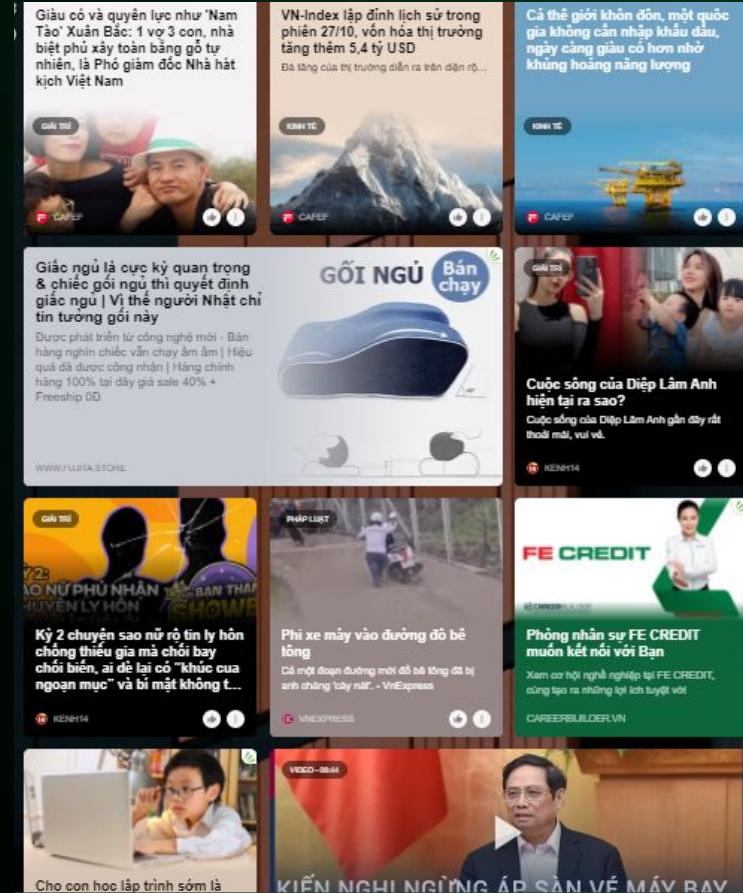
DATA SCIENCE IN CỐC CỐC CHALLENGES & OPPORTUNITIES

- Build ML/AI Prediction model for AdEngine (CTR prediction, Quality Score model)
- Designing and developing machine learning and deep learning systems
- Running **machine learning** tests and experiments
- **Implementing appropriate ML algorithms**, and turn them into microservices on our production environments.
- Improve current targeting model (user interest/user catalogs)
- Build recommendation/suggestion model for our current products (Music Recommendation Engine, B2B, B2C products)
- Others: Analysis data, doing ad-hoc analysis and presenting results in a clear manner.

6 LESSON LEARNED FROM NEWS FEED PROJECT

The personalization distributed news content platform (NewsFeed)

System increased upto **20% CTR** &
5% DAU compared to the human setting





coc coc

/ WEBINAR

Look back our life!

RESEARCH

MEASURE RESULTS

DEVELOP HYPOTHESIS

RUN EXPERIENCE

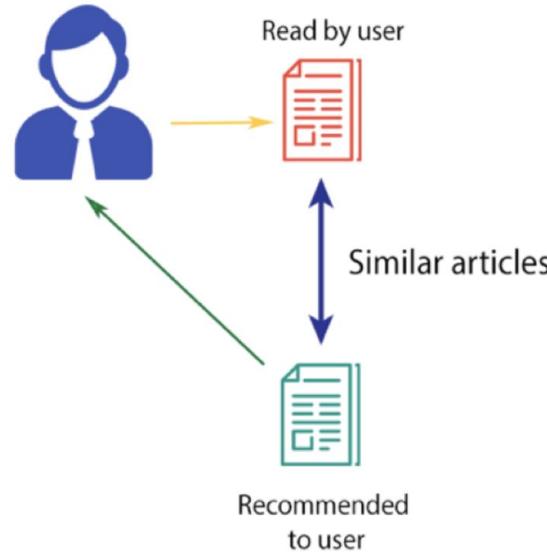
CREATE VARIATIONS



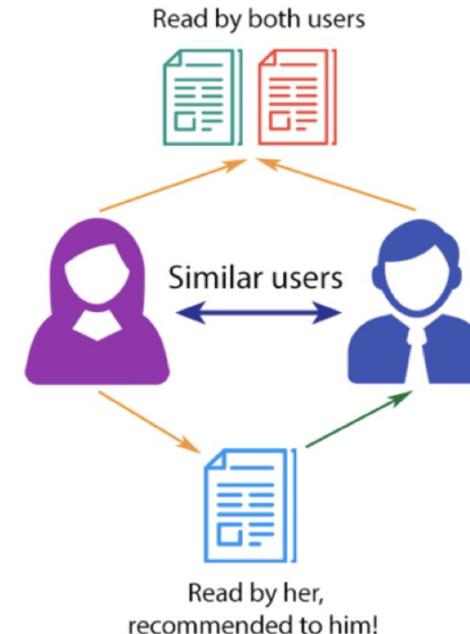
Step 1:

Research - Recommendation Engine

CONTENT-BASED FILTERING

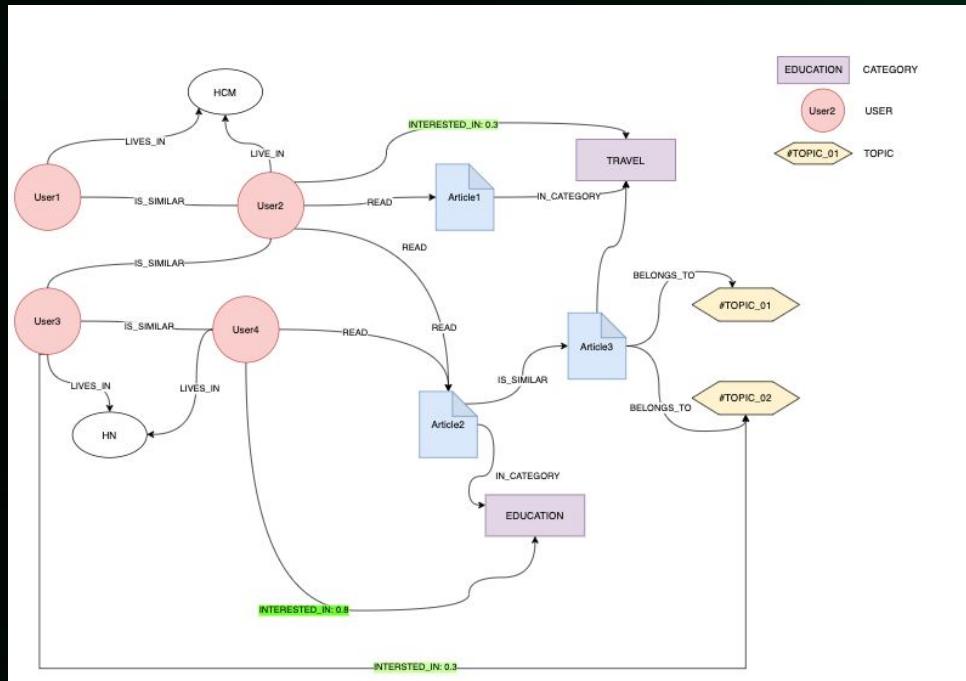


COLLABORATIVE FILTERING



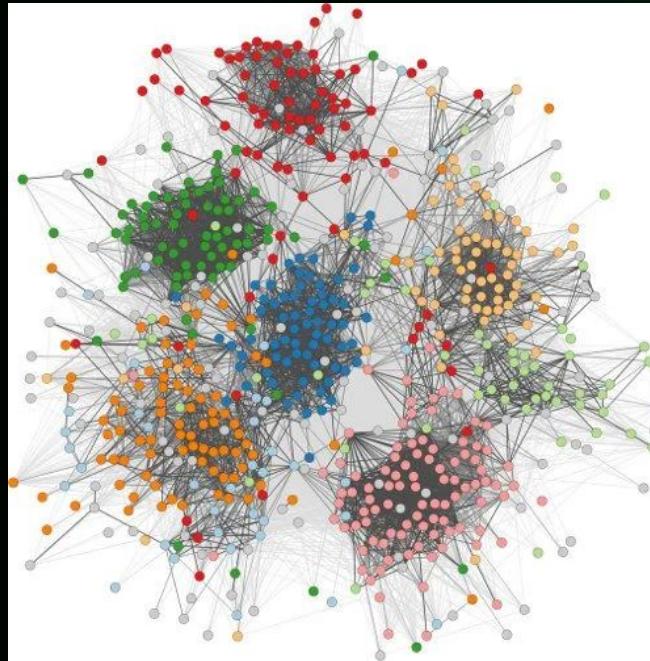
Step 2:

Create Hypothesis about Relationship between User Behavior & User - Articles Representation (GraphDB)



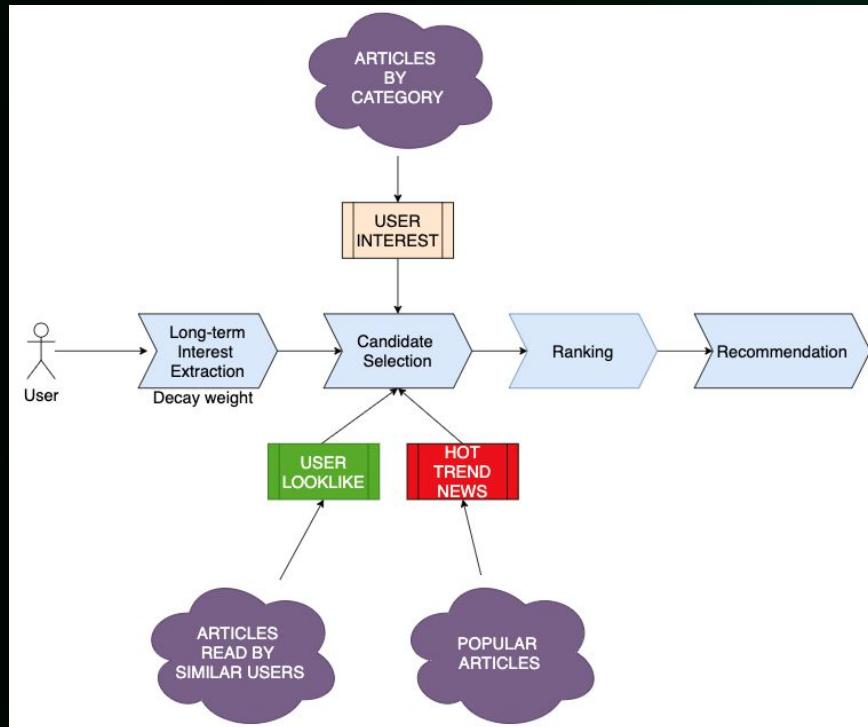
Step 2:

Create Hypothesis about Relationship between
User Behavior & User - Articles Representation (GraphDB) - cont



Step 2:

Create Hypothesis about Relationship between User Behavior & User - Articles Representation (GraphDB) - cont



Step 3:

Create Variations (Code the solutions, Machine Learning Models)

GraphDB Relationship.

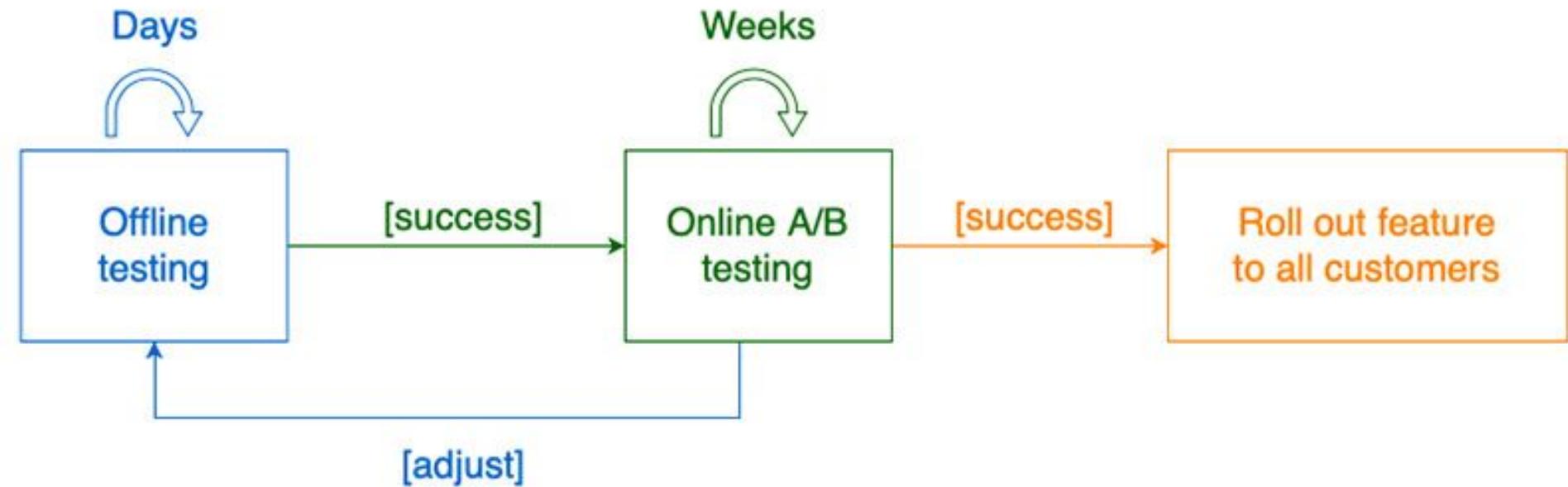
- (user1)-[:IS_SIMILAR]->(user2) based on queries, topicmodel..(cosine similarity)
- (article1)-[:IS_SIMILAR]->(article2)
- (article1)-[:BELONGS_TO]->(#TOPIC_01)
- (article1)-[:IN_CATEGORY]->(CATEGORY)
- (user1)-[:INTERESTED_IN:weight]->(#TOPIC)
- (user1)-[:INTERESTED_IN:weight]->(CATEGORY)
-

Query samples:

- Get Top K articles are in {categories} OR {topics} that we can recommend to our target user recently and their neighbors ? @Đức Lê Huỳnh please check the query grammaticaly, and the benchmark.
- neighbors = MATCH (c:User)-[r:IS_SIMILAR]->() WHERE userId= \$uid RETURN c.userId
 - recommendation = SUM OF
 - (1) similar users MATCH (c:User)-[r:READ]->() WHERE userId= \$uid RETURN c.userId AND
 - (2) MATCH (a:Article)-[r:BELONGS_TO]->(\$lst_categories) OR MATCH (a:Article)-[r:BELONGS_TO]->(\$lst_topics) WHERE (u:INTERSTED_IN:_)>0.5 AND u.uid IN (\$lst_neighbors) RETURN a.articleId AND
 - (3) BY user himself: MATCH (a:Article)-[r:BELONGS_TO]->(\$userid) OR MATCH (a:Article)-[r:BELONGS_TO]->(\$lst_topics) WHERE (u:INTERSTED_IN:_)>0.5 AND u.uid IN (\$lst_neighbors) RETURN a.articleId

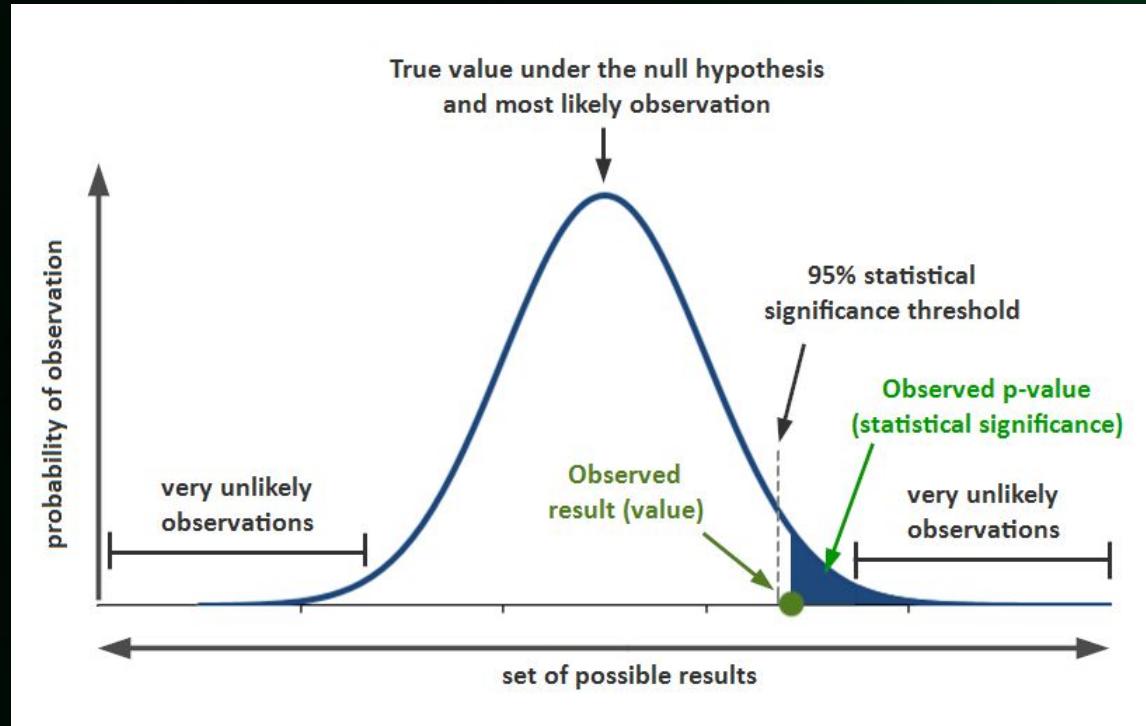
Step 4:

Run Experiment (Online/Offline A/B Testing)



Step 4:

Run Experiment (Online/Offline A/B Testing)



Step 5:

Monitor and measure the results





coc coc

/ WEBINAR

The click through rate increased up to **20%**
compared to the human setting with **99%** of
statistical confidence interval



coc coc

/ WEBINAR

Q&A





cốc cốc

/ WEBINAR

Những dự án phù hợp với newbie để dần làm quen với môi trường công việc thực tế ? Các lỗi cơ bản, sai lầm của newbie trong lúc học và ra làm thực tế ?

Liệu một newbie trái ngành, hay chưa có kinh nghiệm/knowledge nhiều về IT, lập trình, có thể chuyển hướng sang Data Science hay không?

Data science là 1 ngành đa dạng lĩnh vực. Với sinh viên mới bắt đầu, ngoài kiến thức chung về data thì có phải học thêm kiến thức về các ngành đó không?

MINIGAME



coc coc

/ WEBINAR

10

**multiple - choice
questions**

For attendees in Webinar



01 1ST PRIZE

Energizer
Bluetooth Speaker



01 2ND PRIZE

Logitech
Wireless Mouse



05 3RD PRIZE

Xiaomi
Headphone

RULE:

- 1. Click on the link that sent in the chat box**
- 2. Register to join the game**
- 3. Answer 10 multiple-choice questions**
- 4. The winner is who has many correct & fastest answers**



cốc cốc

/ WEBINAR

CONGRATULATIONS

The Winner please send your information (name, phone number, email, address) with a result screenshot to
Fanpage Trình duyệt Cốc Cốc

THANK YOU!

