

Pre-Decisional Support: Accurate, Agile and Active Data Delivery

Lei Xia & Zechen Xu
Data Mining Engineer & Project Manager
NetEase



Lei Xia

Data Mining Engineer

Zechen Xu

Data Product Manager & Project Manager



GDC® GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

ThunderFire UX Team



User Experience Research



User Experience Design



Big Data Service

8 Game Studios

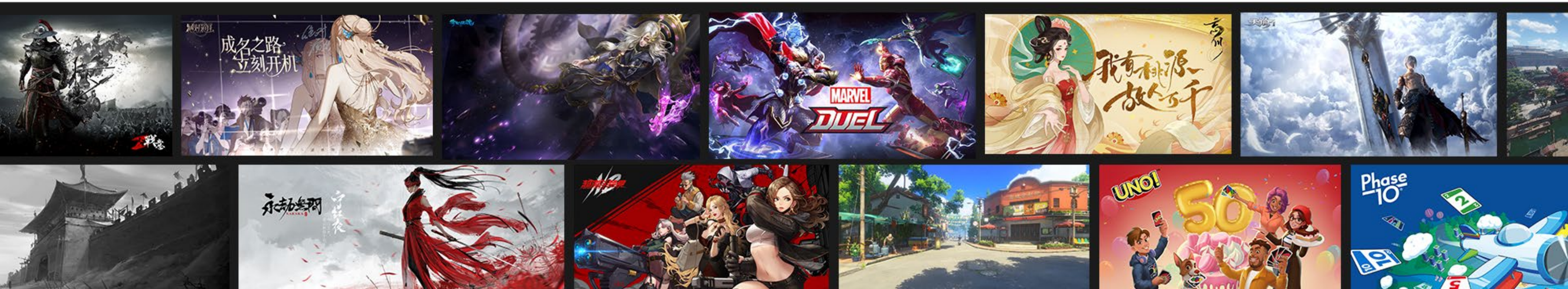
2000+ Developers

26 Games

380 million Players

200+ countries and regions

10+ billion logs are processed everyday



Overview

Part I

- What kind of requirement is data team going to deal with?
- Understand the requirement and product teams
- Make suitable management

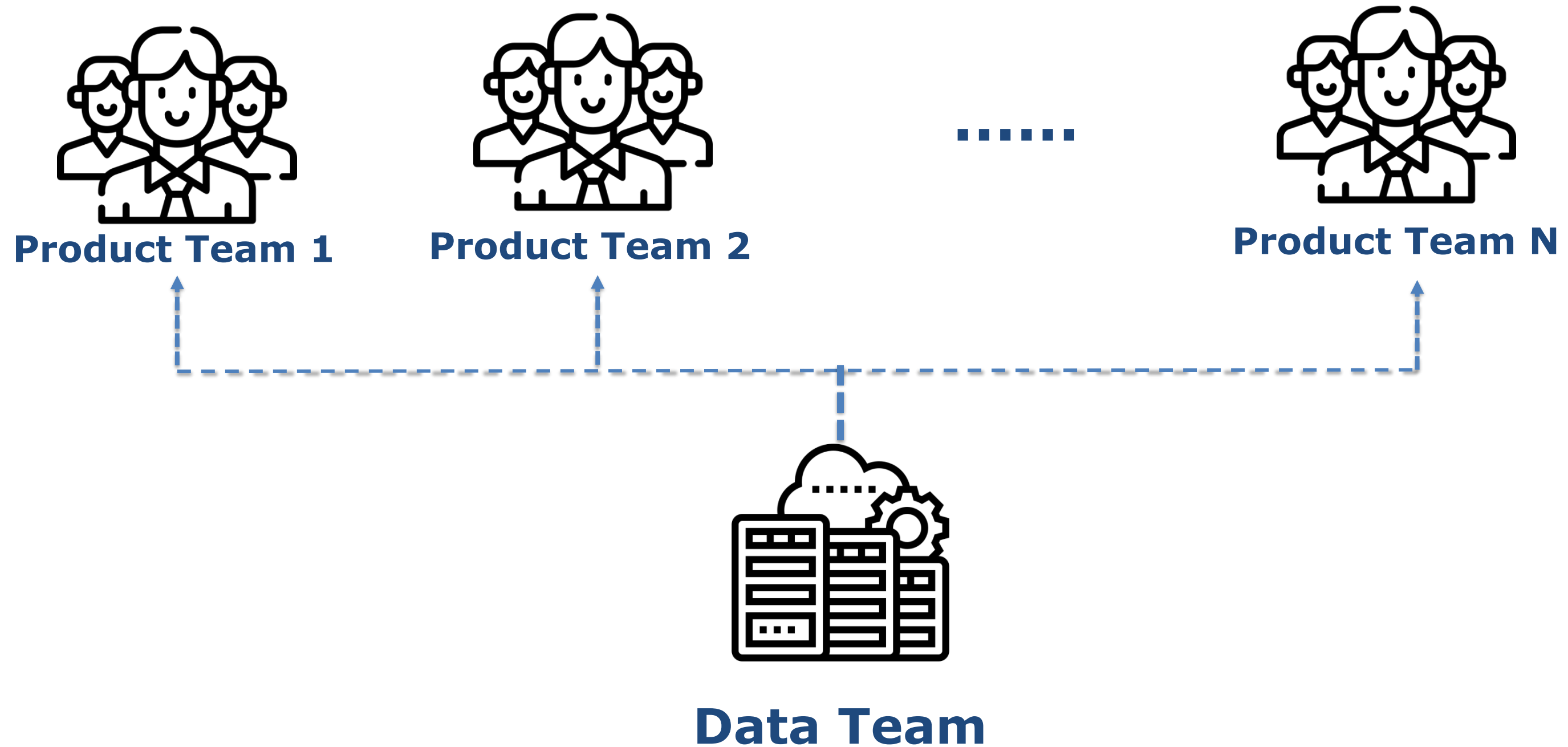
Part II

- Build a Data Team From Scratch
- The Evolution of Data Pipeline
- Quality Assurance for Data Pipeline

Takeaway

Part.1 Role and responsibilities of a data team

Data Team& Product Teams



Responsibilities of a data team

Data (analysis)

Past

Conclusion

Now

Real-time data

Future

Predictions

**Game data-based
products or tools**



GDC

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

Data dashboard : It is a visualization tool used to monitor status of core business



Demo from "sky eye" BI System

- what is going on with my game?
- what adjustments should I make in operation?



GDC

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

1

Data dashboard : provide product teams and decision makers an access to the status quo.

ARPU ...
Pay rate DAU
Core data LTV
ARPPU Retention

Win rate ...
Average damage ...
Other data DPS
Participation Rate

- Status quo
- Target deviations
- Integrate basic information

Data dashboard : Standard modules and custom modules



Standard modules:

- Fast
- Stable
- Low cost



Custom modules:

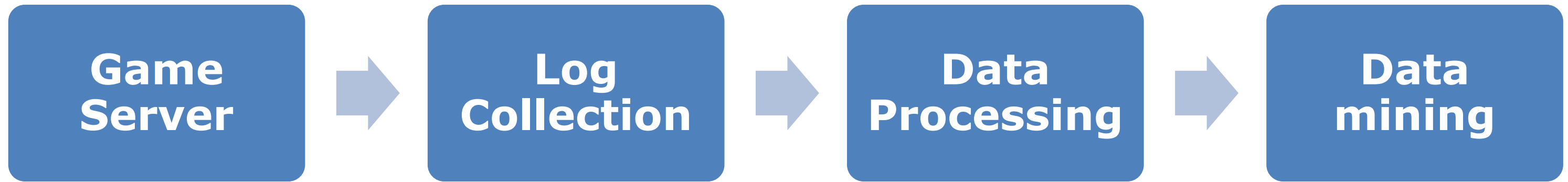
- Customizable
- Precise
- Differentiation



GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

1 Data mining : for a specific purpose



Basic Data Pipeline

- Used for specific situation
- Urgent requirement

1

Data mining : for a specific purpose

■ Data dashboard
■ Data mining



- Highly customized
- Real-time data
- Take up lots of time for developers



GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

0 coding skills needed

+

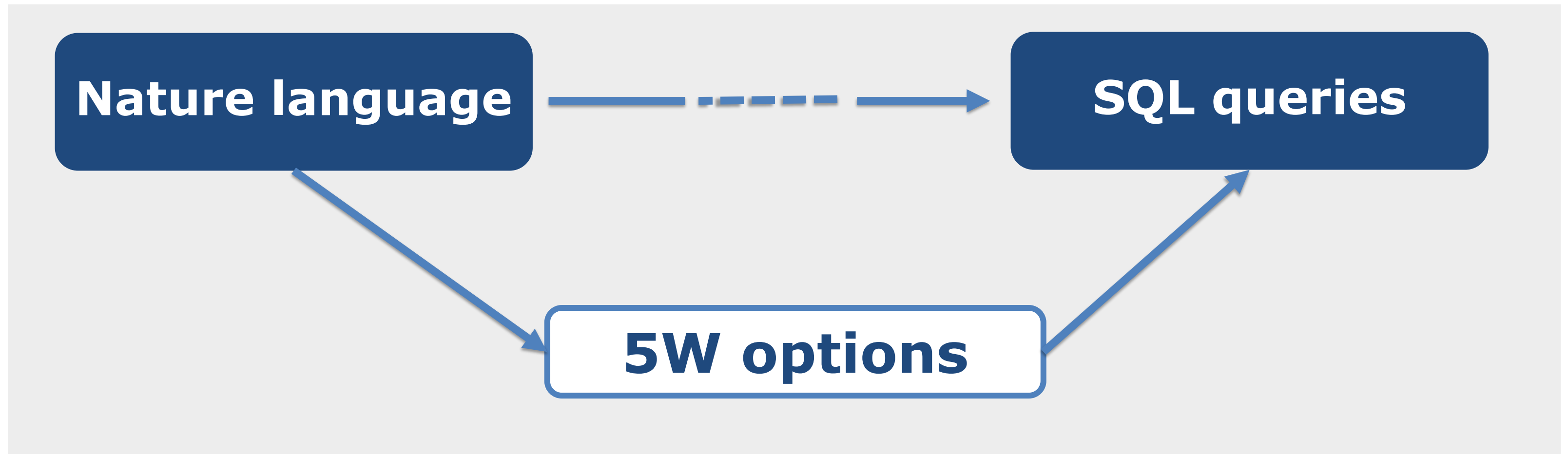
Highly customized results

+

Self-service platform

1

A data mining system that uses natural language



Self-Service BI Platform

1 A data mining system that uses natural language

[1] On July 1, 2021, [2] server 10001, [3] the average damage of the [4] characters [5] participating in the PVP gameplay.

[1]	When	Time range of this query
[2]	Where	Servers range selected for this query
[3]	What	The purpose of this data query
[4]	Who	The primary key to this data query
[5]	Whom	Describe the details of the primary key



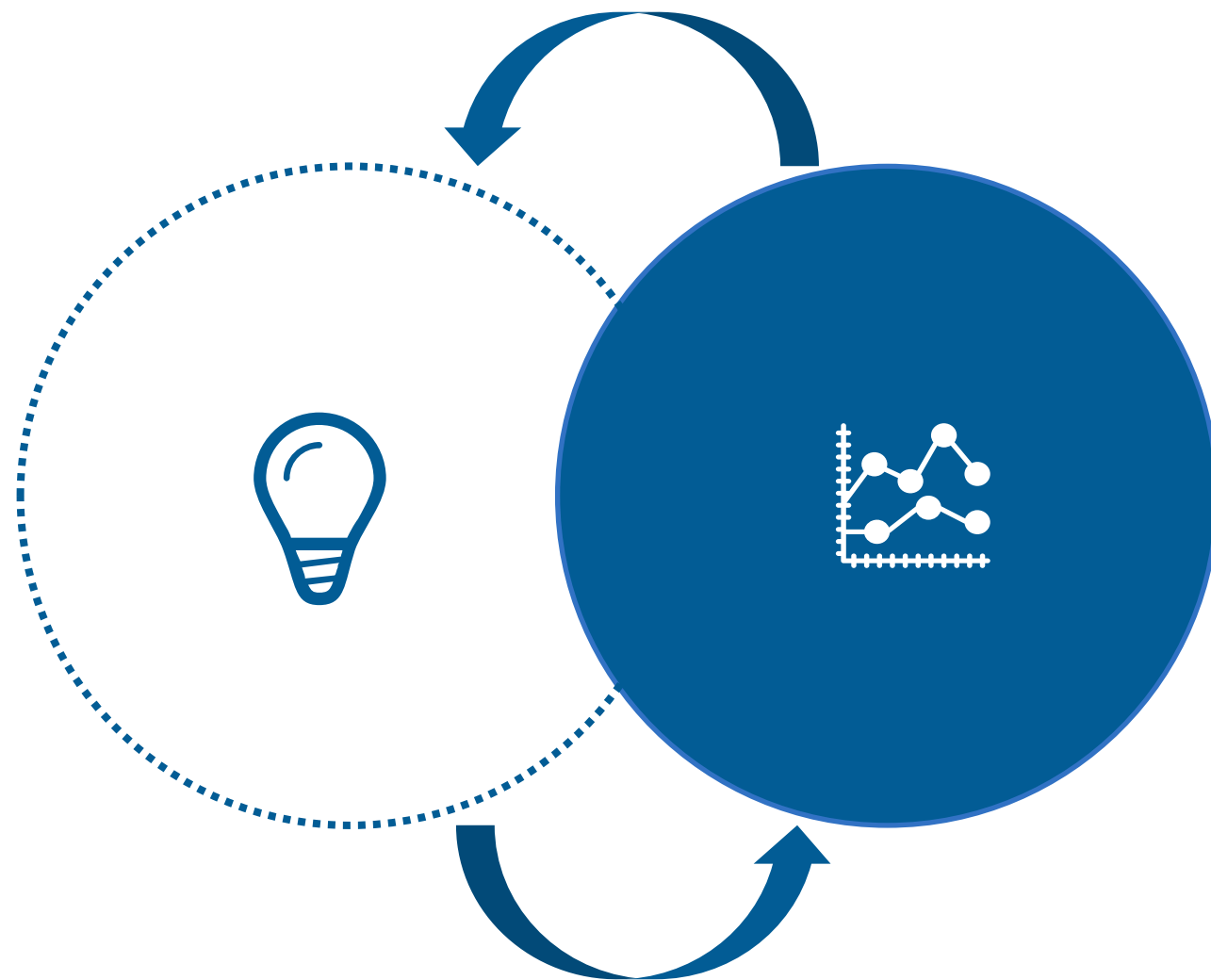
GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

1

Data analysis: for complex requirement

Set up a conjecture and make a data mining plan based on understanding of games



Use the mined data to verify the conjecture

1

Data product

- Data mining
- Data dashboard
- Data analysis

Summarize

Methodology

Still data

Data product

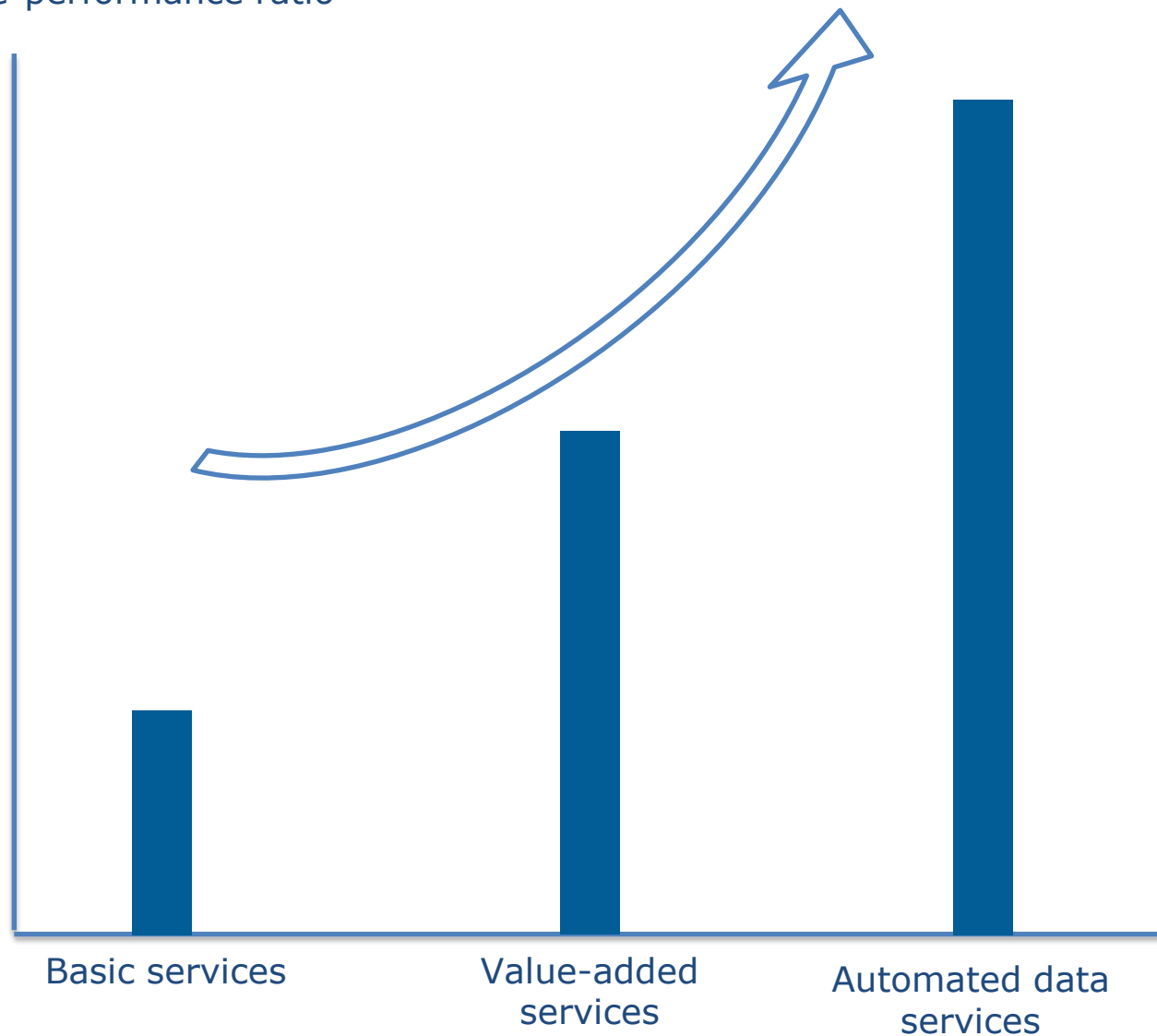


GDC

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

1 Data services

Price-performance ratio



Basic services

- Data mining
- Data dashboard

Value-added services

- Data analysis

Automated data services

- Data products/tools



GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

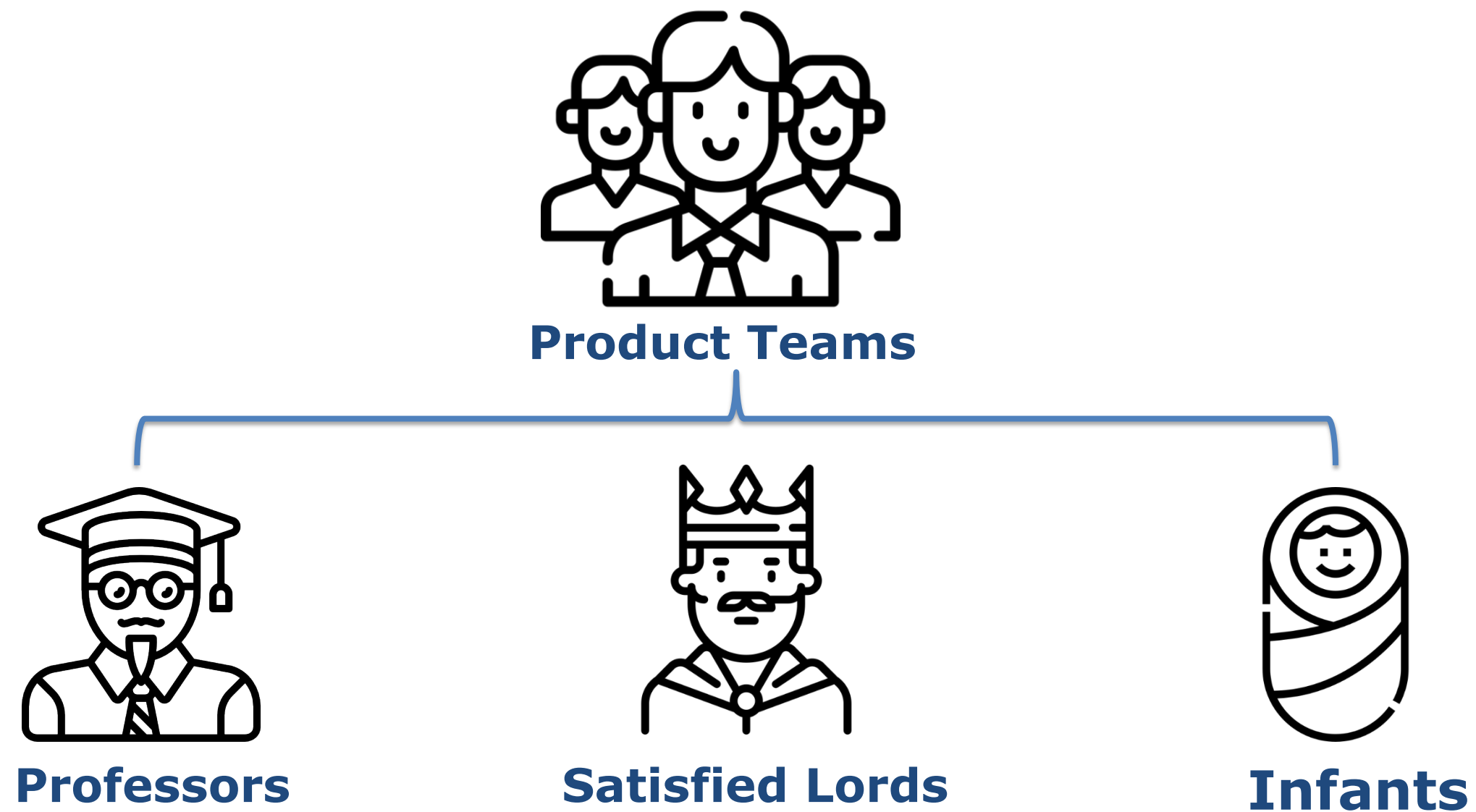
Part.2 Try to understand product teams



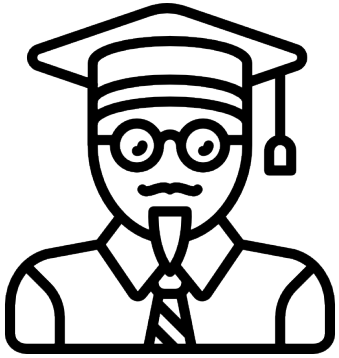
GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

Classification of different product teams



2 Professors teams

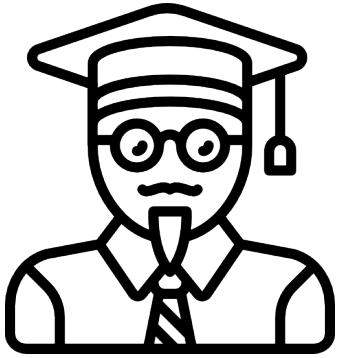


They can clearly understand the importance of data and describe accurate data demand and goals. All data development resources, such as logs, will be prepared in advance.

How to distinguish ?

- The Product team have cooperated with data team before ;
- The leader of the product team recognizes the value of data and hopes to use data to drive product iterations ;
- Game design will also be data-driven.

2 Professors teams



Not necessary to spend a lot of time to communicate. The whole procedure will be smooth. In the CBT period, we can do a good job to match their need of data dashboard.

- **Data dashboard** : Plan in advance
- **Data mining (Manual)** : Very few
- **Data mining (Self-Service)** : Strong data awareness
- **Data analysis** : Regular discussion
- **Data product** : Willing to try different data products

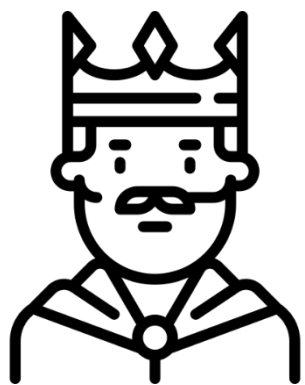


This kind of team may know the importance of data, but they will not take the initiative to propose data requirements.

How to distinguish ?

- High requirements for timeliness ;
- The willingness of spending time to communicate tends to be limited, but their demands on analysis results is usually strong ;
- Often raise requirements temporarily.





To work closely with such a team, it is very necessary to communicate regularly with the person in charge. Keep an eye on the game dynamics, make a first move, rather than being told. Always make a plan ahead.

- **Data dashboard** : Many temporary needs
- **Data mining (Manual)** : Large amount of requirement
- **Data mining (Self-Service)** : Unwilling to try it
- **Data analysis**: Occasionally discussion
- **Data product** : Unwilling to try it





The worst case happens when a team does not know the value of game data at all, and how to cooperate.

How to distinguish ?

- Brand-new product team ;
- They can hardly spend time to make more reasonable data requirements.





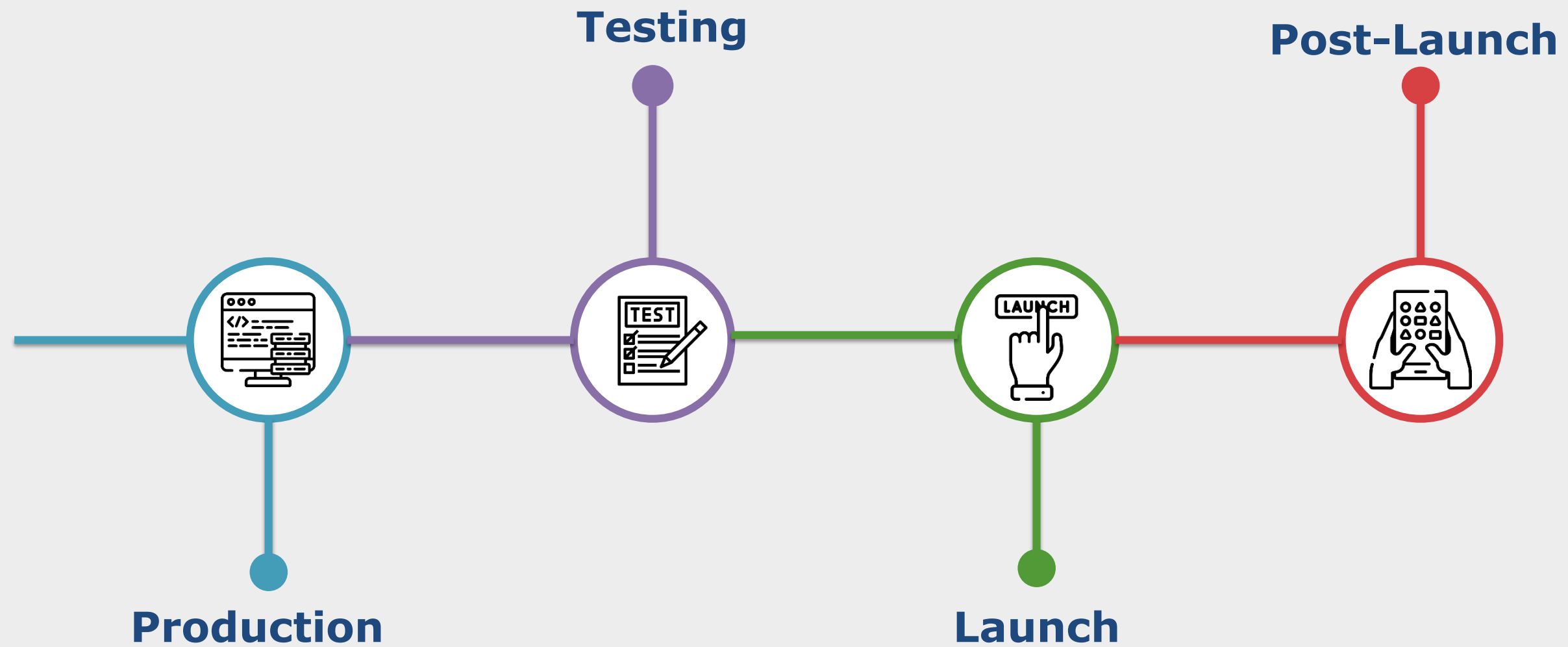
Generally, we help them to discover the value of data, and provide proper training. We tend to begin with basic data requirement, and lower complex requirements.

- **Data dashboard** : Standard module first
- **Data mining (Manual)** : Manual service is more appropriate
- **Data mining (Self-Service)** : User training first
- **Data analysis** : Need to interpret the data result
- **Data product** : Too early to provide



Part.3 Demand Management

Data service in 4 different game development period



Service

Launch

Testing

Post-Launch

Production

Communication

➤ Service

Data team can provide very limited support

➤ Communication

Confirm a unified logging standards

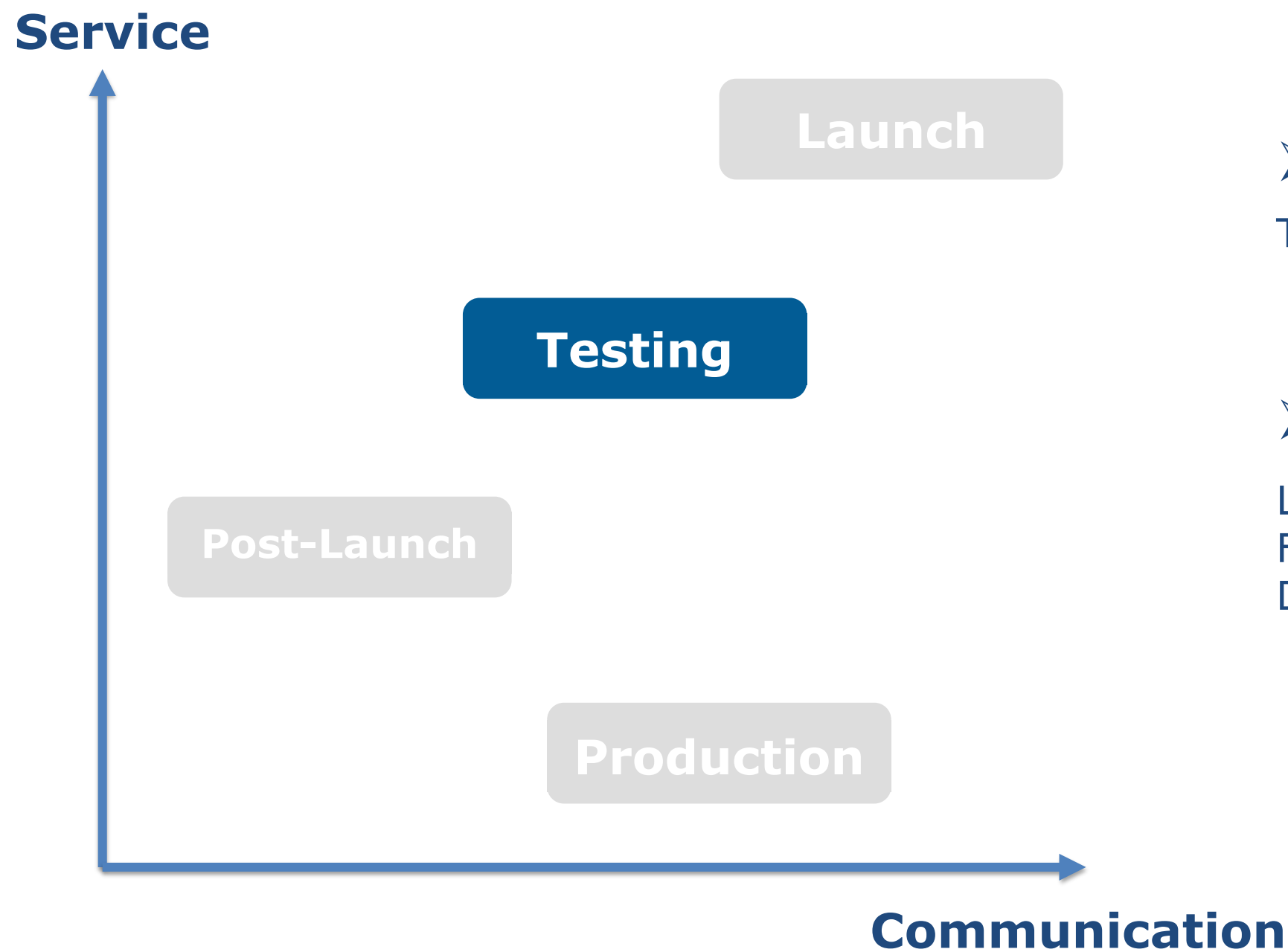
Planning BI system

Recommend data products



GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21



➤ Service

This is the first challenge for data services

➤ Communication

Log exception check
Feedback of data analysis and data mining
Discussion on supplementary demand



Service

Launch

Testing

Post-Launch

Production

Communication

➤ **Service**

The initial launch stage is the most stressful period for data service

➤ **Communication**

Communicate frequently to deal with various emergencies



GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

Service

Launch

Testing

Post-Launch

Production

Communication

➤ Service

Most of the requirements will be developed in the form of a weekly version

➤ Communication

Once a week, synchronize the development plan



GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

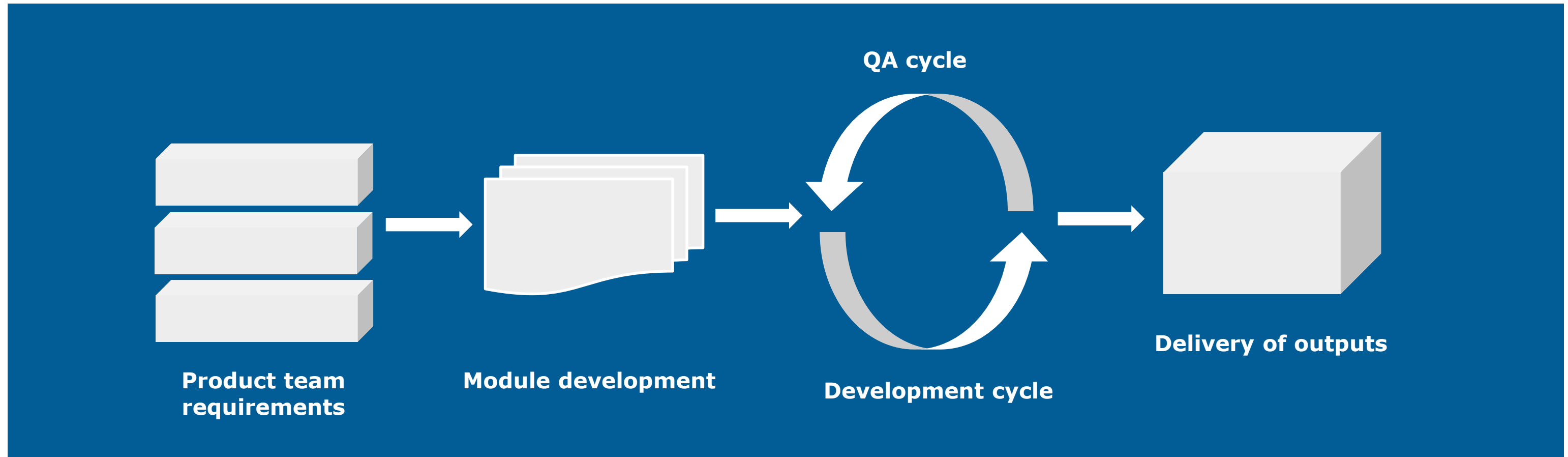
3 Data service process

Product teams : Describe data usage scenarios and functions

Data Product manager : Transfer data requirement into developing requirement

Data engineer : Complete requirement development

Quality Assurance: Ensure the correctness of data service



Part.4 Build a Data Team From Scratch

4 Build a Data Team From Scratch

Game Testing Cycles



4 Build a Data Team From Scratch

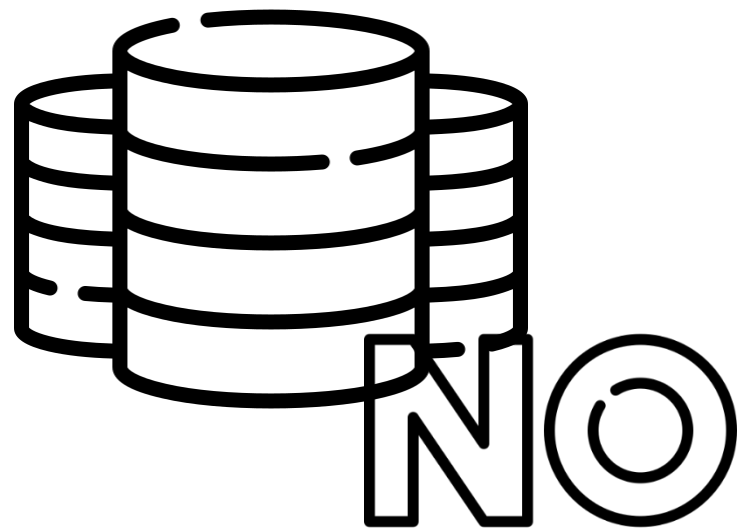
Game Testing Cycles

- Alpha – This is typically the earliest form of game testing and the most exclusive.
- Closed Beta Testing(CBT) – Closed beta testing precedes open beta. Closed beta tests are typically small in scale and often require beta keys.
- Open Beta Testing (OBT) – Open beta follows closed beta and is typically the final round of beta testing.
- Commercial Release – Commences upon the completion of open beta testing. Games in Commercial Release are expected to be much more stable and polished than beta.

4 Build a Data Team From Scratch

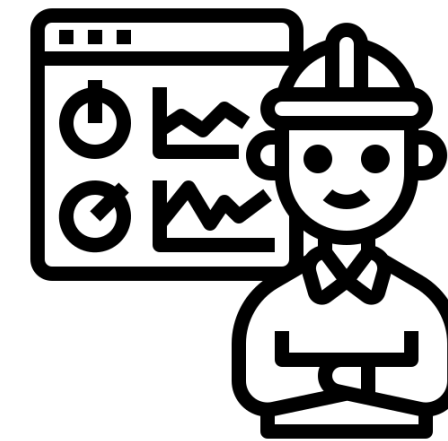
When the game is in the stage of Alpha testing or Closed Beta testing

Big Data?



All you need is

Python
Web developer



Data Processing

Data Visualization



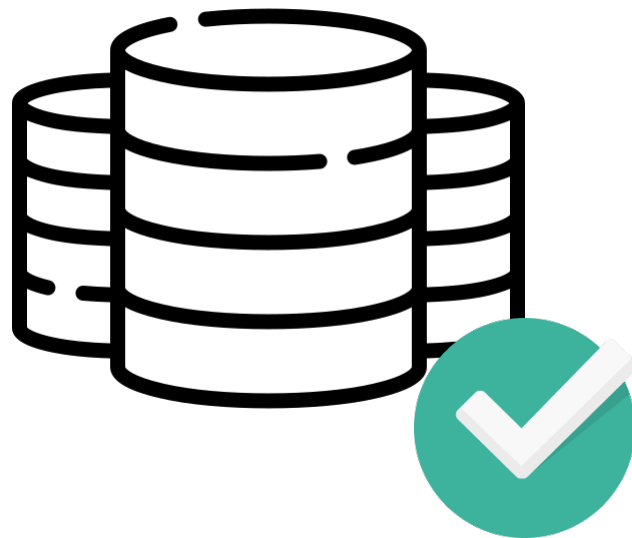
GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

4 Build a Data Team From Scratch

When the game is in the stage of Open Beta testing or Commercial Release

Big Data !



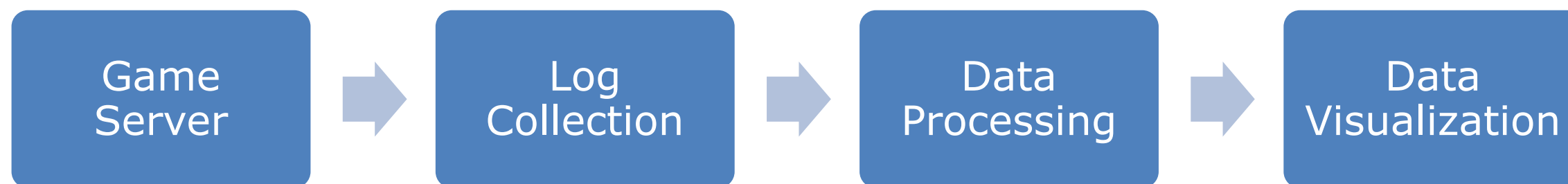
Now you need

Big Data developer



4 Build a Data Team From Scratch

Basic Data Pipeline

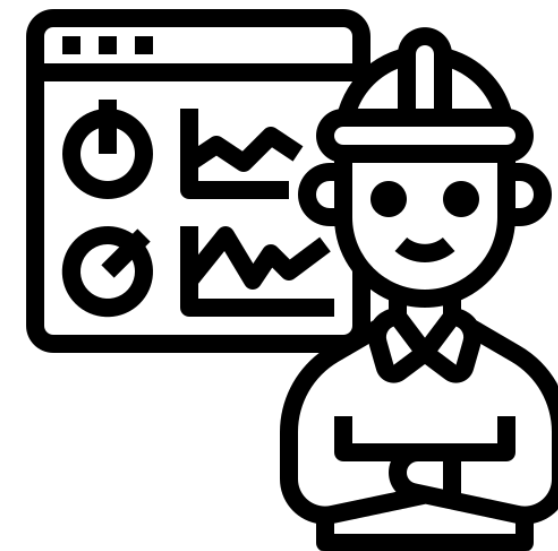


4 Build a Data Team From Scratch

Big Data developer



Web developer



Data Collection and Processing

Data Visualization



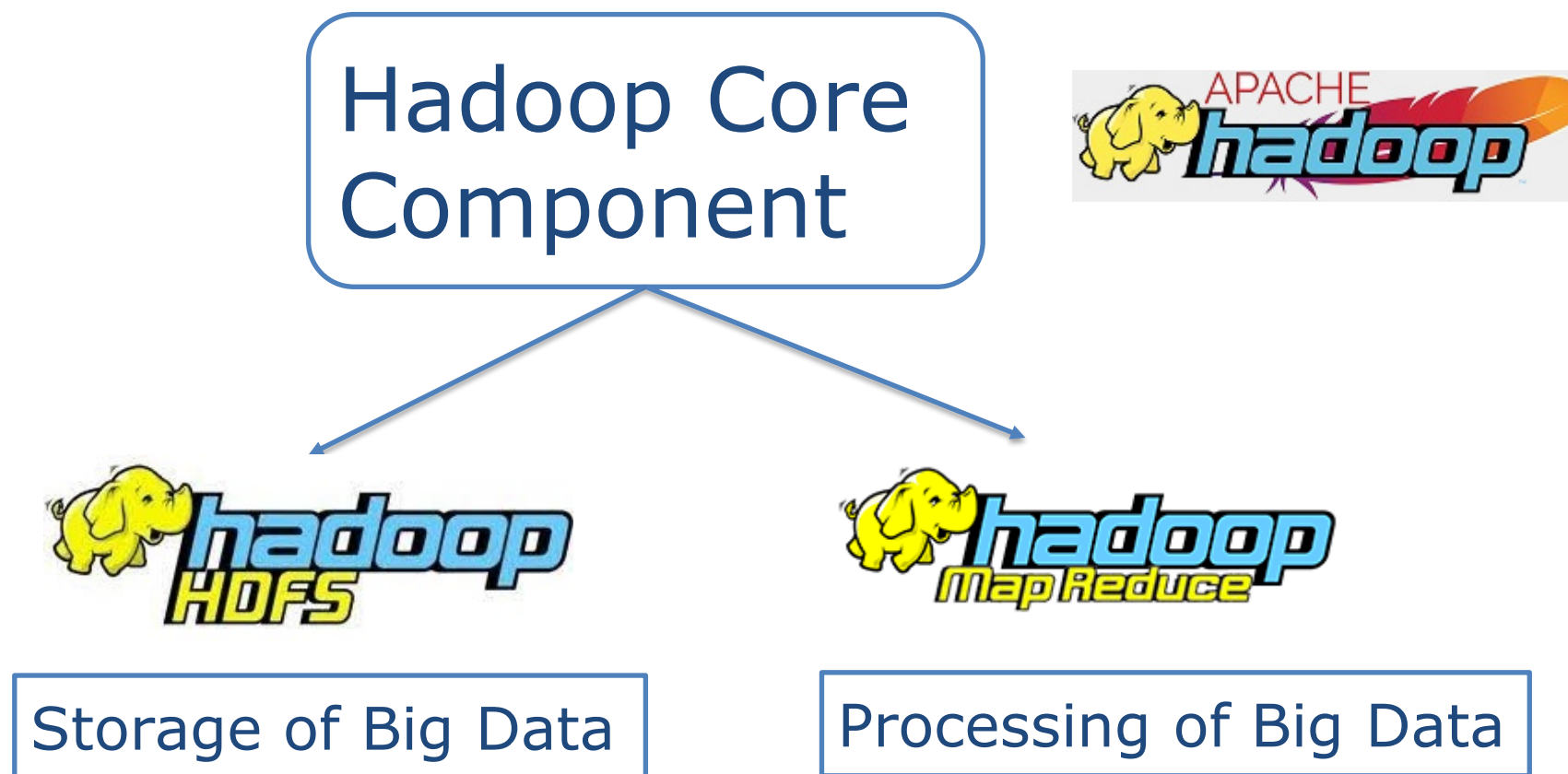
GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

Part.5 The Evolution of Data Pipeline

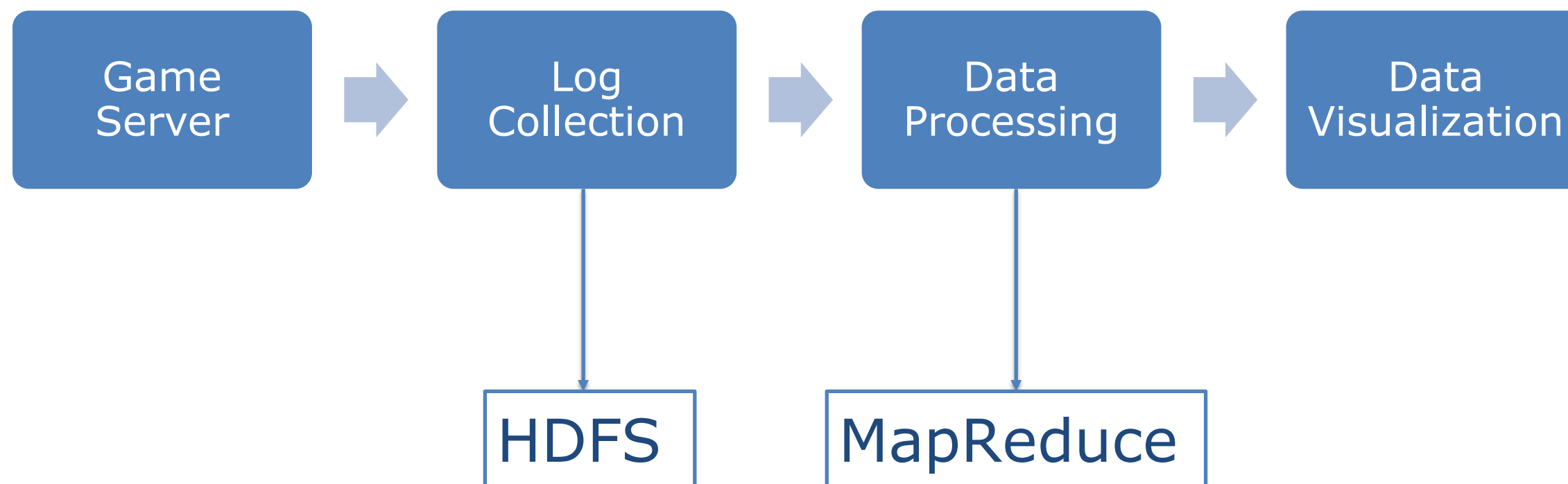
5 Build a Minimal Big Data System

When the data processing capacity exceeds the limit of a single machine



5 Build a Minimal Big Data System

Big Data Pipeline



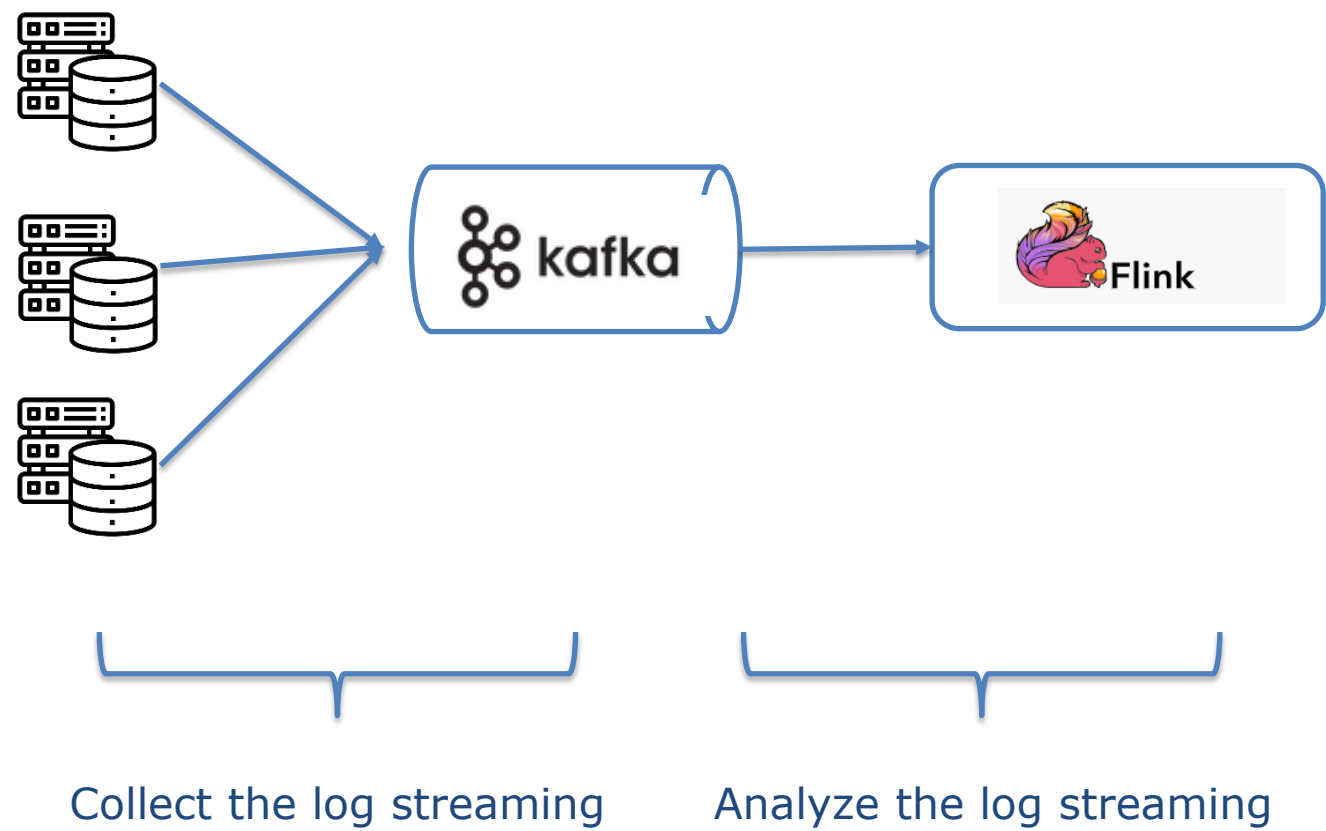
GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

5

Faster Decision Making: Real-Time Data Processing System

When we need to process the data immediately

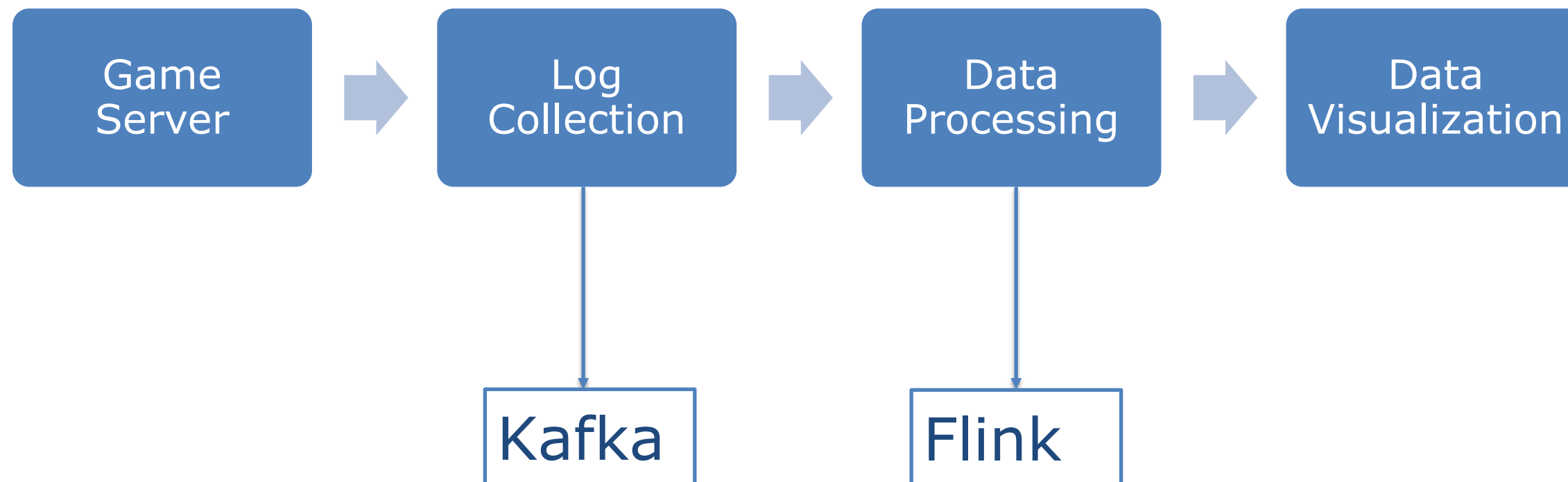


Session Name:	Online Game Technology Summit: Real-Time Data Processing for Multiplayer Online Games
Speaker(s):	Xia Lei, Zechen Xu
Company Name(s):	Netease, NetEase
Track / Format:	Online Game Technology Summit
Overview:	The data generated by MMO games is huge, and real-time data processing can maximize the value of data. Netease's Thunder Fire UX team has extensive experience in processing real-time large data in MMO games. Our talk will introduce the whole system structure of real-time data processing including data recording and collection, data transmission, and data storage. We will also introduce various open source tools of big data which are related to real-time computing, such as Logstash, Kafka, Flink, etc. Tools and processes constitute our real-time computing system. Meanwhile, we will demonstrate the real-time calculation of game data in the practical application of Netease games, which includes real-time monitoring of in-game currency flow, real-time detection of cheating players, real-time processing of chat data. Through this talk, the audience can understand the whole process of real-time processing of game data, be familiar with the big data technology we used, and apply real-time data processing of game data in their own games to improve the value of game data.

Online Game Technology Summit: Real-Time Data Processing for Multiplayer Online Games.

5 Faster Decision Making: Real-Time Data Processing System

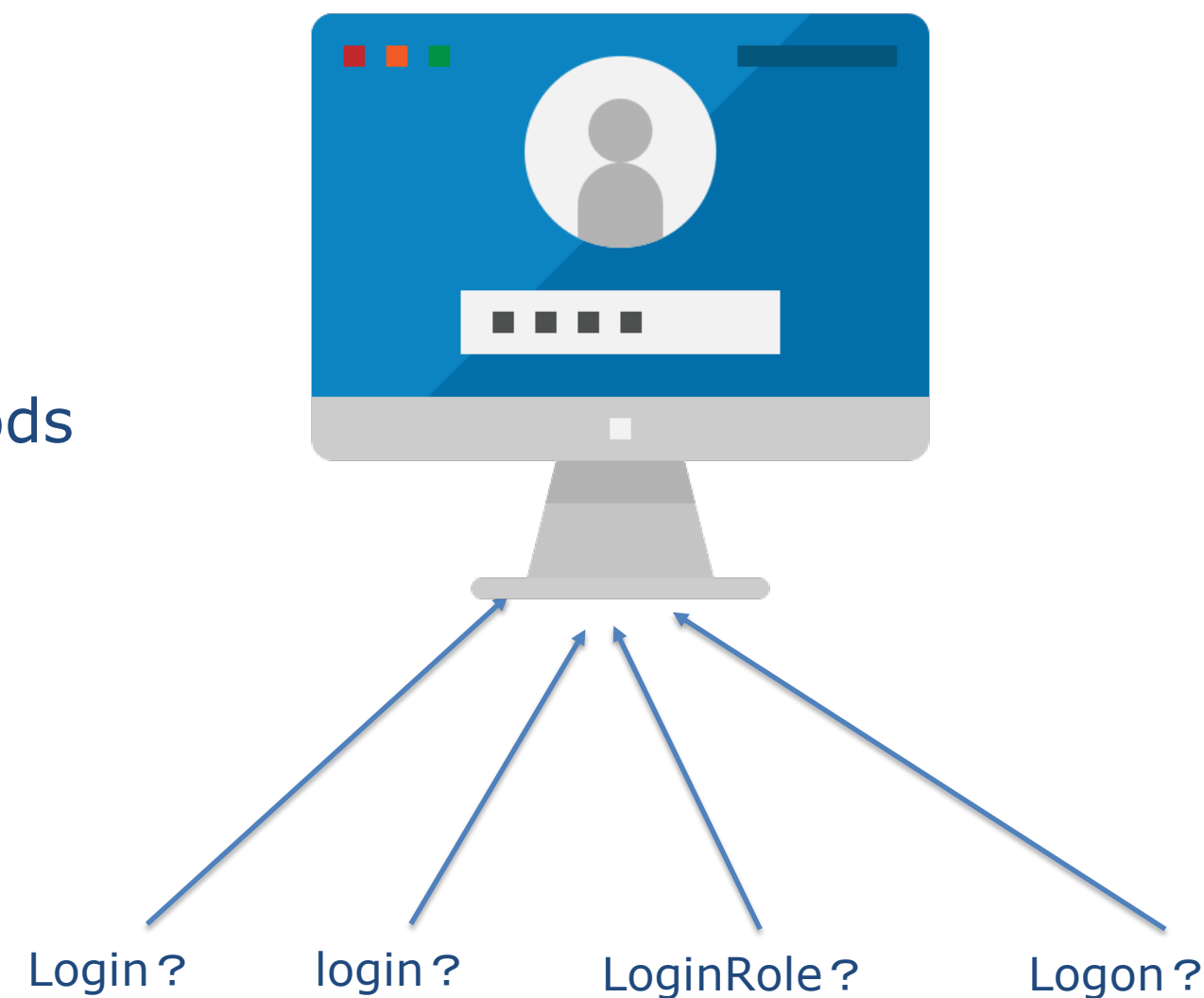
Real-Time Data Processing Pipeline



5 When more games developed

Problems we are facing

- Different Games, Different Logging Methods
- Different Games, Different Log Transmission Methods



GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

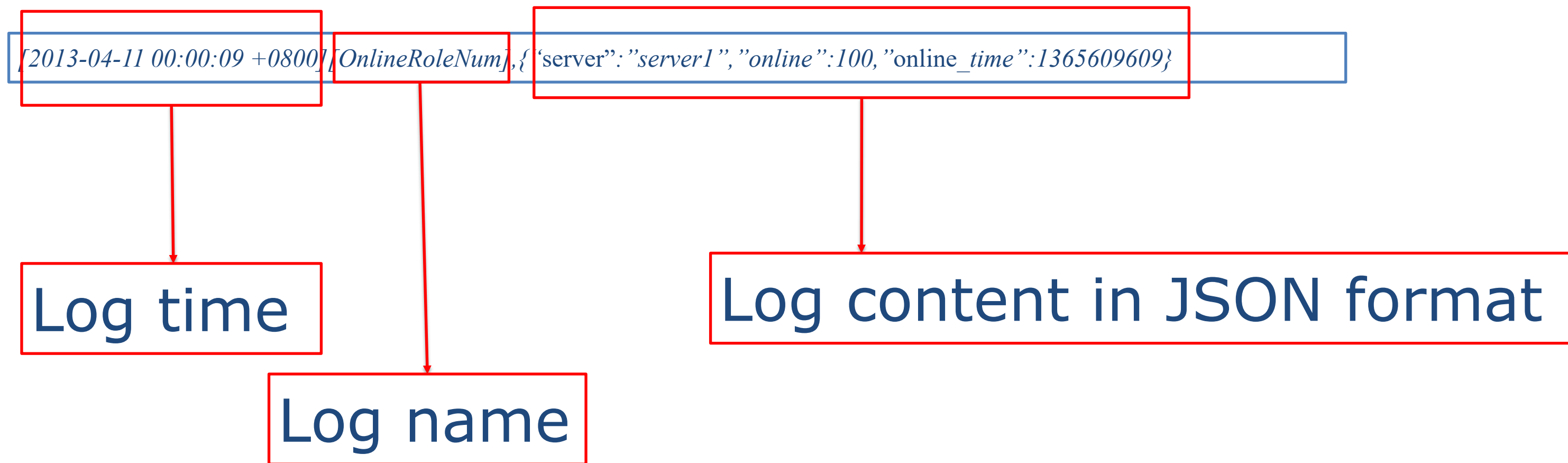
5 Create a unified Logging Standard

All Logs Share Same Format

Same Event in Different Games, Same Log Structure

5 Create a unified Logging Standard

All Logs Share Same Format



GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

5 Create a unified Logging Standard

Same Event in Different Games, Same Log Structure

```
[2013-04-11 00:00:09 +0800][OnlineRoleNum],{"server":"server1","online":100,"online_time":1365609609}
```

When recording the number of players online, always use the same log name and the same json keys

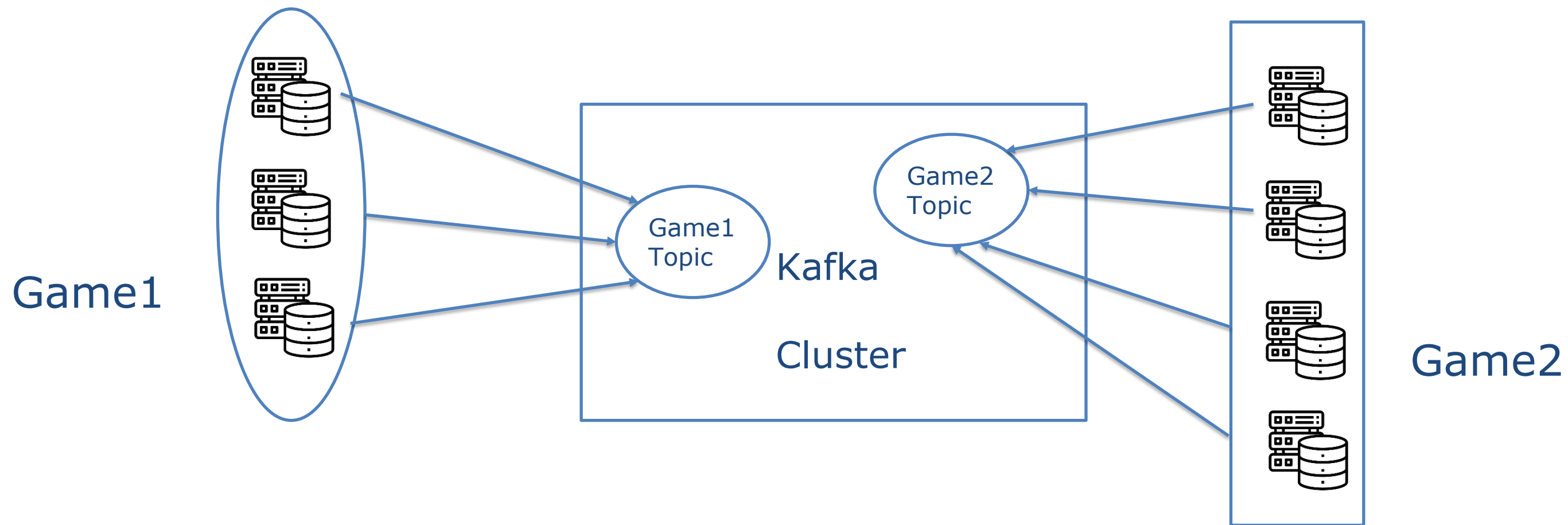


GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

5 Build a unified Log Processing System

All games use the same tool to collect logs, but the logs of different games are transmitted to different topics



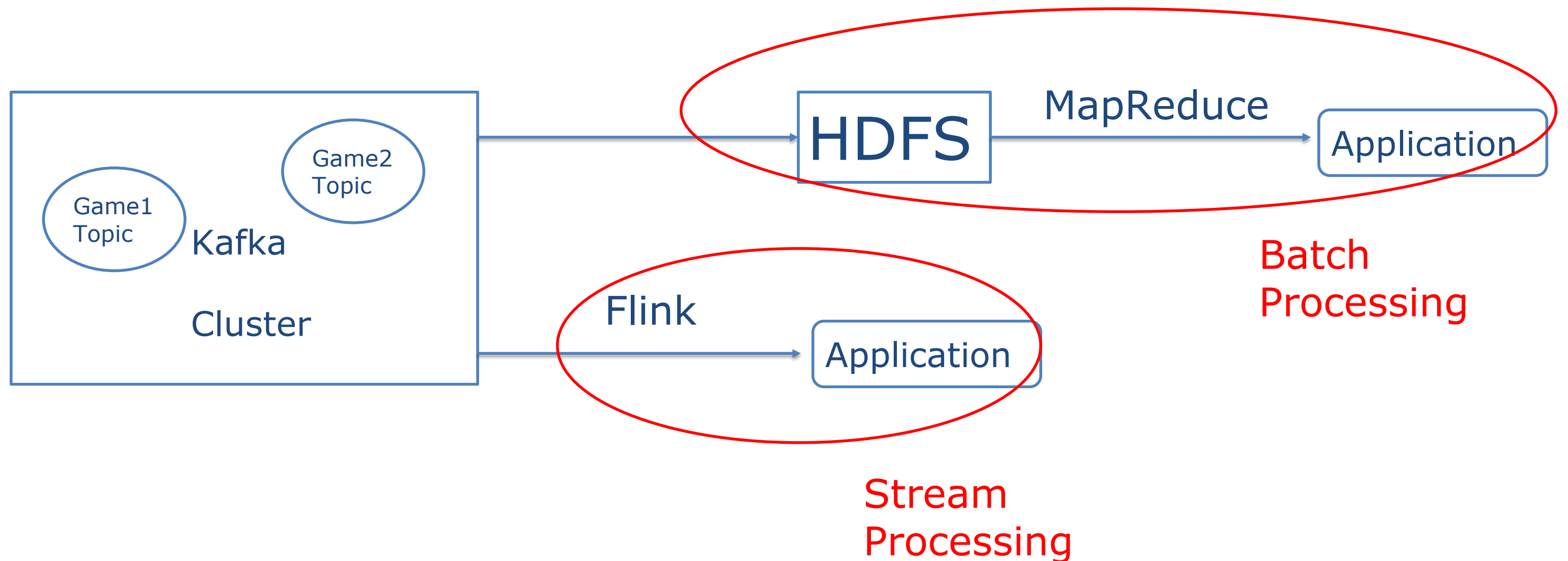
5 Build a unified Log Processing System

- Batch Processing
- Stream Processing

Batch Processing	Stream Processing
Batch processing is lengthy and is meant for large quantities of information that aren't time-sensitive	Stream processing is fast and is meant for information that's needed immediately

5 Build a unified Log Processing System

All the following tasks consume logs from the same Kafka cluster

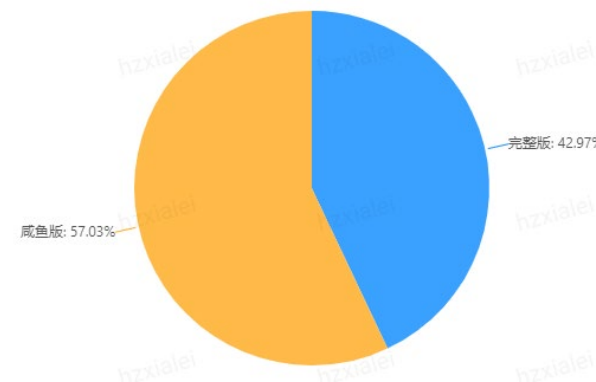


5 Build a low-code BI System

Most of frontend data visualization work can be configured with simple json

Reusable configuration

```
{  
  "chart": {  
    "name": "client_type",  
    "y": "ratio",  
    "pointFormat": "rate: <b>{point.y} %</b>",  
    "format": "{point.type}: {point.y} %",  
    "method": "pie"  
  }  
}
```



Development Efficiency Improved



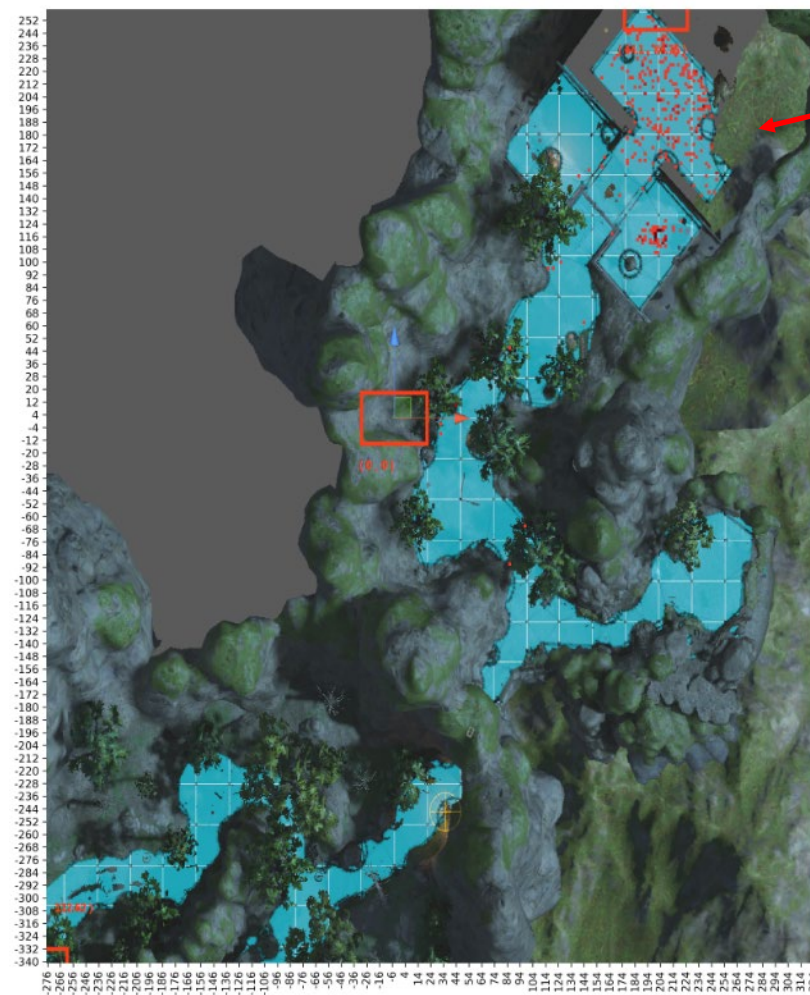
GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

5 Build a low-code BI System

Beyond the basic visualization work: we can do something more complicated

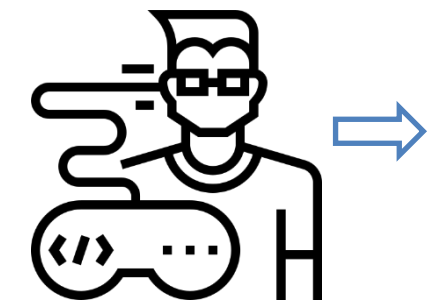
Heat map for player deaths



GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

5 Build a Self-Service BI System for Game Developer



I want to know those players who buy xx items last weekday whether buy yy items yesterday



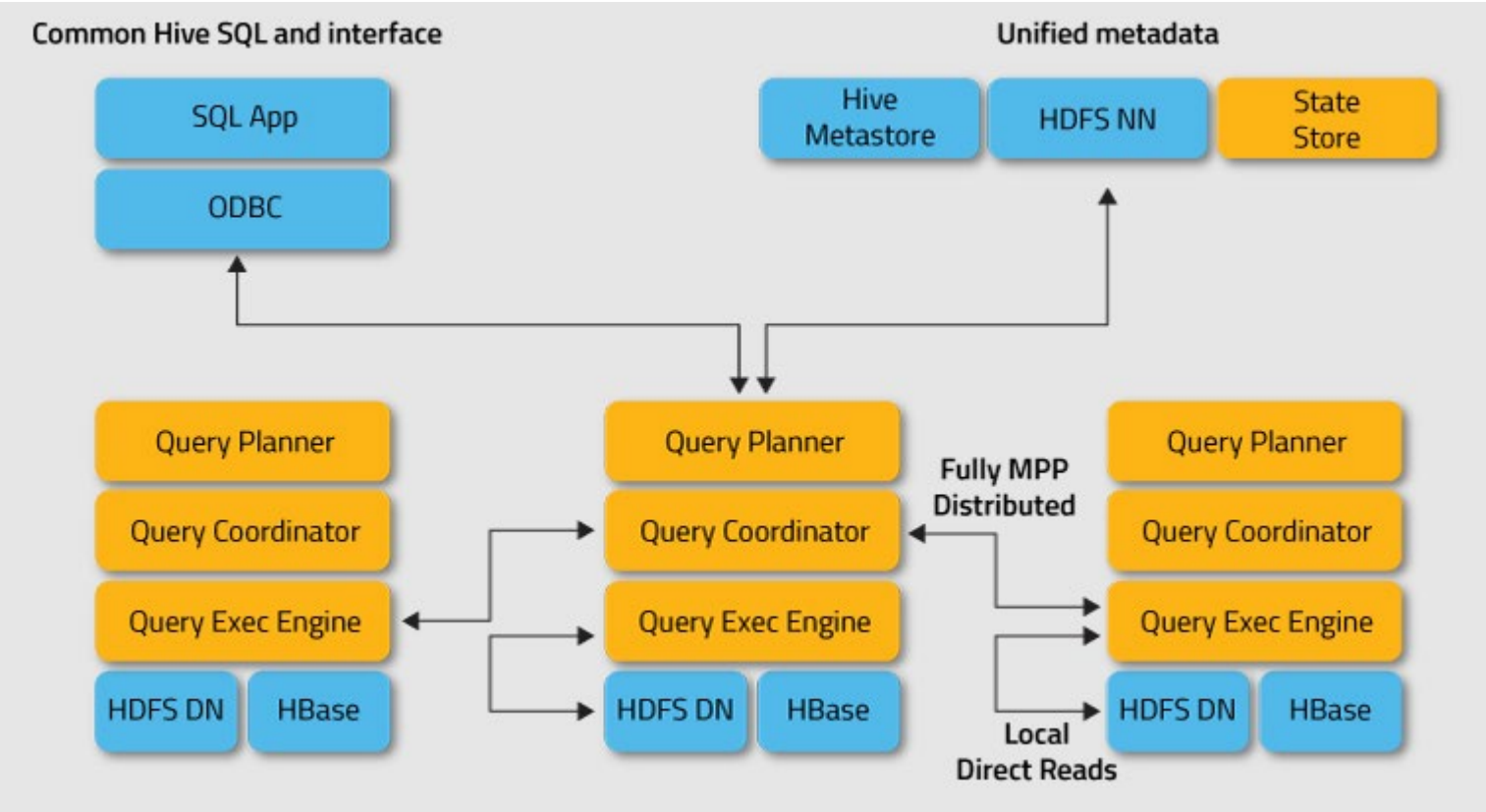
Parse these words to backend SQL queries



Send these SQL queries to ad-hoc query system

5

Ad-hoc query for big data—the massively parallel processing (MPP) system



5 Build a Self-Service BI System for Game Developer

Common query engines

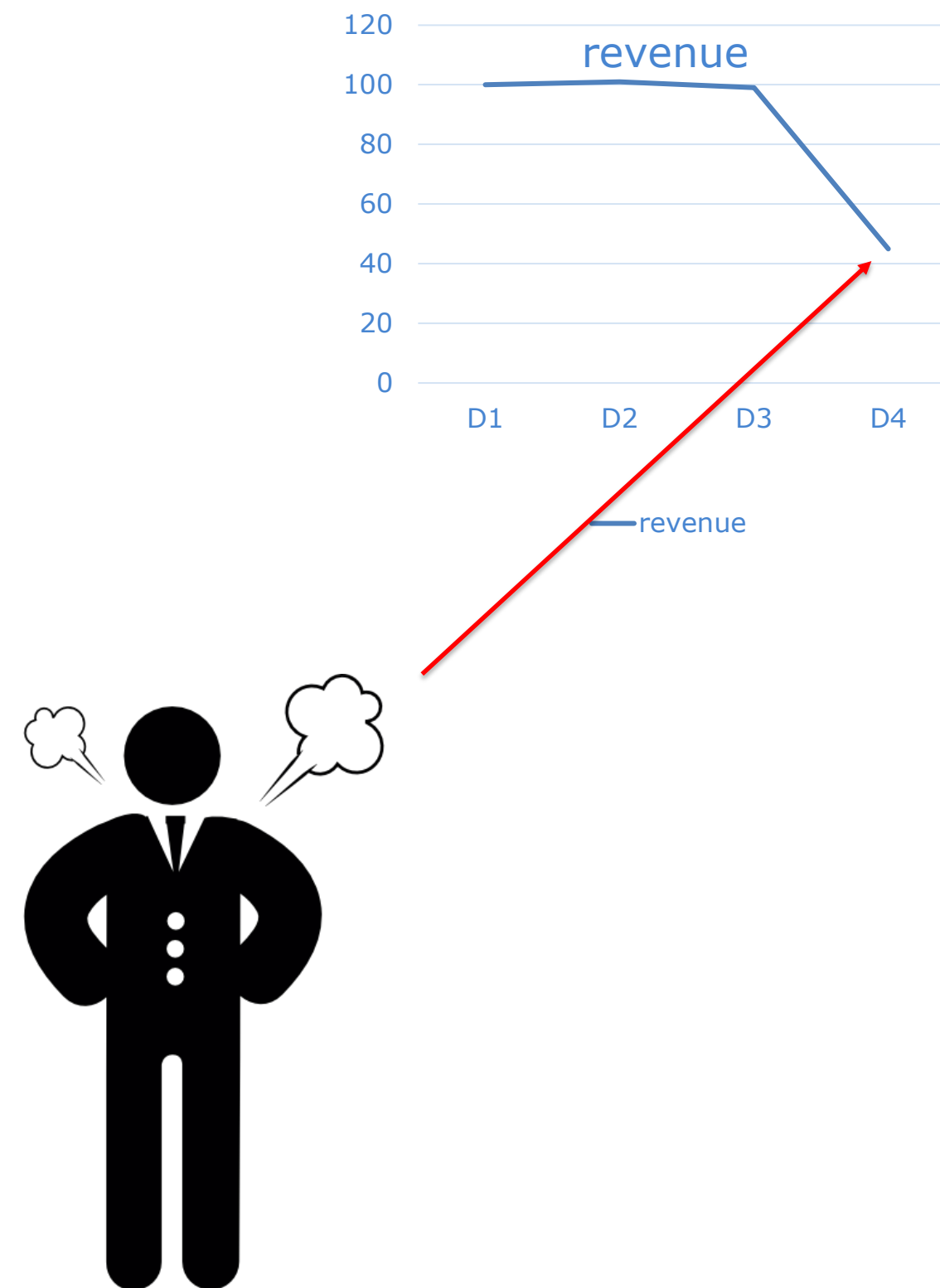
	Ad-Hoc	Join Support	Big Data	Query latency
Druid	✓		✓	Millisecond
Impala	✓	✓	✓	Second
Presto	✓	✓	✓	Second
ClickHouse	✓	weak	✓	Millisecond

Part.6 Quality Assurance for Data Pipeline

6 Quality Assurance for Data Pipeline

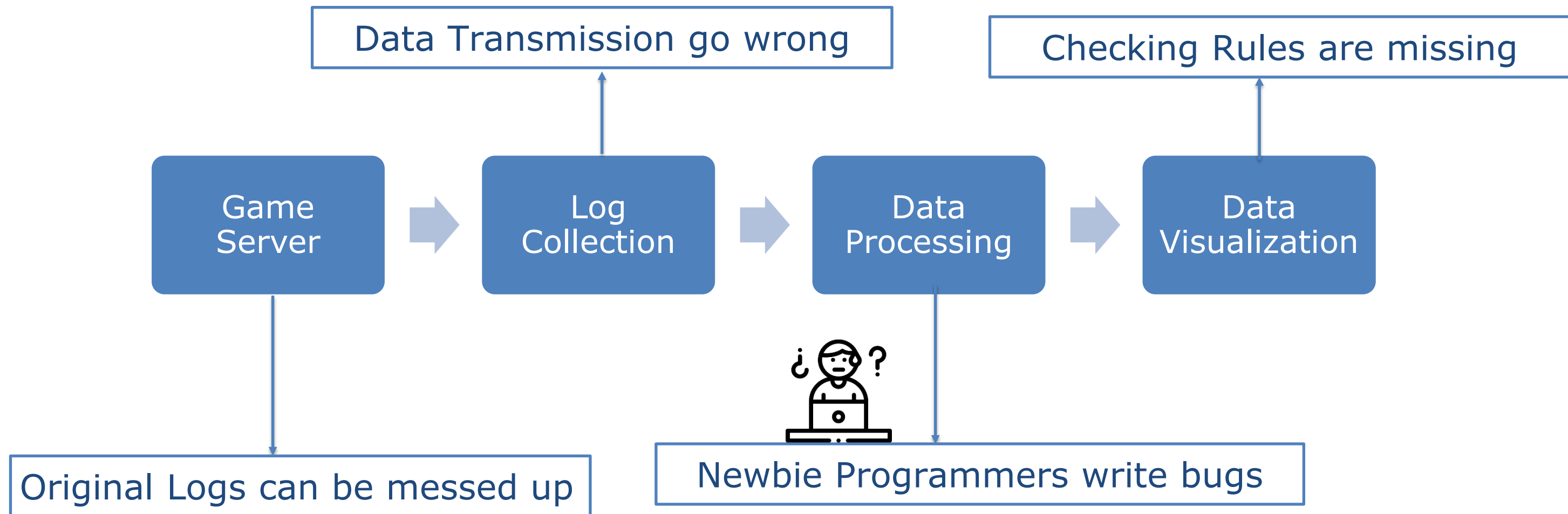
Why do we need QA for the data pipeline?

Imagine the scenario:
When the game producer gets a wrong number about game revenue



6 Quality Assurance for Data Pipeline

Let's review the data pipeline and check anything that can go wrong



6 Quality Assurance for Data Pipeline

List some possible problems with logging

- Log A only records Android information, and iOS information is ignored
- In the global launch, the currency information was omitted
- Forgot to record the time zone information



GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

6 Quality Assurance for Data Pipeline

Developed a log verification tool

Verify and reset

Input the log

校验形式 日志格式

日志格式

```
{ "menus":0,"diamond":0,"exp":0,"gold":1000,"power":0,"grade":0,"vip":0,"exp":0,"fight_num_month":140,"totalBePraiseNum":1,"money_sum":0,"axis":(0,0,0),"role_id":"127418101","online_time":1287,"role_name":"magic","last_operation":1,"fight_num_day":28,"latitude":0,"server":"8101","MaxFightNumMonth":140,"transid":"","ipv6":"","10.74.25.227","scene":"","country":"CN_CN","client_source":"app","app_channel":"netease","Appsflyer_id":"","is_emulator":"","TotalGoldNum":395,"os_ver":"Android OS 9 / API-28 (PKQ1.180904.001/V11.0.3.0.PDCCNXM)","logout_time":1578366000,"network":"4g","location":{"longitude":0,"latitude":0},"old_accountid":"","ip":"10.74.25.227","device_height":1080}
```

校验 重置

校验结果: 未通过

未提供校验的关键P1日志: ['YuanbaoGain', 'OnlineRoleNum', 'LoginUI', 'Update', 'UserCertification', 'Load', 'ItemBuy', 'LoginRole', 'CreateRole', 'YuanbaoUse', 'Identification', 'Prepaid', 'Tutorial', 'Activation']

错误等级	日志名称	信息
ERROR	LogoutRole	日志缺少必要字段: {'oaid', 'engine_ver'}
ERROR	LogoutRole	字段类型不符合规范: ['is_root:<class 'str'>']
ERROR	LogoutRole	字段类型不符合规范: ['last_operation:<class 'int'>']
ERROR	LogoutRole	字段类型不符合规范: ['is_emulator:<class 'str'>']
ERROR	LogoutRole	请使用外网IP
WARNING	LogoutRole	确认是否使用了正确的account_id, 参考《网易游戏手游通用日志格式》

不规范日志:

```
/srv/app/p1/p1-mn1g8101app3-2020010703.log [2020-01-07 03:00:00 +0000][LogoutRole],{"udid":"e359373dd191c64537b26234569a899f","Ads_id":-1,"account_id":"2623","device_width":2340,"can_track":1,"vip_level":0,"isp":"CHN-CT","MonthlyCardDayNum":0,"ChargeDollarTotalNum":0,"aid":2623,"is_root":"","app_ver":"46861p0","VipLevel":0,"MaxFightNumDay":12,"unisdsk_deviceid":"","item_sum":0,"role_enter_sn":"","exp_sum":0,"TotalExpNum":18,"FirstRechargeInfo":0,"mac_addr":"","ff.ff.ff.ff.ff","create_time":1578364231,"longitude":0,"device_model":"1#OPPO PBCM10#8#8","country_code":"CN","gold_num":793,"imei":"e359373dd191c64537b26234569a899f","os_name":"Android OS 8.1.0 / API-27 (OPM1.171019.011/1576243247)","role_level":2,"h5_channel":"","player_message":{"friends":0,"diamond":0,"exp":0,"gold":793,"power":0,"grade":2,"vip":0,"exp":6,"fight_num_month":60,"TotalBePraiseNum":0,"money_sum":0,"axis":(0,0,0),"role_id":"123668101","online_time":1766,"role_name":"冷雁","last_operation":-1,"fight_num_day":12,"latitude":0,"server":"8101","MaxFightNumMonth":60,"transid":"","ipv6":"","192.168.0.2","scene":"","country":"CN_CN","client_source":"app","app_channel":"netease","Appsflyer_id":"","is_emulator":"","TotalGoldNum":248,"os_ver":"Android OS 8.1.0 / API-27 (OPM1.171019.011/1576243247)","logout_time":1578366000,"network":"wifi","location":{"longitude":0,"latitude":0},"old_accountid":"","ip":"192.168.0.2","device_height":1080}
```

日志参考:

```
[2016-01-20 17:05:07 +0800][LogoutRole],{"ip":"218.107.55.253","ipv6":"","device_model":"oneplus a0001","device_height":1920,"device_width":1080,"os_name":"android os","os_ver":"4.3 / api-18 (jls36c/eng.oppo.20140811.134829)","mac_addr":"c0:ee:fb:06:32:2d","udid":"44d78adf1000f6f2","isp":"46007","network":"wifi","app_channel":"netease","app_ver":"fp2_03","server":"bdss","account_id":"aebf1e3cdibuvjp2@ad.netease.win.163.com","old_accountid":"aebf1e3cdibuvjp2@ad.netease.win.163.com","role_id":"1108063","role_name":"索尼卡斯帕","role_level":0,"create_time":1448266141,"vip_level":2,"exp":0,"logout_time":1453280707,"online_time":868,"scene":"","axis":"","money_sum":0,"exp_sum":0,"item_sum":0,"is_root":0,"is_emulator":0,"transid":"F1B7E893-1E34-4A5E-AE69-107DC9F2EFE0_1586318369779_375218232","unisdsk_deviceid":"F1B7E893-1E34-4A5E-AE69-107DC9F2EFE0","oaid":"ab2da4f8-0f26-4b15-be48-98c627e14f86","role_enter_sn":"","engine_ver":"","last_operation":"","imei":""}
```

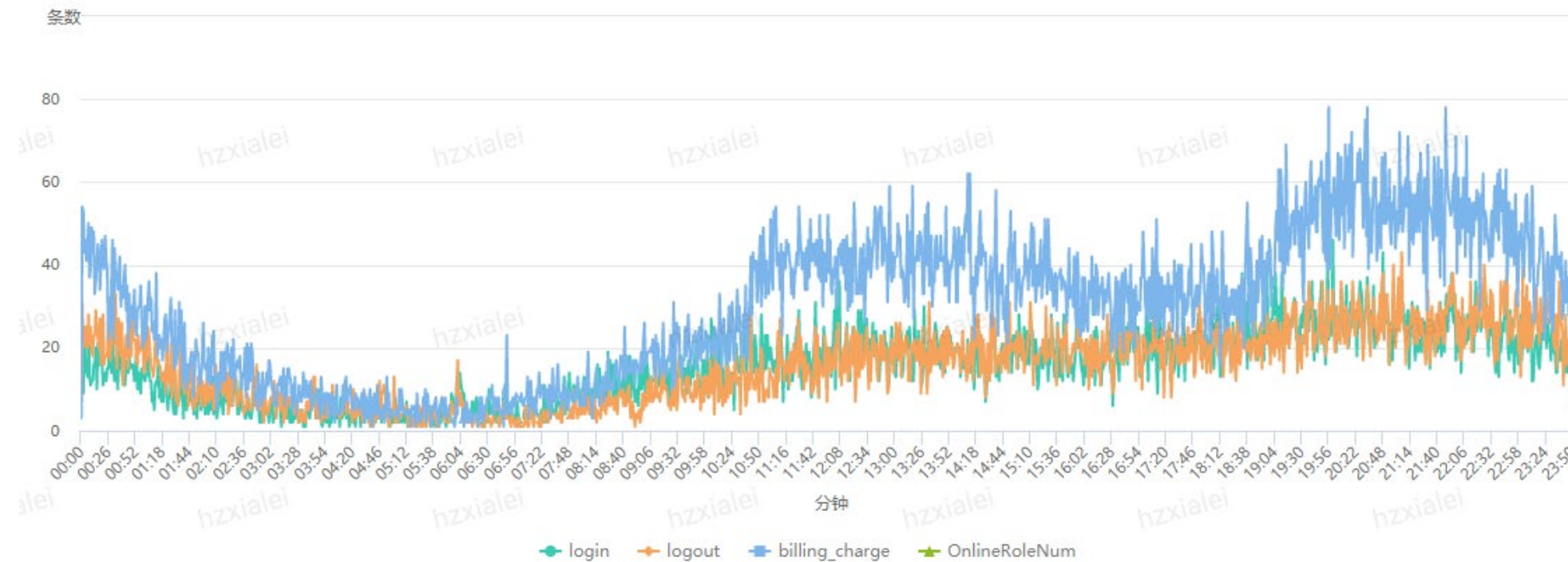
Some fields is missing

Wrong field type

Standard Format for reference

6 Quality Assurance for Data Pipeline

Monitor the log transmission process



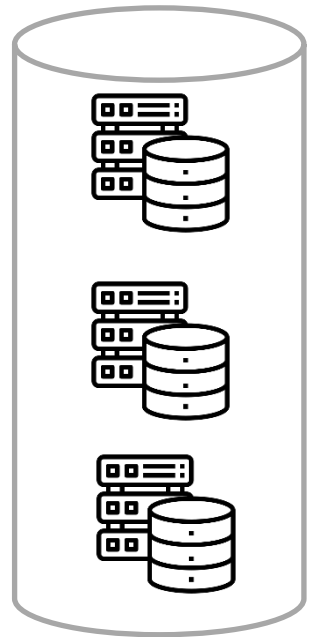
GDC

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

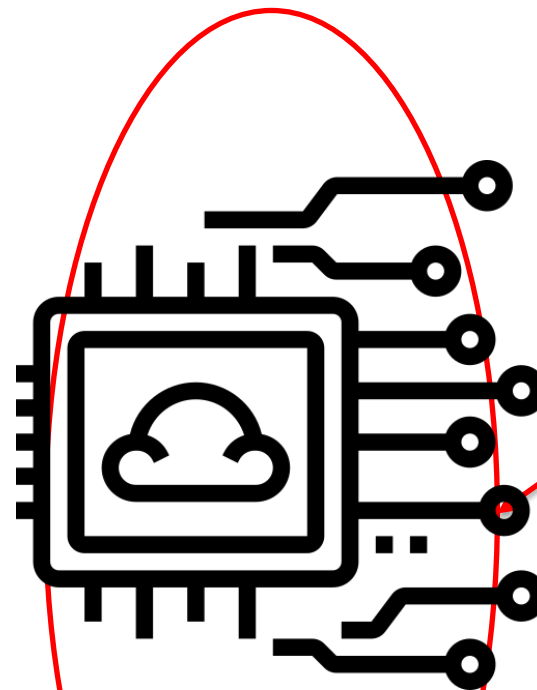
6 Quality Assurance for Data Pipeline

Develop an end-to-end data processing system

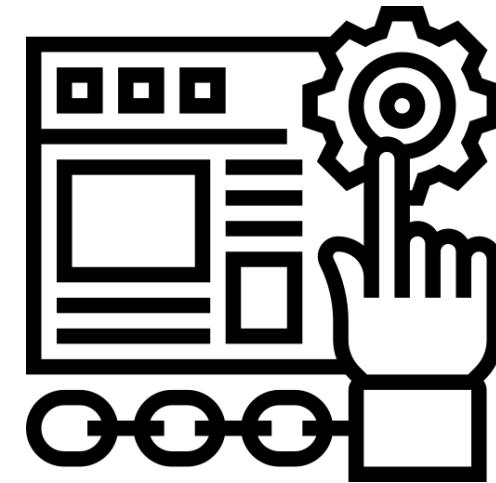
Developers do not need to pay attention to technical details



Standard Log



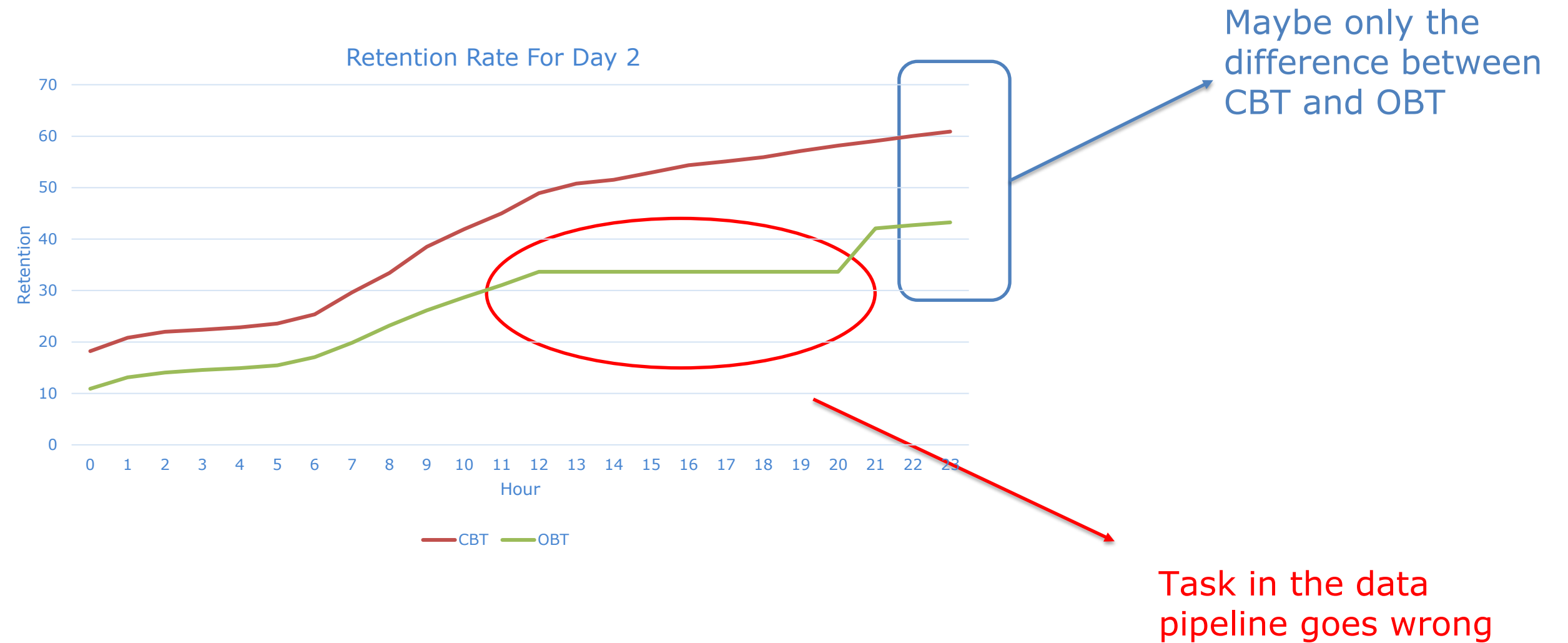
Standard Processing



Standard Application

6 Quality Assurance for Data Pipeline

Examine the result directly



GDC

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

Takeaway

- Get to know the data pipeline and quality assurance for data pipeline
- Come to understand product teams before providing services
- Reduce human resource consumption and provide automated tools to increase the service value and efficiency of the data team

Thanks for your attention!