

# Elastic Stack 을 활용한 Data Dashboard 만들기

Week 1 - Data를 시각화해보자



Fast Campus

# 목차

---

- 용어정리	14
- Elastic stack workflow	15
- Kibana workflow	17
- <u>Index 등록</u>	19
- <u>데이터 탐색</u>	22
- <u>Visualize workflow</u>	25
- <u>Dashboard</u>	62
- Aggregation	
- <u>Metric Aggregation</u>	27
- <u>Bucket Aggregation</u>	32
- <u>Parent Pipeline Aggregation</u>	37
- <u>Sibling Pipeline Aggregation</u>	48

소개

## 강의 소개 - 목표 및 방향

강의가 끝나면	data가 주어지면 dashboard를 구축하고 needs에 맞게 운영할 수 있다.
단,	모든 기능을 100% 마스터 하기보단 dashboard 구축 및 운영을 위한 전반적인 내용과 문제가 생길 시 troubleshoot 하는 방법을 배운다.
그래서	이론 위주가 아니라 실습 위주의 강의가 될 것이다.
그러므로	<ul style="list-style-type: none"><li>• 검색엔진으로서 Elasticsearch</li><li>• Elasticsearch Architecture      <b>는 다루지 않을 것이다.</b></li><li>• (고급) query 및 query 최적화</li></ul>

## 강의 소개 - 운영 방식

<b><u>FAQ</u></b>	자주 물어보는 질문 정리
<b><u>WIKI</u></b>	Elastic Stack 간단한 사용법 정리
<b>Questions</b>	<a href="#">Facebook {Elastic Stack Data Dashboard CAMP} Group</a>
<b>Online Source</b>	<a href="#">Elastic</a> , <a href="#">Facebook Elasticsearch Korea Group</a> , <a href="#">Stack Overflow</a>

## 강의 소개 - 주별 커리큘럼 (1주)

---

- Elastic Stack Work flow 이해하기
- Kibana Work flow 이해하기
- Visualize Work flow 이해하기
- Aggregation 이해하기

WEEK 1

---

## 강의 소개 - 주별 커리큘럼 (2주)

---

- Aggregation 복습 및 (정말 아주 조금 만 더) 깊게 이해하기
- (Visualize) JSON Input으로 Aggregation Parameter 변경해보기
- Managing Field에서 Data Format 변경해보기
- Lucene Query Syntax 이해하기
- Search & Filter 적용해보기
- Script Field 사용해보기

WEEK 2

## 강의 소개 - 주별 커리큘럼 (3주)

---

- (기본적인) Elasticsearch Document API 이해하고 적용해보기
- (기본적인) Elasticsearch Search API 이해하고 검색해보기
- Mapping 이해하고 생성해보기
- AWS에서 Elastic Stack 설치해보기



WEEK 3

## 강의 소개 - 주별 커리큘럼 (4주)

---

- Logstash 이해하기
  - 다양한 Input Source로부터 데이터 전송해보기
  - 다양한 Filter를 적용해서 원하는 형태로 (Elasticsearch에) 데이터 넣어보기
- 

## WEEK 4



## 강의 소개 - 주별 커리큘럼 (5주)

- Sample Dashboard와 똑같은 Dashboard 구축해보기 (Elastic Stack 설치부터 Dashboard 구축까지)
- 공통의 Sample Data로 자신의 만의 Dashboard 구축해보기 (Mapping부터 Dashboard 구축까지)

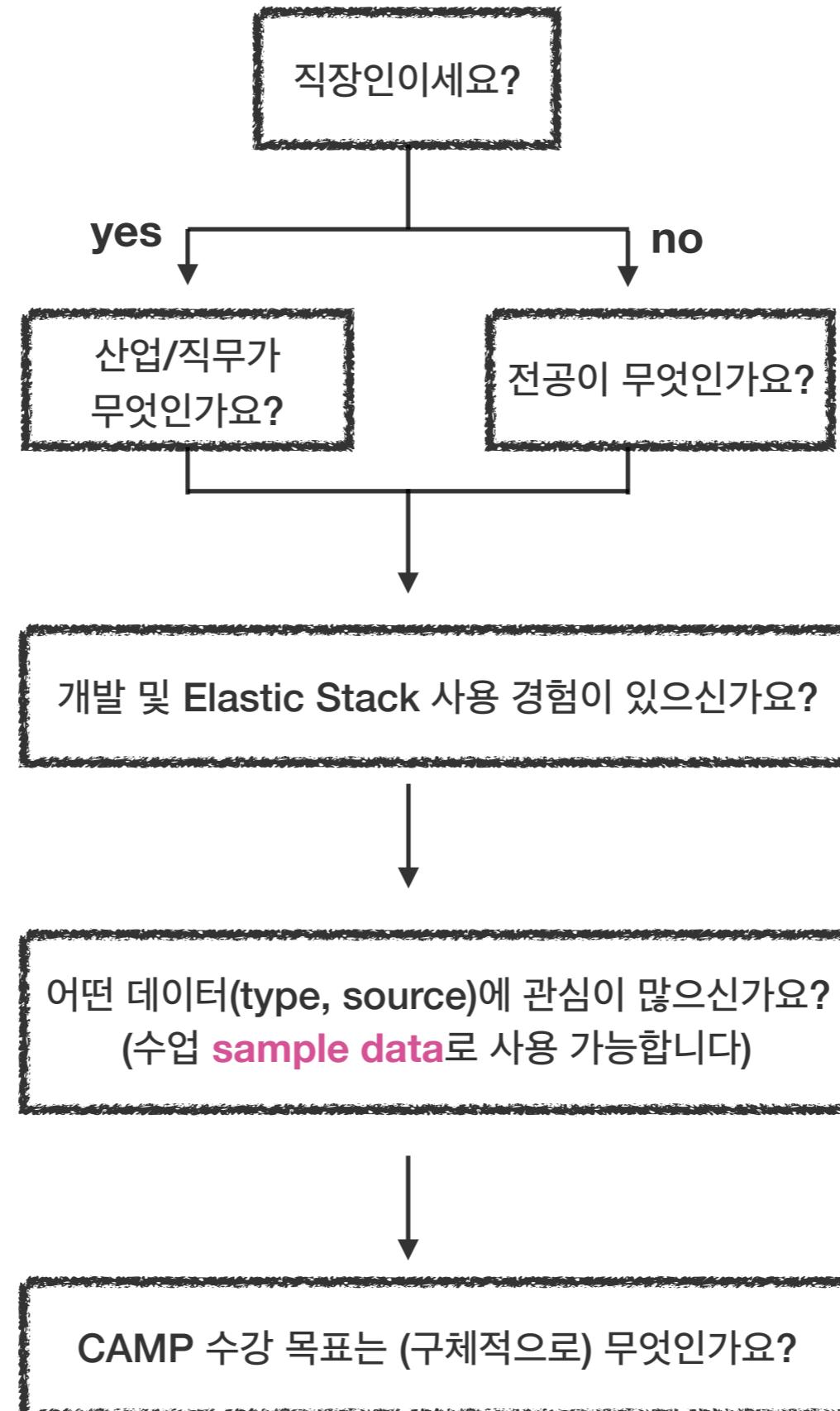
WEEK 1

WEEK 2

WEEK 3

WEEK 4

## 수강생 소개

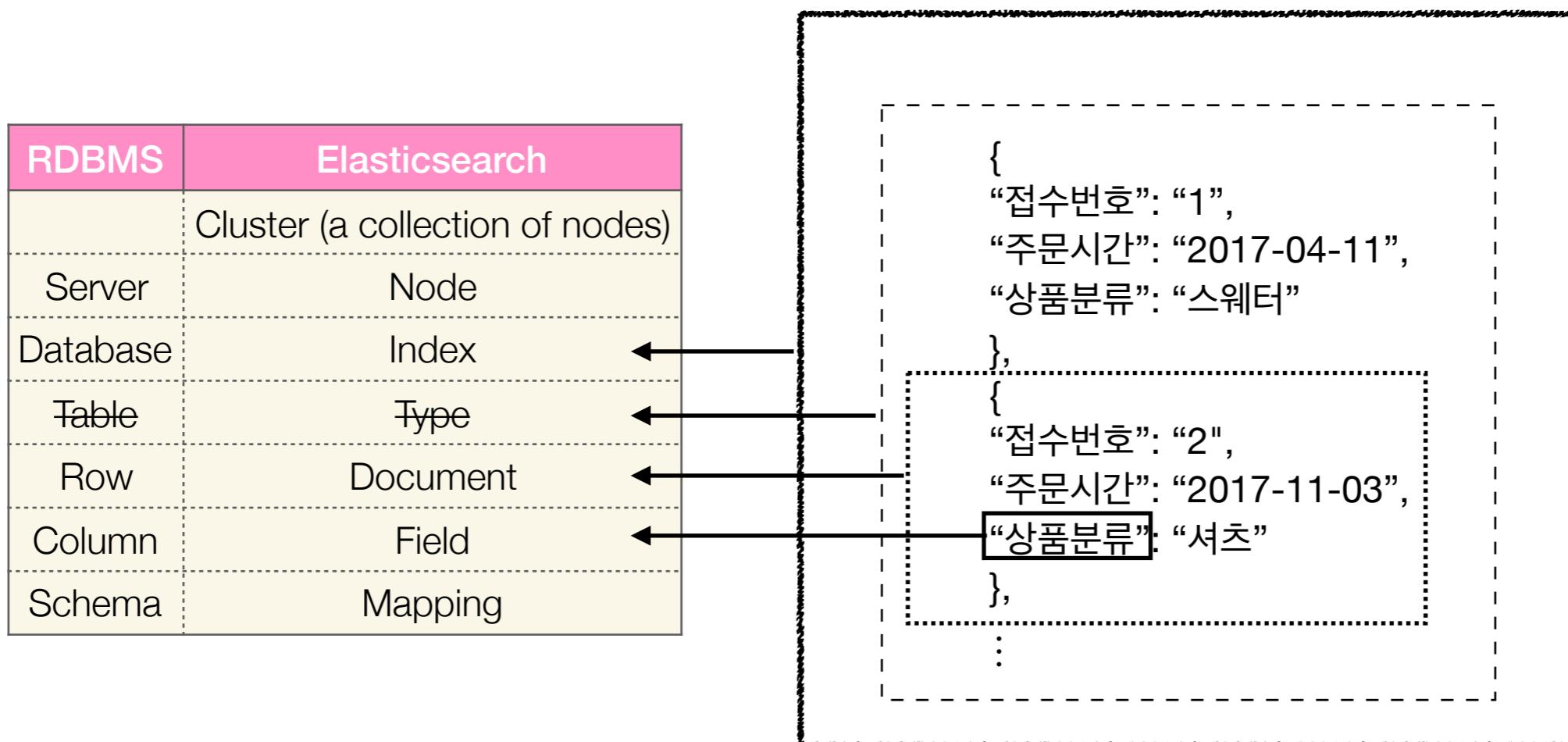


## 오늘의 목표

번호	항목	구체적 내용	(제멋대로) 난이도
1	Elastic Stack Workflow	하나하나가 어떤 역할을 하고 어떤 흐름으로 이어지는지 이해하기	★★★★★☆
1.1	Kibana Workflow	Elastic Stack 중 Kibana에서의 작업 흐름 이해하기	★★★★★☆
1.1.1	Index Pattern	시각화하고자 하는 Index(또는 Indices) 등록하기	★★★★★☆
1.1.2	Discover	Discover를 이용해서 데이터 개략적으로 이해하기	★★☆☆☆☆
1.1.3	Visualize Workflow	원하는 형태로 데이터 시각화하기	★★★★★☆
1.1.4	Dashboard	Visualization을 적절히 배치해서 dashboard 만들기	★★☆☆☆☆
2	Aggregation	원하는 결과를 얻기 위해 어떤 Aggregation 사용해야 될 지 이해하기	★★★★★

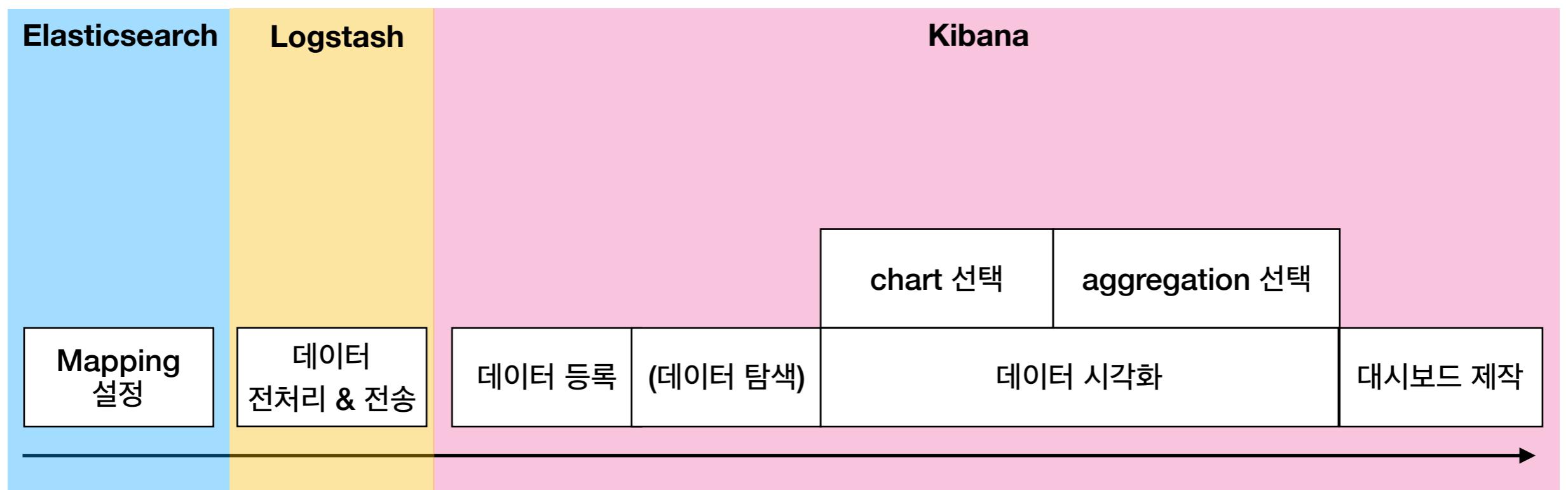
Elastic Stack

## 용어정리

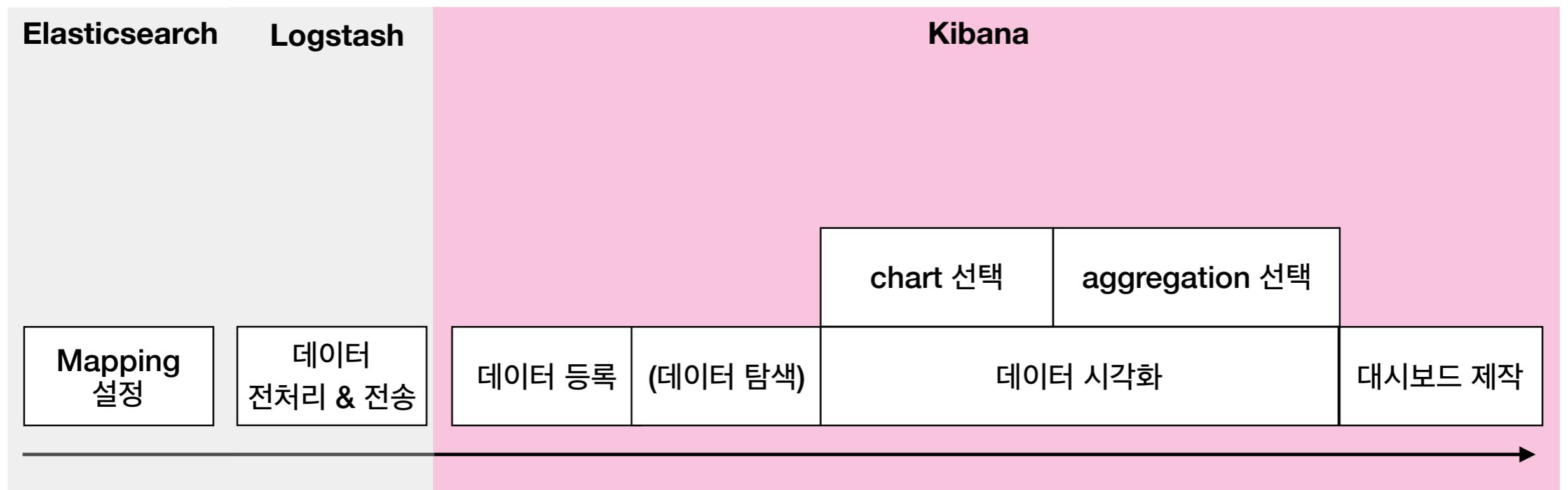


- 6.0.0 이후에는 Type이 사라지고 Index 1개에 Type 1개가 되어 사실상 폐지 ([참고](#))
- 최소한 Index, Document, Field, Mapping 은 무얼 의미하는지 알고 넘어가자!

# Elastic Stack 소개



# Elastic Stack 소개



## Kibana - Workflow

작업	어디에서?	상세
데이터 등록	Kibana Management	Elasticsearch에 저장된 데이터를 Kibana에서 사용할 수 있도록 등록
데이터 탐색	Kibana Discover	시각화하기 전에 데이터에 대한 개괄적 이해
데이터 시각화	Kibana Visualize	dashboard에 배치할 visualization 생성
대시보드 제작	Kibana Dashboard	생성한 visualization을 적절히 배치하여 dashboard 생성

Kibana - Management

## Kibana - Index 등록

The screenshot shows the Kibana Management interface. On the left is a sidebar with icons for Discover, Visualize, Dashboard, Timeline, Dev Tools, Management (which is selected), and Collapse. The main area has a header "Management" and "Version: 5.5.3". Below the header is a navigation bar with tabs: "Index Patterns" (selected), "Saved Objects", and "Advanced Settings".

## Kibana - Index 등록

---

### Elasticsearch에 저장된 데이터를 Kibana로 visualize 하기 위한 (Timelion 예외) 등록 절차

- 보통의 경우, index 이름이 *hello\_world*라면 index 등록할 때도 *hello\_world*라고 한다
- 그런데 만약 비슷한 이름을 가진 index로 데이터가 계속 들어온다면 어떨까?
  - 그럴 때 유용하게 사용할 수 있는게 wildcard(\*)다.
  - 예를 들어, index pattern을 *log-17.10.\** 했다고 하자.
    - 이 뜻은 index 이름이 *log-17.10.* (*log-17.10.01*, *log-17.10.02*, ...) 로 시작하는 index를 묶겠다는 것이다.
    - 이를 통해 log data는 일별로 다른 index에 하되 시각화는 통합해서 볼 수 있다는 장점이 있다.

자세한 사용법은 [클릭](#)

**Kibana - Discover**

## Kibana - 데이터 탐색

The screenshot shows the Kibana Discover interface. On the left sidebar, there are links for Discover, Visualize, Dashboard, Timeline, Dev Tools, Management, and a Collapse button. The main search bar at the top has the placeholder "Search... [e.g. status:200 AND extension:PHP]" and a note "Uses lucene query syntax". The search term "seoul\_wifi" is entered in the search bar. Below the search bar, it says "0 hits". To the right of the search bar are buttons for New, Save, Open, Share, and a time range selector "Last 15 minutes". A "Selected Fields" dropdown menu is open, showing the field "\_source". Below this are sections for "Available Fields" and "Selected Fields". The main content area displays the message "No results found 😞". It includes a note about the search results and suggestions to "Expand your time range" or "Refine your query". Examples of queries are provided in code blocks.

## No results found 😞

Unfortunately I could not find any results matching your search. I tried really hard. I looked all over the place and frankly, I just couldn't find anything good. Help me, help you. Here are some ideas:

### Expand your time range

I see you are looking at an index with a date field. It is possible your query does not match anything in the current time range, or that there is no data at all in the currently selected time range. Click the button below to open the time picker. For future reference you can open the time picker by clicking on the `@time picker` button in the top right corner of your screen.

### Refine your query

The search bar at the top uses Elasticsearch's support for Lucene [Query String syntax](#). Let's say we're searching web server logs that have been parsed into a few fields.

**Examples:**

Find requests that contain the number 200, in any field:  
200

Or we can search in a specific field. Find 200 in the status field:  
status:200

Find all status codes between 400-499:  
status: [400 TO 499]

### 본격적으로 데이터를 시각화 하기 전에 간단히 데이터를 탐색해보는 과정

- 데이터가 주어지면 바로 그래프를 작성하고 dashboard를 구축하는 사람은 드물 것이다.
- 데이터는 몇 개고, 어떤 데이터가 들어 왔고, 데이터 값은 적절히 균형을 이루고 있는지 등에 대해 살펴보지 않는가?
- 그럴 때 유용하게 쓸 수 있는 게 Discover이다.
- 기본적으로 많은 기능이 있으며 다음과 같은 질문에 대답을 해준다.
  - 전체 Documents(=데이터 개수)는 몇 개인가?
  - 시간대별 Documents의 분포는 어떤가?
  - 특정 조건을 만족하는 Documents는 몇 개나 될까?
  - 특정 Field의 값들의 (예를 들어 10대, 20대, 30대 이상) 비율이 어떻게 되지?
  - 특정 Field가 있는/없는 Documents는 얼마나 되지?

자세한 사용법은 [클릭](#)

## Visualize

---

Kibana - Visualize

## Kibana - Visualize Workflow

작업	어디에서?	상세
Visualize 이동	Kibana Management	Elasticsearch에 저장된 데이터를 Kibana에서 사용할 수 있도록 등록
데이터 탐색	Kibana Discover	시각화하기 전에 데이터에 대한 개괄적 이해
데이터 시각화	Kibana Visualize	dashboard에 배치할 visualization 생성
대시보드 제작	Kibana Dashboard	생성한 visualization을 적절히 배치하여 dashboard 생성



작업	어디에서?	상세
Create new visualization	Kibana Visualize	새로운 Visualization 생성하기 위해 Create New Visualization 선택
Select visualization type	Kibana Visualize	목적에 맞는 차트 종류를 선택
Select Search (Index)	Kibana Visualize	Discover에서 Search를 저장하지 않았으면 Index(=New Search) 선택
Select aggregation	Kibana Visualize	목적에 맞는 Aggregation을 선택하여 그래프 작성

자세한 사용법은 [클릭](#)

## Elasticsearch - Aggregation (Metric)

## Aggregation - Metric

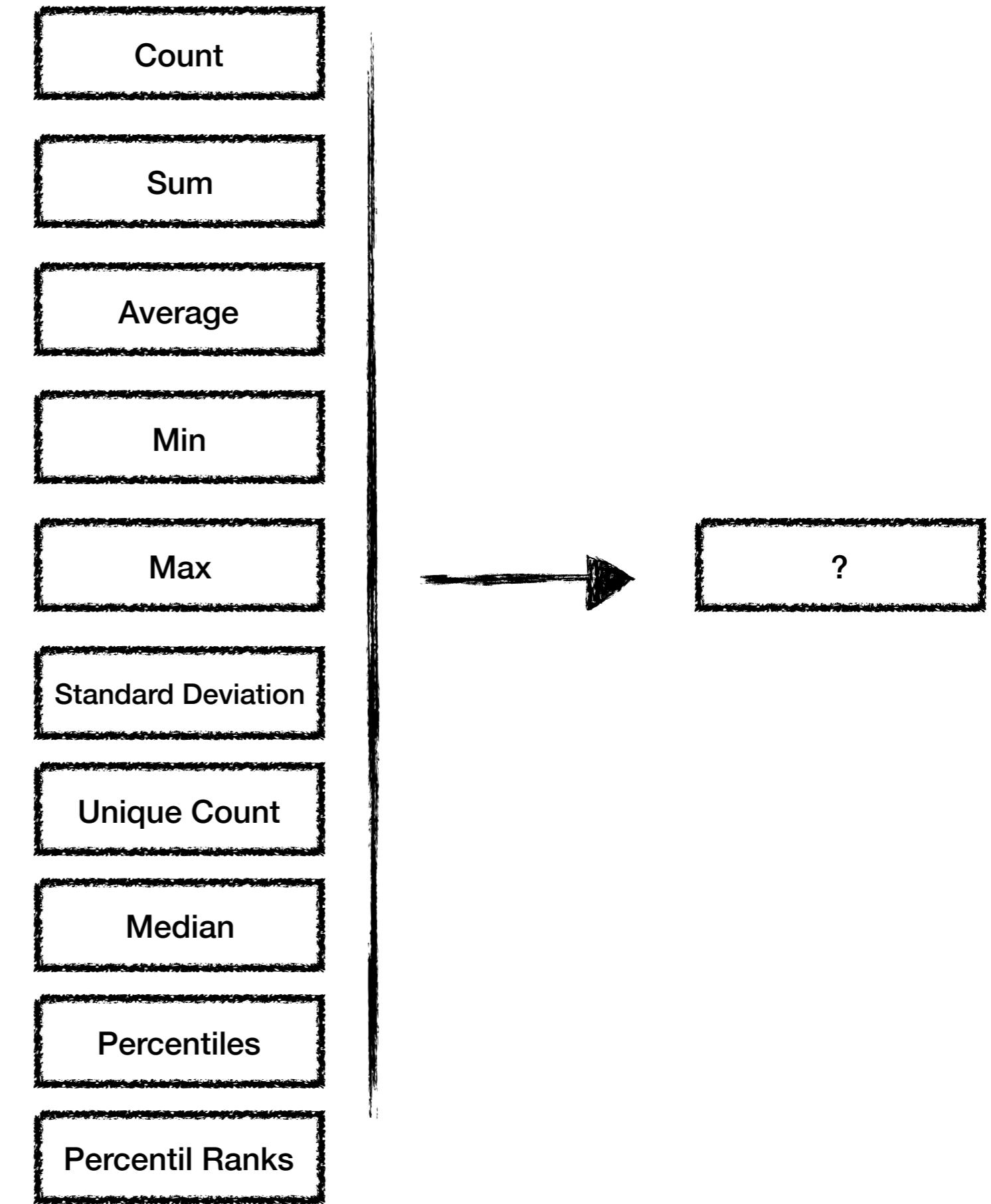
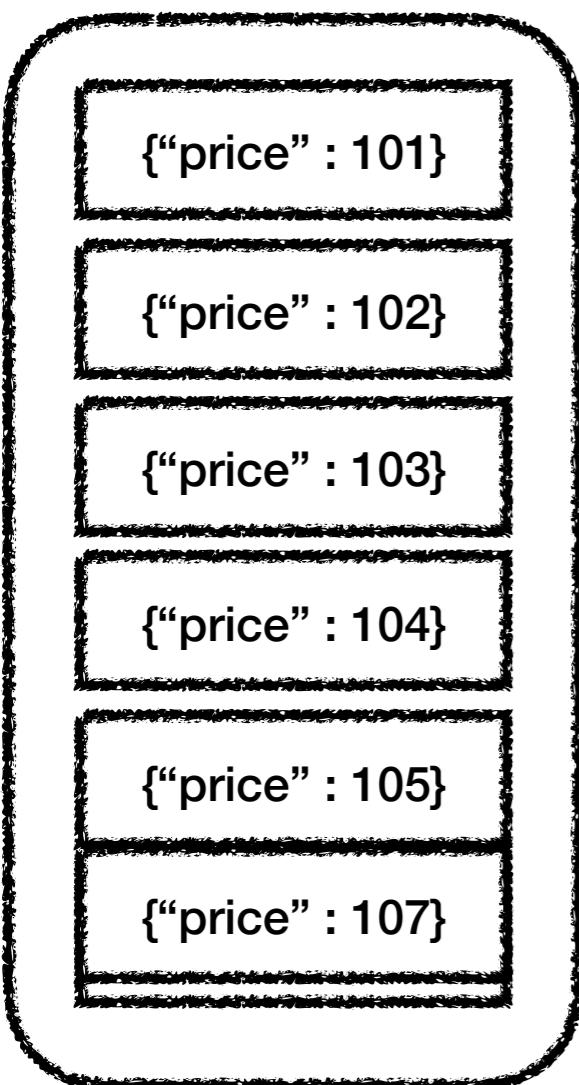
---

Metric Aggregation은 (Bucket 내) 특정 Field 값에 연산을 수행하는 Aggregation이다

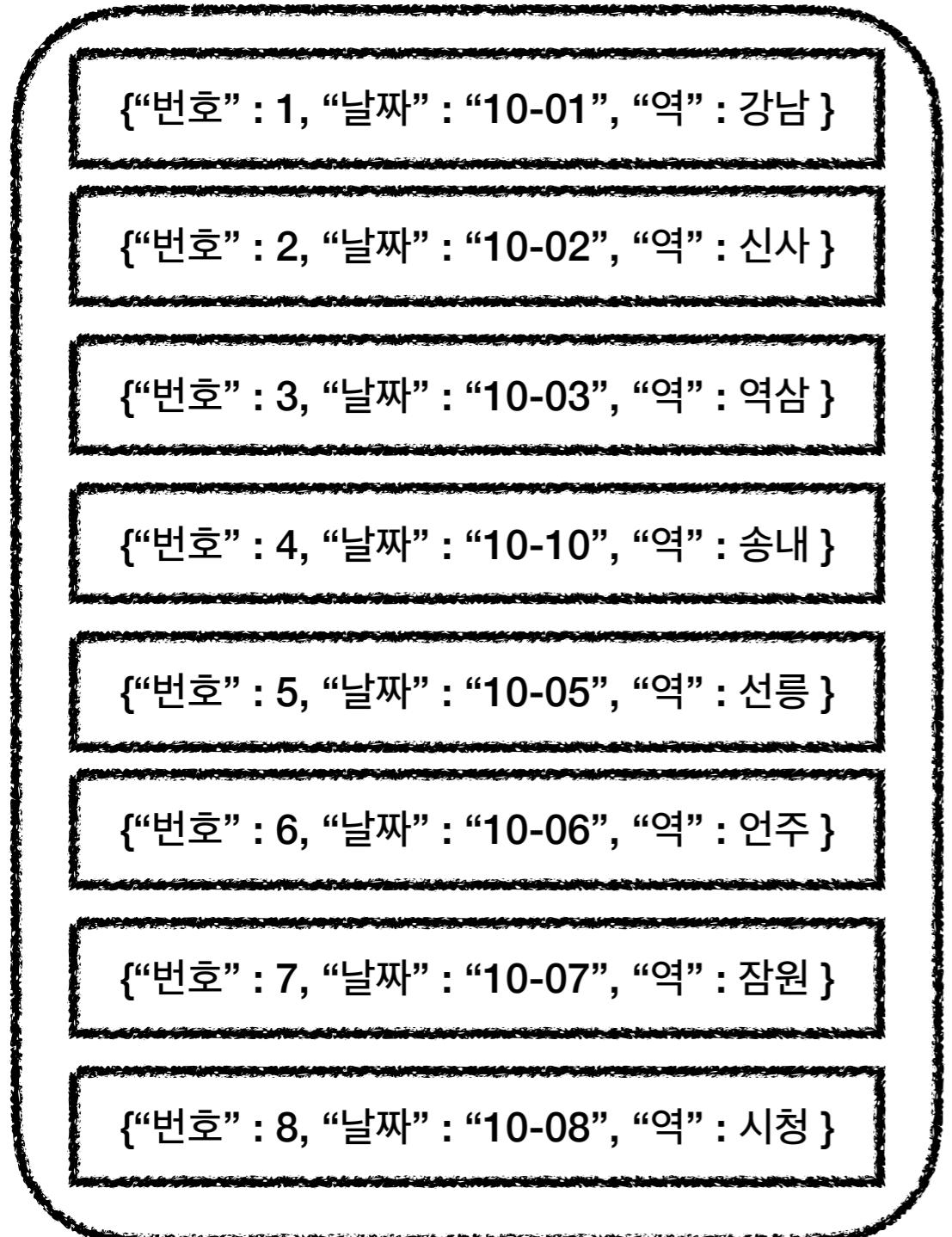
- 기본 연산을 수행하는 Aggregation이다.
- 주의할 점은 Metric Aggregation은 해당 Bucket 내의 Documents에 대해 수행한다는 점이다.
- Kibana에서 제공하는 Metric Aggregation은 다음과 같은 것들이 있다.
  - Value Count Aggregation
  - Avg Aggregation
  - Sum Aggregation
  - Min Aggregation
  - Max Aggregation
  - Extended Stats Aggregation
  - Cardinality Aggregation
  - Percentiles Aggregation
  - Percentile Ranks Aggregation
  - Top Hits Aggregation

## Aggregation - Metric

Single Bucket



## Aggregation - Metric



### Top Hits Aggregation

날짜가	빠른	데이터 3개의	번호	합을 구하세요
번호가	작은	데이터 2개의	역명	모두 나열하세요
역명이	빠른	데이터 2개의	번호	평균을 구하세요

## Aggregation - Metric

종류	적용 가능 Type	(제 멋대로) 한 줄 설명	링크
Avg	Number	(Bucket 내) Document의 특정 Field의 평균 계산	<a href="#">클릭</a>
Sum	Number	(Bucket 내) Document의 특정 Field의 합 계산	<a href="#">클릭</a>
Min/Max	Number	(Bucket 내) Document의 특정 Field의 최소/최대 계산	<a href="#">클릭</a>
Extended Stats	Number	(Bucket 내) Document의 특정 Field의 기초 통계값 계산	<a href="#">클릭</a>
Cardinality	Number	(Bucket 내) Document의 특정 Field의 고유한 개수 계산	<a href="#">클릭</a>
Percentiles	Number	(Bucket 내) Document의 특정 Field의 백분위수 계산	<a href="#">클릭</a>
Percentiles Ranks	Number	(Bucket 내) Document의 특정 Field의 백분위 계산	<a href="#">클릭</a>
Top Hits	All	(Bucket 내) 특정 조건을 만족하는 Documents의 특정 Field의 Agg 반환	<a href="#">클릭</a>
Value Count	All	(Bucket 내) Document의 개수 계산	<a href="#">클릭</a>

실제 데이터로 어떻게 계산되는지 궁금하다면 [클릭](#)

- 
- Number Field : Concat, Sum, Min, Max, Count
  - 기타 Field : Concat

Elasticsearch - Aggregation  
(Bucket)

## Aggregation - Bucket

---

Bucket Aggregation은 데이터를 일정한 기준으로 묶는 Aggregation이다.

- Bucket은 {비슷한 특징을 지닌 Group} 정도로 볼 수 있다.
- Bucket Aggregation 자체는 연산을 수행하지 않고 다른 Aggregation이 사용할 수 있도록 Bucket을 생성한다.
- 그렇다면 Bucket은 어떤 기준으로 생성할 수 있을까?
- Kibana에서 제공하는 Bucket Aggregation(=Bucket 생성 기준)은 다음과 같은 것들이 있다.
  - Date Histogram Aggregation
  - Date Range Aggregation
  - Histogram Aggregation
  - Range Aggregation
  - IPv4 Aggregation
  - Terms Aggregation
  - Significant Terms Aggregation
  - Filters Aggregation
  - Geo Hash Aggregation

## Aggregation - Bucket

---

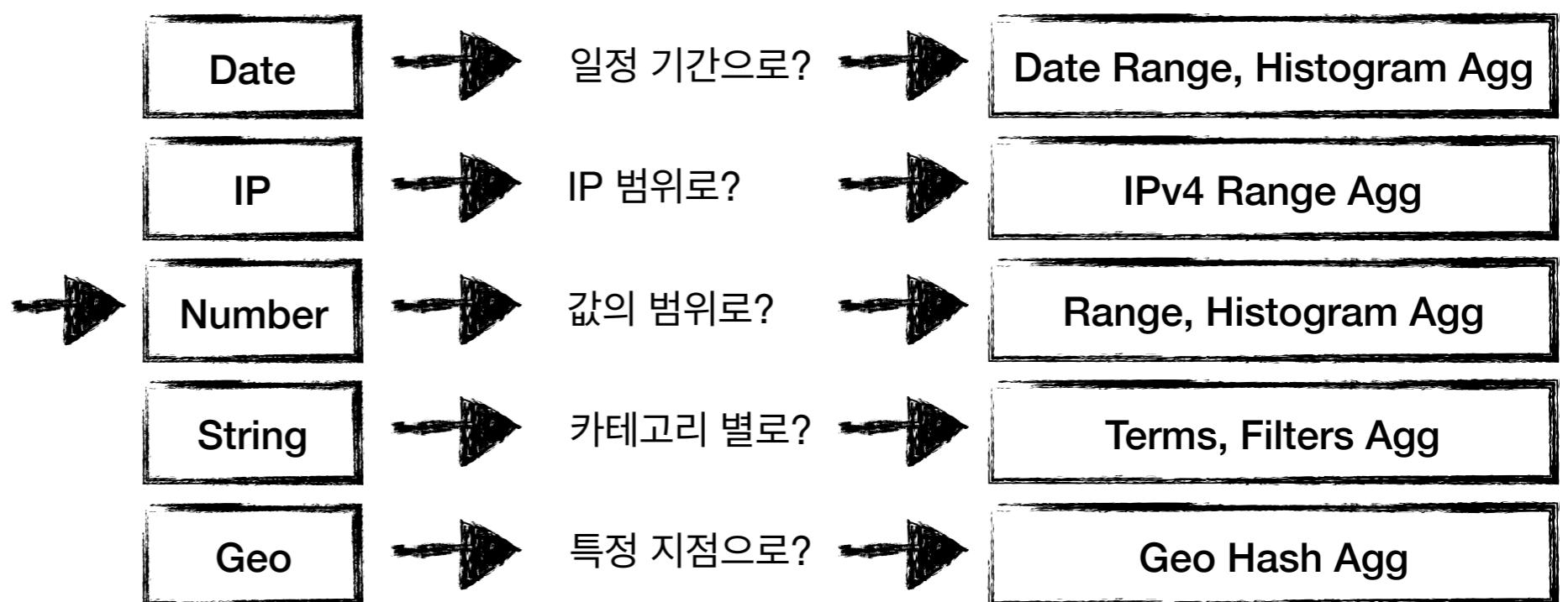
- 아래와 같은 데이터가 있다고 하자
- 어떤 기준으로 Bucket을 생성할 수 있을까?

```
{  
    "시간": "2017-11-06T22:51:39",  
    "ip": "27.119.249.209",  
    "좌표": "37.23486, 126.60655",  
    "가격": 26000,  
    "분류": "청바지",  
    "성별": "여성"  
},  
{  
    "시간": "2017-11-06T22:51:39",  
    "ip": "27.119.249.209",  
    "좌표": "37.23488, 126.60603",  

```

## Aggregation - Bucket

```
{  
  "시간": "2017-11-06T22:51:39",  
  "ip": "27.119.249.209",  
  "좌표": "37.23486, 126.60655",  
  "가격": 26000,  
  "분류": "청바지",  
  "성별": "여성"  
}
```



## Aggregation - Bucket

- pie
- line, area, bar, horizontal
- heat map, coordinate
- tag

종류	적용 가능 Type	(제 멋대로) 한 줄 설명	링크
Date Histogram	Date	날짜/시간을 일정하게 지정하여 구간 내의 Documents로 Bucket 형성	<a href="#">클릭</a>
Date Range	Date	날짜/시간을 임의로 지정하여 구간 내의 Documents로 Bucket 형성	<a href="#">클릭</a>
Histogram	Number	범위를 일정하게 지정하여 구간 내의 Documents로 Bucket 형성	<a href="#">클릭</a>
Range	Number	범위를 임의로 지정하여 구간 내의 Documents로 Bucket 형성	<a href="#">클릭</a>
Terms	Date, IP, Number, String	특정 Field 값을 기준으로 Bucket 생성 (카테고리 데이터에 유용)	<a href="#">클릭</a>
Significant Terms	String	Background 대비 Foreground에서 특별한 값으로 Bucket 생성	<a href="#">클릭</a>
Filters	Date, IP, Number, String	직접 조건을 입력하여 Bucket 생성 (조건 개수만큼 Bucket 생성)	<a href="#">클릭</a>
Geo Hash	Geo Point	특정 지점 (Centroid) 근처에 있는 값들을 모아서 Bucket 생성	<a href="#">클릭</a>
IPv4 Range	IP	IP 주소의 범위로 Bucket 생성	<a href="#">클릭</a>

실제 데이터로 어떻게 계산되는지 궁금하다면 [클릭](#)

## Elasticsearch - Aggregation (Parent Pipeline)

## Aggregation - Parent Pipeline

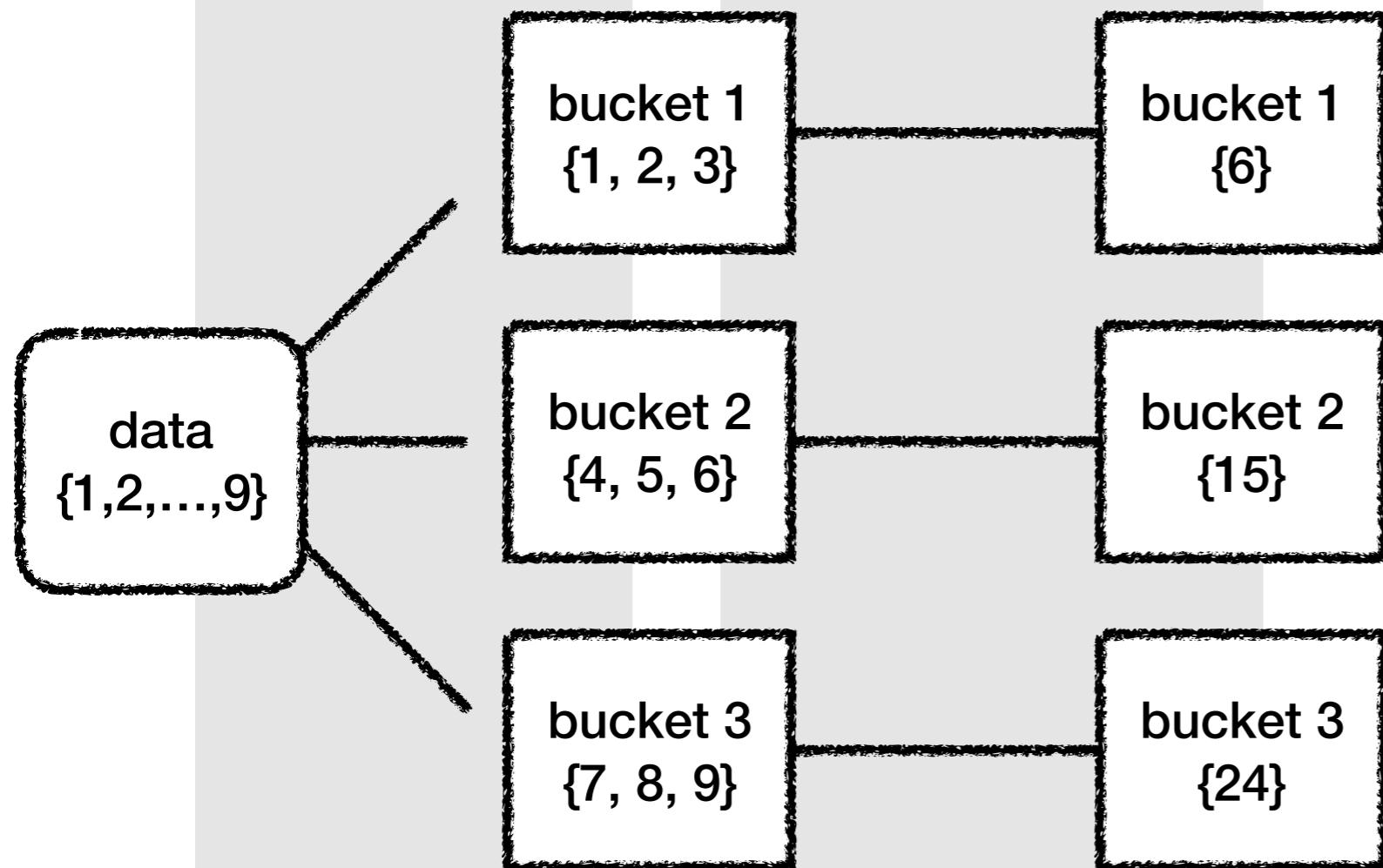
---

Parent Pipeline Aggregation은 Date Range/Histogram Bucket에 대해 Metric Aggregation된 결과에 수행하는 Aggregation이다

- Kibana에서 제공하는 Parent Pipeline Aggregation은 다음과 같은 것들이 있다.
  - Derivative
  - Moving Average
  - Cumulative Sum
  - Serial Diff

## Bucket Aggregation (Range)

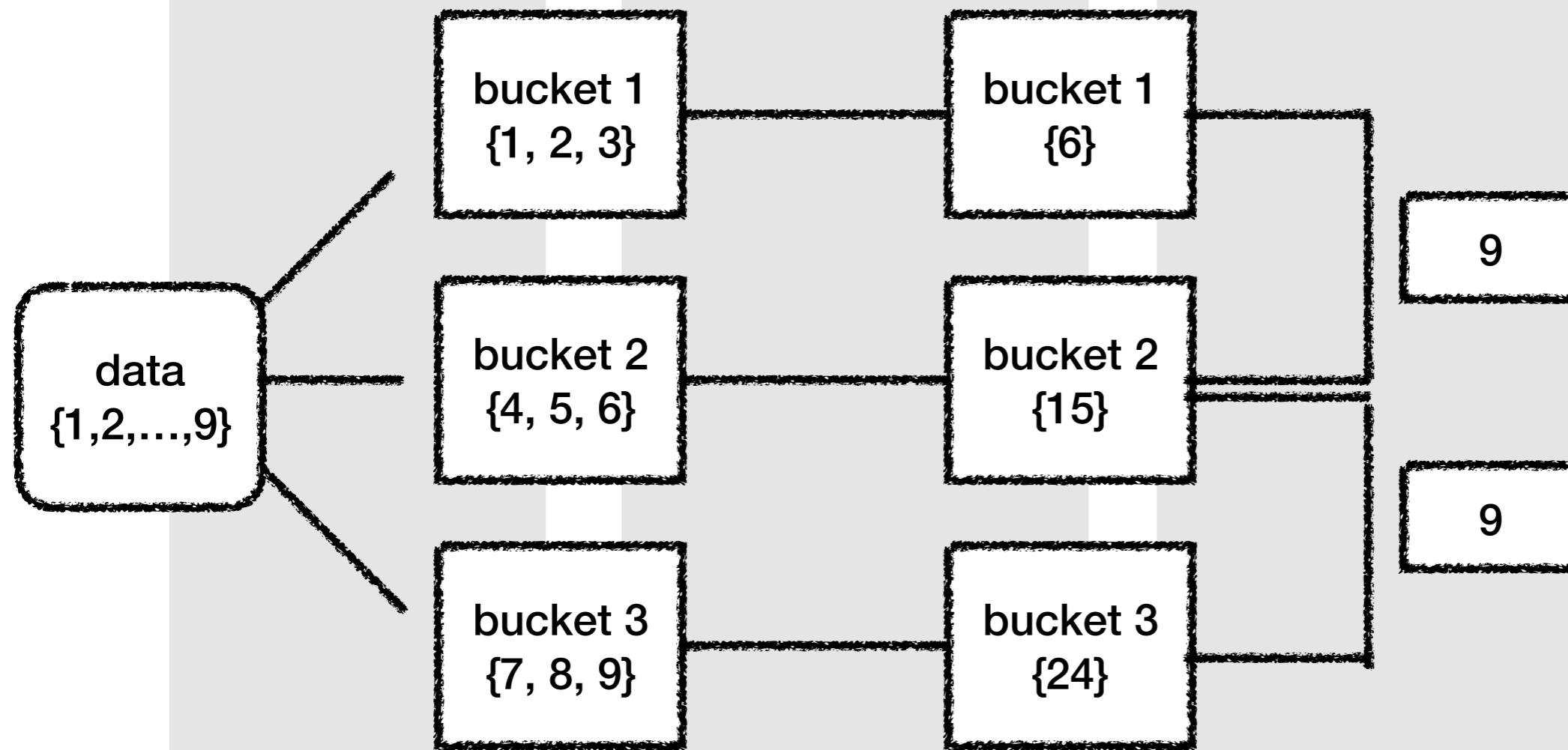
## Metric Aggregation (Sum)



### Bucket Aggregation (Range)

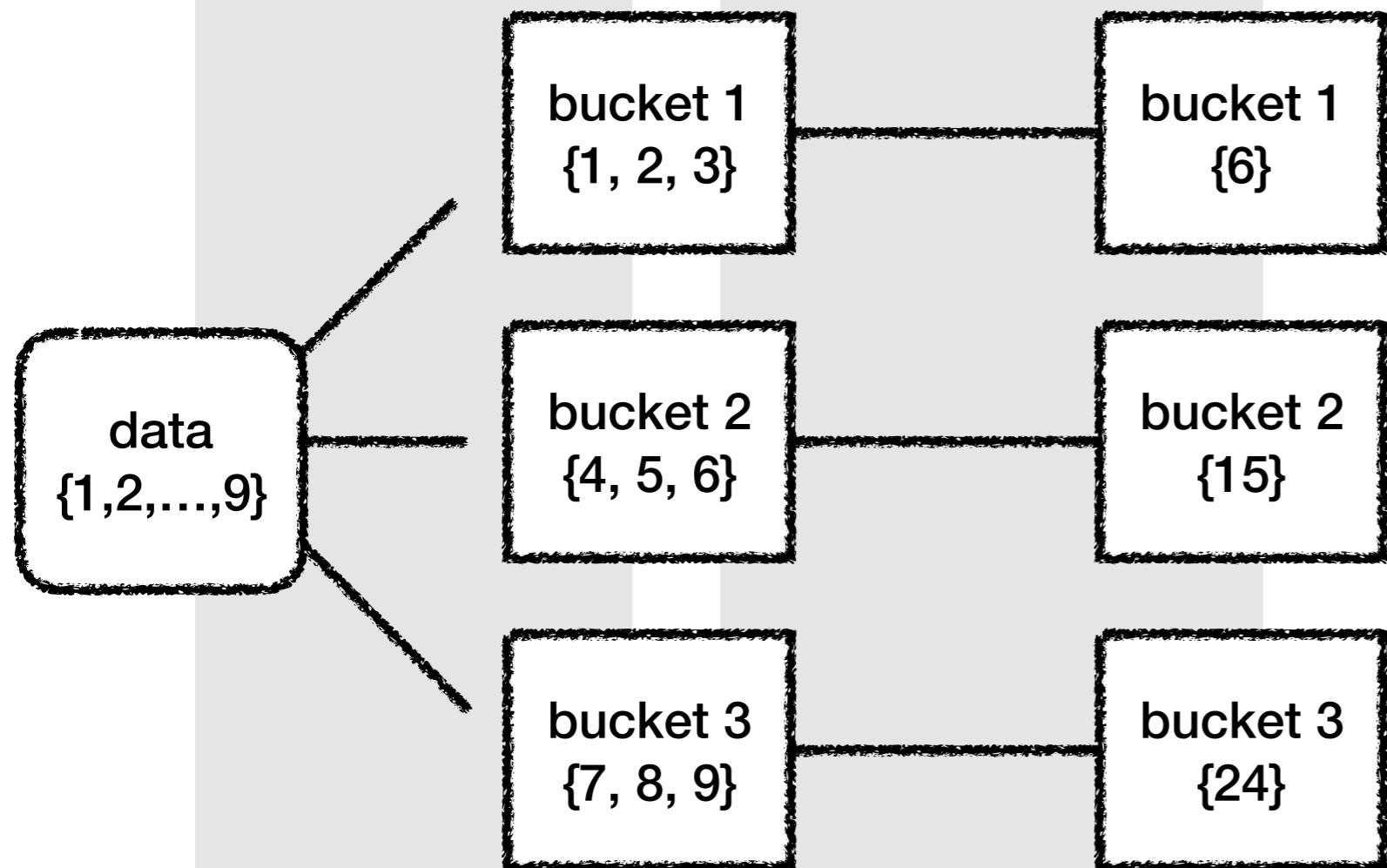
### Metric Aggregation (Sum)

### Parent Pipeline Aggregation (Derivative)



## Bucket Aggregation (Range)

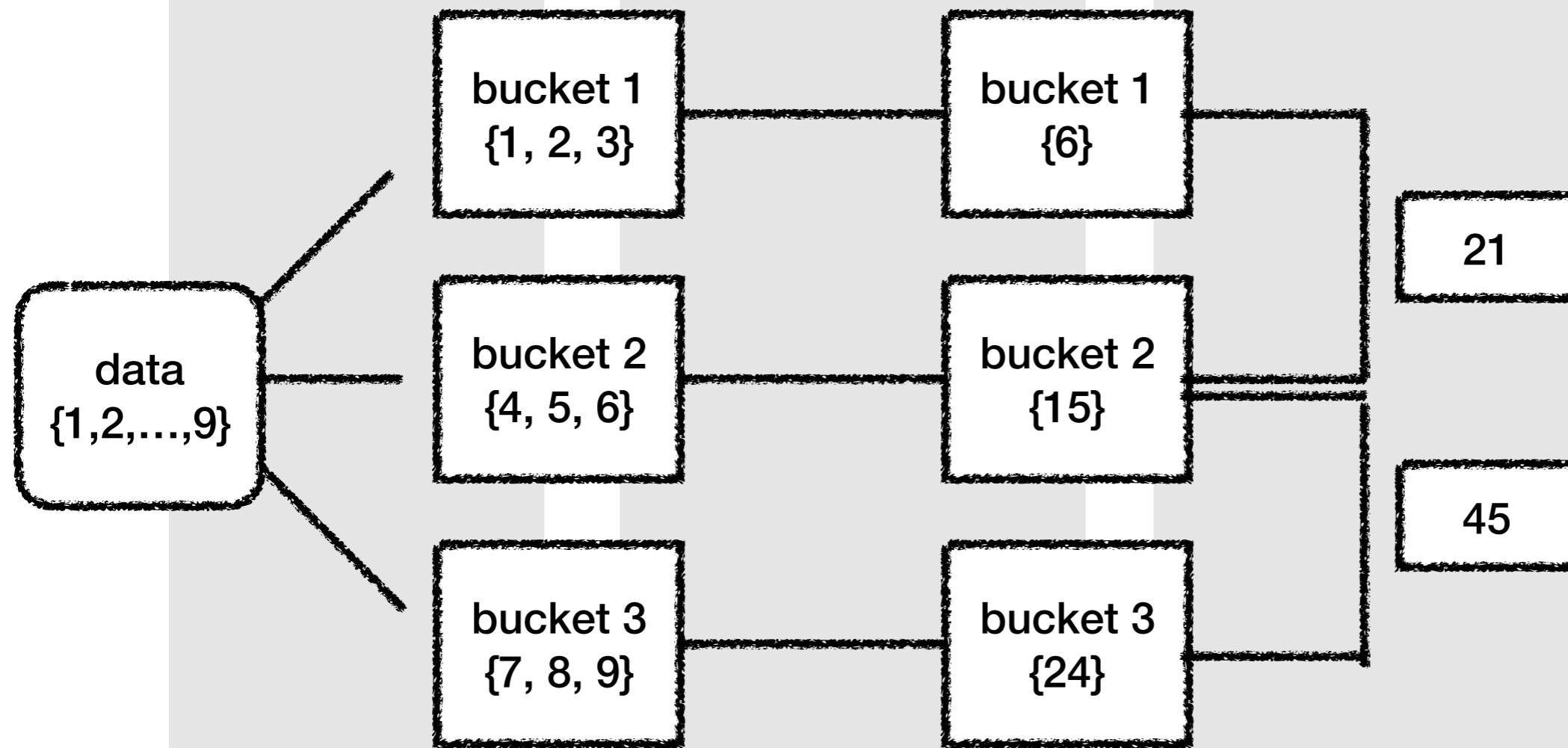
## Metric Aggregation (Sum)



### Bucket Aggregation (Range)

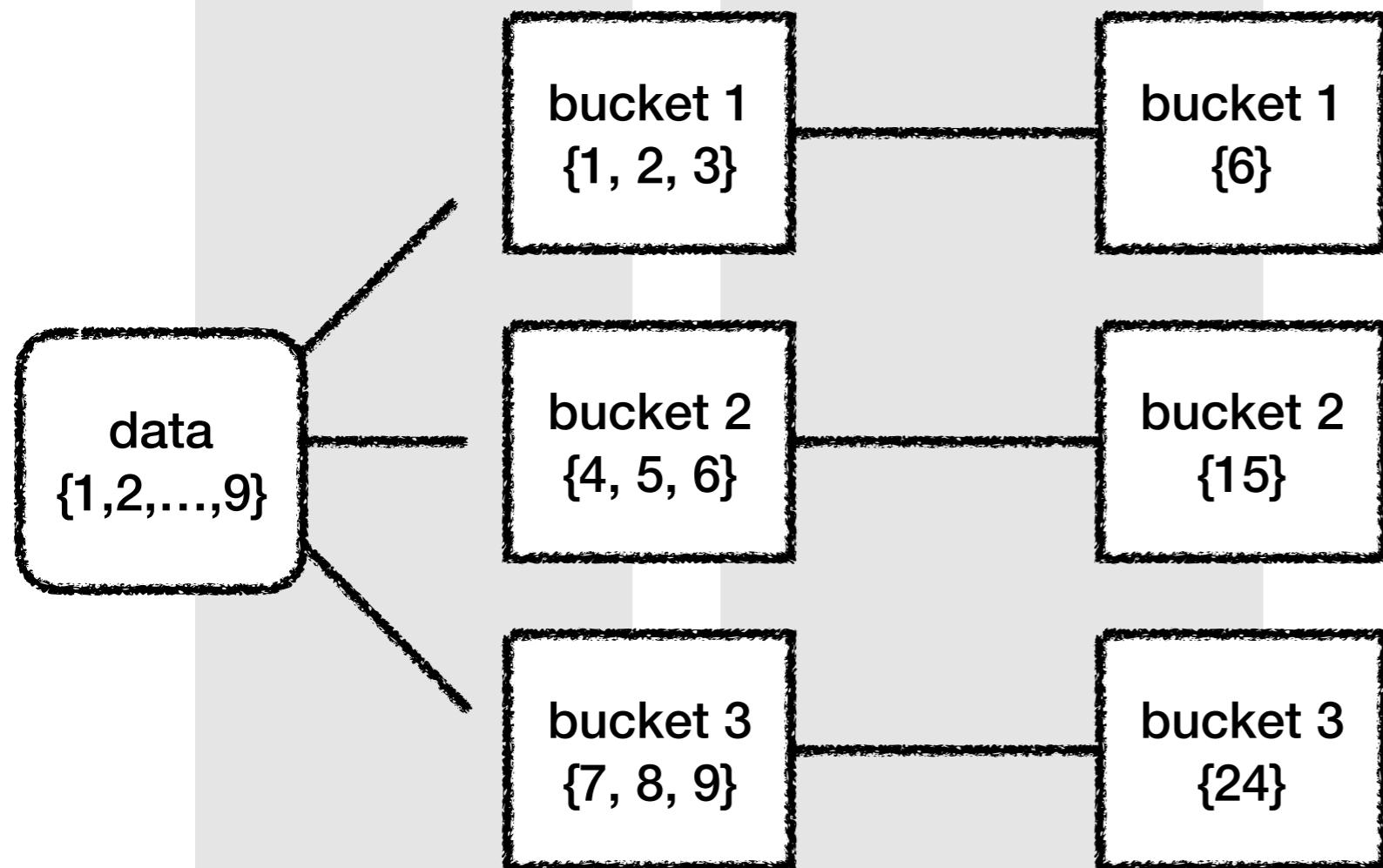
### Metric Aggregation (Sum)

### Parent Pipeline Aggregation (Cumulative Sum)



## Bucket Aggregation (Range)

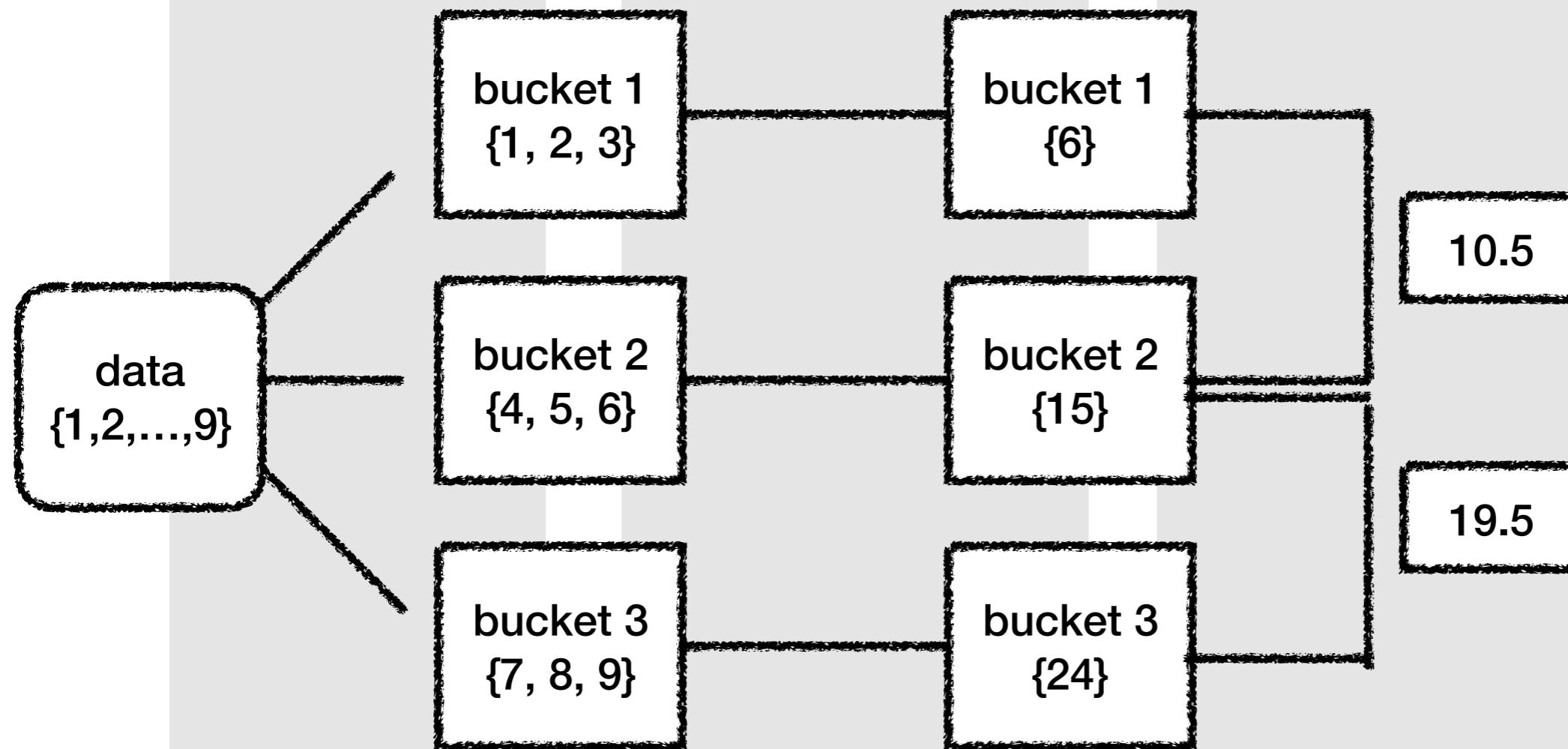
## Metric Aggregation (Sum)



**Bucket Aggregation  
(Range)**

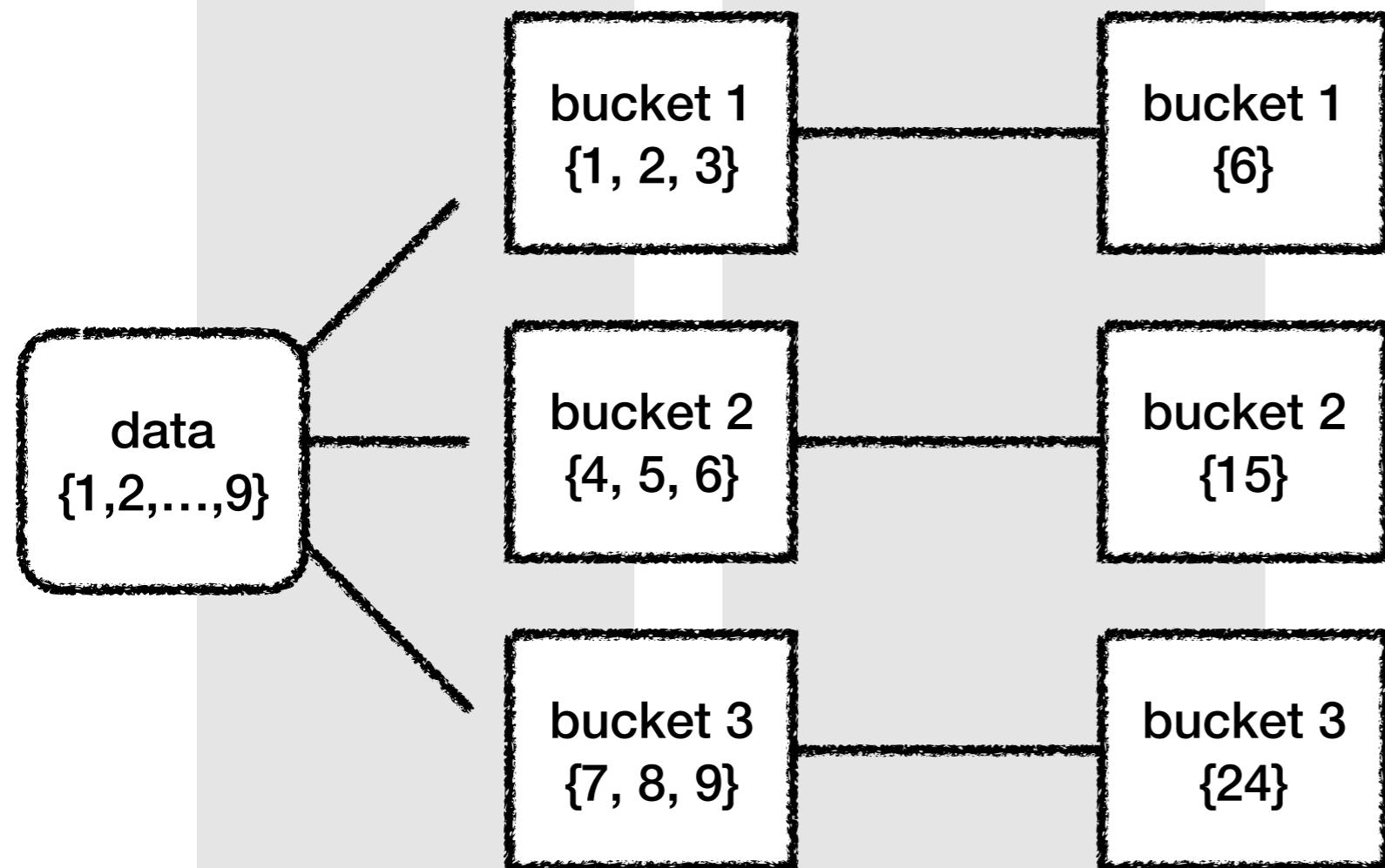
**Metric Aggregation  
(Sum)**

**Parent Pipeline Aggregation  
(Moving Average, window=2)**



## Bucket Aggregation (Range)

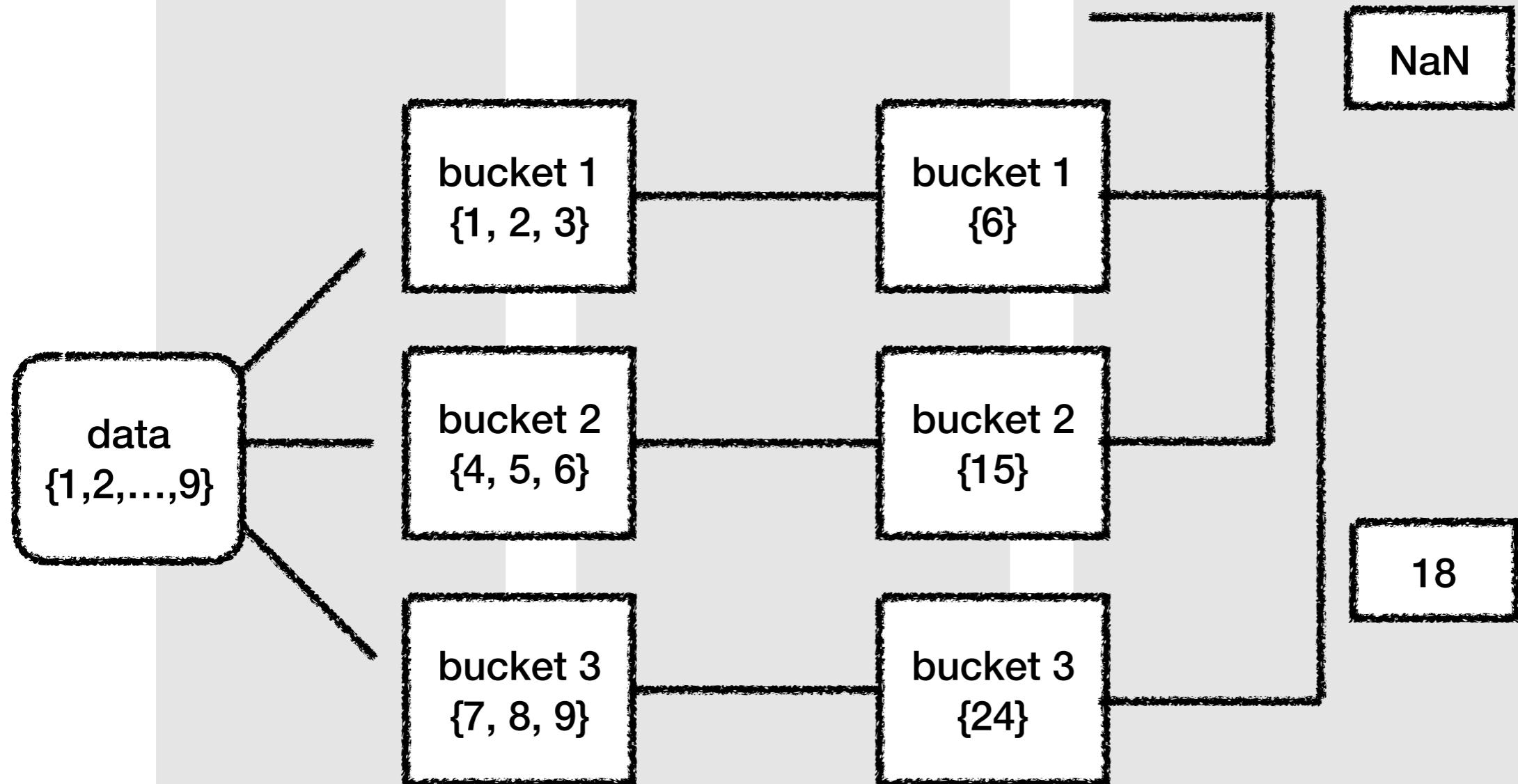
## Metric Aggregation (Sum)



### Bucket Aggregation (Range)

### Metric Aggregation (Sum)

### Parent Pipeline Aggregation (Serial Diff, lag=2)



## Aggregation - Parent Pipeline

종류	(제 멋대로) 한 줄 설명	링크
Derivative	Bucket Agg 후, Bucket 내 Metric Agg 하고 난 후, 연속한 Bucket 간의 차이를 구함	<a href="#">클릭</a>
Cumulative Sum	Bucket Agg 후, Bucket 내 Metric Agg 하고 난 후, Bucket 값들의 누적합을 구함	<a href="#">클릭</a>
Moving Average	Bucket Agg 후, Bucket 내 Metric Agg 하고 난 후, {n개} Bucket 간의 평균을 구함	<a href="#">클릭</a>
Serial Diff	Bucket Agg 후, Bucket 내 Metric Agg 하고 난 후, {n번째 이전} Bucket 과의 차이를 구함	<a href="#">클릭</a>

실제 데이터로 어떻게 계산되는지 궁금하다면 [클릭](#)

Elasticsearch - Aggregation  
(Sibling Pipeline)

## Aggregation - Sibling Pipeline

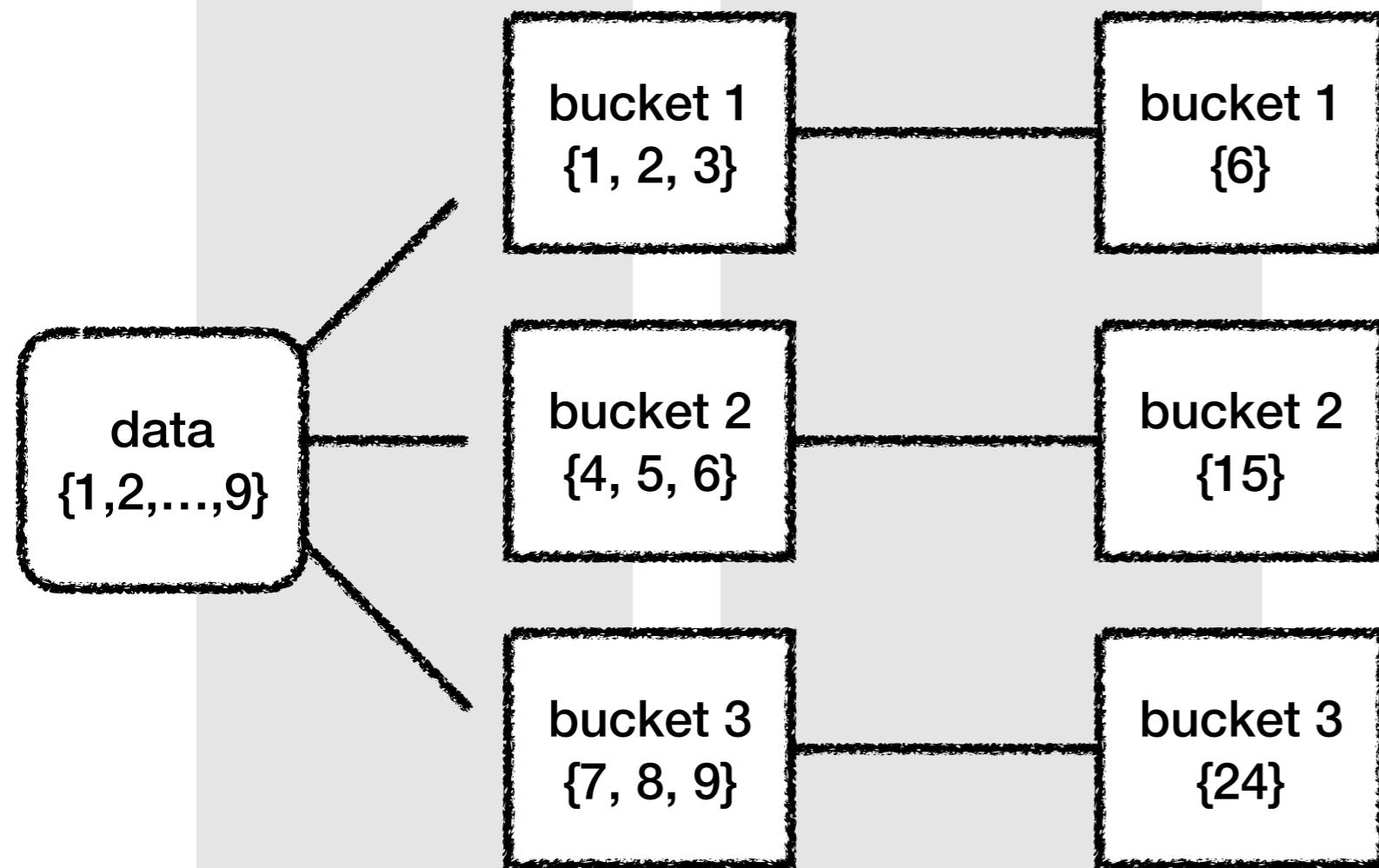
---

Sibling Pipeline Aggregation은 Bucket Aggregation에 Metric Aggregation을 적용하고, 전체 Bucket에 대해 수행하는 Aggregation이다.

- Metric Aggregation과 Sibling Aggregation을 같은 걸 사용하면 의미가 없어진다.
  - {1, 2}의 평균(=1.5)과 {3, 4}의 평균(=3.5)의 평균(2.5)을 구하는 것과 {1, 2, 3, 4}의 평균(=2.5)을 구하는 게 같기 때문이다.
- Sibling Aggregation은 Bucket 간에 최종 연산을 수행하여 하나의 값으로 반환한다.
- 만약 Bucket 간 Agg 값이 보고 싶다면 Terms Agg 같은 Bucket Agg과 Metric Agg를 함께 사용하며 된다
- Kibana에서 제공하는 Sibling Pipeline Aggregation은 다음과 같은 것들이 있다.
  - Average Bucket
  - Sum Bucket
  - Min Bucket
  - Max Bucket

## Bucket Aggregation (Range)

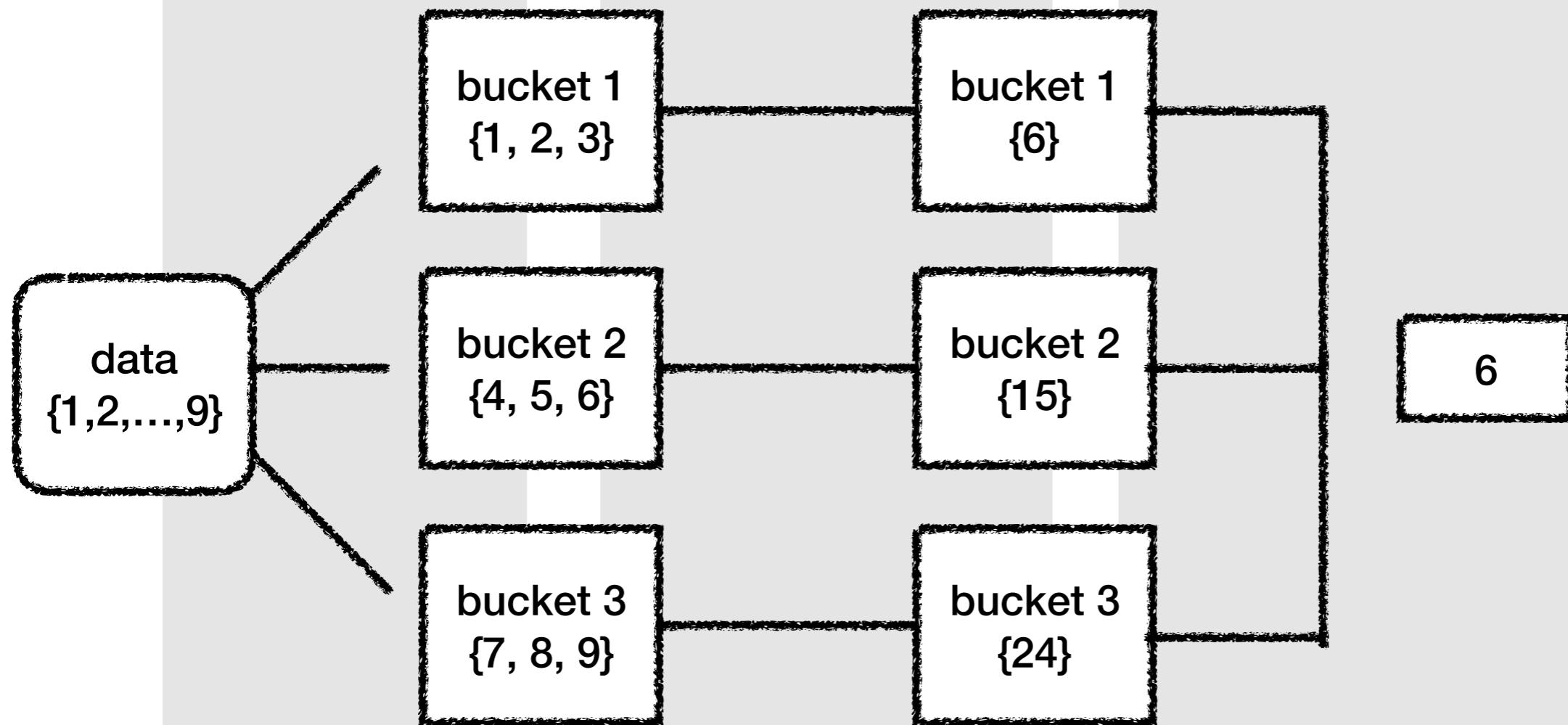
## Metric Aggregation (Sum)



### Bucket Aggregation (Range)

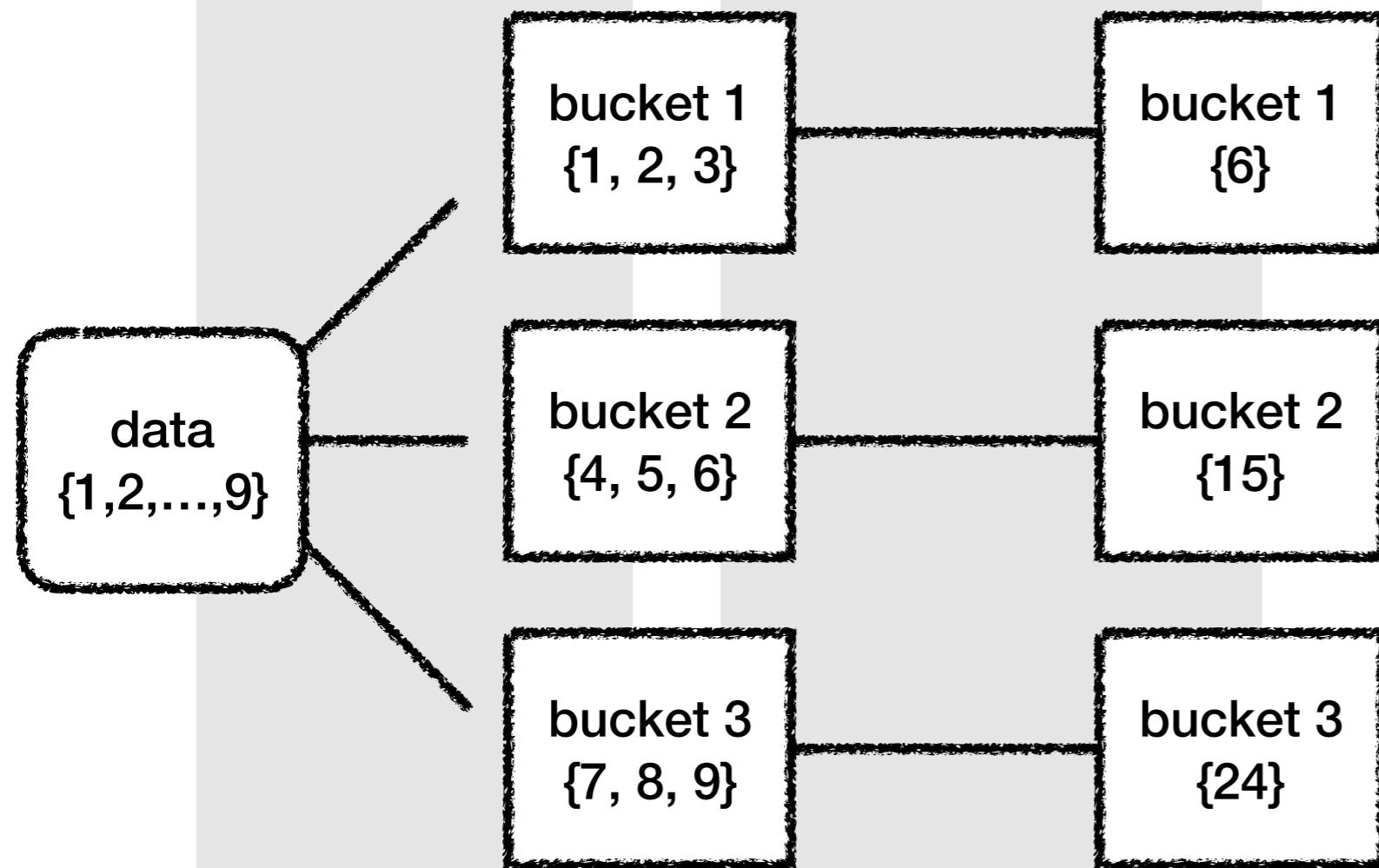
### Metric Aggregation (Sum)

### Sibling Pipeline Aggregation (Min)



## Bucket Aggregation (Range)

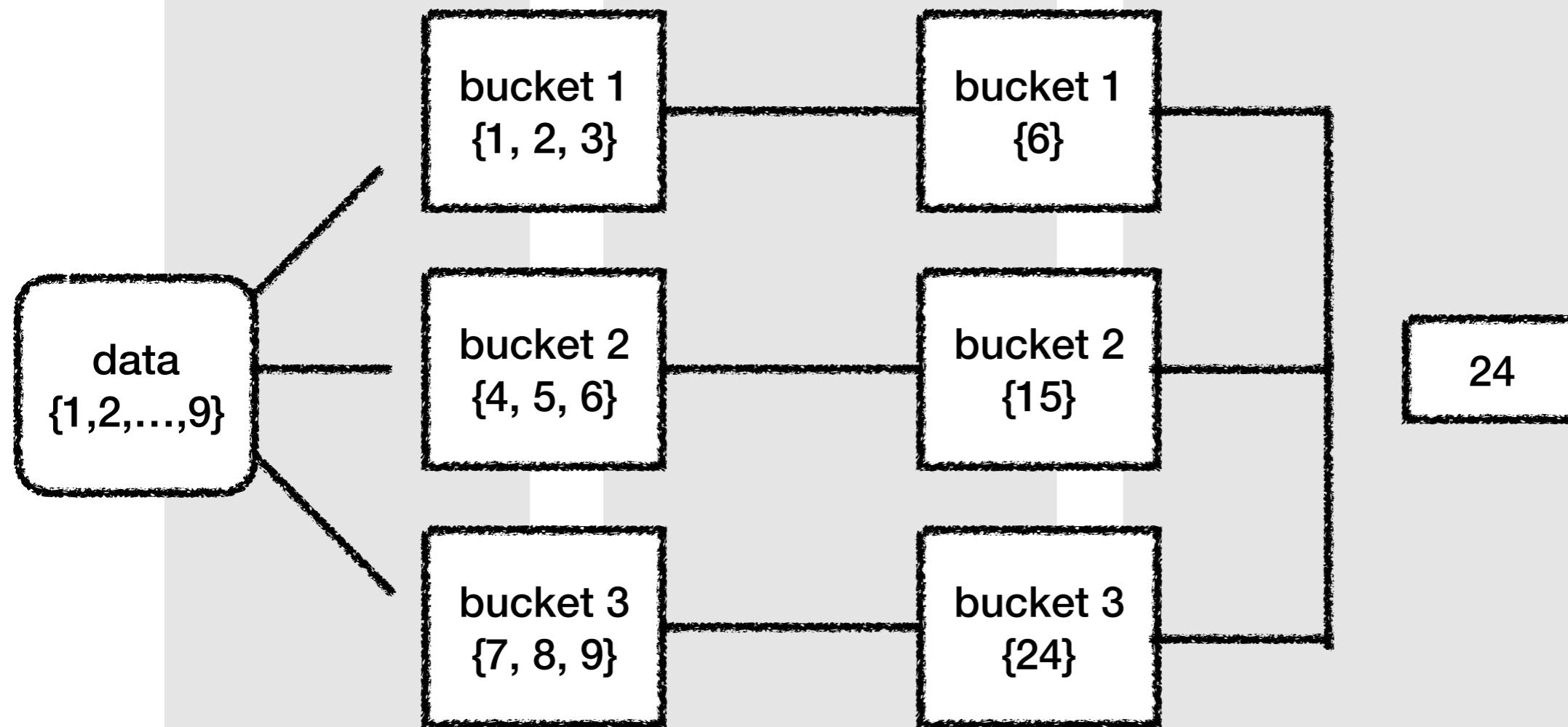
## Metric Aggregation (Sum)



### Bucket Aggregation (Range)

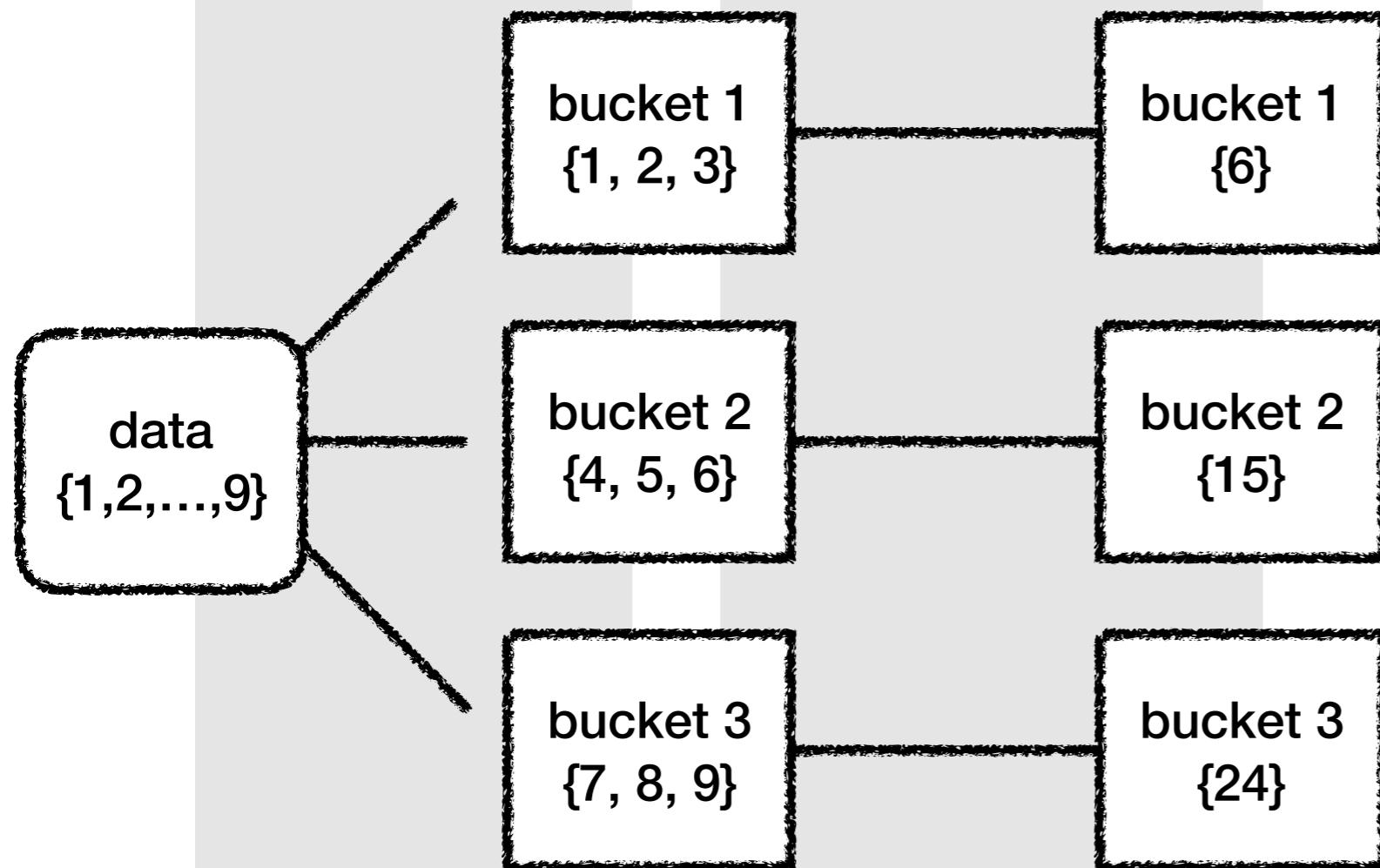
### Metric Aggregation (Sum)

### Sibling Pipeline Aggregation (Max)



## Bucket Aggregation (Range)

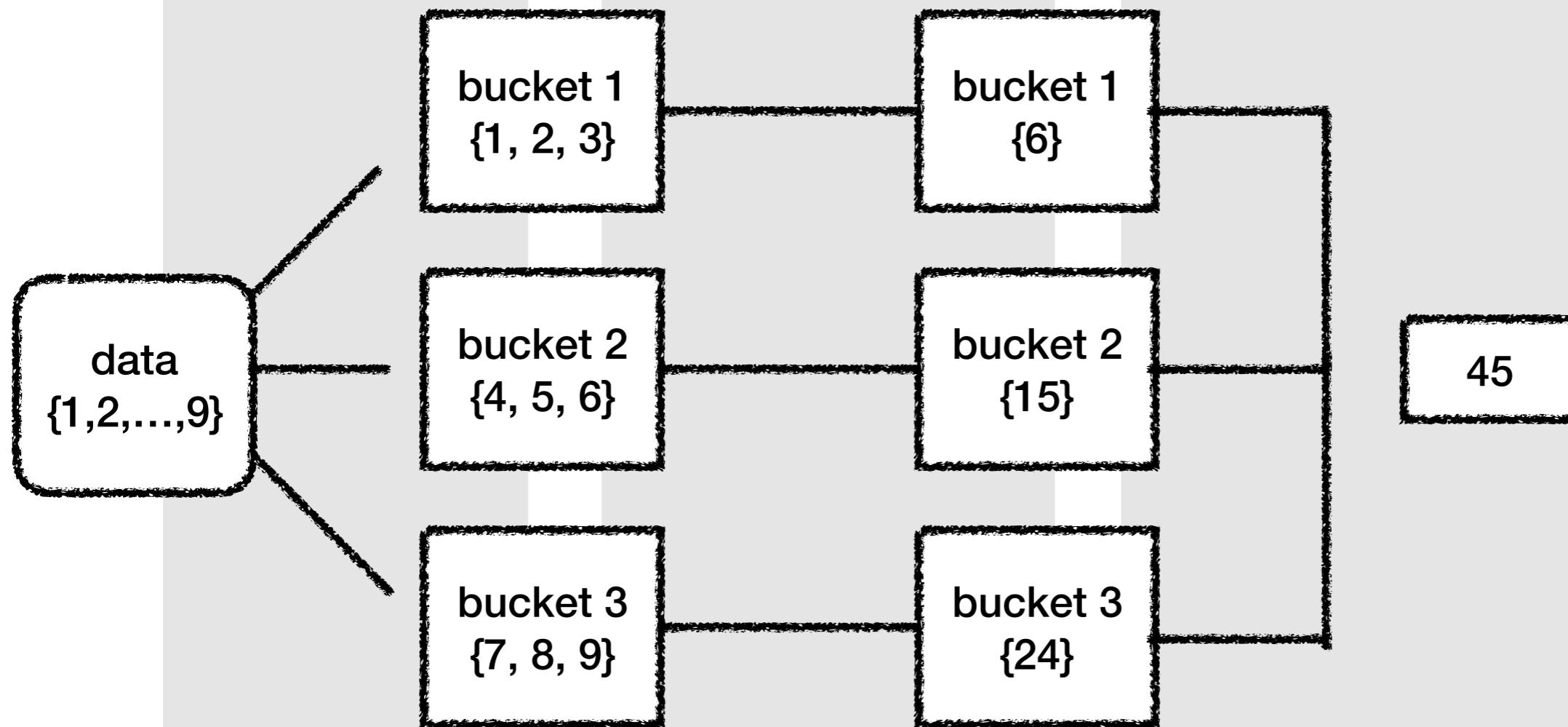
## Metric Aggregation (Sum)



### Bucket Aggregation (Range)

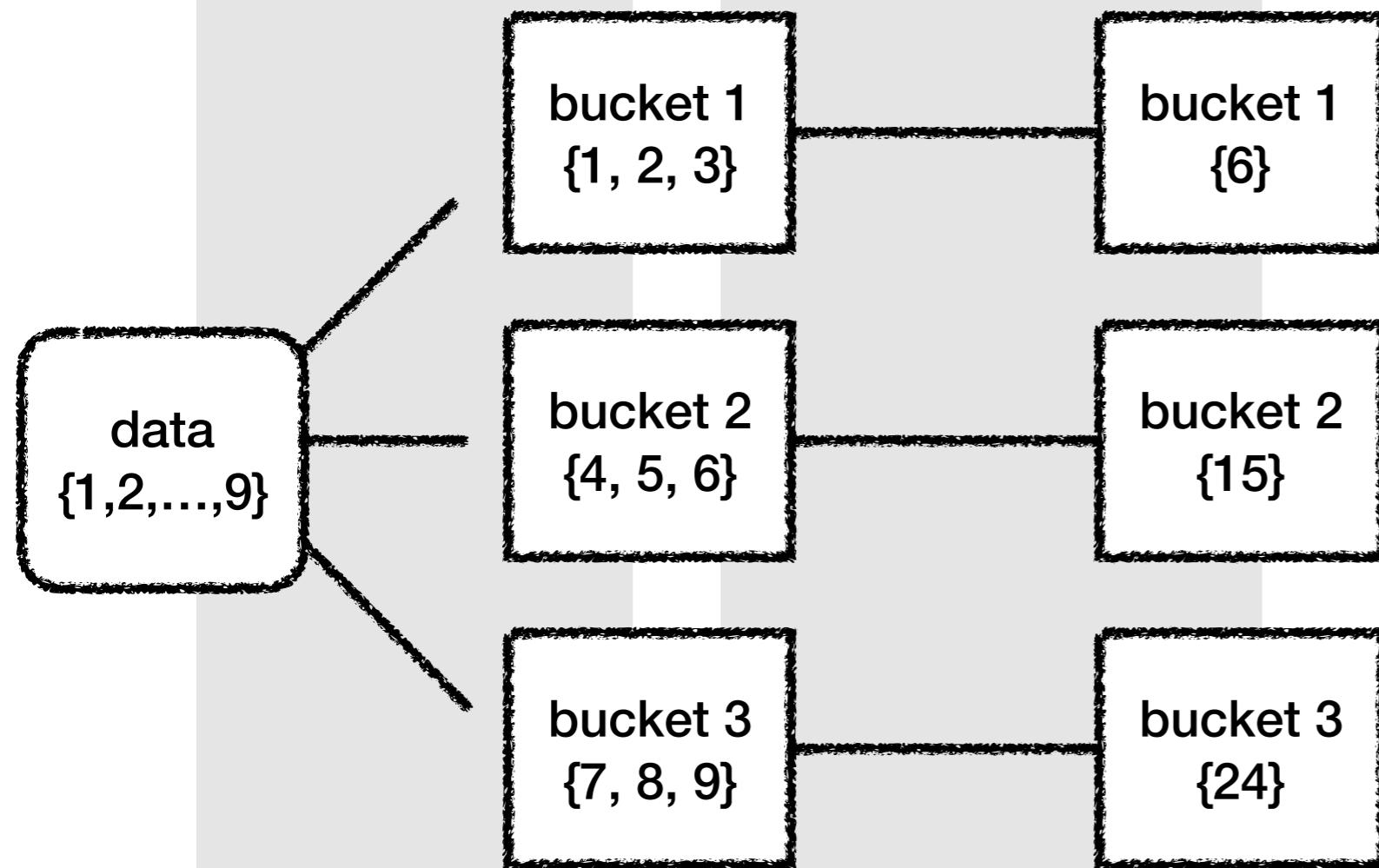
### Metric Aggregation (Sum)

### Sibling Pipeline Aggregation (Sum)



## Bucket Aggregation (Range)

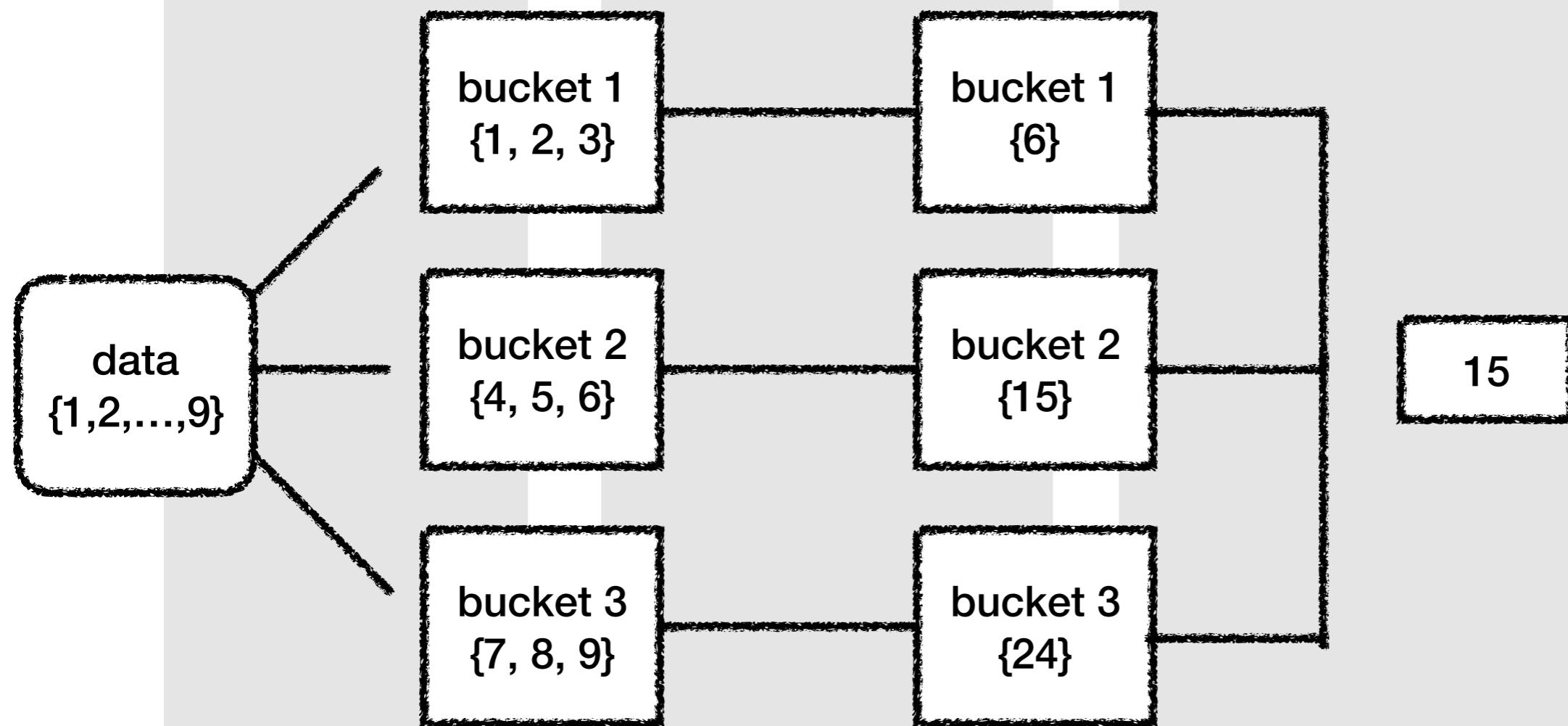
## Metric Aggregation (Sum)



### Bucket Aggregation (Range)

### Metric Aggregation (Sum)

### Sibling Pipeline Aggregation (Average)



## Aggregation - Sibling Pipeline

종류	(제 멋대로) 한 줄 설명	링크
Min Bucket	Bucket Agg 후, Bucket 내 Metric Agg 하고 난 후, Min Aggregation 적용	<a href="#">클릭</a>
Max Bucket	Bucket Agg 후, Bucket 내 Metric Agg 하고 난 후, Max Aggregation 적용	<a href="#">클릭</a>
Sum Bucket	Bucket Agg 후, Bucket 내 Metric Agg 하고 난 후, Sum Aggregation 적용	<a href="#">클릭</a>
Average Bucket	Bucket Agg 후, Bucket 내 Metric Agg 하고 난 후, Average Aggregation 적용	<a href="#">클릭</a>

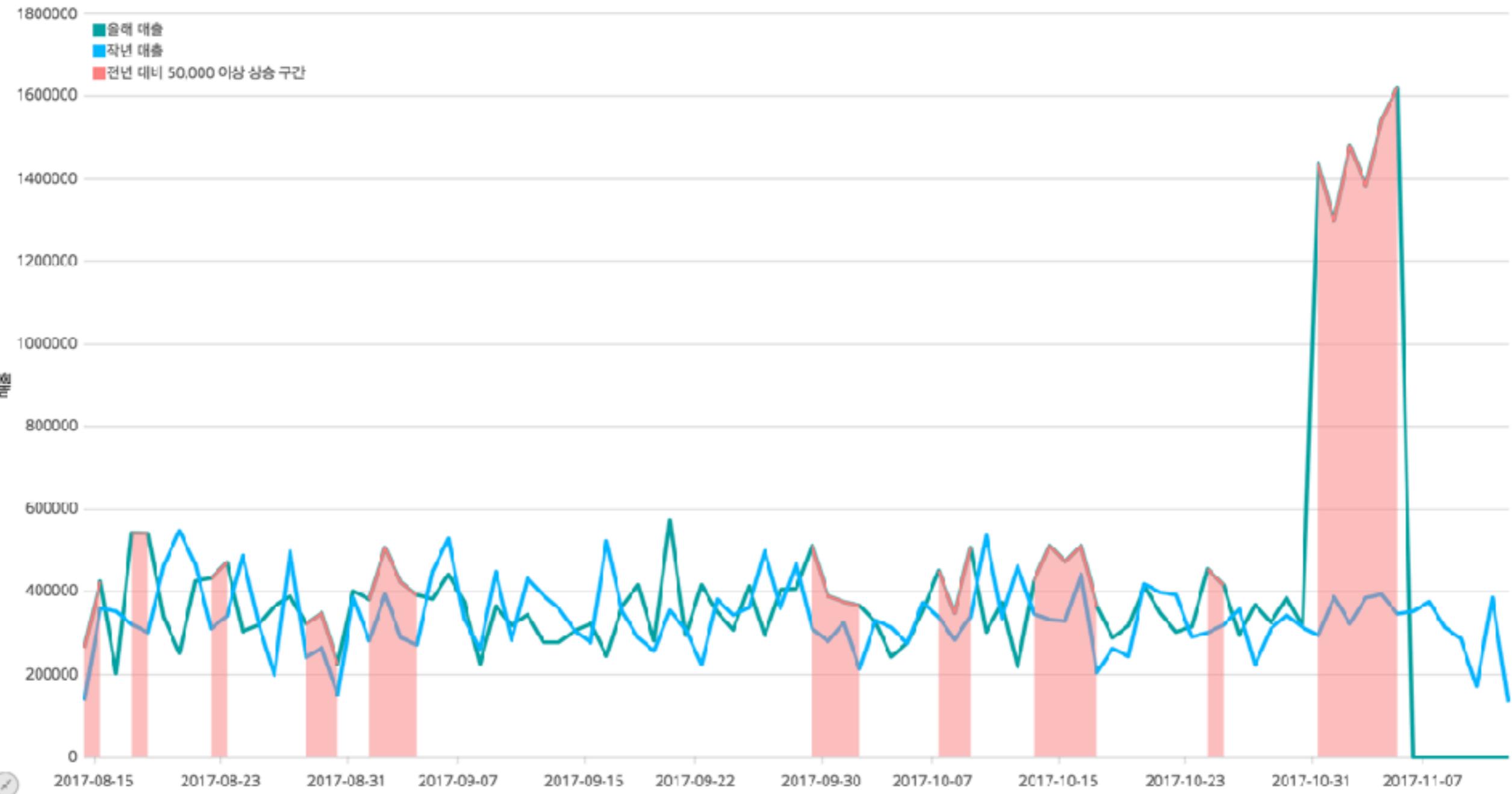
실제 데이터로 어떻게 계산되는지 궁금하다면 [클릭](#)

## Kibana - Timelion

---

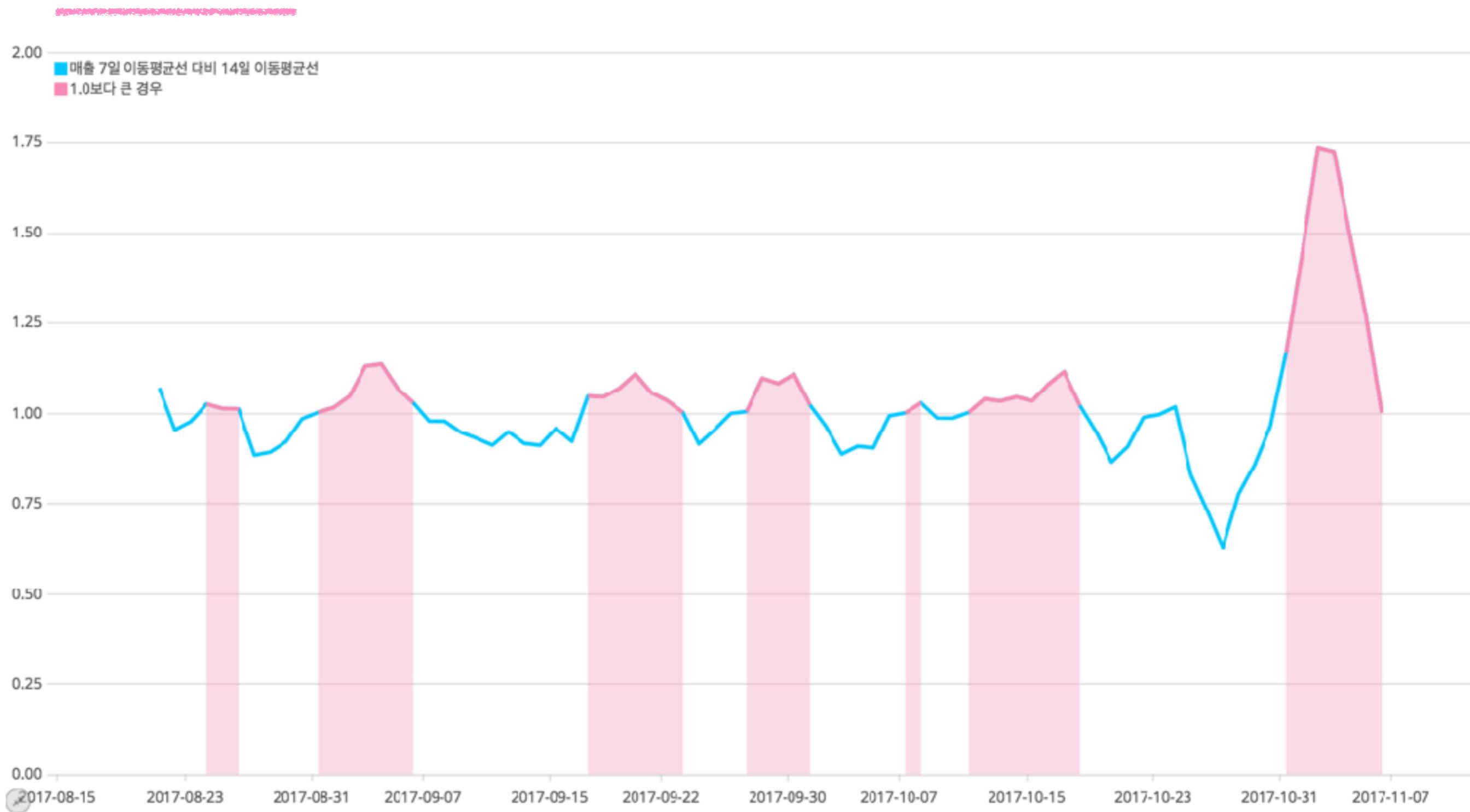
Kibana - Timelion

## Kibana - Timelion



자세히는 [여기](#)를 참조해주세요

## Kibana - Timelion



자세히는 [여기](#)를 참조해주세요

## Dash Board

---

Kibana - Dashboard

## Kibana - Dashboard

The screenshot shows the Kibana interface with the title "Dashboard". On the left, there is a sidebar with icons for Discover, Visualize, Dashboard (which is selected), Timeline, Dev Tools, Management, and a Collapse button. The main area has a search bar at the top with the placeholder "Search...". Below the search bar is a table with one row. The table has two columns: "Name" and "Description". The "Name" column contains the value "shopping", and the "Description" column is empty. At the bottom right of the table, it says "1-1 of 1" with navigation arrows. The overall background is light gray.

자세한 사용법은 [클릭](#)



## 오늘의 목표 - 확인

번호	항목	구체적 내용	(제멋대로) 난이도	확인
1	Elastic Stack Workflow	하나하나가 어떤 역할을 하고 어떤 흐름으로 이어지는지 이해하기	★☆☆☆☆	
1.1	Kibana Workflow	Elastic Stack 중 Kibana에서의 작업 흐름 이해하기	★☆☆☆☆	
1.1.1	Index Pattern	시각화하고자 하는 Index(또는 Indices) 등록하기	★☆☆☆☆	
1.1.2	Discover	Discover를 이용해서 데이터 개략적으로 이해하기	★★☆☆☆	
1.1.3	Visualize Workflow	원하는 형태로 데이터 시각화하기	★★★★☆	
1.1.4	Dashboard	Visualization을 적절히 배치해서 dashboard 만들기	★☆☆☆☆	
2	Aggregation	원하는 결과를 얻기 위해 어떤 Aggregation 사용해야 될 지 이해하기	★★★★★	

## Week2 - 안내

---

이런 거 배우게 (=하게) 됩니다

- Search, Filter
  - Discover에서 특정 조건을 만족하는 Documents만 보고 싶어요.
  - Dashboard에서 전체 UI는 변경하지 않고 특정 조건을 만족하는 값으로만 바뀐 결과를 보고 싶어요
- Scripted Field (Painless) ~~그러나 Painful~~
  - Date Field에서 시간, 요일 등만 뽑아서 사용하고 싶어요.
  - 특정 Number Field 값을 10배로 해서 보고 싶어요.
  - 특정 Number Field의 값에 따라 String Field 만들고 싶어요.
- Visualize 내 JSON Input
  - Moving Average Aggregation을 구할 때 window size를 변경하고 싶어요.
  - Serial Diff Aggregation을 구할 때 3번째 전 값과의 차이를 구하고 싶어요.
- Managing Field
  - Date Field 1은 {YYYY-MM}, Date Field 2는 {MM월DD일} 이렇게 표시하고 싶어요.
- 기타 Kibana에서 알면 좋은 것
- Aggregation 복습 후 Sub Aggregation 등 안내