42. Unique ID Generator

Requirements clarification

Functional requirements

- Mandatory requirements
  - ID must be unique among the whole distributed system.
- Optional requirements
  - IDs can be ordered by time.

Non-functional requirements

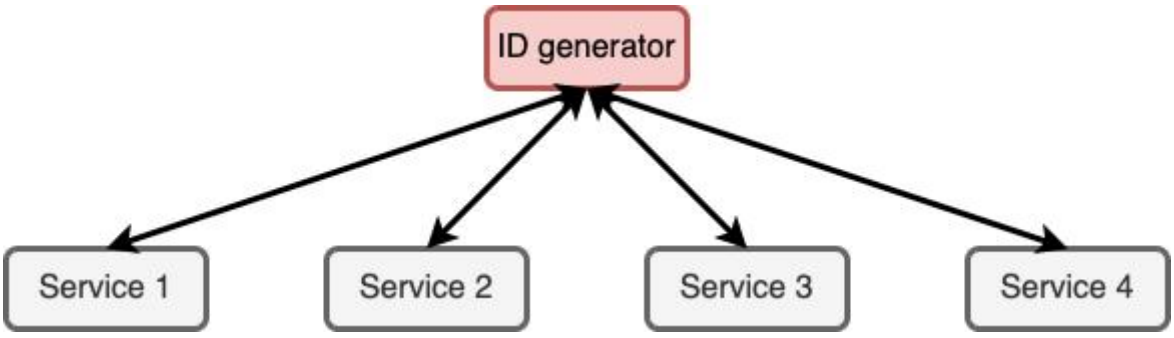
- The generator should be highly available.
- ID generation should happen in real-time with minimal latency.

High-level design

Architecture options

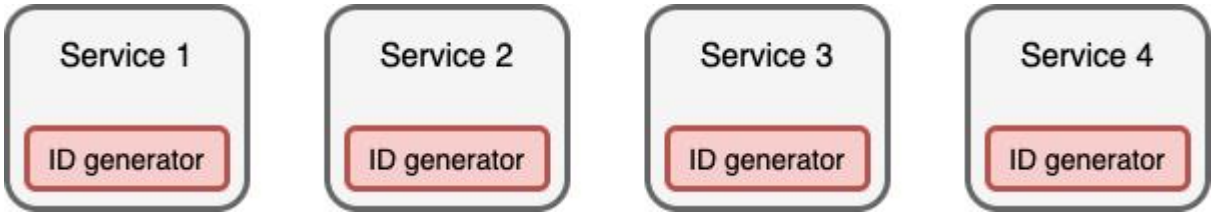
Centralized: Use one centralized service to generate ID.

- Pros
  - Easy to implement.
- Cons
  - Introduce single point of failure.



Decentralized: Each server generates ID respectively.

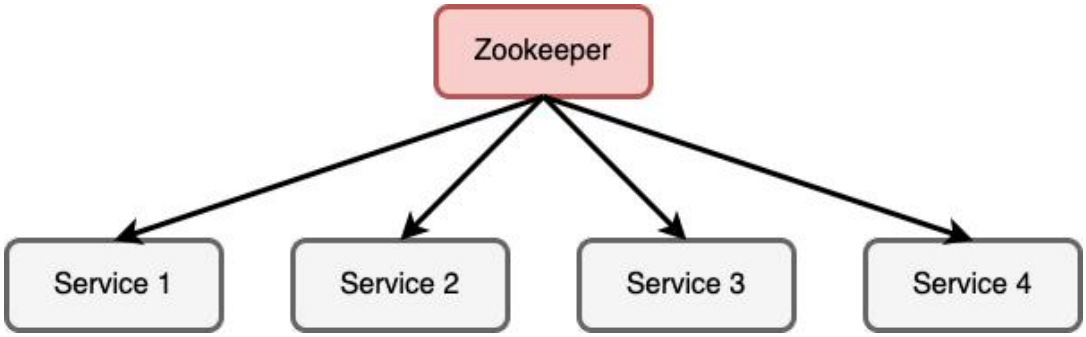
- Pros
  - Improve availability.
  - Cons



Detailed design

Solution options for centralized architecture

- Zookeeper
  - Concept
    - Distributed coordinator to give each server a unique unused range of IDs.



  - Pros
    - Easy to implement.
  - Cons
    - IDs cannot be ordered by time cross multiple servers.
    - Need to consider the size of each range.

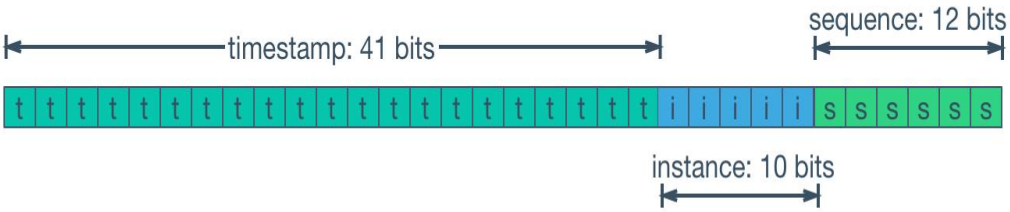
Solution options for decentralized architecture

UUID

- Concepts
  - UUID consists of 32 hexadecimal (base-16) digits, displayed in five groups separated by hyphens.
  - 123e4567-e89b-12d3-a456-426614174000
  - 128-bit.
- Pros
  - Very low probability of getting collision.
  - The ID generators in the services don't need to coordinate each other.
- Cons
  - IDs cannot be ordered by time.

Snowflake ID

- Concepts
  - The format was created by Twitter and is used for the IDs of tweets.
  - Divide an ID into different sections:
    - Sign (1 bit): Always be 0 (Reserved for future used).
    - Timestamp (41 bits): Milliseconds since the epoch or custom epoch.
    - Machine ID (10 bits): Hold up to 1024 machines.
    - Sequence number (12 bits): The sequence number is incremented by 1 and is reset to 0 every millisecond.
    - 64-bit.
    - Each machine can generate IDs by itself.



- Pros
  - IDs can be ordered by time.
  - The binary representation of the timestamp field can be converted to/from a a real date and time.