

# HBVDB-tools

## bvd-merge concept

Jessada Thutkawkorapin

Science for Life Laboratory, Stockholm

August, 2012

# Outline

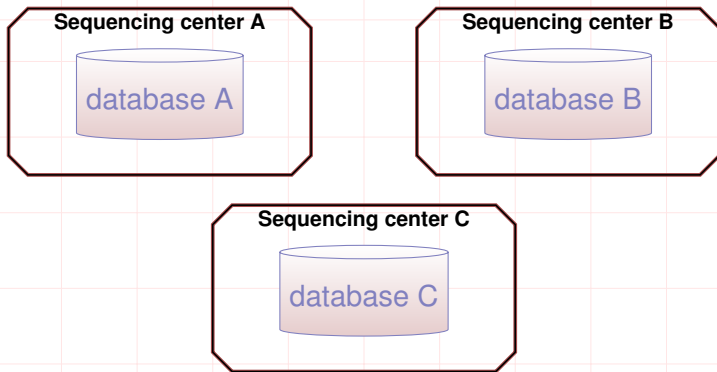
- 1 General idea
  - Overview
  - Read & Write
- 2 Concern & Solution
  - Mutiple exchange & obsolete data
  - Data duplication
  - Confidential issue

# Outline

- 1 General idea
  - Overview
  - Read & Write
- 2 Concern & Solution
  - Multiple exchange & obsolete data
  - Data duplication
  - Confidential issue

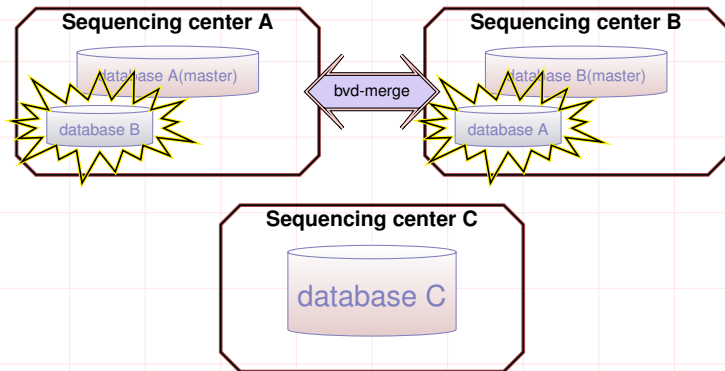
# Overview

We will start the sample exchange scenario by letting every centers only have variant frequencies gathered by themselves.



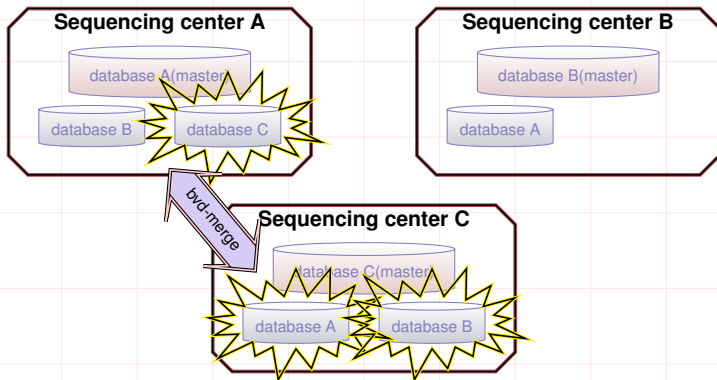
# Overview

First, let center A exchanges the data with center B. Center A will have a copy of database B and center will have a copy of database A.



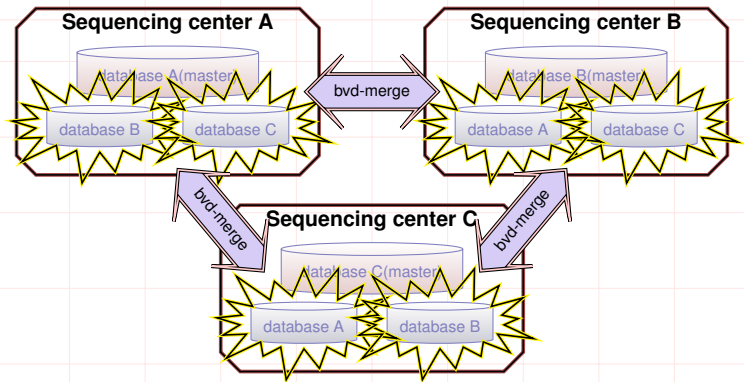
# Overview

Then, center A exchanges the data with center C. Center C will have a copy of databases from center A, which are databases A and B. And center A will have a copy of database C.



# Overview

In conclusion, once the centers exchange their data, they will only make a copy of databases from other centers and keep each of them in the separate databases.



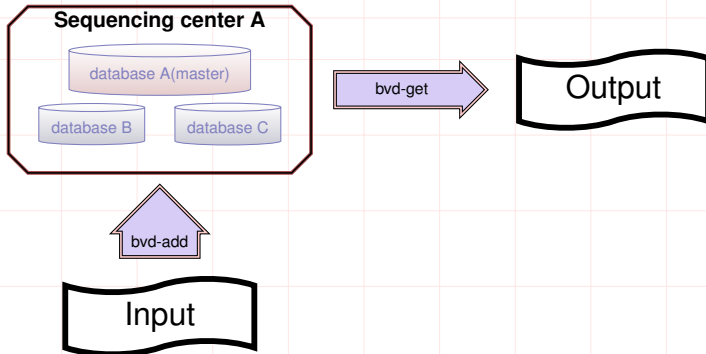
# Outline

- 1 General idea
  - Overview
  - Read & Write
- 2 Concern & Solution
  - Multiple exchange & obsolete data
  - Data duplication
  - Confidential issue



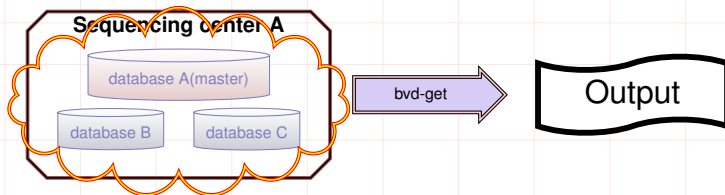
# Read & Write

Reading & Writing the database will be a little complicated under this concept



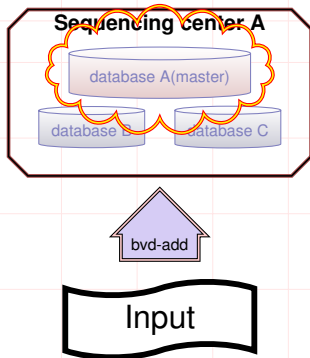
# Read & Write

As it should be, the output, which is retrieved by using `bvd-get`, is the accumulated variant frequencies from all databases, both master and copies, in that center.



# Read & Write

But the update done by `bvd-add` will be done on the master database only.

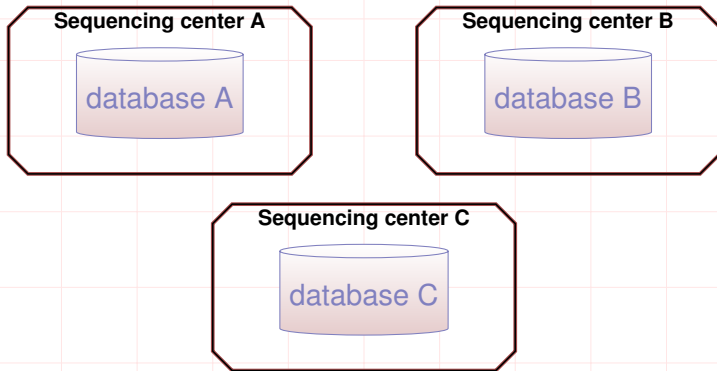


# Outline

- 1 General idea
  - Overview
  - Read & Write
- 2 Concern & Solution
  - Multiple exchange & obsolete data
  - Data duplication
  - Confidential issue

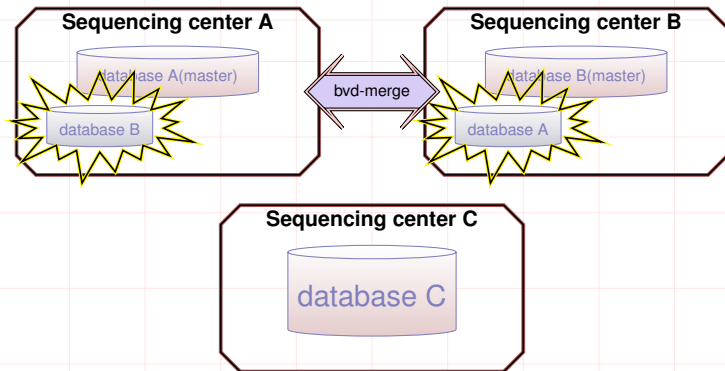
# Multiple exchange & obsolete data

This problem can be easily found and can be simulated by starting from no exchange between any centers.



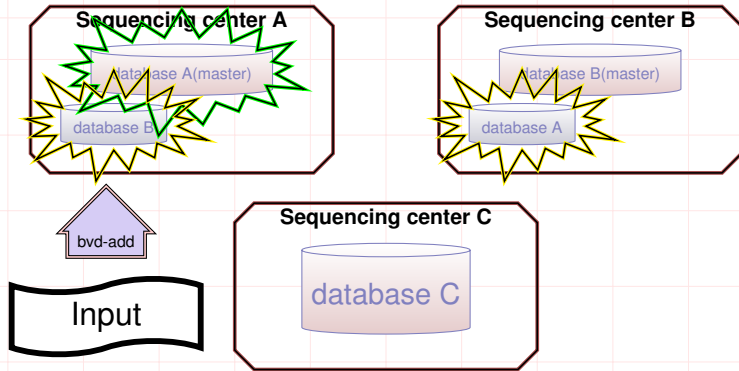
# Multiple exchange & obsolete data

Then center A and B exchange their data.



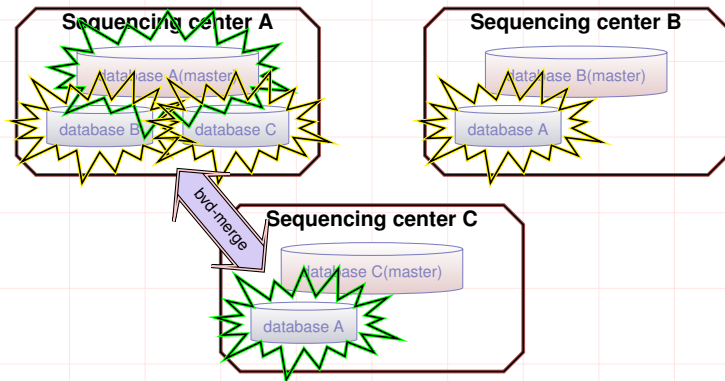
# Multiple exchange & obsolete data

Then another set of variant frequencies is inserted into the database of center A.



## Multiple exchange & obsolete data

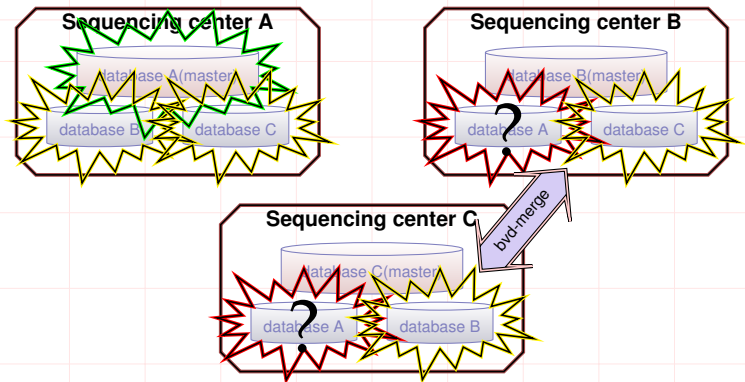
And center A and C exchange their data. With this, the copy of database A in center C is more up-to-date than that in center B.





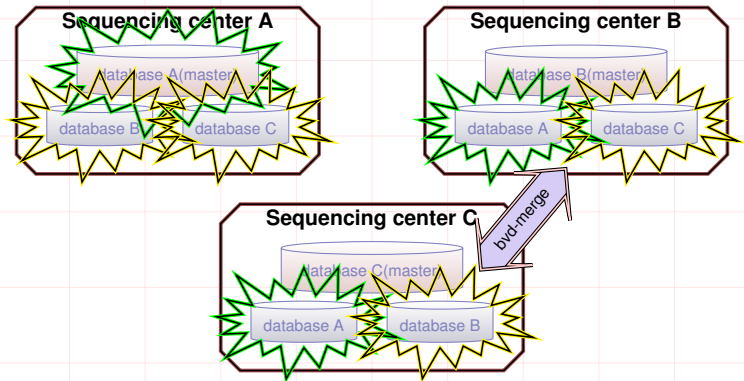
# Multiple exchange & obsolete data

Later when center B and C exchange their data, the database A in center B or C may not be up-to-date if it doesn't have any mechanics to prevent this problem.



# Multiple exchange & obsolete data

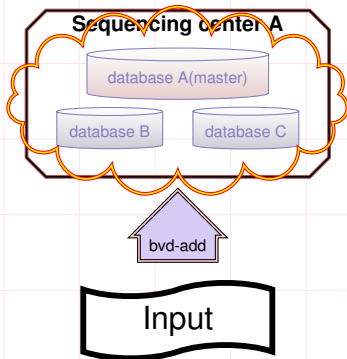
To prevent this, every time that the database is written, **Time Stamp** is recorded to the database so that it can be used later during data exchange to find the most up-to-date database.



# Outline

- 1 General idea
  - Overview
  - Read & Write
- 2 **Concern & Solution**
  - Multiple exchange & obsolete data
  - **Data duplication**
  - Confidential issue

# Data duplication & solution



As bvd-add always did, it still check if inserting content is already in the system. The only difference is that it's not only check with the target database, master, but **bvd-add will check with all the databases in the system, both master and copies, in this case, A, B and C.**

# Outline

- 1 General idea
  - Overview
  - Read & Write
- 2 Concern & Solution
  - Multiple exchange & obsolete data
  - Data duplication
  - Confidential issue

# Confidential issue

This database exchange is a process of copying the whole database, this mean that, as long as there is no confidential issue in each center, there is also no confidential issue in this process.

