



# Loading Genomic Variants

Kim Panjun



# 목차

- Before you begin
- Step1: Create a Google Cloud Storage bucket
- Step2: Upload variants to Google Cloud Storage
- Step3: Import variants to Google Genomics
- Step4: Export variants to Google BigQuery
- <https://cloud.google.com/genomics/v1/load-variants>
- <https://cloud.google.com/products/>

# Before you begin

- Complete the [Quickstart](#).
  - 1. Select or create a Cloud Platform Console project. (Let's try to make own project)
  - 2. Enable the Genomics, BigQuery, and Cloud Storage APIs.
- To get the command-line client, [install the Cloud SDK and Genomics commands](#).
- Enable billing for your project

# Step1: Create a Google Cloud Storage bucket

- Create a bucket using gsutil
  - <https://cloud.google.com/storage/docs/quickstart-gsutil>
  - **Create a bucket**(gsutil mb gs://my-awesome-bucket/ )
  - **Upload an object into your bucket**(Use the gsutil cp command)
  - **Download an object from your bucket**(Use the gsutil cp command)
  - **Copy an object to a folder in the bucket**(Use the gsutil cp command)
  - **List contents of a bucket or folder**(Use the gsutil ls) – detail(gsutil ls -l)
  - **Make your object publicly accessible vs remove the permission**
  - **Give someone access to your bucket vs remove the permission**
  - **Delete objects**(Delete only the objects inside of bucket)
  - **Clean up**(Kill them all!!!!!!!!!!)
- Create a bucket using Cloud Platform Console
  - [https://cloud.google.com/storage/docs/quickstart-console#create\\_a\\_bucket](https://cloud.google.com/storage/docs/quickstart-console#create_a_bucket)

## Step2: Upload variants to Google Cloud Storage

- **Transfer the data**

- To transfer from our Genomic public data!!!  
([http://googlegenomics.readthedocs.io/en/latest/use\\_cases/discover\\_public\\_data/index.html](http://googlegenomics.readthedocs.io/en/latest/use_cases/discover_public_data/index.html))
- To transfer a group of local files using a grouping pattern
- To transfer a local directory tree of files

- **Check the data**

- **Gsutil ls gs://my-bucket-path/my-directory-path**



# Step3: Import variants to Google Genomics

- <https://cloud.google.com/genomics/v1/users-guide>
- **Create a Google Genomics dataset to hold your data(dataset을 만들자)**
  - Ex. `gcloud alpha genomics datasets create --name My-Copy-Of-Platinum-Genomes`
  - Return 결과: Created dataset *dataset\_name*, id: *the-new-dataset-id*
- **Create a variantset(variantset을 만들자)**
  - Ex. `gcloud alpha genomics variantsets create --dataset-id the-new-dataset-id`
  - Return 결과: created variant set id: *variant-set-id*, belonging to dataset id: *the-new-dataset-id*
- **Import your VCFs from Google Cloud Storage to your Google Genomics dataset(Google Cloud Storage에 있던 아까의 vcf파일을 genomics servic로!)**
  - `Gcloud alpha genomics variants import --variantset-id variant-set-id --source-uris gs://이전 스텝에서 실습해서 만들었던 bucket/variants2.vcf, gs://만들었던 bucket/variants_chr*.vcf.` (Note *job\_id*, which you need in the next step. And comma-delimited list of URIs)
- **Check the import job for completion**
  - `gcloud alpha genomics operations describe job_id`(위에서 수행후나온 *job\_id*)

## Step4: Export variants to Google BigQuery

- **Create a BigQuery dataset** in the web UI to hold the data(똑같이 우선 dataset을 !!)
  - (<https://bigquery.cloud.google.com/welcome/>)
- **Export variants to BigQuery**
  - Gcloud alpha genomics variantsets export **variantset-id your-bigquery-table** –bigquer-project **project\_number** –bigquery-dataset **your-bigquery-dataset**
  - Return
    - Done: false
    - Name: **job\_id**
- **Check the export job for completion**
  - Gcloud alpha genomics operations describe **job\_id**
- **Use BigQuery browser tools**