Jun 2021

Google Cloud Healthcare
and Life Sciences

Variant Transforms and BigQuery

https://github.com/googlegenomics/gcp-variant-transforms

Google Cloud

Agenda

- BigQuery overview
- Variant Transforms overview
- Examples

BigQuery

https://cloud.google.com/bigquery

- Highly scalable, columnar storage data warehouse
- Fully managed
- Powered by multiple data centers that each have:
  - Hundreds of thousands of cores
  - Dozens of Petabytes in storage
  - Terabytes of networking bandwidth
- Low cost
  - Storage: $0.02/GB/month (or $0.01/GB/month for long term storage)
  - Query: $5/TB
- Supports standard SQL

Google Cloud

Challenge

How to get variants into BigQuery?

Google Cloud

VCF

standard format
for storing variants

```
##fileformat=VCFv4.3
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS     ID        REF    ALT     QUAL FILTER INFO                              FORMAT      NA00001        NA00002        NA00003
20     14370   rs6054257 G      A       29   PASS   NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:.,.
20     17330   .         T      A       3    q10    NS=3;DP=11;AF=0.017               GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3   0/0:41:3
20     1110696 rs6040355 A      G,T     67   PASS   NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2   2/2:35:4
20     1230237 .         T      .       47   PASS   NS=3;DP=13;AA=T                   GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20     1234567 microsat1 GTC    G,GTCT  50   PASS   NS=3;DP=9;AA=G                    GT:GQ:DP    0/1:35:4       0/2:17:2       1/1:40:3
```

Google Cloud

Challenge

How to get variants into BigQuery?

Solution

Variant Transforms

Google Cloud

Variant Transforms

Open source tool to load VCF files to BigQuery

- Developed by the Google Cloud Healthcare team
- Source of truth on GitHub
- External contributions are welcome!

Highly scalable

- Hundreds of thousands of files
- Millions of samples
- Billions of records

Robustly handles malformed and/or incompatible VCF files

- Fixes missing/incorrect headers
- Gracefully handles invalid records

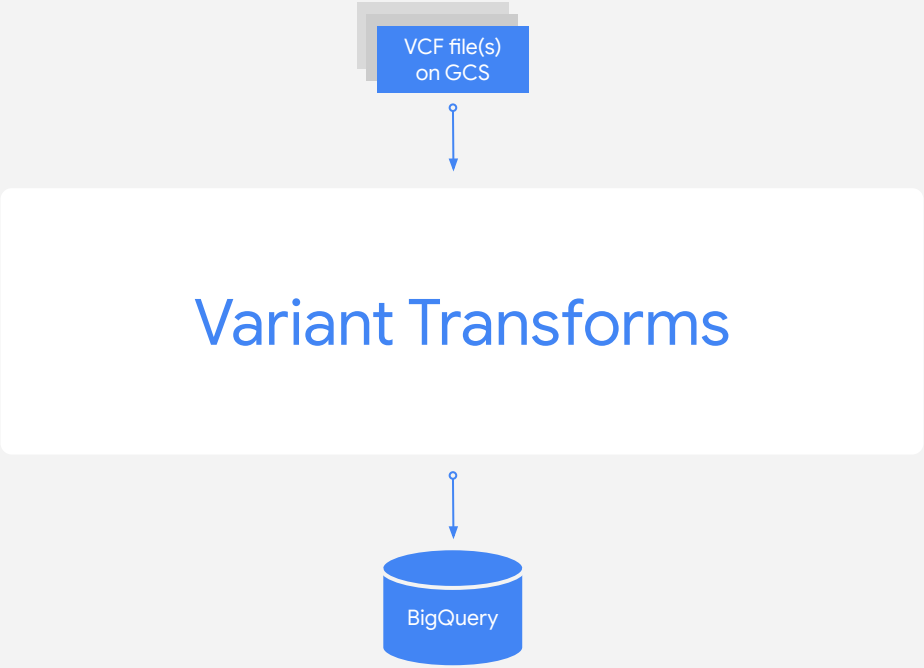
Google Cloud

VCF validator report example

ID	Category	Conflicts	File Paths	Proposed Resolution
GL	FORMAT	num=3 type=Float	gs://.../ALL.chrY.phase1_samtools_si.20	num=None type=Float
		num=None type=Float	gs://.../ALL.wgs.integrated_phase1_v3.2	
			gs://.../ALL.chr18.integrated_phase1_v3	
			gs://.../ALL.chr17.integrated_phase1_v3	
			gs://.../ALL.chr14.integrated_phase1_v3	
			gs://.../ALL.chr7.integrated_phase1_v3.2	
GQ	FORMAT	num=1 type=Float	gs://.../ALL.chrY.genome_strip_hq.2010'	num=1 type=Float
		num=1 type=Integer	gs://.../ALL.chrY.phase1_samtools_si.20	
FT	FORMAT	Undefined header.		num=1 type=String

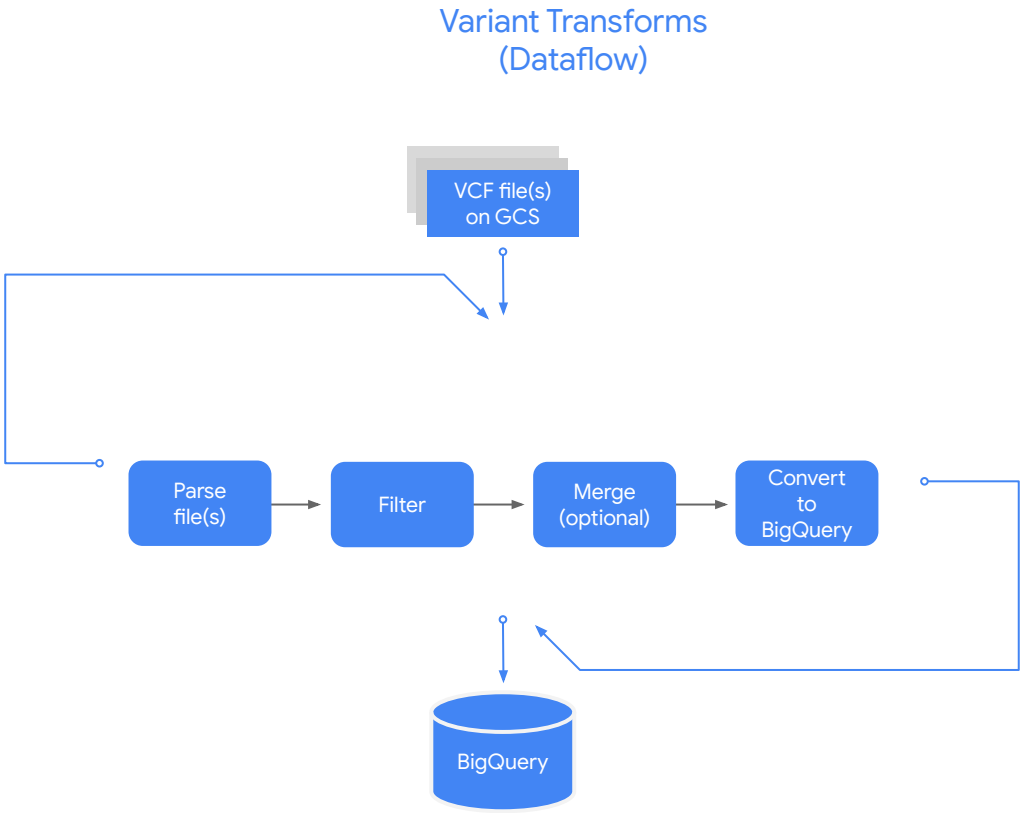
Google Cloud

Architecture



Google Cloud

Architecture



Google Cloud

Example pipeline

Load 2,504 WGS samples from 24 VCF files (800 GiB) into BigQuery tables.

Buckets > genomics-public-data > 1000-genomes-phase-3 > vcf-20150220

Name	Size
ALL.chr1.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.	61.3 GB
ALL.chr10.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotype:	37.8 GB
ALL.chr11.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotype:	38.3 GB
ALL.chr12.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotype:	36.6 GB
ALL.chr13.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotype:	27.1 GB
ALL.chr14.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotype:	25.1 GB
ALL.chr15.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotype:	23 GB
ALL.chr16.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotype:	25.6 GB
ALL.chr17.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotype:	22.1 GB
ALL.chr18.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotype:	21.5 GB
ALL.chr19.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotype:	17.4 GB
ALL.chr2.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.	67.1 GB
ALL.chr20.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotype:	17.2 GB
ALL.chr21.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotype:	10.5 GB
ALL.chr22.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotype:	10.5 GB
ALL.chr3.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.	55.2 GB

Google Cloud

Example pipeline

Load 2,504 WGS samples from 24 VCF files (800 GiB) into BigQuery tables.

```
GOOGLE_CLOUD_PROJECT=myproject
GOOGLE_CLOUD_REGION=us-central1
TEMP_LOCATION=gs://mybucket/variant_import/tmp
OUTPUT_TABLE=myproject:my_dataset.1000_genomes_phase_3_variants
INPUT_PATTERN=gs://genomics-public-data/1000-genomes-phase-3/vcf-20150220/*.vcf

COMMAND="vcf_to_bq  \
  --input_pattern ${INPUT_PATTERN} \
  --output_table ${OUTPUT_TABLE} \
  --job_name genomes \
  --worker_machine_type n1-standard-16 \
  --num_workers 312 \
  --worker_disk_type compute.googleapis.com/projects//zones//diskTypes/pd-ssd \
  --disk_size_gb 500  --infer_headers --runner DataflowRunner"

docker run -v ~/.config:/root/.config \
  gcr.io/cloud-lifesciences/gcp-variant-transforms \
  --project "${GOOGLE_CLOUD_PROJECT}" \
  --region "${GOOGLE_CLOUD_REGION}" \
  --temp_location "${TEMP_LOCATION}" \
  "${COMMAND}"
```

Google Cloud

Dataflow

genomes

STOP | IMPORT AS PIPELINE | SHARE | MAX TIME

JOB GRAPH | EXECUTION DETAILS | JOB METRICS

Job steps view
Graph view

CLEAR SELECTION

ReadFromVcf
Running
12 min 1 sec
0 of 1 stage succeeded

ShardVariants
Running
1 min 3 sec
0 of 1 stage succeeded

ProcessVari...tsresidual
Running
2 sec
0 of 1 stage succeeded

ProcessVariantschr6
Running
6 sec
0 of 1 stage succeeded

ProcessVariantschr20
Running
3 sec
0 of 1 stage succeeded

ProcessVariant
Running
4 sec
0 of 1 stage succ

VariantToAvroresidual
Running
7 sec
1 of 5 stages succeeded

VariantToAvrochr6
Running
1 min 41 sec
1 of 5 stages succeeded

VariantToAvrochr20
Running
35 sec
1 of 5 stages succeeded

VariantToAvro
Running
48 sec
1 of 5 stages suc

Logs | SHOW | 1

Job info

Job name	genomes
Job ID	2021-06-17_16_50_03-14869564267829150444
Job type	Batch
Job status	Running
SDK version	Apache Beam Python 3.7 SDK 2.24.0
Job region	us-central1
Worker location	us-central1-b
Current workers	312
Latest worker status	Autoscaling: Raised the number of workers to 312 based on the rate of progress in the currently running stage(s).
Start time	June 17, 2021 at 4:50:05 PM GMT-7
Elapsed time	3 min 20 sec
Encryption type	Google-managed key

Resource metrics

Current vCPUs	4,992
Total vCPU time	190.241 vCPU hr
Current memory	18.28 TB
Total memory time	713.404 GB hr
Current HDD PD	0 B
Total HDD PD time	0 GB hr
Current SSD PD	152.34 TB

Google Cloud

Dataflow

(after 71 mins)

genomes

+ IMPORT AS PIPELINE

SHARE

MAX TIME

JOB GRAPH | EXECUTION DETAILS | JOB METRICS

Job steps view
Graph view

CLEAR SELECTION

ReadFromVcf
Succeeded
43 days 44 min 31 sec
1 of 1 stage succeeded

ShardVariants
1,298 elements/s
1 day 2 hr 58 min 45 sec
1 of 1 stage succeeded

ProcessVariantschr9
53 elements/s
16 min 13 sec
1 of 1 stage succeeded

ProcessVariantschr21
17 elements/s
7 min 16 sec
1 of 1 stage succeeded

ProcessVariantschr8
70 elements/s
20 min 14 sec
1 of 1 stage succeeded

ProcessVariantschr14
40 elements/s
12 min 19 sec
1 of 1 stage succeeded

VariantToAvrochr9
53 elements/s

VariantToAvrochr21
17 elements/s

VariantToAvrochr8
70 elements/s

VariantToAvrochr14
40 elements/s

Job info

Job name	genomes
Job ID	2021-06-17_16_50_03-14869564267829150444
Job type	Batch
Job status	Succeeded
SDK version	Apache Beam Python 3.7 SDK 2.24.0
Job region	us-central1
Worker location	us-central1-b
Current workers	0
Latest worker status	Worker pool stopped.
Start time	June 17, 2021 at 4:50:05 PM GMT-7
Elapsed time	1 hr 11 min
Encryption type	Google-managed key

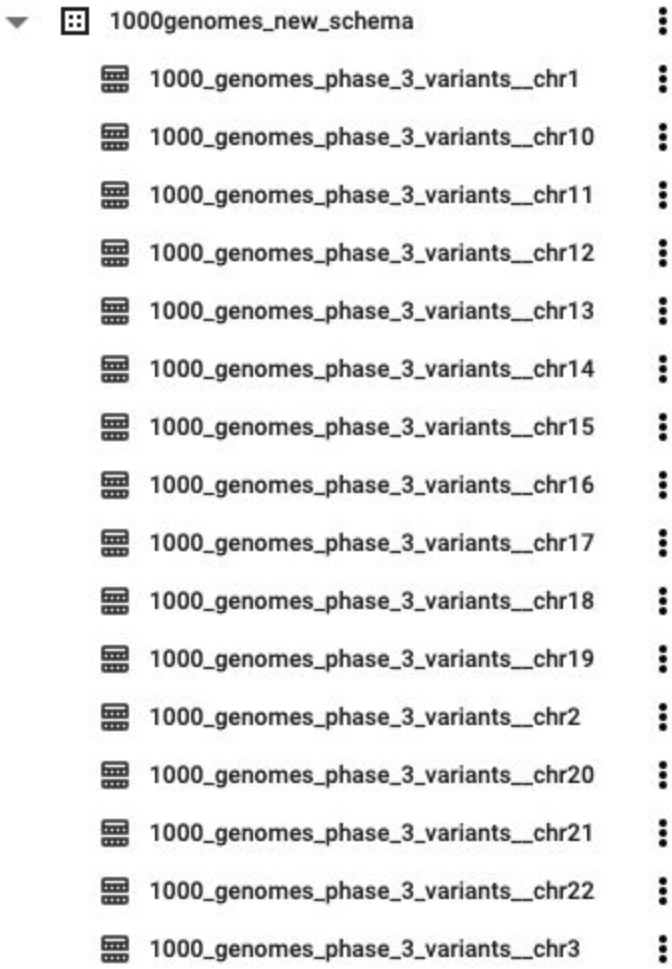
Resource metrics

Current vCPUs	4,992
Total vCPU time	5,582.371 vCPU hr
Current memory	18.28 TB
Total memory time	20,933.891 GB hr
Current HDD PD	0 B
Total HDD PD time	0 GB hr
Current SSD PD	152.34 TB
Total SSD PD time	174,449.095 GB hr

Google Cloud

BigQuery Tables

(1 per chromosome)



Google Cloud

BigQuery Tables

(Partitioned and clustered chromosome 1)

Table info

EDIT DETAILS

Table ID	[redacted]:1000genomes_new_schema.1000_genomes_phase_3_variants__chr1
Table size	438.22 GB
Long-term storage size	0 B
Number of rows	6,468,094
Created	Jun 17, 2021, 6:02:11 PM UTC-7
Last modified	Jun 17, 2021, 7:00:05 PM UTC-7
Table expiration	NEVER
Data location	US
Description	

Table Type	Partitioned
Partitioned by	Integer Range
Partitioned on field	start_position
Partition Range Start	0
Partition Range End	249297660
Partition Range Interval	62340
Partition filter	Not required

Clustered by	start_position end_position

Google Cloud

BigQuery Tables

(Table schema)

SCHEMA DETAILS PREVIEW

Field name	Type	Mode	Policy Tags	Description
reference_name	STRING	NULLABLE		Reference name.
start_position	INTEGER	NULLABLE		Start position (1-based). Corresponds to the first base of the string of reference bases.
end_position	INTEGER	NULLABLE		End position. Corresponds to the first base after the last base in the reference allele.
reference_bases	STRING	NULLABLE		Reference bases.
alternate_bases	RECORD	REPEATED		One record for each alternate base (if any).
names	STRING	REPEATED		Variant names (e.g. RefSNP ID).
quality	FLOAT	NULLABLE		Phred-scaled quality score (-10log10 prob(call is wrong)). Higher values imply better quality.
filter	STRING	REPEATED		List of failed filters (if any) or "PASS" indicating the variant has passed all filters.
call	RECORD	REPEATED		One record for each call.
CIEND	INTEGER	REPEATED		Confidence interval around END for imprecise variants
CIPOS	INTEGER	REPEATED		Confidence interval around POS for imprecise variants
CS	STRING	NULLABLE		Source call set.
IMPRECISE	BOOLEAN	NULLABLE		Imprecise structural variation
MC	STRING	REPEATED		Merged calls.
MEINFO	STRING	REPEATED		Mobile element info of the form NAME,START,END<POLARITY; If there is only 5' OR 3' support for this call, will be NULL NULL for START and END
MEND	INTEGER	NULLABLE		Mitochondrial end coordinate of inserted sequence
MLEN	INTEGER	NULLABLE		Estimated length of mitochondrial insert
MSTART	INTEGER	NULLABLE		Mitochondrial start coordinate of inserted sequence
SVLEN	INTEGER	REPEATED		SV length. It is only calculated for structural variation MEIs. For other types of SVs; one may calculate the SV length by INFO:END-START+1, or by finding the difference between lengthes of REF and ALT alleles
SVTYPE	STRING	NULLABLE		Type of structural variant
TSD	STRING	NULLABLE		Precise Target Site Duplication for bases, if unknown, value will be NULL

Google Cloud

BigQuery Table (Data preview)

Row	reference_name	start_position	end_position	reference_bases	alternate_base	names	quality	call.sample_id	call.genotype	call.phaseset	call.CN	call.CNL	call.CNP	call.CNQ
1	1	207782714	207782714	G	T	rs201224574	100.0	1050000895148064400	0	*	null			null
									0					
								7229853199564655891	0	*	null			null
									0					
								4516189868442699828	0	*	null			null
									0					
								7320282954593162247	0	*	null			null
									0					
								2333474091774200307	0	*	null			null
									0					
								5741251685300651846	0	*	null			null
									0					
								4207861292840416466	0	*	null			null
									0					

Google Cloud

Example query

(sex inference)

RUN | MORE | SAVE | SCHEDULE | This query will process 163.3 GiB when run.

```
WITH filtered_snp_calls AS (
  SELECT
    c.sample_id,
    CAST((SELECT LOGICAL_AND(g > 0) FROM UNNEST(c.genotype) AS g) AS INT64) AS hom_AA,
    CAST(EXISTS (SELECT g FROM UNNEST(c.genotype) AS g WHERE g > 0)
      AND EXISTS (SELECT g FROM UNNEST(c.genotype) AS g WHERE g = 0) AS INT64) AS het_RA
  FROM
    `[redacted].1000genomes_new_schema.1000_genomes_phase_3_variants__chrX` AS v, UNNEST(v.call) AS c
  WHERE
    # Only include biallelic snps.
    reference_bases IN ('A','C','G','T')
    AND alternate_bases[ORDINAL(1)].alt IN ('A','C','G','T')
    AND (ARRAY_LENGTH(alternate_bases) = 1
      OR (ARRAY_LENGTH(alternate_bases) = 2 AND alternate_bases[ORDINAL(2)].alt = '<*>'))
)

SELECT
  sample_id,
  ROUND(SAFE_DIVIDE(SUM(het_RA), SUM(hom_AA) + SUM(het_RA)), 3) AS perct_het_alt_in_snvs,
  ROUND(SAFE_DIVIDE(SUM(hom_AA), SUM(hom_AA) + SUM(het_RA)), 3) AS perct_hom_alt_in_snvs,
  SUM(hom_AA) AS hom_AA_count,
  SUM(het_RA) AS het_RA_count
FROM filtered_snp_calls
GROUP BY
  sample_id
ORDER BY
  sample_id
```

Google Cloud

Example query

(sex inference)

Query complete (15.7 sec elapsed, 163.3 GB processed)

Job information | **Results** | JSON | Execution details

Row	sample_id	perct_het_alt_in_snvs	perct_hom_alt_in_snvs	hom_AA_count	het_RA_count
1	583477157765007	0.668	0.332	56504	113712
2	25242456645083931	0.659	0.341	47518	91876
3	30245695142467316	0.05	0.95	81264	4288
4	32858136970890294	0.63	0.37	44453	75536
5	44073384264225292	0.512	0.488	55696	58415
6	46029689092401278	0.053	0.947	79231	4442
7	47812019554178849	0.057	0.943	75211	4563
8	52138393317078276	0.655	0.345	57346	108735
9	52195536888289121	0.044	0.956	83179	3861
10	57078017347152026	0.048	0.952	77307	3886
11	59245965522197874	0.047	0.953	82744	4125
12	63709839326451994	0.057	0.943	89209	5430
13	68374849871561002	0.035	0.965	82311	3019
14	69431804903828540	0.058	0.942	79482	4904
15	76538193160765155	0.54	0.46	52603	61782

Rows per page: 100 | 1 - 100 of 2504 | First page

Google Cloud

Annotations

Native support for parsing annotation fields from VEP

```
CSQ=G|upstream_gene_variant|MODIFIER|PSMF1|ENSG00000125818|Transcript|ENST00000333082|protein_coding||||||||||2567|1||HGNC|HGNC:9571|1|P1|ENSP00000327704||||||||||||||||||||||||||,T|upstream_gene_variant|MODIFIER|PSMF1|ENSG00000125818|Transcript|ENST00000333082|protein_coding||||||||||2567|1||HGNC|HGNC:9571|1|P1|ENSP00000327704||||||||||||||||||||||||||||,G|upstream_gene_variant|MODIFIER|PSMF1|ENSG00000125818|Transcript|ENST00000381899|protein_coding||||||||||2582|1|cds_end_NF|HGNC|HGNC:9571|2||ENSP00000371324||||...
```

alternate_bases.**CSQ**	RECORD	REPEATED	List
alternate_bases.CSQ.**Consequence**	STRING	NULLABLE	Des
alternate_bases.CSQ.**IMPACT**	STRING	NULLABLE	Des
alternate_bases.CSQ.**SYMBOL**	STRING	NULLABLE	Des
alternate_bases.CSQ.**Gene**	STRING	NULLABLE	Des
alternate_bases.CSQ.**Feature_type**	STRING	NULLABLE	Des
alternate_bases.CSQ.**Feature**	STRING	NULLABLE	Des
alternate_bases.CSQ.**BIOTYPE**	STRING	NULLABLE	Des
alternate_bases.CSQ.**EXON**	STRING	NULLABLE	Des
alternate_bases.CSQ.**INTRON**	STRING	NULLABLE	Des

Google Cloud

Spec: http://snpeff.sourceforge.net/VCFannotationformat_v1.0.pdf

Example query using annotations

Find all high impact variants in BRCA1 genes:

RUN MORE This query will process 6.7 GiB when run.

```
SELECT
    reference_name AS CHROM,
    start_position AS POS,
    reference_bases AS REF,
    alternate_bases.alt AS ALT,
    vep.IMPACT AS Impact,
    vep.SYMBOL AS Symbol,
    vep.Gene AS Gene,
    vep.Consequence AS Consequence,
FROM
    `bigquery-public-data.gnomAD.v3_genomes__chr17` AS main_table,
    main_table.alternate_bases AS alternate_bases,
    alternate_bases.vep AS vep
WHERE
    Symbol ="BRCA1"
    AND Impact = "HIGH"
GROUP BY 1,2,3,4,5,6,7,8
```

Google Cloud

Example query using annotations

Find all high impact variants in BRCA1 genes:

Query complete (4.0 sec elapsed, 6.7 GB processed)

Job information | **Results** | JSON | Execution details

Row	CHROM	POS	REF	ALT	Impact	Symbol	Gene	Consequence
1	chr17	43093568	CT	C	HIGH	BRCA1	ENSG00000012048	frameshift_variant
2	chr17	43051117	C	G	HIGH	BRCA1	ENSG00000012048	splice_acceptor_variant&NMD_tran
3	chr17	43094415	C	T	HIGH	BRCA1	ENSG00000012048	stop_gained
4	chr17	43067623	A	ATGAG	HIGH	BRCA1	ENSG00000012048	frameshift_variant
5	chr17	43119266	T	A	HIGH	BRCA1	ENSG00000012048	splice_acceptor_variant
6	chr17	43110580	C	A	HIGH	BRCA1	ENSG00000012048	splice_acceptor_variant&NMD_tran
7	chr17	43091399	A	AGACTG	HIGH	BRCA1	ENSG00000012048	frameshift_variant
8	chr17	43091487	CCT	C	HIGH	BRCA1	ENSG00000012048	frameshift_variant
9	chr17	43074432	G	A	HIGH	BRCA1	ENSG00000012048	stop_gained
10	chr17	43124026	ACT	A	HIGH	BRCA1	ENSG00000012048	frameshift_variant
11	chr17	43045710	G	GT	HIGH	BRCA1	ENSG00000012048	stop_gained&frameshift_variant
12	chr17	43092524	GT	G	HIGH	BRCA1	ENSG00000012048	frameshift_variant
13	chr17	43063344	TTTTC	T	HIGH	BRCA1	ENSG00000012048	frameshift_variant
14	chr17	43125275	C	A	HIGH	BRCA1	ENSG00000012048	splice_donor_variant&NMD_transcr
15	chr17	43091923	G	A	HIGH	BRCA1	ENSG00000012048	stop_gained
16	chr17	43091433	C	T	HIGH	BRCA1	ENSG00000012048	splice_donor_variant&NMD_transcr

Google Cloud

Example query using Annotations

Find all high impact variants involved in double-strand break repair (GO:0006302):

RUN MORE This query will process 6.2 GiB when run.

```
SELECT
    reference_name AS CHROM,
    start_position AS POS,
    reference_bases AS REF,
    alternate_bases.alt AS ALT,
    vep.IMPACT AS Impact,
    vep.SYMBOL AS Symbol,
    vep.Consequence AS Consequence,
FROM
    `bigquery-public-data.gnomAD.v3_genomes__chr8` AS main_table,
    main_table.alternate_bases AS alternate_bases,
    alternate_bases.vep AS vep
WHERE
    Symbol IN (SELECT DB_Object_Symbol
               FROM `isb-cgc.genome_reference.GO_Annotations`
               WHERE GO_ID = 'GO:0006302')
    AND Impact = "HIGH"
GROUP BY 1,2,3,4,5,6,7
```

Google Cloud

Example query using Annotations

Find all high impact variants involved in double-strand break repair (GO:0006302):

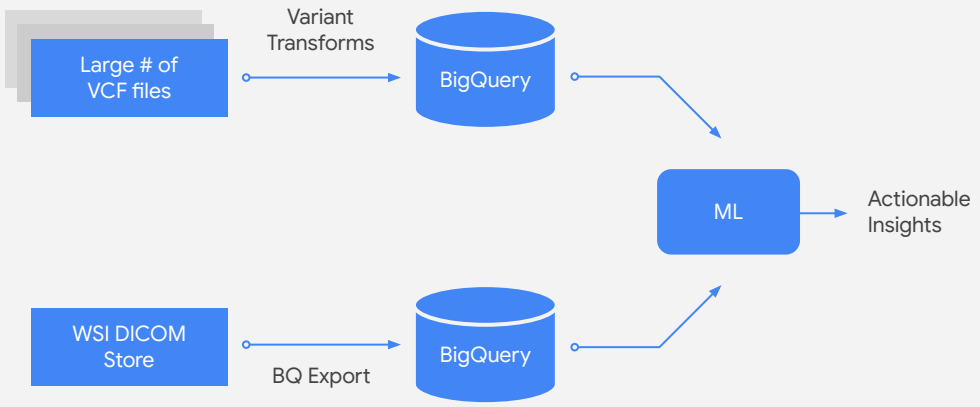
Query complete (4.2 sec elapsed, 6.2 GB processed)

Job information | Results | JSON | Execution details

Row	CHROM	POS	REF	ALT	Impact	Symbol	Consequence
1	chr8	144516246	GC	G	HIGH	RECQL4	frameshift_variant
2	chr8	71216736	T	A	HIGH	EYA1	stop_gained
3	chr8	89964437	G	T	HIGH	NBN	stop_gained
4	chr8	116852706	AGCT	A	HIGH	RAD21	splice_acceptor_variant&coding_sequence_variant
5	chr8	116852708	C	A	HIGH	RAD21	splice_acceptor_variant
6	chr8	31076197	GAACA	G	HIGH	WRN	frameshift_variant
7	chr8	27784047	AT	A	HIGH	ESCO2	frameshift_variant
8	chr8	31064912	A	G	HIGH	WRN	splice_acceptor_variant
9	chr8	89946250	C	CT	HIGH	NBN	frameshift_variant
10	chr8	144512289	G	A	HIGH	RECQL4	stop_gained&NMD_transcript_variant
11	chr8	89982734	GAA	G	HIGH	NBN	frameshift_variant
12	chr8	27787981	A	AT	HIGH	ESCO2	frameshift_variant
13	chr8	89984513	TCTGCCCTTACCTC	T	HIGH	NBN	splice_donor_variant&coding_sequence_variant&intron_variant
14	chr8	71216689	G	A	HIGH	EYA1	splice_donor_variant
15	chr8	71404915	T	C	HIGH	EYA1	splice_acceptor_variant&non_coding_transcript_variant
16	chr8	144512410	C	T	HIGH	RECQL4	stop_gained

Google Cloud

Case Study: Color



Blog post:
https://cloud.google.com/blog/big-data/2018/03/how-color-uses-the-new-variant-transforms-tool-for-breakthrough-clinical-data-science-with-bigquery

Google Cloud

Questions?

Visit our GitHub page for roadmap & feature requests:
github.com/googlegenomics/gcp-variant-transforms