Re-engineering the Hazard Calculation: toshi-hazard-post and toshi-hazard-store v2

Chris DiCaprio and Chris Chamberlain

Te Tauira Matapae Pūmate Rū i Aotearoa
NSHM The New Zealand National Seismic Hazard Model
A GNS Science Led Research Programme

E mahi ana me
In collaboration with

Ngā hoa tuku pūtea
Funding partners
MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT HĪKINA WHAKATUTUKI
Toka Tū Ake EQC

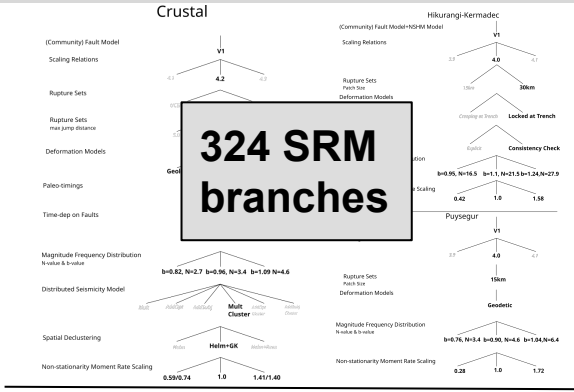
Motivation

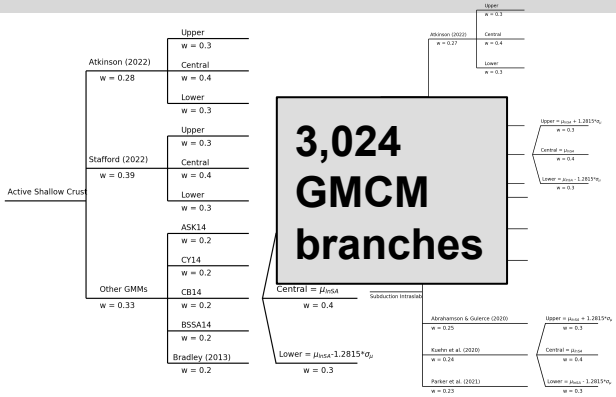
`toshi-hazard-store` and `toshi-hazard-post` facilitate calculation of large logic trees

Issues

- **Development was done under considerable time pressure**
  - We used tools already used by project and at our disposal: ToshiAPI, dynamoDB
  - Significant technical debt incurred: “it ain’t pretty, but it works”
- **Not user friendly:**
  - Burdensome workflow
  - Requires mimicking NSHM project’s IT stack; not realistic for non-GNS users
- **Performance left on the table**

Refresh: Hazard Calculation





979,776 total branches

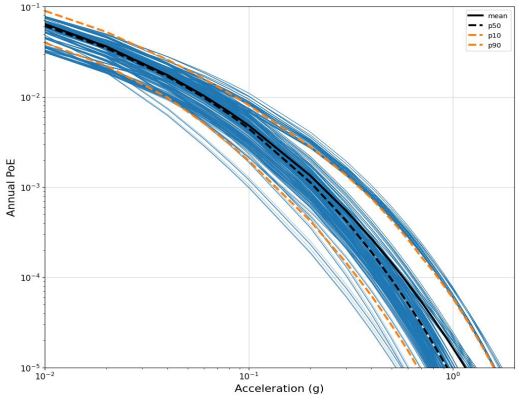
Decompose the model and break calculation into 2 stages

Stage 1

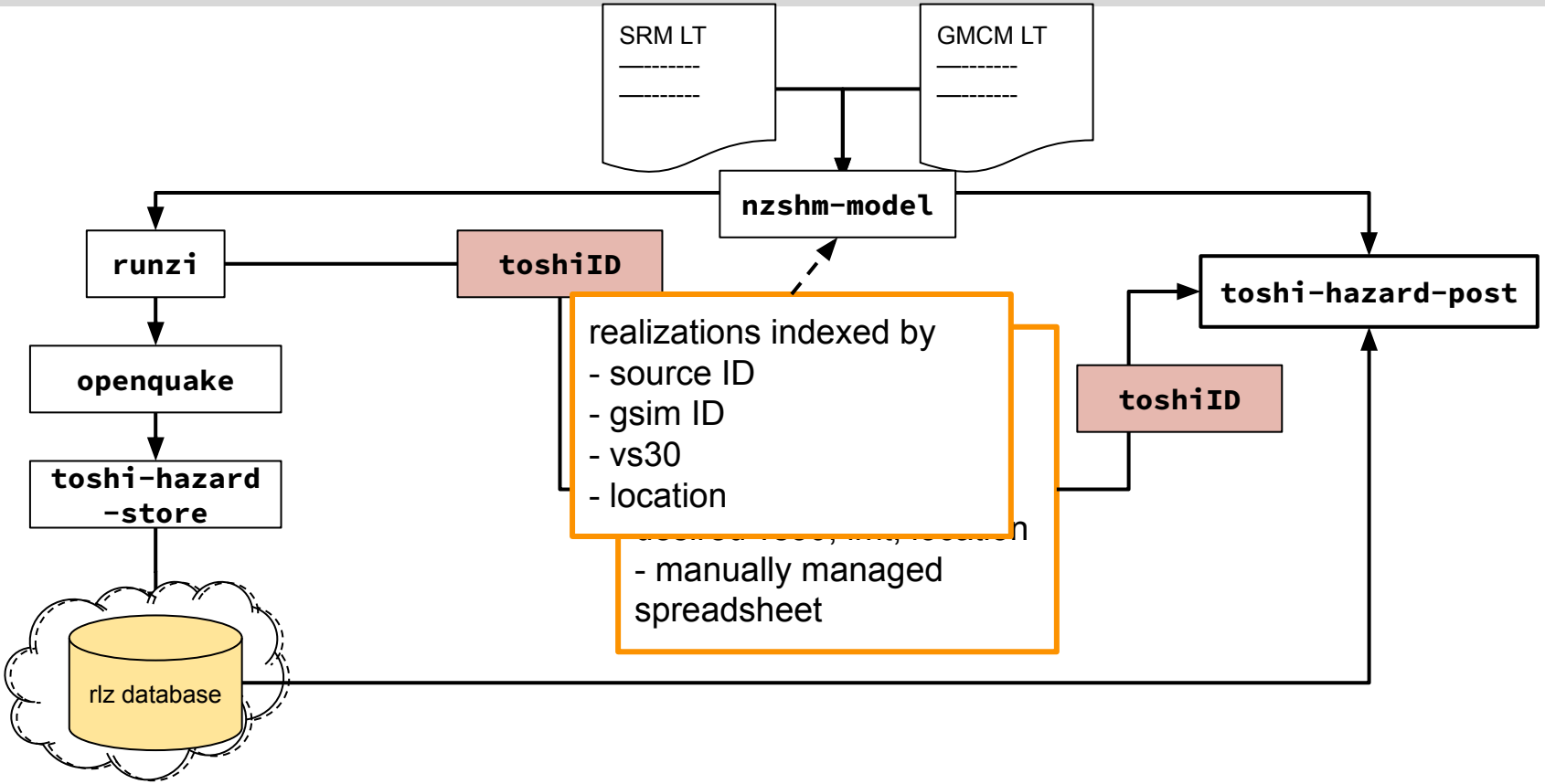
- Calculate hazard for independent components concurrently
- Store **component realizations**

Stage 2

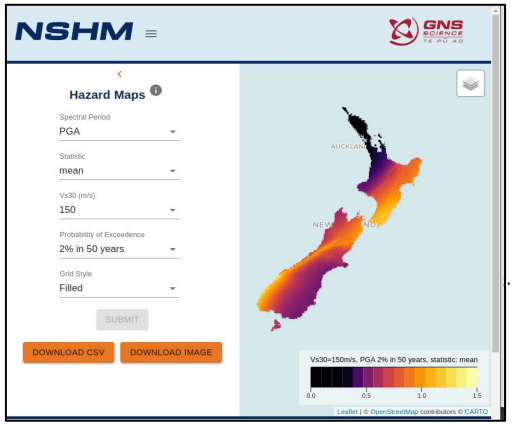
- Combine independent components to form 979,776 realizations
- Calculate aggregate statistics (e.g. weighted mean and fractiles)



Hazard Workflow: Old vs New



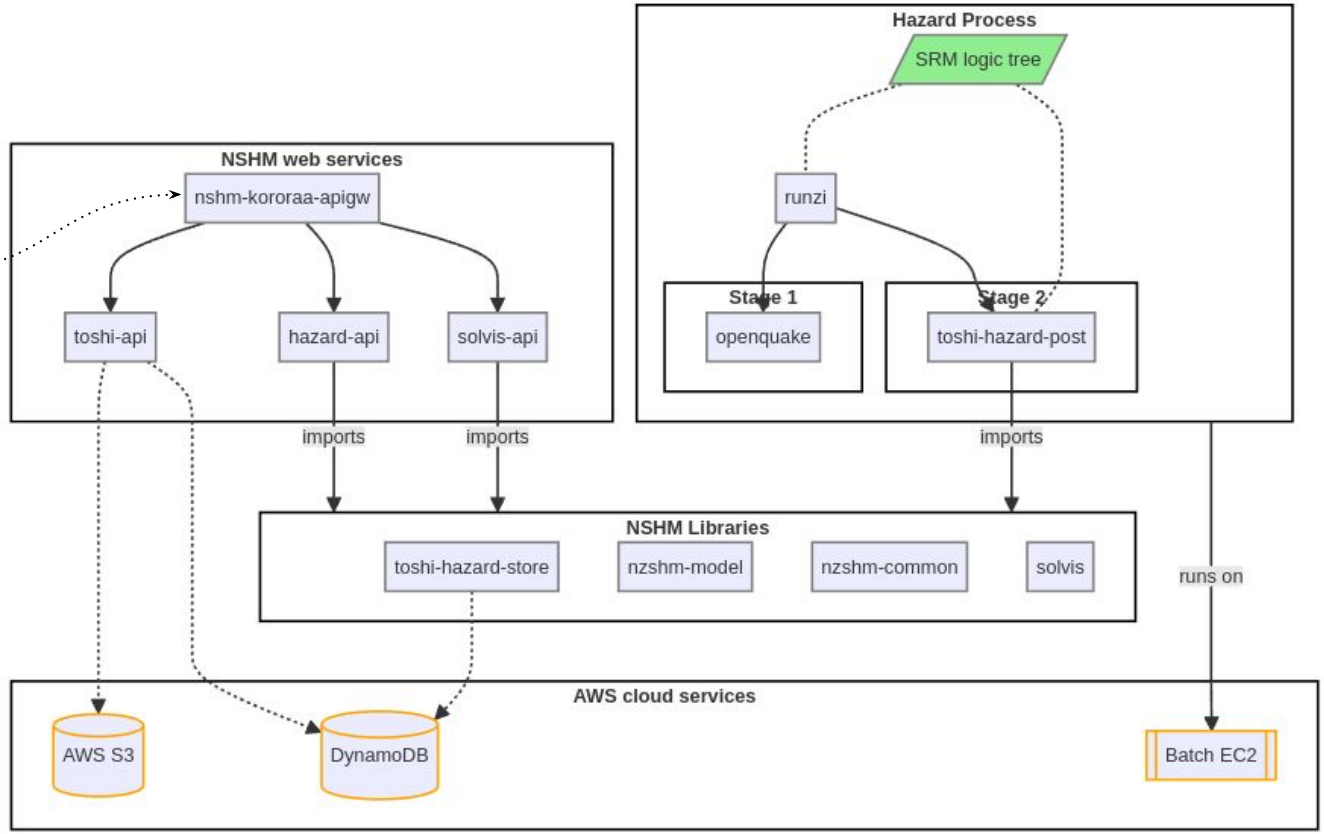
Calculating and consuming Hazard Curves



Kororā Web Application

toshi-hazard-post

pyarrow



NSHM 2022 = the PSHA branch realisations rodeo

With:

- 3991 sites,
- 20 vs30 values,
- 27 Intensity Measure Types (IMTS) PGA, SA(0.5) etc
- ~900 source/gsim model permutations

=> 1,965,487,680 individual realisation curves, of 44 points each.

In DynamoDB, approx 2 Terabytes and 200 million objects

In Parquet, approx 0.6 Terabyte and 2 billion rows

Database Types

- **DynamoDB** *still used for website*
- **Sqlite3** *an available anywhere sql DB,*
- **introducing Arrow** *https://arrow.apache.org/docs/python/install.html*

Small demo:

- **reading the NSHM dataset just using pyarrow/S3.**
- **Tables, filtering, to_pandas(), shape**
- **Exploring the data**
- **Identifying branches**

The NSHM realisation dataset schema

```
schema = pa.schema(
    [
        ("compatible_calc_fk", dict_type),   # id for calculation equivalence, for PSHA engine interoperability
        ("calculation_id", dict_type),       # a reference to the original calculation that produced this item
        ("nloc_001", dict_type),             # the location string to three places e.g. "-38.330~17.550"
        ("nloc_0", dict_type),               # the location string to zero places e.g.  "-38.0~17.0")
        ('imt', dict_type),                  # the IMT label e.g. 'PGA', 'SA(5.0)'
        ('vs30', vs30_type),                 # the vs30 value e.g 400
        ('rlz', dict_type),                  # the rlz id from the the original calculation eg “rlz-001”
        ('sources_digest', dict_type),       # a unique hash id for the NSHM LTB source branch
        ('gmms_digest', dict_type),          # a unique hash id for the NSHM LTB gsim branch
        ("values", values_type),             # a list of the 44 IMTL values
    ]
)
```

Introducing nzshm-component-rlz-dataset (from June 2024)

A collection of standardised PSHA realizations to facilitate research and model development, allows for:

- create new models from existing branch realisations (recombine, reweight)
- add new realisations, compare, merge into new models
- make detail comparisons to NSHM_v1.0.4 baseline.

Accessible:

- public access with option to download dataset OR query directly against the cloud store (S3 bucket)
- uses lightweight and simple python libraries
- Productive using only workstation compute - no server-side resources

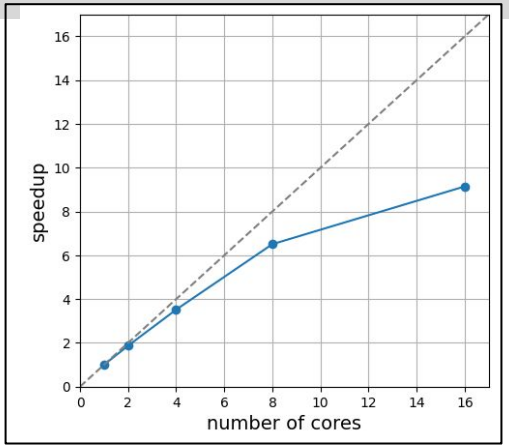
toshi-hazard-post

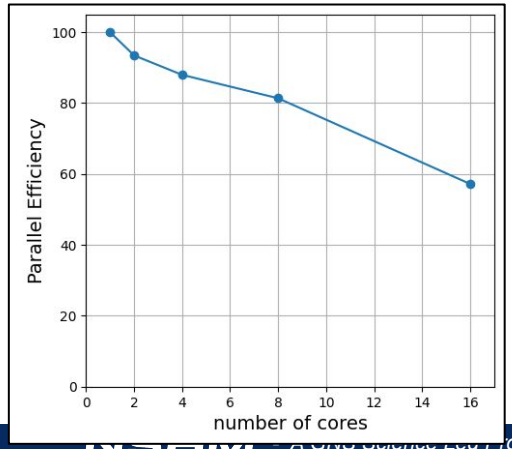
Simplification: 3,485 down to 1,077 lines of code

- Supporting libraries nzhsm-model and nzshm-common
- Removal of toshiAPI dependency

Performance Improvement: 3x faster

- Makes better use of numpy vectorization
- Simplification of logic





36 Core, 64GB workstation

Possible shell demos of libraries?

- common
- model
- solvis

ALL DONE!

Demo

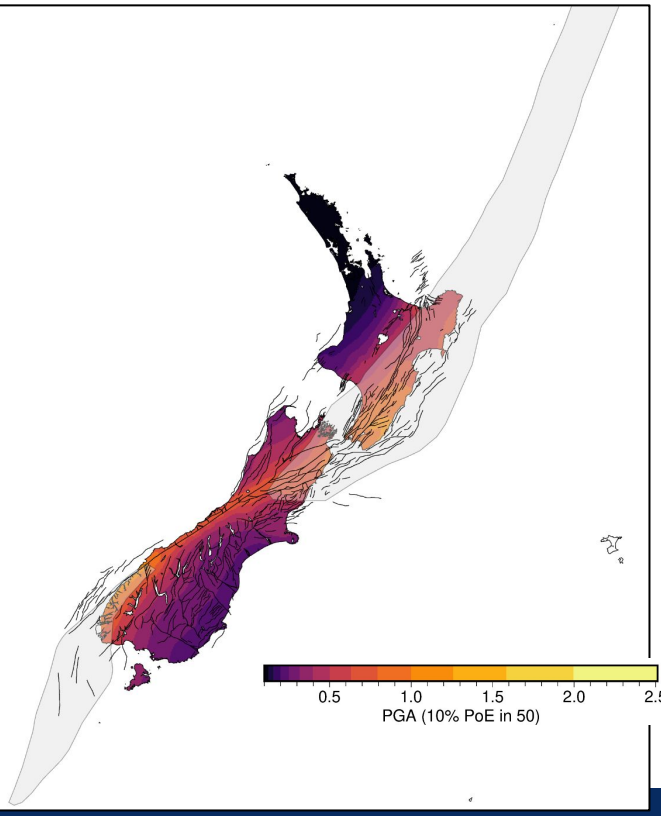
1. **Calculate NSHM_v1.0.4 for 2 locations and 3 imts**
   a. Show how you specify the model by its name
   b. Run in parallel
2. **Export the logic trees of NSHM_v1.0.4 for editing**
   a. Show that the logic tree structure is included in the file (for SRM)
   b. Choose highest weighted branch from all

```
>>> from nzshm_model import get_model_version
>>> model = get_model_version('NSHM_v1.0.4')
>>> model.source_logic_tree.to_json('/home/chrisdc/NSHM/thp-demo/srm-NSHM_v1.0.4
>>> model.gmm_logic_tree.to_json('/home/chrisdc/NSHM/thp-demo/gmcm-NSHM_v1.0.4.j
```

```
>>> from nzshm_model import get_model_version
>>> model = get_model_version('NSHM_v1.0.4')
>>> model.source_logic_tree.to_json('/home/chrisdc/NSHM/thp-demo/srm-NSHM_v1.0.
>>> model.gmm_logic_tree.to_json('/home/chrisdc/NSHM/thp-demo/gmcm-NSHM_v1.0.4.
```

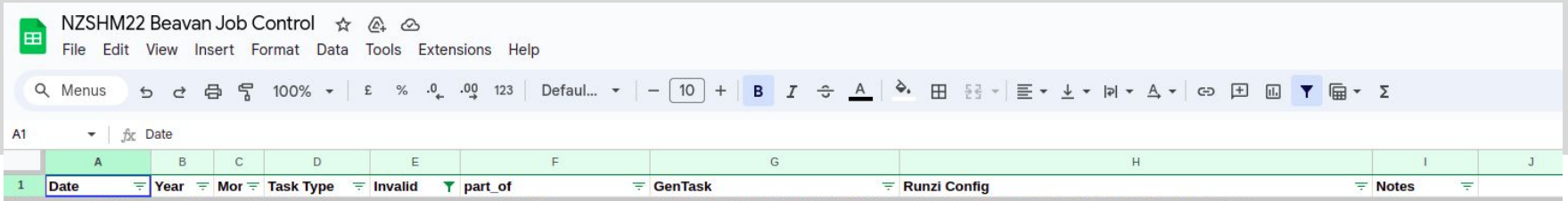
Hazard modelling at scale

The NZ NSHM 2022 uses a large logic tree to better capture epistemic uncertainty ....



- 324 (SRM) x 3024 (GMCM) = 979,776 total branches / hazard realisations
- >1.1 Million sources per source branch
- 3774 sites (0.1 deg calculation grid , NZ cities, SWG)
- 27 spectral periods
- 20 site conditions (Vs30)

This cannot be run as a single openquake job due to memory and time constraints.



NZSHM22 Beavan Job Control

File Edit View Insert Format Data Tools Extensions Help

Menus 100% £ % .0 .00 123 Defaul... 10 B I S A

A1 Date

	A	B	C	D	E	F	G	H	I	J
1	Date	Year	Mor	Task Type	Invalid	part_of	GenTask	Runzi Config	Notes	
96	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzc2Mg==	vs30 = 350, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
97	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzc2Nw==	vs30 = 300, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
98	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzc2OA==	vs30 = 450, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
99	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzg0Mw==	vs30 = 200, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
100	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzkwMA==	vs30 = 900, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
101	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzkwMQ==	vs30 = 500, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
102	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk1OQ==	vs30 = 1000, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
103	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk2Nw==	vs30 = 1500, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
105	9/6/2023	2023	6	OQ Hazard	0	NSHM_v1.0.4_mcverry	R2VuZXJhbFRhc2s6NjUzMDM5Nw==	McVerry vs30=400 oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
106	9/6/2023	2023	6	OQ Hazard	0	NSHM_v1.0.4_mcverry	R2VuZXJhbFRhc2s6NjUzMDM5OA==	McVerry vs30=250 oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
107	21/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzNjgwNg==	vs30 = 250, transpower critical sites		
109	21/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzNzI0OQ==	vs30 = 150, transpower critical sites, orig IMTs only		
110	22/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzNzM5Nw==	vs30 = 175, transpower critical sites, orig IMTs only		
111	22/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzNzU0NQ==	vs30 = 200, transpower critical sites, orig IMTs only		
112	22/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzNzY5Mw==	vs30 = 225, transpower critical sites, orig IMTs only		
113	22/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzNzg0MQ==	vs30 = 275, transpower critical sites, orig IMTs only		
114	22/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzNzk4OQ==	vs30 = 300, transpower critical sites, orig IMTs only		
115	22/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzODEzNw==	vs30 = 350, transpower critical sites, orig IMTs only		
116	23/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzODI4NQ==	vs30 = 375, transpower critical sites, orig IMTs only		
117	23/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzODQzMw==	vs30 = 400, transpower critical sites, orig IMTs only		
118	23/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzODU4MQ==	vs30 = 450, transpower critical sites, orig IMTs only		
119	23/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzODcyOQ==	vs30 = 500, transpower critical sites, orig IMTs only		
120	24/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzODg3Nw==	vs30 = 600, transpower critical sites, orig IMTs only		
121	24/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzOTAyNQ==	vs30 = 750, transpower critical sites, orig IMTs only		
122	24/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzOTE3Mw==	vs30 = 900, transpower critical sites, orig IMTs only		
123	24/08/2023	2023	8	OQ Hazard	0	TEST_AGAINST_OQ_V2	R2VuZXJhbFRhc2s6NjUzOTMyMQ==	Test Against OQ v2	verifying OQ vs THP aggregat	
124	24/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzOTM0MA==	vs30 = 1000, transpower critical sites, orig IMTs only		
125	24/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzOTQzOQ==	vs30 = 1500, transpower critical sites, orig IMTs only		
126	3/10/2023	2023	10	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjUzOTkzMA==	vs30 = 400, crustal only, fix min mag 6.8, NZ sites, orig IMTs		
127	19/10/2023	2023	10	OQ Hazard	0	NSHM_v1.0.4_IFMonly	R2VuZXJhbFRhc2s6NjU0MDAzOQ==	vs30 = 400, test just IFM with NZ sites		
128	19/10/2023	2023	10	OQ Hazard	0	NSHM_v1.0.4_IFMonly	R2VuZXJhbFRhc2s6NjU0MDE4NA==	vs30 = 400, test just IFM with NZ, NZ_0_1_NB_1_1		
129	3/11/2023	2023	11	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6NjU0MDQ2Nw==	vs30 = 1000, cave sites for Jeff Lang		
160	18/3/2024						R2VuZXJhbFRhc2s6NjgwMjUyOA==	PGA for geonet sites (Emily WS) vs30=250		
161	18/3/2024						R2VuZXJhbFRhc2s6NjgwMjYyNw==	vs30=400		

Old Hazard Workflow

SRM LT

GMCM LT

---

runzi

toshAPI

toshiID

toshiID
- Vs30
- imt
- OQ version
- GMCM
- SRM

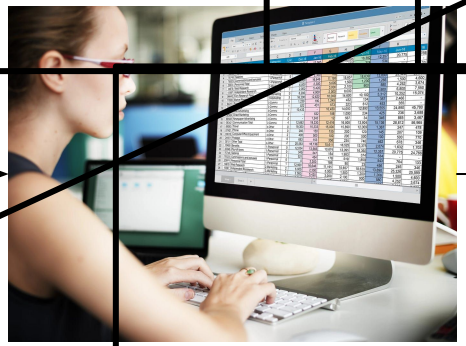
openquake

toshi-hazard-post

toshi-hazard-store

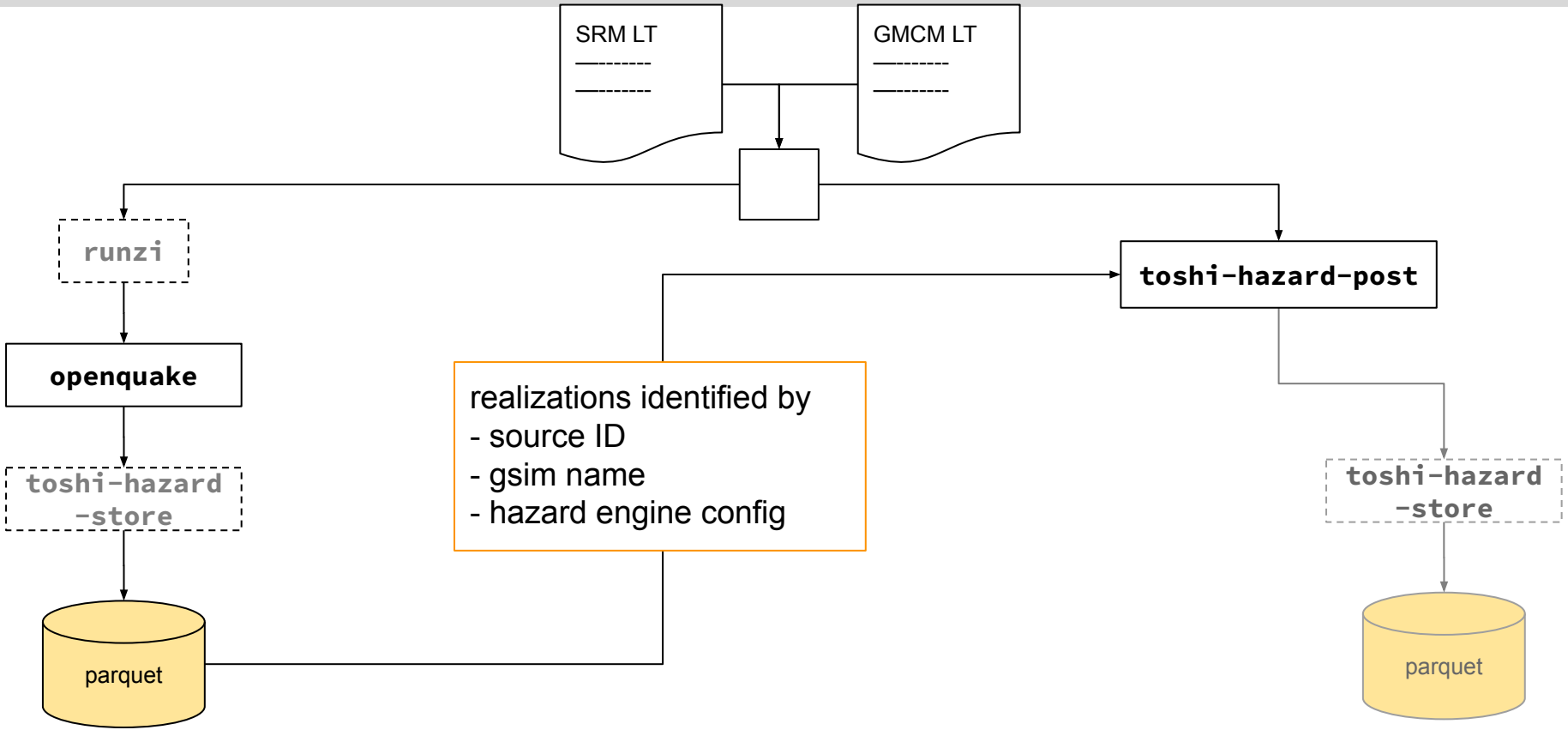
toshAPI

toshi-hazard-store



DynamoDB

New Hazard Workflow



NSHM python libraries for research

Library	Hazard pipeline	NSHM user website	Researcher
nzshm-model	Yes	Yes	identify, filter and build branches and logic trees. Build hazard engine inputs and configs.
nzshm-common	Yes	Yes	handle coded locations, bins and grids.
solvis	Yes	Yes	geospatial investigation of NSHM inversion sources and rates (both individual inversions and the composite model)
toshi-hazard-post	Yes	Not yet	Building all realizations and calculating aggregate statistics (hazard curves or disaggregations)
toshi-hazard-store	Yes	Yes	not needed if using parquet, but maybe useful in other cases.

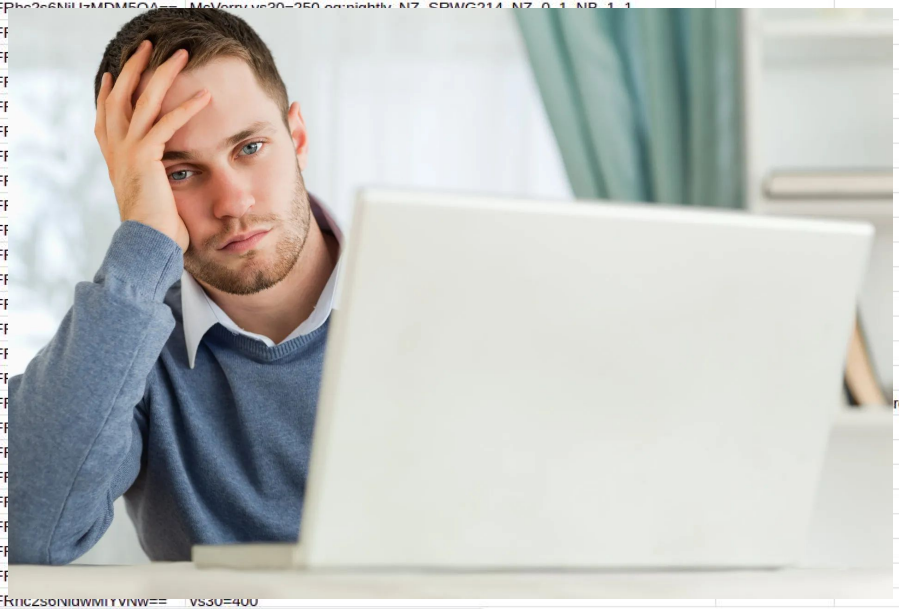
NZSHM22 Beavan Job Control

File Edit View Insert Format Data Tools Extensions Help

Menus 100% £ % .0 .00 123 Defaul... 10 B I

A1 Date

	A	B	C	D	E	F	G	H	I	J
1	Date	Year	Mor	Task Type	Invalid	part_of	GenTask	Runzi Config	Notes	
96	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzc2Mg==	vs30 = 350, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
97	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzc2Nw==	vs30 = 300, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
98	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzc2OA==	vs30 = 450, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
99	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzg0Mw==	vs30 = 200, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
100	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzkwMA==	vs30 = 900, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
101	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzkwMQ==	vs30 = 500, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
102	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk1OQ==	vs30 = 1000, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
103	18/4/2023	2023	4	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk2Nw==	vs30 = 1500, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
105	9/6/2023	2023	6	OQ Hazard	0	NSHM_v1.0.4_mcverry	R2VuZXJhbFRhc2s6NjUzMDM5Nw==	McVerry vs30=400 oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
106	9/6/2023	2023	6	OQ Hazard	0	NSHM_v1.0.4_mcverry	R2VuZXJhbFRhc2s6NjUzMDM5OA==	McVerry vs30=250 oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
107	21/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk3Mg==	vs30 = 600, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
109	21/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk3Nw==	vs30 = 700, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
110	22/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk4MA==	vs30 = 800, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
111	22/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk4OA==	vs30 = 1100, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
112	22/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk5MA==	vs30 = 1200, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
113	22/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk5OA==	vs30 = 1300, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
114	22/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk5OQ==	vs30 = 1400, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
115	22/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk5Mg==	vs30 = 1500, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
116	23/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk5Mw==	vs30 = 1600, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
117	23/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk5NA==	vs30 = 1700, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
118	23/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk5NQ==	vs30 = 1800, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
119	23/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk5Nw==	vs30 = 2000, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
120	24/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk5OA==	vs30 = 2200, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
121	24/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyMzk5OQ==	vs30 = 2400, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
122	24/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyNDAw	vs30 = 2600, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
123	24/08/2023	2023	8	OQ Hazard	0	TEST_AGAINST_OQ_V2	R2VuZXJhbFRhc2s6MjkyNDAx	vs30 = 2800, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
124	24/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyNDAy	vs30 = 3000, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
125	24/8/2023	2023	8	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyNDAz	vs30 = 3300, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
126	3/10/2023	2023	10	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyNDA0	vs30 = 3700, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
127	19/10/2023	2023	10	OQ Hazard	0	NSHM_v1.0.4_IFMonly	R2VuZXJhbFRhc2s6MjkyNDA1	vs30 = 4100, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
128	19/10/2023	2023	10	OQ Hazard	0	NSHM_v1.0.4_IFMonly	R2VuZXJhbFRhc2s6MjkyNDA2	vs30 = 4600, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
129	3/11/2023	2023	11	OQ Hazard	0	NSHM_v1.0.4	R2VuZXJhbFRhc2s6MjkyNDA3	vs30 = 5100, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
160	18/3/2024						R2VuZXJhbFRhc2s6MjkyNDA4	vs30 = 5700, oq:nightly, NZ, SRWG214, NZ_0_1_NB_1_1		
161	18/3/2024						R2VuZXJhbFRhc2s6NTUwMTY1Nw==	vs30=400		



regat

ramme