ミーティングメモ

2013-09-04

辰巳大輔

新学術公募研究「低温干渉計型重力波検出器における突発性雑音低減」の一環として
関連するシステム KAGRA- DMG と DetChar について
現状と問題点を神田、端山に 2013-09-04 に聞いたので
そこで挙がった To be decided な項目をリストする。

話の前提
o 新学術領域 A04 班の目指す
  GW alert system + External trigger GW search
  に関して
o 新学術公募研究がどう寄与し、どこを研究・開発すべきかを明らかにする
ことを目的とする。

国立天文台
辰巳大輔

目次

(1) データフローに関する現状案

1.1 Low-latency GW search のためのデータフロー

神田さんから説明されたものをまとめた。

1.2 DetChar のデータフロー

端山くん、神田さんから提示された案をまとめた。

(2) LIGO S5 の現状と S6 への改善案

S6 からだいぶん時間も経つので、現状 or aLiGO 時点でのシステムと乖離している可能性大。

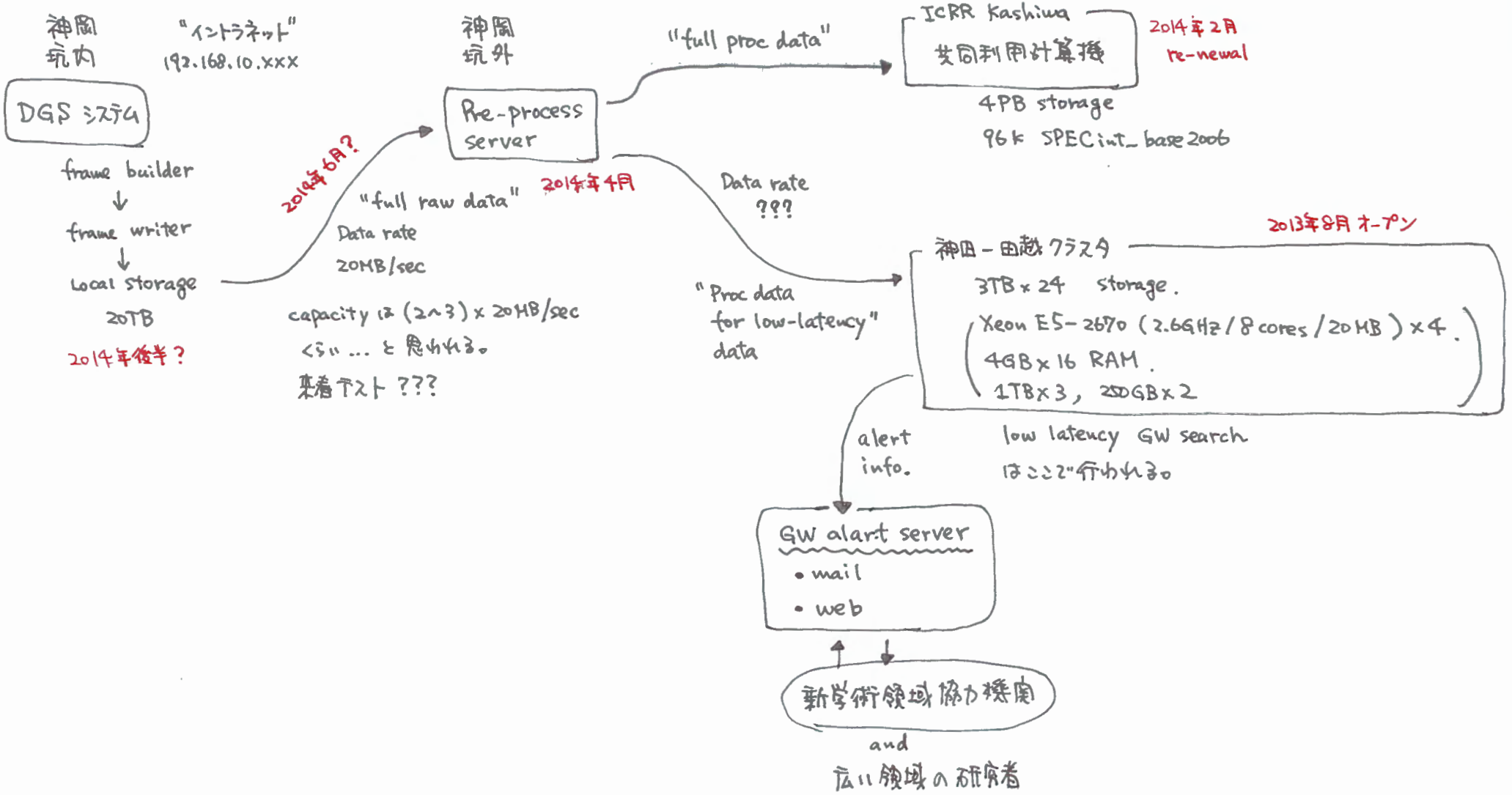
(3) Virgo: 2010 年頃の話

(4) KAGRA における今後の課題

2013-09-04

Low latency analysis system

神田さんによる現状説明



Detector Characterization System
(DetChar)

(現状案の説明を元に長もがまとめた)

2013-09-04

(神田案)

神岡坑内

(frame builder??)

各DGSマシン | ADC | DAC

各サブグループ(各DGSマシン)毎にDet Charプログラムが稼動しており、veto flag (Data Quality Flag)を出力する。

raw data +α

frame writer

最低限、全て??のraw dataがmargeされて"full raw data"が作られる。

★上記プラス各サブグループが出力するveto flag (Data Quality flag)が含まれる...?

flagが出力されるまで何秒かかる?

1分以内、10秒以内?

★中間生成データ(spectrum, cross colleration など)も上記に含むか?

要ニデータ量試算!!! ·····> 下記raw data量から考えてnegligible.

"full raw data" + "veto flag"

Pre-process server

例えば、channel間相関解析(他のサブシステムとの)などは各DGSマシンでは出来ないので pre-process serverで実行することが考えられる。

環境モニターのraw data量は

100Hz sampling × 200 channels × 2 bytes (16 bit) = 40 kB/s

16kHz sampling dataが本当にないのか?要確認

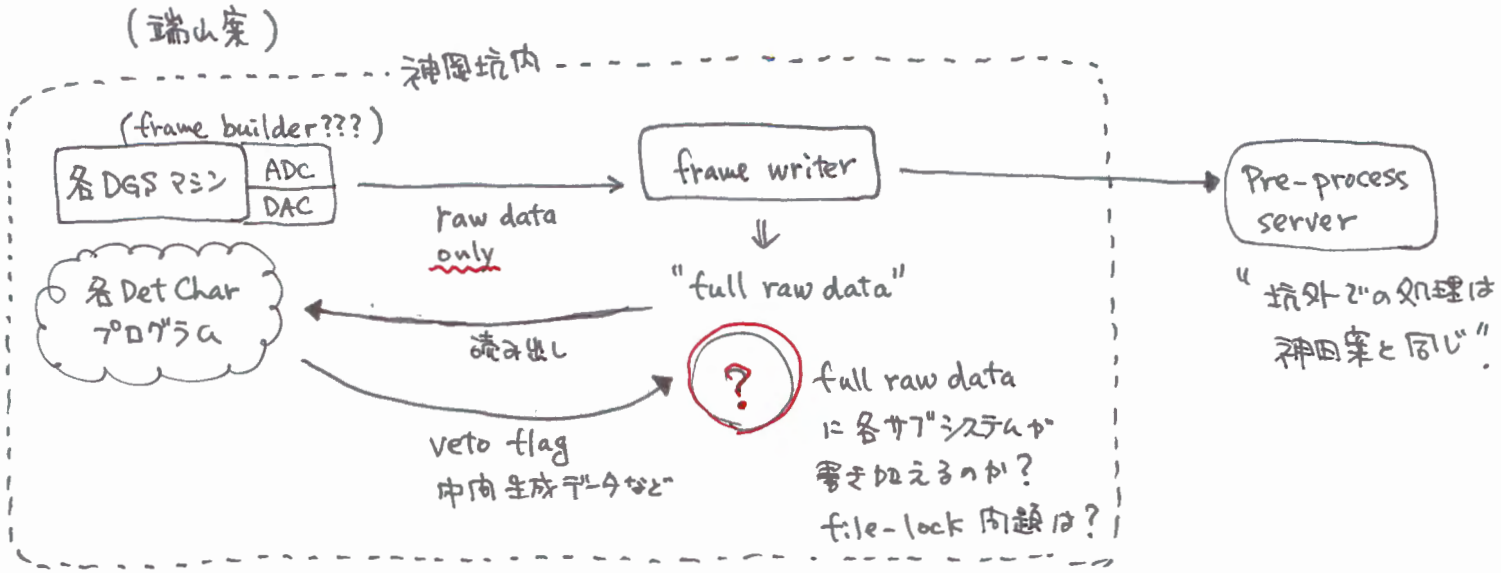
1チャンネルあると + 32 kB/s

神田案の問題点

frame builderが"raw data+α"のファイルを出力するタイミングはバラバラ。だれがこの"raw data+α"を統合する仕組みを作るのか?

Det Char system

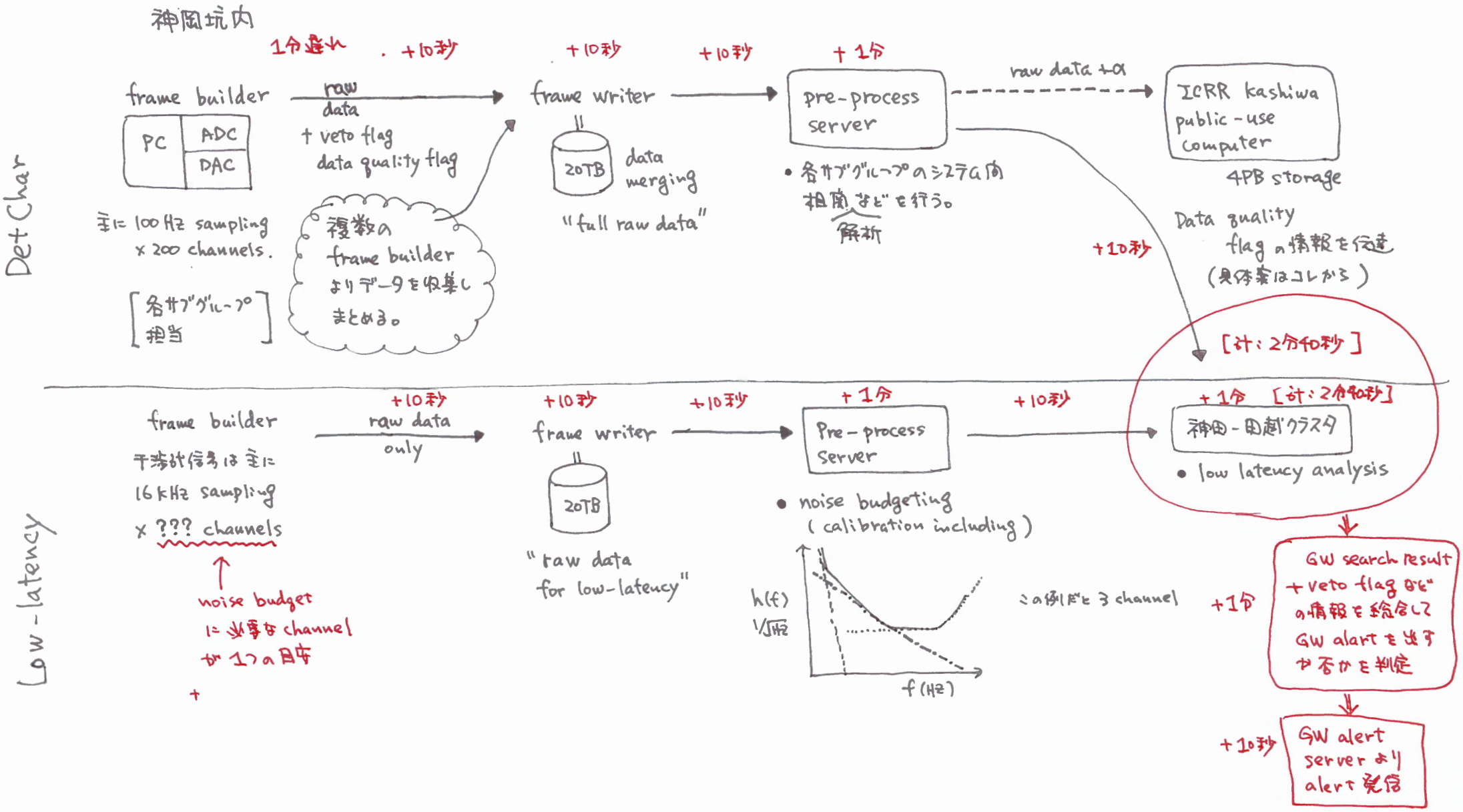
2013-09-04



私見:
う〜ん。この案は破綻してる気がする

以上を元にまとめると.....

2013-09-04



神田さんの言ってることが腑に落ちないので
調べました。添付 Excel file 参照

LIGO-G080538-00-Z によると
1) S5 の段階で、pipeline search 開始まで約 20 分掛かっていた。
2) S6 に向けて、10.5 分まで短縮する予定。
3) 行程 10-17 の辺りでのさらなる改良も模索していて、
   8 分まで短縮できるかも。(2.5 分の短縮)

(A) LDR (Data Grid を使った Data Replication) の役割
~~~~~~

基本的には GW international network search (Coherent search)
のために必要と認識したほうが良さそうである。
ただ、S6 後に上記 3) の low-latency 対応 LDR が開発されたので
それを使えということではないかと思う。
S5: 6m40s --> S6: 4m50s --> now?: 2m20s

(B)   彼らが S6 に向けて改善した点
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

ファイルをディスクに書き終わってから、次の作業開始まで
２分掛けていたのを、きちんと書き終わりフラッグを次の作業へ
引き継ぐ仕組みを作っただけ。

自分でプログラムが掛ければ大した技術ではないが、
詳細仕様を知らないと DGS との間の情報受け渡しが出来ない。
DetChar はどうやって新しいファイルの書き込み終了タイミングを
知っているのか？ 何秒くらいの遅延があるのか？(精度 10 秒程度で)

(C) データ長
~~~~~~~~~~~~~~

S5 では、RAW data は 32 sec, RDS data (proc data) は 128 sec
端山情報によれば、S6 の RDS data は 128 sec

以上

> o 具体的なデータ転送の仕組み
>   公募研究ではどこから情報を得て、どこに情報を出力すべきか。
>
>   ** GWF files (Frame format data)
>   (a) 坑内 --> 坑外 (pre-process server)
>   (b) pre-process server --> Kanda-Tagoshi cluster
>   ** XML files
>   ** Databasee related
>
>   FDR とは (LIGO Data Replicator とは別物？)
~~~~~~~~~~~~~~

"LIGO Data Replication Status and Plans for S6"

LIGO-G080538-00-Z

			As of S5			Plan for S6			
		Process	wasted time	acuumlated		wasted time	acuumlated		comments
	1	**frame builder write a file in /frames directory**	00:32	00:32		00:32	00:32		
KAGRA DGS		**CDC reads back, validates**	00:10	00:42		00:10	00:42		
担当部？	2	**Diskcache detects frames**	02:00	02:42		00:00	00:42		skipped
		records metadata & path		02:42	**02:42**		00:42	**00:42**	
	3	**Disk2Disk copies to /archive directory**	01:00	03:42		00:00	00:42		skipped
	4	**Diskcache detects new location on /archive**	02:00	05:42		00:10	00:52		on /frames
	5	**LDAS queries Diskcache**	01:00	06:42		01:00	01:52		
		run FrCheck	00:10	06:52		00:10	02:02		
KAGRA		**md5sum**	00:10	07:02		00:10	02:12		
Pre-Process Server	**6**	**Every 4 frames are merged into a file**	02:08	09:10		02:08	04:20		
担当部		**LDAS computes h(t) and RDS 1,3 data**		09:10			04:20		
	7	**Diskcache detects new h(t) and RDS frames**	02:00	11:10		00:10	04:30		
	8	**Pub scruipts query Diskcache**	01:00	12:10		00:10	04:40		
		Publish to local LDR		11:10			04:40		
	9	**Local RLS updates remote RLS**	01:00	13:10	**09:28**	01:00	05:40	**04:58**	
	10	**Remote LDR pulls new metadata**		13:10			05:40		
	11	**Remote LDR schedules transfer**	01:00	14:10		01:00	06:40		
KAGRA では未だ見通しの	**12**	**Remote LDR pulls file**	01:00	15:10		01:00	07:40		
立たない	**13**	**Remote LDR computes checksums**	00:10	15:20		00:10	07:50		
(Data Grid Server)	**14**	**Diskcache detects transferred data**	02:00	17:20		00:10	08:00		
担当部	**15**	**Diskcahe hash dumped to ASCII**	01:00	18:20		01:00	09:00		
	16	**Dumps mirrored to different nodes**	00:30	18:50		00:30	09:30		
	17	**Dumps loaded into "LDRdataFindServer"**	01:00	**19:50**	**06:40**	01:00	**10:30**	**04:50**	
		--- Analysis pipelines start ---							
ここからが									
田越ー神田クラスタ									
担当部									

A robust and GRID compliant system for Virgo data transfer

https://workarea.ego-gw.it/ego2/virgo/data-analysis/vdas/TransferFramework_VDASMeeting101110.pdf

(1) Data transfer tools

as of 2010

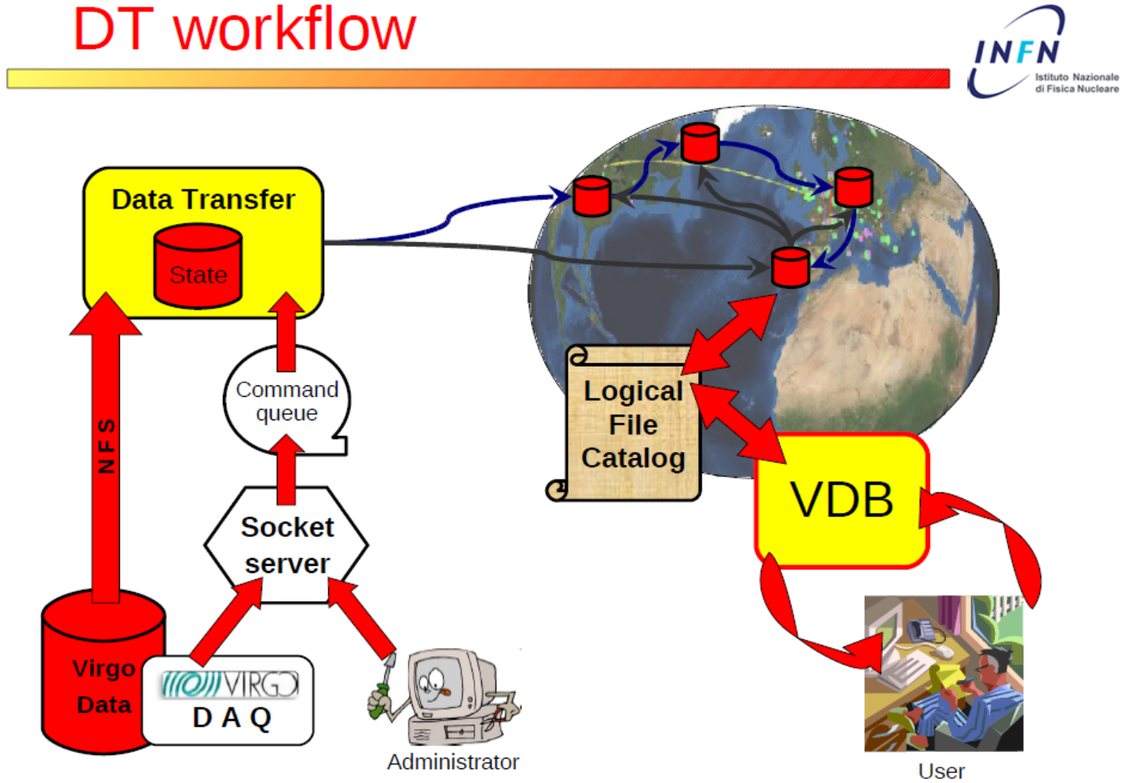
TOOL name		Developer			
(LIGO Data Replicator	LDR **GridFTP**	LIGO **Globus**		U.S.)	
Storage Resource Broker	**SRB** **bbftp**	IN2P3 CNAF	Bologna	France Italy	RFC1323 implementation
~~File Transfer Service~~	~~**FTS**~~				
LCG client tool	**lcg-cp**	**open science grid**			

(2) Metadata server

as of 2010

Virgo metadata catalogue	**VDB**	
Logical File Catalogue	**LFC**	

DT workflow



Virgo Data Management at CNAF Bologna
https://workarea.ego-gw.it/ego2/virgo/data-analysis/vdas/virgo-data-management-at-cnaf-bologna

Virgo Data Analysis Software (VDAS) wiki
https://workarea.ego-gw.it/ego2/virgo/data-analysis/vdas/vdas-main/

Data transfer with Grid tools: Basic User Guide
https://workarea.ego-gw.it/ego2/virgo/data-analysis/vdas/data-transfer-with-grid-tools-basic-user-guide

(4) KAGRA における今後の課題

[[[DAS]]]
o GW alert を出す時に重要な天体源位置情報を得るためには
  GW international network を用いた解析が必須。
  この解析に如何に、公募研究が寄与出来るか。

o CBC search (for low latency analysis) で必要なデータ長。
  (**A の付帯項目)

o Burst search はデータ長に制限があるのか？ないと思うが。
  (確認事項)

[[[DMG]]]
o GW alert に関わる部分についての所要時間は、
  Data Transfer 10sec
  Data process (analysis) 1min
  を大凡の目安として、各システム・プログラムの仕様策定
  をしたらと考えていることが神田さんから説明された。
  DGS, DMG, DAS, DetChar で広く共通認識を作ることが肝心。

o GWF (frame format data) のファイル名：命名規則
  地味だけど、コードを書くには重要。
  LIGO 式には [S]-[data description part]-[G]-[T].gwf と決まっていて
  [S] = K ... for KAGRA は勝手に決めて良い？まぁ O.K. だろう。
  [G] = (GPS start time in integer second)
  [T] = (Data length of the file)
  ここまでは議論の必要なし。残り部分を要確認＋議論。

o データ長 (for low-latency analysis) の決定 (**A)
  データ長の関数として、転送時間を調べること。
  各 GW search で必要とする要件を確認。

o Data Quality Flag の定義
  FrameL の中で定義されているのは、1 frame 毎のフラッグ。
  しかも GW detector 間の情報共有を目的に各検出器に bit 割り当て
  されているので、多くの情報が載せられない。

  metadata server や、FrameL 内に KAGRA 独自のデータ定義を作るなど
  やり方は幾つか考えられるが、公募研究に取ってこれはシステムの
  大事な output なので早急に決める必要あり。

o 具体的なデータ転送の仕組み
  公募研究ではどこから情報を得て、どこに情報を出力すべきか。

  ** GWF files (Frame format data)
  (a) 坑内 --> 坑外 (pre-process server)
  (b) pre-process server --> Kanda-Tagoshi cluster
  ** XML files
  ** Databasee related

  FDR とは (LIGO Data Replicator とは別物？)

[DetChar]
o DGS, DetChar (+DMG) 間で、Data Flow に関して認識の確認。

o GW alert system にどういう方法とデータ形式で情報を渡すか。
  Data Quality Flag の定義と密接に関連する議題。

o Low-latency analysis の結果を、DetChar (各サブグループ) に
  どうやってフィードバックするか。
  例えば、Cryogenic system が重力波探査に多大な影響を与える
  雑音を出していた場合、サブグループのメンバーに情報をフィードバック
  することで検出器が改善されることが期待できる。

以上

大学共同利用機関法人　自然科学研究機構
国立天文台　光赤外研究部　重力波プロジェクト推進室
助教　辰巳大輔

---- Daisuke Tatsumi ----

National Institutes of Natural Sciences
National Astronomical Observatory of Japan

< Address > 2-21-1 Osawa, Mitaka, Tokyo 181-8588

< Tel > +81-422-34-3990 (TAMA300 site)
                -3620 (Office)
< FAX > +81-422-34-3793