



2017 ICT Workshop

Giga Tera Peta Dati: Ma siamo pronti?

C.Baffa
INAF

Un sasso nello stagno



Il diluvio prossimo venturo



Già nel 2001 Brunner, Djorgovski, Prince, Szalay ponevano il problema in “Massive Dataset in Astronomy”

Gaia (2014-2019) avrà un catalogo di 1-2 PB

Lofar produrrà immagini tra 0.2 e 1TB ciascuna

SKA PSS avrà un flusso dati di circa 1TB/sec

<e la lista continua...>

Dimensioni immaneggiabili



Ma il problema è già tra noi!

5 ore di osservazioni di Alma hanno prodotto
300GB di dati

Un progetto Muse ha prodotto 90000 spettri: un
cluster apposito per la riduzione

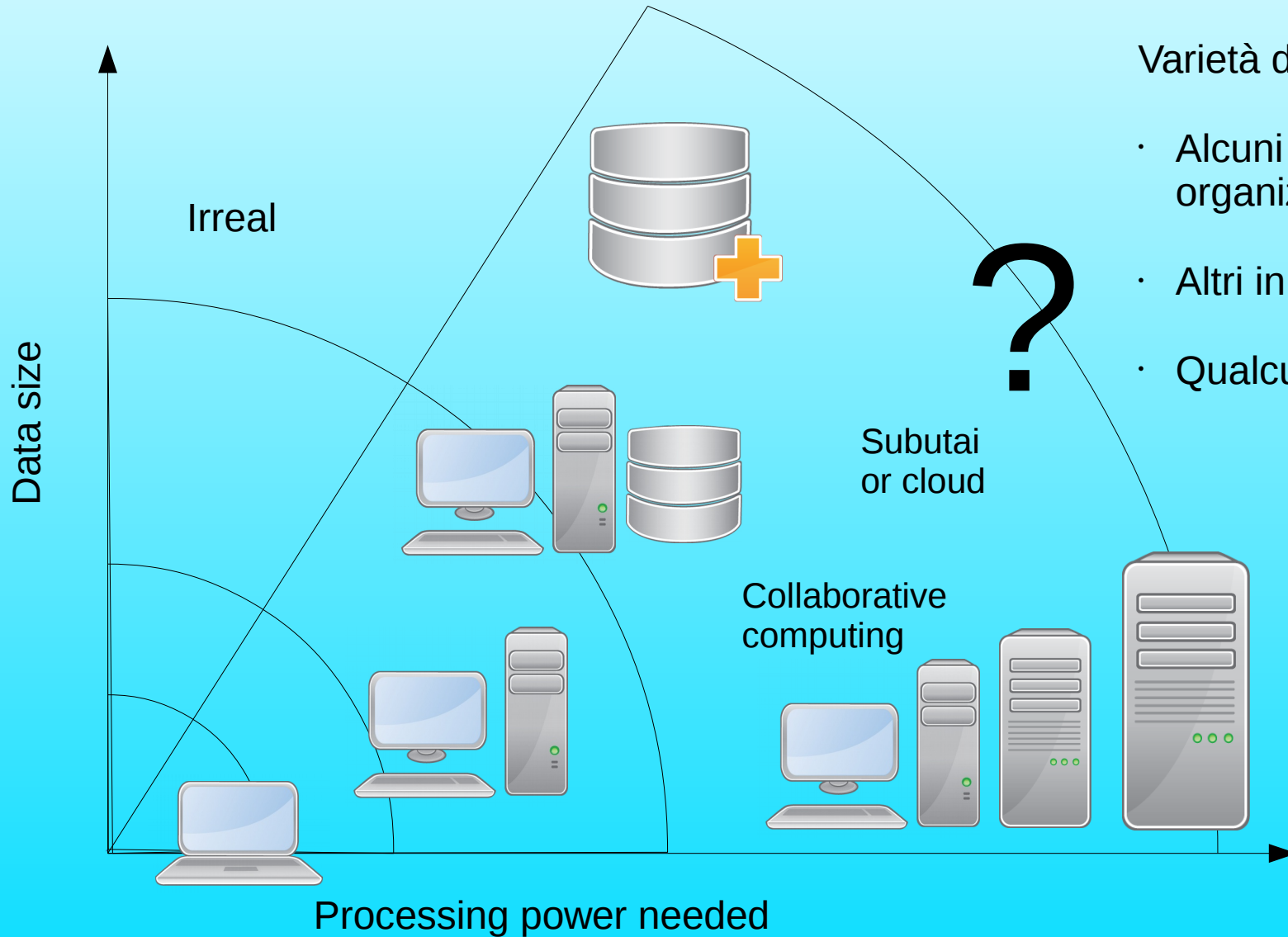
Speciali spettri di Alma richiedono enormi quantità
di Ram per essere ridotti

Il confronto tra le osservazioni e più di 10^6 modelli
richiede super-calcolo parallelo

Alcuni tipi di limiti

- Problemi di immagazzinamento (10TB OK, 100TB ?)
- Problemi di potenza di calcolo (NVIDIA TitanX 10TFlops single, 0.25TFlops double. Reali <10%)
- Problemi di memoria centrale: ora già casi da > 64GB RAM.
- Problemi di rete: accesso a database remoti nel range 1-50TB. In futuro > 1PB

Ranges e soluzioni?



Varietà di situazioni:

- Alcuni gruppi ben organizzati
- Altri in affanno
- Qualcuno in difficoltà

Centralità della rete

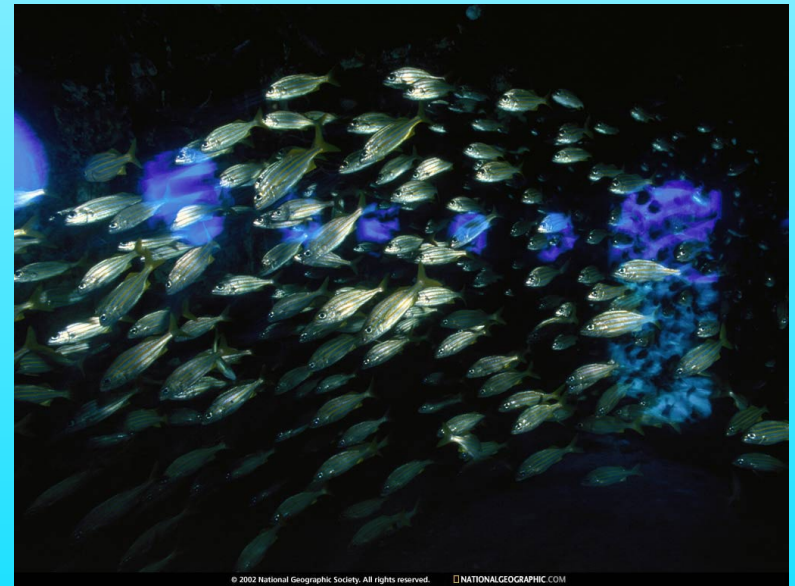


Possibili approcci

Big Data Center



Versus



Small and diffuse 9

Un problema culturale diffuso

Differenze culturali tra aree di ricerca

Differenze di richieste

Molteplicità di approcci:

Una *Taglia Unica* non funziona.

→ **Necessità di expertise diffusa** ←