

**I HC QUC GIA THNH PH H CHỜ MINH
TRNG I HC BỒCH KHOA
KHOA KHOA HC VỊ K THUT MÒY TỜNH**



**BỒO CỒ
ỒN CHUYỒN NGỊNH**

**XỒY DNG C S D LIU TỒCH HP
CHO CỒC H THNG D BỒO NGN HN
TRONG KHỒ TNG THU VN**

Ngịnh: Khoa hc mỳ tờnh

HI NG: Hi ng 5

GVHD1: Lổ Hng Trang

GVHD2: Trng Qunh Chi

TKH: Lổ Th Bo Thu

—o0o—

SVTH1: Trn Hị Tun Kit (2011493)

SVTH2: Nguyn c Thu (2012158)

Thnh ph H Chờ Minh, Thờng 12/2023



Lời cảm ơn

Chúng tôi cảm ơn ông Nguyễn Lê Quốc Anh cũng với những người khác, đặc biệt là những người đã tận tâm PGS. TS. Lê Hồng Trang tại Trường Đại học Bách Khoa - Đại học Quốc gia TP.HCM. Chúng tôi xin cảm ơn ông và các thành viên trong nhóm đã ủng hộ cũng như vị trí của mình trong công trình này.

Chúng tôi cảm ơn ông và những người khác đã tận tâm và tận tâm cũng với những người khác, đặc biệt là những người đã tận tâm PGS. TS. Lê Hồng Trang tại Trường Đại học Bách Khoa - Đại học Quốc gia TP.HCM.

Chúng tôi xin cảm ơn ông và các thành viên trong nhóm đã tận tâm và tận tâm cũng với những người khác, đặc biệt là những người đã tận tâm PGS. TS. Lê Hồng Trang tại Trường Đại học Bách Khoa - Đại học Quốc gia TP.HCM.

Chúng tôi xin cảm ơn ông và các thành viên trong nhóm đã tận tâm và tận tâm cũng với những người khác, đặc biệt là những người đã tận tâm PGS. TS. Lê Hồng Trang tại Trường Đại học Bách Khoa - Đại học Quốc gia TP.HCM.

Thị trấn Hồ Chí Minh, Ngày 28 tháng 12 năm 2023

NGUYỄN THỊ HẠ

Trần Thị Tú Anh

Nguyễn Thị Thu



Chứng tử tin rơ òn nị lị mt bc tin quan trng trong s phòt trin ca chứng tử vị khững th t c mị khững cú s h tr vị úng gúp ca tt c mị nị.

Tóm tắt nội dung

Trong tài liệu này, chứng cứ giải thích một số tồn tại về hợp, ưu phi vị gièm sớt d lưu hồ tng thy vn ti Vit Nam. H thng c xut cú th lện nn tng cho vic xóy dng mt kho d lưu quc gia chuyển sốu v thũng tin hồ tng.

lện rử xut ca chứng cứ, chứng cứ ô xóy dng mt bn th nhim toện đin, trong ú chứng cứ ô the hìn mữ phng v quy trnh thu thp vị truyên phòt d lưu t trm thi tit ti Nhị Bó. ng thi, chứng cứ ô tp trung vạo quò trnh chuyển i vị t chc lu tr d lưu m bo s minh bch vị hiu sut ti u trong quò trnh x lý thũng tin.

Bị nghiổn cu nặ lị biu hìn ca nhng n lc òng k trong vic ti u hóa qun lý d lưu hồ tng ti Vit Nam. H thng xut nặ khũng ch giú quyit còc thòch the v tồch hp vị iu phi mị cùn t nn tng cho mt kho d lưu quc gia vng mn, phc v cho nhng nhu cu a dng ca nghiổn cu vị ng dng trong lnh vc hồ tng.

Bng phón cũng cũng vic

STT	H vị tổn	MSSV	Phón chia cũng vic
1	Trn Hị Tun Kit	2011493	<ul style="list-style-type: none"> - Phón tồch d liu - Xóy dng kin trức h thng - Nghiổn cu, tởm hiu cũng ngh vị òp dng - Xóy dng API
2	Nguyñ c Thu	2012158	<ul style="list-style-type: none"> - Phón tồch d liu - Xóy dng kin trức h thng - Nghiổn cu, tởm hiu cũng ngh vị òp dng - Xóy dng Data Flow

Mục lục

Danh sách hình vẽ

Danh sách bảng

Chương 1

Gii thiu

1.1 Phòt bii bii toàn

Trong bi cnh bin i khờ hu ngi cng đin bin phc tp, vic xóy dng c s d liu tồch hp cho còc h thng d bào ngn hn trong lnh vc khờ tng thu vn tr thnh mt thòch thc quan trng. Bii toàn t ra li lìm th nio chũng ta cú th ti u húa qun lý thũng tin thi tit, t nhieu ngun vị lu tr chũng trong mt c s d liu hii qu, nhm cung cp ngun thũng tin ng b vị cht lng cao h tr còc h thng d bào.

1.2 ng lc thc hin

i mt vi s phc tp ngi cng tng ca bin i khờ hu, còc mũ hnh d bào cng khũng ngng ùi hi tồnh chònh xòc ngi cng cao. Vic tồch hp thũng tin t nhieu ngun vị lu tr chũng trong mt c s d liu hii qu tr thnh chóa khúa quan trng òp ng mi thòch thc trong lnh vc ngy. Nhn thc c tm quan trng ca d bào thi tit vị thu vn trong vic m bo an toin vị phòt trin cng ng, chũng tũ xut d òn "Xóy dng c s d liu tồch hp cho còc h thng d bào ngn hn trong khờ tng thu vn".

1.3 Mc tiúu d òn

Mc tiúu ca d òn lĩ ngiỏn cu vị xóy dng mt c s d liu linh hot, cú kh nng tồch hp d liu, tng hp thũng tin t nhieu ngun còc trm ra-a vị quan trc khờ tng thu vn, ng thi lu tr chũng mt còch hii qu. D òn nhm to ra mt ngun thũng tin òng tin cy vị toin đin nóng cao chònh xòc ca mũ hnh d bào thi tit. Bng còch ngy, chũng ta cú th h tr hii qu trong vic d òn còc bin ng thi tit ngn hn. C s d liu s khũng ch linh hot mị cùn hii qu, cú kh nng tng hp thũng tin mt còch hii qu vị lu tr chũng mt còch cú t chc. Qua ú, d òn khũng ch mang li ngun thũng tin òng tin cy mị cùn úng gúp vò vic hii rũ hn v bin ng thi tit, c bit lĩ trong bi cnh bin i khờ hu ngi cng tr nỏn quan trng.

1.4 Phm vi d òn

D òn s tp trung vò Thnh ph H Chờ Minh, mt ù th ln vi mũ trng khờ hu c bit vị cú nh hng ln n cuc sng hng ngi ca cng ng.

Chũng tũ s ngiỏn cu vị thu thp d liu t nhieu trm ra-a vị trm quan trc khờ tng thu vn trong thnh ph m bo tồnh a dng vị i đin cho còc iu kin thi tit a phng. Sau khi ò thu thp d liu cn thit, chũng tũ s tin hnh thc hin còc bc tin x lý vị chun hoò d liu m bo chònh xòc vị tồnh nht quòn



ca tập dữ liệu. Cuối cùng, chứng tỏ sự xói mòn của dữ liệu tích hợp lưu trữ và quản lý thông tin từ các nguồn khác nhau, tạo ra một nguồn dữ liệu đồng nhất và đáng tin cậy. Nguồn thông tin tích hợp sẽ cung cấp dữ liệu đa chiều và chi tiết về nhu cầu thị trường và vị trí của doanh nghiệp. Mục tiêu là giúp công nghệ vị trí của nó vượt qua chướng ngại vật cho bất kỳ ai, bất kể họ đang thay đổi hay không.

Chúng tôi mong rằng việc sử dụng dữ liệu tích hợp sẽ giúp hiểu rõ hơn về các mô hình dữ liệu thị trường, giúp giải quyết các vấn đề kinh doanh và các vị trí thị trường khác nhau một cách thông minh và an toàn.

Chương 2

C s lý thuyết

2.1 H C s d liu

Database systems perform vital functions for all sorts of organizations because of the growing importance of using and managing data efficiently. A database system consists of a software, a database management system (DBMS) and one or several databases. DBMS is a set of programs that enables users to store, manage and access data. In other words database is processed by DBMS, which runs in the main memory and is controlled by the respective operating system

A database is a logically coherent collection of data with some inherent meaning and represents some aspects of the real world. A random assortment of data cannot be referred to as a database. Databases draw a sharp distinction between data and information. Data are known facts that can be recorded and that have implicit meaning. Information is data that have been organized and prepared in a form that is suitable for decision-making. Shortly information is the analysis and synthesis of data. The most fundamental terms used in database approach are entity, attribute and relationship. An entity is something that can be identified in the users' work environment, something that the users want to track. It may be an object with a physical or conceptual existence. An attribute is a property of an entity. A particular entity will have a value for each of its attributes. The attribute values that describe each entity become a major part of data stored in the database

Database Management System is a general-purpose software system designed to manage large bodies of information facilitating the process of defining, constructing and manipulating databases for various applications. Specifying data types, structures and constraints for the data to be stored in the database is called defining a database. Constructing the database is the process of storing data itself on some storage medium that is controlled by the DBMS. Querying to retrieve specific data, updating the database to reflect changes and generating reports from the data are the main concepts of manipulating a database. The DBMS functions as an interface between the users and the database ensuring that the data is stored persistently over long periods of time, independent of the programs that access it [3]. DBMS can be divided into three subsystems; the design tools subsystem, the run time subsystem and the DBMS engine.

The design tools subsystem has a set of tools to facilitate the design and creation of the database and its applications. Tools for creating tables, forms, queries and reports are components of this system. DBMS products also provide programming languages and interfaces to programming languages. The run time subsystem processes the application components that are developed using the design tools. The last component of DBMS is the DBMS engine which receives requests from the other two components and translates those requests into commands

to the operating system to read and write data on physical media. [4]

Database approach has several advantages over traditional file processing in which each user has to create and define files needed for a specific application. In these systems duplication of data is generally inevitable causing wasted storage space and redundant efforts to maintain common data up-to date. In database approach data is maintained in a single storage medium and accessed by various users. The self-describing nature of database systems provides information not only about database itself but also about the database structure such as the type and format of the data. A complete definition and description of database structure and constraints, called meta-data, is stored in the system catalog. Data abstraction is a consequence of this self-describing nature of database systems allowing programdata independence. DBMS access programs do not require changes when the structure of the data files are changed hence the description of data is not embedded in the access programs. This property is called program-data independence. Support of multiple views of data is another important feature of database systems, which enables different users to view different perspective of database dependent on their requirements. In a multi-user database environment users probably have access to the same data at the same time as well as they can access different portions of database for modification. Concurrency control is crucial for a DBMS so that the results of the updates are correct. The DBMS software is to ensure that concurrent transactions operate correctly when several users are trying to update the same data

Using a DBMS also eliminates unnecessary data redundancy. In database approach each primary fact is generally recorded in only one place in the database [6]. Sometimes it is desirable to include some limited redundancy to improve the performance of queries when it is more efficient to retrieve data from a single file instead of searching and collecting data from several files, but this data duplication is controlled by DBMS so as to prohibit inconsistencies among files. By eliminating data redundancy inconsistencies among data are also reduced [5]. Reducing redundancy improves the consistency of data while reducing the waste in storage space. DBMS gives the opportunity of data sharing to the users. Sharing data often permits new data processing applications to be developed without having to create new data files. In general, less redundancy and greater sharing lead to less confusion between organizational units and less time spent resolving errors and inconsistencies in reports. The database approach also permits security restrictions. In a DBMS different types of authorizations are accepted in order to regulate which parts of the database various users can access or update.

2.2 H thng d liu

2.2.1 Data Flow

Dùng D liu (Data Flow) lị s chuyñ ng ca d liu t mt v trờ n v trờ khòc hoc t mt quy trờnh nịy n mt quy trờnh khòc trong h thng [?]. Data Flow thng cú nị d liu bt ngun vị nị nú c tiỏu th hoc lu tr. Trong quò trờnh Data Flow c thc thi, d liu cú th xy ra còc thay i thịn nh dng hoc cu trũc khòc.

Chuyñ i d liu (Data Transformation):

Chuyñ i d liu lị mt phñ quan trng ca quò trờnh dùng d liu, nị d liu c thay i òp ng yỏu cu c th ca quy trờnh hoc h thng. Cú hai hng chònh cho vic thc hìn chuyñ i d liu: theo lũ vị theo thi gian thc.

Trong chu kỳ xử lý theo luồng, dữ liệu được xử lý theo từng tập, thông tin lớn được xử lý vào các khối không gian thời gian thực. Điều này thúc đẩy hợp tác cho các tác vụ yêu cầu xử lý dữ liệu lớn và phục vụ các ứng dụng ngay lập tức.

Ngoài ra, xử lý theo thời gian thực là quá trình xử lý dữ liệu ngay khi nó đến, mà không cần đợi khi có một lượng lớn dữ liệu xử lý. Điều này thông qua các công cụ trong các ứng dụng để hỗ trợ, như xử lý sự kiện thời gian thực.

ETL:

ETL là một phương pháp quan trọng được sử dụng cho việc quản lý dữ liệu trong hệ thống lưu trữ dữ liệu. Về cơ bản ETL bao gồm ba bước chính trong quy trình này.

1. **Extract (Trích Xuất):** Dữ liệu được trích xuất từ các nguồn khác nhau, chẳng hạn như cơ sở dữ liệu, tệp tin, hoặc các dịch vụ trực tuyến.
2. **Transform (Biến đổi):** Dữ liệu được biến đổi để đáp ứng yêu cầu của hệ thống đích. Điều này có thể bao gồm việc làm sạch dữ liệu, chuyển đổi định dạng, hoặc thêm các trường tổng hợp.
3. **Load (Lưu trữ):** Dữ liệu đã biến đổi được nạp vào hệ thống lưu trữ, thông qua một kho dữ liệu hoặc data warehouse.

Data Pipe:

Ngành dữ liệu là một khái niệm quan trọng trong việc khai thác dữ liệu hiệu quả. Các công cụ chuyển đổi và tích hợp dữ liệu đóng vai trò như cầu nối giữa các nguồn dữ liệu khác nhau.

Thông qua việc sử dụng ngành dữ liệu, các tổ chức có thể xử lý một cách linh hoạt và hiệu quả. Các tác vụ như xử lý, theo dõi hiệu suất, và thêm các trường tổng hợp có thể được thực hiện một cách dễ dàng, giúp giảm thiểu sự can thiệp thủ công và tăng cường hiệu quả của hệ thống.

Thách thức:

- **Tính Nhất Quán của Dữ Liệu:** Đảm bảo tính nhất quán của dữ liệu qua các giai đoạn khác nhau của dòng dữ liệu có thể là một thách thức, đặc biệt là trong các hệ thống phân tán.
- **Trễ:** Dòng dữ liệu thời gian thực đòi hỏi phải có một thách thức trong việc duy trì tính kịp thời.
- **Xử lý Lỗi:** Xử lý lỗi trong quá trình dòng dữ liệu là một vấn đề quan trọng để đảm bảo tính chính xác của dữ liệu.
- **Khả năng Mở Rộng:** Khi lượng dữ liệu tăng, việc mở rộng các quy trình dòng dữ liệu trở nên quan trọng để duy trì hiệu suất.

Mặc dù vậy, dòng dữ liệu vẫn đóng vai trò quan trọng trong việc quản lý và xử lý dữ liệu trong các hệ thống khác nhau, và việc hiểu rõ về các khái niệm và công cụ liên quan là cần thiết để tối ưu hóa hiệu suất của hệ thống.

2.2.2 Data Orchestration:

Data Orchestration (Điều phối Dữ liệu) là quá trình phối hợp và quản lý nhiều quy trình dữ liệu, quy trình làm việc hoặc dịch vụ khác nhau để tạo ra một kết quả cuối cùng.

Khởi niệm cơ bản:

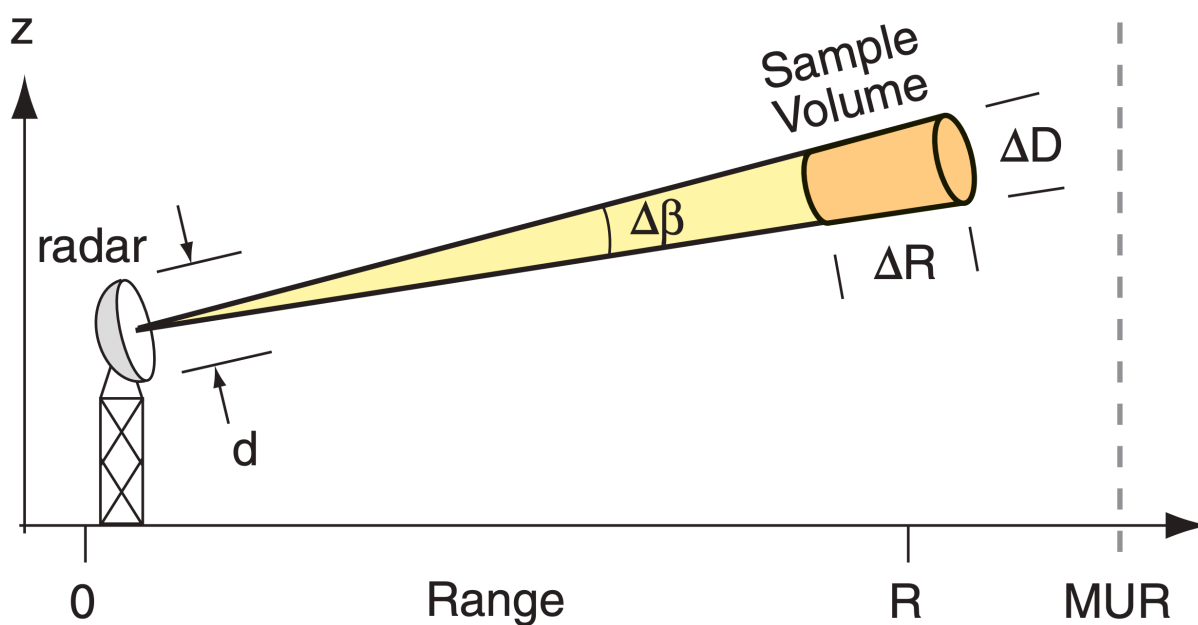
- **Quản lý Quy trình làm việc:** Data Orchestration bao gồm việc nhúng và quản lý các quy trình làm việc, xác định thời gian và thực thi các quy trình đó một cách linh hoạt.
- **Lập lịch Công việc:** Các bước khi lập lịch việc thực hiện các công việc vào thời gian phù hợp với thời gian thực tế và kỳ vọng.
- **Quản lý Sản phẩm:** Các bước khi quản lý sản phẩm các công việc, mở rộng mô hình công việc để thực hiện khi các công việc phức tạp và cần cập nhật.
- **Giám sát và Ghi log:** Hệ thống lưu trữ và cung cấp các công việc giám sát tiến trình của các quy trình làm việc và ghi nhận thông tin liên quan để quyết định.
- **Phân rã Công việc:** Các bước khi phân tích và chia nhỏ các công việc thành nhiều nhiệm vụ và thực hiện chúng song song, cải thiện hiệu suất thực thi.

Thách thức:

- **Phức tạp:** Quản lý các quy trình làm việc phức tạp vì nhiều sản phẩm và logic của việc kết nối các bước thực thi.
- **Môi trường Phức tạp:** Việc lưu trữ các quy trình làm việc trong môi trường phân tán và hệ thống các vấn đề về bảo mật và các vấn đề khác.
- **Quản lý Phiên bản:** Quản lý các thay đổi trong quy trình làm việc và mở rộng tính năng thách thức khi cập nhật lưu trữ các bước thực thi và các bước.
- **Thời gian:** Lưu trữ và thay đổi trong khi lưu trữ các công việc học hỏi và lưu trữ các thách thức trong lưu trữ dữ liệu.

Hiện nay, Data Orchestration đóng một vai trò quan trọng trong việc tích hợp quy trình làm việc và lưu trữ mở rộng hiệu suất của hệ thống. Việc triển khai những nguyên lý này một cách chính xác liên quan đến việc sử dụng các công nghệ và doanh nghiệp trong thị trường hiện nay.

2.3 Lý thuyt v khò tng



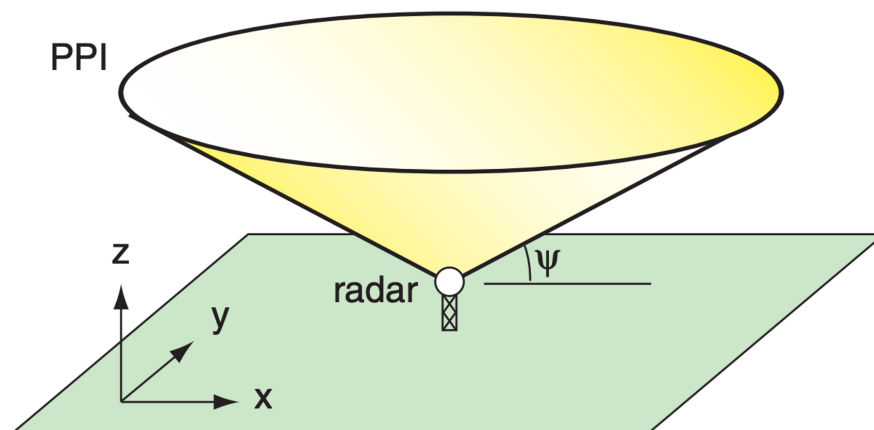
Hình 2.1: H thng radar thí tit - [?]

2.3.1 Còc Khòì Nim C Bn

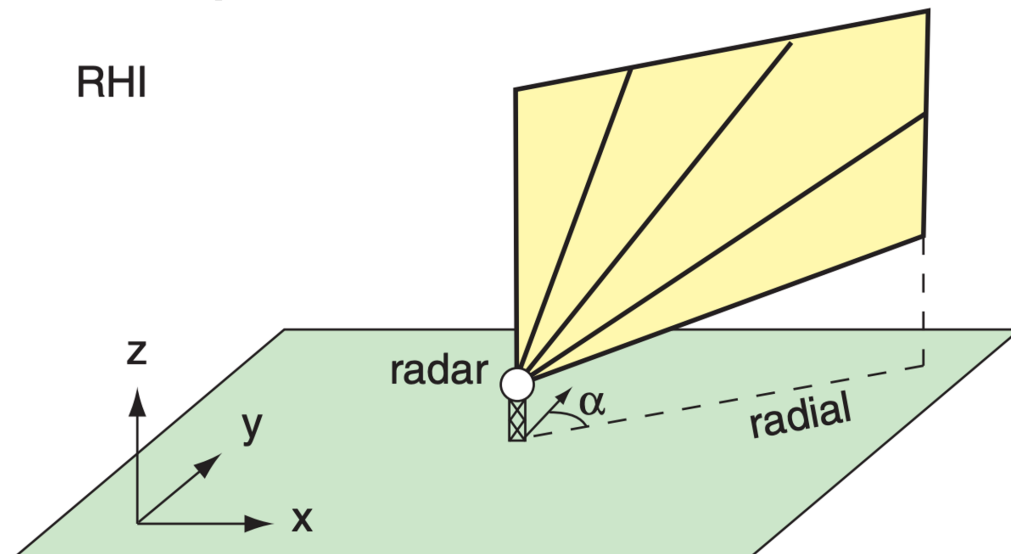
2.3.1.1 Radar thí tít

¹ Radar thí tít lĩ mt loi cm bin cú kh nng phòt sũng vũ tũyn (bc sũng trong phm vi t 250 - 1000 kW) [?]. gia tng cng sũng, mt cho antent (antenna dish) hõnh parabol c s dng nhm hi t bc sũng. Radar cú th nõng vị h (tu theo yõu cu) thu nhp thũng tin tĩ còc v trờ ch nh trong khũng gian 3 chiu.

Thũng thng, còc radar c lp trõnh quĩt theo gũc hng (azimuth) 360° , mĩ vùng s quĩt mt gc nõng khỏc nhau. Nh vỹ, radar s mt khong t 4 n 10 phũt hoĩn thĩnh mt ln quĩt.



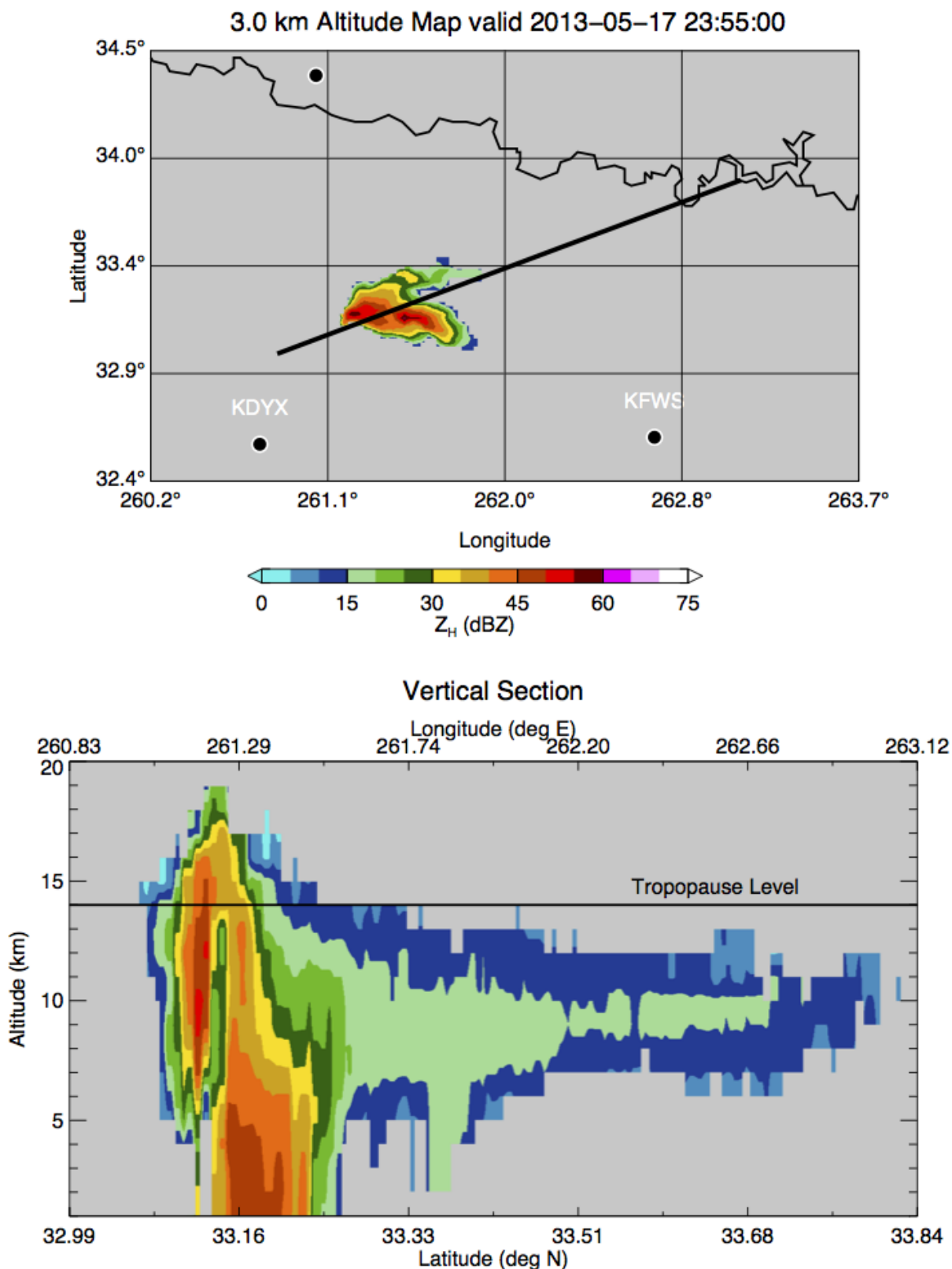
(a) Sn phm quĩt trũn vị gũc nõng c nh (Plan-Position Indicator - PPI) - [?]



(b) Sn phm quĩt thng ng mt gc phng v nhĩ nh (Range Height Indicator) - [?]

i vị biu ãn PPI, radar s quĩt toĩn b gũc hng, nhĩng ch mt gũc nõng nhĩ nh. Kt qu thu c tng t mt bn trũn mt phng. Vĩ RHI, radar s gĩ nguyõn gũc hng nhĩng thay i v gũc nõng. Kt qu thu c giũp ngĩ xem cú cõi nhõn rũ nõĩ hĩn v chiu cao, kỏch thc cĩ hĩn tng khỏ tng.

¹Tổn tĩng Vĩt cĩ còc thut ng s cĩn c ã trũn TCVN 12636-12 : 2021 [?]



Hình 2.3: So sánh kt qu thu c t phng phòp PPI v RHI - [?]

2.3.1.2 Phương trình Radar vị pha hi vũ tuyền

Tiếp tục từ bài trước, radar sẽ phát ra một xung sóng trong không gian ngắn ($\Delta t = 0.5 - 10 \mu s$). Lúc này, tu thủc tiếp xúc phón t t do trong khũng khờ (hi nh, khủi bi, ...), nng lng ca bc sũng nỳ s b hp th mt phn. Cng bc sũng mĩ radar nhn c s nh hn cng sũng ban u. T l nỳ c th hin thũng qua **Phương trình radar** [?]:

$$\left[\frac{P_R}{P_T} \right] = [b] \cdot \left[\frac{|K|}{L_a} \right]^2 \cdot \left[\frac{R_1}{R} \right]^2 \cdot \left[\frac{Z}{Z_1} \right]$$

Trong ú, cõc bin ca phng trnh gm cú:

- $|K|$ khũng cú n v:
 - $|K|^2 \approx 0.93$ cho cõc ht nh lng
 - $|K|^2 \approx 0.208$ cho tinh th bng
- $R(\text{km})$: khong cõch
- $R_1 = \sqrt{Z_1 \cdot c \cdot \Delta t / \lambda^2}$: h s khong cõch
- Z : H s phn hi vũ tuyền ca Radar
- $Z_1 = 1 \text{ mm}^6 \text{ m}^{-3}$: h s n v phn hi vũ tuyền

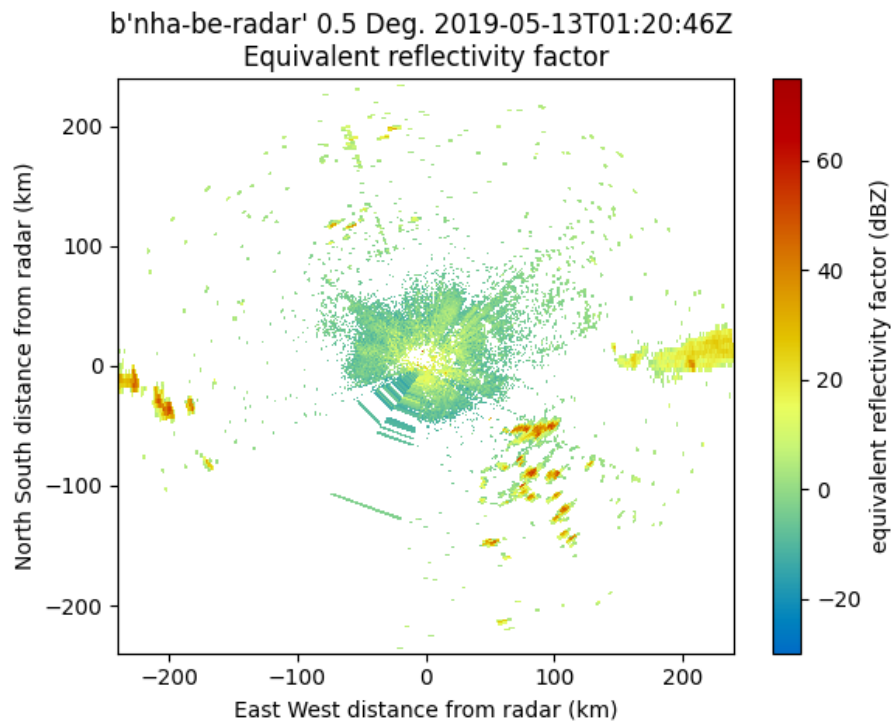
T phng trnh Radar, ta suy ra c cũng thc tõnh phn hi vũ tuyền:

$$\text{dBZ} = 10 \left[\log \left(\frac{P_R}{P_T} \right) + 2 \log \left(\frac{R}{R_1} \right) - 2 \log \left| \frac{K}{L_a} \right| - \log(b) \right]$$

Cõc nhĩ khờ tng thu vn hc thng quan tĩm n con s nỳ vớ nú t l thun vi mc giõng thu (precipitation).

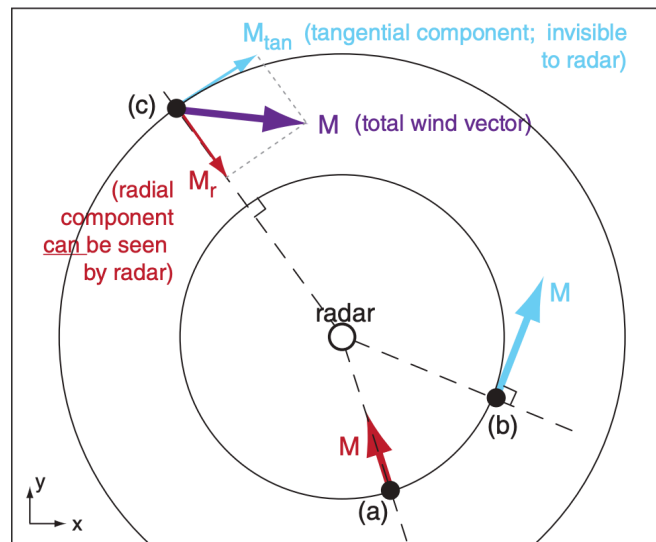
Giõ tr (dBZ)	Thĩ tit
-28	Sng mĩ
-12	Khũng khờ trong lĩnh
25 - 30	Tuyt khũ / ma nh
40 - 50	Ma ln
75	Ma ò khng l

Bảng 2.1: Tng quan h s phn x ca radar vị giõng thu - ?]



Hình 2.4: Minh hoa h s phn x t d liu radar Nhì Bô

2.3.1.3 Vn tc xuyón tótm



Hình 2.5: Minh ho còc tòngh hung vn tc mị radar Doppler cú th quan sòt. (a) Phng ca giú trúngh tì M trúngh vớ ng kòngh ng trúngh cú tótm tì radar, radar cú th xòc nh c vn tc tì óy. (b) Phng ca giú trúngh vớ tìp tuyón ca ng trúngh, radar khũng th xòc nh c vn tc. (c) Phón tòngh hng giú tì M thnh 2 vn tc vuũng gúc nhau, radar ch xòc nh c vector vn tc theo M_r .

Khi còc súngh vũ tuyón t còc radar Doppler nịy tuyón n còc phón t trong khũng khờ, s chuyón dch v trở ca còc ht nịy lch pha gia tòngh hũu tuyón i vị tòngh hũu nhn lĩ c. Còc radar s cn c vộ thũng tìn nịy tòngh toòn vn tc giú tì còc im trong khũng gian.


```
● → titan2023 ncdump -h radar.nc
netcdf radar {
dimensions:
    time = UNLIMITED ; // (1748 currently)
    range = 1198 ;
    sweep = 5 ;
    string_length = 32 ;
```

Hình 2.6: *Thữg tin radar nh dng NETCDF. S chiu ca b d liu tng cng lị 2975 chiu, c phón nhúm cho 4 nhón khòc nhau.*

Bt u t phiỏn bn 4.0, API netCDF gii thiu kh nng s dng nh dng d liu HDF5. S tồch hp quan trng nịy cho phỏp ngi dứg netCDF to tp tin HDF5, m khúa nhng li ờch nh kờch thc tp tin ln hn òng k vị h tr cho nhiu chiu khũng gii hn. Bc tin nịy ònh du mt bc quan trng hng tỉ vic tn dng nhng u im m rng ca nh dng HDF5.

NetCDF Classic vị nh dng 64-bit Offset lị tiỏ chun quc t ca Open Geospatial Consortium[?], th hin s chc chn vị tin cy trong vic m bo kh nng tng thờch vị m rng ca nh dng netCDF trởn toịn cu.

2.5 Cũng ngh s dng