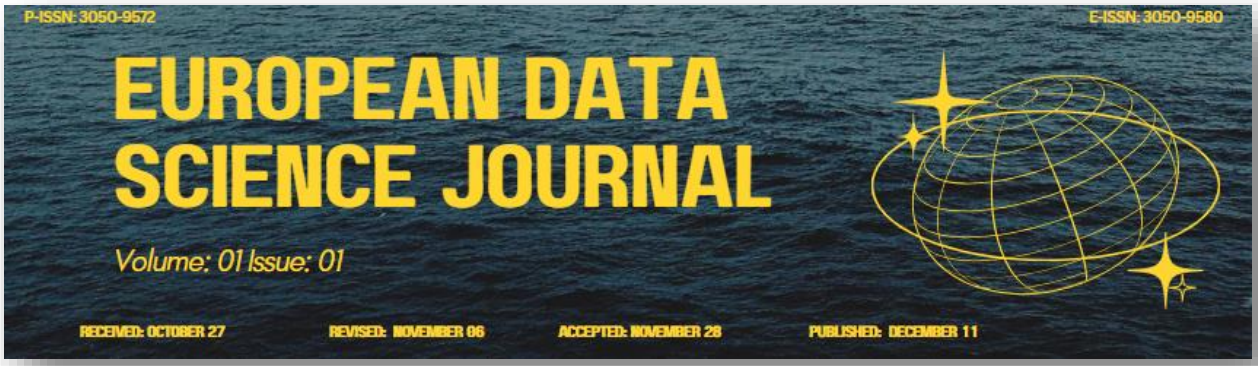


Cloud-Based Big Data Fusion Architecture for National Disaster Management

Dhanaraj Sathiri
Independent Researcher
dhanrajsathiri@gmail.com

Abstract

Cloud-based Big Data integration Frameworks suitable for national Disaster Management Systems are examined, with particular emphasis on national and international Disaster scenarios and the requirements they impose. Key concepts and definitions are reviewed, together with important Cloud deployment and service models. Existing Cloud-based Big Data integration frameworks, able to address the particular demands of national Disaster Management Systems, are identified, with special consideration for Data Ingestion, Processing and Integration capabilities. The evaluation confirms that such frameworks exist but are in part still maturing and require consolidation and independent validation before they can be recommended for wide use.

National and international Disaster Management Systems face many challenges, including high operational costs, the management of data sourced from heterogeneous Data Providers and the lack of a coherent Disaster Analytics framework that integrates these issues. The use of Cloud-based Big Data offers potential solutions to these difficulties, but much work remains to be done before it is possible to recommend a solution suitable for these scenarios. The analysis of Cloud-based Big Data integration Frameworks to address the needs of a national Disaster Management System, highlighted the importance of considering the national and transnational dimensions of specific Disasters. Existing Frameworks exhibit the required properties but are still maturing: Data Processing security and resiliency properties need independent validation before they can be aptly applied in this domain.

Keywords : Disaster, crisis, National Disaster Management System, National Disaster Management Authority, Ontario Disaster Relief Program, Data Processing and Analysis Center, Internet of Things, sensor networks, and post disaster information and data collection.

1. Introduction

Disasters, natural or man-made, claim numerous lives every year. The problem has intensified with increasing population density and climate change. Their impact on life can be minimized but not eliminated. By utilizing the cloud, a nation's resources for disaster management can be developed as a common utility. The cloud may be delivered as a public, private, community, or hybrid service model. National Disaster Management Systems (NDMS) provide reliable, available, and resilient services while meeting security and privacy requirements.

Cloud-based Big Data integration frameworks can act as NDMS and support Disaster Management Systems for a nation or region. Cloud services operate over shared infrastructure, but big data applications process and analyze large-scale heterogeneous data for applications such as disaster management and detection. Efficient NDMS deployment may use Data Ingestion and Streaming Frameworks to ingest data from Weather Data & Sensor Networks, data-processing framework for Data Storm

Management, Automated Detection and Disaster Surveillance Systems based on Social Media Data, and framework for Sensors and Distributed Lights-out Systems.

1.1. Background and Significance

Cloud computing has gained momentum in the area of disaster management and response. It provides the required flexibility and scalability to run disaster simulation, prediction, and analysis systems in a cost-effective way. National governments, state authorities, and organizations involved in disaster management use a number of software applications to prepare a cloud-based big data integration framework for simulating natural calamities and analyzing disaster management and response strategies. These applications fulfill different requirements and run on different hardware and software. However, in most cases, resources are not used efficiently because cloud resources are not shared and the available capacity is not used fully. A common integration framework can thus provide more efficiency and hence support the goals of cloud computing—decreasing operational and investment costs while improving maintenance—while also tackling the inefficiency of using resources that are not coordinated.

Cloud-based big data integration frameworks are analyzed for disaster management based on the integration of various data sources. Sensor networks and the Internet of Things (IoT) are used to gather real-time data such as seismic readings or flood forecasts from sensor nodes that generate large data volumes. Social media, such as Twitter, can be monitored to understand current public sentiment, while government agencies and other organizations can be queried for public information related to a disaster. Datasets from previous disasters can also be archived to support simulation studies. Both data ingestion, to push data into the cloud-based big data platform from producer side applications, and data streaming, to provide real-time computations for decision making, are discussed along with the integration of heterogeneous data sources.

P-ISSN: 3050-9572 E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 01 Issue: 01
RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 28 PUBLISHED: DECEMBER 11

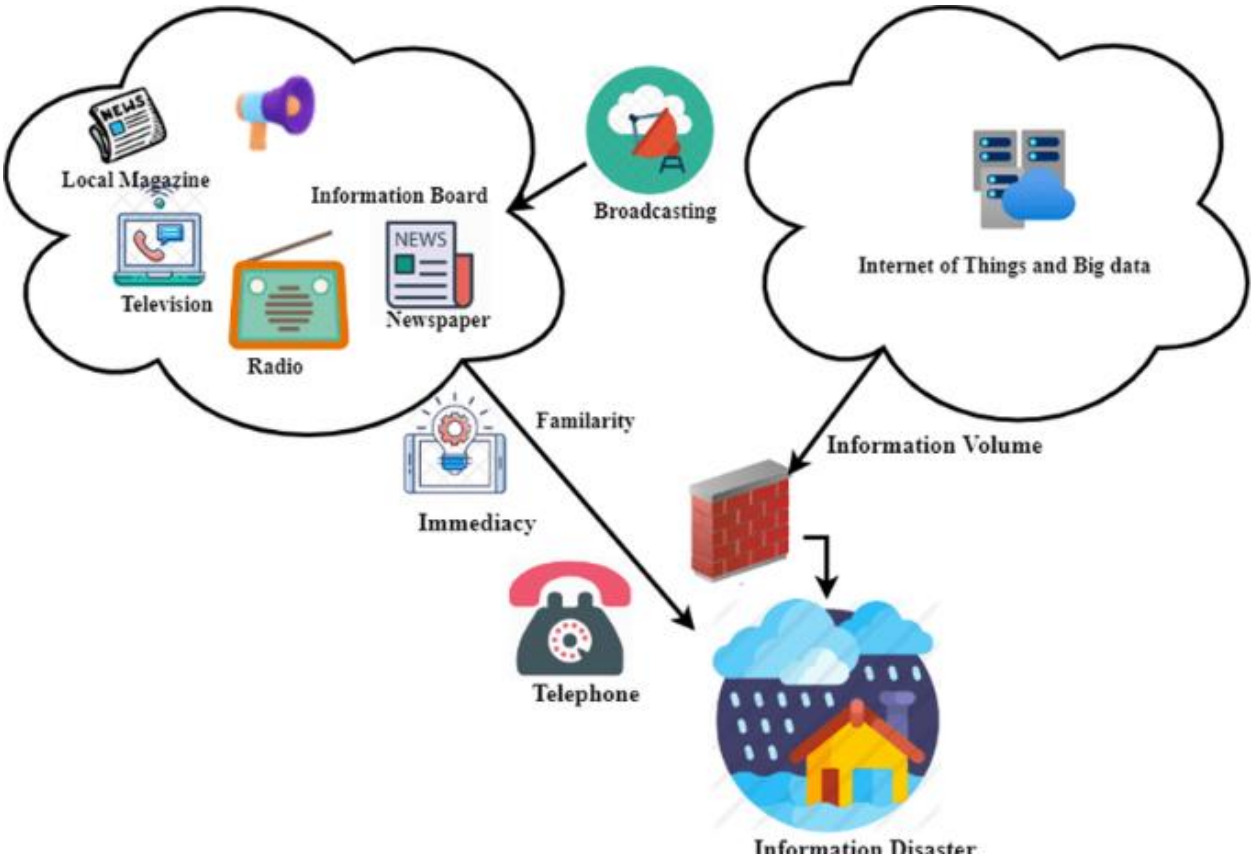


Fig 1: Disaster management based on internet

1.2. Research design

The analytical study examines cloud-based big data integration frameworks for national disaster management systems. First, key terms and concepts are defined, and architectural patterns supporting national disaster management systems are discussed. Next, the requirements of national disaster management systems are analyzed, focusing on operational reliability, availability and resilience, as well as security, privacy, compliance and legal issues. The cloud deployment models and service models relevant to disaster analytics are reviewed, and big data integration frameworks supporting disaster-related scenarios are analyzed. The possible integration of heterogeneous data sources is explored, with specific attention to geo-spatial data, social media data, sensor networks and the Internet of Things. Cloud-based big data integration frameworks for national disaster management systems are proposed in the light of the foregoing analysis. It identifies data loading and streaming, data processing and analytics, and the dissemination of master-analysis results among partner agencies as the principal components for integration. The exploration concludes by pinpointing contours for future research focused on the use of cloud-based big data integration in disaster management and on disaster-resilient smart cities.

Likewise, data from public sources, such as warnings released by local authorities or emergency centers, can also be analyzed for disaster forensics, but they are not always present for every disaster. In the future, NDMSs will need to offer dependable yet highly scalable solutions for large-scale disasters, enabling reliable disaster prediction and disaster forensic analysis for every event.

Equation 1: Derivation of total ingestion rate and storage requirement

Let the framework ingest data from n heterogeneous sources such as sensors, IoT devices, social media, public agencies, and satellite feeds.

Suppose source i produces data at rate r_i(t). The total incoming rate is the sum of all source rates.

- Step 1: For each source, write its contribution: r_1(t), r_2(t), ..., r_n(t).
- Step 2: Add all rates because all flows enter the common ingestion layer.

Therefore, D_total(t) = Σ_{i=1}^{n} r_i(t).

- Step 3: Storage consumed over a time window [0, T] equals the integral of incoming rate over time.

Thus, S(T) = ∫_0^T D_total(t) dt = ∫_0^T Σ r_i(t) dt.

- Step 4: If the rates are approximately constant over the interval, r_i(t)=r_i, then:

D_total = Σ r_i, and S(T) = T Σ r_i.

2. Conceptual Foundations of Cloud-Based Big Data Integration

Cloud-based big data integration frameworks that support various disaster management scenarios, particularly national disaster management systems (NDMSs). Several disaster management processes, including risk identification, forensics, and post-disaster analysis, require access to a plethora of heterogeneous data sources. Cloud-based big data integration frameworks are crucial for these processes because they rely on near-real-time access to large amounts of data and often differ in scale and structure across disaster events. However, NDMSs are also subject to stringent reliability, availability, and security requirements. Therefore research investigates the architectural patterns employed by existing cloud-based big data integration frameworks in order to identify suitable frameworks for NDMS scenarios.

A disaster affects many life aspects and incurs high economic losses. Thus, disaster management processes require detailed and extensive investigations of disasters, which rely on different data types. Although many data types are typically available for all disaster events, these data types differ in scale and distribution across events. Data from sensor networks and the Internet of Things (IoT) are generally available for all events, but the amount of data from social networks such as Facebook, Twitter, and Weibo varies with the scale and publicity of the disaster.

Cloud-based big data integration frameworks play a vital role in supporting disaster management by enabling the collection, processing, and analysis of large volumes of heterogeneous data from multiple sources. In national disaster management systems (NDMSs), processes such as risk identification, disaster forensics, and post-disaster assessment require timely access to diverse datasets, including information from sensor networks, Internet of Things (IoT) devices, satellite imagery, and social media

platforms. These frameworks leverage cloud computing capabilities to provide scalable storage, high processing power, and near-real-time data integration, which are essential during rapidly evolving disaster scenarios. However, NDMSs also demand strict reliability, availability, and security to ensure continuous system operation and protection of sensitive information. Since disasters impact multiple aspects of society and often result in significant economic losses, effective disaster management relies on comprehensive data analysis. While certain data sources, such as IoT sensors, are consistently available across events, others—like social media data from platforms such as Facebook, Twitter, and Weibo—vary depending on the scale and public attention of the disaster. Therefore, understanding and selecting appropriate architectural patterns for cloud-based big data integration frameworks is crucial for building efficient, resilient, and responsive disaster management systems.

Disasters impact many aspects of human life and often lead to significant economic losses, making effective disaster management essential. To respond properly and reduce damage, disaster management processes require detailed and comprehensive investigations of disaster events. These investigations depend on collecting and analyzing different types of data from multiple sources. However, although many data types are generally available for most disaster events, their scale, quantity, and distribution can vary widely from one event to another. For instance, data generated from sensor networks and Internet of Things (IoT) devices—such as environmental sensors, monitoring systems, and smart infrastructure—are usually consistently available across most disasters. In contrast, data from social media platforms like Facebook, Twitter, and Weibo often vary significantly depending on the disaster's scale, location, and the level of public attention it receives. Large and highly publicized disasters typically generate vast amounts of social media data, while smaller or less visible events may produce limited online information. Therefore, integrating and analyzing diverse data sources is crucial for improving disaster monitoring, response, and management strategies.

P-ISSN: 3050-9572
E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 01 Issue: 01
RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 28 PUBLISHED: DECEMBER 11

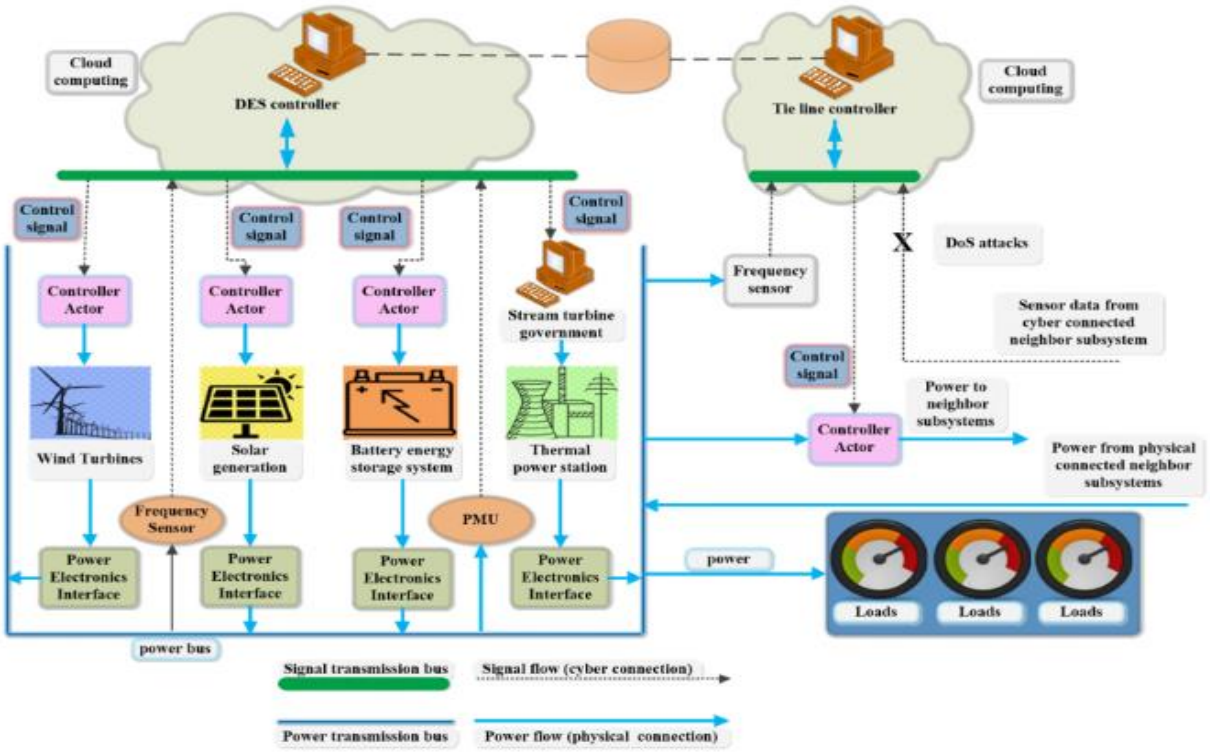


Fig 2: Big Data Analytics Using Cloud Computing

2.1. Definitions and Key Concepts

Cloud technology and services have transformed the way organizations use information technology for business processes, making information technology infrastructure and services easily available, and complementing capacity and management skills. Cloud-based big data integration frameworks enable the integration of heterogeneous data sources, support complex analytic tasks, and optimize service delivery and performance. Cloud-based big data integration frameworks can support national Disaster Management Systems, enabling disaster-related organizations to obtain analytics as a service during natural disasters, especially when disaster conditions overwhelm their existing capabilities.

The cloud model supports five essential characteristics: on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. The Big Data concept characterizes large volumes of structured or unstructured data that have the potential to be mined for information and used in the decision-making process. The three main characteristics of Big Data are volume, variety, and velocity. National Disaster Management Systems are defined as the institutional arrangements for managing the risks of natural disasters at different levels of government in a country, supported by a dedicated body that provides the overall direction and coordinates its implementation at all levels of government. Disaster management is defined as the

continuous process of planning, organizing, training, equipping, exercising, evaluating, and taking corrective action in an overall risk management program so that the impact of disasters is minimized.

2.2. Architectural Patterns for Disaster Management

Today's natural disasters cause not only fatalities and injuries but also social, economic, and environmental losses. In 2014-2019, the average number of natural disasters was annually 337, with about 78% having negative impacts. To overcome these threats, countries must establish a national disaster and emergency management system (NDMS). NDMS quality depends on many attributes, such as reliability, availability, resilience, security, privacy, and compliance. To continuously and fully achieve these attributes, many countries use large-scale cloud computing systems, where Services-50 and Services-51 cloud computing unite the computational and storage resources into a unified logical pool.

Cloud-based big data integration frameworks provide NDMS with a direct and effective environment for natural disaster response. They offer comprehensive approaches for disaster-data processing, information quality assessment, disaster-situation prediction, heterogeneous data-source integration, and data-service streaming and sharing. The frameworks have advantages in monitoring natural and social disaster processes. Moreover, cloud frameworks possess significant disaster-object sensing and disaster-situation warning functions, afford services for nation participants' sensitive information surveillance and integration, and provide a reason to believe that real-time disaster information may be daily collected and shared.

Equation 2: Derivation of processing utilization and stability condition

Let λ be the incoming workload measured in messages per second.

- Step 1: Define utilization as workload divided by capacity.

$\rho = \lambda / (m\mu)$.

- Step 2: For a stable queue, the system must process faster than arrivals on average.

So $\lambda < m\mu$.

- Step 3: Divide both sides by $m\mu$ (positive quantity).

This gives $\rho < 1$.

3. Requirements for National Disaster Management Systems

National disaster management systems require reliability, availability, and resilience, and they also demand security, privacy, and compliance with various government and legal regulations. Reliability and availability need to be established from the perspective of the end users. Resilience is not a requirement of every N-DMS component but is very important for all disaster management centers of a country. A specific sensor installed in a given location may not be available; however, other sensors near that location may still be used. Moreover, different sensors (like wireless sensor networks, IoT sensors, etc.) can be used for similar purposes in different places, then the common information can still be gathered in some way by N-DMS.

P-ISSN: 3050-9572 E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 01 Issue: 01
RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 28 PUBLISHED: DECEMBER 11

Government N-DMS must handle several privacy-related issues, especially because sensitive data from government-sponsored risk detection sensors, auto-detection images, and confidential data from public agencies, such as the police department, need to be integrated with public social media communications. Regulatory compliance also represents a big challenge because the data come from multiple clouds belonging to different countries and using different rules and laws. In addition to national and regional information on data privacy, financial-related authorities must certify that the data-processing procedure does not depict or leak vital security information related to the financial sector.

National Disaster Management Systems (N-DMS) must be designed with a strong focus on reliability, availability, resilience, security, privacy, and regulatory compliance to effectively support disaster response and management. From the perspective of end users, reliability and availability ensure that critical information and services remain accessible during emergencies when they are most needed. While resilience may not be required for every component of an N-DMS, it is essential for disaster management centers to maintain continuous operations even when certain elements fail. For example, if a sensor installed at a specific location becomes unavailable, nearby sensors or alternative technologies such as wireless sensor networks and IoT devices can still provide similar data, ensuring that essential information is not lost. At the same time, government-operated N-DMS must address significant privacy concerns, as they integrate sensitive data from risk detection sensors, automated surveillance images, and confidential information from public agencies like police departments, along with data from public social media platforms. Managing this diverse data securely requires strict adherence to privacy regulations and legal frameworks. Additionally, regulatory compliance becomes complex because data may originate from multiple cloud platforms located in different countries, each governed by its own laws and policies. Financial authorities must also ensure that data-processing procedures do not expose or leak sensitive information related to national financial security, making compliance and secure data handling a critical aspect of N-DMS design and operation.

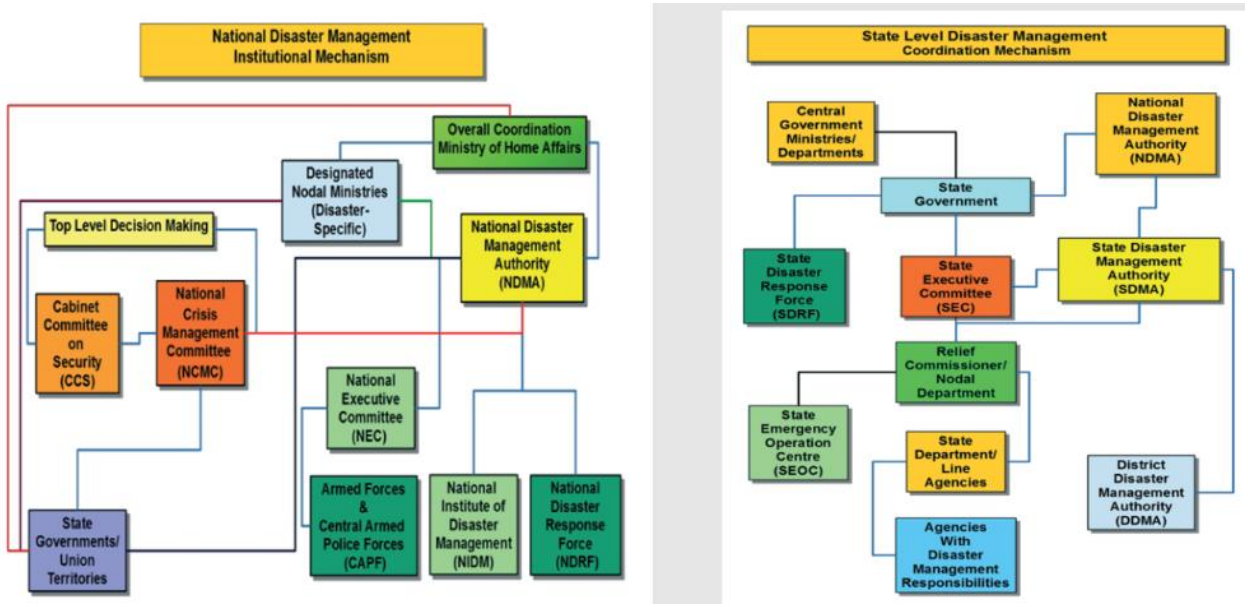


Fig 3: NATIONAL DISASTER MANAGEMENT POLICY

3.1. Reliability, Availability, and Resilience

Numerous issues affect the quality and performance of National Disaster Management Systems. Reliability, availability, and resilience receive particular attention. Reliability is the probability that a system can deliver the required services without failure. Reliability is essential because disaster management systems and applications may be in use for long periods without routine management, and reliability problems may go unreported until services are required in a disaster situation. System failures during such situations can lead to grave consequences.

Availability refers to the proportion of time the system is functional and accessible when required by users. Because public cloud infrastructures span many physical locations and incorporate numerous physical resources, they are often more available than private installations maintained by a single organisation. However, public clouds may not be available during a disaster, as floods, typhoons, earthquakes, and tsunamis can disrupt network, power, and communication services and limit access to the cloud services. Consequently, either hybrid clouds or dedicated and well-resourced private clouds are required for National Disaster Management Systems. Resilience is the capability of a system to recover from disruptions.

Equation 3: Derivation of availability from failure and repair behaviour

Availability is the fraction of time the service is operational.

During one operating cycle, the system is up for MTBF and down for MTTR.

- Step 1: Total cycle time = MTBF + MTTR.
- Step 2: Useful service time = MTBF.
- Step 3: Availability is useful time divided by total cycle time.

A = MTBF / (MTBF + MTTR).

- Step 4: Unavailability is the complement:

U = 1 - A = MTTR / (MTBF + MTTR).

3.2. Security, Privacy, and Compliance

The deployment of cloud-based big data integration for national disaster management systems can enhance operational effectiveness, improve service quality and reduce costs. However, such integration must also meet rigorous security and privacy requirements, comply with applicable legislation, and respect the interests of affected communities and their use of public and private infrastructure. Security refers to security controls and measures. Privacy goes a step further, restricting disclosure of members of a public to other members of the public or to the authorities. Both security and privacy are fundamentally related to the compliance of the operating procedure with local, national and international laws. These can include, amongst others, communication secrecy laws, provisions to establish abuse in criminal proceedings, protections of indigenous peoples, and the GDPR in the European Union.

Cloud service providers devote considerable talent and resources to security beyond what a single organisation could afford. However, the buck stops with the system administrator: “All computers are vulnerable to attack; they are compromised by the

fallibility of their users or the systems on which they execute. The annoying truism of UNIX is that if you are to be a position to read all of a computer's data then you can by definition change it in the way you want". The full entertainment of the cloud conceptualisation—cut price, rapid growth in resources but no control of data—does not square with these requirements. There is considerable luck involved, with how long any service remains free from a security disaster given that, at some time, at least, backups will be exposed as source-accessible. Cyberattacks are a growing security threat, with a peculiarity: the absence of a border. More extensive national and international legislation is required to address these dangers.

4. Cloud Deployment Models and Service Models

Cloud computing offers different deployment models that suit different requirements. National disaster management systems based on Cloud Computing and Big Data Analytics must choose the proper deployment model (Public, Private, Hybrid, or Community Cloud) according to their requirements. At the same time, its disaster analytics can be implemented using Infrastructure as a Service (IaaS) — deploying the analytics tools on virtual machines in the Cloud — or using Platform as a Service (PaaS) or Software as a Service (SaaS) provided by third-party service providers.

Public cloud is based on the services offered by the Cloud Service Providers for the general public. This is the most popular version of Cloud Computing. The services are offered for a fee and the data are stored on the entire cloud infrastructure of the CSP. Several Disaster Management Organizations can use similar services together. The spam information from social media should be filtered. Sensitive applications requiring data protection cannot be deployed in the Public Cloud. Private Cloud is constructed by an organization for its own use. The systems are available only to its members. Disaster Management Organizations worldwide can use Hybrid Cloud. The Interactive Web Map Services and the Data Storage can be deployed on Public Cloud. The operations requiring data protection, such as filtering the information from social media, can be performed in the Private Cloud. Organizations with similar data protection requirements can share the operating systems in Community Cloud. The dynamic nature of Social Media enables the use of the Cloud for handling information.

Cloud computing provides several deployment models—Public, Private, Hybrid, and Community Cloud—that can be selected according to the requirements of national disaster management systems. Public cloud services are offered by cloud service providers for the general public and are widely used due to their scalability and cost-effectiveness. In this model, services are available for a fee and data are stored on the provider's infrastructure, allowing multiple disaster management organizations to use the same resources. However, sensitive applications that require strict data protection are not suitable for deployment in a public cloud. A private cloud, on the other hand, is built and maintained by an organization for its exclusive use, ensuring greater control, security, and privacy for critical operations. Hybrid cloud combines both public and private cloud environments, enabling organizations to store interactive web map services and large datasets on the public cloud while performing secure tasks, such as filtering social media information, within the private cloud. Community cloud is designed for organizations with similar security and operational requirements, allowing them to share infrastructure and services. Additionally, disaster analytics in such systems can be implemented through Infrastructure as a Service (IaaS) by deploying analytics tools on virtual machines, or through Platform as a Service (PaaS) and Software as a Service (SaaS) offered by third-party providers. The dynamic nature of social media data further supports the use of cloud computing to efficiently process and manage disaster-related information.

Hybrid clouds combine public and private clouds, allowing secure transactions for sensitive data through a private infrastructure while simultaneously using public infrastructure for non-sensitive transactions. Demand spikes may be managed by hiring public

P-ISSN: 3050-9572 E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 01 Issue: 01
RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 28 PUBLISHED: DECEMBER 11

cloud capacity as necessary. A community cloud is a variation on the cloud concept where the physical infrastructure is shared among several organizations from a specific community with common concerns, such as security, policy, or compliance, and is managed internally or by a third party.

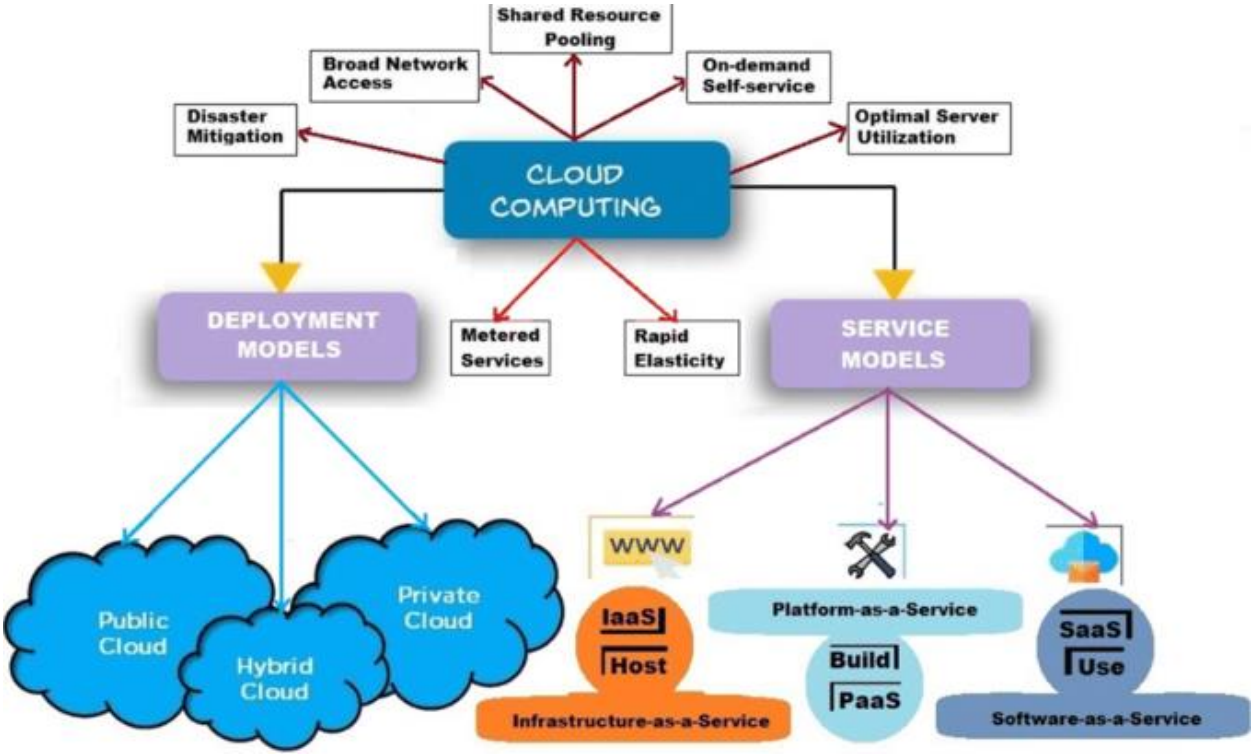


Fig 4: Cloud Computing Deployment and Service model

4.1. Public, Private, Hybrid, and Community Clouds

Cloud-based systems can be deployed in four ways: in public, private, hybrid, or community clouds.

Public clouds are owned and operated by cloud service providers who deliver services and applications to customers over the internet. Resources are shared among various organizations, optimizing capacity costs. Public clouds enable rapid scaling and variable costs. However, with growing national data protection regulations and the security concerns cloud customers share, many applications that require secure data handling may find it impossible to use public clouds.

Private clouds are operated exclusively for a single organization. They can be located in an organization's on-site data center or can be hosted by third-party service providers. A private cloud can serve a single organization or group of organizations, offering higher levels of security and data protection. However, dedicated resources may result in higher costs than public clouds, and

having the same capacity as a public cloud may take too long. Consequently, a company may have to periodically purchase additional resource capacity from a public cloud provider.

4.2. IaaS, PaaS, and SaaS for Disaster Analytics

National Disaster Management Systems (NDMS) may utilize Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) during disaster preparedness, response and reconstruction phases. An IaaS model can provide the required computing power and flexible data storage needed for disaster data ingestion and processing, which is usually done at large scales. IaaS can also be used to support the hosting of the NDMS itself when it is in demand. NGI, a national geospatial cloud for disaster response in India, uses IaaS to host services that cater to disaster response operations. Supporting applications and devices can also use IaaS to store and process large volumes of continuously generated data from the detection phase. Detection data are monitored and verified by humans connected to NDMS so that only truly detected events are considered in the subsequent phases. Sensor node failures and false detections are also verified in this phase.

Moving on to PaaS, it can help NDMS, the supporting services and other disaster analytics applications streamline the configurations needed for real-time distributed-data analytics at a massive scale. Its deployment at national or community levels can greatly simplify the development and acceleration of applications that require real-time and/or large-scale data-intensive processing, such as transforming sensing layer and telecommunication data into structured information. Using its analytics cloud, NDMS and other disaster management agencies can deploy and host machine-learning models that continuously and incrementally learn from inputs streaming from the sensing phase as part of decision-support systems. SaaS enables ready-to-use NDMS supporting applications for the common public and regular Android-based communications devices enabling natural communication with NDMS.

Equation 4: Derivation of reliability over time

Let λ_f be the failure rate of a component in failures per hour.

For a small interval Δt, the probability of surviving is approximately $1 - \lambda_f \Delta t$.

Applying this repeatedly over time leads to the exponential survival model.

Therefore, component reliability is $R(t) = e^{-\lambda_f t}$.

Series system case: if all k components must work, then overall reliability is the product:

$R_{series}(t) = \Pi_{j=1}^{k} R_j(t)$.

If all components have the same failure rate λ_f, then:

$R_{series}(t) = (e^{-\lambda_f t})^k = e^{-k \lambda_f t}$.

Parallel redundancy case: if one of k replicated nodes is enough, then failure occurs only if all replicas fail.

$R_{parallel}(t) = 1 - \Pi_{j=1}^{k} (1 - R_j(t))$.

5. Big Data Integration Frameworks for Disaster Scenarios

Big data integration frameworks are proposed for disaster scenarios. The main processing modules are described: the ingestion layer provides data pipelines and offers support for data streaming; the processing layer integrates batch, micro-batch, and stream processing for data analysis; and the analytics layer provides sampling methods for quality assessment of disaster data.

Recent natural disasters have revealed deficiencies in many national disaster management systems, with inadequate preparation, response, and recovery due to lack of required infrastructure. Cloud-based big data integration frameworks can address this problem. Cloud should be used not just for data storage but also for processing and real-time disaster analysis, and the four main disaster scenarios—before, during, immediately after, and after a disaster—must be considered. Data collection, management, and integration should be performed for each scenario. To prepare for natural disasters, the administration needs to predict the location, time, and strength of the event. As the event approaches, social media becomes the primary means of sharing information, with conventional communication channels usually remaining normal. Indeed, during an event, social media can provide instantaneous information regarding the condition on the ground. Consequently, much of the information provided through geospatial sensor networks, such as weather, hydrological, and seismic data, can become redundant.

During the disaster, data collected from social media can aid both the public and administration. Data posted by the public can facilitate operational operations by pointing to hotspots in distress, while information posted by administration agencies can warn the public regarding route danger points, shelter availability, and food conditions at various places. As the disaster passes, administering bodies put out requests for data for damage assessment. After the disaster, live information on site can help identify rescue and rehabilitation needs, while information on damaged infrastructures can provide the appropriate caretaking authorities with recovery needs.

Equation 5: Derivation of weighted data-fusion score

Let x_i be the normalized contribution of source i and w_i be its trust/importance weight.

- Step 1: Normalize weights so that $\Sigma w_i = 1$.
- Step 2: Form the weighted sum:

$F = \Sigma_{i=1}^{n} w_i x_i.$

- Step 3: If the weights are not normalized initially, say raw weights α_i, convert them using:

$w_i = \alpha_i / \Sigma \alpha_i.$

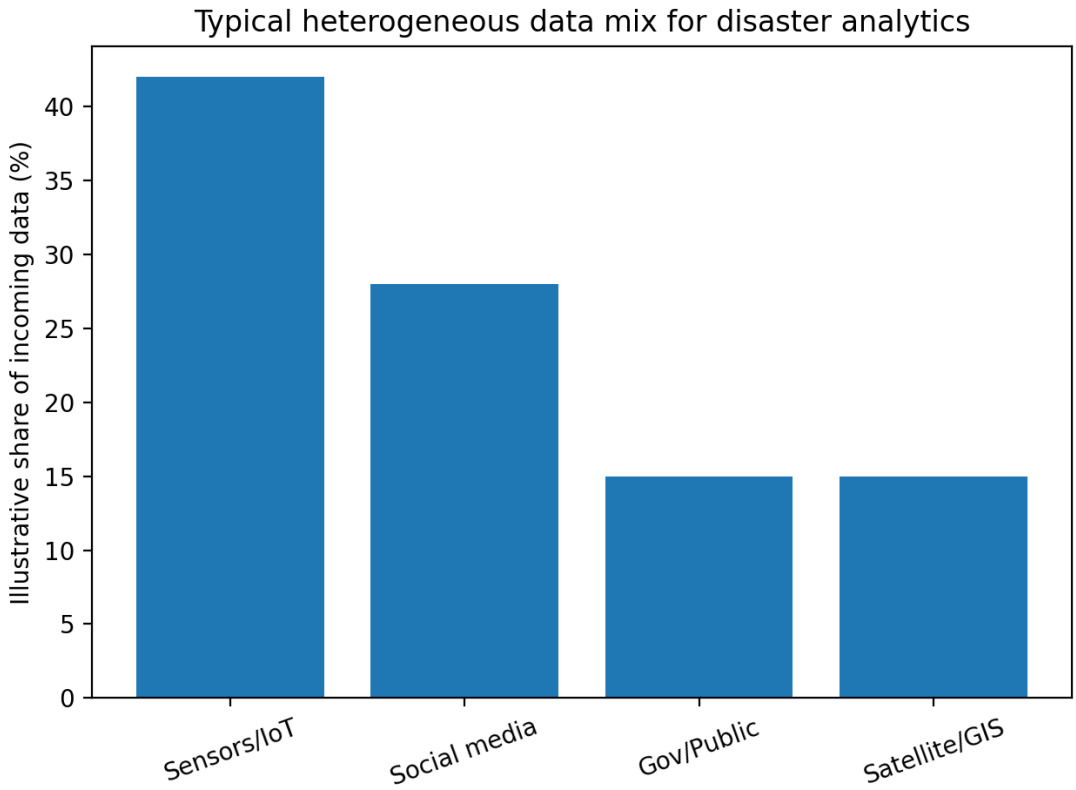
Then $F = \Sigma (\alpha_i / \Sigma \alpha_i) x_i.$

5.1. Data Ingestion and Streaming

The ingestion of real-time data produced during disaster scenarios is a key function in any cloud-based national disaster management system. Cloud-based platforms must support high-throughput data streaming services that enable these data to be made available to interested applications and services in a timely manner. The data produced during disasters are often event-driven and do not respect predefined schedules. Therefore, it is necessary to move away from traditional batch-oriented data ingestion methods based solely on data polling. An architecture tailored to the specific needs of disaster management has to be developed, one in which data produced during a disaster are streamed into the cloud by a set of producers and made available to cloud consumers through a corresponding set of data streaming services. At its core, the architecture relies on publish-subscribe messaging.

P-ISSN: 3050-9572
E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 01 Issue: 01
RECEIVED: OCTOBER 27
REVISED: NOVEMBER 06
ACCEPTED: NOVEMBER 28
PUBLISHED: DECEMBER 11

Several cloud-based, publish-subscribe messaging services are available today. These services enable flexible and decoupled communication between cloud applications, allow routing by content, and provide excellent Quality of Service (QoS) support. They can be functional for disaster data from sensor networks and the Web in general, with dedicated and optimized wrappers absorbing the data from each category of producer and publishing it to the corresponding cloud messaging service. The wrappers can communicate with one or more cloud consumers by means of subscription requests and can enable the streaming of disaster-related Social Media data, making them directly available for analysis and data fusion together with other feeds such as those from sensor networks.



5.2. Data Processing and Analytics

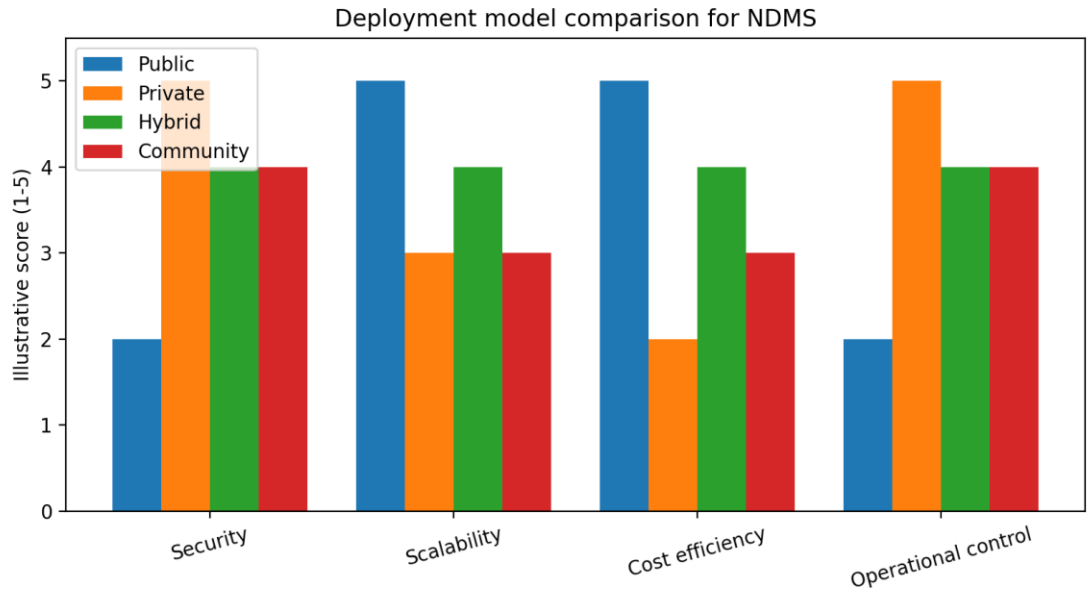
The data generated by disaster-related applications is a mix of both structured and unstructured data. For such applications, the relevant data could consist of visual surveillance videos or audio streams from major cities, signals posted in a social networking web sites, Wikipedia updates, and meteorological models in addition to images taken by satellites. Therefore, the processing phase for such applications should be able to process diverse forms of data. Managed cloud computing environments provide tools for processing unstructured data, such as visual surveillance videos or audio data, more efficiently.

P-ISSN: 3050-9572
E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 01 Issue: 01
RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 28 PUBLISHED: DECEMBER 11

6. Integration of Heterogeneous Data Sources

Several heterogeneous data sources contribute to big data analysis during disaster situations. Data collected from sensor networks or the emerging Internet of Things (IoT) is often used through sensor data access and resource management services, such as those provided by Sensor as a Service (SaaS) or IoT as a Service (IoTaaS). These services enable cloud users to access sensor data in an easy and flexible way. Data from social media and various public communication sources, including satellite images, play important roles in big data analysis for disaster situations. Moreover, users' personal observations posted on social media platforms like Twitter and Facebook can be crucial for various disaster scenarios. For instance, in the 2015 Nepal earthquake, the pictures and videos uploaded to social media by people in the affected area provided significant support to disaster management. However, due to the enormous volume of data produced on social media platforms, it is challenging to utilize them effectively in big data disaster analysis.

As previously discussed, the services provided by social media platforms, including streaming APIs, offer a solution for real-time data collection. Twitter's streaming API allows authorized developers to access Twitter data in real time. Twitter employs the Publish-Subscribe Messaging pattern for these streaming APIs, which optimizes the operating time and utilization level. The Published Data model provides metadata about the content uploaded by users, while the Subscribers model delivers a dynamic subscribing mechanism that allows other users to actively collect data related to their interests. When a user follows a certain account, they automatically receive updates based on their interests and the profile of the account they follow.



6.1. Sensor Networks and IoT

National Disaster Management Systems (NDMS) require integration of heterogeneous data sources to support proactive, reactive,

and post-disaster analytics. During emergencies, incoming data is often uncertain or temporally and spatially inaccurate. These streams must be checked, cleaned, and correctly sensed in order from disparately-located sources. Sensor data generated by devices, smartphones, vehicles, and social networks can be merged for disaster prevention, as can publicly available government information. Many users sharing nearby disaster content over social networks can be effectively utilized to watch, report, and warn others as the situation develops.

The volume of data generated by the Internet of Things (IoT) and sensor networks is rapidly increasing because of the extensive use of networks, sensors, and data collection devices. These data sources provide near real-time data relevant to disaster management. Sensor networks are being deployed to monitor environmental conditions, providing important data to NDMS. IoT-based sensor networks and smart city application environments are key sources for proactive disaster management solutions. As a response to floods in southern India, the torrent monitoring, early warning, and smart rescue plan demonstrate a novel application architecture based on such technologies. During crises, data sizes rapidly increase, with public social media platform APIs facilitating useful real-time data collection for disaster management and reduction.

6.2. Social Media and Public Communications

The ever-changing dynamics of social media introduce additional challenges for disaster management organizations in filtering real-time posts and reaching out to users for valuable information. There are two approaches to address these difficulties in managing social media data. One approach is to use textual information retrieval to extract information from posts based on some keywords and phrases like rescue or help. Associated with these keywords, geotagged tweets can also be retrieved and then analyzed for a given region. In addition, more advanced machine learning and natural language processing models are being developed to automatically classify important posts that require attention, like requests for rescue or medical help, especially in urgent situations. The second approach is to actively engage the social media user community in providing real-time information to assist NDMOs. For example, the Government of India initiated a project called Ayushman Bharat, aiming to provide affordable medical facilities to every individual. It has established a dedicated 24×7 WhatsApp number for people to share free medical consultations, significant health hazards, and incidents.

Emergency management agencies around the world are increasingly adopting social media to communicate, share information, and interact with the public before, during, and after emergencies. They use emerging social media platforms to engage the public not only as recipients of information but as active participants. Social media facilitates the dissemination of information during emergencies, allows emergency managers to crowdsource information to supplement publicly available channels, and enables a two-way dialogue for increased public awareness. However, the dynamic nature of social interactions introduces multiple challenges in managing such information, including massive data collection, noise, relevance, and validation, and the effective use of social media for emergency notification.

7. Conclusions

Cloud-based big data integration frameworks that harness the capabilities of cloud computing and big data are promising for the creation of national disaster management systems. Such systems must monitor potential natural and man-made disasters, manage periodic simulations, perform disaster planning, coordinate emergency rescue and evacuation operations, and provide post-disaster assistance. A national disaster management system must reliably serve tens of thousands of users of various data classification and access levels.

The integration of various technology domains enables a highly scalable and cost-effective framework. Large numbers of data sources for disaster scenarios are integrated, including data streams from sensor networks and the Internet of Things, public communications such as social media, online training and examination systems, and disaster management websites. Data sources are integrated in a multidimensional data model using an enterprise service bus architecture. The demand for reliability, availability, and resilience; security, privacy, and compliance requirements; and appropriate cloud service models is validated for disaster processing and analytics.

Moving forward, the integration of neural networks with cloud computing and big data technologies for disaster management will enable the performance of diverse prediction and classification tasks. Investigations into the hybrid analysis of social network and news feed data for disaster management are still in their infancy. The performance of multiple types of data processing during a single disaster event forms a distinct research domain. The modeling of hybrid cloud-based frameworks using descriptive, predictive, and prescriptive analytics is an emerging trend. The integration of neural networks with cloud computing and big data technologies is expected to significantly enhance disaster management systems in the coming years. By leveraging large-scale data processing and advanced machine learning techniques, these systems can perform a wide range of prediction and classification tasks, enabling faster and more accurate disaster response. However, the hybrid analysis of social network data and real-time news feeds for disaster management is still in its early stages, presenting new opportunities for research and innovation. Managing and processing multiple types of data simultaneously during a single disaster event creates a unique research domain that requires sophisticated analytical approaches. In this context, the development of hybrid cloud-based frameworks that incorporate descriptive, predictive, and prescriptive analytics is emerging as a promising trend, offering improved decision-making capabilities and more effective disaster preparedness and response strategies.

Article theme	Derived equation	Why it matters
Data ingestion and streaming	$D\_total(t) = \Sigma\ r\_i(t)$, $S(T)=\int D\_total(t)dt$	Quantifies multi-source data flow into the cloud ingestion layer
Processing and analytics	$\rho = \lambda/(m\mu)$, stability when $\rho<1$	Shows when stream processing remains sustainable during a disaster surge
Reliability and resilience	$R(t)=e^{-\lambda\_f\ t}$, $R\_parallel = 1-\Pi(1-R\_j)$	Captures redundancy benefits and failure behaviour
Availability	$A = MTBF/(MTBF+MTTR)$	Measures service uptime needed for national systems
Heterogeneous integration	$F = \Sigma\ w\_i\ x\_i$	Explains weighted fusion across sensors, social media, and official feeds
Risk-oriented decision support	Risk* = H E V (1-P)	Supports prioritization for response and mitigation
Cloud deployment economics	$C\_total = C\_priv + c\_pub\ B + C\_sec$	Represents hybrid-cloud burst handling and governance costs

Table: Summary table linking the mathematics to the article

7.1. Future Trends

Future trends include an increasing use of social media for disaster management and follow-up analysis without user privacy invasion through user profile analysis. Real-time, timely, and on-line integration of big data for supporting disaster scenarios over different regions and countries with different languages using mutual geographical location is necessary. To realize this concept and support national disaster management system readiness over a region or country, integration of huge and heterogeneous data sources like sensor networks, social media, public communications, and mass communication over network remains a challenge. Most of these sources are massive, dynamic, and vulnerable to attacks, so a reliable and efficient approach to integrate these sources, considering the above aspects, is extremely essential.

Once the big data remains available, there is a need for big data processing and analytical support for supporting different disaster-related tasks. Disaster-related big data are also massive and complex to consume and need special support for accessing, and these special accesses can be offered through data analytics support. Recent trends in every research discipline suggest that cloud computing and virtualisation are the regions of interest. Nevertheless, for disaster scenarios, disaster management as a service based on all the elements and support available from a cloud-based framework (from cloud deployment models as well as service models) is also a promising concept. A national disaster management system can be presented as the most disaster-oriented data analytics cloud service model.

8. References

1. Chaudhuri, N., & Bose, I. (2020). Exploring the role of deep neural networks for post-disaster decision support. Decision Support Systems, 130, 113234.
2. Fan, C., Esparza, M., Dargin, J., Wu, F., Oztekin, B., & Mostafavi, A. (2020). Spatial biases in crowdsourced data: Social media content attention concentrates on populous areas in disasters. Computers, Environment and Urban Systems, 83, 101514.
3. Yandamuri, U. S. (2021). A Comparative Study of Traditional Reporting Systems versus Real-Time Analytics Dashboards in Enterprise Operations. Universal Journal of Business and Management, 1(1), 1-13.
4. Fathi, R., Thom, D., Koch, S., Ertl, T., & Fiedrich, F. (2020). VOST: A case study in voluntary digital participation for collaborative emergency management. Information Processing & Management, 57(4), 102174.
5. Gazi, T., & Gazis, A. (2020). Humanitarian aid in the age of COVID-19: A review of big data crisis analytics and the General Data Protection Regulation. International Review of the Red Cross, 102(913), 75–94.
6. Gerdes, A. (2020). A moderate interpretation of group privacy illustrated by cases from disaster management. Journal of Contingencies and Crisis Management, 28(4), 446–452.

7. Iglesias, C. A., Favenza, A., & Carrera, Á. (2020). A big data reference architecture for emergency management. Information, 11(12), 569.
8. Inala, R. Designing Scalable Technology Architectures for Customer Data in Group Insurance and Investment Platforms.
9. Jin, X., & Spence, P. R. (2020). Understanding crisis communication on social media with CERC: Topic model analysis of tweets about Hurricane Maria. Journal of Risk Research.
10. Kankanamge, N., Yigitcanlar, T., Goonetilleke, A., & Kamruzzaman, M. (2020). Determining disaster severity through social media analysis: Testing the methodology with South East Queensland flood tweets. International Journal of Disaster Risk Reduction, 42, 101360.
11. Khan, A., Gupta, S., & Gupta, S. K. (2020). Multi-hazard disaster studies: Monitoring, detection, recovery, and management, based on emerging technologies and optimal techniques. International Journal of Disaster Risk Reduction, 47, 101642.
12. Kwapong Baffoe, B. O., & Luo, W. (2020). Humanitarian relief sustainability: A framework of humanitarian logistics digital business ecosystem. Transportation Research Procedia, 48, 363–387.
13. Inala, R. (2020). Building Foundational Data Products for Financial Services: A MDM-Based Approach to Customer, and Product Data Integration. Universal Journal of Finance and Economics, 1(1), 1-18.
14. Lin, A., Wu, H., Liang, G., Cardenas-Tristan, A., Wu, X., Zhao, C., & Li, D. (2020). A big data-driven dynamic estimation model of relief supplies demand in urban flood disaster. International Journal of Disaster Risk Reduction, 49, 101682.
15. Liu, Z., Du, Y., Yi, J., Liang, F., Ma, T., & Pei, T. (2020). Quantitative estimates of collective geo-tagged human activities in response to Typhoon Hato using location-aware big data. International Journal of Digital Earth, 13(9), 1072–1092.
16. Malawani, A. D., Nurmandi, A., Purnomo, E. P., & Rahman, T. (2020). Social media in aid of post disaster management. Transforming Government: People, Process and Policy, 14(2), 237–260.
17. Nagendra, N. P., Narayanamurthy, G., & Moser, R. (2020). Management of humanitarian relief operations using satellite big data analytics: The case of Kerala floods. Annals of Operations Research.

18. Park, M., Jung, D., Lee, S., & Park, S. (2020). Heatwave damage prediction using random forest model in Korea. Applied Sciences, 10(22), 8237.
19. Mangalampalli, B. M. (2021). Scalable Data Warehouse Architecture for Population Health Management and Predictive Analytics. World Journal of Clinical Medicine Research, 1(1), 1-18.
20. Qayum, A., Ahmad, F., Arya, R., & Singh, R. K. (2020). Predictive modeling of forest fire using geospatial tools and strategic allocation of resources: EForestFire. Stochastic Environmental Research and Risk Assessment, 34(12), 2259–2275.
21. Romascanu, A., Ker, H., Sieber, R., Greenidge, S., Lumley, S., Bush, D., Morgan, S., Zhao, R., & Brunila, M. (2020). Using deep learning and social network analysis to understand and manage extreme flooding. Journal of Contingencies and Crisis Management, 28(3), 251–261.
22. Sarker, M. N. I., Peng, Y., Yiran, C., & Shouse, R. C. (2020). Disaster resilience through big data: Way to environmental sustainability. International Journal of Disaster Risk Reduction, 51, 101769.
23. Song, X., Zhang, H., Akerkar, R. A., Huang, H., Guo, S., Zhong, L., Ji, Y., Opdahl, A. L., Purohit, H., Skupin, A., Pottathil, A., & Culotta, A. (2020). Big data and emergency management: Concepts, methodologies, and applications. IEEE Transactions on Big Data, 8(2), 397–419.
24. Talley, J. W. (2020). Disaster management in the digital age. IBM Journal of Research and Development, 64(1–2), 1:1–1:5.
25. Aljumah, A., Kaur, A., Bhatia, M., & Ahamed Ahanger, T. (2021). Internet of things-fog computing-based framework for smart disaster management. Transactions on Emerging Telecommunications Technologies, 32(8), e4078.
26. Mukesh, A., & Aitha, A. R. (2021). Insurance Risk Assessment Using Predictive Modeling Techniques. International Journal of Emerging Research in Engineering and Technology, 2(4), 68-79.
27. Bag, S., Gupta, S., & Wood, L. (2021). Big data analytics in sustainable humanitarian supply chain: Barriers and their interactions. Annals of Operations Research.
28. Bell, D., Lycett, M., Marshan, A., & Monaghan, A. (2021). Exploring future challenges for big data in the humanitarian domain. Journal of Business Research, 131, 453–468.

29. Fan, C., Zhang, C., Yahja, A., & Mostafavi, A. (2021). Disaster City Digital Twin: A vision for integrating artificial and human intelligence for disaster management. International Journal of Information Management, 56, 102049.
30. Ding, Z., Jiang, S., Xu, X., & Han, Y. (2022). An Internet of Things based scalable framework for disaster data management. Journal of Safety Science and Resilience, 3(2), 136–152.