UNIVERSITY OF PIRAEUS

DEPARTMENT OF DIGITAL SYSTEMS

POSTGRADUATE PROGRAMME:

“TECHNO-ECONOMIC MANAGEMENT AND SECURITY OF DIGITAL SYSTEMS”

AERA OF STUDY:

TECHNO-ECONOMIC MANAGEMENT OF DIGITAL SYSTEMS

MASTER THESIS

TITLE:

USING BIG DATA TECHNOLOGIES IN MARITIME SHIPPING INDUSTRY TO ACHIEVE COST EFFECTIVENESS

Student Pavlos Vernikos

Proffessor Panagiotis Demestixas – Konstantinos Tsagaris

Athens, October 2014

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ:

"ΤΕΧΝΟΟΙΚΟΜΟΜΙΚΗ ΔΙΟΙΚΗΣΗ ΚΑΙ ΑΣΦΑΛΕΙΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ"

ΚΑΤΕΥΘΥΝΣΗ:

ΤΕΧΝΟΟΙΚΟΜΟΜΙΚΗ ΔΙΟΙΚΗΣΗ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΙΤΛΟΣ:

ΧΡΗΣΗ ΤΕΧΝΟΛΟΓΙΩΝ ΕΠΕΞΕΡΓΑΣΙΑΣ ΜΕΓΑΛΩΝ ΔΕΔΟΜΕΝΩΝ ΣΤΗΝ ΝΑΥΤΙΛΙΑ ΓΙΑ ΜΕΙΩΣΗ ΟΙΚΟΝΟΜΙΚΩΝ ΠΟΡΩΝ

Φοιτητής **Παύλος Βερνίκος**

Καθηγητής **Παναγιώτης Δεμέστιχας – Κωνσταντίνος Τσαγκάρης**

Αθήνα, Οκτώβριος 2014

I would like to thank Mr. Demestixas for his supervision, Dr. Konstantinos Tsagaris for his support and guidance, Mr. Nikos Papanikolaou for his maritime knowledge, Mr. Nikos Kapinakis for his knowledge in Handoop implementation, Mr. Dimitris and Panos Tsikopoulos for the all the experience provided all those years in maritime industry.

Abstract

Information has always been one of the key points in knowledge production process. Everything that we know about most things in our world today is based on the results of observation and experimenting. Those results are nothing more than pure information, so those procedures can be described as information production producers.

Since information is a key point in the knowledge production process it's vital this information be collected, kept and processed. Collection of all this information, in all aspects of knowledge as we mentioned, is the creation of data. Data is nothing more than a collection of a big amount of information. The format of those data, which has passed several stages through the years as well as several technological revolutions, has led us today to electronic format and electronic storage. Analyzing of those data was the only way to create knowledge in all sciences at the early years. As time passed by, data, which is still produced in higher levels, overcame the early stages of basic knowledge creation and has been widely introduced to all kind of industries trying to provide great benefits and great challenges. Most companies have always tried to get the pragmatic approach of the collected data, using them for beneficial decision making.

Data has increased their complexity a lot and several models of storing, managing have been implemented during the last decades. We have data stored in simple serial data bases to new relational databases, for structured data, while other unstructured data (due to their contents that were pictures, videos etc.) was not able to be stored. During the years the information sources have also changed and new ones have been added, such as data from social media, text streaming data from sensors and other sources.

When you are dealing with so much information in so many different forms, it's not possible to use traditional data management ways. The size of the data was always increasing; the difference today is that we have to deal with different types of information and timeliness.

For Maritime shipping industry the main issue has always been the transmission of the information from the vessels due to low speed and at the same time extremely expensive transmission cost based on the existing technology. The last decade revolution in maritime communications is a fact. The innovation, as a main reason, in maritime communications is huge and we have passed from Morse code to high bandwidth terminals that can transmit big amount of data in short time and low cost. This technological

breakthrough has inspired all maritime related manufactures to upgrade their components by adding several sensors in order to produce a very important amount of information for the operation of the vessel. So vessels today can produce a big amount of data that are very important for the proper operation of the vessel and better management which is guiding to cost reducing decisions.

This dissertation is focused on Big Data technology and how the implementation of this technology can be cost effective on the Maritime Shipping Industry. The data, its production and its importance, in Maritime Shipping industry is going to be described in depth. The goal of this dissertation is to explain thoroughly the information integration that has been applied the last decade and analyze how it affected the Maritime Shipping Industry. The future of the Maritime information integration and use of data will also be mentioned. An important issue that will also be mentioned is the Maritime communications as this has always been the key point for the data collection.

Subsequently, a detailed explanation of the need that created the Big Data technology will also be included along with what Big Data technology is all about and all the issues that we must take under consideration in order to proceed with the implementation of such technology.

Περίληψη

Η πληροφορία υπήρξε ανέκαθεν ένα από τα βασικά σημεία της διαδικασίας παραγωγής γνώσης. Όλα όσα γνωρίζουμε για τα περισσότερα πράγματα στον κόσμο μας σήμερα είναι με βάση τα αποτελέσματα της παρατήρησης και του πειραματισμού. Τα αποτελέσματα αυτά δεν είναι τίποτα περισσότερο από απλές πληροφορίες, για αυτό τον λόγο και αυτές οι διαδικασίες μπορούν να περιγραφούν ως διαδικάσιες παραγωγής πληροφοριών.

Δεδομένου ότι οι πληροφορίες είναι ένα βασικό σημείο στη διαδικασία παραγωγής γνώσης η συλλογή, αποθήκευση και επεξεργασία αυτων των πληροφορίων είναι ζωτικής σημασίας. Η συγκέντρωση όλων αυτών των πληροφοριών, σε όλα τα γνωσιακά αντικειμενα, όπως είπαμε, είναι η δημιουργία των δεδομένων. Τα δεδομένα δεν είναι τίποτα περισσότερο από μια συλλογή από μεγάλη ποσότητα πληροφοριών. Η μορφή αυτών των δεδομένων, η οποία έχει περάσει από διάφορα στάδια μέσα από τα χρόνια, καθώς και αρκετές τεχνολογικές επαναστάσεις, μας οδήγησε σήμερα στην ψηφιακή μορφή και στον ψηφιακό τρόπο αποθήκευσης. Ο μόνος τρόπος για την δημιουργία της γνώσης σε όλες τις επιστήμες, στα πρώτα χρόνια, ηταν η αναλυση και επεξεργασία αυτων των δεδομένων. Καθώς ο χρόνος περνούσε, τα δεδομένα, τα οποία εξακολουθουσαν να παράγονται σε υψηλότερα γνωσιακά επίπεδα, ξεπέρασαν τον σκοπο της ν δημιουργίας βασικών γνώσεων και έχουν εισαχθεί ευρέως σε όλα τα είδη των βιομηχανιών με την χρηση των οποιον προσπαθούν να δημιουργησουν μεγάλα οφέλη και αντιμετοποίζοντας μεγάλες προκλήσεις. Οι περισσότερες εταιρείες πάντα προσπάθουσαν να βρουν την πραγματιστικοτητα στην τρεχουσα αγορά δια μεσω τησ συλλογης δεδομένων και τη χρήση τους για ληψη ευεργετικών αποφασεων.

Η αυξημένη πολυπλοκότητα των δεδομένων δημιουργησε πολλά και διάφορα μοντέλα για την αποθήκευση, διαχείριση τους κατά τη διάρκεια των τελευταίων δεκαετιών. Έχουμε λοιπον δεδομένα που αποθηκεύονται σε απλές σειριακές βάσεις δεδομένων αλλα και στις νέες σχεσιακές βάσεις δεδομένων, για δομημένα δεδομένα, ενώ άλλα αδόμητα δεδομένα (ως προς το περιεχόμενο τους, που ήταν εικόνες, βίντεο κλπ), δεν ήταν σε θέση να αποθηκευτούν. Κατά τη διάρκεια των ετών, οι πηγές πληροφοριών έχουν επίσης αλλάξει και έχουν προστεθεί νέες, όπως δεδομένα από τα μέσα κοινωνικής δικτύωσης, τα δεδομένα ροής κειμένου από τους αισθητήρες και άλλες πηγές.

Όταν έχουμε να κάνουμε με τόσες πολλές πληροφορίες σε τόσες πολλές διαφορετικές μορφές, δεν είναι δυνατή η χρήση των παραδοσιακών τρόπων διαχείρισης δεδομένων. Το

μέγεθος των δεδομένων πάντα αυξανόταν, η διαφορά σήμερα είναι ότι έχουμε να κάνουμε με διαφορετικούς τύπους πληροφοριών.

Για την εμπορική ναυτιλιακή βιομηχανία το κύριο ζήτημα ήταν πάντα η διαβίβαση των πληροφοριών από τα πλοία λόγω της χαμηλής ταχύτητας και ταυτόχρονα του εξαιρετικά ακριβού κόστους μεταφοράς, με βάση παντα την υπάρχουσα τεχνολογία. Η επανάσταση την τελευταία δεκαετία στις θαλάσσιες επικοινωνίες είναι γεγονός. Η καινοτομία, ως κύριο λόγο, στις θαλάσσιες επικοινωνίες είναι τεράστια και έχουμε περάσει από τον κώδικα Morse σε τερματικά υψηλού εύρος ζώνης που μπορουν να μεταδώσουν μεγάλο όγκο δεδομένων σε σύντομο χρονικό διάστημα και με χαμηλό κόστος. Αυτό το τεχνολογικό επίτευγμα έχει εμπνεύσει όλες τις εταιριές που σχετίζονται με τον κλάδο ωστε να αναβάθμισουν τα προιόντα τους με την προσθήκη διαφόρων αισθητήρων, ώστε να παράγουν ένα πολύ σημαντικό ποσό πληροφοριών σχετικο με την λειτουργία του πλοίου. Έτσι, τα πλοία σήμερα μπορεί να παράγουν μια μεγάλη ποσότητα δεδομένων που είναι πολύ σημαντική για τη σωστή λειτουργία του πλοίου και την καλύτερη διαχείριση η οποία οδηγεί στην ληψη αποφάσεων μείωσης του κόστους .

Αυτή η πτυχιακή επικεντρώνεται στην τεχνολογία Big Data και πώς η εφαρμογή αυτής της τεχνολογίας μπορεί να είναι οικονομικά αποτελεσματική στην βιομηχανία της εμπορικής ναυτιλίας. Στόχος της παρούσας πτυχιακης είναι να εξηγήσει λεπτομερώς την ενσωμάτωση των πληροφοριών που έχει εφαρμοστεί την τελευταία δεκαετία και να αναλύσει πώς επηρεάζεται η βιομηχανια της εμποτικης ναυτιλίας . Ένα σημαντικό θέμα που θα πρέπει επίσης να αναφερθεί είναι οι θαλάσσιες επικοινωνίες, καθώς αυτό ήταν πάντα ενα απο τα βασικα στοιχεία για τη συλλογή δεδομένων.

Table of Contents

Chapter 1

Maritime Shipping Industry

1.1 Introduction of Maritime Shipping Industry

Maritime Shipping Industry includes companies that are focused on the shipping function and we can say that they are split in three major categories:

Ship-owner: He is an individual who owns a controlling interest in one or more ships. The ships are usually registered as one-ship companies.

Shipping Company: It is a legal organization which owns ships. It may be a legal partnership, Company or corporation in jurisdiction with enforceable laws of corporate governance, with an audited balance sheet, showing its controlling interests in the ships it operates and the status of its other assets, liabilities and bank accounts. Its executive officers are responsible for running the business and taking investment decisions.

Shipping Management Company: Companies specialized in the management of a ship. These are in depended companies that undertake the management of ships from Shipping companies and ship-owners by paying a management fee.

1.2 Shipping function of Maritime Industry

Shipping is the conveyance of goods from a place of lower utility to a place of higher utility. Goods may consist of raw materials, equipment components, and a whole range of consumer products.

Shipping is as old as the trade itself. There is evidence even from 5,000 B.C that sea trade emerged in Mesopotamia between Babylonians and Indians. In the following centuries, the world trade was concentrated almost exclusively within Mediterranean. Egyptians, Phoenicians, Greeks and Romans dominated the sea trade. In the late medieval years, new powers took over such as Venice & Genoa.

The four key factors for shipping expansion in 19th century where:

a) The development of the steam engine

b) The construction of vessels having iron hull

c) The appearance of screw propellers

d) The deep sea cable network for communications

1.3 Vessel Types

In order for shipping to be implemented several types of vessels are needed based on the type of goods that we need to transfer as you can see on diagram 1.

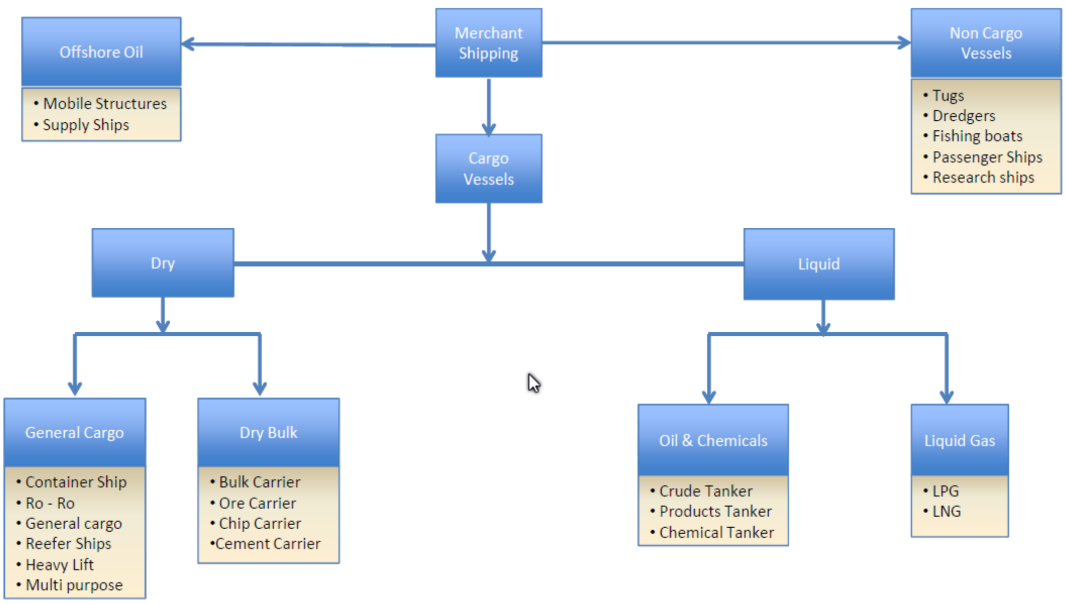


Diagram 1: Type of vessels

All those vessels have different sizes and also two main types of Tonnage measurement:

Deadweight Tonnage (dwt): Expresses the number of tons (2240 lbs) a vessel can transport of cargo, stores and bunker fuel. It is the difference between the number of tons of water a vessel displaces "light" and the number of tons of water a vessel displaces when submerged to her loadline.

Gross Tonnage: It applies to vessels and not cargo. It is determined by dividing by 100 the volume incubic feet of the vessel's closed-in spaces. It is used as a basis for pilotage and dry dock dues, and sometimes tonnage dues. So we have vessels with different sizes as you can see on Table 1.

Dry Bulk and Ore Carriers	
Large capesize bulk carrier	150,000 dwt plus
Small capesize bulk carrier	80,000 – 149,999 dwt; moulded breadth > 32.31 m
Panamax bulk carrier	55,000 – 84,999 dwt; moulded breadth < 32.31 m
Handymax bulk carrier	35,000 – 54,999 dwt
Handysize bulk carrier	10,000 – 34,999 dwt
Post –Panamax container ship	moulded breadth > 32.31 m
Panamax container ship	moulded breadth < 32.31 m

Table 1: Vessel's sizes

Crude Oil Tankers	
ULCC	350,000 dwt plus
VLCC	200,000 – 349,999 dwt
Suezmax crude tanker	125,000 – 199,999 dwt
Aframax crude tanker	80,000 – 124,999 dwt; moulded breadth > 32.31m
Panamax crude tanker	50,000 - 79,999 dwt; moulded breadth < 32.31m

Table 1: Vessel's sizes

1.4 Shipping Companies economy

The income of the Shipping companies always depends on the freight rates combined with the operation cost of the company (including the operation cost of the vessels). The freight rates are based on supply and demand economic model of price determination in a market. In conclusion, in a competitive market, the unit price for a particular good will vary until it settles at a point where the quantity demanded by consumers (at current price) will equal the quantity supplied by producers (at current price), resulting in an economic equilibrium of price and quantity.

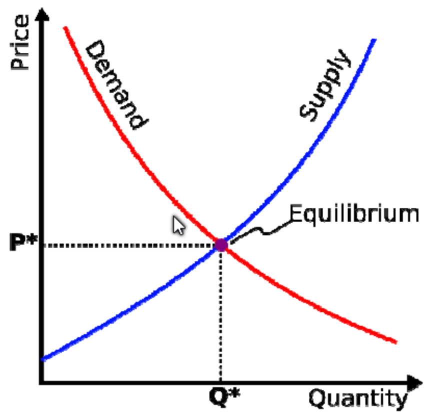


Diagram 3 : Supply and Demand

There are four basic laws for Supply & Demand:

1. If demand increases and supply remains unchanged, a shortage altogether, thus leads to a higher equilibrium price.
2. If demand decreases and supply remains unchanged, a surplus altogether, thus leads to a lower equilibrium price.
3. If demand remains unchanged and supply increases, a surplus altogether, thus leads to a lower equilibrium price.
4. If demand remains unchanged and supply decreases, a shortage altogether, thus leads to a higher equilibrium price.

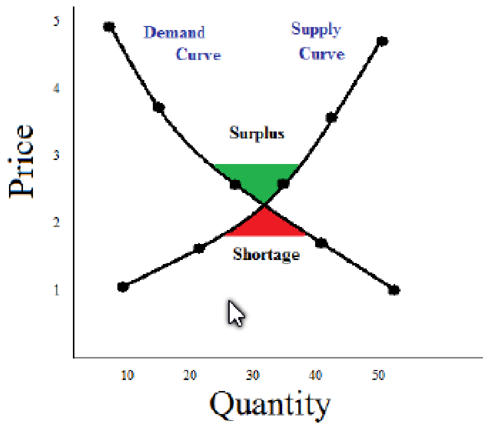


Diagram 4 : Price and Quantity

Freight contracts cannot be made directly with the ship owner but via a third person that is called charterer or managing company. Chartering or managing companies have to deal with shipbrokers in order to proceed with the proper vessel.

1.4.1 Shipbroking

Shipbroking is a financial service, which forms part of the global shipping industry. Shipbrokers are specialist intermediaries/negotiators (i.e. brokers) between shipowners and charterers who use ships to transport cargo, or between buyers and sellers of ships.

Shipbrokerage has been developed and large companies are specialized in various sectors, e.g. Dry Cargo Chartering, Tanker Chartering, Container Chartering, Sale & Purchase, Demolition and Research.

Shipbroking can be categorized as follows:

i. Sale and purchase

S&P brokers handle the buying and selling of existing vessels in the secondary market or new ships (called newbuildings, in industry parlance). S&P brokers discuss opportunities and market trends with shipowners, charterers, investors and bankers, they also report on market sales, vessel values, market trends and activity. When a shipowner has a vessel for sell or is looking for a vessel to acquire, the shipbroker will source the market for suitable tonnage, discuss with the counterparty or their broker the main points of a transaction and they eventually negotiate all the details of the transaction in discuss. During the negotiations, the shipbrokers do not only negotiate the price of the vessel on behalf of the principals but also all the logistical details for the transfer of the title and the vessel herself to the buyers. During the negotiations minor disputes may have to be handled given that the market may be moving in favor of the buyer (vessel price is softening) or in favor of the seller (vessel price is strengthening) giving each party a reason to cancel the transaction. When shipbrokers act on behalf of passive investors or financial buyers, they may also have to advise and provide employment for the vessels, a vessel manager and oversee such agreements. Some S&P brokers specialize in the sale of ships for scrapping, which requires a different set of skills.

ii. Dry cargo broking

Dry cargo brokers are typically specialists in the chartering of Bulk carriers, and are appointed to act either on behalf of a shipowner looking for employment for a ship, or a charterer with a cargo to be shipped. Dry cargo chartering brokers have to maintain large databases of vessel positions, cargoes and rates and pay close attention to the direction of the markets in order to advise their clients accurately on how to maximize profits or minimize expenses. Dry cargo shipping can be categorized by Vessel size: namely, Bulkers such as Capesizes, Panamaxes and Handysize are the main sectors. Each size of vessel suits different types of cargo and trade routes. Thus owners, charterers and brokers tend to specialize on one or the other of these sectors.

iii. Tanker broking

Tanker brokers specialize on the chartering of Tankers, which requires different skills and knowledge from Dry cargo broking. Tanker brokers may specialize in crude oil, gas, oil products or chemical tankers. Tanker brokers negotiate maritime contracts, known as Charter Parties. The main terms of negotiation are freight/hire and demurrage. Due to oil being a fast moving trade, freight rates for crude oil tanker charters are most commonly based on the Worldscale Index; the Worldscale Association publishes flat rates annually. Some specific voyages, such as named voyages (i.e., from A to B) and for specialist ships, like LNG tankers (a highly specialized sector of the tanker market), freight rates can be agreed at a fixed rate between the parties.

iv. Container broking

Container brokers specialize on the chartering of container ships and provide container ship owners and charterers with market-related information.

1.4.2 Chartering

Chartering is the main activity in shipping industry and include all necessary actions in order for goods to be transferred. Charterer’s main concern is to contact end customers and make the necessary arrangements for the shipping, sometimes he may own a cargo and employ a shipbroker to find a ship to deliver the cargo for a certain price, called freight rate.

Freight rates may be on a per-ton basis over a certain route (e.g. for iron ore between two specific locations), in Worldscale points (in case of oil tankers) or may alternatively be expressed in terms of a total sum - normally in U.S. dollars - per day for the agreed duration of the charter.

A charterer may also be a party without a cargo who takes a vessel on charter for a specified period from the owner and then trades the ship to carry cargoes at a profit above the hire rate, or even makes a profit in a rising market by re-letting the ship out to other charterers.

Charter types:

i. Voyage charter is the hiring of a vessel and crew for a voyage between a load port and a discharge port. The charterer pays the vessel owner on a per-ton or lump-sum basis. The owner pays the port cost (excluding stevedoring), fuel cost and crew cost. The payment for the use of the vessel is known as freight. A voyage charter specifies a period, known as laytime, for loading and unloading the cargo. If laytime is exceeded, the charterer must pay demurrage. If laytime is saved, the charter party may require the shipowner to pay despatch to the charterer.

ii. Contract of Affreightment is a contract similar to a voyage charter, but ship-owner undertakes to carry a number of cargoes within a specified period of time on a specified route. Agreed frequency of cargoes may require more than one ship.

iii. Time charter is the hiring of a vessel for a specific period of time; the owner still manages the vessel but the charterer selects the ports and directs the vessel. The charterer pays for all the fuel consumed by the vessel, port charges, commissions, and a daily hire to the owner of the vessel.

iv. Trip time charter is a comparatively short time charter agreed for a specified route only (as opposed to the standard time charter where charterer is free to employ the vessel within agreed trading areas).

v. Bareboat charter or demise charter is an arrangement for the hiring of a vessel whereby no administration or technical maintenance is included as part of the agreement. The charterer obtains possession and full control of the vessel along with the legal and financial responsibility for it. The charterer pays for all operating expenses, including fuel, crew, port expenses and P&I and hull insurance. In commercial demise chartering, the charter period may last for many years; and may end with the charterer acquiring title (ownership) of the ship. In this case, a demise charter is a form of hire-purchase from the owners, who may well have been the shipbuilders. Demise chartering is common for tankers and bulk-carriers.

Chapter 2

Maritime communications

Communication was one of the fundamental challenges since the dawn of life. A lot of sciences have made several efforts to implement it in all fields and major innovations have guided communication to today's level. Communication in Maritime Industry was always vital for the ship's safety and survival. As far as small distances are concerned solutions were easy and were accomplished via various ways. The real problem was to overcome the long distance communication.

In order to overcome the communication issue with long distances they had to be a bit more creative. The ideas of Heinrich Hertz, in the late 1800s, on electromagnetic radiation have been used by Royal Navy and scientists such as Guglielmo Marconi worked so as to use radio waves as a method of communication between ships and ship-to-shore in Morse code.

In the mid-1800s, James Clerk Maxwell, a well-respected physicist and mathematician developed his theory on electromagnetism. He showed that light waves consisted of oscillating, perpendicular electric and magnetic fields.

By converting sound waves into radio waves, ships could send information across vast stretches of ocean to communicate with their allies. They would then decode the radio waves back into sound waves, and listen to the message they had received. By the early 20th Century, the Royal Navy had started to adopt transmitters fitted with alternators and used magnetic detectors with headphones.

After the discovery of microwaves, the U.S. Navy realized that they could bounce these waves off the surface of the Moon, the Earth's natural satellite. This would allow them to send information to the other side of the world with much less interference. After the USSR sent Sputnik, the first artificial satellite to contain a radio transmitter into space in 1957, the US followed soon after. This "space-race" led to more efficient satellite communication systems. Since then, technology has advanced at an impressive rate, improving the way in which ships can communicate with each other.

2.1 Satellite Communications

There has been a huge amount of different communication types for the maritime industry but the main focus of this dissertation will be on the satellite that eventually overcame most of the challenges and today has become the point of reference.

Throughout the years there have been several satellite implementations some of them were successful and some of them had a sort period of life. One of the more successful implementations in Maritime Industry is by a company named Inmarsat PLC.

Inmarsat is a British satellite telecommunications company, offering for several years now satellite communication services to the Maritime Industry. Started with global and limited coverage services, analog, slow at the beginning and during the last two decades global, digital and faster services. It provides telephone, fax and data services to ships worldwide via mobile terminals, installed on board of the ships, communicating to ground stations through geostationary telecommunications satellites. Inmarsat's network except maritime industry provides communications services to a range of governments, aid agencies, media outlets and businesses in need of communication in remote regions or where there is no reliable terrestrial network.

The company was originally founded in 1979 as the International Maritime Satellite Organization (Inmarsat), a non-profit international organization, set up at the behest of the International Maritime Organization (IMO), a UN body, for the purpose of establishing a satellite communications network for the maritime community.

The commercial part was initiated in 1982 with the acronym "Inmarsat". The intent was to create a self-financing body which would improve safety of life at sea. The name was changed to "International Mobile Satellite Organization" when it began to provide services to aircraft and portable users, but the acronym "Inmarsat" was kept. When the organization was converted into a private company in 1999, the business was split into two parts: The bulk of the organization was transferred to the commercial company, Inmarsat plc, and a small group became the regulatory body, IMSO.

The International Mobile Satellite Organization (IMSO) is the intergovernmental organization that oversees certain public satellite safety and security communication services provided via the Inmarsat satellites. Some of these services concern:

- Global Maritime Distress Safety System (GMDSS) established by the International Maritime Organization (IMO)

- Search and rescue coordinating communications
- Maritime safety information (MSI) broadcasts
- Aeronautical mobile satellite (route) service, or AMS(R)S, in compliance with the Standards and Recommended Practices (SARPs) established by the International Civil Aviation Organization (ICAO)
- General communications

Based on the above Inmarsat provides global maritime distress and safety services (GMDSS) to ships and aircraft at no charge, as a public service.

Inmarsat first try for commercial maritime satellite services was via the Inmarsat A terminal. Inmarsat-A is now 20 years old. Yet it has operated excellently and actually benefited from improved and newer technologies. This means that the infrastructure was capable of fulfilling modern needs totally into the 21 st century.

2.2 Satellite Coverage

In satellite communications technology the challenge of coverage was a little bit easier than other technologies, technically, but the cost was extremely high for both the service provider and the manufacturers of the communication terminals. So coverage focused on the global coverage from the service provider. The terminal manufacturers created a variety of terminals with global and not coverage ability. This was related with the cost of the production and also combined with the cost of the airtime from the service provider. It has been noticed the combination of low cost terminals and low airtime rates has helped a lot the penetration of the satellite communications to the maritime Industry.

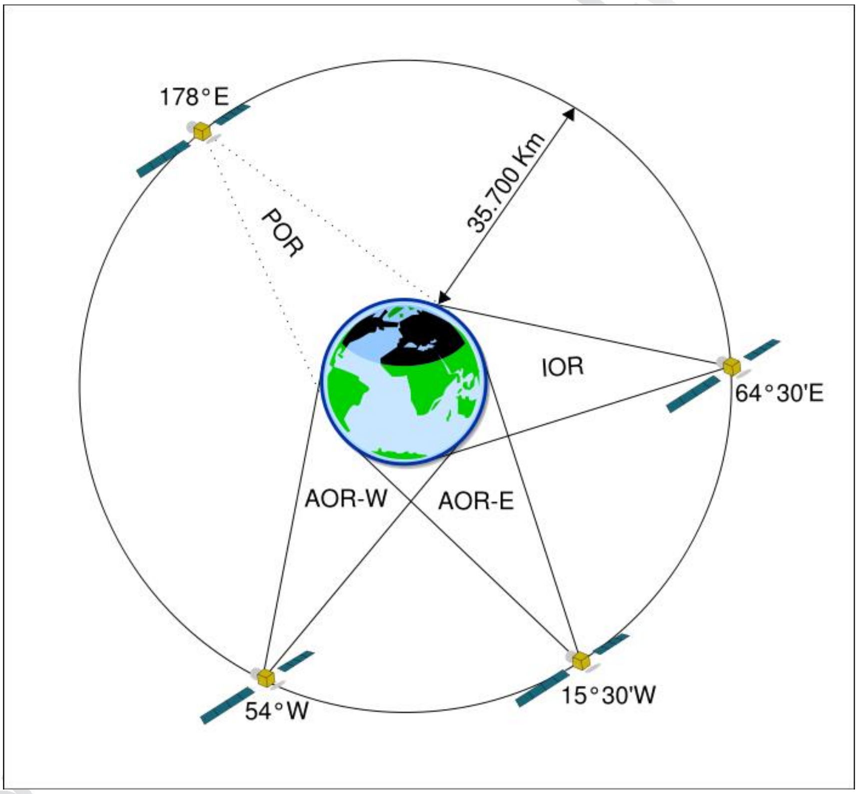
For terminals with world coverage we have some technical limitation and by global we mean almost all areas except the areas of earth poles. This limitation is related with the position of the satellites and the reception angle that makes impossible to have signal near the poles. Global satellite coverage consists of global beams. Beam is a specific

Coverage that was not global was name spot beam coverage. Spot beam coverage includes specific areas that are covered from specific satellites. In the following pages you can find detailed pictures of coverage from all satellites and from each one separately. The separate view is very important to be mentioned as there are several areas close to a satellite you may assume that are covered from but they are covered from another satellite.

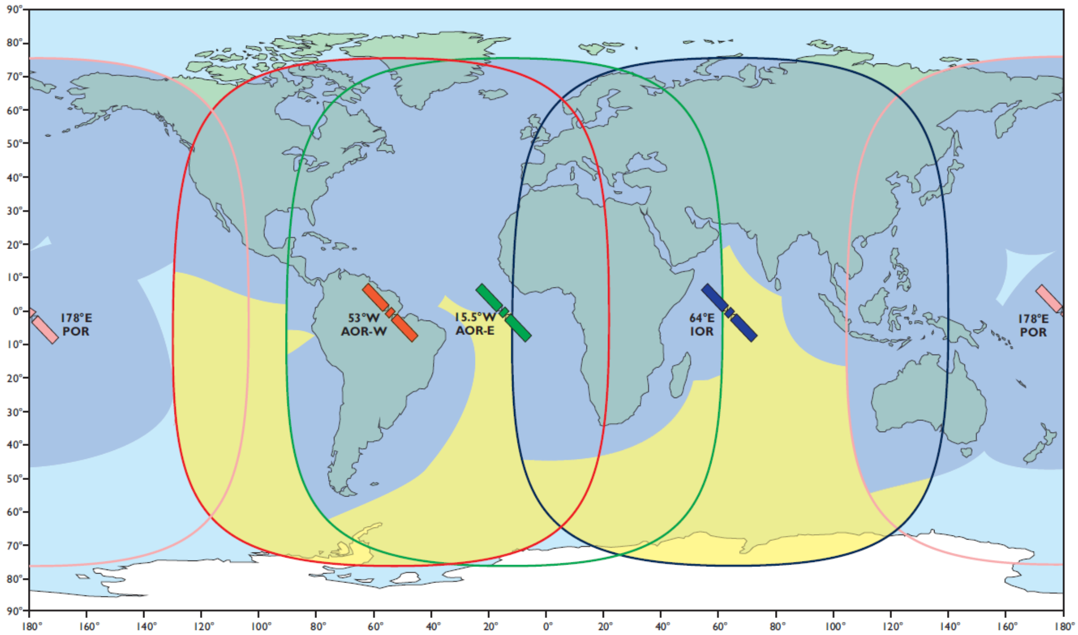
It's also important to mention that world coverage terminal as also using the spot beams

Inmarsat as a maritime communications operator was the first with wholly-owned constellations – the Inmarsat-2 and Inmarsat-3 series – are located above the world's sea-lanes form four ocean regions:

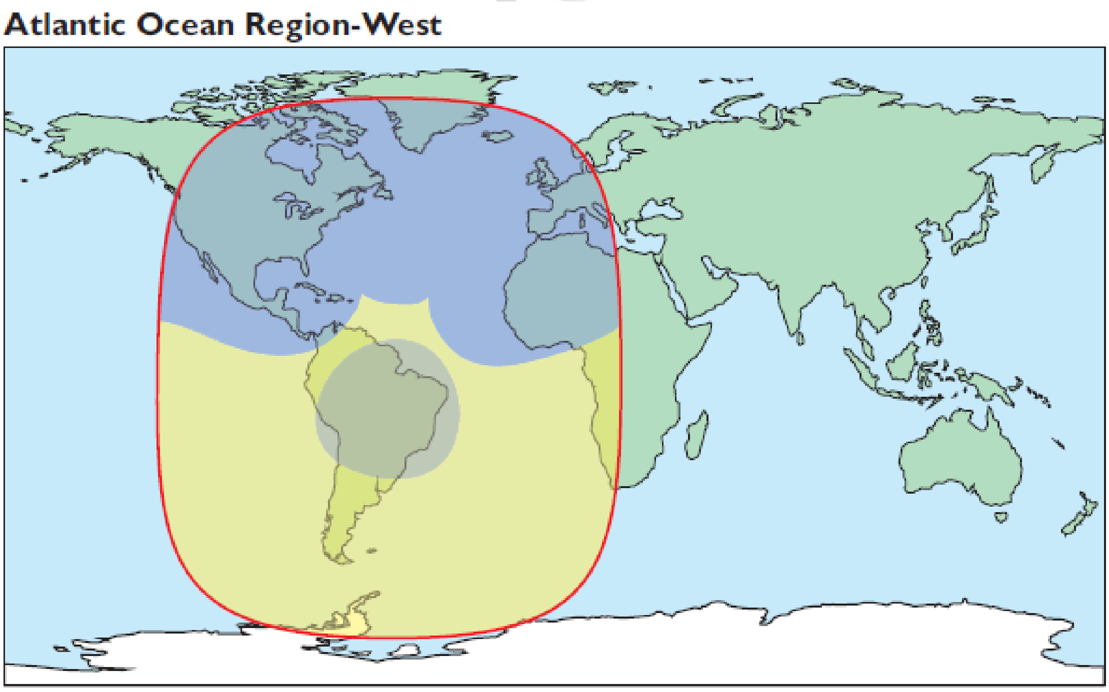
- Atlantic Ocean Region West (AOR-W), at 54 degrees West
- Atlantic Ocean Region East (AOR-E), at 15.5 degrees West
- Indian Ocean Region (IOR), at 64 degrees East
- Pacific Ocean Region (POR), at 178 degrees East.



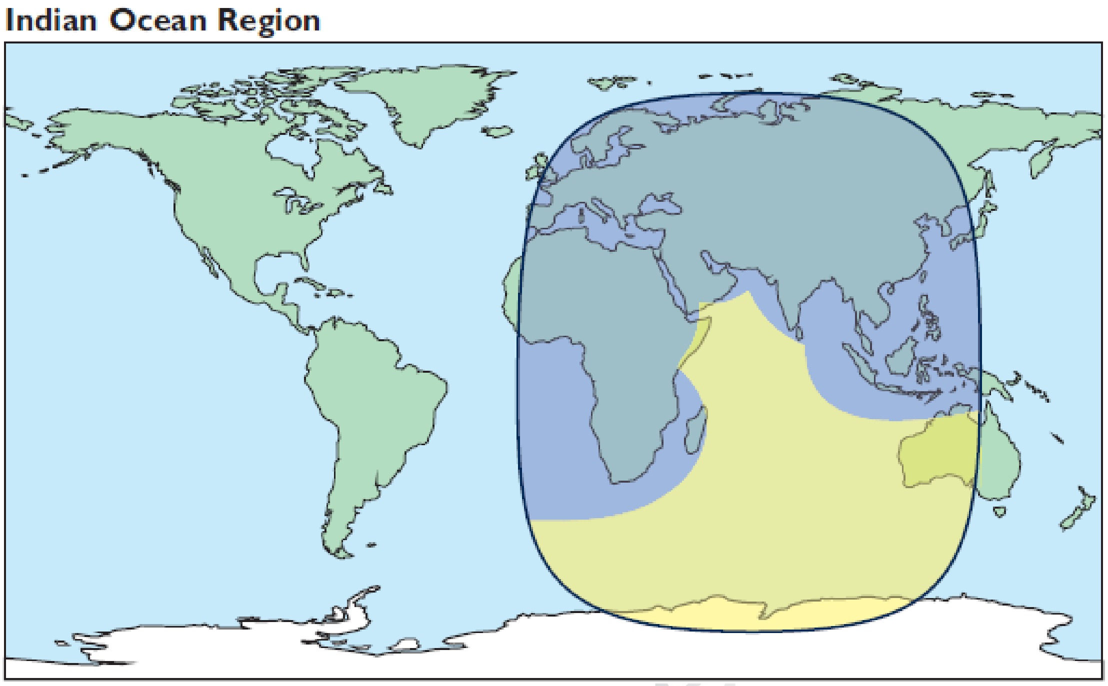
Four Geostatic Satellites



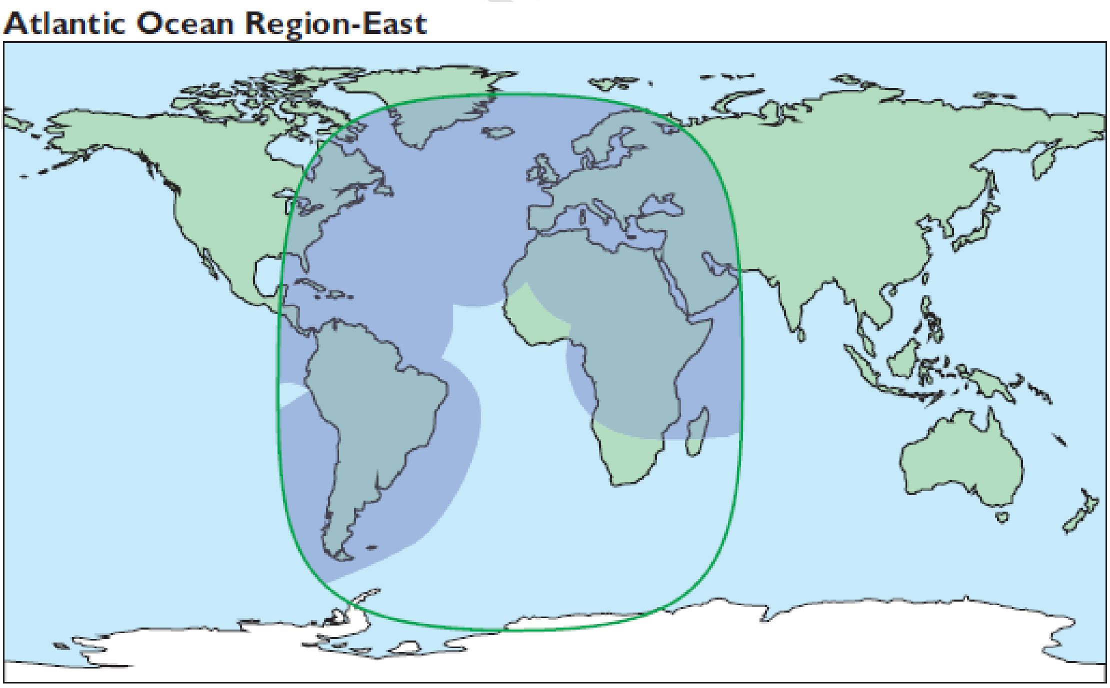
Spot beam Maritime I-3 satellite coverage



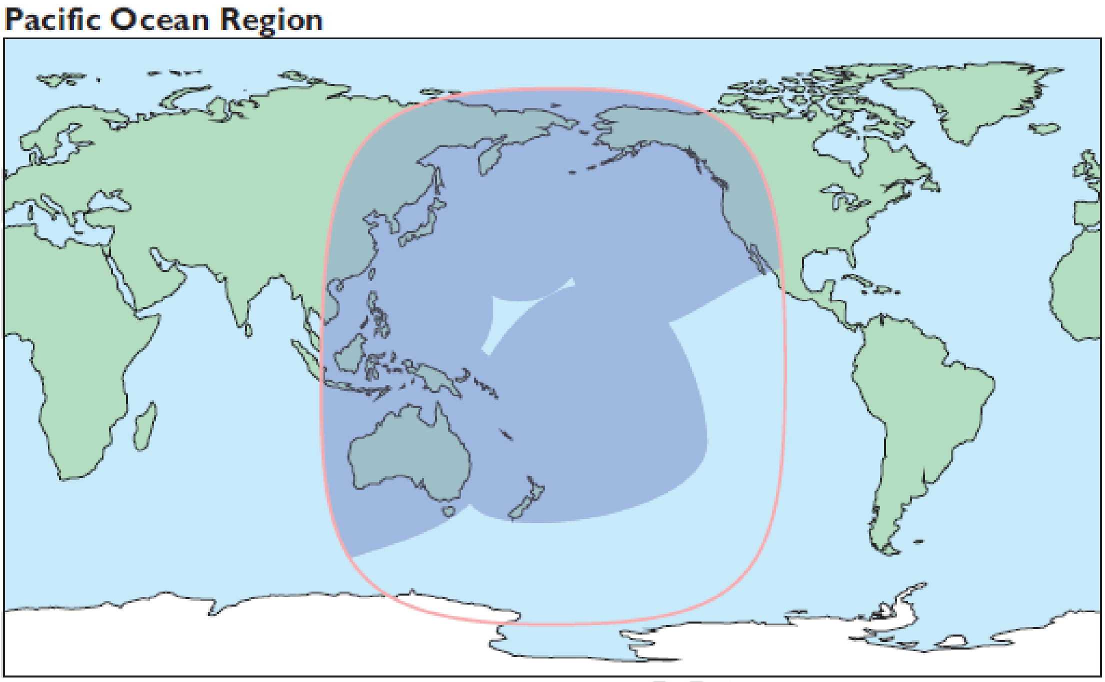
Spot beam Maritime I-3 AORW satellite coverage



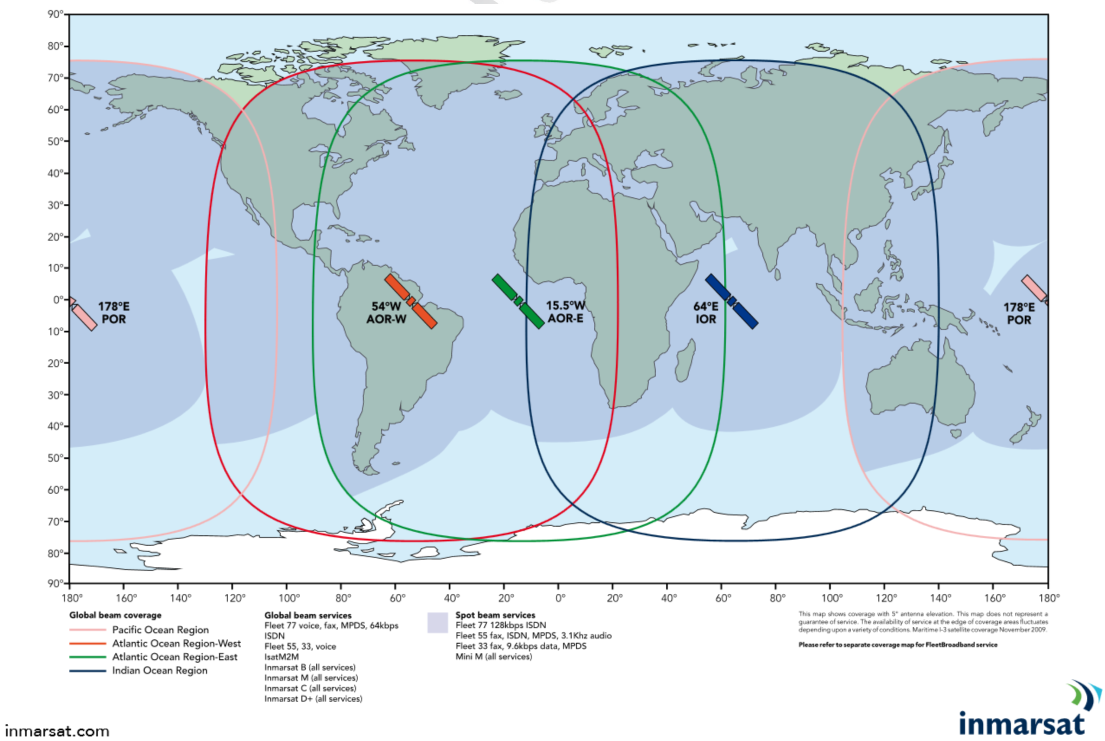
Spot beam Maritime I-3 IOR satellite coverage



Spot beam Maritime I-3 AORE satellite coverage



Spot beam Maritime I-3 POR satellite coverage



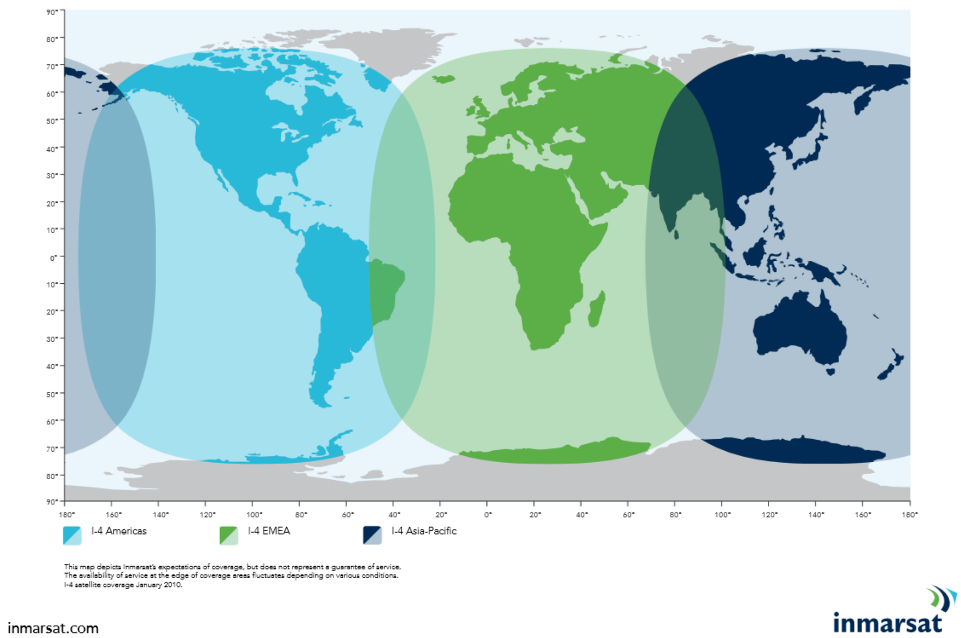
Global Maritime I-3 satellite coverage

Following the full deployment of the Inmarsat-4 (I-4) satellite constellation in 2009, a further three satellite regions were created over the Earth's major landmasses. They are:

I-4 Americas, at 98 degrees West

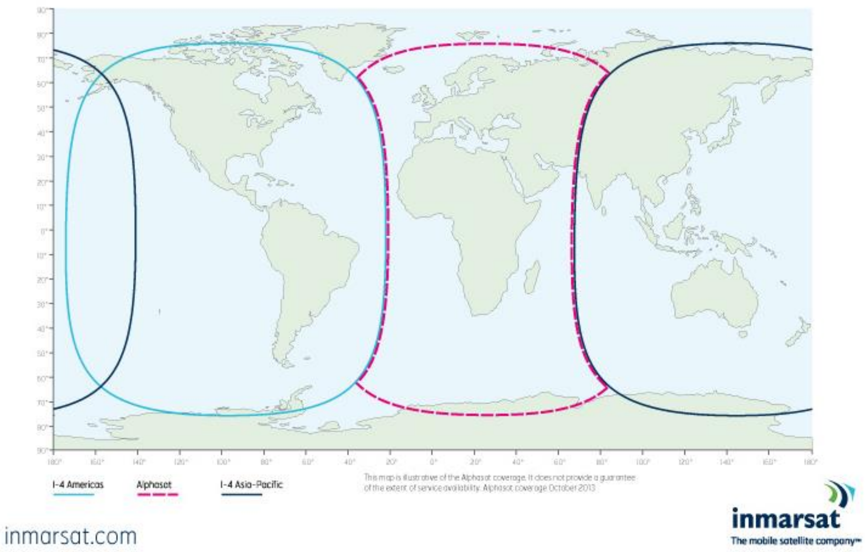
I-4 EMEA (Europe, Middle East and Africa), at 25 degrees East

I-4 Asia-Pacific, at 143.5 degrees East.



I-4 satellite coverage

From November 2013, I-4 EMEA is covered by Alphasat (I-4A F4) with a new footprint, as illustrated in the map below.



Alphasat Coverage

2.3 Inmarsat Satellite Terminals

The principals that Inmarsat-A mobile Satellite communications was based are still the same in most todays' satellite terminals so it’s important to proceed with a small analysis on Inmarsat A. (2, n.d.)

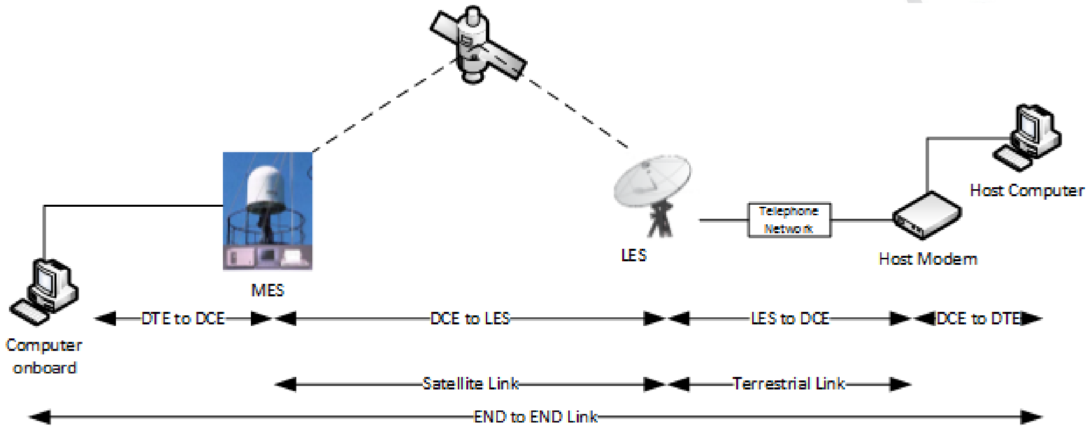
Inmarsat A system provided two-way direct-dial phone (high quality voice), fax, telex, electronic mail and data communications to and from anywhere in the world with the exception of the poles. It also provided distress communication capabilities. It was based upon analogue technology. It support data rates of between 9,600 bps through up to 64,000 bps depending upon different elements of your end-to-end connection.

The Inmarsat-A service was comprised by 3 components:

The mobile-earth station (MES): An Inmarsat-A terminal was a small, self-contained satellite earth station comprising a lightweight parabolic antenna, electronic units, power supply interface, and direct-dial telephone, fax and telex connections.

The satellites: The transmission and reception of signals are co-coordinated by four network co-ordination stations (NCS), one for each satellite coverage region - Atlantic Ocean East and West, Indian Ocean and Pacific Ocean.

The Land-Earth station (LES): A call from a mobile or transportable Inmarsat-A terminal is routed via the Inmarsat satellite system to a land earth station (LES) for connection to the national and international phone and data networks.



Terminals receive and transmit in the L-band (1.5/1.6GHz). Each NCS monitors the flow of communications traffic through its satellite to ensure that calls are set up correctly and that all LES's are working correctly.

The maritime terminals typically consist of the above-deck equipment comprising the stabilized antenna (enabling it to stay locked onto the satellite even in heavy sea conditions) and the below-deck equipment comprising the telephone, telex, fax and data interfaces.

Systems that followed Inmarsat A and most of them as operational till today are:

2.3.1 Inmarsat-B

This was the next terminal trying to replace Inmarsat A providing high quality voice services at 16kbits/s, telex services, medium speed fax/data services at 9.6 Kbit/s and high speed data services at 56 (for USA), 64 Kbit/s. (1, n.d.)

Inmarsat B terminal

2.3.2 Inmarsat-C

One of the mandatory terminals that all vessels must have today and based on safety regulation two of them must be on up and running in order for the vessel to sail. This is a "satellite telex" terminal with store-and-forward, polling etc. capabilities. Certain models of Inmarsat-C terminals are also approved for usage in the MDSS system, equipped with GPS. Via this terminal the Master of the vessel can communicate with plain text messages that can be transformed to emails and faxes also except of telex messages. (1, n.d.)

Inmarsat C terminal

2.3.3 Inmarsat D/D+/IsatM2M

Inmarsat's version of a pager, although much larger than terrestrial versions. Some units are equipped with GPS. The original Inmarsat-D terminals were one-way (to mobile) pagers. The newer Inmarsat-D+ terminals are the equivalent of a two-way pager. The main use of this technology nowadays is in tracking. (1, n.d.)

Inmarsat D Antenna

2.3.4 Inmarsat-M

Provides voice services at 4.8 Kbit/s and medium speed fax/data services at 2.4 Kbit/s. It paved the way towards Inmarsat-Mini-M. (1, n.d.)

2.3.5 Mini-M

Provides voice services at 4.8 Kbit/s and medium speed fax/data services at 2.4 Kbit/s. One 2.4kbit/s channel takes up 4.8kbit/s on the satellite. The most popular terminal in Maritime Communications that is till today operational. Mini-m was a small revolution as it was the first terminal with small dimensions and with low cost. The main purpose of this terminal was to provide low cost voice service. (1, n.d.)

Mini-M terminal

2.3.6 Fleet 33/55/77

Actually a family of networks that includes the Inmarsat-Fleet77, Inmarsat-Fleet55 and Inmarsat-Fleet33 members (The numbers 77, 55 and 33 come from the diameter of the antenna in centimeters). It provides a selection of low speed services like voice at 4.8 Kbit/s, fax/data at 2.4 Kbit/s, medium speed services like fax/data at 9.6 Kbit/s, ISDN like services at 64 Kbit/s (called Mobile ISDN) and shared-channel IP packet-switched data services at 64 Kbit/s (called Mobile Packet Data Service MPDS). However, not all these services are available with all members of the family. The latest service to be supported is Mobile ISDN at 128 Kbit/s on Inmarsat-Fleet77 terminals. MPDS (Mobile Packet Data Service): Previously known as IPDS, this is an IP-based data service in which several users share a 64kbit/s carrier in a manner similar to ADSL. (1, n.d.)

Fleet 33 Terminal

Fleet 55 Terminal

Fleet 77 terminal

2.3.7 Fleet Broadband

Since now all satellite terminals were able to provide one service per use, meaning that the user was not able to make a data connection and at the same time to make a voice call. Fleet broadband is the first terminal that is able to proceed with data and voice service at the same time. This new evolution allows the user to be always data active/connected and at the same time to use the voice service. Now days there is a new feature applied named multi-voice that provides the ability to the user to be able to have multiple voice calls.active. This generation of terminals can be described as the "Maritime ADSL" as the speeds of the terminals are from 150kbit/sec to 432kbit/sec which is in the range of the first land ADSL connections.

FleetBroadband features:

Global coverage
Fleet Broadband is accessible globally, except in the extreme polar regions

SMS
Send and receive SMS text messages to and from other Fleet Broadband terminals and land-based cellular networks through your PC.

Enhanced voice services
Voicemail and other enhanced call management options are available, including caller identification, caller barring and call forwarding.

Emergency calling

Our 505 Emergency Calling facility puts vessels directly in contact with a maritime rescue co-ordination centre (MRCC) free of charge.

General applications

- ECDIS
- FTP
- GPS location data look-up-and-send
- IP SCADA
- ISDN
- SMS
- Crew Welfare – Chat
- Crew welfare prepaid / postpaid
- Database sychronisation
- Email
- File transfer
- Handheld support
- Hotspots
- Internet access
- Intranet access
- Multi-voice
- Notice To Mariners
- Remote access
- Remote monitoring and maintenance
- Safety 505
- Safety red button
- Secure VPN access
- Shore-to-ship connectivity
- Store and forward video
- Surveillance
- Telemetry
- Telephony VoIP
- Telephony circuit
- Time-critical data transfer

- Vessel routeing
- Video conferencing
- Weather updates

Bespoke applications

- GSM pico cell
- IT security
- POS / ATM access onboard
- X-Stream live video
- Anti-piracy / citadel deployment
- Asset tracking
- Banking / online shopping
- Cargo management
- Crew management
- E-logs
- Encryption
- Fish catching applications
- Geospatial intelligence
- Government / regulatory / control
- Intranet and secure integration
- Live broadcast
- Live (high-data-rate) onboard-to-shore telemetry (text streaming)
- Maritime / Port Regulations
- Multi-WAN support and integration
- Online collaboration
- Optimised social solutions
- Personnel security
- Planned / predictive maintenance
- Remote access / control / management / support
- Remote manned video
- Remote / security video surveillance
- Remote telemechanical engineering
- Remote unmanned video
- Research and operations onboard

- Situational awareness
- Telemedicine
- Telemetry – SCADA
- Tele-training / certification
- Traditional Fax over IP
- Voyage management
- Walkie-talkie (ROIP) VHF/UHF radio integration
- Weather forecasts

Fleet Broadband 150 (1, n.d.)

Standard IP data

Always-on IP data service up to 150kbps for office applications such as email and internet access, real-time electronic charts and weather reporting.

Satellite telephony

Access up to three additional telephone lines from a single FB150 terminal with Fleet Broadband Multi-voice. Make clear phone calls to terrestrial and mobile networks, as well as other satellite terminals, anytime.

Enhanced voice services

Voicemail and other enhanced call management options are available, including caller identification, caller barring and call forwarding.

Fax

FleetBroadband 150 provides a circuit-switched data connection for legacy applications such as Group 3 fax.

Sailor Fleet Broadband 150

Intellian Fleet Broadband 150

Satlink Fleet Broadband 150

<u>*Fleet Broadband 250*</u> (1, n.d.)

Standard IP data

An always-on connection at up to 284kbps for applications such as email and internet access, real-time electronic charts and weather reporting.

Satellite telephony

Access up to nine telephone lines from a single FB250 terminal with FleetBroadband Multi-voice, and make crystal-clear phone calls to terrestrial and mobile networks, as well as other satellite terminals, anytime.

Streaming IP

Guaranteed rates up to 128kbps are available on demand for live applications such as video conferencing and database synchronisation.

Emergency calling

Our 505 Emergency Calling facility puts vessels directly in contact with a maritime rescue co-ordination centre (MRCC) free of charge.

Fax

FleetBroadband 250 provides a circuit-switched data connection for legacy applications such as Group 3 and Group 4 fax.

Sailor Fleet Broadband 250

Satlink Fleet Broadband 250

Fleet Broadband 500 (1, n.d.)

Standard IP data

An always-on connection at up to 432kbps for applications such as email and internet access, real-time electronic charts and weather reporting.

Satellite telephony

Access up to nine telephone lines with FleetBroadband Multi-voice and make crystal-clear phone calls to terrestrial and mobile networks, as well as other satellite terminals, anytime.

Streaming IP

Guaranteed connection rates of up to 256kbps available on demand, for live applications such as video conferencing and database synchronisation.

Fax

FleetBroadband 500 provides a circuit-switched data connection for legacy applications such as Group 3 and Group 4 fax.

Sailor Fleet Broadband 500

JRC Fleet Broadband 500

Furuno Fleet Broadband 500

2.3.8 XpressLink

This terminal integrates the industry leading FleetBroadband service with VSAT service to give a solution with increased performance and flexibility – and unlimited data.

You can enjoy seamless global coverage and reliable FleetBroadband connectivity, as well as always-on data speeds of up to 768Kbps and a Committed Information Rate (CIR) of 192Kbps.

High-speed broadband

Access always-available IP data at up to 768kbps, with a committed information rate (CIR) of 192kbps.

Voice services

Crystal-clear phone calls to terrestrial and mobile networks as well as other satellite terminals are provided. Voicemail and other enhanced call management options are available, including caller identification, caller barring and call forwarding.

Integrated solution

Xpresslink is a fully-integrated Ku-band and L-band solution with VSAT and FleetBroadband terminals (FB500 or FB250)*. It offers unlimited data for a fixed monthly fee, which includes FleetBroadband data usage.

Fully-redundant service

With VSAT and FleetBroadband terminals operating on separate networks, XpressLink is a fully-redundant and resilient service.

Seamless mobility

Includes an intelligent below-deck control system, the XpressLink G5 unit, that automatically manages switching between Ku-band and L-band and selects the optimum available Ku-band coverage area. In addition, the system manages and deploys updates to installed satellite equipment.

Invoice management

XpressLink includes a secure customer portal that allows ship management to track invoices for voice usage and compare invoices across vessels in the fleet.

Vessel tracker

Keep track of vessels at all times through a vessel tracker interface, using Google Earth maps. The vessel tracker provides real-time access to information on vessel locations anywhere in the world. It helps reduce manual reporting and inaccuracies and increases efficiency.

Crew calling

Available as a post-paid or pre-paid service. With pre-paid, crew can make low-cost calls using calling cards or codes ordered and allocated through a secure customer portal.

Global support

In addition to 24/7 global support, a maintenance network is located in key ports around the world. Request maintenance and product upgrades easily by submitting request forms via a secure customer portal onboard or ashore. (1, n.d.)

General applications

- Antivirus
- Bandwidth reports
- Black/white listing
- Category and application filtering
- Crew services
- Data history reporting
- Firewall
- Invoice management
- Maintenance and product upgrades orders

- Order management
- Threat management
- Vessel tracker
- Web reputation

Intellian XpressLink Terminal

2.3.9 Global Xpress

Global Xpress (GX) is the next generation technology in maritime communications. This new technology will be available for use in 2015. Global Xpress operates in the Ka-band, which is already more efficient in its use of bandwidth to provide higher bandwidth on a consistent, end-to-end global basis.

Inmarsat GX will deliver higher performance consistently – with a downlink up to 50Mbps and an uplink to 5Mbps. Inmarsat GX will enable seamless, real-time operations between vessels and land offices, plus enhanced communications and entertainment for the crew.

Besides more bandwidth, GX will give you exclusive access to specialist maritime applications and content. Download the latest chart update, engine monitoring and weather-routing applications, all from a centralized source. And distribute real-time to all the vessels in the fleet simultaneously.

GX terminals will be lower cost than those currently available for Ku-band, and services will be available for a fixed monthly fee.

This technology and this generation of terminals will allow the big data implementation for the maritime shipping industry that will be analyzed.

Intellian GX terminal

Chapter 3

Big Data

3.1 Defining Big Data

One of the greatest challenges for organizations has always been the managing and analyzing of data. Data management has passed several phases during the years and each phase tried to solve a specific type of data management problem.

In the late 1960's data was stored in flat files with no structure. The needs upon that time was only to store the information and use it without applying any analysis. When companies needed to get some information in order to understand more regarding their customers they had to apply new methods based on new programming models. This new need has caused the creation of relational data model and relational database management system in the 1970's. The relational model added a level of abstraction so that it was easier for programmers to satisfy the business demands to extract the required value from data.

The relational model offered tools that helped companies to better organize their data and be able to compare their transactions. It also helped business managers who wanted to be able to examine information and compare it for decision-making purposes. But a problem emerged from this exploding demand for answers: Storing this growing volume of data was expensive and accessing it was slow. Making matters worse, lots of data duplication existed, and the actual business value of that data was hard to measure.

The Entity-Relationship (ER) model was the solution, which added additional abstraction to increase the usability of the data. In this model, each item was defined independently of its use. Therefore, developers could create new relationships between data sources without complex programming. It was a huge advance at the time, and it enabled developers to create more complex models requiring complex techniques for joining entities together. It is especially important for transactional data management of highly structured data.

When the volume of data that was needed to manage from the companies grew out of control, a solution was provide from the creation of the data warehouses. The data warehouse was intended to help companies deal with increasingly large amounts of structured data that they needed to be able to analyze by reducing the volume of the data to something smaller and more focused on a particular area of the business. It filled the need to separate operational decision support processing and decision support — for performance reasons. In addition, warehouses often store data from prior years for understanding organizational performance, identifying trends, and helping to expose patterns of behavior. It also provided an integrated source of information from across various data sources that could be used for analysis. Data warehouses were commercialized in the 1990s, and today, both content management systems and data warehouses are able to take advantage of improvements in scalability of hardware, virtualization technologies, and the ability to create integrated hardware and software systems, also known as appliances. Data warehouses due to their structure they were too complex and was not able to offer the speed and the agility that companies required. As always a new solution was needed and this was called data marts. The data marts were focused on specific business issues and were much more streamlined and supported the business need for speedy queries than the more massive data warehouses.

Data warehouse was updated with new data on a daily or weekly bases that was fine for planning, financial reporting, and traditional marketing campaigns, but is too slow for increasingly real-time business and consumer environments. The problem was not so easy this time and took a long time to be solved as companies began to store unstructured data. An unstructured data element would be stored in a relational database as one contiguous chunk of data. This object could be labeled (that is, a customer inquiry) but you couldn't see what was inside that object.

The object database management system (ODBMS) was the solution. The object database stored the BLOB (binary large objects) as an addressable set of pieces so that we could see what was in there. Unlike the BLOB, which was an independent unit appended to a traditional relational database, the object database provided a unified approach for dealing with unstructured data. Object databases include a programming language and a structure for the data elements so that it is easier to manipulate various data objects without programming and complex joins. The storing and managing of unstructured data has moved ahead with the rise of the web in the 1990s. The data were consisted of web content, images, audio, and video.

Here is the point that big data technologies have been initiated. A definition of big data can be described as the tool to manage huge volume of different type of data, at an acceptable speed meaning within the right time frame in order to allow real-time analysis and decision making.

Big Data big data is typically broken down by four characteristics:

- Variety – Includes unstructured data of all varieties: text, audio, video, click streams. Log files and more
- Velocity – Frequently time-sensitive, how fast the data is processed
- Volume – Huge amount of data in size or terabytes and even petabytes
- Veracity – To screen out spam and other data that is not useful for making business decisions (5)

Before we start to analyze the data in order to solve the addressed problem, first data must be captured, organized and integrated. So the correct procedure is displayed below:

Data must first be captured from various sources so we need various physical interfaces as data are in different format. After the caption of data due to their difference we need to organize them and then integrate them.

As soon as integration is completed then big data are ready to be analyzed, based on the problem issued, to be able to proceed with beneficial decision making.

In order to further understand big data, it helps to lay out the components of the architecture. Below is a diagram with the components based on their relationship:

<u>Operational data sources</u>

When a company is deciding to proceed with a classic data management, then of course has to take under consideration all related data sources in order to include them. This action secures that the company will have all relevant data for the issue to be solved. Till now the data sources were in classic format and usually from internal data sources. This has changed nowadays and in order for a company to proceed with a proper solution or a beneficial decision making, has to add more sources, sometimes external ones, with data in various format like content management systems so that company can have a 360-degree business view.

All operational data sources must include the below characteristics:

- Keep track of critical data required for real-time, day-to day operation
- To be updated continuously
- Contain structured and unstructured data
- Must be able to scale to support a huge amount of users in consistent basis

Organizing data services and tools

As a growing amount of data is produced from a variety of sources most likely most of them are not straightforward and need to be organized. Till now even if such information was captured from some companies there were no tools to provide results at a reasonable time frame. Now with the latest technological achievements is possible to manage this kind of data as computing cost has been dropped and new software solutions have been invented from emerging companies allowing to store, access and analyze huge amounts of data in real time.

Organizing data services are, in reality, an ecosystem of tools and technologies that can be used to gather and assemble data in preparation for further processing. As such, the tools need to provide integration, translation, normalization, and scale.

Analytical data warehouses and data marts

Analytical warehouses and marts provide compression, multilevel partitioning, and a massively parallel processing architecture. This allows companies to reveal patterns from their data and put it into a form that is available to their business. Typically, data warehouses and marts contain normalized data gathered from a variety of sources and assembled to facilitate analysis of the business. Data warehouses and marts simplify the creation of reports and the visualization of disparate data items. They are generally created from relational databases, multidimensional databases, flat files, and object databases — essentially any storage architecture.

Traditional and advanced analytics

The capability to manage and analyze petabytes of data enables companies to deal with clusters of information that could have an impact on the business. This requires analytical engines that can manage this highly distributed data and provide results that can be optimized to solve a business problem. Analytics can get quite complex with big data.

Reporting and visualization

Companies always create reports to give them an understanding of the collected data about everything from monthly sales figures to projections of growth. Big data changes the way that data is managed and used. If a company can collect, manage, and analyze enough data, it can use a new generation of tools to help management truly understand the impact not just of a collection of data elements but also how these data elements offer context based on the business problem being addressed. With big data, reporting and data visualization become tools for looking at the context of how data is related and the impact of those relationships on the future.

Big data applications

Big data applications are able to manage huge volumes, velocities and velocities of data and provide the requested result. In manufacturing, a big data application can be used to prevent a machine from shutting down during a production run. A big data traffic management application can reduce the number of traffic jams on busy city highways to decrease accidents, save fuel, and reduce pollution.

3.2 Types of Big data

One of the main characteristics of big data is the variety of data. Big data stores various types of information imported from various data sources. Therefore, taking advantage of big data requires that all this information be integrated for analysis and data management. Big data includes two main types of data structured and unstructured.

Structured data

Structured data are defined as data that has a defined length and format. Structure data usually include numbers, dates addresses e.t.c. This type of data is the most common that were used and can be easily managed from current database systems like a relational database and you can get the requested information using a language like structured query language (SQL).

The sources that are producing this kind of information is basically two:

- Computer/Machine generated

  This type of data can include:

  1. Sensor Data
  2. Web log data
  3. Point of sale data
  4. Financial data

- Human generated

  This type of data can include:

  1. Input Data
  2. Click stream data
  3. Gaming - related data

Unstructured data

Unstructured data are defined as data that does not follow a specific format. This type of data is not the most common but for sure takes almost the biggest amount of data generated. As a percentage of data connected the unstructured are calculated as 80% of data and structure as a 20%. The reason that are not the most common used from the companies is that technology didn't really support doing much with it.

The sources that are producing this kind of information is basically two:

- Computer/Machine generated

  This type of data can include:

  1. Satellite Images
  2. Scientific data
  3. Photograph and video
  4. Radar or sonar data

- Human generated

  This type of data can include:

  1. Social media data
  2. Mobile data
  3. Website content
  4. Text within documents, logs, surveys and e-mails

3.3 Redundant Physical Infrastructure

The first level of a Big data architecture is the physical infrastructure. The physical infrastructure includes all hardware elements, physical cable connections, network connection, interfaces, mechanical sensors, etc.

To proceed with a Big data implementation your physical infrastructure must include the below principles:

- Performance
  Performance is the measured speed of a single transaction or query usually measured end to end. Sometimes is also referred as latency, low latency is translated to high speed and high speed to great performance. Of course since a big data implementation consists of a variety of elements we must be careful to avoid bottlenecks. The best scenario is all parts of the implementation to be at the same or similar level of performance as the measured performance is always the overall performance of the implementation.
- Availability
  A big data implementation especially in terms or real-time data management must be always operational. The measurement of availability is the percentage of guaranteed uptime of service and the restoration time of a hardware failure. This demand all systems related to have spare unit for immediate replacement as minimum but the standard procedure is for all systems to be in pairs in working condition, mirrored so in the event that the primary system will go down the second system will continue without any service interruption. The restoration time for the failed system in this case is not vital but still important as the service has continue without any interruption. This technique is also called failovering.

- Scalability

  This is the ability of the infrastructure to be able to handle more data without applying major changes. The infrastructure must not been designed for limited amount of data.
- Flexibility

  How fast and easy we can apply new resources to our system.
- Cost

  As the infrastructure is a set of components that each one has a specific cost translated to a specific performance. The important issue in this case is the wise selection of the components based on a specific budget. This is the more difficult part of a big data implementation as a good balance of cost and performance is needed.

Most big data implementations need to be highly available, so the networks, servers, and physical storage must be both resilient and redundant. Resiliency and redundancy are interrelated. An infrastructure, or a system, is resilient to failure or changes when sufficient redundant resources are in place, ready to jump into action. In essence, there are always reasons why even the most sophisticated and resilient network could fail, such as a hardware malfunction. Therefore, redundancy ensures that such a malfunction won't cause an outage. Resiliency helps to eliminate single points of failure in your infrastructure.

Usually physical infrastructure based in companies' premises are limited and costly. So a solution to this problem came from another technology which also has made great improvements, the Cloud-Computing technology.

Cloud computing refers to a broad set of computing and software products that are sold as a service, managed by a 3rd-party provider and delivered over a network. Infrastructure-as-a-Service (IaaS) is a flavor of cloud computing in which on-demand processing, storage or network resources are provided to the customer. Sold on-demand with limited or no upfront investment for the end-user, consumption is readily scalable to accommodate spikes in usage. Customers pay only for the capacity that is actually used (like a utility), as opposed to self-hosting, where the user pays for system capacity it is are used or not.

This means that the technical and operational complexity is masked behind a collection of services, each with specific terms for performance, availability, recovery, and so on. These terms are described in service-level agreements (SLAs) and are usually negotiated between the service provider and the customer, with penalties for noncompliance.

3.4 Security Infrastructure

Big data management systems has similar security and privacy requirements as the conventional data management systems. The security requirements have a minimum level but must always be customized to specific business needs. The most important arising challenges when big data becomes part of the strategy are below:

- Data Access
  Access for users to raw or computed big data has about the same level of technical requirements as non-big data implementations. The data should be available only to relevant personnel who have a legitimate business need for examining or interacting with it. Most core data storage platforms have strong security schemes and are often applied in several points, providing appropriate access across the many layers of the architecture.
- Application access
  Application access to data is also needed. Most application programming interfaces (APIs) offer protection from unauthorized usage or access. This level of protection is probably adequate for most big data implementations.
- Data Encryption
  Data encryption is the most challenging aspect of security in a big data environment. In traditional environments, encrypting and decrypting data really stresses the systems' resources. With the volume, velocity, and varieties associated with big data, this problem is more complicated. A more wise approach is to identify the data elements requiring this level of security and to encrypt only the necessary items.
- Threat detection
  The use of mobile devices and social networks in some cases increases both the amount of data and the possibilities for security threats. It is therefore important that organizations take a more complete approach to security. Sometimes is may not be a direct part of a big data architecture the security of each device having access to the data and the security of the mentioned device could not be managed from the administrator of the big data management platform but it must be under consideration and sometimes a minimum level of security can be requested from the device vendor.

3.5 Operational Databases

At the core of any big data system are the database engines containing the collections of data elements. These engines need to be fast, scalable, and rock solid. There are more than one database engines each one with specific advantages and disadvantages resulting big data systems having better performance with one engine than another, or more likely with a mix of database engines. No single right choice exists regarding database languages. Although SQL is one of the most common database query language in use today, other languages may provide a more effective or efficient way of solving specific big data challenges.

It is very important to understand what types of data can be manipulated by the database and whether it supports true transactional behavior. Database designers describe this behavior with the acronym ACID. It stands for:

- Atomicity
  Requires that each transaction is "all or nothing": if one part of the transaction fails, the entire transaction fails, and the database state is left unchanged. An atomic system must guarantee atomicity in each and every situation, including power failures, errors, and crashes. To the outside world, a committed transaction appears (by its effects on the database) to be indivisible ("atomic"), and an aborted transaction does not happen.
- Consistency
  Ensures that any transaction will bring the database from one valid state to another. Any data written to the database must be valid according to all defined rules, including but not limited to constraints, cascades, triggers, and any combination thereof. This does not guarantee correctness of the transaction in all ways the application programmer might have wanted (that is the responsibility of application-level code) but merely that any programming errors do not violate any defined rules.
- Isolation
  Ensures that the concurrent execution of transactions results in a system state that would be obtained if transactions were executed serially, i.e. one after the other. Providing isolation is the main goal of concurrency control. Depending on concurrency control method, the effects of an incomplete transaction might not even be visible to another transaction

- Durability

  Means that once a transaction has been committed, it will remain so, even in the event of power loss, crashes, or errors. In a relational database, for instance, once a group of SQL statements execute, the results need to be stored permanently (even if the database crashes immediately thereafter). To defend against power loss, transactions (or their effects) must be recorded in a non-volatile memory.

Below is a comparison of these characteristics of SQL and NoSQL databases:

Engine	Query Language	MapReduce	Data Types	Transactions	Examples
Relational	SQL, Python, C	No	Typed	ACID	PostgreSQL, Oracle, DB/2
Columnar	Ruby	Hadoop	Predefined and typed	Yes, if enabled	HBase
Graph	Walking, Search, Cypher	No	Untyped	ACID	Neo4J
Document	Commands	JavaScript	Typed	No	MongoDB, CouchDB
Key-value	Lucene, Commands	JavaScript	BLOB, semityped	No	Riak, Redis

3.6 Organizing Data Services and Tools

Organizing data services and tools capture, validate, and assemble various big data elements into contextually relevant collections. As big data is massive, techniques have evolved to process the data efficiently and seamlessly.

Organizing data services is a set of tools and technologies that can be used to gather and assemble data in preparation for further processing. As such, the tools need to provide integration, translation, normalization, and scale.

Such technologies include the below:

- A distributed file system
  Necessary to accommodate the decomposition of data streams and to provide scale and storage capacity
- Serialization services
  Necessary for persistent data storage and Multilanguage remote procedure calls (RPCs)
- Coordination services
  Necessary for building distributed applications (locking and so on)
- Extract, transform, and load (ETL) tools
  Necessary for the loading and conversion of structured and unstructured data into Hadoop
- Workflow services
  Necessary for scheduling jobs and providing a structure for synchronizing process elements across layers

3.7 Analytical Data Warehouses

Specialized data storage system that combines desperate data sources into a clean, business centric, historically accurate, user friendly, cohesive view of an enterprise's business information. Analytical Data Warehouses turn raw data into business friendly information that allows business users to easily preform enterprise analytical functions, such as performance management or business optimization. Analytical Data Warehouses contain an enterprise wide view of information that combines many different views of business unit information

The role of an Analytical Data Warehouse or Analytical Data Mart is to support decision making. The methodology used for the Data Warehouse does not change the purpose, that of decision making.

To enable decision making, data is sourced from multiple operational systems, like HR, CRM, Sales, ERP systems. The operational data is combined and transformed into a more business friendly format to support decision making. An example of the transformation that occurs in an Analytical Data Warehouse to turn raw data into information is taking reference codes, such as UPCs, and transforming the codes into descriptive information, such as a product name.

Both people and other software programs act as the decision makers who leverage Analytical Data Warehouse information. The information in the Analytical Data Warehouse is typically delivered through specialized software tools that provide a friendly user interface, through the use of consumer-level web technologies, data visualizations, and drag and drop interfaces. These tools are typically referred to as Business Intelligence or Decision Support tools.

To reiterate the role of an Analytical Data Warehouse is to collect and consolidate multiple operational sources of data, transform that data into business-friendly information, and make the information available for users or other software systems to consume. All this is done to enable fact-based decision making, either by end users or automated systems.

3.8 Big Data Analytics

Big data analytics is the process of examining big data to uncover hidden patterns, unknown correlations and other useful information. The primary goal of big data analytics is to help companies make better business decisions by enabling data scientists and other users to analyze huge volumes of transaction data as well as other data sources that may be left untapped by conventional business intelligence (BI) programs.

We list three classes of tools that they can be used independently or collectively. The three classes of tools are as follows:

- Reporting and dashboards
  These tools are used to provide a "user-friendly" presentation of the information from various sources. Some of the tools that are being used are traditional ones that can now access the new kinds of databases collectively called NoSQL (Not Only SQL).
- Visualization
  These tools are the next step in the evolution of reporting. The output tends to be highly interactive and dynamic in nature. Another important distinction between reports and visualized output is animation. Business users can watch the changes in the data utilizing a variety of different visualization techniques, including mind maps, heat maps, infographics, and connection diagrams. Often, reporting and visualization occur at the end of the business activity. Although the data may be

imported into another tool for further computation or examination, this is the final step.

- Analytics and advanced analytics

  These tools reach into the data warehouse and process the data for human consumption. Advanced analytics should explicate trends or events that are transformative, unique, or revolutionary to existing business practice. Predictive analytics and sentiment analytics are good examples of this science.

3.9 Big Data Applications

Custom and third-party applications offer an alternative method of sharing and examining big data sources. These applications are either horizontal, in that they address problems that are common across industries, or vertical, in that they are intended to help solve an industry-specific problem.

The most prevalent categories as of this writing are log data applications (Splunk, Loggly), ad/media applications (Bluefin, DataXu), and marketing applications (Bloomreach, Myrrix). Solutions are also being developed for the healthcare industry, manufacturing, and transportation management. Like any other custom application development initiative, the creation of big data applications will require structure, standards, rigor, and well-defined APIs. Most business applications wanting to leverage big data will need to subscribe to APIs across the entire stack. It may be necessary to process raw data from the low-level data stores and combine the raw data with synthesized output from the warehouses. As you might expect, the operative term is custom, and it creates a different type of pressure on the big data implementation.

Software developers need to create consistent, standardized development environments and devise new development practices for rapid rollout of big data applications.

Chapter 4

Handoop Implementation

The Term Big Data is getting more and more Attention. Looking up the Term will state that Big Data is very large, loosely structured that defies traditional Storage.

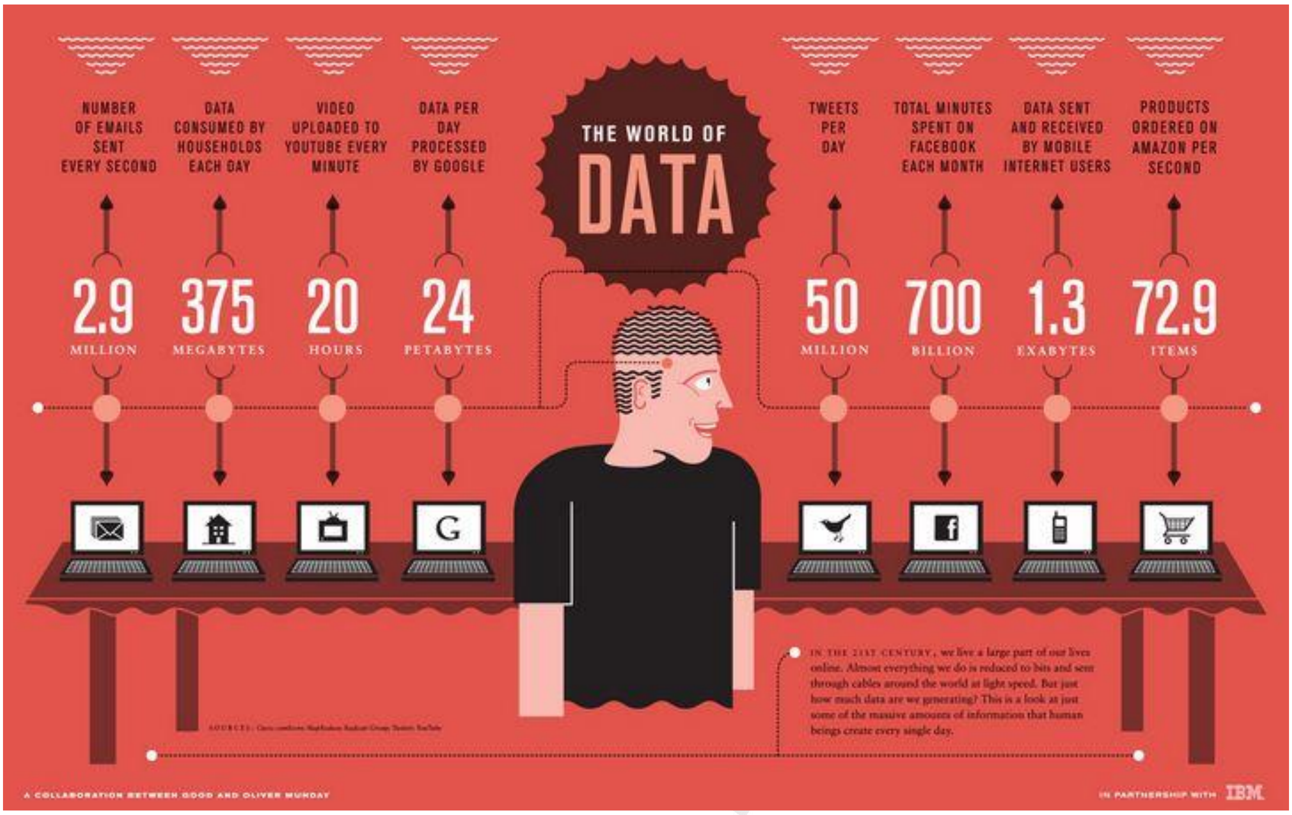
Where does all the data come from? Until now most part of the enormous information was Human generated Data. It includes emails, photos, documents, videos even social media related data like Facebook, Twitter, Linked In, etc. In the recent years another type of Data is been produced: Machine Generated Data. It consists of sensor data, various logs of all kinds. This new breed of information is orders of magnitude larger than the Human generated Data. Due to its sheer magnitude and lack of structure, most of the machine generated Data was mostly ignored even not captured. Any attempt to utilize the data was disgarded as not possible or not cost effective.

To visualize the Volume of Information produced:

Facebook stores each day about 100 TB adding to the whooping 40 PB of Storage.

Yahoo the largest data storage owns over 60 PB of data.

Ebay as the largest e commerce provider, captures over 50 TB per day adding to its 40 PB od stored Information.



What are the challenges using Big Data?

The most obvious fact is the size. It renders the traditional Databases or storage filers obsolete due to the cost of such an endevor.

Another big issue is the nature of the data. Most of it is unstructured or semi structured. A good example might be log files which structure is origin related.

Every major player involved in Information technologies today knows that Big Data is part of the game. One needs to process it in order to harvest all the intelligence. Until now traditional data processing models used to store its information in a storage cluster which in turn was copied to a compute cluster for processing and the results were written back to the storage cluster. It is obvious by now that this approach is too time consuming and doesnt quite Work.

Hadoop was created as a solution to the problems mentioned above. The major difference to previous models and approaches is that Hadoop provides distributed storage and distributed processing all in one. What is Hadoop?

Hadoop is an open source software stack that runs on a cluster of machines. It was originally developed and open sourced by Yahoo. Today Hadoop has evolved to an Apache Foundation project and has numerous renown contributors which refine its unique characteristics:

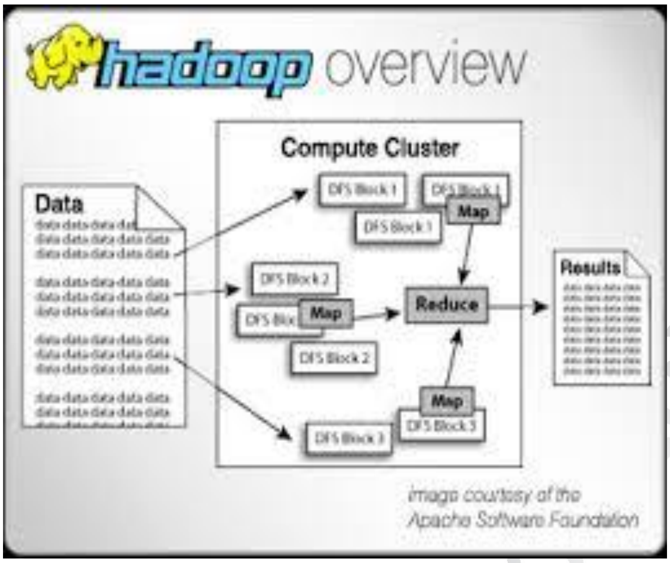
Hadoop clusters scale horizontally which means that more storage and compute power can be achieved by adding additional nodes to the cluster. This eliminates the need of more powerful and more expensive hardware by utilizing commodity hardware, reducing the overall cost.

Hadoop was designed to handle instructed or semi structured Data. Due to the fact that it doesn't enforce schemas on the data it stores, it can handle any kind of Information (text, binary data)

Hadoop is an open source implementation of Googles distributed computing framework which is proprietary. It consists of two parts:

Hadoop Distributed File System (HDFS) , which is modeled after Google's GFS and Hadoop MapReduse, which is modeled after Google's MapReduce.

In Hadoop speak, storage is provided by Hadoop Distributed File System (HDFS). Compute is provided by MapReduce.

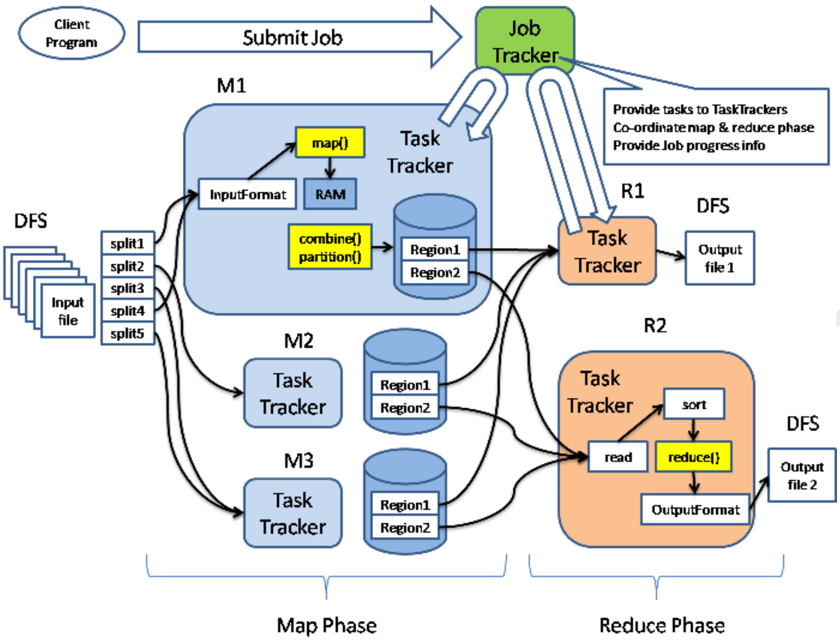


MapReduce is a programming framework. Its description was published by Google in 2004 [http://research.google.com/archive/mapreduce.html]. Much like other frameworks, such as Spring, Struts, or MFC, the MapReduce framework does some things for you, and provides a place for you to fill in the blanks. What MapReduce does for you is to organize your multiple computers in a cluster in order to perform the calculations you need. It takes care of distributing the work between computers and of putting together the results of each computer's computation. Just as important, it takes care of hardware and network failures, so that they do not affect the flow of your computation. You, in turn, have to break your problem into separate pieces which can be processed in parallel by multiple machines, and you provide the code to do the actual calculation.

MapReduce takes care of distributed computing. It reads the data, usually from its storage, the Hadoop Distributed File System (HDFS), in an optimal way. However, it can read the data from other places too, including mounted local file systems, the web, and databases. It divides the computations between different computers (servers, or nodes).

It is also fault-tolerant. If some of your nodes fail, Hadoop knows how to continue with the computation, by re-assigning the incomplete work to another node and cleaning up after the node that could not complete its task. It also knows how to combine the results of the computation in one place:

- MapReduce has a master and workers, but it is not all push or pull, rather, the work is a collaborative effort between them.
- The master assigns a work portion to the next available worker; thus, no work portion is forgotten or left unfinished.
- Workers send periodic heartbeats to the master. If the worker is silent for a period of time (usually 10 minutes), then the master presumes this worker crashed and assigns its work to another worker. The master also cleans up the unfinished portion of the crashed worker.
- All of the data resides in HDFS, which avoids the central server concept, with its limitations on concurrent access and on size. MapReduce never updates data, rather, it writes new output instead. This is one of the features of functional programming, and it avoids update lockups.
- MapReduce is network and rack aware, and it optimizes the network traffic.
- MapReduce splits computation into multiple tasks. They are called Mappers and Reducers.
- Each phase (map phase and reduce phase) can be parallelized independently.



What are the benefits of MapReduce programming? it summarizes a lot of the experiences of scientists and practitioners in the design of distributed processing systems. It resolves or avoids

several complications of distributed computing. It allows unlimited computations on an unlimited amount of data. It actually simplifies the developer's life. And, although it looks deceptively simple, it is very powerful, with a great number of sophisticated (and profitable) applications written in this framework.

HDFS, or the Hadoop Distributed File System, gives the programmer unlimited storage (fulfilling a cherished dream for programmers). However, here are additional advantages of HDFS:

- *Horizontal scalability*: Thousands of servers holding petabytes of data. When you need even more storage, you don't switch to more expensive solutions, but add servers instead.
- *Commodity hardware*: HDFS is designed with relatively cheap commodity hardware in mind. HDFS is self-healing and replicating.
- *Fault tolerance*: Every member of the Hadoop zoo knows how to deal with

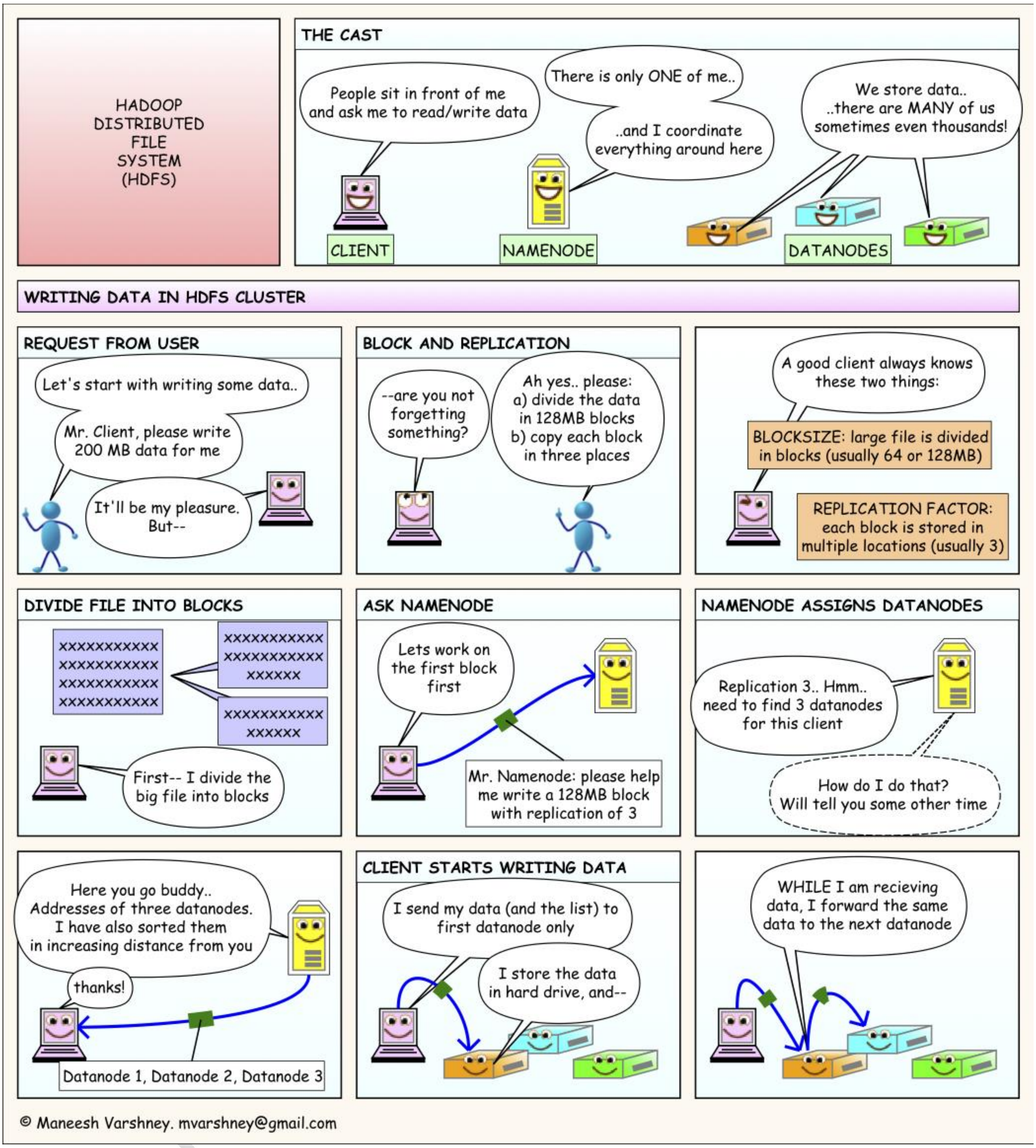
hardware failures. If you have 10 thousand servers, then you will see one server fail every day, on average. HDFS foresees that by replicating the data, by default three times, on different data node servers. Thus, if one data node fails, the other two can be used to restore the third one in a different place.

HDFS implementation is modeled after GFS, Google Distributed File system, thus you can read the first paper on this, to be found here: http://labs.google.com/papers/gfs.html.

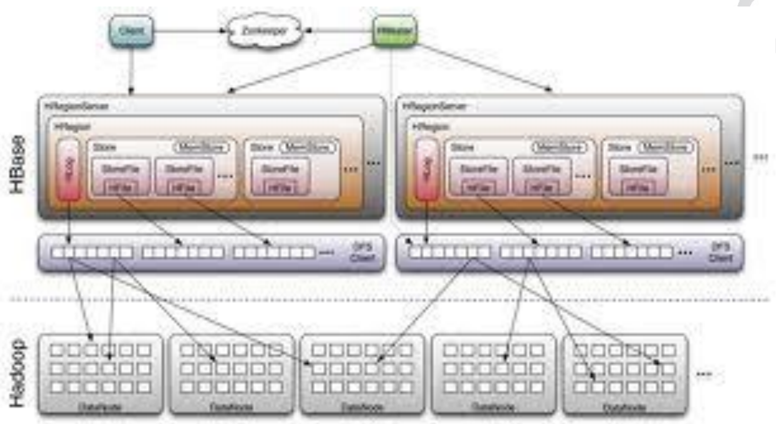
In an HDFS cluster, there is ONE master node and many worker nodes. The master node is called the Name Node (NN) and the workers are called Data Nodes (DN). Data nodes actually store the data. They are the workhorses. Name Node is in charge of file system operations (like creating files, user permissions, etc.). Without it, the cluster will be inoperable. No one can write data or read data. This is called a Single Point of Failure. As we saw hadoop doesn't need fancy, high end hardware. It is designed to run on commodity hardware. The Hadoop stack is built to deal with hardware failure and the file system will continue to function even if nodes fail. The file system will continue to function even if a node fails. Hadoop accomplishes this by duplicating data across nodes. So how does Hadoop keep data safe and resilient in case of node failure? Simple, it keeps multiple copies of data around the cluster.

To understand how replication works, lets look at the following scenario. Data segment #2 is replicated 3 times, on data nodes A, B and D. Lets say data node A fails. The data is still accessible from nodes B and D.

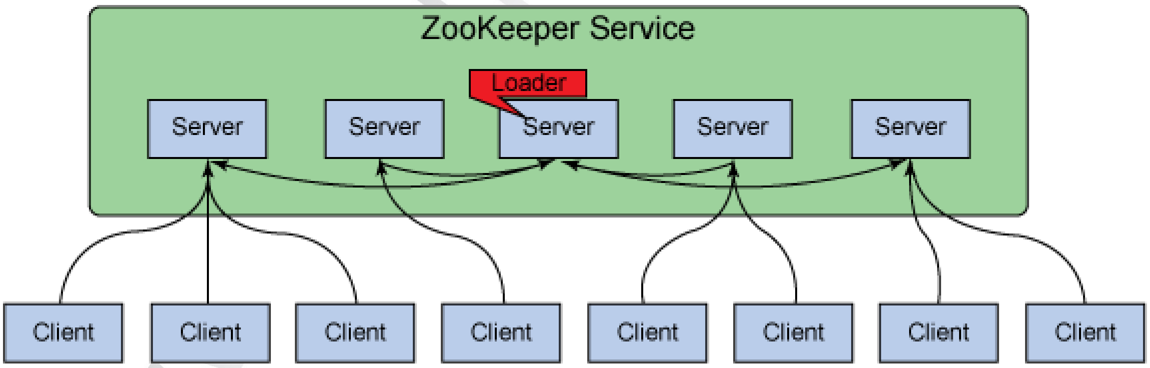
HDFS supports writing files once (they cannot be updated). This is a stark difference between HDFS and a generic file system (like a Linux file system). Generic file systems allows files to be modified. However appending to a file is supported. Appending is supported to enable applications like Hbase.


HADOOP DISTRIBUTED FILE SYSTEM (HDFS)
THE CAST
People sit in front of me and ask me to read/write data
There is only ONE of me..
..and I coordinate everything around here
We store data.. ..there are MANY of us sometimes even thousands!
CLIENT
NAMENODE
DATANODES
WRITING DATA IN HDFS CLUSTER
REQUEST FROM USER
Let's start with writing some data..
Mr. Client, please write 200 MB data for me
It'll be my pleasure. But--
BLOCK AND REPLICATION
--are you not forgetting something?
Ah yes.. please: a) divide the data in 128MB blocks b) copy each block in three places
A good client always knows these two things:
BLOCKSIZE: large file is divided in blocks (usually 64 or 128MB)
REPLICATION FACTOR: each block is stored in multiple locations (usually 3)
DIVIDE FILE INTO BLOCKS
xxxxxxxxxxx xxxxxxxxxxx xxxxxxxxxxx xxxxxxxxxxx
xxxxxxxxxxx xxxxxxxxxxx xxxxxx
xxxxxxxxxxx xxxxxx
First-- I divide the big file into blocks
ASK NAMENODE
Lets work on the first block first
Mr. Namenode: please help me write a 128MB block with replication of 3
NAMENODE ASSIGNS DATANODES
Replication 3.. Hmm.. need to find 3 datanodes for this client
How do I do that? Will tell you some other time
Here you go buddy.. Addresses of three datanodes. I have also sorted them in increasing distance from you
thanks!
Datanode 1, Datanode 2, Datanode 3
CLIENT STARTS WRITING DATA
I send my data (and the list) to first datanode only
I store the data in hard drive, and--
WHILE I am recieving data, I forward the same data to the next datanode
© Maneesh Varshney. mvarshney@gmail.com

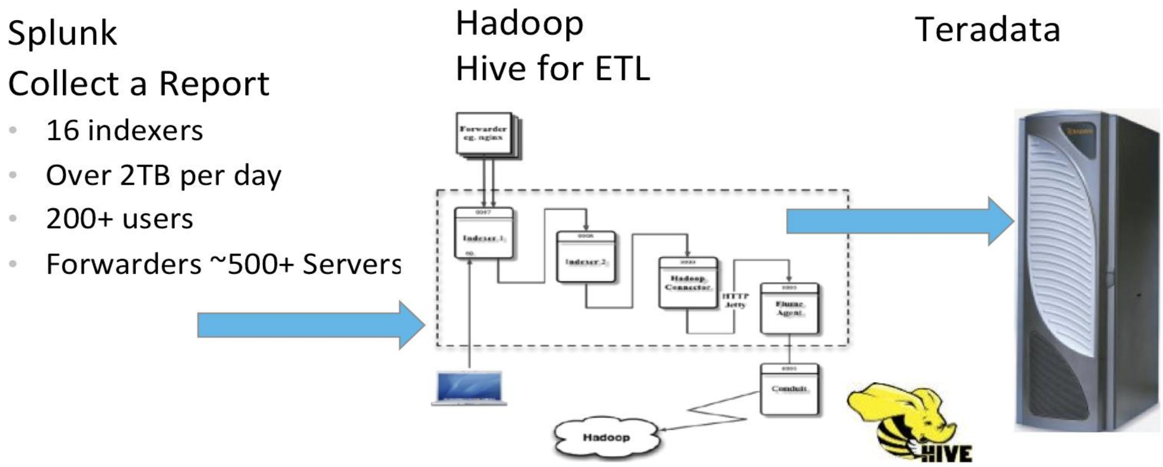
Hadoop utilizes and depends on more projects than the 2 described above, which deal with a certain aspect of big Data:



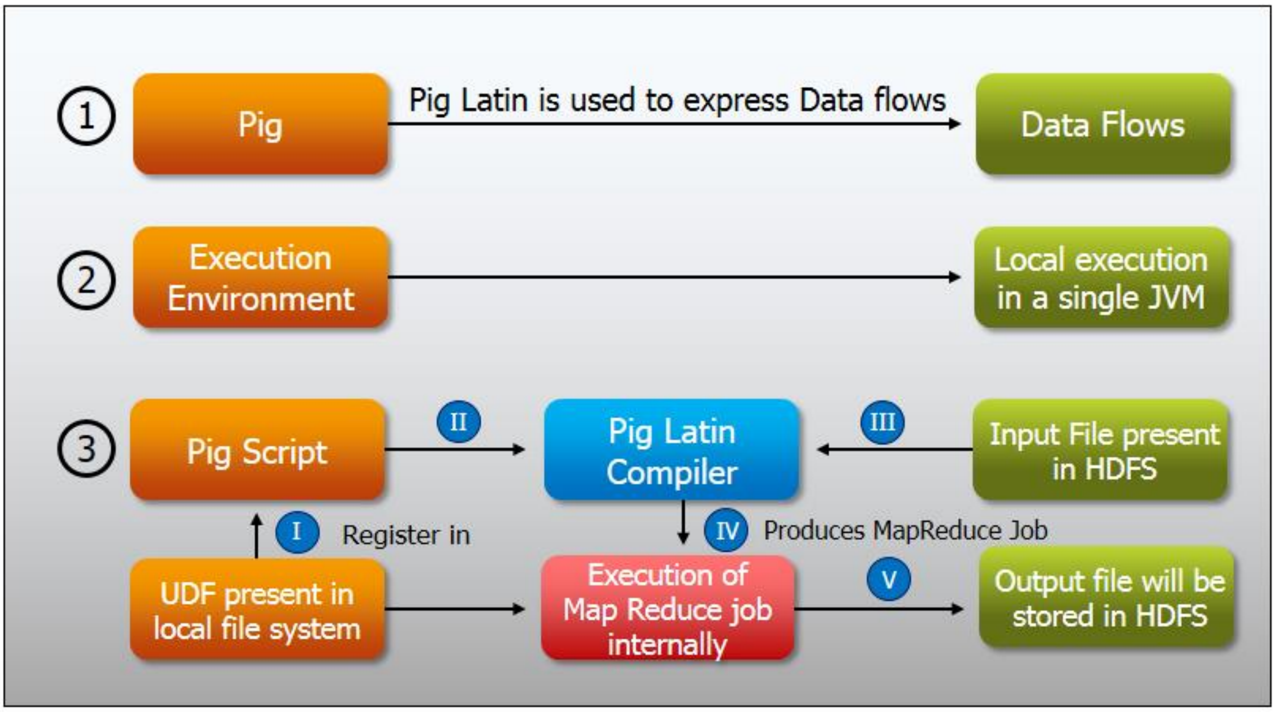
HBase is a database for Big Data, up to millions of columns and billions of rows. Another feature of HBase is that it is a key-value database, not a relational database.



ZooKeeper is a centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services. ZooKeeper is also fault-tolerant.



Hive defines a simple SQL-like query language, called QL, that enables users familiar with SQL to query the data. At the same time, if your Hive program does almost what you need, but not quite, you can call on your MapReduce skill. Hive allows you to write custom mappers and reducers to extend the QL capabilities.



Pig: If Hive is the SQL of Big Data, then Pig Latin is the language of the stored procedures of Big Data. It allows you to manipulate large volumes of information, analyze them, and create new derivative data sets. Internally it creates a sequence of MapReduce jobs and the programmers can use this simple language to solve pretty sophisticated large-scale problems.

Chapter 5

Big data in Maritime Shipping Industry

In many industries data processing has already created significant value. For the maritime shipping industry the data processing has to overcome low speed and high cost of data transmission. This was caused due to technological limitations. The marine industry has historically followed other industries in terms of technology and never been ahead. Technology solutions are either coming from the naval defense world or land-based solutions are being converted to marine applications.

Current generation of marine satellite terminals are capable of data transmission speed in terms of downloading/uploading up to 512kbits/sec. The data traffic of such terminals is billed per amount of data in rather high cost rates. The maritime company is able to select one of the below packages:

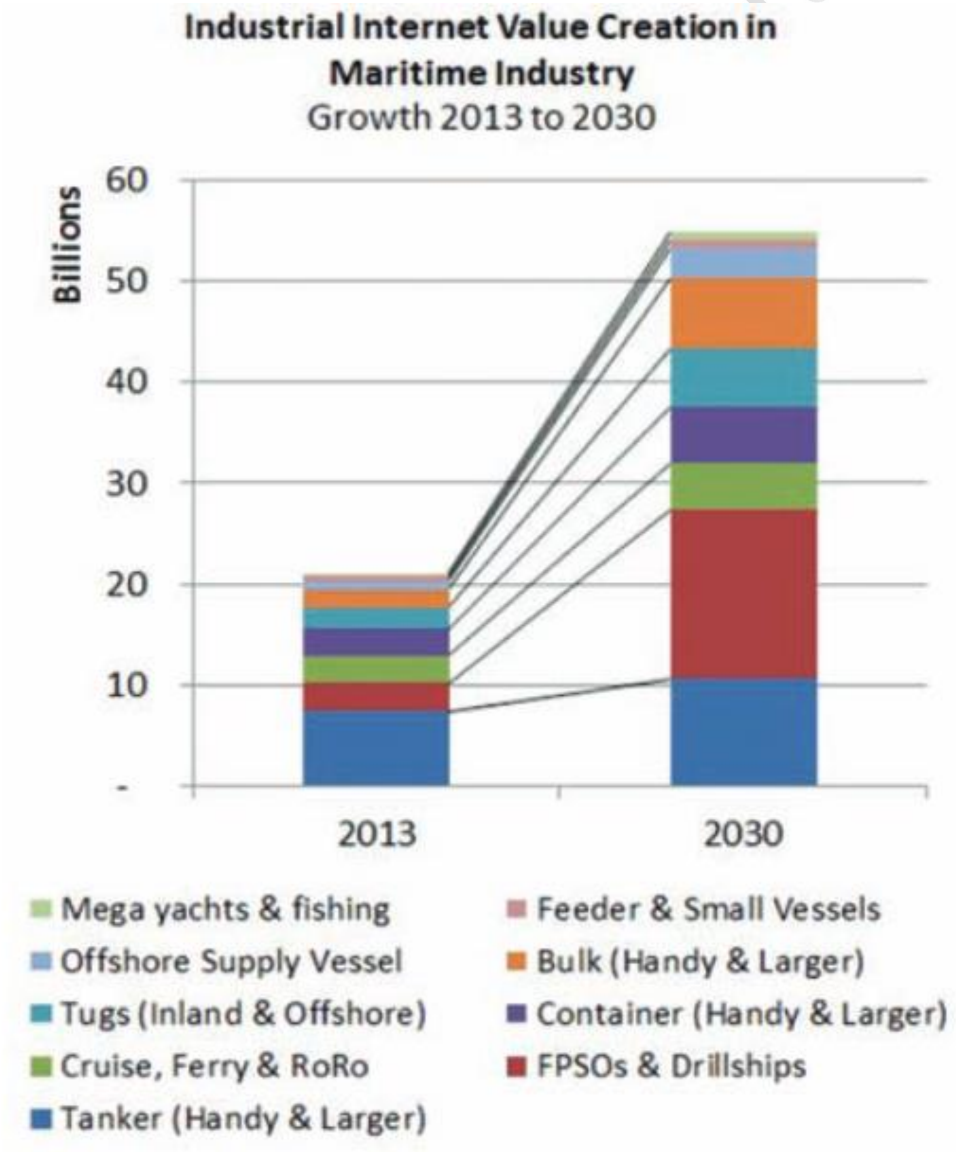
PLANS	Standard	Entry	2GB	6GB	AYCE
ALLOWANCE	10 MB	200MB	2GB	6GB	Unlimited
CONTRACT	1-month	1-year	2 -years	2 -years	2 -years

All plans consists of a monthly fee and a cost per Mbytes for data volume exceeding the allowance. The cost per Mbytes is not the same in any plan but is getting lower in bigger plans.

With the new generation of Ka band terminals the limitations of low speed and high cost of data transmission have been solved. With the new generation of satellites that will be launched from Inmarsat by the end of 2015, maritime shipping industry will be able to use download speed up to 50mbit/sec and upload speed up to 5 mbit/sec in terms bandwidth. Regarding the cost the maritime company will pay a monthly fee that will billed per bandwidth and not per amount of data. So the maritime company will select the needed bandwidth of 1 up to 50 Mbit/sec that will be charged regardless the amount of data passed.

Bringing the right, intuitive technology solution enables an easier transition for the maritime shipping industry to begin incorporating with data processing into how they operate their assets and businesses.

It has been estimated that the opportunity across industries only from internet will exceed US$10 trillion per year in the next 15 years, the opportunity for asset owners, operators and managers to reduce costs, improve fuel efficiency, and increase uptime and reliability is approximately US$20 billion today and will exceed US$50 billion by 2030 (4).



Estimated growth in value creation for industrial internet application in maritime

This opportunity exists across sectors, ranging from super-tankers to inland tugboats and offshore platforms. In order to capture this opportunity, asset owners and operators will need to bring together equipment with sensors, software analytics and well trained people. Equipment is being installed with more and more sensors and often that data is being integrated into more comprehensive monitoring and control systems.

Not only will new technology be needed, both in terms of equipment, data integration and software analytics, maritime shipping companies will have to evaluate their employees and make investments in people. While traditional maintenance and operations skillsets will not become obsolete, they will need to be augmented by data analytics capabilities. Training of existing personnel will be needed to provide some level of data analytics capability and many maritime shipping companies will need to bring in new resources, likely from other industries, to provide deeper data analytics expertise.

The basic concept is connecting machines with each other and with people to get more out of assets, help people be more productive, make supply chains more efficient, enhance customer experience and drive innovation.

Newer ships launched recently are equipped with more sensors, providing more and more performance and condition data that can be used to operate and maintain that equipment at a higher performance level and lower cost. This wealth of data, while it can create value, does create a challenge as it is overwhelming without analytical tools. For example, a new vessel today might have over 1,000 data points, which would create 2.6 billion pieces of data over a month, and when extrapolated across a fleet of 100 assets would equal over 3 trillion data points per year. Software analytics can integrate a variety of data sources in a variety of formats and use automated algorithms to help users make sense of the data and turn that data into actionable information.

The new information available from this data is consumed by people as they make more informed decisions ranging from planning maintenance to optimizing equipment configuration to prioritizing resources across an enterprise. In order to transform the data into actionable information, the people involved need to have domain expertise in how the machines operate, how the business works, and how to analyze data. These people need access to the data and information through multiple channels, including web, mobile, intelligent reports, and enterprise applications.

Across segments, the maritime shipping industry will benefit from using the data to reduce fuel consumption, increase equipment reliability, decrease maintenance costs, and ensure environmental compliance. Based on the global fleet as of 2013, the value creation potential is estimated at approximately US$20 billion.

With more new-builds being equipped with smarter machines and more robust technology infrastructure, that value creation potential is projected to grow at 15-20 per cent per year for the next 5 years. While all sectors in marine would like equipment that operates more efficiently and does not fail, some sectors have specific opportunities to create value. (3)

Container ships are one particular example. Container ships can apply the concept to increase equipment reliability and thus increase schedule reliability and deliver better value to their customers and differentiate themselves. They can also use the existing data on board to operate their ships more efficiently such as optimising the configuration of generators based on the actual electrical load and the actual performance of those specific generators on board that specific ship. Container ships can also utilise the realtime electrical load on the ship (with changes primarily driven by the number of refrigerated containers being transported) and optimise the vessel speed to minimize total fuel consumption, including both propulsion fuel consumption as well as electrical generation consumption. This can translate to millions of dollars per year in fuel consumption reduction and hundreds of thousands of dollars per year in maintenance cost avoidance. Not to mention the incremental value of more revenue and greater customer value with a more reliable asset. Outside of just container shipping, the benefits for most sectors can be categorized a few ways. First, there are benefits by shifting to Condition Based Maintenance (CBM), both reducing maintenance costs as well as energy consumption. It is easy for maintenance planners to defer maintenance when there is not a clear reason for why it is necessary now.

For example, what could be a $5,000 maintenance event (fuel injector maintenance) could easily be lined out to save $5,000, only to have that issue propagate into a $750,000 top end overhaul during dry dock. By doing the right maintenance, at the right time, at the right cost, additional surprises could be avoided during extended maintenance periods. In this example, there is additional value created by returning the cylinder to 'like new' condition and minimizing excessive fuel consumption. In addition to CBM there is also CBO – Condition Based Operations. CBO is where operators use real-time data and analytics to make better operational decisions to optimize the operation of the asset. Examples of CBO include using real-time performance data to optimize which generators are being utilized for different electric demand situations, and using real-time performance data to optimize speed, trim and draft. Both the CBM and CBO philosophies can be acted upon by onboard and on shore users. Onboard users can use that information in real-time to make better maintenance and operations decisions, while shorebased users can use the information to plan more effectively at the enterprise level as well as manage operator performance across assets. In the marine space, not only will operators and managers on board and on shore benefit, all levels of leadership as well as external stakeholders stand to see benefits. Senior management can have a better understanding of the health of their enterprise, prioritize issues across a fleet and can see and measure the impact of their policy decisions. Vessel lifecycle will also be an important factor in how the big data technologies are adopted in the marine space. New vessels are often being built with significant sensoring and a strong technology infrastructure. This will make it easier for that data to be consumed by analytics applications in order to convert it into actionable information. On board new builds, the required investment could be very low, enabling fast payback times and therefore fast adoption of the big data concept. Older ships, without electronic engines and little sensorisation, will face a different required investment than newer ships with built in sensors in order to capture value. Owners of older, existing ships will have to weigh the increased investment in sensors, data integration, networking and communications with the potential return. While the benefits will likely significantly overcome the required investment for higher value assets, it will be less clear cut for older, lower value vessels with little existing technology infrastructure. Based on an analysis of the approximately 100,000 ships in the global fleet today, there are 20,000-30,000 ships sailing the oceans that already have some technology infrastructure in place and would be able to easily justify the investment required to start taking advantage of the onboard data. This number is expected to grow at 3,000-7,000 per year, with most newbuilds incorporating a solid technology infrastructure in the shipyard during construction.

Historically, original equipment manufacturers (OEM) involved in data integration would each have their own data protocol for transferring data or making it available to other systems. This increased the integration costs to bring together data from relevant equipment, for example a main engine, fuel flow meter, torque meter and ECDIS system. While other industries, such as the power generation industry, have already moved to common open standards data protocols, the marine industry is just starting to move in this direction.

Instead of a different proprietary data interface for each piece of equipment, more and more marine equipment is starting to communicate either with Modbus, NMEA or OPC standards. This is enabling new-build ships to more easily transition to the big data technologies.

In addition to enabling communications across systems, open standards will also allow for the creation of standard equipment hierarchies across maintenance management programs, Classification equipment registries, and OEM equipment databases.

These standard hierarchies will help facilitate seamless data sharing between different potential data consumers in order to maximise value creation for vessel owners. While most OEMs are adopting this approach, there are some who are maintaining their own internal, proprietary data protocols, thus limiting the amount of integration that can be achieved.

If other industries, which have moved in this direction previously, provide a good prediction of potential marine industry dynamics, than the OEMs who do not adopt open standards data protocols will be penalised by the market, as their solutions will be less valuable than if they are were able to easily integrate with the broader big data technologies

Equipment OEMs and software providers are also focusing on how to utilise the available data to create actionable information to help drive better operational and maintenance decisions.

Many OEMs are starting to use the data from their specific piece of equipment to help ship owners and operators either operate or maintain their equipment more effectively. If this trend continues with each OEM providing their own 'big data' solution, only for their equipment, owners and operators will become overwhelmed with information technology and not be able to get the value out of each system.

In addition, each component on the ship is not operating in isolation and needs to be analysed as part of the larger system.

For instance, an increase in fuel consumption could be driven by decreased engine condition, change in fuel pressure, increased fuel temperature, lack of maintenance, increased marine growth on the hull, a fouled propeller, heavier than expected sea state, higher speed requirements, headwinds, or a different fuel type among many other potential reasons.

If the engine is analysed in isolation, it may lead to an issue not being identified or an incorrect diagnosis, and a decision being made without the most accurate and comprehensive information.

Integrated analytics that can consume data from multiple sources and help users transform that data into actionable information are the logical next step in the marine industry. As many vessels already have sensors and some level of data integration, monitoring and analytics solutions should be flexible enough to leverage the existing onboard systems such as integrated control systems, stand-alone sensors, navigation systems, and ballast and tank management systems.

This is necessary in order to minimize up-front investment in sensors and data integration in order to maximise the owner's return. Early adopters are starting to move in this direction, with the MV Gabriel Schulte described above being a good example.

Another challenge in the marine industry regarding analytics is how to deal with real-time data analytics on board a vessel, near-real-time analytics on-shore, and the additional ship and shore based business systems (i.e. maintenance planning, purchasing, supply chain, fuel management, cargo management, etc.). The software package used needs to enable ship and shore based users to interact with the data and create actionable information.

In addition to the analytics to transform data into actionable information, data validation and qualification must also occur.

Unlike many land-based industries and other transportation industries, marine assets are operated in a variety of different ways. While a cargo airplane essentially has four modes, with all engines being used for all four (taxi, take-off, cruise and landing), a commercial container ship can operate at a variety of speeds, plant configurations, displacement/drafts, electrical loads, acceleration/deceleration, weather conditions, etc.

In the airplane example, it may be fairly simple to compare a data point from multiple flights. In the marine industry, data should be qualified and validated to ensure like data points are being compared, both over time and across vessels.

These analytics must not only take the manual effort out of moving from raw data to actionable information, they must also be easy to use with minimal training. For example, the user interface should be simple and intuitive to enable an onboard operator or a shore-based manager to begin using the automated analytics with as little as an hour of training.

As discussed below, there is a role for skilled data analysts – however, the software must facilitate those without data analytics skills being able to engage with the analytics to understand equipment health and performance and use the analytics to make better decisions.

In addition to common focus areas of overall maintenance, equipment reliability, general fuel efficiency and environmental compliance, specific sub-sectors within the marine industry will require 'custom applications' in order to realize the full potential of the big data.

These custom applications might be focused on specialty equipment that is only used in a specific sector or optimizing within a specific business model.

While there will be a need for sector specific customer applications, there will also still be the need for the open standards platform that will encompass the common infrastructure (data collection, management, qualification, automated analytics) and functionality (analytics focused on reliability, fuel efficiency and the environment) as well as provide an open standards interface to allow integration of custom applications.

These applications might be supplied as add-on modules by the open standards platform provider, or they may be developed either in house or by a sector focused supplier. Outlined below are a few examples of the types of custom applications that would create additional value in specific sectors.

Tow configuration optimization for tugs – By utilizing actual performance data from a variety of sources, including engine performance, power, speed through water,

tow diagrams, etc., tug operators will be able to better understand exactly what factors cause a specific configuration of barges to be more efficient than others. Automated analytics could enable the tug operator to clearly understand how water conditions, tug power, and operational constraints impact the fuel efficiency of a given barge configuration and use that information to optimize profit in their specific business situation, whether that be reducing fuel consumption or reducing the time to detach and add barges along a voyage.

Combined engine & compressor analytics in offshore

– Many offshore platforms utilize a combined engine and compressor set for compressing and storing natural gas.

In order to better understand both the efficiency and health of the overall system, the system should be assessed as a whole instead of just standard analytics for the engine and standard analytics for the compressor.

Generator configuration for container ships

– Container ships often experience high fluctuations in the electrical load due to refrigerated containers being loaded and unloaded in various ports. With 3-5 generators on board many container ships, there are often multiple combinations of generators that will be able to produce the required electricity. Most power management systems today will help determine the number of generators needed based on OEM guidance. In practice, specific generators will perform differently over time and at various loads. This results in fuel efficiency differences between generator combinations, even with the same number of generators online (for example, using generators 1, 2 & 3 or using 1, 2 & 4 or 2, 3 & 4).

By better understanding the actual, recent performance of those specific generators, operators can have transparency into what the optimal configuration is for a specific load level at that point in time.

The above are just three examples of the types of custom applications that can be used in specific marine sectors. Other custom applications could be developed to better understand hull & propeller efficiency for specific ship types, efficiency/effectiveness of dynamic positioning systems, fuel efficiency of refrigeration, dredge productivity, optimizing LNG-tanker speed with real-time boil-over analysis, etc. These should all be able to interact with open standards platforms that will integrate, validate, analyze and share the data.

Bringing together the data from the variety of shipboard equipment is only the first step in creating value with the big data technologies. Value is actually created by transforming that data into actionable information that can be used by onboard operators as well as shore-based management to make better operational, maintenance and management decisions.

To facilitate the transformation of raw data into actionable information, skilled data analysts (as mentioned above), with deep marine domain expertise, will be needed to either conduct analysis or configure automated analytics in software.

Data analysts will need to be able to combine the skills and experience gained as a Chief Engineer or a Technical Superintendent (e.g. how equipment operates, troubleshooting, repairs, etc.) along with the ability to analyze data. This data analysis skill set is very different and will be new to many Chief Engineers and Technical Superintendents. Automated analytics that are incorporated into a software program will make it easier to bridge the gap between the current capabilities and the required capabilities to be able to maximize value capture from ships connected to the internet.

Organizations will also look to augment their existing technical teams with analysts who may not have the deep marine domain expertise, but will be able to help the Chief Engineers and Technical Superintendents maximize the value of the data and analytics. Skilled data analysts combined with a robust technical system would be able to take the troubleshooting findings from one vessel and create new algorithms that can instantly be applied across a fleet – all from the shore-based office. This data analyst role will also extend beyond just technical ship management. Fleet managers and trade-managers can use this data to better optimize route planning or vessel dispatch, both in order to save fuel, as well as optimize the entire system to use assets based on their current performance and expected reliability. Serving different stakeholders the marine industry presents another challenge with the broad range of stakeholders involved. There is the owner of the vessel who purchases the vessel and may or may not play a part in operations. Often, there is a ship-management company, who is responsible for crewing the vessel, maintenance, fuel and other operational aspects. Vessels are often chartered by another company who will often pay a daily rate plus the fuel consumption. In addition, there is the onboard crew, usually hired by the ship-management company with varying levels of experience in the industry and on that specific ship. Outside of the ship ownership and operational stakeholders, there are classification societies, who help ensure ships are seaworthy and safe. OEMs stay engaged beyond the initial building with service contracts and support. Each of these stakeholders has different incentives, and sometimes these incentives are at odds with each other. With this wide range of stakeholders, the value of the big data technologies will likely be shared across many of them, thus creating inefficient decision making as to whether to invest in new sensors, technology infrastructure, software or people. Organizations that are able to align all of the different stakeholders to focus on delivering their service most efficiently will be able to make the right investments and outperform those who are not able to reach a consensus across their stakeholders.

In addition to aligning stakeholders in order to make better investment decisions, a monitoring and analytics solution must allow for stakeholders in different roles to interact with the technology in different ways. Onboard operators must be able to use data and analytics to make better operational and maintenance decisions, often in real-time. Shore based managers must be able to look across a fleet and across time to understand variations in performance and the impact of maintenance and efficiency initiatives. Senior executives must be able to gain transparency into enterprise health and prioritise resources more effectively across the enterprise.

Conclusions

The big data technology presents a huge opportunity to the marine industry, with the potential to create over 20 billion dollars of value annually.

While only a minority of vessels are currently positioned to begin to capture the big data technology benefits, that number will grow significantly, as almost every new build ship is having technology built in to capture these benefits.

The benefits to marine stakeholders are significant. Substantial fuel savings, reduction in maintenance and repair costs, and greater assurance of environmental compliance are the largest drivers. Organizations need to start thinking now about how they are going capture benefit from the big data technology. Many marine organizations will need to bolster their technology and data analysis capabilities in order to take advantage of the opportunity. Those that don't make these organizational investments risk being left behind.

The marine industry has the opportunity to learn from other industries which are further down the 'big data' road, such as commercial aviation and power generation. Learning from these industry examples will help marine organizations mitigate the challenges and minimize costs.

Many companies are already beginning to invest in 'data collection'. This is causing as many challenges as is value creation.

Often, this data, if just collected, is overwhelming and will either paralyses decision making or simply be ignored. Real-time, automated analytics on the vessel and on shore are necessary to transform the raw data into actionable information that can be used to make better operational and maintenance decisions.

Proven technologies that leverage the established infrastructure of GPS, communications satellites, control systems, and the internet can use automated analytics to help make better decisions and be internalized within an organization to ensure intellectual property that is developed with the analytics is protected as a competitive advantage.

Stakeholders must work together to achieve these benefits. Since the benefits are often split amongst multiple stakeholders, multiple stakeholders must align on the business case and come together to share in the initial investment, as well as the follow-on savings.

This applies to industry providers, especially equipment OEMs. Open standards will enable value creation for a variety of players and will be one factor that contributes to who the market rewards and who the market penalizes.

Non-traditional stakeholders, such as software companies and technology companies from outside of the marine space, will introduce new capabilities to marine customers.

With such significant value creation possible for a wide variety of stakeholders, and a different set of capabilities required to capture that value, some companies will make sound investments and further differentiate themselves in the marketplace, while others are at risk of not acting, or delaying action, and potentially falling behind their competition.

Each organization will need to define how it is going to approach the big data technologies, what capabilities will be needed, and how to make the right investment trade-offs in order to achieve specific strategic objectives. The common imperative for all organizations, whether it is a ship owner, operator, manager, shipyard, technology company, or OEM, is to begin addressing this now and not be left behind as others move forward.

Bibliography

1. www.Inmarsat.com

2. www.wikipedia.com

3.General Electric press release, June 18, 2013

4. ESRG analysis based on multiple data sources, including: Lloyd's Register, IMO,

5. AlphaLine, World Shipping Organisation, Clarkson Capital Markets, MADiesel

Engine Outlook 2010, proprietary research

6. HANDOOP: THE ENGINE THAT DRIVES BIG DATA by LARS NILSEN

7. VSEA Warfare Centers Philadelphia

8. L-3 SAM Electronics GmbH press release October 15, 2012

9. Ground Control Global Satellite Internet Solutions