18DCSOOT RUDRA BARAD DSA UNIT FEST - I 03/09/2021



81.	Formula for covariance $(ov(x,y) = \Sigma_{i=1}^{n}(x_i - x_i)$) (41-	· ÿ)
	n-		
	Mean of $x = (98+87+90+85+95+75)$ 6 = 88.3	3	
	Mean of y = (15+12+10+10+16+7) 16 = 11-67		
	Subtract each value from its respective med	8 nr	then
	multiply these new values together		
		RODUCT	
	$(x-x) \qquad (y-y) \qquad (x-x)$	-ド)(メー	(F
	9.67 3.33	32.20	
		-0.44	
	1.67 -1.67	-2.79	
	-3.33 -1.67	5.56	
	6.67 4.33	28.88	
	-13.33 -4.67	62.25	
	Dalalia and Han		
	Adding all the products together it gields	the	
	value 125.66		
	Final step divide by (2-1) - (1		
	Final step, divide by (n-1) = 6-1 = 5		
	: 125.66 5 = 25.132		
	123.00 3 23.132		

18 DCS007 RVDRA BARAD



- 82. HADOOP is a software framework from apache software foundation that is used to store & process big data.
 - It is like a platform on a suite which provides voorlous services to solve the big bala problems.
- > In includes open source projects as well as a complete mange of one complementary tools.
- -> Major component of Madoop include HDFS, VARM, MADOOP COMMON, HIVE, etc-
 - HDFS HDFS stands for Madoop distributed file

 system. It is major component of Madoop

 ecoystm & is responsible for storing large

 data sets of soft structured on unstructed

 data access various node & there by maintaing

 the metadata in form of by log files

WORKING OF HDES

- -> It enables napled toransfor of data between nodes.
- -> It is closely coupled with map neduce 8 it onganizes 8 condenses the nesult into a cohesive
 - answers to query
- them to different nodes in cluster.



>	It keeps the track of where file data is kept		
idotugo	inothe sociasterior eli il almongramo betarpetai		
	engine that can schedule, distribute and month		
	MDES ARCHITECTURE enotholiggo slaiting		
200	NAME HODE JOHN JOHN HELD STALL		
relati	cone functionality		
	BLOCK OPS		
NACTO!	SPARK SEL - H is buil on top of spank (THEND)		
	Support for shurtwird data		
n	DATA MODES DATA MODES		
-			
MI BI			
	lainer soicher deuriford saidom) tothe		
	nock1 LCUENT nock 2		
adopeu	Merephy - it is library that is used to manipulate of		
	material planto graph parelled &		
	Fraturies of Spank and		
	- lightning fact processing speed		
	→ ease of use		
	enissiona monde smit Insie		
	oldixol (
	appliance belondeldage real traggue andle college		
	dianamos buspasas sandro e		



23 SPARK ELOSYSTEM consists of different types of tightly integrated components. At its cone, spank is a computational engine that can schedule, dist nitute and monitor multiple applications. Hautifully 270H SPARK CORE - It is heart of spark and performs the core functionality SPARK SAL - It is built on top of spank come It provide support for structured data SPARK STREAMING- It supponts scalable and faulttolerant processing of streaming data MLIB - It is machine library that contains various machine learning algorithms GRAPHX - It is library that is used to manipulak graphs 8 perform graph parallel computations. Features of Spank are: -> Lightning fast processing speed ease of use -> real time stream processing -> flexible offines support for sophisticated analysis -> active & expanding community