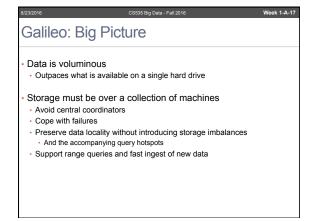


About me

Research area
Big Data
Storage, retrievals, analytics and visualization

Projects
Synapse
A Framework for Ad Hoc Model Construction in Data Streaming Environments
Galileo
Distributed data storage system for large scale geospatial time-series datasets
Columbus
Visual Analytics, cloud based workflow engine
Mendel
Genomic data storage system

Research group meeting
1:00PM ~ 2:00PM every Friday



Galileo: key features

Support for large numbers (10<sup>9</sup>) of small files

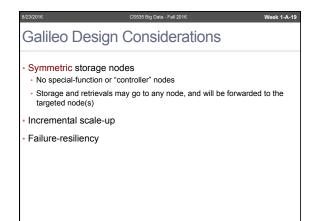
High throughput storage and retrieval

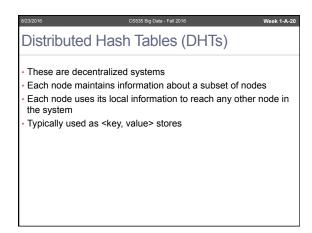
Data is multidimensional with multiple *types*Time-series data

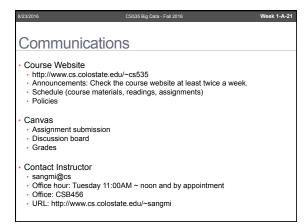
Support for exact match and range queries (with wildcards) along multiple dimensions

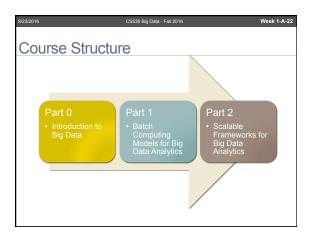
Support for multiple data formats

netCDF, BUFR, HDF 4/5, and data from the Defense Meteorological Satellite Program



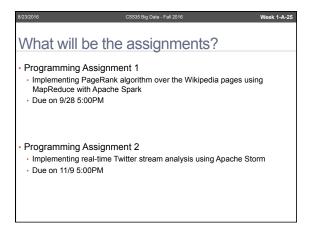




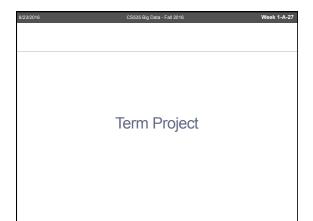


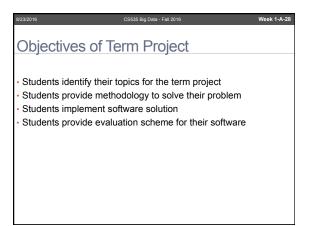
<sup>12016</sup>	crading Policy				
	ategory	Percentage of Final Grade			
Р	rogramming Assignments	40 % Assignment 1: 20% Assignment 2: 20%			
Te	erm Project	30 % Deliverable 0: 1% Deliverable 1: 5% Deliverable 2: 18% Deliverable 3: 6 %			
Q	uizzes and participation	10 %			
M	lidterm Exam	20%			

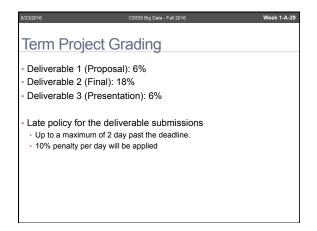


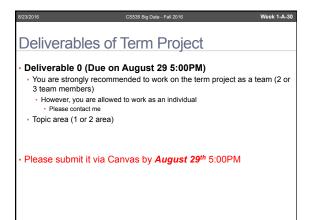


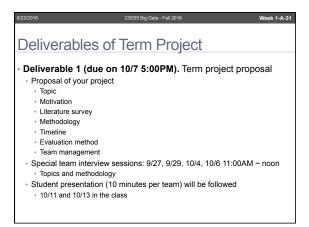
8/23/2016	CS535 Big Data - Fall 2016	Week 1-A-26
How do I s	ubmit the assignme	nts?
	mming assignments are individual	dual submission
	uld be via canvas	
<ul> <li>Late policy for the</li> </ul>	ne assignment submissions	
<ul> <li>Up to a maximun</li> </ul>	n of 2 day past the deadline.	
<ul> <li>10% penalty per</li> </ul>	day will be applied	
<ul> <li>Each student wi in CSB120</li> </ul>	II provide demo of the program	ming assignment
<ul> <li>Each assignmer</li> </ul>	nt will count 20% of total score	of this course
_		

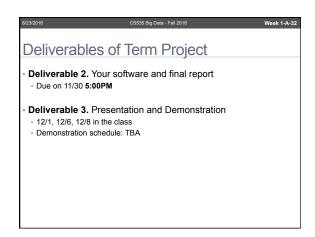


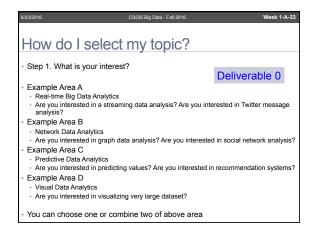


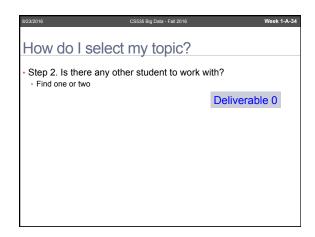


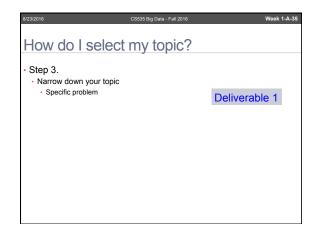


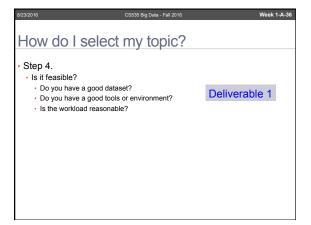






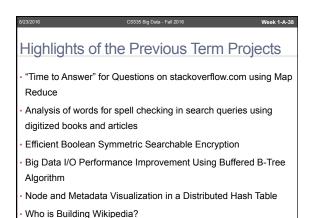


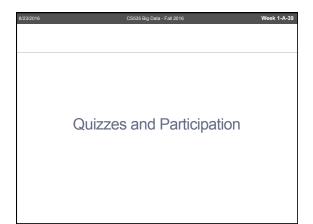


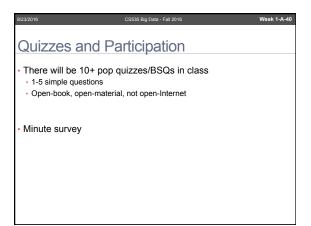


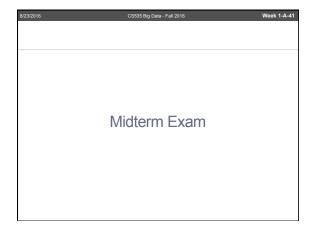
## Highlights of the Previous Term Projects

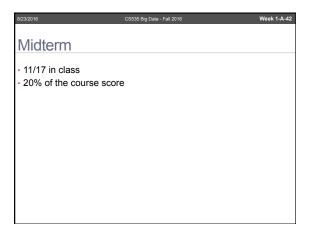
- Supporting Emergency Response During Natural Disasters with Twitter Data
- · Winning Words in the Supreme Court
- Mendel: A Distributed Storage System for Efficient Sequence Alignment and Similarity Searching (published in IEEE IPDPS 2016)
- Processing Smart Grid Data In Real Time (DEBS grand challenge 2014)

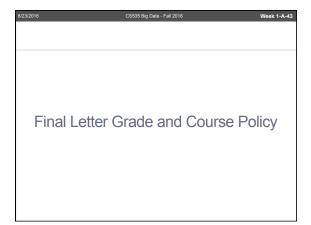


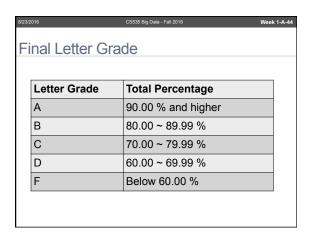












eek 1-A-45
ce to

Course Policy

No Cell-phones in the class.

No Laptops in the class.

If you need to use a laptop during lectures, please sit in the back row.

I will ask you to turn off your laptop if needed.

8/23/2016	CS535 Big Data - Fall 2016	Week 1-A-4
More im	portantly,	
• Δttend the c	class, ask questions, and discuss	
	course web page and canvas regularly	
	hnologies and apply them	
• Share your	experiences with other students in class	S

