

SE ZG522: Big Data Systems

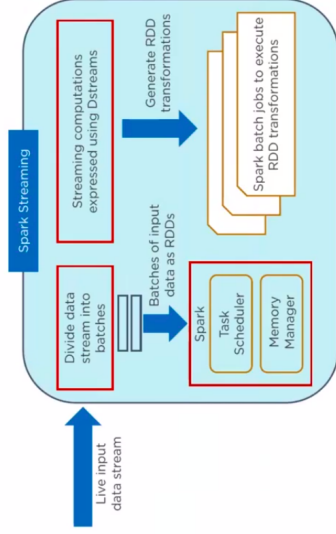
16

Spark Streaming



Dr. Kamlesh Tiwari
Assistant Professor, Department of CSIS,
BITS Pilani, Pilani Campus, Rajasthan-333031 INDIA
May 15, 2022 **ONLINE** (WILP @ BITS-Pilani Jan-May 2022)
<http://ktiware.in/bds>

Working



DStreams: Discretized Streames
Transformations: map(func), flatMap(func), filter(func),
union(otherStream), transform(func), count(), join()

Thank You!

Spark Streaming

- Extension enabling scalable, high-throughput, fault tolerant stream processing
- Inout data stream goes to streaming component the breaks in batches which subsequently processed using engine.



Input could come from Twitter, Kafka, mangoDB, MySQL etc...

Windowing and checkpointing

- Windowing
- Checkpointing for failure resilience Metabdata Level Data Level
- Accumulators or shared variables
- Broadcast variables
- Demo 1

¹ <https://www.youtube.com/watch?v=JUrREnqghRM>

Thank you very much for your attention!