

# Big Data, Real-time Processing and Storm

Prashanth Babu

<http://About.Me/Prashanth>

Please let me know your feedback and / or comments on this Outline for the Workshop.

# Outline of the session

- Big Data
- Batch processing
- Real-time processing
- Real-time vs. Batch processing
- Tools / frameworks
- Complex Event Processing
- Event Stream Processing
- CEP vs. ESP
- Storm overview
- Use cases of Storm
- Comparison with other open source Big Data solutions
- Storm vs. Hadoop

# Outline of the session

- Storm Dependencies

- Storm Architecture

- Storm Components

  - Nimbus

  - Supervisor

  - ZooKeeper

- Storm Concepts

  - Tuples

  - Streams

  - Spouts

  - Bolts

  - Topologies

# Outline of the session

- Storm UI
- Modes of operation
  - Local Mode
  - Production Mode
- Storm through an example
  - Word Count demo
- Trident
- How the combination of Storm and Hadoop can help analyze data
- Other features & Community Support
- Real-time analysis of tweets with Storm
- References
- Q n A

# Storm Overview

- Developed at BackType by Nathan Marz and team
- Open sourced by Twitter [after acquiring Backtype] in August, 2011
- Stream Processing
- Distributed RPC
- Continuous computation
- Guaranteed data Processing
- Fault tolerance
- Horizontal scalability
- CEP
- Further discussion in the session. 😊