Streaming and Steering Applications: Requirements and Infrastructure Summary

Overview:

We propose a workshop covering a class of applications -- those associated with streaming data and related (near) real-time steering and control – that are of growing interest and importance. The goal of the workshop is to identify application and technology communities in this area and to clarify challenges that they face. We will focus on application features and requirements as well as hardware and software needed to support them. We will also look at research issues and possible future steps after the workshop. We will produce a report on workshop discussions that will be delivered in 2015 based on a workshop held in Indianapolis. Two full meeting days (October 1-2) will be followed by a report writing day on October 3, 2015. The proposed budget covers travel and meeting expenses for 35 attendees. The meeting will be streamed with live video and screen sharing.

Intellectual Merit:

We have surveyed the field and identified application classes including Internet of People, wearables, Social media, Twitter, cell phones, blogs, financial transactions, Industrial Internet of Things, Cyberphysical Systems, Satellite and airborne monitors, National Security, Justice, Military, Astronomy, Light Sources, and Instruments like the LHC, Sequencers, Data Assimilation, Analysis of Simulation Results, Steering and Control. We also survey technology developments across academia and Industry where all the major commercial clouds offer significant new systems aimed at this area. We have identified an organizing committee expanding the core group – Fox (Indiana), Jha (Rutgers) and Ramakrishnan (LBNL) proposing this workshop. We have given a preliminary list of over 25 experts who are potential attendees. We argue that this field needs such an interdisciplinary workshop addressing the big picture: applications, infrastructure, research and futures.

Broader Impacts:

In selecting list of attendees we will reach out to underrepresented communities; in particular women and ethnic minorities. An initial list of attendees includes 8 women and 2 ethnic minority institutions.and we will aim to broaden attendee characteristics as we prepare final list. Real time streaming of sessions will enhance opportunities for a broad community to engage at the meeting and we will support questions and comments from remote participants. A web site will be established to support this workshop. It will contain the final report, presentations, position papers, archival copies of streamed video and a repository of useful documents and links.

Streaming and Steering Applications: Requirements and Infrastructure.

1. Introduction

This proposal is to support a workshop covering a class of applications -- those associated with streaming data and related (near) real-time steering and control. The goal of the workshop is to identify application, infrastructure and technology communities in this area and to clarify challenges that they face. We will focus on application features and requirements as well as hardware and software needed to support them. We will also look at research issues. In this proposal we cover typical application areas (Section 2) and some approaches to software (Section 3). Section 4 covers broad features of workshop including related (previous) activities, a summary of why it's needed and our plan to make this a sustainable activity within this community. Section 5 covers our plan for recruitment to workshop and suggested attendees. It also covers the report generated by workshop describing findings and identifying future activities building the streaming and steering community. Section 6 discusses broader impact and Section 7 concludes proposal with the appendix in Section 8 giving proposed schedule and venue details.

2. Streaming and Steering Application Areas

In Table 1, we identify eight problem areas that involve streaming data. We argue that the applications of Table 1 are critical for next-generation scientific research and thus need research into a unifying conceptual architecture, programming models as well as scalable run time. All problem areas are actively used today but without agreed software models focused on streaming.

Streaming/Steering Application Class		Details and Examples	Features
1	Internet of People: wearables	Smart watches, bands, health, glasses, telemedicine	Small independent events
2	Social media, Twitter, cell phones, blogs, financial transactions	Study of information flow, online algorithms, outliers, graph analytics	Sophisticated analytics across many events; text and numerical data
3	Industrial Internet of Things, Cyberphysical Systems, Control	Software Defined Machines, Smart buildings, transportation, Electrical Grid, Environmental and seismic sensors, Robotics, Autonomous vehicles, Drones	Real-time response often needed; data varies from large to small events
4	Satellite and airborne monitors, National Security; Justice, Military	Surveillance, remote sensing, Missile defense, Anti-submarine, Naval tactical cloud	Often large volumes of data and sophisticated image analysis
5	Astronomy, Light Sources, Instruments like LHC, Sequencers	Scientific Data Analysis in real time or batch from "large" sources. LSST, DES, SKA in astronomy	Real-time or sometimes batch, or even both. large complex events
6	Data Assimilation	Integrate typically distributed data into simulations to enhance quality.	Link large scale parallel simulations with time dependent data. Sensitivity to latency.
7	Analysis of Simulation Results	Climate, Fusion, Molecular Dynamics, Materials. Typically local or in-situ data	Increasing bottleneck as simulations scale in size.
8	Steering and Control	Control of simulations or Experiments. Data could be local or distributed	Variety of scenarios with similarities to robotics

Table 1: Eight Streaming and/or Steering Application Classes

As we illustrate in Table 1, these applications are not of course new but they are growing rapidly in size and importance. Correspondingly it becomes relevant to examine the needed functionality and performance of hardware and software infrastructure that could support these applications. We can identify such applications within academic, commercial and government areas. Examples in Table 1 include the Internet of Things projected to reach 30 to 70 billion devices in 2020 [1] with particular examples including wearables, brilliant machines [2] and smart buildings; these myriad of small devices contrasts with events streaming from larger scientific instruments such light sources, telescopes, satellites and sequencers. There is the social media phenomena that adds over 20,000 photos online every second [3] with an active research program studying structure and dynamics of information. In National Security, one notable example comes from the Navy which is developing Apache Foundation streaming (big data) software for missile defense [4]. A NIST survey [5] of big data applications found that 80% involved some of streaming [6] and the AFOSR DDDAS initiative [7] looks at streaming and control (steering). Data assimilation and Kalman filters have been extensively to incorporate streaming data into analytics such as weather forecasts and target tracking.

Most of the applications involve linking analysis with distributed dynamic data and can require real-time response. The requirements of distributed computing problems which couple HPC and cloud computing with streaming data, are distinct from those familiar from large scale parallel simulations, grid computing, data repositories and workflows which have generated sophisticated software platforms. Scientific experiments are increasingly producing large amounts of data that need to be processed on HPC and/or cloud platforms. These experiments often need support for real-time feedback to steer the instruments. Thus, there is a growing need to generalize computational steering to include coupling of distributed resources in real-time, and a fresh perspective on how streaming data might be incorporated in this infrastructure. The analysis of simulation results or visualizations has been explored significantly in last few years and is recognized to be a serious problem as simulations increase their performance towards exascale. The in-situ analysis of such data shares features with streaming applications but the data is not distributed if simulation and analysis engines are identical or co-located.

One goal of the workshop will be to identify those features that distinguish different applications in the streaming/steering class. Five categories we have already identified are:

a) **Set of independent events where precise time sequencing unimportant.** *e.g. independent search requests or smartphone or wearable cloud accesses from users.*
b) **Time series of connected small events where time ordering important.** *e.g. streaming audio or video; robot monitoring.*
c) **Set of independent large events where each event needs parallel processing with time sequencing not critical** *Example: processing images from telescopes or light sources with material science.*
d) **Set of connected large events where each event needs parallel processing with time sequencing critical** *e.g. processing high resolution monitoring (including video) information from robots (self-driving cars) with real time response needed.*
e) **Stream of connected small or large events that need to be integrated in a complex way.** *e.g. streaming events being used to update model (e.g. clustering) rather than being classified with an existing static model which fits category a).*

These 5 categories can be further considered for single or multiple heterogeneous streams. we will refine and expand these categories as part of workshop

3. Software Models for Streaming and Steering

Although the growing importance of these application areas has been recognized, we see that the needed hardware and software infrastructure is not as well studied. Particular solutions such as for the analysis of events from the LHC or imagery from telescopes and light sources have been developed.

The distributed stream processing community has produced frameworks to deploy, execute and manage event based applications at large scale and these are one important class of streaming and steering software. Examples of early event stream processing frameworks include, Aurora[8], Borealis[9], StreamIt[10] and SPADE[11]. With the emergence of Internet Scale applications in the recent years, new distributed map-streaming processing models like, Apache S4[12], Apache Storm[13], Apache Samza[14], Spark Streaming[15], Twitter's Heron[16] and Granules[17] with commercial solutions including Google Millwheel[18] Azure Stream Analytics[19] and Amazon Kinesis[20] have been developed.

Although these academic and commercial approaches are effective, we suggest a more integrated approach that spans many application areas and many solutions and evaluates applications with current and future software. This could lead to new research directions for a scalable infrastructure, and clearer ideas as to appropriate infrastructure to support a range of applications. Note in the grid solutions for problems like LHC data analysis, events tend not be streamed directly but rather batches are processed on distributed infrastructure. In Table 2 below, we contrast some well know scientific computing paradigms with streaming and steering.

Paradigm		**Features and Examples**
1	Multiple Loosely Coupled Tasks	Grid computing, largely independent computing/event analysis, many task computing
2	MapReduce	Single Pass compute and collective computation.
3	BSP and Iterative MapReduce	Iterative staged compute (map) and computation includes parallel machine learning, graph, simulations. Typically Batch
4	Workflow	Dataflow linking functional stages of execution
5	Streaming	Incremental (often distributed) data I/O feeding to long running analysis using other computing paradigms. Typically interactive
6	Steering	Incremental I/O from computer or instrument driving possibly real-time response (control)

Table 2: Six Computing Paradigms with Streaming and Steering contrasted with four other paradigms common in scientific computing.

In the first four paradigms of the above table, data is typically accessed systematically either at the start of or more generally at programmatically controlled stages of a computation. In workflow, multiple such data-driven computations are linked together. On the other hand, the streaming paradigm absorbs data asynchronously throughout the computation while steering feeds back control instructions.

Identifying research directions will be one of the goals of the workshop. We can already identify the need to study the system architecture including balance between processing on source, fog (local) and cloud (backend), online algorithms, storage, data management, resource management, scheduling, programming models, quality of service (including delay in control responses) and fault tolerance. Optimizations like operator reordering, load balancing, fusion, fission etc. have been researched to reduce the latency of the stream processing applications [21].

4. Workshop Goals Objectives and Organization

4.1 The Four Workshop Goals

The purpose of this workshop is to explore the landscape sketched above and identify the application and technology communities and converge on the immediate and long-term challenges. We propose a workshop that will examine four aspects of this landscape:

- *Application Study:* Table 1 is a limited sampling of applications that critically depend upon Steering HPC. It is necessary to extend and refine Table 1 with broader set of application characteristics and requirements. We need to improve the set of features at the end of Section 2 and identify which aspects are important in determining software and hardware requirements. Perhaps a set of benchmarks will be important
- *System Architectures:* A critical challenge that follows is to understand scalable architectures that will support the different types of streaming and steering applications in table 1, i.e., firm up the vague concept of "ubiquitous sensors and Internet of Things" to match the range of applications types and infrastructure types. In particular we should identify where HPC, accelerators, and clouds are important.
- *Research Directions:* There is a need to integrate features of traditional HPC such as scientific libraries and communication with the rich set of capabilities found in the commercial streaming ecosystem. This general approach has been validated for a range of traditional applications, but not for rich class of streaming and steering problems. Interesting questions are centered on the data management requirements while the NRC study [22] stressed the importance of new online (streaming -- look at each data point once) algorithms.
- *Next Steps Forward:* We hope this workshop starts a process that will identify and bind the community of applications and systems researchers and providers in the streaming and steering areas. We intend a thorough report with the final day of workshop devoted to writing this. As well as covering findings of workshop, the report will suggest next steps forward. These could include a second workshop to dig deeper in some areas, and other studies such as collection of benchmarks to move us forward

4.2 Previous Events

We are not aware of any meeting in this area that juxtaposes infrastructure, applications and systems (software). There are many Internet of Things workshops and conferences – the online list WikiCFP [23] for example lists 66 IoT events with 11 still open this year. Robotics at this site has 18 open and Sensor Networks 13. These meetings would not attract the interdisciplinary mix we aim at in our proposed workshop. DDDAS meetings [7] also cover some topics proposed here.

This workshop will be unique in that it will focus on understanding big picture as opposed to discuss specific solutions. It will also bring together resource and infrastructure providers with academic community.

4.3 Organizing Committee

The core organizing committee consists of Geoffrey Fox (Indiana University), Shantenu Jha (Rutgers) and Lavanya Ramakrishnan (LBNL). These three have worked together over the last six months to understand workshop area and produced a report on HPC Streaming for DoE [24]. Fox has worked extensively on streaming problems for last 15 years starting with the publish-subscribe system NaradaBrokering [25] and now focused on cloud control of robotics [26]. Jha has extensive experience in computational steering, analysis of large scale simulations and distributed computing and middleware. Ramakrishnan workflow research has covered several DoE streaming applications.

If approved, we will expand the organizing committee with 3-5 others considering for example George Djorgovski or Alex Szalay for astronomy, an LHC experiment representative, Tom Hacker for NEES, Roger Barga (Amazon) for commercial, Shrideep Pallickara (Colorado State) for software, Maryam Rahnemoonfar (Texas A&M) for drones and minority outreach, Madhav Marathe (Virginia Tech) or Alessandro Flammini (Indiana University) for Network Science.

5. Workshop Recruitment and Possible Participants

5.1 Recruitment, Outreach to Community

The relevance and long-term impact of the workshop outcomes are critically dependent on the participation in this workshop by a group of community members that represent a broad and diverse range of stakeholders.

We will ensure that the participants cover as many of the sub disciplines and groups in the broad area of cyberinfrastructure as possible. We propose a two-prong strategy to broaden participation and access to this workshop. We have assembled a preliminary list of invitees that includes persons identified by the PIs. We will work together with several NSF program managers to finalize the initial round of invitees to ensure a good mix of disciplines and experience, targeting as many early career scientists as possible and members from underrepresented groups. We will expressly leave a portion of the workshop attendees unassigned to allow those we have not considered or are outside of our peer networks to gain access.

As seen from tentative list of participants, we will spread the list of invitees from a broad range of sub-disciplines including both early career as well as more senior and accomplished investigators. Other criteria for selection of participants by the organizing committee will include research interests and potential contributions to the workshop. This information will be gathered from all workshop applicants as part of the open registration process. We will use XSEDE, OSG mailing lists as well as the standard HPDC conference & workshop announcement lists, such as hpc-announce, distributed-systems etc.

The workshop will host approximately 35 in-person participants. Although we will invite and support participants from across the country to attend in person, we will also invite community members to participate virtually. We will provide web-based access to the plenary presentations and major breakouts for remote participants via screen sharing and video streaming, and we will try to establish virtual participation in the breakout sessions. We will have a special outreach activity to early career scientists (newly established faculty and post-doctoral researchers) as well as graduate students. We will look for appropriate connection to recent NSF ACI CAREER awardees.

We will ask all participants for a 1-2 page vision statement as a pre-condition for a formal invitation and access to registration to attend. These statements will be shared and distributed to all participants. The organizers will read and review these statements.

5.2 Potential list of participants organized by Project and Discipline

We have identified several potential invitees based on projects, individual expertise and organization. We will chose the attendees to cover most of application categories in Table 1. One set of possible invitees is given below and this doesn’t fully cover Table 1 and we can think of substitutes in listed areas. The expertise below is very broad and currently we intend to choose after consultation with NSF a subset of these topics. We intend a second workshop to follow up issues raised in first and broaden topics. This second event is not funded in this proposal. We also have not listed the three core organizers (Fox, Jha, Ramakrishnan) and local (Indiana and Purdue particularly) faculty who will participate. Also note some invitees are not expected attendees but rather senior contacts who can suggest suitable attendees. We will supplement this list as described in Section 6 with outreach to under-represented communities.

Possible Invitee	Institution	Expertise/Project
Research Applications		
Deb Agarwal	LBNL	Ameriflux And Fluxnet environmental sensors
Charlie Catlett	Chicago	Array of Things [27]
George Djorgovski	Caltech	Astronomy (Sky surveys, CRTS, LSST, DES)
Tom Hacker	Purdue	Earthquake Engineering, NEES
Kris Hauser	Duke	Robotics; planning and control
Linda Hayden	ECSU	Remote Sensing ans UA
Kerstin Kleese Van Dam	PNNL	Analysis in Motion, Chemical Imaging
Madhav Marathe	Virginia Tech	Network Science
Richard Mount	SLAC	LHC physics, ATLAS
Maryam Rahnemoonfar	Texas A&M	Drones at FAA site [28]
Alex Szalay	JHU	Astronomy, Simulation Visualization, Sensors
TBD	Light (5) and Neutron (2) Sources at DoE laboratories	
Technology		
Alok Choudhary	Northwestern	Data mining and HPC
Jack Dongarra	UTK/ORNL	BDEC. Big Data and Exascale
Dennis Gannon	Independent	HPC and Clouds
Scott Klasky	ORNL	Simulation Visualization, ADIOS
Shrideep Pallickara	Colorado State	Streaming software and health applications
Ioan Raicu	IIT	HPC for Big Data
Karsten Schwan	Georgia Tech	High performance QoS sensitive streaming
XiaoFeng Wang	Indiana Univ.	Systems Security
Research Infrastructure		
PSC, SDSC, TACC, UTK, SP-Forum		XSEDE Service Providers
Pete Beckman	ANL	DoE Hardware and Software
Kate Keahey,Robert Ricci	Utah, Chicago	NSF Cloud: Chameleon, CloudLab
Ruth Pordes	Fermilab	Open Science Grid
John Towns	NCSA	XSEDE
Industry Applications and Infrastructure		
Roger Barga	Amazon	Kinesis Real Time Stream Processing
Bobby Evans	Yahoo	Apache Storm Project Management Committee
TBD	Google	Brillo (IoT) MillWheel (Streaming) [29]
TBD	Microsoft	Azure Event Hubs, Stream Analytics
Bill Ruh, VP GE Software	General Electric	IIoT, Software Defined Machines and Prefix software
Youngchoon Park	Johnson Controls	Internet of Buildings
Government		
Many Program Managers	NSF	PM's covering fields in Table 1 + CISE including Big Data (Chaitan Baru)
Richard Carlson	DoE	DoE Instruments, Visualization
Wo Chang	NIST	Big Data Public Working Group
Frederica Darema	AFOSR	Dynamic Data-Driven Application Systems (DDDAS)
David Horner (John West)	DoD	HPC Modernization Program (ERDC)
Piyush Mehrotra	NASA	Big Data Initiative
TBD	ONR	Navy Tactical Cloud (built on Storm)

Table 3: Possible Invitees to Proposed Workshop on Streaming and Steering

5.3 Preparation of Workshop Report

Breakout (working) groups will be asked to collaboratively author their reports in real time via shared collaborative tools (probably Google Documents), which allow multiple users to view and edit a document simultaneously, while saving and tracking edits by user. The breakout reports will be presented to the plenary, and made accessible online to all breakout groups for further discussion and edits. We will video record all major sessions of the workshop.

All participants will be encouraged to stay for the 3rd writing day to refine notes, synthesize main findings and formulate key report sections. A pre-workshop organizational conference call will select track and theme leads (who will double as editors). The writing team, comprised of the organizers, track and theme leads will be required to stay.

The writing team/editors will continue to engage after the workshop to finalize the report. We will deliver a draft final report within 30 days of the workshop. Whereas a bulk of the writing will occur on Day 3, the editors will meet via a remote conferencing system within 30 days of the workshop to prepare a final draft of the workshop report and findings. We have found this to be an effective pathway from the immediate aftermath of a workshop to a quick report.

The draft report will be disseminated to all workshop participants and posted on the workshop web site; it will be distributed on mailing lists such as XSEDE, OSG, DOE welcoming and soliciting comments and feedback within a 45-day timeframe. We will thus deliver a final report to NSF 90 days from the workshop.

The report will be a live document e.g., arXiv repository, with the main material and essentially complete first version, but one that is updated with incremental refinements. Taking advantage of the live document, in addition to bringing the community to the report, we will examine the possibility of taking the draft of the report to the community, whilst respecting the time constraint.

6. Broader Impact of this proposal

In selecting list of attendees we will reach out to underrepresented communities; in particular women and ethnic minorities. The latter are represented currently by Elizabeth City State University (Linda Hayden) and Texas A&M, Corpus Christi (Maryam Rahnemoonfar), while 8 women are in list of Sections 4.3 and 5.2. In preparing final list we will aim to broaden attendee characteristics. We noted in Section 5.1 that we will use the real time streaming of sessions to enhance opportunities for broad community to engage at the meeting.

As described in data management plan, a web site will be established to support this workshop. It will contain the final report, working group reports, presentations, position papers, archival copies of streamed video and a repository of useful documents and links.

7. Conclusions: Workshop Impact

Streaming data and steering are well established fields but just as data turned into a deluge with profound impact, now with the Internet of Things and new experimental instruments, we see a streaming deluge requiring new approaches to control or steering. This workshop will bring together interdisciplinary experts on applications and infrastructure to address the three conceptual goals of what are the driving applications, what are actual and needed hardware and software and what are research challenges. The community identified for this workshop needs to work together on an ongoing basis and this will come out from the fourth "futures" goal of workshop. We are not aware of any closely related activity and suggest the streaming deluge can only be addressed by a set of activities such as those proposed here.

8. Appendix: Workshop Schedule and Venue

The meeting is proposed to be held from September 30 to October 2 in Indianapolis at the IUPUI (combined Indiana University, Purdue Indianapolis campus) using their event facilities [30] which are located in the center of campus which is itself in downtown Indianapolis. It is an easy (14 mile) taxi ride from Indianapolis airport and near many downtown hotels including Marriott (nearest), Hyatt and Hilton. We have available the Tower Ballroom (see picture below with “oval” seating style) with seating for 60 in conference style and breakout rooms. The rooms are equipped with video conferencing/streaming presentation support.

We will provide lunch and refreshments (coffee) to the participants plus a reception on the evening of September 30.

The meeting is organized as two days (October 1, 2) for main discussions plus a final day (October 3) for organizers to work on meeting report. We only provide two small rooms on the final day to support the 6-10 people expected to attend that day.

The meeting is organized around four goals described in Section IV: Application Study, Systems Architecture, Research Directions, Next Steps Forward with the first two goals covered on day one (September 30) and the second two goals on day two. A proposed schedule is given below.

Note that we will be streaming sessions and questions and comments will be solicited from those attending remotely.

Day One Morning: Introduction and Plenary on Architectures and Systems

- Attendees Introduction: 2 slide presentations by those not on panels
- Application Requirements Panel and discussion
- System Architecture Panel and discussion

Day One Afternoon: Breakout Sessions

- Breakout Sessions: Application Requirements and System Architecture
- Plenary Summary

Day Two Morning: Plenary on Research Directions and Next Steps Forward

- Recap and lessons from Day One
- Research Directions Panel and discussion
- Next Steps Forward Panel and discussion
- Breakout Sessions: Research Directions and Next Steps Forward

Day Two Afternoon: Breakout Sessions and Planning

- Breakout Sessions: Research Directions and Next Steps Forward continued
- Plenary Summary
- Plenary discussion of findings in all four goals
- Organize report writing and discussion of follow up activities

Day Three: Report Writing Day

Make as much progress as possible with workshop report

NSF funded conferences are required to address child care services. These are available to our workshop attendees through “Sitters to the Rescue” established in 1996 with good credentials. The charge is $20 per hour per sitter. If needed by any participant, we will rent another room at the IUPUI facility to satisfy this requirement.

The proposed facilities satisfy federal accessibility requirements.

Streaming and Steering Applications: Requirements and Infrastructure
References

1. Cisco Internet Business Solutions Group (IBSG) (Dave Evans). *The Internet of Things: How the Next Evolution of the Internet Is Changing Everything*. 2011 April [accessed 2013 August 14]; Available from: http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf.
2. Chauhan, N. *Modernizing Machine-to-Machine Interactions*. 2014 Available from: https://www.gesoftware.com/sites/default/files/GE-Software-Modernizing-Machine-to-Machine-Interactions.pdf.
3. Kimberlee Morrison. *How Many Photos Are Uploaded to Snapchat Every Second?* 2015 June 9 [accessed 2015 June 15,]; Available from: http://www.adweek.com/socialtimes/how-many-photos-are-uploaded-to-snapchat-every-second/621488.
4. ONR. *Data Focused Naval Tactical Cloud (DF-NTC): ONR Information Package*. 2014 June 24 [accessed 2015 June 15]; Available from: http://www.onr.navy.mil/~/media/Files/Funding-Announcements/BAA/2014/14-011-Attachment-0001.ashx.
5. NIST. *NIST Big Data Public Working Group (NBD-PWG) Home Page*. 2013 [accessed 2015 March 31]; Available from: http://bigdatawg.nist.gov/home.php.
6. Geoffrey C. Fox, Shantenu Jha, Judy Qiu, and Andre Luckow, *Towards an Understanding of Facets and Exemplars of Big Data Applications*, in *20 Years of Beowulf: Workshop to Honor Thomas Sterling's 65th Birthday* April 13, 2015. Annapolis http://dsc.soic.indiana.edu/publications/ogre.v1.2.pdf.
7. Craig Douglas. *DDDAS Dynamic Data-Driven Application Systems Home Page*. [accessed 2015 June 15]; Available from: http://www.dddas.org/.
8. Cherniack, M., H. Balakrishnan, M. Balazinska, D. Carney, U. Cetintemel, Y. Xing, and S.B. Zdonik. *Scalable Distributed Stream Processing*. in *CIDR* 2003.
9. Abadi, D.J., Y. Ahmad, M. Balazinska, U. Cetintemel, M. Cherniack, J.-H. Hwang, W. Lindner, A. Maskey, A. Rasin, and E. Ryvkina. *The Design of the Borealis Stream Processing Engine*. in *CIDR* 2005.
10. Thies, W., M. Karczmarek, and S. Amarasinghe. *StreamIt: A language for streaming applications*. in *Compiler Construction* 2002: Springer.
11. Gedik, B., H. Andrade, K.-L. Wu, P.S. Yu, and M. Doo. *SPADE: the system s declarative stream processing engine*. in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data* 2008: ACM.
12. Neumeyer, L., B. Robbins, A. Nair, and A. Kesari. *S4: Distributed stream computing platform*. in *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on* 2010: IEEE.
13. Anderson, Q., *Storm Real-time Processing Cookbook*. 2013: Packt Publishing Ltd. ISBN:178216443X
14. Kamburugamuve, S., *Survey of distributed stream processing for large stream sources*. 2013. http://grids.ucs.indiana.edu/ptliupages/publications/survey_stream_processing.pdf.
15. Zaharia, M., T. Das, H. Li, S. Shenker, and I. Stoica. *Discretized streams: an efficient and fault-tolerant model for stream processing on large clusters*. in *Proceedings of the 4th USENIX conference on Hot Topics in Cloud Ccomputing* 2012: USENIX Association.
16. Sanjeev Kulkarni, Nikunj Bhagat, Maosong Fu, Vikas Kedigehalli, Christopher Kellogg, Sailesh Mittal, Jignesh M. Patel, Karthik Ramasamy, and Siddarth Taneja, *Twitter Heron: Stream Processing at Scale*, in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*. 2015, ACM. Melbourne, Victoria, Australia. pages. 239-250. DOI: 10.1145/2723372.2742788.
17. Shrideep Pallickara. *Granules Home Page*. 2015 [accessed 2015 JUne 12]; Available from: http://granules.cs.colostate.edu/.
18. Akidau, T., A. Balikov, K. Bekiroğlu, S. Chernyak, J. Haberman, R. Lax, S. McVeety, D. Mills, P. Nordstrom, and S. Whittle, *MillWheel: fault-tolerant stream processing at internet scale.* Proceedings of the VLDB Endowment, 2013. 6(11): p. 1033-1044.

19. Microsoft. *Azure Stream Analytics*. 2015 [accessed 2015 June 12]; Available from: http://azure.microsoft.com/en-us/services/stream-analytics/.
20. Varia, J. and S. Mathew. *Overview of amazon web services*. 2013 Available from: http://docs.aws.amazon.com/gettingstarted/latest/awsgsg-intro/intro.html.
21. Hirzel, M., R. Soulé, S. Schneider, B. Gedik, and R. Grimm, *A catalog of stream processing optimizations.* ACM Computing Surveys (CSUR), 2014. 46(4): p. 46.
22. Committee on the Analysis of Massive Data; Committee on Applied and Theoretical Statistics; Board on Mathematical Sciences and Their Applications; Division on Engineering and Physical Sciences; National Research Council, *Frontiers in Massive Data Analysis*. 2013: National Academies Press. http://www.nap.edu/catalog.php?record_id=18374
23. *WikiCFP: Listing of IOT Conferences and Workshops*. [accessed 2015 June 15]; Available from: http://www.wikicfp.com/cfp/call?conference=IOT.
24. Geoffrey Fox, Shantenu Jha, and Lavanya Ramakrishnan. *Scalable HPC Workflow Infrastructure for Steering Scientific Instruments and Streaming Applications*. 2015 April 14 [accessed 2015 June 15]; Available from: http://dsc.soic.indiana.edu/publications/StreamingHPC-WhitePaper-WorkflowWorkshop.pdf.
25. NaradaBrokering. *Scalable Publish Subscribe System.* 2010 [accessed 2010 May]; Available from: http://www.naradabrokering.org/.
26. Supun Kamburugamuve, Leif Christiansen, and Geoffrey Fox, *A Framework for Real Time Processing of Sensor Data in the Cloud.* Journal of Sensors, 2015. 2015: p. 11. DOI:10.1155/2015/468047. http://grids.ucs.indiana.edu/ptliupages/publications/iotcloud_hindavi_two_column_final_2.docx
27. Charlie Catlett. *Array of Things*. 2014 [accessed 2015 June 15]; Available from: https://arrayofthings.github.io/.
28. *Six FAA UAS Test Sites*. 2014 [accessed 2015 June 15]; Available from: https://www.faa.gov/uas/legislative_programs/test_sites/.
29. Tyler Akidau, Alex Balikov, Kaya Bekiroglu, Slava Chernyak, Josh Haberman, Reuven Lax, Sam McVeety, Daniel Mills, Paul Nordstrom, and Sam Whittle, *MillWheel: fault-tolerant stream processing at internet scale.* Proc. VLDB Endow., 2013. 6(11): p. 1033-1044. DOI:10.14778/2536222.2536229
30. *IUPUI Indianapolis Event Facilities*. 2015 [accessed 2015 June 15]; Available from: http://eventservices.iupui.edu/facilities.asp.