

بسم الله الرحمن الرحيم

داده‌های حجیم



سیما نادری

دانشجوی کارشناسی ارشد

فهرست مطالب

- مقدمه
- نگاشت و کاهش
- Hadoop
- Storm
- Heron
- مقایسه Storm و Heron
- Pregel
- منابع

مقدمه

- داده‌های حجیم دارای سه ویژگی است

- حجم

- سرعت

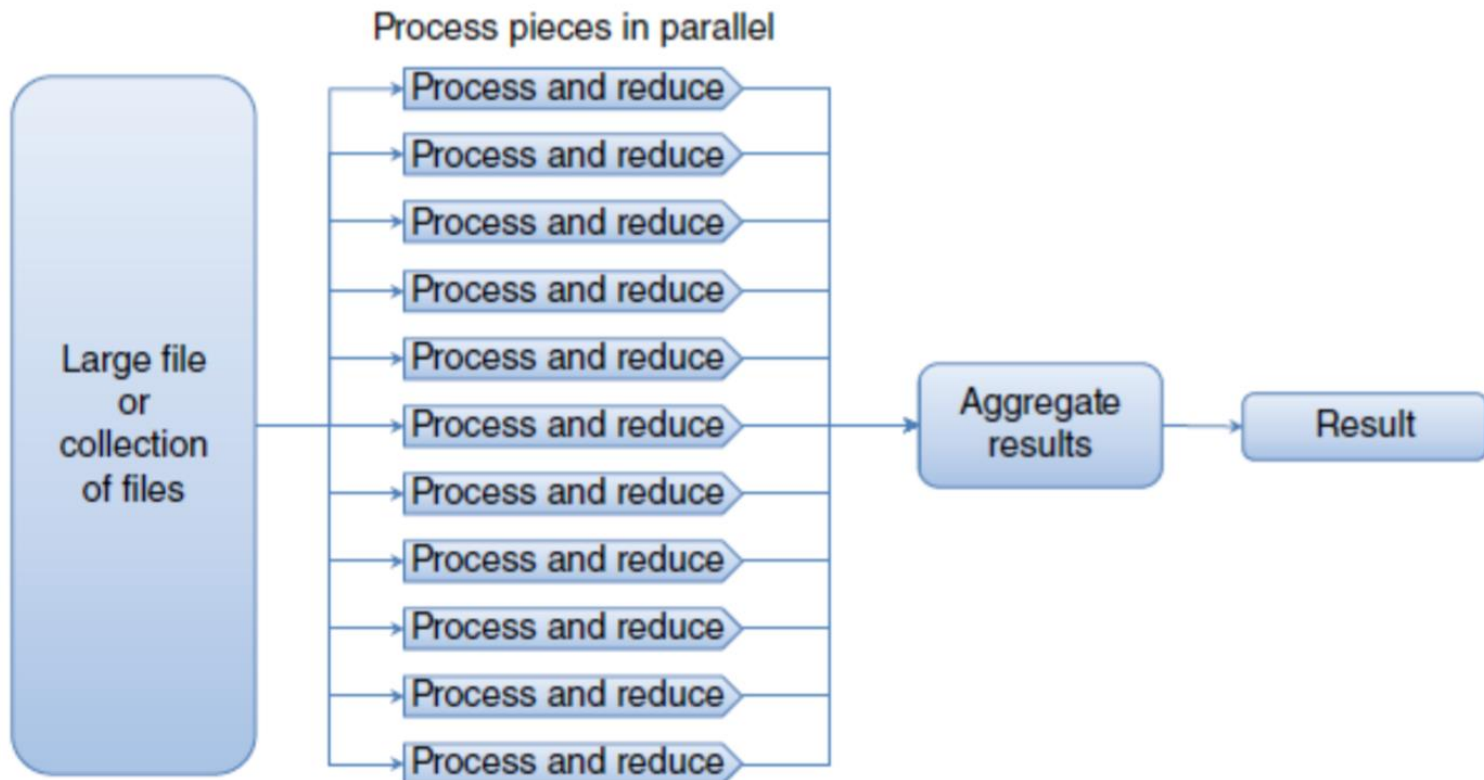
- تنوع

- شرایط رشد

- توانایی ذخیره کردن

- پردازش داده‌ها

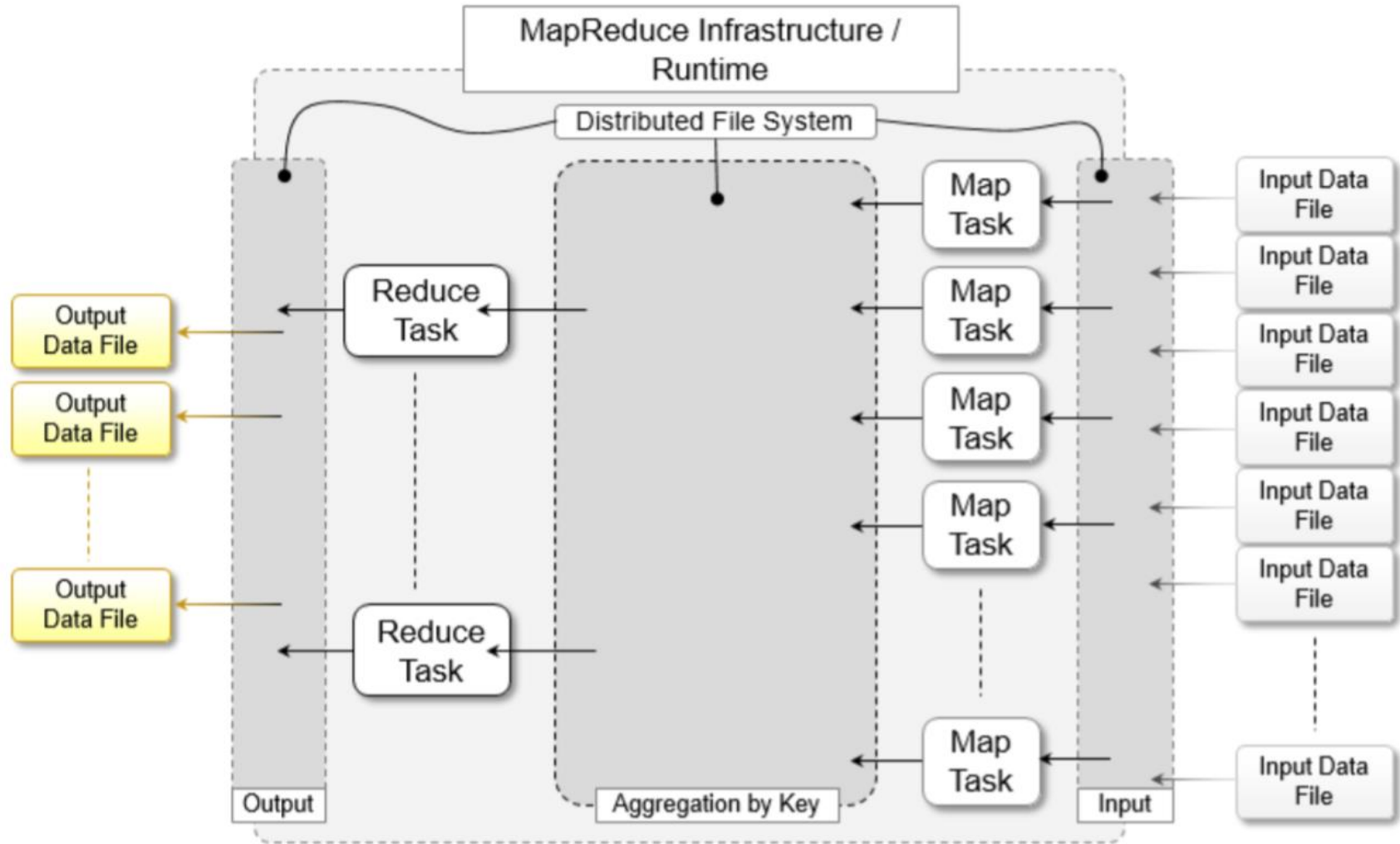
- وجود تولیدکنندگان



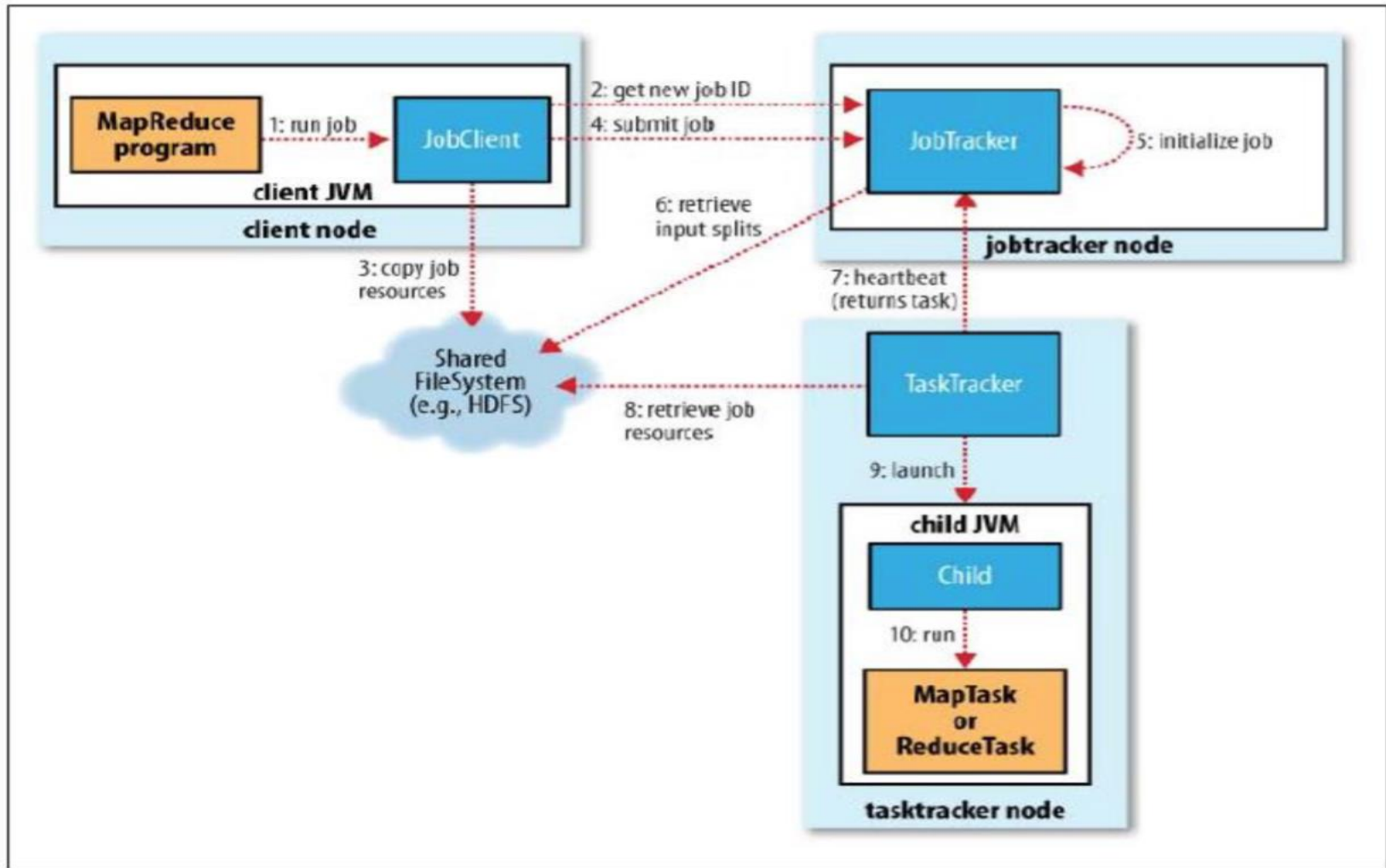
نگاشت و کاهش

- مهمترین مدل برای پردازش داده‌های حجیم
 - برروی Hadoop
 - دو تابع نگاشت و کاهش
 - انجام سایر عملیات توسط خود سیستم
- تابع نگاشت: یک جفت کلید مقدار
- تابع کاهش: یک کلید و لیستی از مقادیر دارای آن کلید

نگاشت و کاهش



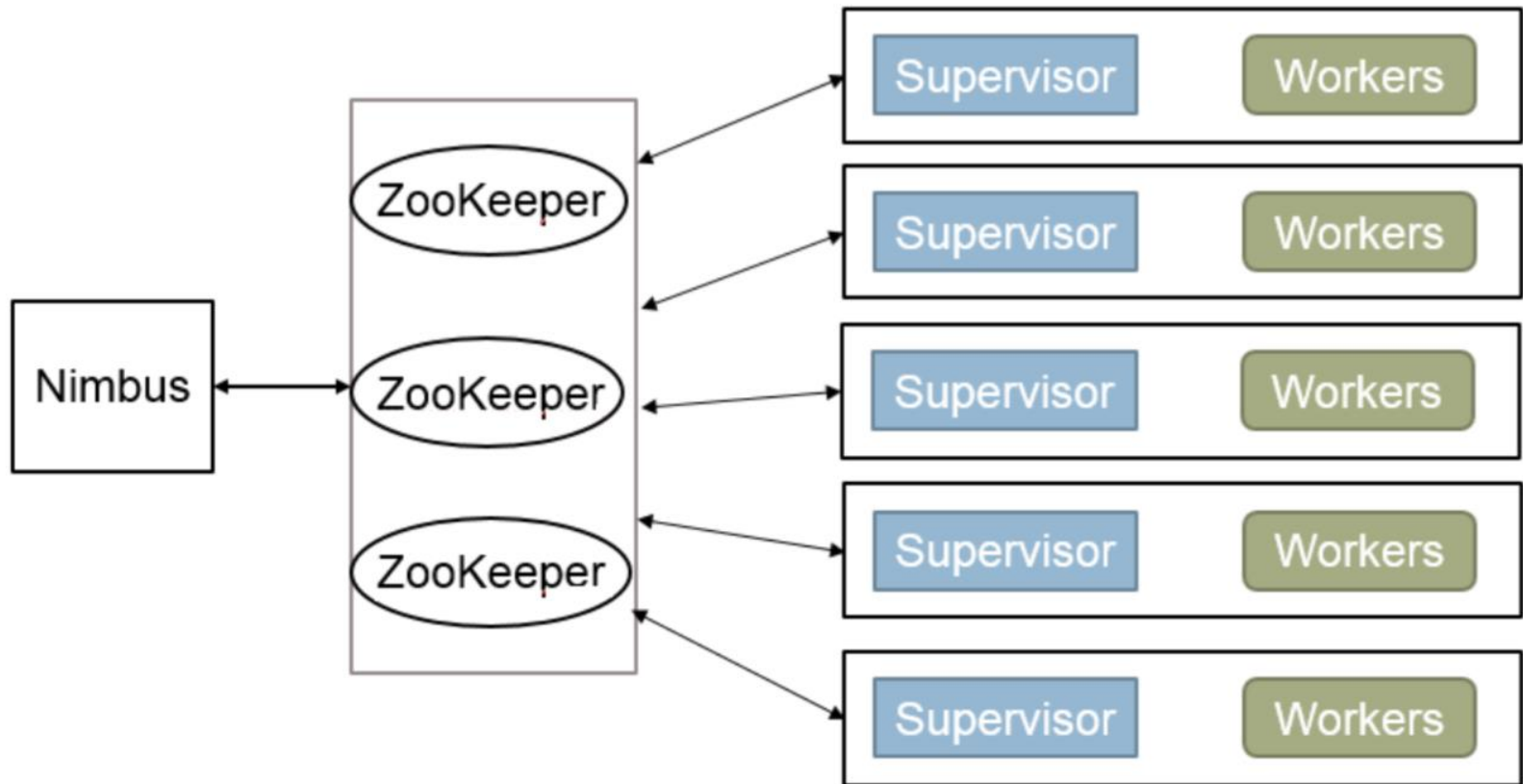
hadoop



Storm

- پردازشگر جریانهای داده بهصورت آنی بر روی سیستمهای توزیع شد
- برنامه *storm* یک جریان کاری است
- ورودی به سیستم جریانی از چندگانه است که به آن توپولوژی می گویند
- Spout: مبدا دادهها
- bolt: پردازش بر روی دادهها

معماری storm

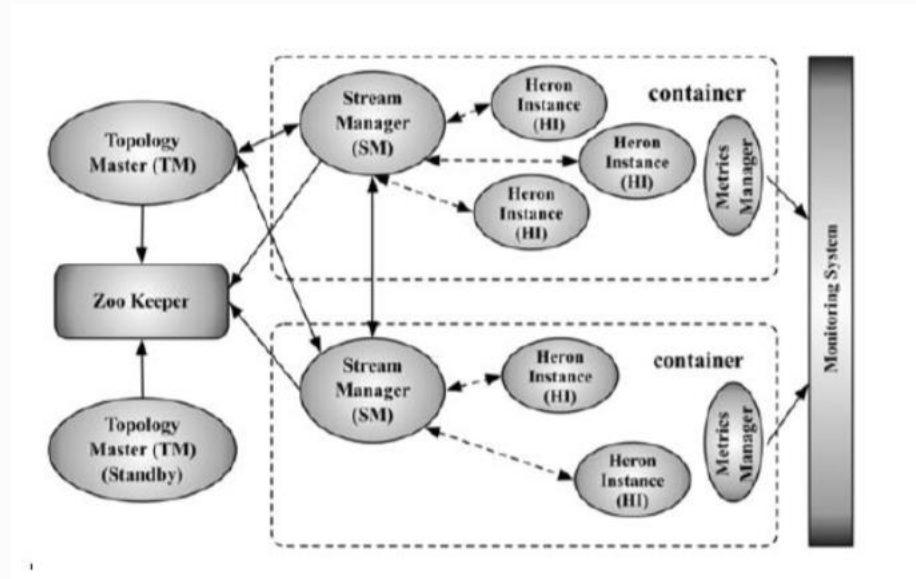


storm

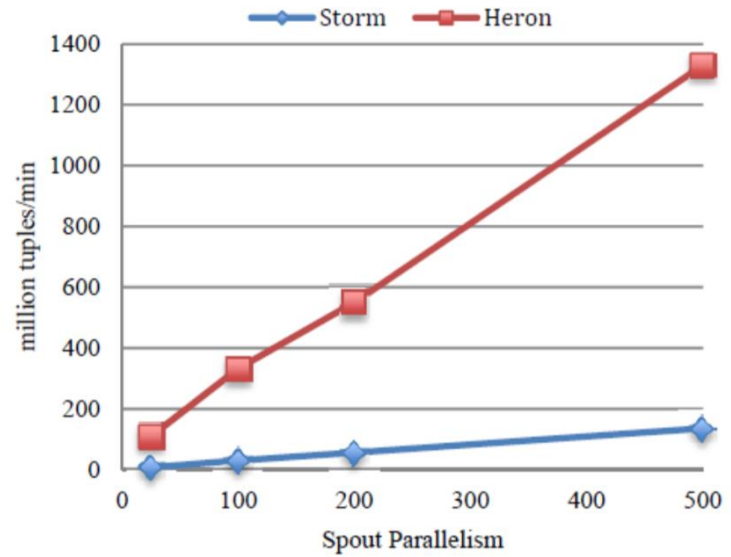
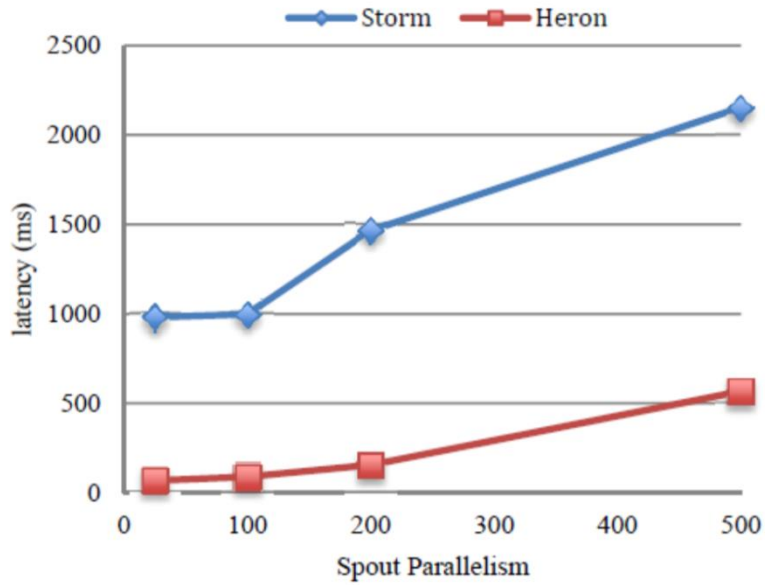
- انواع تقسیم‌بندی در storm:
- گروه‌بندی مختلط: که به صورت تصادفی چندگانه‌ها را بخش‌بندی می‌کند.
- گروه‌بندی موضوعی: که با استفاده از تابع هش بر روی زیرمجموعه‌ای از چندگانه‌ها بر اساس موضوع یا ویژگی بخش‌بندی را انجام می‌دهند.
- گروه‌بندی کل: که کل جریان را برای تمام کارهای مصرف‌کننده تکرار می‌کند.
- گروه‌بندی سراسری: که تمام جریان را برای یک bolt می‌فرستد.
- گروه‌بندی محلی: که چندگانه را به bolt‌های مصرف‌کننده در همان اجراکننده می‌فرستد.

Heron

- مشکلات storm
 - مدیریت
 - رفع اشکال
 - زمانبندی



مقایسه Storm و Heron



Pregel

- مشکلات گراف بزرگ
 - محلی بودن دسترسی به حافظه پایین
 - پردازش در رئوس هر گراف پایین

- Pregel
 - در هر راس یک تابع
 - راس دارای یک شناسه و یک مقدار
 - راس یک موجودیت مستقل
 - لبه‌ها مقدار دارند اما نام ندارند
 - تابع بر روی تمامی رئوس به طور همزمان اجرایی شود
 - هر راس میتواند مقدار خود و لبه‌های مربوط به خود را عوض کند
 - رای به غیرفعال شدن هر راس
 - زمان خاتمه الگوریتم
- که تمامی رئوس رای به غیرفعال شدن خودشان داده باشند

- Kaisler, S., et al. Big data: Issues and challenges moving forward. in System sciences (*HICSS*), *2013 46th Hawaii international conference on*. 2013. IEEE.
- Toshiwal, A., et al. Storm@ twitter. in Proceedings of the 2014 ACM SIGMOD *international conference on Management of data*. 2014. ACM.
- Kulkarni, S., et al. Twitter heron: Stream processing at scale. in Proceedings of the *2015 ACM SIGMOD International Conference on Management of Data*. 2015. ACM
- Malewicz, G., et al. Pregel: a system for large-scale graph processing. in Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. 2010. ACM.

