**DataFlow**

Windowing: where in event time data are grouped

Triggering: when in processing time groups are emitted

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

**Naiad**

No support for stale update

Checkpointing needs pause all workers and flush message queues

A close up of a newspaper

Description automatically generated

A close up of a newspaper

Description automatically generated

**Spark Streaming**

Unify Batch & Interactive Processing

Use speculative execution for straggler

Lineage for parallel workers recovery

Connect to new master when old fails

Fine to lose some running tasks on reconnect

Based on immutable RDD

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A close up of a newspaper

Description automatically generated

A close up of a logo

Description automatically generated**PowerGraph**

Fine-grained parallelism

Not all vertices are activated

Vertex-cuts more suitable with power rule

Quickly shatter a graph by cutting a small fraction of the very high degree vertices

Checkpoint at the end of super-step for sync

No straggler mitigation

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated**GraphX**

Immutable data

Overhead of distribution is high

Join strategy: send vertices to the edge site

Automatic Join Elimination

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A close up of text on a white background

Description automatically generated

A screenshot of a cell phone

Description automatically generated**Weld**

Hard to debug and scale to more libraries

Fully deterministic, cannot express asnyc algo

Falut tolerance restricted to a single machine

Primitives: scalar, structure, vector, dict

Builder: vecbuilder, merger, etc.

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated**PyWren**

Breakdown computation into stateless funcs

Schedule on serverless containers

Use external storage for state management

No need for explicit fault tolerance

More elastic. Good for streaming workload

Overhead from fetching state. Cold start

Network bandwidth is pretty good

Blob storages more suitable for large file

Scheduler might be the bottleneck

Information gap between user and server

No sharing inventive

A screenshot of a cell phone

Description automatically generated

A close up of a map

Description automatically generated

A screenshot of a cell phone

Description automatically generated**TPU**

No features to improve the average case

No caches, branch pred, out-of-order execution

Simple design with MACs, Unified Buffer

TPU is not energy proportional

systolic execution to save energy by reducing reads and writes of the Unified Buffer

A screenshot of text

Description automatically generated

A screenshot of a cell phone

Description automatically generated A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated