



University of
South Australia

An overview of the Apache ecosystem

1



2



- Maintains
- configuration information
 - file naming
 - distributed synchronization

Makes applications more robust to changes and errors

APACHE ZooKeeper™



Data warehouse software to
facilitate reading, writing and
managing large datasets residing
in distributed storage using SQL.



- Platform for analyzing large datasets
- Uses its own language called Pig Latin
- Pig Latin is really easy to write!
- Automatically optimizes code execution

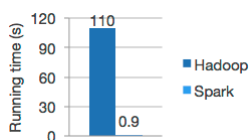
Apache Pig

A unified analytics engine for large-scale data processing.

A rich set of higher-level tools such as:

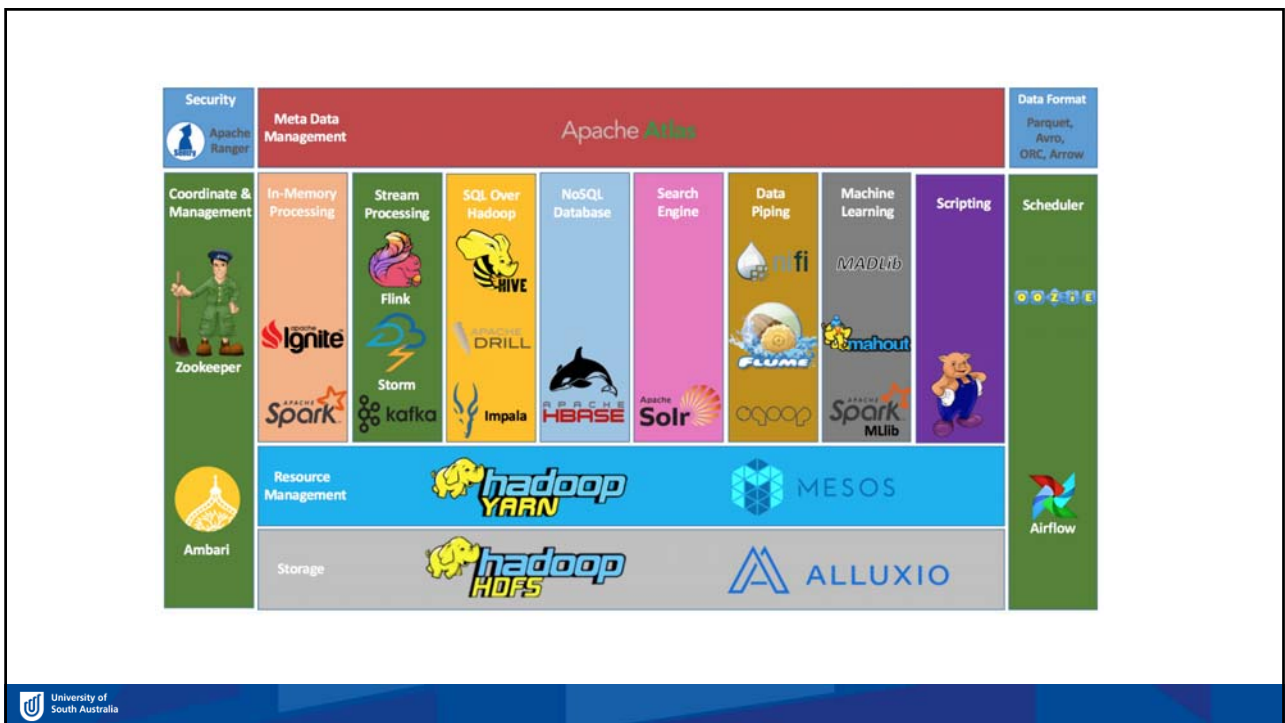
- Spark SQL (structured data)
- Mlib (Machine learning)
- GraphX (Graph Processing)

It's really fast!





7



8

WARNING

This material has been reproduced and communicated to you by or on behalf of the **University of South Australia** in accordance with section 113P of the *Copyright Act 1968* (**Act**).

The material in this communication may be subject to copyright under the Act. Any further reproduction or communication of this material by you may be the subject of copyright protection under the Act.

Do not remove this notice